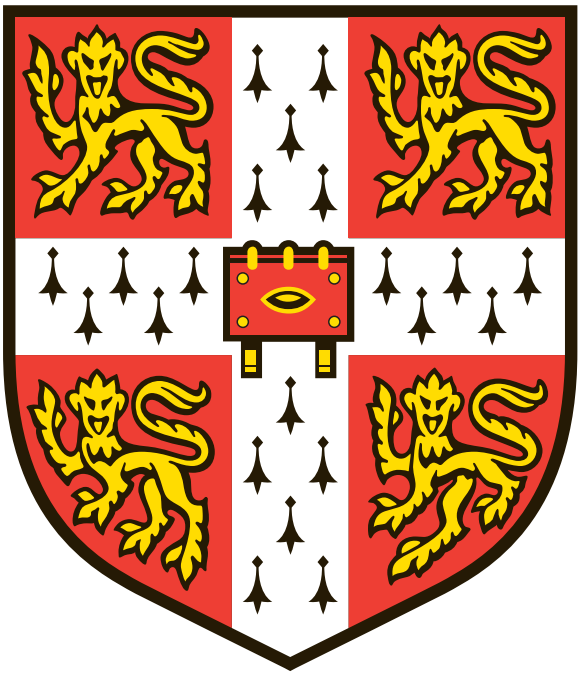
# Application of CRISPR/Cas9 screening to study cancer drivers and to identify novel cancer vulnerabilities



**Gemma Turner**

Wellcome Sanger Institute

University of Cambridge

This dissertation is submitted for the degree of

*Doctor of Philosophy*

Churchill College

September 2019

I would like to dedicate this thesis to maw, paw and Danielle;
I hope I can make you all as proud of me as I am of you.

# Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

This dissertation does not exceed 60,000 words in length, exclusive of tables, footnotes, bibliography, and appendices.

**Gemma Turner**

September 2019

# Acknowledgements

When I started my PhD four years ago, I underestimated how challenging it would be, both scientifically and personally. I'd like to take this opportunity to thank the many people who have helped me get to this stage.

I would firstly like to thank my supervisor, Dave, who has always found time for me despite his crazy schedule. I really appreciate him giving me this opportunity and putting his trust in me. Thank you for your advice, patience, understanding and for just caring.

I would also like to thank Luca Crepaldi, Leo Parts, Helen Davies and Peter Campbell, who all gave me their time and support during my rotation projects in my first year at Sanger. I learned a lot from them, including many skills that were invaluable throughout my PhD.

This project would not have been possible without the generosity of various people who have given up their time to teach me or help me with my experiments. Thank you to the Gene Editing facility for producing my iPSC lines. Thank you to the members of CGaP, especially Rebecca and Verity, for making the project run so smoothly and for the many, many hours spent in cell culture. Thanks to Vivek Iyer, Francesco Iorio and Emanuel Gonçalves for helping me with my data analysis. Thanks also to Fiona Behan and Kosuke Yusa for providing CRISPR reagents and advice.

I was lucky to have two brilliant scientists as mentors during the first few years of my PhD. Marco taught me many things and working with him has undoubtedly made me a better scientist. I really appreciate the time, support and friendship he has offered me since the day I joined Sanger. I'm very grateful to Clara for sharing her knowledge and experience with me, and for being so friendly and welcoming. She put a lot of work into this project before I started and it would've been a much steeper, more painful learning curve without her supervision.

Finding the motivation to go to the lab every day would've been near impossible if I hadn't been surrounded by such lovely people. Louise, thank you for letting me be involved in your project and always providing kind words and light relief with your hilarious stories. Your dedication and passion for science is inspiring. Carmen, we come as a pair and I'm so grateful that I got to spend the past 3 years with you sat behind me! Thank you for the life chats, snack breaks, gym trips, many laughs and endless encouragement; you've helped me more than you'll ever realise. Vicky, you're the most selfless person I know and I'll be forever grateful to you for giving up your weekends to help me. I'm going to miss our chats and your friendly smile! Aravind, thanks for just always being there for me and for being a great flatmate and passenger

- it would've been a lot less fun without you. Annie, thanks for the constant supply of amazing sweet treats - they powered me through many a tough time! Thanks also to Agnes, Saskia, Andrea and the rest of team 113 for being such great friends and colleagues.

Arian, thank you for always understanding what goes on in my head. It means so much that you're always there (with an appropriate gif), no matter how long it's been. Katie, thank you for being by my side every day for the past 6 years, through the highs and many lows. I really would be lost without you and I'm proud that we got through this challenge together.

Finally, I'd like to thank my family for being my life-long cheerleaders. Especially my mum, dad and Danielle - I can't put into words how grateful I am for everything. I wouldn't be the person I am today without you. Thank you for never letting me give up. Your constant love and support mean the world and I know you'll get me through whatever life brings. Not forgetting my big bear, Rox, for always cheering me up.

*"Whit's fur ye'll no go by ye."*

# Abstract

The development of targeted therapies has had a significant impact on cancer survival rates. However, targeting cancers that are driven by loss of tumour suppressor genes remains a major challenge. One promising approach to treat these cancers is the exploitation of synthetic lethal interactions. Synthetic lethality describes an interaction between two genes, where loss of one gene alone does not affect viability but loss of both genes induces cell death. Inhibiting the synthetic lethal partner of a tumour suppressor gene should specifically kill tumour cells, and so these represent potential therapeutic targets. However, very few synthetic lethal interactions have been well-established.

The aim of this project was to systematically screen for synthetic lethal partners of known tumour suppressor genes. To do so, isogenic human induced pluripotent stem cell lines were generated, each carrying a loss-of-function mutation in a single tumour suppressor gene. These cells have a normal genetic background, thus making it simpler to accurately identify interactions. CRISPR/Cas9 technology was applied as it allows for large-scale, unbiased screening of genetic interactions. A genome-wide guide RNA library was prepared and implemented for knockout screening in the isogenic cell line panel. Analysis was performed to identify genes that were specifically essential for cell fitness/survival in the mutant lines. Particular focus was placed on four tumour suppressor genes that encode subunits of the PBAF/BAF complexes. Approximately 20% of human cancers harbour mutations in subunits of these complexes, so identifying dependencies associated with these could have broad therapeutic potential. Candidate synthetic lethal interactions with these genes were investigated using low-throughput assays in the stem cells and in a cancer cell line. The data obtained suggests that screening in stem cells produces highly variable results. Although potential vulnerabilities associated with all of the tumour suppressor genes were identified, further work is required to validate these and to assess the quality of the results.

In addition to genome editing, CRISPR/Cas9 has been adapted as a tool for controlling gene regulation. In collaboration with Dr Louise van der Weyden, I applied this technology to address another challenging area of cancer biology. Metastasis is the main cause of cancer mortality, yet we still have a poor understanding of the genes that control this process. Considering this, an *in vivo* CRISPR activation screen was performed to identify novel drivers of metastatic colonisation. A mouse melanoma cell line was transduced *in vitro* with a library designed to up-regulate expression of membrane proteins, which represent ideal drug targets.

These cells were then used in an *in vivo* experimental metastasis assay. Enrichment of guide RNAs in the lungs was assessed to identify genes that increased pulmonary metastatic colonisation when activated. Candidate genes were selected using three analysis strategies, and hits from each were tested. Several genes were successfully validated using the experimental metastasis assay. The most robust hit was studied further to explore its potential as a therapeutic target.

Collectively, the work described in this thesis demonstrates how CRISPR/Cas9 screening can be applied in different model systems to study genes that drive cancer and to explore novel therapeutic strategies.

# Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| AUC | area under the curve |
| BAF | BRG-/BRM-associated factor |
| BAGEL | Bayesian Analysis of Gene EssentiaLity |
| BF | Bayes factor |
| BFP | blue fluorescent protein |
| bp | basepair |
| Cas | CRISPR associated protein |
| CGaP | Cellular Genotyping and Phenotyping facility |
| CRISPR | Clustered Regularly Interspaced Palindromic Repeats |
| CRISPRa | CRISPR activation |
| CRISPRi | CRISPR interference |
| crRNA | CRISPR RNA |
| dCas9 | deactivated/dead Cas9 |
| DSB | double-strand break |
| FACS | fluorescence-activated cell sorting |
| FBS | fetal bovine serum |
| FDR | false discovery rate |
| gDNA | genomic DNA |
| GFP | green fluorescent protein |
| gRNA | guide RNA |
| GSH | glutathione |
| HDR | homology-directed repair |
| hESCs | human embryonic stem cells |
| indel | insertion/deletion |
| iPSCs/iPS cells | induced pluripotent stem cells |
| IRES | internal ribosomal entry site |
| KO | knockout |
| LB | Luria broth |
| LC-MS | liquid chromatography mass spectrometry |
| LOF | loss-of-function |
| M-FISH | multiplex fluorescence *in situ* hybridisation |

| | |
|---|---|
| MAGeCK | Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout |
| MEFs | mouse embryonic fibroblasts |
| MOI | multiplicity of infection |
| MRT | malignant rhabdoid tumour |
| MSI | microsatellite instability |
| MUT | mutant |
| ncBAF | non-canonical BAF |
| NHEJ | non-homologous end joining |
| NTC | non-targeting control |
| OCCC | ovarian clear cell carcinoma |
| PAM | protospacer-adjacent motif |
| PBAF | polybromo-associated BAF |
| PBS | phosphate-buffered Saline |
| PFA | paraformaldehyde |
| PrRc | precision recall |
| RNAi | RNA interference |
| ROC | receiver operator curve |
| RRA | robust rank aggregation |
| shRNA | short hairpin RNA |
| siRNA | short interfering RNA |
| SLI | synthetic lethal interaction |
| TBS-T | Tris-Buffered Saline supplemented with 0.1% Tween-20 |
| TCGA | The Cancer Genome Atlas |
| tracrRNA | trans-activating crRNA |
| TSG | tumour suppressor gene |
| WSI | Wellcome Sanger Institute |
| WT | wildtype |

# Chapter 1

# Introduction

## 1.1 Targeted therapies for cancer treatment

In the UK, it is estimated that 1 in 2 people will be diagnosed with cancer in their lifetime.[1] Survival rates are still as low as 1% for some cancer types, but massive improvements have been made over the past 40 years.[2] These increases in survival are largely due to advances in early detection and treatment, particularly targeted therapies. Traditional chemotherapeutic agents act by interfering with cell division and inducing cell death. They do not specifically target the tumour but have a greater impact on cancer cells as they divide more rapidly than most normal cells. Targeting of fast-dividing normal cells, such as hair follicles and cells in the stomach, can lead to negative side effects (e.g. hair loss and nausea). Chemotherapies are still widely used and have various mechanisms of action. Alkylating agents bind to and modify DNA, inducing damage and inhibiting division (e.g. cisplatin forms crosslinks between and within DNA strands[3]). Topoisomerase inhibitors inhibit enzymes that are involved in the separation of DNA strands to allow replication (e.g. topotecan binds to topoisomerase I and induces double-strand breaks [DSBs][4]). Antimetabolites compete with, replace, or inhibit the function of metabolites required for DNA synthesis (e.g. fluorouracil is a an analogue of uracil which is incorporated into DNA/RNA and causes inhibition of a nucleotide synthetic enzyme, thymidylate synthase[5]). Unlike these chemotherapies, targeted therapies are designed to specifically eliminate cancer cells by targeting molecules that are required for tumour growth and progression.[6] Targeted therapies can be broadly categorised as either small molecules or monoclonal antibodies.[7] Small molecules can penetrate the cell membrane and act on intracellular targets. Monoclonal antibodies bind to tumour-specific antigens presented on the cell surface. These therapies have many different mechanisms of tumour cell killing, targeting the various hallmarks of cancer.[8]

Signal transduction pathways are often hyperactivated in tumour cells and inhibitors can be designed to block these (e.g. Cetuximab, an EGFR inhibitor[9]). Hormone therapies block the production of hormones or interfere with their function to impede the growth of

hormone-dependent tumours (e.g. Trastuzumab binds to HER2[10]). Tumour cells develop mechanisms to evade cell death and so drugs can be designed to induce apoptosis (e.g. Venetoclax, a BCL2 inhibitor[11]). Angiogenesis inhibitors block the growth of new blood vessels, which are required by tumours to gain oxygen and nutrients (e.g. Bevacizumab, a VEGF inhibitor[12]). Gene expression is often altered significantly in tumour cells, and drugs can be designed to target expression modulators (e.g. Vorinostat, a histone deacetylase inhibitor[13]). A major focus of current research is immunotherapy, which functions by activating immune-mediated killing of tumour cells (e.g. Ipilimumab, a CTLA-4 inhibitor[14,15]).

Typically, targeted drugs are approved for treatment of a specific tumour type, sometimes with an additional indication regarding mutation status or hormone expression. However, 2018 saw the first initial tissue-agnostic approval based on genetic mutation; Larotrectinib is indicated for any solid tumour carrying an *NTRK* gene fusion.[16]

Design of these therapies requires an understanding of the genetic and molecular basis of tumour development. At a genetic level, cancer drivers can be broadly classified as oncogenes or tumour suppressor genes (TSGs). Proto-oncogenes encode proteins that function to stimulate cell division and differentiation, and inhibit cell death.[17] Activation (i.e. gain-of-function) of these genes transforms them into oncogenes, which drive abnormal cell proliferation and lead to tumour development. Examples of oncogenes include *BRAF*[18] and *H-/N-/K-RAS.*[19] TSGs encode proteins that function to control cell growth; they inhibit progression through the cell cycle, repair DNA errors, and induce apoptosis.[20] Inactivation (i.e. loss-of-function, LOF) of these genes can result in tumour development. Examples of TSGs include *TP53*[21] and *RB1.*[22]

## 1.2 Targeting tumour suppressor genes

Despite the fact that TSGs are frequently inactivated by mutation, deletion, or silencing in many cancers, the majority of targeted therapies are oncogene inhibitors.[23] This is likely because it is inherently more challenging to restore the normal function of a gene than it is to inhibit it. However, several strategies have emerged which hold great promise for improving therapeutic targeting of TSG-driven cancers.

The most conceptually simple strategy is to reintroduce a functional copy of the TSG using gene therapy techniques. In practise, this has proven to be challenging due to inefficiencies in delivery and maintenance of wildtype (WT) protein expression, in addition to issues with safety.[24] An alternative approach is to target regulators of the inactivated TSG. One of the functions of the first identified tumour suppressor, RB1, is to inhibit DNA replication

by binding to and repressing E2F transcription factors.[25] This can be reversed by phosphorylation of RB1 by cyclin-CDK complexes, which often occurs aberrantly in tumour cells.[26] It has been shown that inhibition of oncogenic kinases can indirectly reactivate RB1 and restore its repressive functions.[27,28] Similarly, targeting of epigenetic modulators that inactivate TSGs has shown potential. Many genes are silenced in tumours due to hypermethylation or hypoacetylation.[29,30] Inhibitors of DNA methyltransferases and histone deacetylases are available and can reverse this silencing. However, currently these are not specific and induce global changes rather than targeted reactivation of TSGs, which can also affect non-malignant cells.[24] Another option is to move downstream and inhibit pathways that have been activated as a result of TSG function being lost. For example, inactivation of *PTEN* leads to hyperactivation of the PI3K/AKT/mTOR pathway and so downstream inhibition of this signalling cascade may reverse the effects.[31] However, pathways like this are complex and interference would likely alter other associated networks.

The most promising strategy to date has been exploitation of vulnerabilities induced by TSG inactivation, or so-called synthetic lethal interactions (SLIs). Synthetic lethality refers to an interaction between two genes, where cells can survive with loss of either gene but loss of both induces cell death.[32] When a TSG is lost, tumour cells can become dependent on the function of a second protein for survival (Fig. 1.1).[33] Pharmacological inhibition of this second protein would be lethal to tumour cells but, in theory, should have little impact on non-malignant cells as they have maintained function of the tumour suppressor. The most clinically advanced example of this strategy is inhibition of PARP in *BRCA1/2*-mutant cancers.[34]

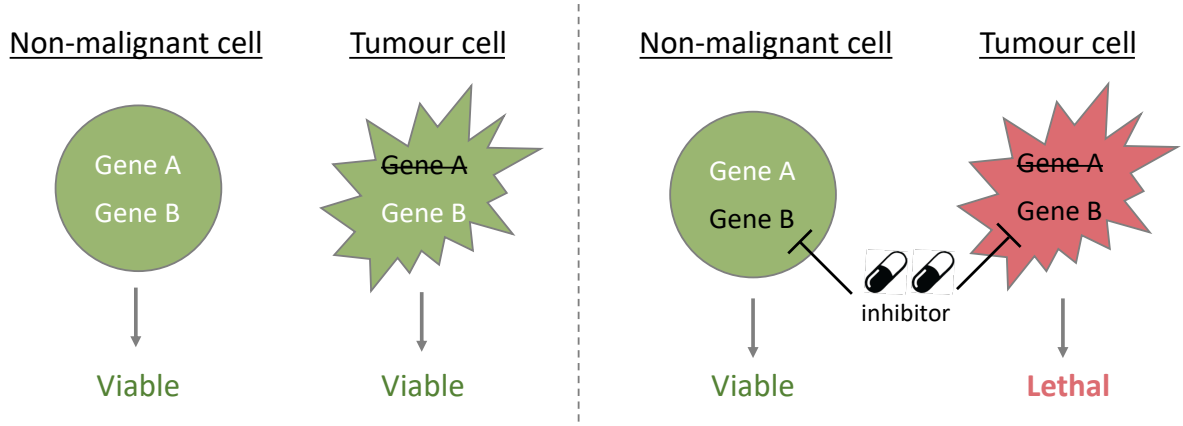



**Figure 1.1. Exploiting synthetic lethality for cancer treatment.** Tumour cells often lose the function of a TSG (gene A), but loss of this gene alone does not affect viability. If a synthetic lethal partner of gene A is known (gene B), then pharmacological inhibition of this gene should induce death specifically in the tumour cells. Inhibition of gene B should not affect viability in non-malignant cells as the function of gene A is maintained.

This concept has been expanded to incorporate similar genetic interactions. Synthetic dosage lethality describes the situation where overexpression or overactivation of one gene induces a vulnerability to loss of another gene.[35,36] These interactions could be exploited in tumours that are driven by oncogenes, such as *KRAS*-mutant lung tumours, where pharmacological inhibition of the oncoprotein is challenging.[37] Deletion of TSGs is often accompanied by loss of other genes that are in close proximity in the genome.[38] Whilst these are considered to be 'passenger' genes, as their loss usually has no clear role in tumour progression, they can confer specific vulnerabilities that could be targeted by drugs.[38,39] This concept is referred to as collateral sensitivity/lethality. This was first demonstrated in glioma cells, where loss of a passenger gene *ENO1* (commonly deleted in glioblastomas) induced a specific sensitivity to inhibition of ENO2.[38] These vulnerabilities offer the potential to develop novel therapies for cancers that do not respond to standard treatment and/or are 'undruggable' using current targeted approaches.

### 1.2.1 DNA damage response pathways

One of the key areas of interest in developing synthetic lethal therapies is targeting the various DNA damage response (DDR) pathways. Cells have evolved a range of mechanisms to detect and repair different types of DNA damage, induced by both endogenous and environmental factors. Interference with these mechanisms can cause mutations and genomic aberrations that are associated with many human diseases, including cancer.[40]

Mismatch repair (MMR) is used to detect and repair mismatches and insertions/deletions (indels) that occur during DNA replication (as reviewed by Li, 2008).[41] The presence of these errors is initially detected by MutS complexes (MSH2-MSH6 and MSH2-MSH3), which then recruit MutL complexes (MLH1-PMS2, MLH1-PMS1, MLH1-MLH3). A single strand incision is made and EXO1 nuclease, polymerases δ and ε, and ligase I act to repair the DNA error. Other protein components involved include PCNA, RFC, and RPA.

Base excision repair (BER) occurs when DNA glycosylases detect and remove bases that have been damaged by oxidation, deamination or alkylation.[42] Repair is undertaken by APE1 endonuclease, polymerases β, δ and ε, and ligase I or III. Single-strand break repair (SSBR) has many overlapping features with BER, as single-strand breaks occur indirectly as a result of this process. However, when the break is induced directly, other factors are involved in SSBR. Breaks can be detected by PARP1 binding and activation. PARP1 functions by adding poly(ADP-ribose) chains to itself and other proteins. It recruits and stabilises the complex of proteins required for repair. One of the key proteins involved is XRCC1, which acts as a molecular scaffold to stabilise and activate various enzymes involved in SSBR.

Nucleotide excision repair (NER) is used to repair damage that disrupts the DNA helical structure.[43] There are two sub-pathways: transcription-coupled NER acts on lesions that block transcription and global-genome NER surveys the whole genome. In transcription-coupled NER, CSA, and CSB displace stalled RNA polymerase to allow for repair. In global-genome NER, the XPC-hHR23B complex screens the genome for disrupted basepairing. In both pathways, the helicase components of transcription factor TFIIH (XPB and XPD) unwind ~30 basepairs of DNA around the damage. RPA stabilises the intermediate, then endonucleases (XPG and ERCC1/XPF), DNA polymerases, and ligase I act to repair the damage.

Non-homologous end-joining (NHEJ) is one of the main mechanisms used to repair DSBs.[44] The Ku protein recognises DSBs and activates the protein kinase DNA-PKcs. This leads to recruitment of end-processing enzymes, polymerases, and ligase IV. NHEJ can operate at any cell cycle phase. It is an error-prone process as repair often results in small indels, which can be deleterious if they cause a frameshift.

Homologous recombination (HR) is another mechanism used in response to DSBs and it is also used to repair stalled replication forks and inter-strand DNA cross-links.[45] HR generally only occurs in S and G2 phase as the sister chromatid sequence is used as a homologous template for repair. In HR, proteins such as the MRE11-RAD50-NBS1 complex initiate ssDNA generation. BRCA1 is a key player in HR, where it is involved in 5'-end resection of the DSB. BRCA1 also interacts with BRCA2 and PALB2 to recruit the

recombinase RAD51. This is important for invasion of the undamaged homologous template. Polymerases, nucleases, helicases, and ligases act to repair the damage using the template sequence. Unlike NHEJ, HR results in a faithful repair of the DNA to its original sequence.

The gold standard example of synthetic lethality involves BRCA1/2 and PARP, which are major DDR components.[46,47] When PARP is inhibited, SSBs cannot be repaired and this can lead to stalled replication forks and DSB induction. In normal cells, HR would be employed to repair these. However, cells lacking BRCA1/2 have defective HR and are therefore more sensitive to PARP inhibition. Signalling pathways mediated by the kinases ATM and ATR can also be involved in the repair of DSBs and ssDNA, respectively. Identifying synthetic lethal partners of genes such as these important players is a key focus in the field currently.

## 1.3 Screening for synthetic lethal interactions

Hartwell *et al.* first proposed the concept of exploiting synthetic lethality to develop cancer therapeutics in 1997.[33] Despite two decades of research, the only clinical success has been in exploiting the interaction between BRCA/PARP. Four PARP inhibitors have been FDA-approved for the treatment of *BRCA*-mutant cancers; three of these are indicated for ovarian cancer and two for breast cancer.[48-52] For many TSGs, synthetic lethality has not been thoroughly explored. Since this concept was first introduced, a range of methods have been developed to screen for genetic interactions (Table 1.1). The functional genomics tools we currently have available make it more feasible to systematically interrogate synthetic lethality on a large scale.

**Table 1.1. Advantages and disadvantages of approaches to screen for synthetic lethal interactions.**

| Screening method | Advantages | Disadvantages |
|---|---|---|
| **Yeast screen** | Small genome allows testing of a large interaction space | Divergence of gene functions/interactions in humans |
| | Easy to genetically manipulate | |
| **Chemical screen** | Easier to translate hits to clinic | Drugs often have multiple targets or unknown targets |
| | | Limited to screening 'druggable' targets |
| | | Must be performed in plate format |
| **RNAi screen** | Possible to target every gene in the genome | More off-target effects than CRISPR |
| | Incomplete knockdown may reflect drug inhibition more than complete knockout | Difficult to achieve complete knockdown |
| | Does not induce DNA damage | On-target activity of si/shRNAs may be variable |
| **CRISPR screen** | Can allow for complete knockout of gene function | Off-target effects |
| | Can be used to target non-coding regions, allowing gene inhibition and activation | On-target activity of gRNAs may be variable, particularly in amplified genomic regions |
| | | Complete knockout may not reflect effects of drug inhibition |
| | | Induces DNA damage which is problematic for screening normal cell lines |
| **Computational analysis** | Combines data from multiple sources (e.g. sequencing, expression data, experimental screen data) to provide more support to findings | Additional experiments are still required for validation |

### 1.3.1 Screening for synthetic lethality in yeast

Many of the first screens for SLIs were performed in the budding yeast, *Saccharomyces cerevisiae.* Mapping of genetic interaction networks has advanced more in *S. cerevisiae* than in any other organism, due to the fact that its genome is small (12 megabases)[53] and relatively easy to manipulate.[54] The development of Synthetic Genetic Array analysis[55] and similar methodologies (e.g. dSLAM[56,57] and GIM[58]) has allowed for high-throughput screening of synthetic lethality in yeast. The basic concept is to introduce a mutation in a gene of interest into a set of single mutant strains (carrying mutations for ~4800 nonessential genes) that are each tagged with a unique DNA barcode. This can be done by mating a haploid single mutant pool with a different haploid strain that is mutant for the gene of interest,[58] or by introducing the mutation of interest into a pool of diploid heterozygous mutant strains.[56,57] These screens rely on competitive growth between the resulting double mutants. The relative abundance of each double mutant is analysed by a barcode microarray. If an SLI exists between a pair of genes, growth of the double mutant for these genes should be impaired and thus would be depleted in the population. One large study involved genome-wide screening in 132 mutant strains using SGA technology.[59] Approximately 4000 interactions were identified, involving ~1000 genes. They found that generally, genes that had a negative (synthetic lethal) interaction were less likely to physically interact at the protein level. Many associations were identified between functionally-distinct pathways. For example, members of the sister chromatid cohesion complex were found to be synthetic lethal with genes in spindle checkpoint pathways and various pathways involved in DNA repair, damage, and replication.

A major limitation of using yeast as a model is that not all genes are evolutionarily conserved, and so only a fraction of the findings can be mapped to the human genome.[60] Some yeast genes have more than one ortholog in humans, and often the functions and interactions of the encoded proteins have diverged.[60] This can make it difficult to accurately translate findings between the species. Despite this, it has been shown that some SLIs are conserved between yeast and humans (e.g. *CHEK1/2* and *WEE1*; *RAD17* and *TOP1*[61]). Therefore, these screens can be informative for human studies, and the simplicity with which a very large interaction space can be tested makes this a valuable model system.

### 1.3.2 Screening for synthetic lethality in human cancer cell lines

Whilst we can learn a lot from studying model organisms, human cancer cell lines offer a more clinically relevant experimental system to identify SLIs. Two general approaches are commonly used: screening a cell line panel or screening an isogenic cell line pair.[62] Hits from screens in a large panel of cell lines can be cross-referenced with the mutational status of the lines (Fig. 1.2a). If a vulnerability is consistently identified in cell lines that have a mutant TSG, but not in lines that are WT for this gene, this indicates that an SLI may exist. Cell lines must be carefully selected to ensure that mutant and WT genotypes are both well-represented. To obtain an isogenic pair, a WT parental cell line can be engineered to carry a LOF mutation in a TSG, or a parental cell line that is already mutant can be engineered to express the WT protein (Fig. 1.2b). Screening of both the parental and derivative can be compared to identify vulnerabilities that occur specifically when the gene of interest is lost.

There are advantages and disadvantages of both strategies. SLIs identified in isogenic pairs may be specific to that cell line, whereas screening in a diverse panel would ensure that hits are relevant across different backgrounds. Multiple isogenic pairs could be screened to remove this context-specificity. However, having an identical genetic background can be beneficial as we can confidently infer that any observed effect is due to an interaction between the two disrupted genes. In contrast, cell lines in a panel may share other genetic aberrations in addition to the one being studied, possibly confounding the results and making it difficult to deduce which genes interact. Another caveat to consider is that engineering a mutation in a cell line may not accurately reflect the true context in which this genetic change would occur, compared to lines that have acquired it naturally. Despite these differences and limitations, both approaches have been applied successfully to identify SLIs using various screening technologies.

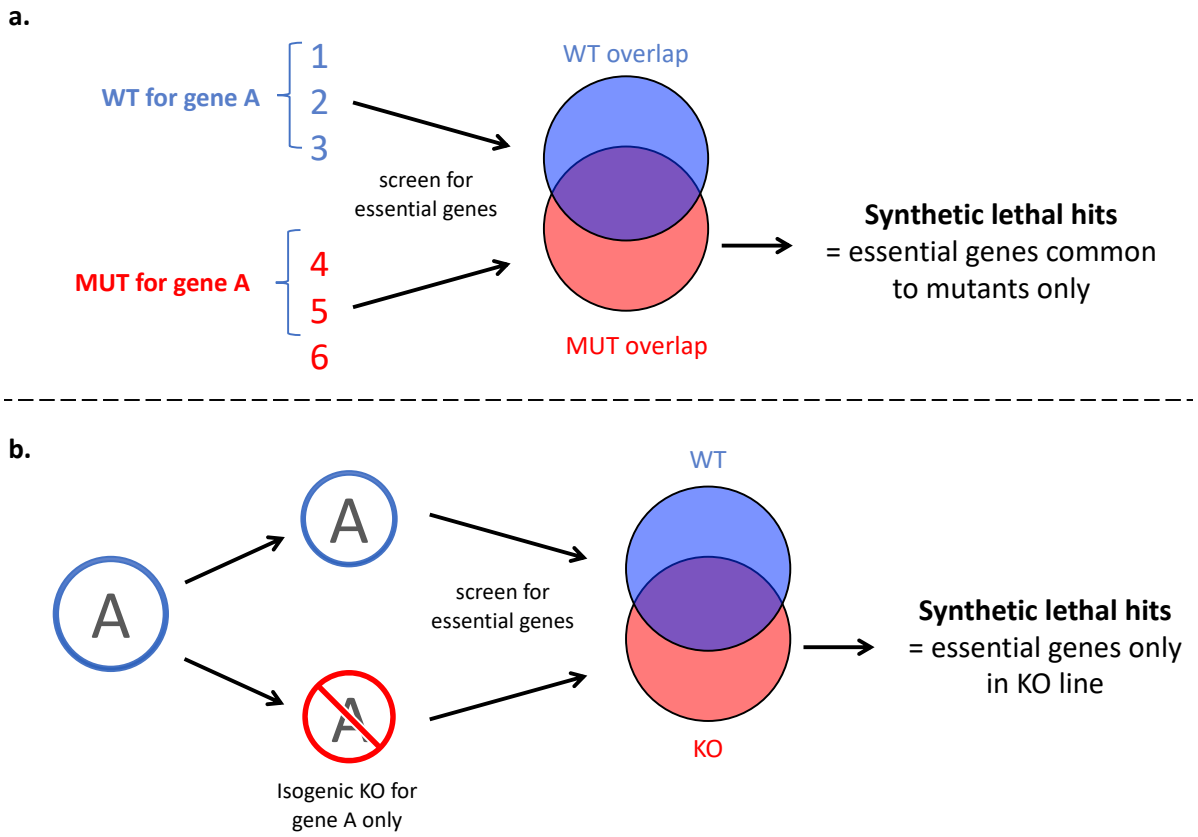



**Figure 1.2. Approaches to synthetic lethality screening in human cell lines. a)** Cell lines can be grouped into those that are WT or mutant (MUT) for a gene of interest. Screening is performed to identify genes that are essential for cell fitness/survival in each line. The essential genes that overlap in all WT lines or overlap in all MUT lines are compared. Synthetic lethal candidates are those that are specifically essential in the MUT lines; **b)** An isogenic derivative of a cell line can be engineered by creating a single knockout (KO) in gene A (alternatively, an existing mutation could be corrected). Both the parental and KO lines are screened and the results are compared. The synthetic lethal candidates are those that are specifically essential in the KO line.

### 1.3.3 Chemical screening to identify synthetic lethal interactions

For some time, the only way to screen for SLIs in human cell lines was to use chemical compound libraries.[63,64] Chemical screens have been used to identify specific vulnerabilities in both isogenic cell line pairs and panels of cell lines with common genetic features. For example, screening in an isogenic renal cancer cell line pair revealed that *VHL*-mutant cells were specifically vulnerable to a small molecule inhibitor, STF-62247.[65] A high-throughput screen of colorectal and gastric cancer cell lines identified that loss of *MRE11A* and *ATM* are both synthetically lethal with FEN1 inhibition.[66] The use of candidate molecules provides direct clinical relevance to any hits, which is beneficial compared to genetic screens where targets may not be druggable. However, chemical libraries can contain molecules with unknown

targets and those that are annotated often have multiple targets, both by design and due to off-target activity.[67,68] This can make it challenging to identify which targets are responsible for the observed synthetic lethality and to understand the mechanism of action. It could also mask potential SLIs, as targeting multiple proteins may lead to general cytotoxicity, where inhibition of one alone may have induced synthetic lethality.

### 1.3.4 RNAi screening to identify synthetic lethal interactions

The discovery of RNA interference (RNAi)[69] made it feasible to study gene-gene interactions in human cells on a large scale[70] and hence systematically screen for synthetic lethality. This can be done using short interfering RNAs (siRNA) or short hairpin RNAs (shRNA), both of which inhibit protein translation by promoting degradation of specific messenger RNAs.[71] An arrayed format can be used with siRNA and shRNA, where each gene is targeted in a separate well and the desired phenotype can be measured.[72] Alternatively, shRNAs can be labelled with barcodes, pooled, and packaged into a single viral library for transduction into cells.[73] Changes in the relative abundance of the barcodes can be measured after a period of time and used as a proxy for shRNA expression.[73] Depletion of a given shRNA indicates that knockdown of the targeted gene impaired cell fitness. This technology has been widely applied to identify SLIs, most notably in two large-scale studies: ProjectDRIVE[74] and Project Achilles.[75]

Project DRIVE involved shRNA screening of ~8000 genes across 398 cancer cell lines. To overcome issues with off-target effects and limited statistical power, a median of 20 shRNAs per gene was used. The effects of each shRNA on cell viability were assessed after 14 days. The study investigated a range of things, including synthetic lethality. They identified subsets of interactions where synthetic lethality was associated with a pathway, a paralog, or collateral lethality. Reduced expression of a cell death inhibitor, *BCL2L1*, and increased expression of a pro-apoptotic protein, *BIM*, were predictive of sensitivity to knockdown of the anti-apoptotic protein MCL1. Synthetic lethality was also identified in members of separate pathways that share downstream connections e.g. cells with LOF in *APC* were sensitive to knockdown of *CTNNB1*. Synthetic lethal interactions between paralogous genes were also identified, based on both mutation and expression. *ARID1A* mutants were more sensitive to *ARID1B* depletion, and *RPL22* mutants were more sensitive to *RPL22L1* depletion. Low expression of *ARF5* predicted a dependency on *ARF4*. Several cases of collateral lethality were identified, including a sensitivity to knockdown of *PRMT5* in cells that have lost *MTAP*, a gene that is co-deleted with the tumour suppressor *CDKN2A*. Some of these findings overlapped

with results from the Project Achilles shRNA screens. For this project, 501 cancer cell lines were screened using a genome-wide shRNA library. Achilles identified over 80 paralog synthetic lethal interactions. For example, cells with low *FERMT2* expression had a dependency on *FERMT1*, which has a role in integrin and cytoskeleton regulation. *SMARCA2* was identified as being essential in cancer cell lines carrying a mutation in its paralog *SMARCA4,* with both genes acting as core subunits in the SWI/SNF chromatin remodelling complexes. Interestingly, the analysis of the RNAi dataset from Achilles found that most dependencies they identified were predicted by differences in gene expression rather than DNA mutation. This may be because they had a limited number of cell lines representing mutation of any given gene. This bias towards more commonly mutated genes is a major limitation of screening projects like DRIVE and Achilles, as they do not offer the opportunity to identify SLIs with genes that are mutated less frequently in cell lines.

RNAi can also be combined with chemical screening to identify synthetic lethal partners of drug targets.[76] As with drug screens, RNAi also has several limitations. Despite intended targeting of one gene, si/shRNA molecules can have off-target activity which may result in false positive results.[77,78] Combining multiple molecules to target the same gene can reduce the likelihood of this. Additionally, as RNAi functions at a post-transcriptional level, it is difficult to achieve complete KO of a target and often only partial knockdown is achieved.[79] Synthetic lethal hits may be missed as a result, but it may represent the clinical context more accurately as complete inhibition with a drug can be challenging.[80]

## 1.3.5 CRISPR/Cas9 screening to identify synthetic lethal interactions

### 1.3.5.1 CRISPR/Cas9 technology

Whilst RNAi has contributed massively to advances in screening for synthetic lethality, the recent development of CRISPR/Cas9 technology has offered an improved and more versatile approach. Clustered regularly interspaced palindromic repeats (CRISPR) were first identified in *E.coli* in 1987[81], but it was not until 2005 that the function of these loci started to become clear.[82,83] CRISPR/Cas is used as an adaptive immune system by bacteria; they integrate phage DNA as CRISPR loci, allowing them to recognise these foreign bodies and prevent further infection.[82-85] In the years that followed, researchers identified the components and exact mechanism of the CRISPR/Cas9 system.[86] By 2013, CRISPR/Cas9 was adapted for genome editing in human and mouse cells,[87,88] and has since been harnessed by scientists across many

fields. Advances in our understanding and application of this technology have been rapid and constant; it has revolutionised our approach to functional genomics.

A variety of CRISPR/Cas systems have been identified but researchers most commonly use CRISPR/Cas9, the type II-A system employed by *Streptococcus pyogenes.*[89] This has been simplified to a two-component system, requiring Cas9 protein and a guide RNA (gRNA) to cleave DNA in mammalian cells (Fig. 1.3). In the natural system, two RNA molecules (a CRISPR RNA [crRNA][90] and a trans-activating crRNA [tracrRNA][91]) are transcribed separately and form a duplex that binds to Cas9. A fusion of these can be engineered to produce a single gRNA molecule.[92] The crRNA provides a sequence homologous to the target DNA, and the tracrRNA interacts with Cas9. Cas9 is an endonuclease which creates blunt-end DSBs in the targeted DNA region.[84,93] The protein recognises a protospacer-adjacent motif (PAM) sequence ('NGG') adjacent to the gRNA target, unwinds the DNA and cleaves at a position three basepairs upstream of the PAM.[93] DSBs trigger endogenous repair mechanisms in the cell; the two most prominent are non-homologous end joining (NHEJ) and homology-directed repair (HDR) via homologous recombination.[94] NHEJ is error-prone and introduces indels in the DNA, often leading to frameshift mutations which can cause LOF of a gene.[44] HDR is less efficient but repairs DNA in an error-free manner; this requires a donor template with homology to the regions surrounding the break site.[45] A donor template can be provided simultaneously with the Cas9 and gRNA to enable introduction of specific sequence changes via HDR.[88]

In addition to genome editing, this technology has been adapted for many applications, including gene regulation.[95] Mutation of the RuvC and HNH domains in Cas9 deactivates the nuclease function of the protein, but still allows it to be guided to a target.[96] This derivative, called deactivated or dead Cas9 (dCas9), can be fused to transcriptional activation or repression domains.[97] Targeting of these to gene promoters and enhancers can allow for overexpression (CRISPR activation, CRISPRa) or silencing (CRISPR interference, CRISPRi), respectively.[97] Researchers have also generated Cas9 variants that recognise different PAM sequences, increasing the targeting capacity.[98,99] Other CRISPR/Cas systems have different functions and investigation of these is widening the applications of this technology even further.[100]

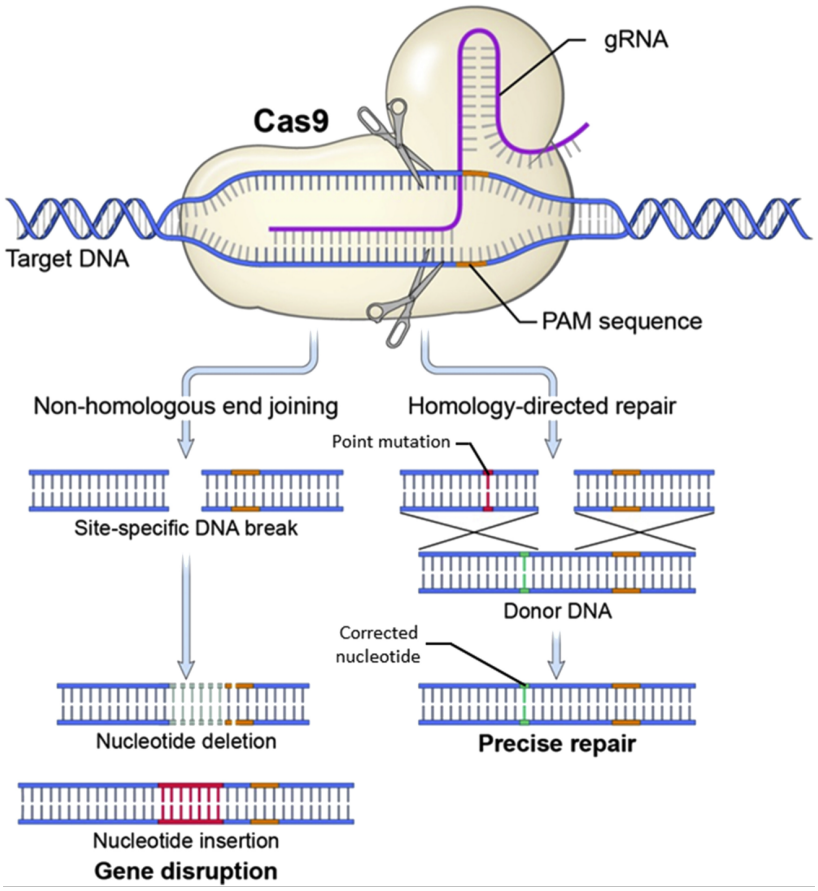



**Figure 1.3. CRISPR/Cas9 genome editing in mammalian cells.** Cas9 forms a complex with a gRNA and is guided to a specific target DNA region. The gRNA binds to a homologous 20 nucleotide DNA sequence positioned immediately downstream of an 'NGG' PAM sequence. Cas9 cleaves the DNA producing a DSB. The cell then activates endogenous repair pathways. Non-homologous end joining can result in insertions and/or deletions which disrupt the gene. Alternatively, in the presence of a donor template, homology-directed repair can replace or insert a specific sequence. Figure taken from [101].

### 1.3.5.2 Pooled CRISPR/Cas9 screening

One of the most powerful applications of CRISPR/Cas9 is genome-scale, high-throughput screening in mammalian cells. CRISPRa and CRISPRi technologies have been applied in screens (as reviewed by Kampmann, 2018),[102] but design of gRNAs for these is more challenging as regulatory regions in the genome are not as well annotated as protein-coding regions. Use of the original CRISPR/Cas9 system to screen for the effects of gene knockout is more well-established. This can be done in an arrayed format with a single gene targeted per well,[103] but is most commonly used in a pooled format where gRNAs targeting all genes are combined.

A pooled library of gRNAs targeting all genes of interest can be designed and packaged into lentiviral vectors, then transduced into a single population (Fig. 1.4).[104,105] Cas9 can be

introduced simultaneously with the gRNA library[104,106,107] or cells can be engineered to stably express Cas9 prior to screening.[105,107-111] The transduced population must then be maintained for a period of time to allow for proliferation and for the gRNAs to induce a phenotype. Screens most commonly focus on cell fitness/survival as a primary phenotype. At the endpoint, the abundance of each gRNA can be compared to the abundance in the initial library.[104,105] If loss of a gene impairs cell fitness, cells carrying gRNAs that target that gene should be under-represented in the final population and hence have a relative depletion. If loss of the targeted gene confers a growth advantage, a relative enrichment of these gRNAs would be observed.

This technology has made it relatively simple to screen every protein-coding gene in the genome in a single experiment. Various analyses can be used to interpret CRISPR/Cas9 screen data; these are constantly evolving and improving as more data becomes available[112-118]. Each one has a different approach, but the primary aim is generally to determine the gene-level significance of any observed changes in gRNA abundance. Two of the most widely used packages, Bayesian Analysis of Gene EssentiaLity (BAGEL)[114] and Model-based Analysis of Genome-wide CRISPR-Cas9 Knockout Gene (MAGeCK),[116] will be discussed further in Chapter 3.

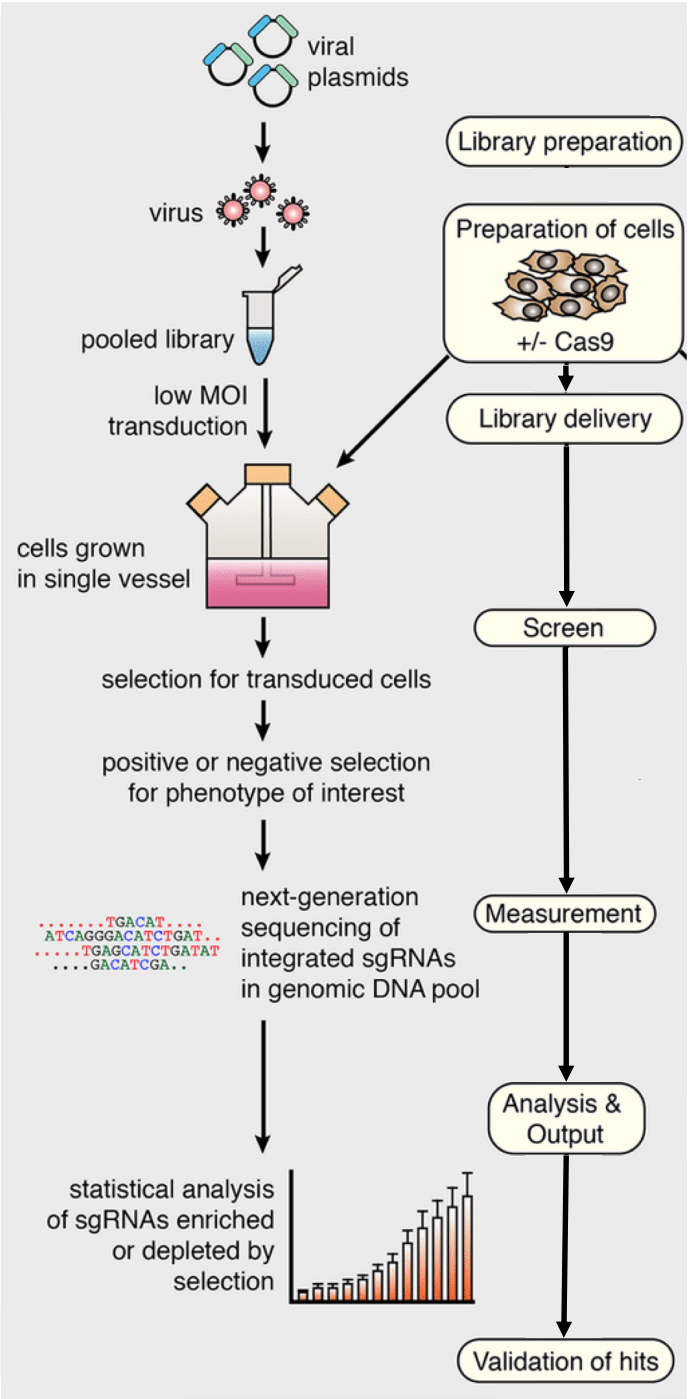



**Figure 1.4. Strategy for pooled CRISPR/Cas9 screening.** A viral library containing a pool of gRNAs is prepared. If Cas9 is not encoded by the library backbone, cells must be engineered to express Cas9. Cells are transduced with the pooled gRNA virus at a low multiplicity of infection (MOI). Library backbones often have an antibiotic resistance marker to allow for selection of transduced cells. Selection for the phenotype of interest is then performed; if the phenotype is cell fitness/survival, cells are simply maintained in culture. After a period of time, genomic DNA is extracted, then PCR amplification and next-generation sequencing of the integrated gRNAs is performed. Analysis is carried out to identify depletion or enrichment of gRNAs, and hits are validated. Figure taken from [119].

Whilst it has many advantages, there are challenges associated with CRISPR/Cas9 screening. The most prominent are related to gRNA design, in terms of both on-target and off-target activity. Many algorithms have been developed to design gRNAs with optimal efficacy, considering factors such as position-specific nucleotides and GC content (as reviewed by Cui *et al.*, 2018).[120] However, sequence-independent factors such as chromatin accessibility and epigenetic markers can affect gRNA activity, but these vary between cell lines and are more difficult to predict.[121] Algorithms have also been designed to predict and reduce the off-target activity of gRNAs, with factors such as the position in the gRNA influencing tolerance to mismatches.[120] Using multiple gRNAs (typically 5-10) to target each gene can help tackle these issues, but it is imperative to always consider the potential for false positives and false negatives.

Despite these limitations, CRISPR/Cas9 has been applied successfully to identify SLIs in cancer cell lines using both isogenic[122,123] and cell line panel strategies.[109,124,125] By screening two sets of isogenic cell lines that differed by *VHL* status, one group identified novel SL partners, including members of DDR pathways, that could be targeted in clear cell renal cancers with inactivated *VHL*.[123] This type of experiment is more low-throughput and targeted; this is beneficial as interactions with a specific gene of interest can be identified in clinically relevant cell lines. Large cell line panel screens by both the Broad Institute[125] and WSI[109] independently identified that the DNA helicase WRN is a synthetic lethal target in cells that have microsatellite instability, which is caused by a defect in DNA mismatch repair. These pan-cancer cell line studies are limited in the same way as the RNAi projects, in that they depend on good representation of a mutant gene to identify potential interactions.

As with RNAi, CRISPR/Cas9 screens can also be performed in combination with drugs.[126] Wang *et al.* (2019) performed screens in cells treated with an ATR inhibitor and identified that RNASEH2 deficiency causes increased sensitivity both *in vitro* and *in vivo*.[126] Additionally, the recent development of paired gRNA systems has offered a novel approach that highlights the versatility of CRISPR/Cas9 technology. Several groups have designed vectors that encode two gRNAs under the control of separate promoters in a single construct.[127-131] By pairing gRNAs targeting different genes, two genes can be perturbed simultaneously in a single cell. A paired gRNA library can be applied across a range of cell lines to study genetic interactions without the requirement for any existing or engineered mutations.[128,129,131] However, it is necessary to select candidate synthetic lethal pairs as it is unfeasible to screen every possible gene combination. For example, one group performed a screen for interactions using a library which paired TSGs with drug targets[129], and another

randomly paired drug targets.[128] Paired library design is more complex than in single gRNA screens and requires many gRNA combinations plus additional controls. Analysis and interpretation of data from these screens is also challenging. This is a relatively new approach but it has already shown promise, and with further development it is likely to be an invaluable tool for synthetic lethal screening.

### 1.3.6 Computational prediction of synthetic lethal interactions

In addition to experimental approaches, *in silico* methods have been developed to predict SLIs. The majority of these have focused on applying information gained from genetic interaction mapping in yeast to predict synthetic lethality in humans.[132-135] Validation of interactions predicted in this way has been demonstrated in human cell lines, for example between *SMARCB1* and *PSMA4*.[135] Another approach is to take advantage of large datasets such as The Cancer Genome Atlas (TCGA),[136] which profile mutations, gene expression, and many other features of cancer cell lines and human tumours. Mutual exclusivity identified within these datasets has been used to predict SLIs.[137] If LOF mutations or deletion of two genes co-occur less frequently than expected by chance, this may indicate that losing both genes is detrimental to cell fitness and confers a selective disadvantage. Whilst predictions of this kind have been validated[138], the majority do not. Also, this approach is biased towards pairs of genes that are both mutated at a high frequency, which is often not the case as synthetic lethal partners may not be drivers themselves.[62] Perhaps the most thorough strategy is to integrate several 'omics and experimental datasets. Jerby-Arnon *et al.* developed the 'DAISY model' which infers SLIs using three methods: 1) mutual exclusivity of gene inactivation events, 2) associations between under-expression/low copy number of genes and essentiality of another gene (from shRNA screen data), 3) co-expression of genes, as synthetic lethal pairs tend to be involved in similar biological processes and so may have similar expression patterns.[139] Known SLIs were recognised using this model and they validated novel predicted synthetic lethal partners of *VHL* in human cell lines. Whilst *in silico* predictions can be useful to identify and prioritise hits, experimental validation is still required to confirm any interactions.

# 1.4 PBAF/BAF complexes in cancer

## 1.4.1 PBAF/BAF composition and function

ARID1A, ARID1B, ARID2, PBRM1 and SMARCA4 are all subunits of the BRG-/BRM-associated factor (BAF) or Polybromo-associated BAF (PBAF) chromatin remodelling complexes.[140] These complexes use ATP to change the position of histones and other factors on chromatin, and hence regulate transcription.[141] BAF complexes are composed of an ATPase (SMARCA2 or SMARCA4), a DNA-binding protein (ARID1A or ARID1B), and various other subunits (Fig. 1.5).[140] PBAF complexes differ slightly, most notably with the presence of ARID2 in place of ARID1A/ARID1B, and the incorporation of PBRM1 which contains six bromodomains.[140] Recent analysis of fitness correlations between PBAF/BAF subunits in RNAi and CRISPR/Cas9 screen datasets also revealed a previously uncharacterised configuration, referred to as non-canonical BAF (ncBAF).[142,143] Biochemical investigation revealed that ncBAF is composed of common subunits such as SMARCD1 and SMARCA2/4, and specific subunits GLTSCR1/1L and BRD9 (Fig. 1.5).[143,144]

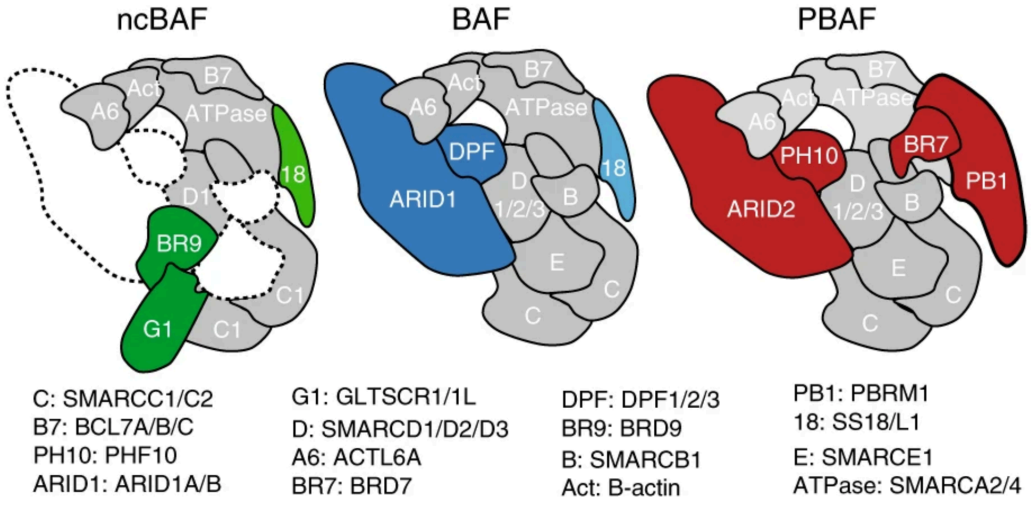



**Figure 1.5. Composition of mammalian PBAF/BAF/ncBAF complexes.** Schematic showing the subunits present in the mammalian ncBAF, BAF and PBAF complexes. Coloured subunits represent those that are specific to a single complex (ncBAF = green, BAF = blue, PBAF = red). Grey subunits are those that are shared between multiple complexes. Figure taken from [143].

As epigenetic regulators, PBAF/BAF complexes have important roles in a range of biological processes including neural differentiation,[145,146] cardiac[147] and brain[148] development, self-renewal and pluripotency in embryonic stem cells,[149] and metabolism.[150,151] They also have non-transcriptional roles in DNA repair and PBAF complexes specifically have been associated with processes that maintain genomic stability.[152] The function of these complexes varies depending on the subunit composition, and inactivation of individual subunits can lead to the development of different cancer types.[152] The reason for this variation is not clear, but some subunits are mutated more frequently than others. It is estimated that collectively, the genes encoding PBAF/BAF subunits are mutated in ~20% of human tumours, making them one of the most commonly mutated complexes in cancer.[153,154]

### 1.4.2 BAF-specific subunits: ARID1A and ARID1B

*ARID1A* (also known as *BAF250a*) is the most commonly mutated BAF gene.[152,153] It encodes a protein that contains an ARID DNA-binding domain[155] and an uncharacterised domain, which may have ubiquitin ligase activity.[156] Inactivating mutations in *ARID1A* are found in many different cancer types. Approximately 50% of ovarian clear cell carcinomas (OCCC) and endometriosis-associated ovarian carcinomas harbour LOF mutations in *ARID1A*.[157,158] It is also frequently mutated in uterine endometrial carcinoma (34%),[159] stomach cancer (34%)[160] and bladder cancer (29%),[161] amongst others. Although missense mutations do occur, mutations are generally truncating (nonsense or frameshift).[152] No apparent 'hotspot' has been identified, with mutations spread throughout the gene.

*ARID1B* (or *BAF250b*) is very similar to *ARID1A*, with ~60% sequence homology and it also encodes a DNA-binding subunit.[162] ARID1A and ARID1B are mutually exclusive; only one of these subunits is present in a single BAF complex, but a mixture of complexes containing either of them can exist.[162] However, mutations in *ARID1B* are much less frequent.[152] This may be due to variation in the expression or function of these subunits across different cell types; opposing roles for these subunits have been shown.[163]

### 1.4.3 PBAF-specific subunits: PBRM1 and ARID2

PBRM1 (or BAF180) is a protein containing six bromodomains and is specific to the PBAF complex. *PBRM1* is the second most commonly mutated gene in renal clear cell carcinoma, with mutation or loss occurring in ~41% of cases.[164] It is primarily inactivated by truncating

mutations or deletion. The exact role that PBRM1 plays in the PBAF complex is unclear, but it has been shown to be important for genomic stability, with roles in sister chromatid cohesion[165] and re-priming stalled replication forks.[166]

Another PBAF-specific subunit is encoded by *ARID2* (or *BAF200*). Although not a homolog of ARID1A/1B, it also has an ARID DNA-binding domain and these subunits are mutually exclusive.[152] *ARID2* mutations occur in a range of cancer types including melanoma,[167-169] non-small-cell lung cancer[170] and hepatitis-associated hepatocellular carcinoma.[171] *ARID2* mutations are rarely found in renal clear cell carcinoma, suggesting that the functional importance of these PBAF-specific subunits varies with cell type.[152,153]

## 1.4.4 Targeting PBAF/BAF-mutant cancers

Considering the frequency and range of PBAF/BAF mutations, there is huge interest in finding new therapeutic approaches for cancers driven by these complexes. Malignancy is generally associated with inactivation of the subunits and so targeting these deficiencies is challenging. The main focus of ongoing research is to identify tumour cell vulnerabilities that are induced when PBAF/BAF complexes are impaired. The majority of studies thus far have centred around *ARID1A*, but dependencies induced by loss of other subunits have been identified (Table 1.2).

### 1.4.4.1 Dependencies associated with *ARID1A* mutation

Using shRNA data from the large-scale Project Achilles screens, *ARID1B* was identified as an essential gene required for growth specifically in *ARID1A*-mutant cancer cell lines.[172] This dependency was more pronounced when considering only lines with inactivating mutations in *ARID1A*, excluding missense mutations. In that study, the interaction was validated experimentally, with *ARID1B* knockdown causing impaired proliferation and colony formation in *ARID1A*-mutant OCCC cell lines but not in WT lines. Various other studies have supported this SLI, and this will be discussed further in Chapter 4.

A recent study found that ARID1A plays a role in the metabolism of glutathione (GSH) by enhancing transcription of SLC7A11.[173] This gene encodes a cystine transporter and low expression in *ARID1A*-deficient cells is associated with low basal GSH levels. Depletion of GSH causes increased reactive oxygen species which can induce apoptosis. Cells lacking ARID1A were shown to be specifically vulnerable to inhibition of the GSH metabolic pathway.[173] In that study, researchers focused on using buthionine sulfoximine, an inhibitor

against GCLC which is a rate-limiting enzyme in GSH synthesis. GCLC depletion specifically impaired the growth of *ARID1A*-mutant ovarian cancer cell lines both *in vitro* and *in vivo*.

High-throughput drug screening in OCCC lines revealed that *ARID1A* mutation is associated with increased sensitivity to dasatinib, a kinase inhibitor[174]. Depletion of ARID1A in OCCC, breast and colorectal cancer cell lines confirmed this increased sensitivity. Further investigation with siRNAs suggested that this may be due to an SLI with YES1, one of the dasatinib targets[174]. Dasatinib treatment increased G1-S cell cycle arrest and caspase activity in *ARID1A*-deficient cells. Sensitivity to the drug appeared to be dependent on p21 and RB1 activity. A preliminary experiment with an OCCC tumour xenograft indicated that dasatinib impaired growth of *ARID1A*-mutant tumours *in vivo*.

Various studies have associated ARID1A with the PI3K/AKT/mTOR pathway. Significant enrichment of activating mutations in *PIK3CA* and loss of *PTEN* have been observed in *ARID1A*-mutant endometrial cancer and OCCC.[175] Project Achilles identified *PIK3CA* depletion as the second strongest hit for synthetic lethality with *ARID1A* mutation.[75] Additionally, an mTORC1/2 inhibitor that targets downstream signalling of PI3K/AKT was significantly more effective in *ARID1A*-mutant lines compared to WT.[176] PI3K and mTOR inhibitors were also screened in a large panel of OCCC cell lines and xenograft models.[177] In contrast to previous findings, the mutational status of *ARID1A* was not sufficient to discriminate the sensitivities of the cell lines. These conflicting data suggest that targeting of the PI3K/AKT/mTOR pathway may not be broadly applicable as a synthetic lethal approach in *ARID1A*-mutant tumours and further investigation is required.

It was recently shown that ARID1A has a role in homologous recombination via an interaction with ATR, a regulator of the DNA damage response.[178] *ARID1A*-deficient cells have impaired checkpoint signalling and reduced repair of DNA DSBs. PARP inhibitors are known to be lethal in cancer cells that are deficient in DSB repair pathways.[46,47] Using an isogenic system, it was found that breast, colorectal and ovarian cancer cell lines have increased sensitivity to PARP inhibitors when ARID1A is depleted.[178] Treatment with a PARP inhibitor also specifically suppressed growth of *ARID1A*-mutant breast and colorectal xenografts *in vivo*. Another study identified increased sensitivity to ATR inhibitors in *ARID1A*-mutant cancer cell lines both *in vitro* and *in vivo.*[179] Loss of ARID1A function was associated with reduced progression through the cell cycle and defects in recruitment of TOP2A to chromatin. Inhibition of ATR would affect the repair of DNA damage associated with these defects, leading to apoptosis and hence could explain the increased sensitivity to ATR inhibitors.

Together, these findings suggest that exploitation of a DSB repair deficiency could be an effective strategy to target *ARID1A*-mutant cancers.

Using 3D OCCC models, it was shown that *ARID1A*-mutant cells are more sensitive to EZH2 inhibition, with increased induction of apoptosis.[180] A similar effect was also observed using tumour xenografts. PIK3IP1 is a negative regulator of the PI3K/AKT/mTOR pathway and is down-regulated when ARID1A is lost. Subsequent silencing of *PIK3IP1* by EZH2-mediated methylation activates this pathway and increases proliferation. Use of an EZH2 inhibitor resulted in increased expression of this regulator, reduced proliferation and increased apoptosis.[180] This would suggest that ARID1A-deficient cells are dependent on *PIK3IP1* inactivation. Further experiments were performed to investigate synthetic lethality with EZH2 in other tumour cell types.[181] Lung, adrenal gland and renal carcinoma cell lines carrying mutations in *ARID1A*, *PBRM1* and *SMARCA4* were found to be vulnerable to EZH2 inhibition. This dependency appeared to be specifically associated with destabilisation of the PRC2 complex when EZH2 was disrupted. Further to this, repression of *SMARCA2* has been shown to be a potential biomarker for the efficacy of EZH2 inhibition in *SMARCA4* and *ARID1A* mutants.[182] These studies suggest that dependency on EZH2 occurs more generally across PBAF/BAF-mutant cancers, not just *ARID1A* mutants.

Studies have also identified synthetic lethality between *ARID1A* and other epigenetic regulators. One group found that depletion or inhibition of HDAC2 caused re-expression of *PI3KIP1* in *ARID1A*-mutant cells, with reduced proliferation and increased apoptosis.[183] HDAC2 is a binding partner of EZH2-containing PRC2 complexes[184] and this interaction is dependent on ARID1A.[183] ARID1A has also been identified as a transcriptional repressor of HDAC6.[185] HDAC6 represses p53, and inactivation of *ARID1A* leads to up-regulation of *HDAC6* and a reduction in p53-mediated apoptosis. *ARID1A*-mutant OCCC cell lines and xenografts are specifically susceptible to pharmacological inhibition of HDAC6.

**Table 1.2. Candidate synthetic lethal targets in PBAF/BAF-mutant cancers.**

| Tumour status | Vulnerability | Reference |
|---|---|---|
| ***ARID1A*-deficient** | EZH2 inhibition | 180-182 |
| | GSH/GCLC inhibition | 173 |
| | ARID1B depletion | 172 |
| | YES1 inhibition/Dasatinib | 174 |
| | PI3K/AKT/mTOR inhibition | 75,175,176,177 |
| | PARP inhibition | 178 |
| | ATR inhibition | 179 |
| | HDAC2 inhibition | 183 |
| | HDAC6 inhibition | 185 |
| ***PBRM1*-deficient** | EZH2 inhibition | 181 |
| | TIP60 deficiency | 186 |
| ***SMARCA4*-deficient** | CDK4/6 inhibition | 187,188 |
| | *MAX* deficiency | 189 |
| | OXPHOS inhibition | 190 |
| | *SMARCA2* deficiency | 191,192 |
| | EZH2 inhibition | 181,182 |
| ***SMARCB1*-deficient** | *SMARCA4* deficiency | 193 |
| | BRD9 inhibition | 143 |
| | ATR inhibition | 194 |
| | EZH2 inhibition | 181,195-197 |

#### 1.4.4.2 Dependencies associated with loss of other PBAF/BAF genes

In addition to EZH2 inhibition, various vulnerabilities have been associated with mutations in other PBAF/BAF subunits (Table 1.2). *SMARCA4*-deficient tumours are specifically vulnerable to inhibition of CDK4/6 in subtypes of ovarian and lung cancer,[187,188] and to inhibition of oxidative phosphorylation in lung cancer.[190] A recent study identified a synthetic lethal interaction between *SMARCA4* and *MAX*, with mutually exclusive mutations present in small cell lung cancer.[189] *MAX*-deficient cells were shown to be specifically vulnerable to SMARCA4 depletion; it would be interesting to investigate whether MAX is a targetable vulnerability in *SMARCA4*-mutant lung cancers.

Using an RNAi screen based on a competitive growth assay, TIP60 was implicated as a potential synthetic lethal partner of PBRM1.[186] Treatment with a *TIP60* siRNA caused a greater loss of cells expressing an shRNA targeting *PBRM1* compared to those expressing a control shRNA. TIP60 is a histone acetyltransferase which has a role in response to DNA double strand breaks.[198] Inhibition of both genes lead to increased micronuclei formation, a feature often associated with DNA damage[199], which could explain why the double knockdown population was depleted.

Synovial sarcoma and malignant rhabdoid tumour (MRT) cell lines show selective sensitivity to suppression of the ncBAF subunit, BRD9.[143] In most synovial sarcomas, a reciprocal translocation results in an oncogenic fusion between *SS18* and one of *SSX1/2/4.*[200] SS18-SSX1 fusion proteins disrupt BAF complexes by displacing WT SS18 and also another subunit, SMARCB1.[201] As a result, SMARCB1 is degraded. Homozygous loss of *SMARCB1* is a common feature of MRT.[202-204] Researchers found no dependencies associated with any other BAF or PBAF subunits, suggesting that the vulnerability in these BAF-perturbed cancers is specific to ncBAF disruption.[143] This could have clinical potential as inhibitors of BRD9 have recently been developed.[205,206] Expression of SS18-SSX fusion proteins, leading to depletion of SMARCB1, has also been shown to induce sensitivity to ATR inhibitors in cell lines.[194]

Similar to the *ARID1A/ARID1B* interaction, dependencies between other PBAF/BAF subunits have been identified. Various studies have highlighted an SLI between two other mutually exclusive components, *SMARCA4* and *SMARCA2.* This was initially shown in non-small-cell lung carcinoma, where *SMARCA2* depletion suppressed growth of *SMARCA4*-deficient cell lines *in vitro* and tumour xenografts *in vivo.*[191,192] Sensitivity to *BRM*-targeting shRNAs was also observed in *SMARCA4*-mutant ovarian and liver cancer cell lines, suggesting

it was not specific to the lung.[192] Another study found that tumour cells deficient in SMARCB1 were dependent on the function of SMARCA4.[193] Inactivation of *Smarca4* also caused a marked reduction in tumour formation in mice that were already mutant for *Smarcb1*. Researchers proposed that tumourigenesis in *SMARCB1* mutant cells was not driven by the loss of PBAF/BAF function, but rather by an oncogenic effect of residual complexes containing SMARCA4.[193]

### 1.4.5 Clinical potential of vulnerabilities in PBAF/BAF-mutant cancers

Many of the SLIs discussed here were identified in a small number of cell lines, often only in one tumour type. Further investigation is required to determine how robust and widely applicable these are, however some of them do have clinical potential. Several EZH2 inhibitors are in clinical trials and could be tested in a range of PBAF/BAF-mutant cancers.[207] Alternatively, the broad HDAC inhibitor Vorinostat could be used for HDAC2 inhibition.[13] A HDAC6-specific inhibitor, ACY1215, has been tested in multiple myeloma[208] and may be effective in *ARID1A*-mutant OCCC patients. Dasatinib and other compounds that target YES1 are being investigated for several cancer types.[209] There are also many clinically advanced inhibitors against ATR, PARP and the PI3K/AKT/mTOR pathway which could be re-purposed to target PBAF/BAF-deficient tumours.[209]

# 1.5 Induced pluripotent stem cells as a model

Induced pluripotent stem cells (iPS cells or iPSCs) are adult cells that have been genetically reprogrammed to a state similar to embryonic stem cells.[210] They have the potential to differentiate into cells belonging to all three germ layers (endoderm, mesoderm, ectoderm). Yamanaka *et al.* were the first to engineer iPSCs by exogenously expressing four genes in mouse skin cells: *Oct3/4, Sox2, c-Myc* and *Klf4*.[210] Soon after this, iPSCs were derived from human cells and one group demonstrated that *OCT4, SOX2, NANOG* and *LIN28* were also sufficient for reprogramming.[211,212] This discovery brought new hope for regenerative medicine which has unfortunately yet to be realised. In 2014, the first human trial began using cells derived from iPSCs to treat macular degeneration. However, this trial stopped after only one participant was treated because two genetic changes were identified in the cells and their potential effects were unclear. In the last few years several small-scale trials have been initiated to use iPS-derived cells in patients with Parkinson's disease, spinal cord injury and heart disease.[213-215] Although progress in the therapeutic field has been slower than expected, iPSCs have become an invaluable research tool with many applications.

## 1.5.1 Genome editing of induced pluripotent stem cells

Zinc-finger nucleases[216] and transcription activator-like effector nucleases[217] have been used successfully for targeted genome editing in iPSCs, but CRISPR/Cas9 has become the dominant technology due to its relative simplicity and versatility. Various strategies for delivering Cas9 and gRNA to iPSCs for genome editing have been implemented.[218] The potential for off-target activity is still a concern when using CRISPR/Cas9, but various studies have reported low off-target effects in human iPSCs.[219-227] These studies used both targeted and whole genome sequencing approaches to identify differences such as single nucleotide variants and small indels in edited iPSCs that were not present in the parental cells.

CRISPR/Cas9 technology has been used in iPSCs to study the molecular and cellular pathological mechanisms of many diseases with a genetic basis.[228] Diseases can be modelled by knocking out a gene: KO of *DNMT3B* in iPSCs results in hypomethylation similar to that observed in patients with a rare autosomal recessive disorder caused by mutations in *DNMT3B*.[229] Disease-associated mutations can also be corrected using HDR in patient-derived iPSCs: correction of *CYBB* mutations in cells from patients with chronic granulomatous disease resulted in restoration of ROS activity in iPS-derived monocytes.[230] Engineering isogenic

derivatives can allow for comparison of WT and mutant iPSCs (or iPS-derived cells) to identify genes responsible for disease phenotypes and/or to elucidate disease mechanisms. For example, an isogenic line was generated by correcting a *SOD1* mutation in iPSCs derived from an amyotrophic lateral sclerosis patient. RNA-seq analysis of mutant and WT motor neurons revealed both up-regulated and down-regulated genes associated with this mutation.[231]

### 1.5.2 Screening induced pluripotent stem cells

The ability to accurately model disease phenotypes using iPSCs has made them an ideal tool for both target-based and phenotypic chemical screening. Over one thousand compounds have been screened in iPSC models for various diseases, including candidates that have progressed to clinical trials (as reviewed by Shi *et al.,* 2017).[232] However, large-scale genetic screening has been limited in comparison, with focus primarily on targeted approaches. There have been various reports that gene targeting in iPSCs/ESCs is much less efficient than in transformed cell lines, which may explain the lack of published high-throughput screens.[88,233-235]

Several groups have performed genome-scale RNAi screening[236,237] and CRISPR/Cas9 screening[238-240] in human embryonic stem cells (hESCs), but no iPSC screens have been published. One shRNA screen identified regulators of self-renewal and pluripotency in hESCs, with further validation and functional assays confirming the role of a transcription factor, PRDM14.[126] Another study carried out differentiation after transduction of hESCs with an shRNA library to identify genes required for neural lineage development.[125]

Three CRISPR/Cas9 KO screening studies have been performed with the aim of identifying genes essential for pluripotent stem cell fitness.[238-240] Two of these investigated the same hESC line and another used a haploid hESC line. Additionally, one of these studies performed screening to identify regulators of pluripotency and to identify genes that suppressed dissociation-induced death.[238] Although hits from these screens were successfully validated, several issues arose in the stem cell screens which had not been described previously in cancer cell line screens. A study published prior to these demonstrated that stem cells are highly sensitive to DSB-induction mediated by Cas9.[241] This correlated with the findings in the genome-wide screens, and may further explain the relative lack of literature regarding CRISPR/Cas9 screening in iPSCs.

## 1.6 Summary

The development of targeted therapies has been instrumental in improving cancer survival, but this area is dominated by oncogenic inhibitors. Therapies to target cancers driven by inactivation of TSGs are relatively scarce. Exploiting vulnerabilities associated with the loss of these genes is proving to be a promising therapeutic approach. The development of CRISPR/Cas9 technology has revolutionised our ability to screen for genetic interactions on an unprecedented scale, and has already improved our understanding of cancer dependencies. Many of the large studies thus far have focused on identifying SLIs by associating genetic dependencies with mutation/expression in panels of cancer cell lines.

Isogenic models may offer a more reliable system to identify dependencies, but few TSGs have been systematically studied using this approach. Indeed, the 'gold standard' SLI between BRCA and PARP was first identified in an isogenic model.[47] Interestingly, this finding was in mouse ESCs rather than in a cancer cell line. Human iPS cells have proved to be a very useful tool for disease modelling; a normal genetic background allows for interrogation of genetic interactions in the absence of many other aberrations. Considering the lack of robust SLIs identified to date, a new approach using a model like iPSCs may address the issues caused by the genetic complexities of cancer cell lines.

The overarching aim of this project is to identify novel synthetic lethal partners of known TSGs. To do so, CRISPR/Cas9 screens will be performed in a panel of isogenic human iPSCs carrying inactivating mutations in a range of TSGs. Particular focus will be placed on PBAF/BAF complex subunits, as therapies targeting these genes are lacking despite the fact that they are collectively mutated in ~20% of cancers. Ultimately, we hope to gain a broader understanding of the vulnerabilities associated with TSG loss and to highlight novel targets for cancers with an unmet clinical need.

# 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]

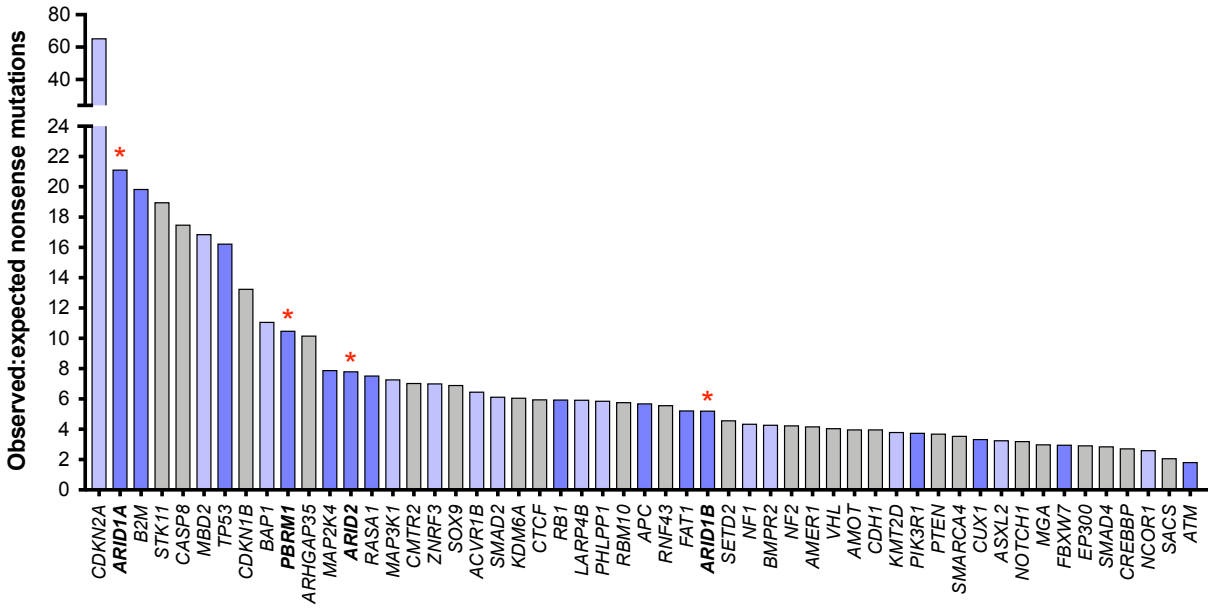



**Figure 2.1. Candidate tumour suppressor genes.** Wong *et al.* performed analysis on 7,651 genome sequences (352 whole genomes, 7,299 exomes) across 28 tumour types.[242] The genes shown had a high ratio of observed to expected nonsense mutations (Benjamini-Hochberg's false discovery rate adjusted 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).

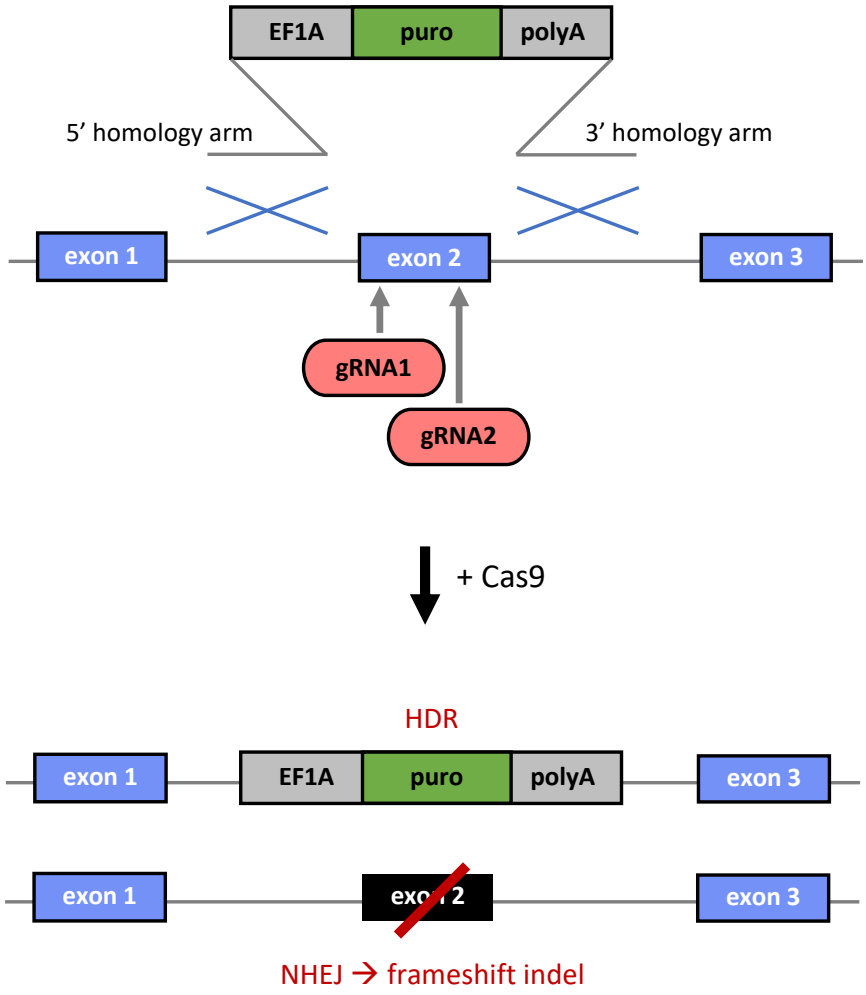



**Figure 2.2. Engineering isogenic KO iPSCs.** Schematic showing the strategy used to engineer isogenic KO iPSCs. A template containing a puromycin resistance cassette flanked by 5' and 3' homology arms, two gRNAs targeting an exon, and Cas9 were transfected into BOB cells. Edited clones had the puromycin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

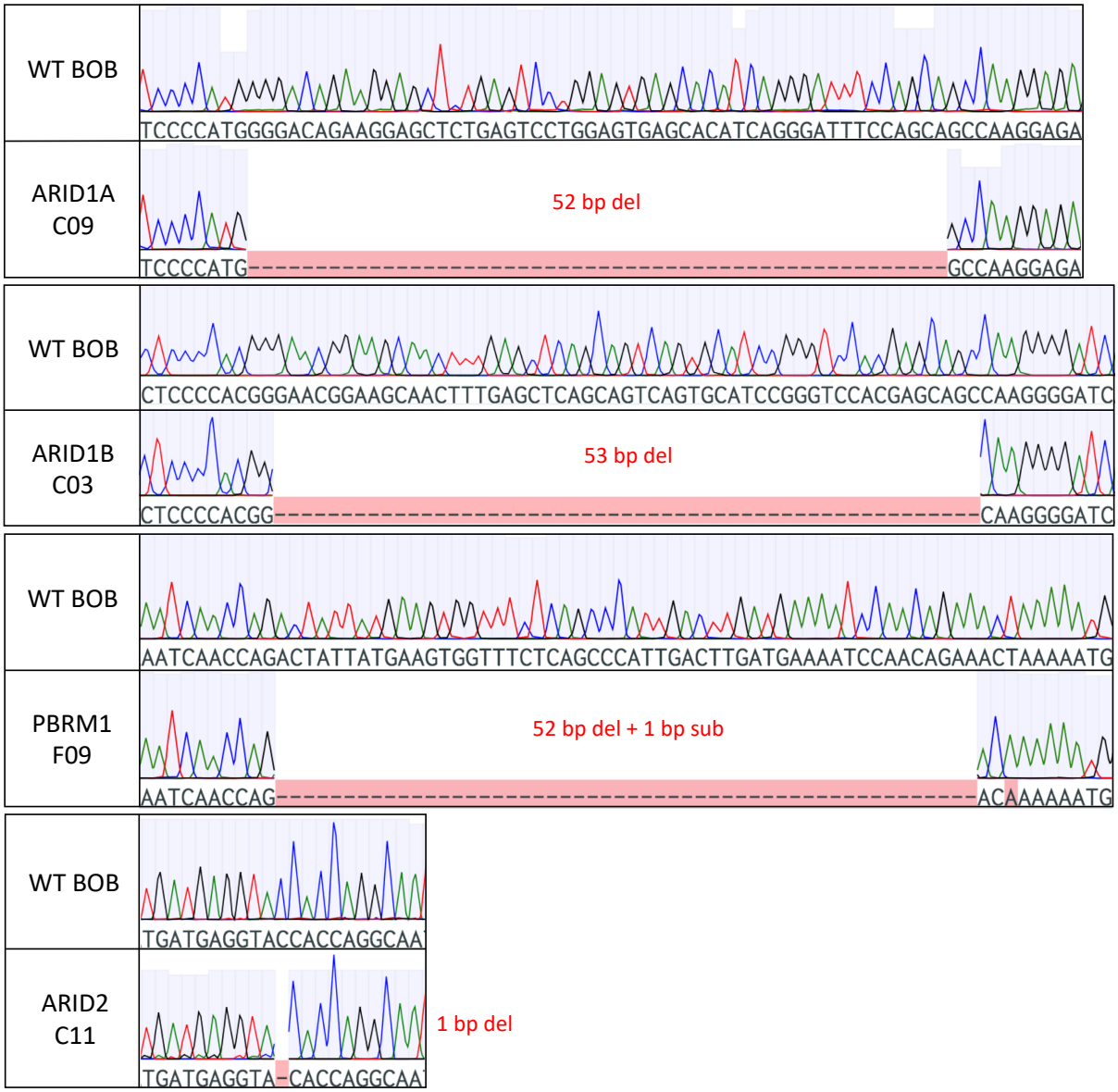



**Figure 2.3. Sequencing of PBAF/BAF mutations in KO iPSCs.** PCR and Sanger sequencing were performed to confirm the frameshift deletions (shown in red) in *ARID1A, ARID1B, PBRM1* and *ARID2*. The WT sequence in the parental was also sequenced for comparison. Traces are shown for one clonal line per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

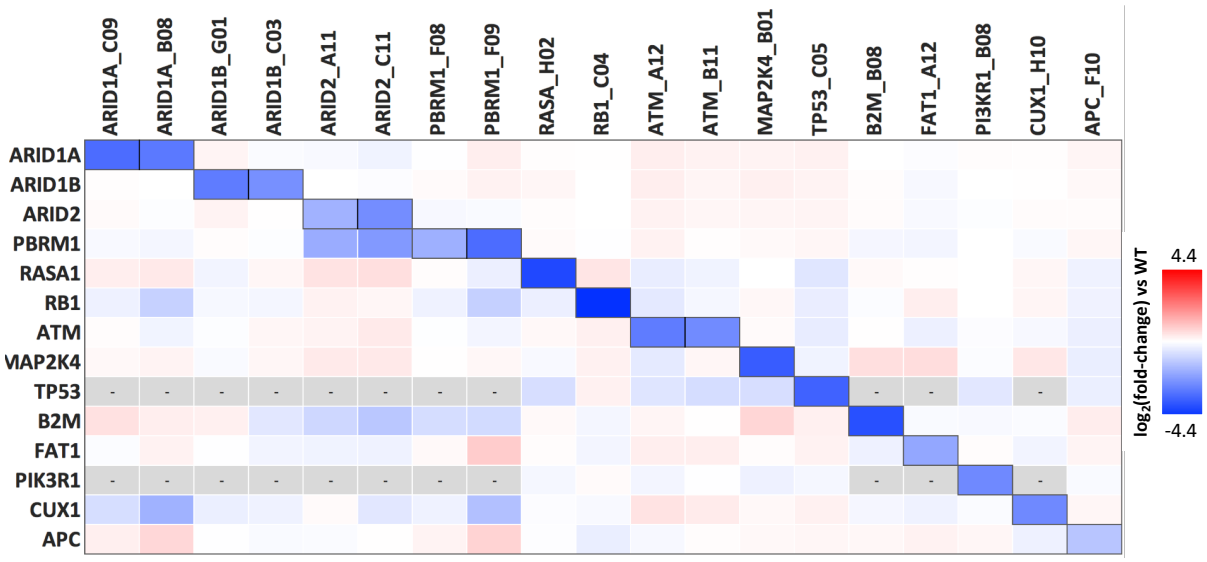



**Figure 2.4. Proteomics of knockout iPSC lines.** $Log_2$ (fold-change) of protein abundance in KO lines relative to the parental BOB line, as measured by LC-MS. The parental was always analysed in the same run as the KO line that it was compared to. (–) indicates that a protein was not detected. FBXW7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

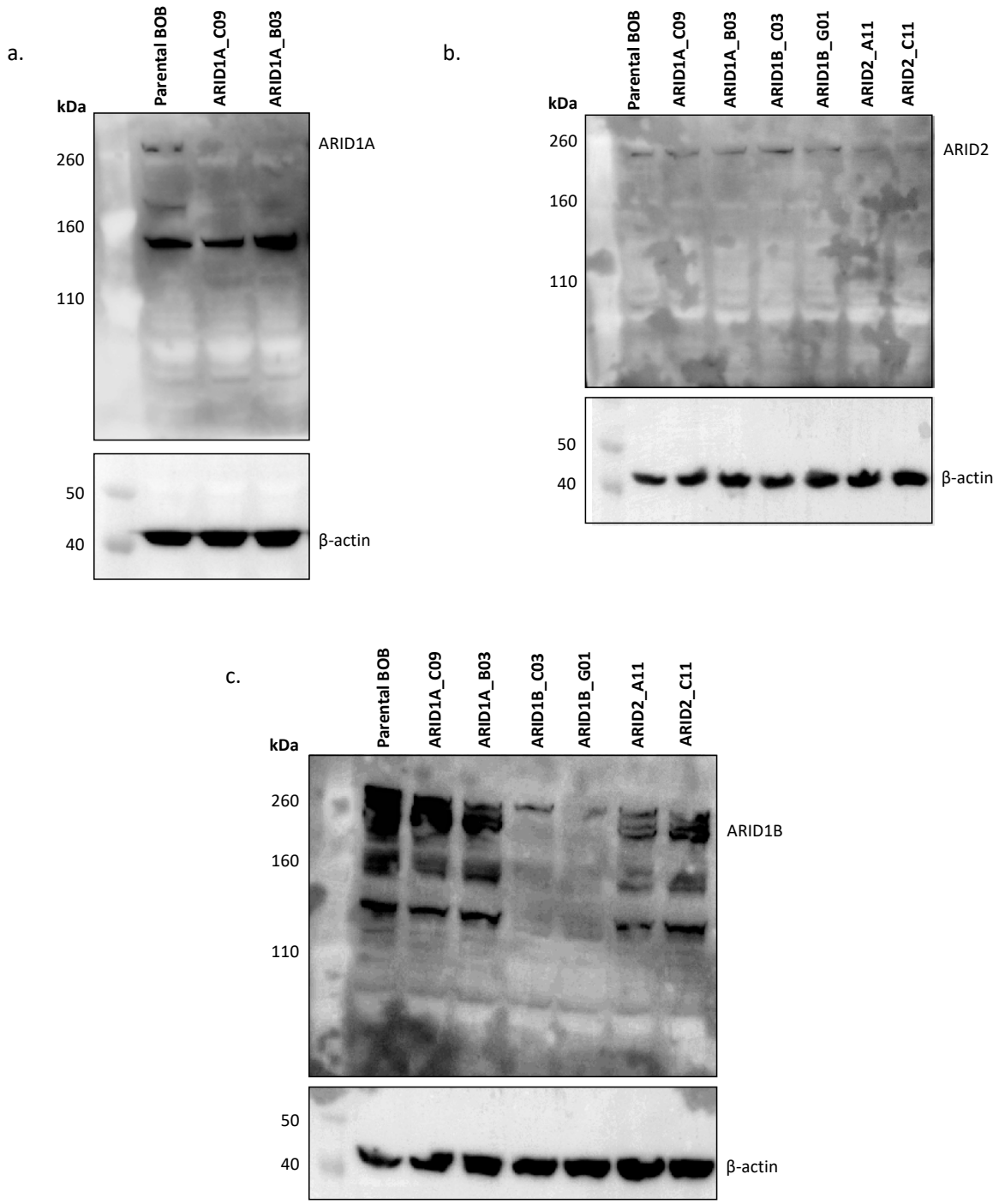



**Figure 2.5. Western blotting of ARID1A, ARID2 and ARID1B in iPSCs.** Protein lysates were extracted from parental BOB cells and both knockout clones for *ARID1A, ARID1B* and *ARID2*. Western blotting was performed to measure expression of ARID1A (a), ARID2 (b) and ARID1B (c) proteins. In each case, 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).

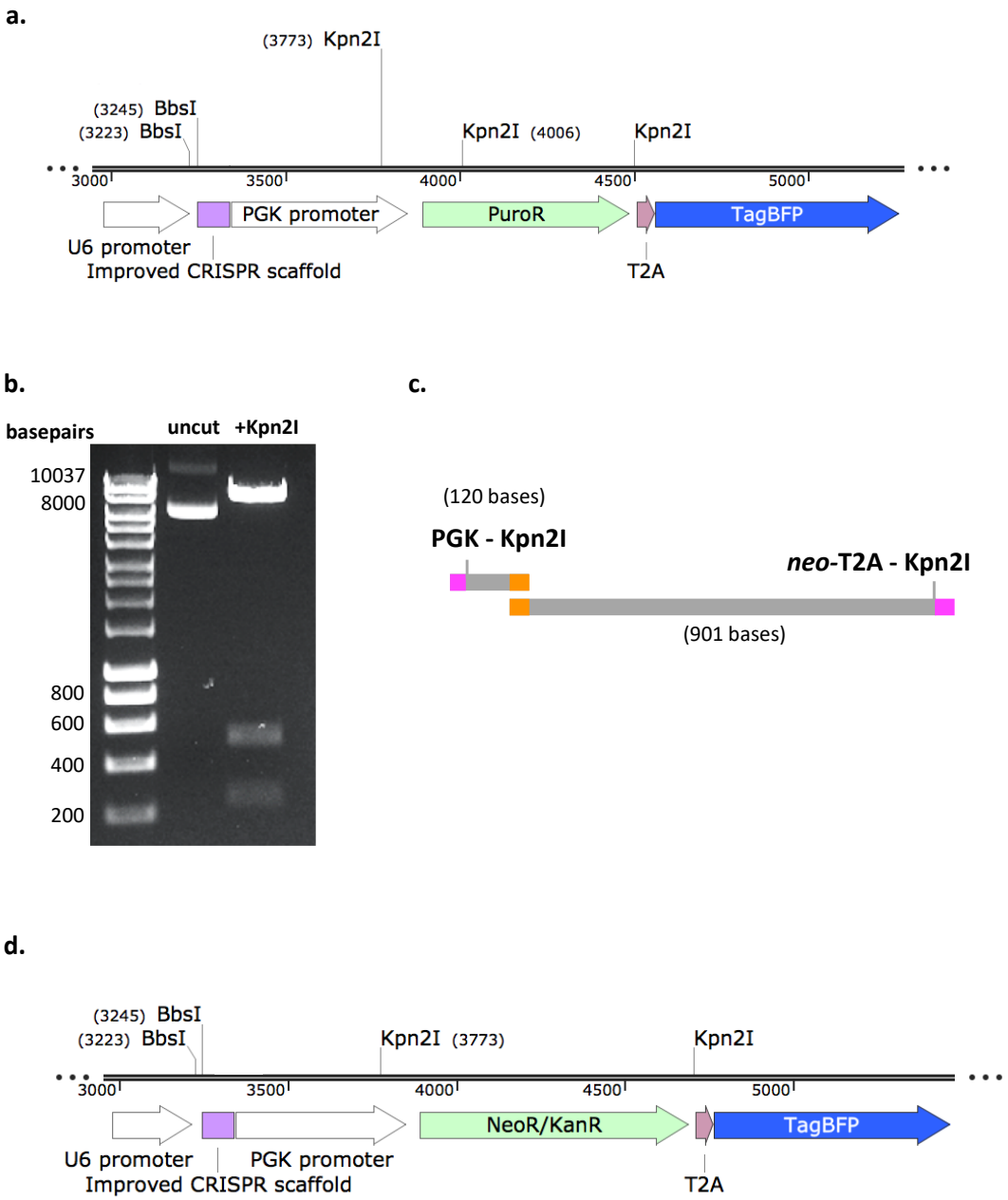



**Figure 2.6. Cloning of a neomycin-resistant library backbone. a)** Schematic showing key features of the Yusa v1.1 library backbone, **b)** agarose gel image showing Kpn2I digestion of the Yusa v1.1 backbone (marker sizes indicated on left), **c)** illustration of the oligonucleotides designed for 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.

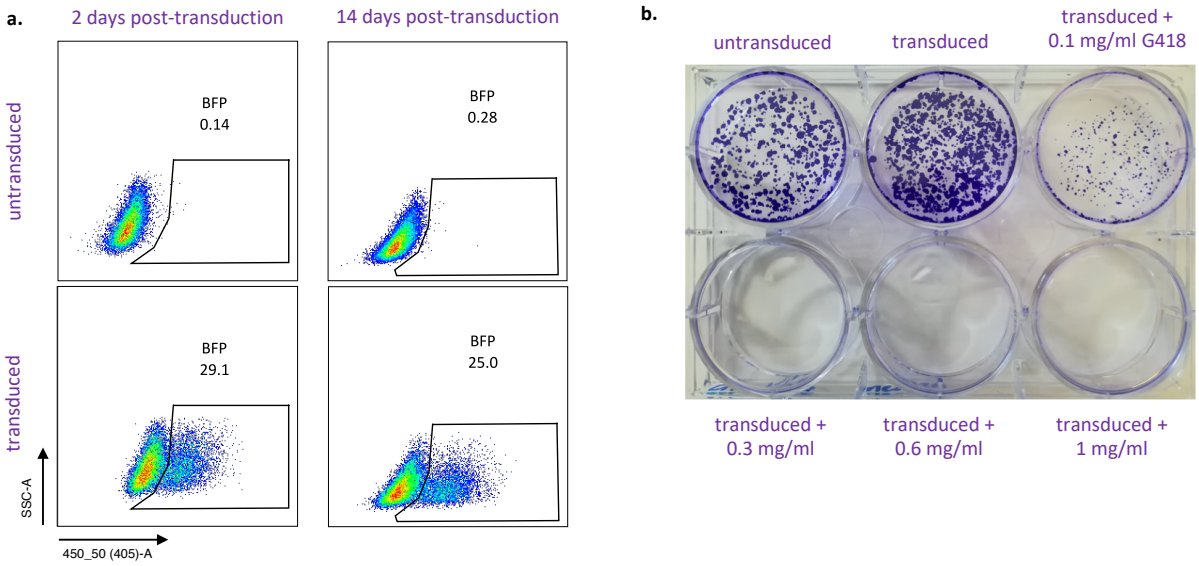



**Figure 2.7. Testing neoR backbone resistance to G418. a)** Flow cytometry plots showing BFP expression (450_50 (405)-A) in BOB-Cas9 cells at 2- and 14-days post-transduction with the neoR backbone, **b)** clonogenic assay showing BOB-Cas9 cells stained with crystal violet on day 14 post-transduction 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.

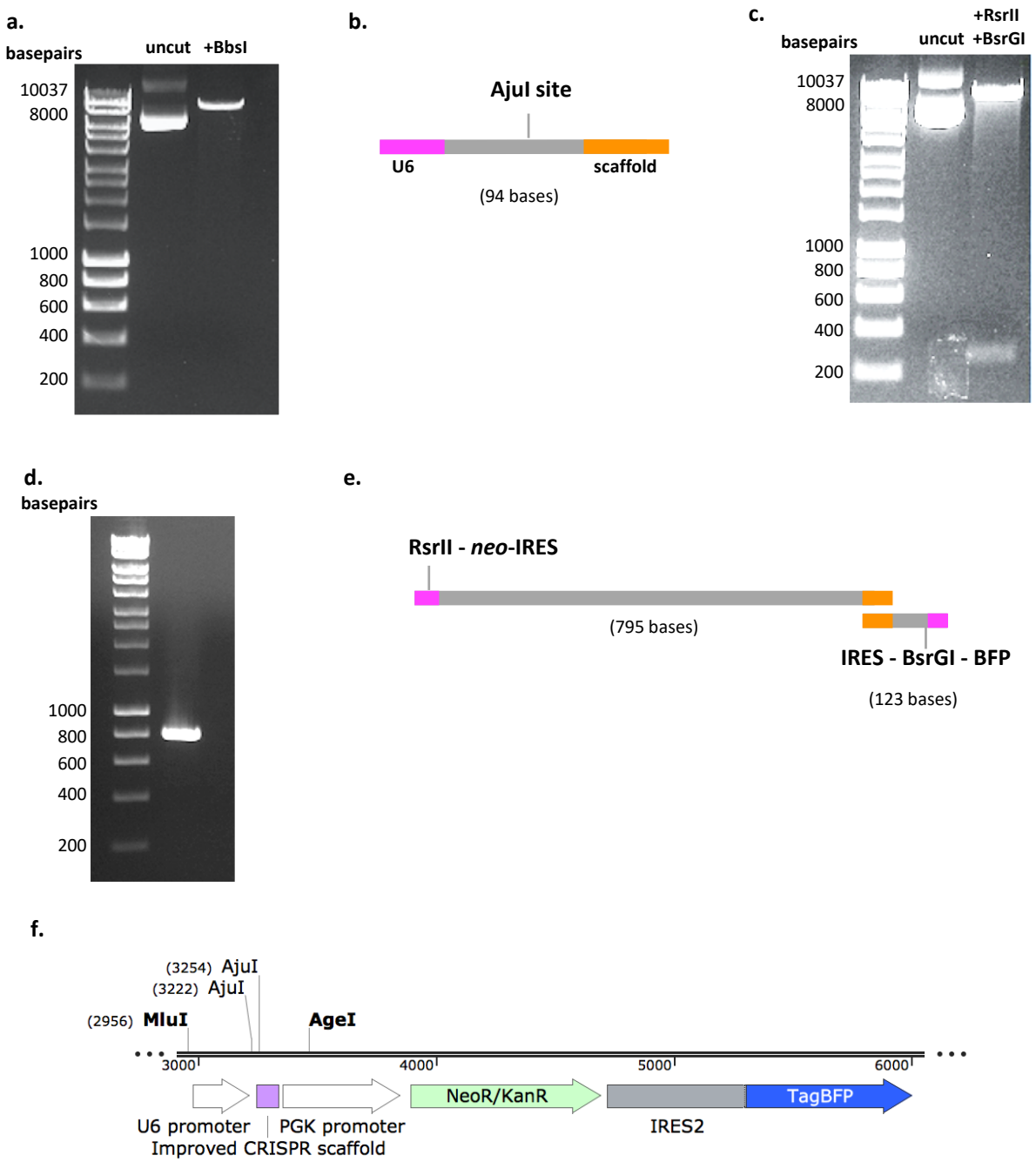



**Figure 2.8. Cloning in AjuI recognition site and an IRES element. a)** BbsI digest of neoR backbone, **b)** illustration of the oligonucleotide designed to insert AjuI recognition site via Gibson Assembly, **c)** RsrII/BsrGI double-digest of neoR (AjuI) backbone, **d)** PCR amplification of neoIRES 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.

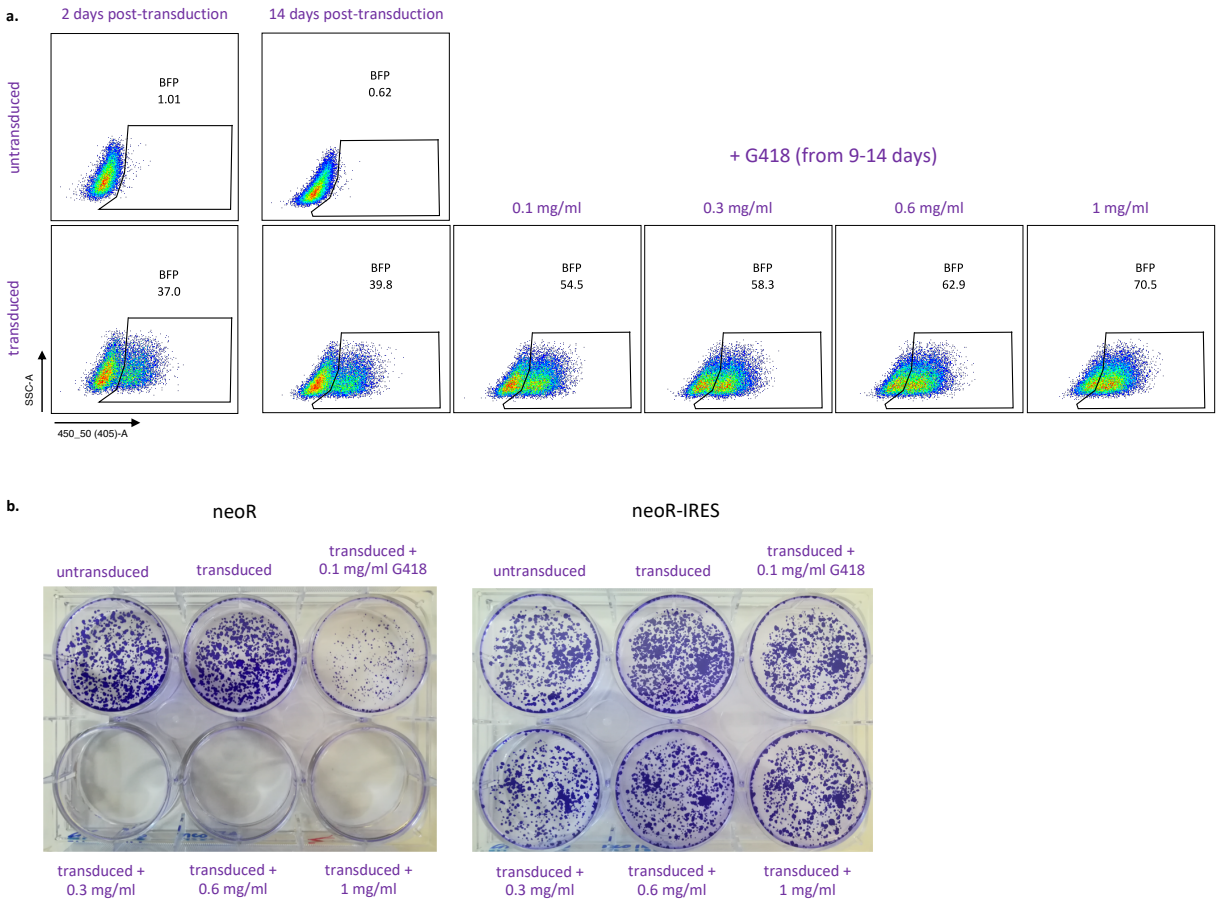



**Figure 2.9. Testing neoR-IRES backbone resistance to G418. a)** Flow cytometry plots showing BFP expression in BOB-Cas9 cells at 2- and 14- days post-transduction with the neoR-IRES backbone. Cells were selected with various doses of G418 from day 9-14, **b)** Clonogenic assay showing BOB-Cas9 cells 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.

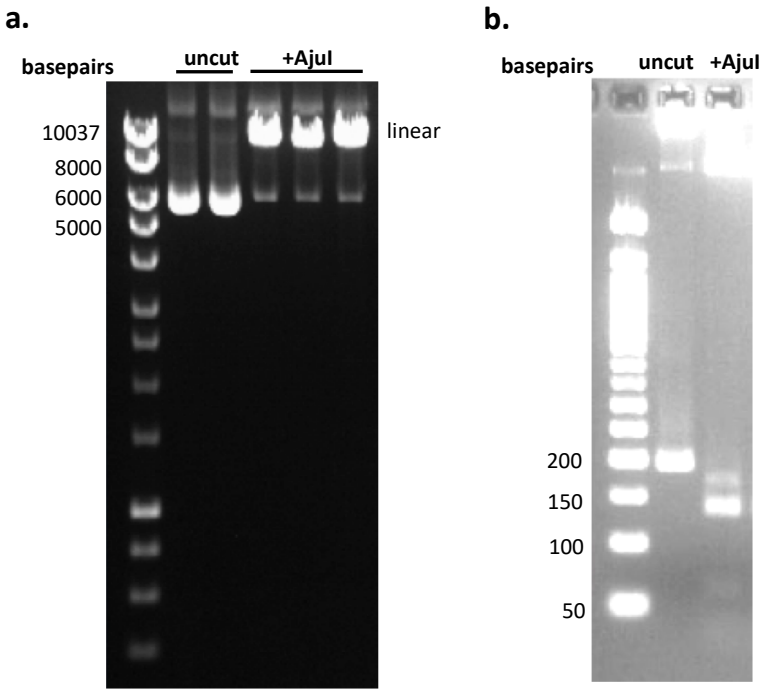



**Figure 2.10. Digestion of neoR-IRES backbone with AjuI. a)** neoR-IRES plasmid digested with AjuI for (left-right) 15 minutes, 1 hour and 2 hours, **b)** AjuI digestion of a PCR amplicon across the restriction sites. Uncut amplicon = 194 bp; expected bands with 2 cuts = 132 bp, 32 bp, 30 bp; expected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

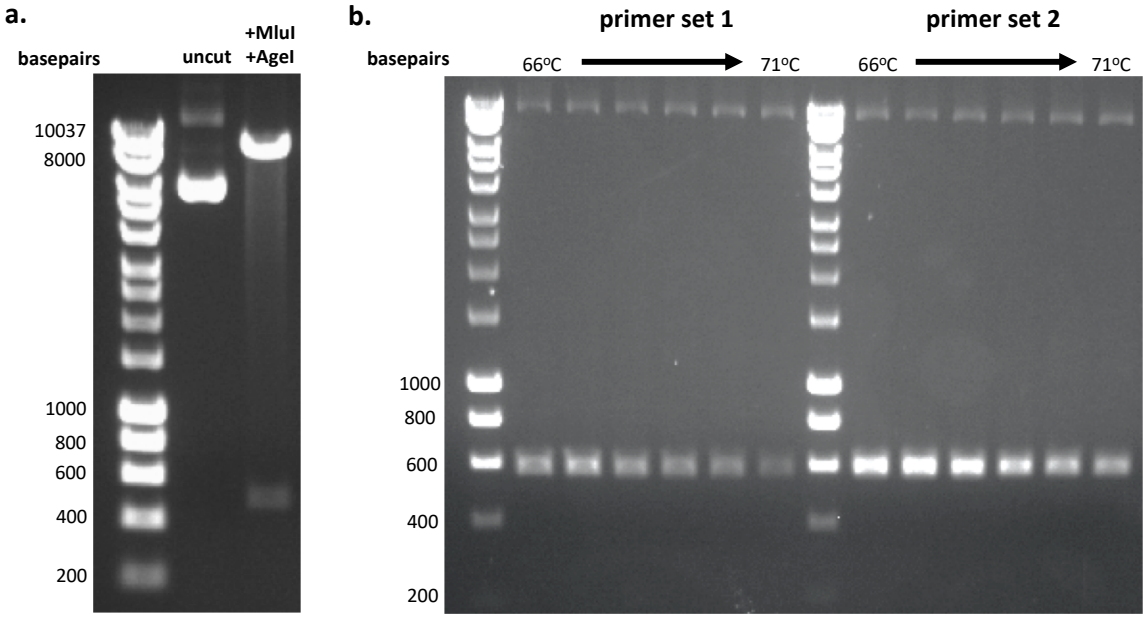



**Figure 2.11. Cloning strategy for gRNA library transfer. a)** MluI/AgeI double-digest of the neoR-IRES backbone, **b)** PCR amplification of gRNAs from the Yusa v1.1 library with two primer sets and a gradient of annealing temperatures.

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.

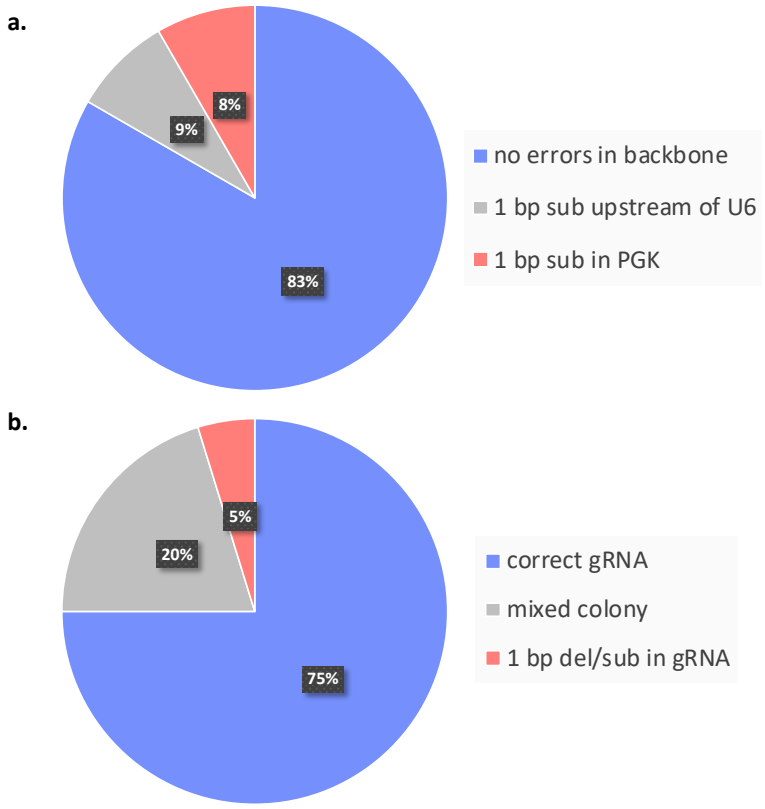



**Figure 2.12. Sequencing of colonies from gRNA library transfer.** Results from Sanger sequencing of 84 colonies cultured after the transfer of gRNAs into the neoR-IRES backbone. The error rate was calculated for **(a)** the plasmid backbone (in the area altered by cloning) and **(b)** the gRNA sequence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

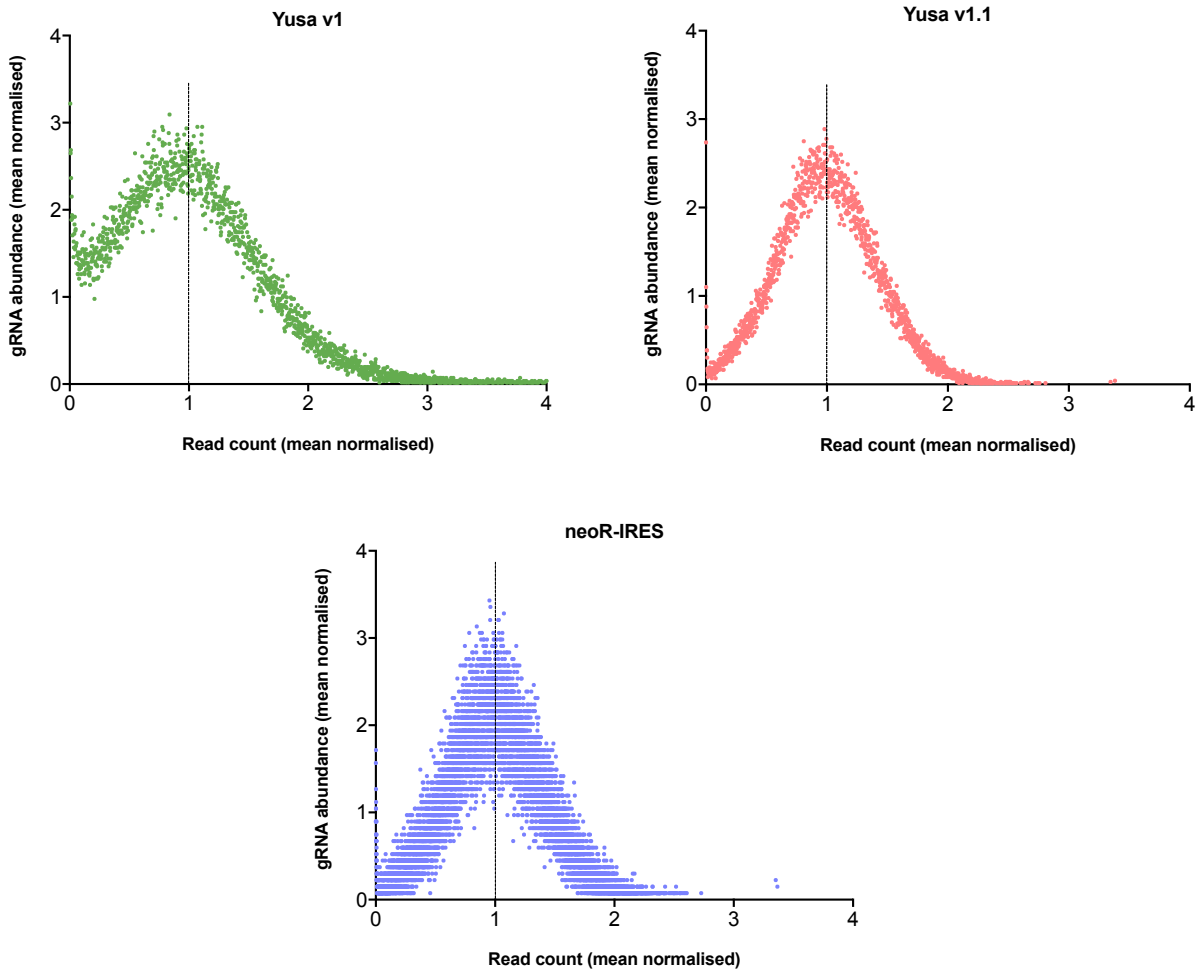



**Figure 2.13. Comparison of gRNA distribution between libraries.** Abundance of gRNAs (normalised against the mean) with each read count (normalised against the mean) is shown for the Yusa v1, Yusa v1.1 and neoR-IRES libraries. Read counts were obtained from sequencing each plasmid on a HiSeq 2500.

## 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%).

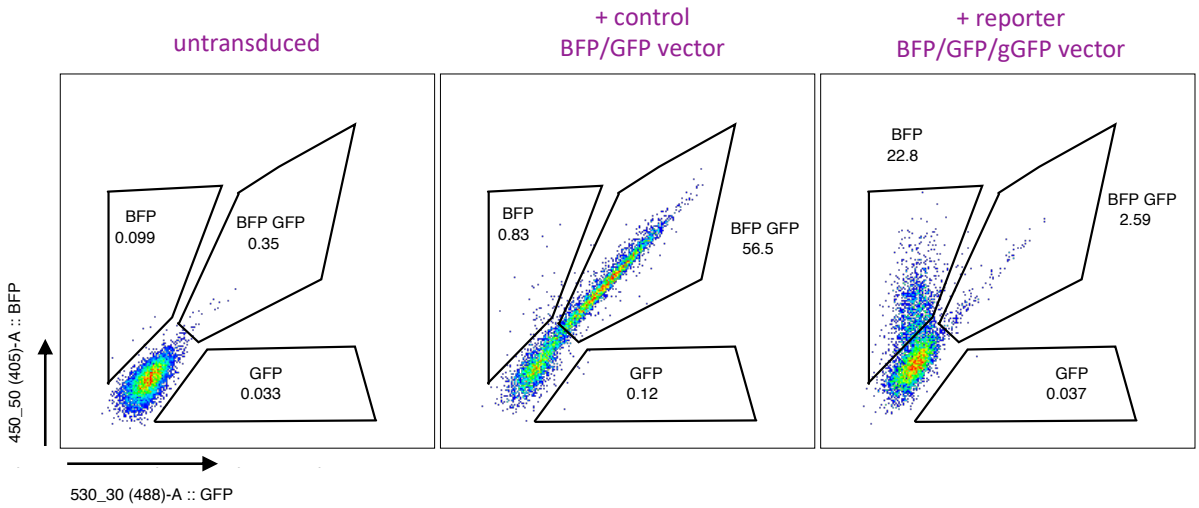



**Figure 2.14. Cas9 activity in iPSCs.** BOB-Cas9 cells were transduced with a control BFP/GFP vector (middle), a reporter BFP/GFP/gGFP vector (right) or untransduced (left). BFP and GFP expression were measured by flow cytometry at 3 days post-transduction. Untransduced and control cells were used for 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.

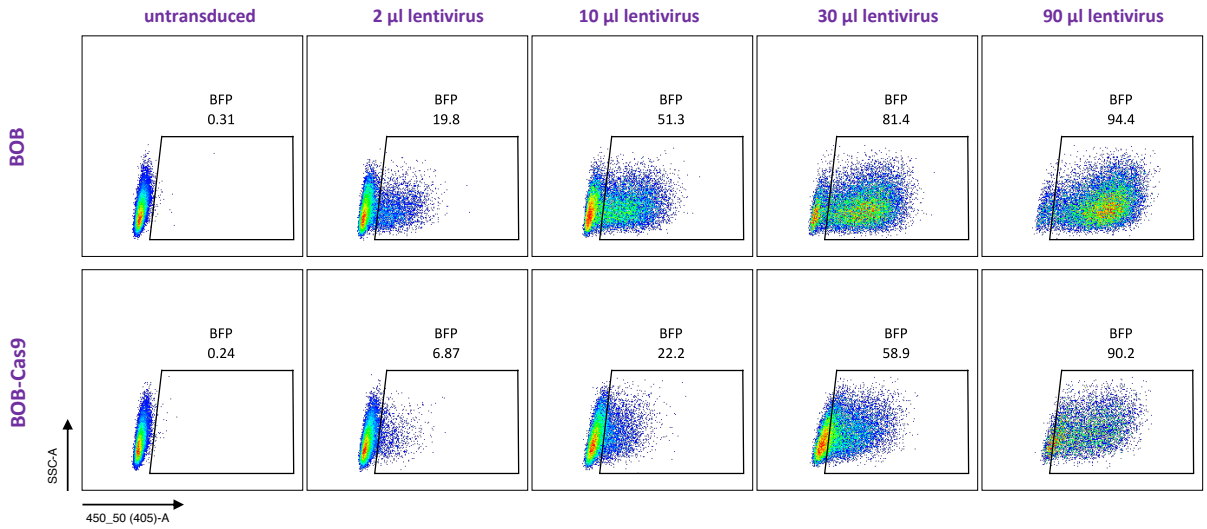



**Figure 2.15. Library titration in BOB vs BOB-Cas9.** Flow cytometry plots showing BFP expression (450_50 (405)-A) in BOB and BOB-Cas9 cells 3 days post-transduction with various volumes of the neoR-IRES library lentivirus. Untransduced cells were used as a control for both lines.

### 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.

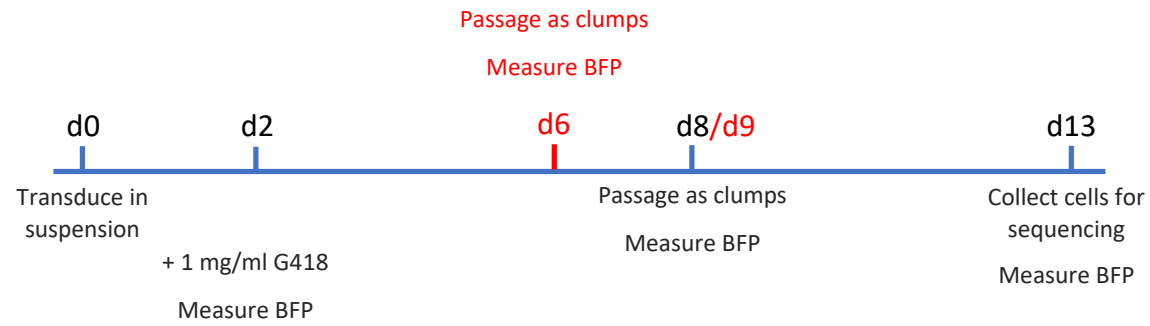



**Figure 2.16. iPSC screening protocol.** Stable Cas9 cells were transduced in suspension with the neoR-IRES library lentivirus then seeded. After 48 hours, 1 mg/ml of G418 was added. A sample of cells (from a small dish seeded in parallel at the point of transduction) was taken for flow cytometry analysis 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

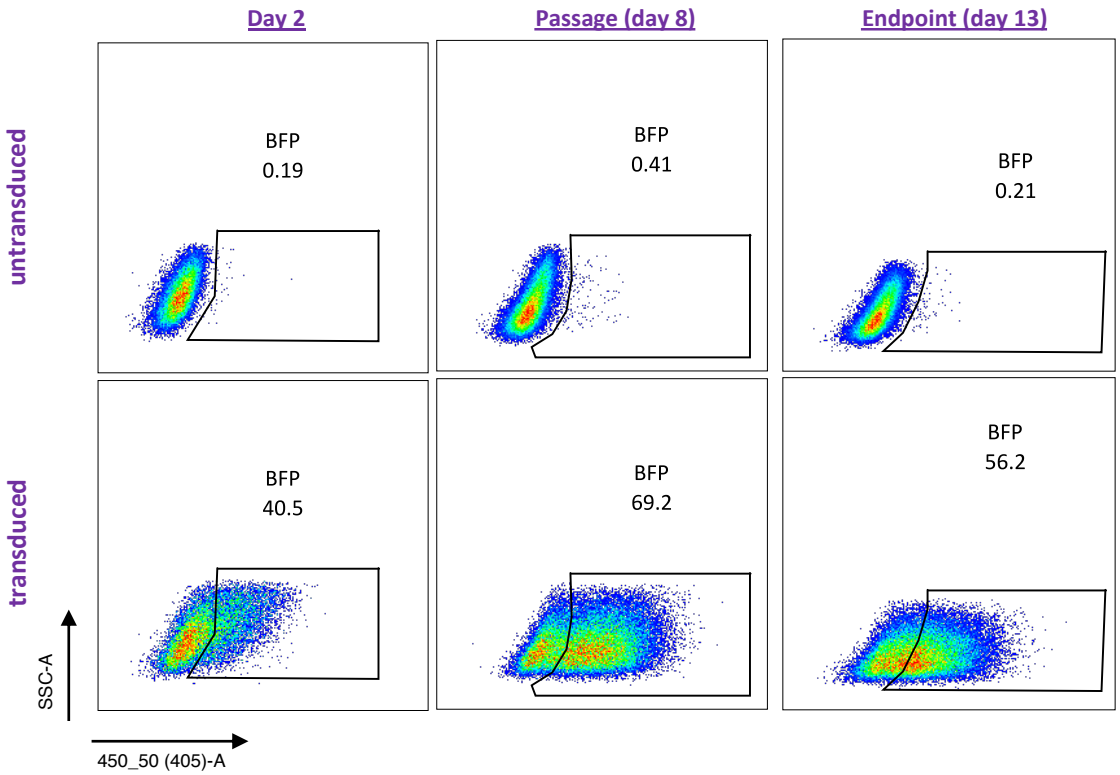



**Figure 2.17. BFP expression throughout BOB-Cas9 screen.** Flow cytometry plots showing BFP expression (450_50 (405)-A) in BOB-Cas9 cells on day 2, day 8 and day 13 post-transduction with the neoR-IRES library. Untransduced cells were used as a control for gating. This analysis was performed for each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

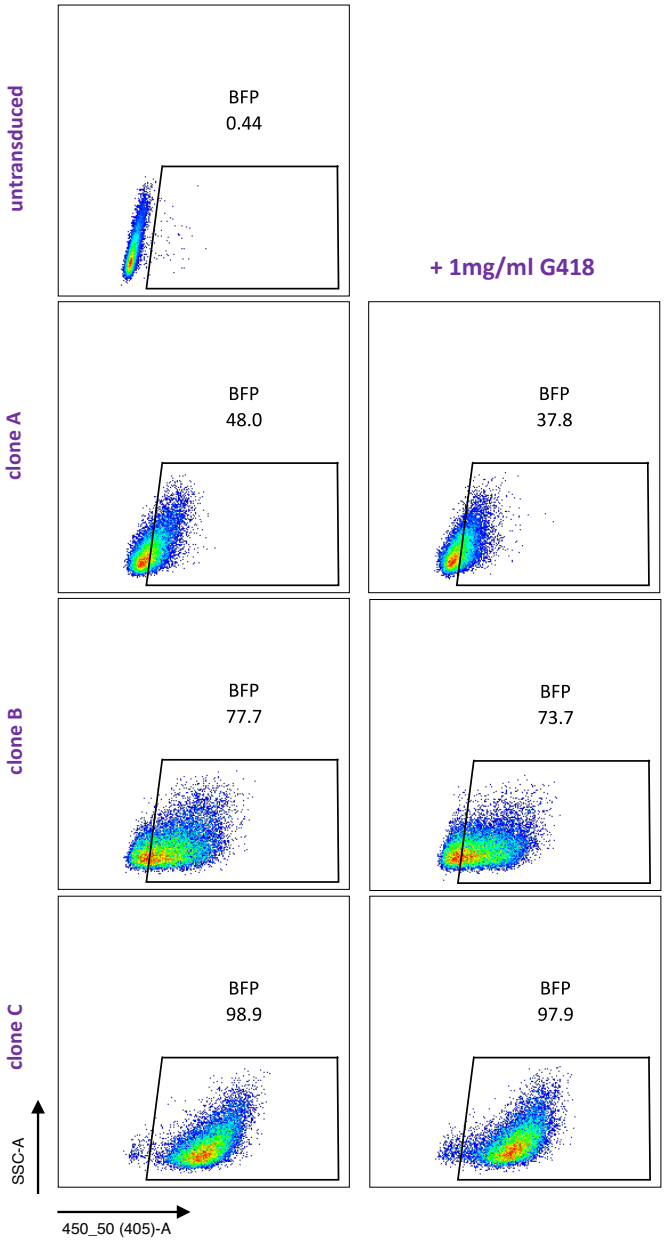



**Figure 2.18. Variable BFP expression in clonal transduced lines.** Single-cell colonies were picked from populations of BOB-Cas9 cells that were either untransduced or had been screened with the neoR-IRES library. All transduced cells had been selected with 1 mg/ml G418 throughout the screen. Each 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.

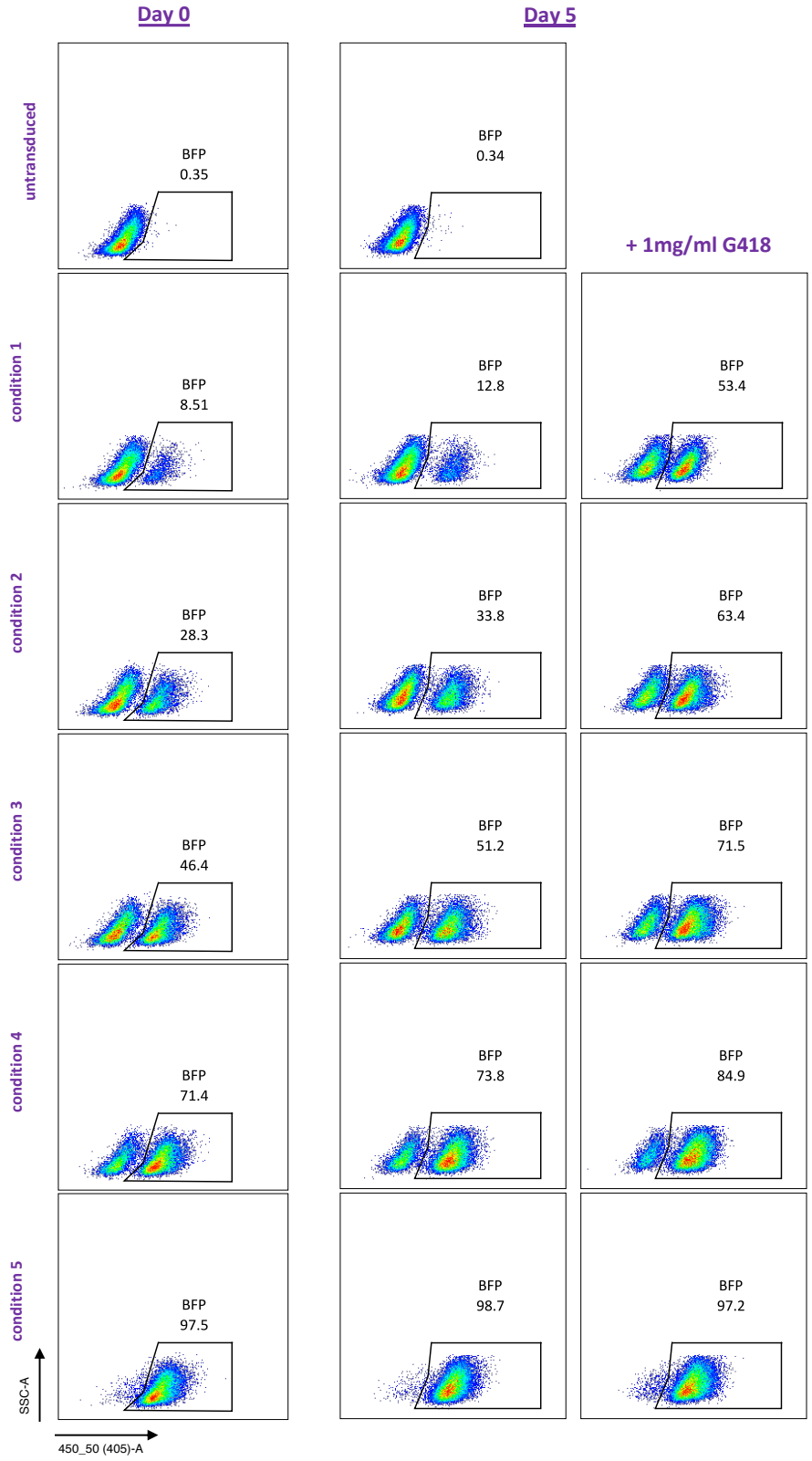



**Figure 2.19. Efficiency of G418 selection.** Untransduced BOB-Cas9 cells were mixed with a clonal transduced BOB-Cas9 line that was previously confirmed to have ~100% BFP expression. Populations containing 10%, 30%, 50%, 70% and 100% of the transduced line were prepared. A fraction of each population 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).

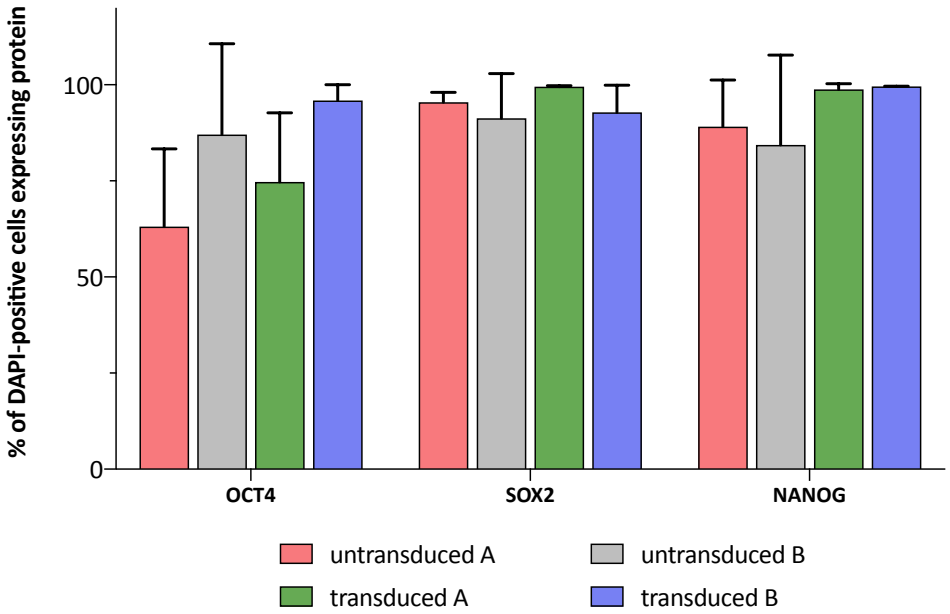



**Figure 2.21. Expression of pluripotency markers in BOB-Cas9 cells.** Cells from two replicates of the BOB-Cas9 neoR-IRES screen were maintained separately after the screen endpoint (transduced A/B). Untransduced cells from each replicate population were cultured in parallel (untransduced A/B). Cells 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.

# Chapter 3

# Analysis of iPSC CRISPR/Cas9 screens

## 3.1 Introduction

A variety of methods are available to analyse CRISPR/Cas9 knockout screens. There are no examples in the literature of genome-wide screens in iPSCs and few in hESCs, but a vast amount of data has been published for hundreds of cancer cell line screens. Using insights gained from these studies, we analysed data from our screens in the parental BOB and KO derivatives. We initially performed basic quality control measures and tested screen performance, using published datasets as a reference. To evaluate reproducibility, we compared the data from all lines including the results of biological replicates of both the parental and *TP53* KO lines. Using various filtering strategies, we identified candidate SLIs for all 15 of the TSGs studied. In addition to exploring genes that were required for cell fitness in the iPSCs, we also identified genes that, when lost, appeared to provide a proliferative advantage.

### 3.1.1 Aims of this chapter

- To assess the quality of our iPSC screen data and evaluate the performance of the screens in terms of recall of known fitness genes.
- To assess the reproducibility of the screens by comparing the results of all lines, with a particular focus on biological replicates of the same lines.
- To filter for dependencies that were specific to KO lines and identify candidate SLIs.
- To investigate enrichment of genes in the iPSC screens.

## 3.2 Screen data quality control

### 3.2.1 Sequencing coverage

In total, 24 screens were performed as part of this project: I performed 5 screens and a further 19 were performed by Rebecca McRae and Verity Goodwin (CGaP, WSI). As discussed in Section 2.7.2, these screens were performed using the same protocol and library but added an additional passage was added by CGaP. Each screen was carried out in technical triplicate, with the exception of the *FAT1* KO line which had only two replicates due to unexplained cell death in one replicate. All samples were processed in the same way and gRNAs were sequenced on a HiSeq 2500, with 6 samples multiplexed per run. There was slight variation in sequencing depth between different samples and runs, which was accounted for by normalisation prior to further analysis. An average of $4.07 \times 10^7$ read counts per replicate mapped to the gRNA library (Fig. 3.1). This was equivalent to ~400x coverage of the library, with no samples dropping below 200x coverage (Fig. 3.1).

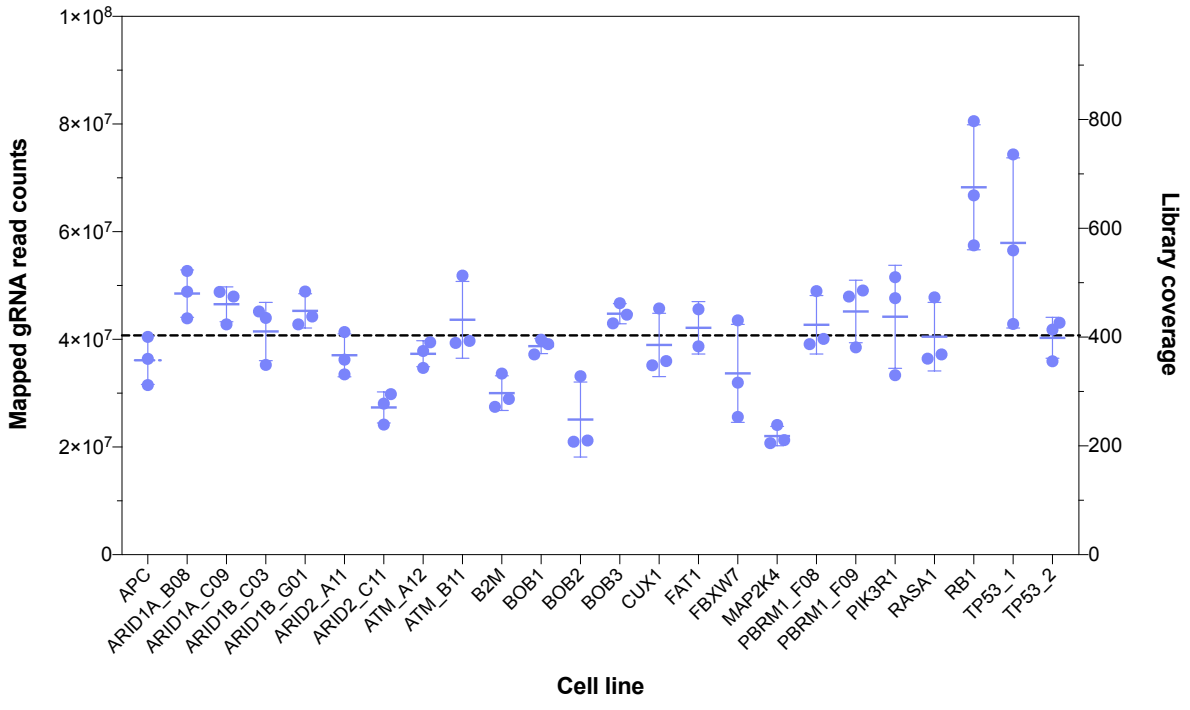



**Figure 3.1. Sequencing coverage across all iPSC screens.** PCR was performed to amplify gRNAs present in the genomic DNA of each screened cell line. gRNAs were then sequenced on a HiSeq 2500 and mapped to the library. The number of mapped reads is plotted for each replicate in every screened cell line (left y-axis). The corresponding library coverage that was achieved is also shown (right y-axis). The dotted line indicates the mean across all samples. BOB, BOB_2 and BOB_3 refer to replicate screens of the parental BOB line. TP53 and TP53_2 refer to replicate screens of the *TP53* KO line.

### 3.2.2 Enrichment of non-targeting controls

When looking at the initial gRNA read count data prior to analysis, it was noted that there was an enrichment of non-targeting control (NTC) gRNAs in the screen samples compared to the library plasmid. The $\log_2$(fold-change) in abundance between plasmid and screen for the targeting gRNAs was distributed around 0, whereas NTC gRNAs were enriched (an average of 1.2 for the screens shown in Fig. 3.2). The NTC gRNAs do not target any region in the genome, hence Cas9 would not induce DSBs in cells expressing these gRNAs. In line with our previous observation of Cas9-induced toxicity (Section 2.7.1), we hypothesised that cells expressing NTC gRNAs had a proliferative advantage due to the lack of DNA damage. Therefore, widespread depletion of cells expressing targeting gRNAs but not those expressing NTC gRNAs would cause this observed enrichment of the controls. This effect was also observed in the *TP53* KO line screen (Fig. 3.2), suggesting that depletion of TP53 was not sufficient to avoid this toxicity. Other studies have since reported similar findings (discussed in Section 5.2.2). We decided to remove the NTC gRNAs from all further analysis as they skewed the results rather than acting as controls.

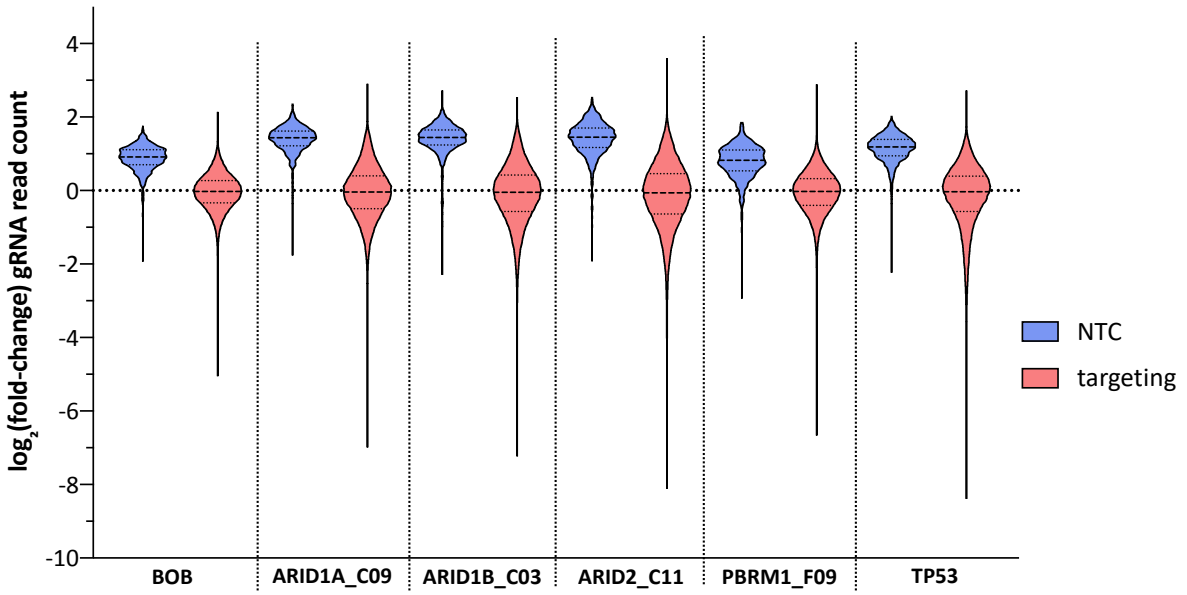



**Figure 3.2. Fold-changes of non-targeting and targeting gRNAs.** The $\log_2$(fold-change) in read count between the screen and the library plasmid was calculated for NTC gRNAs and all other targeting gRNAs. Results are shown for screens in the BOB, ARID1A_C09, ARID1B_C03, ARID2_C11, PBRM1_F09 and TP53 cell lines. The values shown were calculated using the average read count of the technical replicates in each screen.

### 3.2.3 Correlation between technical replicates

The correlation of gRNA read counts between technical replicates was measured for each cell line using Pearson's Correlation Coefficient. The average R value was calculated for each line, with a median of 0.82 across all screens (range 0.6-0.9) (Fig. 3.3a). In a recent study of 324 cancer cell lines using the Yusa v1.1 library, a median R of 0.8 was achieved.[109] However, it was noted that this correlation was not sufficient to distinguish between replicates of the same cell line and any two random cell lines. To gain a better measure of reproducibility, in that study they selected gRNAs that had an average pairwise Pearson's Correlation of > 0.6 across all screens when comparing the count fold-changes at the screen endpoint vs the plasmid library. For each replicate, they then calculated the average gene-level fold-change for only the genes targeted by these 'reproducible gRNAs'. Pearson's Correlation of these fold-changes was assessed between all replicates across all screens, and a reproducibility threshold was defined that would allow distinction between replicates of the same cell line and random lines. We repeated this analysis on our screens and identified 279 gRNAs that were reproducible (average R > 0.6). However, this measurement assumes that the cell lines being compared are independent, but the cell lines we screened were almost genetically identical. Unsurprisingly, the correlation between cell lines was still not distinct from that observed between replicates. Based on the reproducibility threshold identified in the cancer cell line screens (R=0.68), the majority of our screens passed this quality control test (Fig. 3.3b). A more thorough approach could be to repeat this analysis using a combined set of data from the iPSC screens and cancer cell line screens. Comparing the iPSCs to independent lines may provide a more reliable threshold to assess replicate reproducibility.

Notably, the initial 5 screens had the lowest correlations in both analyses, suggesting that the screen quality was improved by the addition of a passage. This may be due to a reduction in the cell death that occurred when the cells approached confluency. We also noted that replicate A of the ARID1A_C09 screen negatively impacted the average correlation. Replicates 1 and 2 had R = 0.83 for all gRNA counts and R = 0.76 for reproducible gRNA fold-changes. During this screen, the level of transduction in replicate 1 was higher and more cell death was observed, which may explain the lower correlation. We therefore decided to exclude replicate 1 of the ARID1A_C09 screen from further analysis.

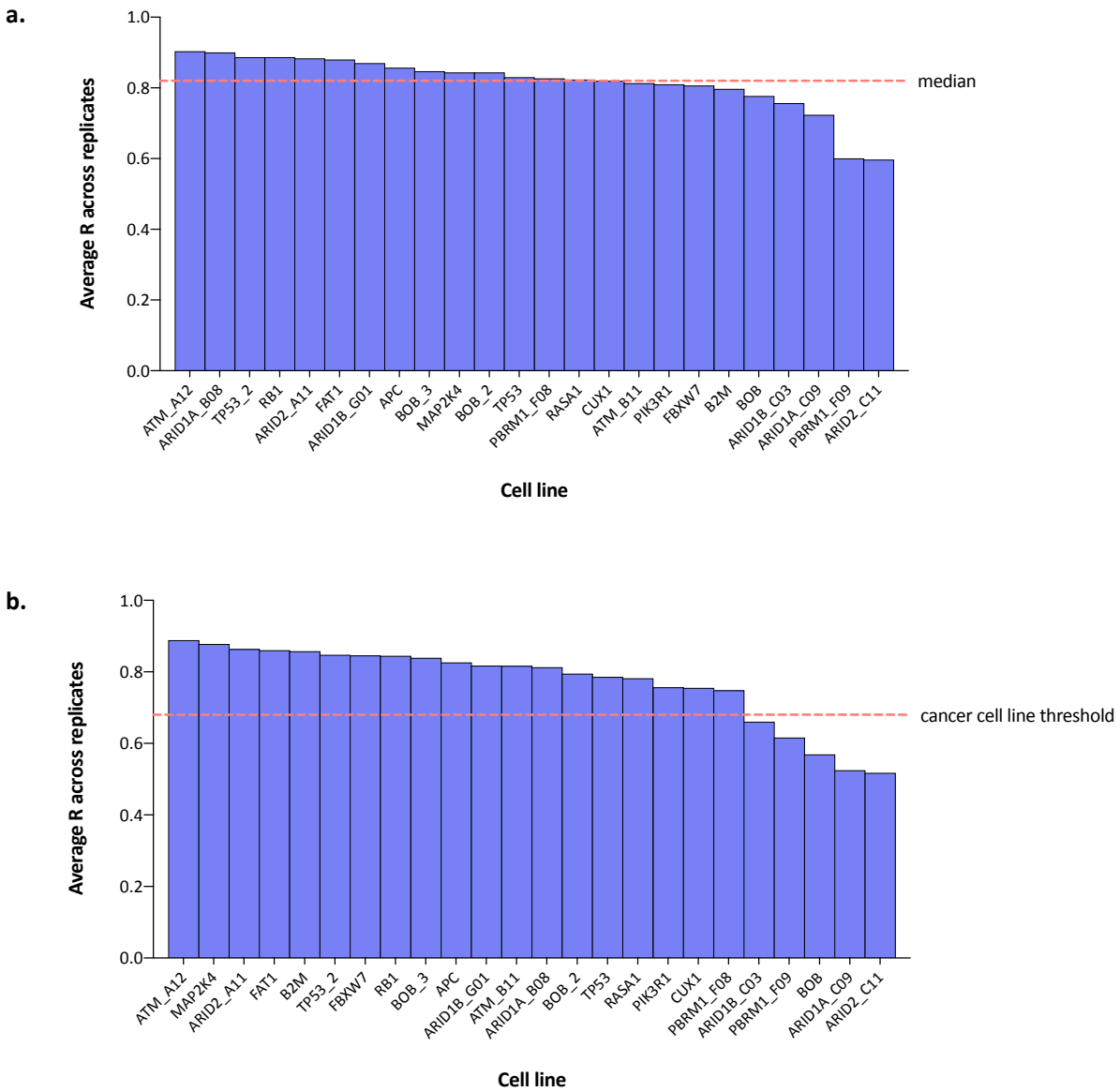



**Figure 3.3. Correlation between screen technical replicates. a)** Pearson's Correlation values were computed between technical replicates of each cell line using gRNA read counts. The average R value for each screen is shown, with the median represented by a dotted line, **b)** Pearson's Correlation values were computed between technical replicates of each cell line using the gene-level fold-changes only for reproducible gRNAs. The average R value for each screen is shown. The dotted line represents the reproducibility threshold that was defined by Behan *et al.* based on screens in cancer cell lines.[109]

## 3.3 Identifying effects of gene loss on cell fitness

Several methods are available for analysis of CRISPR/Cas9 knockout screen data. We chose two of the most well-established methods, described below, to identify genes that were significantly depleted in our screens. These analyses allowed us to compare our data to previously published cancer cell line screens that were processed in the same way after screening with the same gRNAs.[109] Our library had an average of 5 gRNAs/gene but, for a subset of ~2000 cancer-related genes, there were 10 gRNAs/gene. This may have caused a bias towards genes that were better represented and we were unclear how to account for this, so we removed the data for all additional gRNAs before further analysis.

### 3.3.1 Bayesian Analysis of Gene EssentiaLity

Bayesian Analysis of Gene EssentiaLity (BAGEL) is an algorithm developed to analyse genetic perturbation screens using *a priori* known training sets of 'essential' and 'nonessential' genes.[114] The 360 essential genes were defined based on their essentiality in at least 50% of a set of shRNA screens, and constitutive expression in a panel of cell lines. Using the same panel, the 972 nonessential genes were defined as those which generally lacked expression in these lines. Firstly, median-ratio normalisation is performed on all raw read counts to account for differences in sequencing coverage. A $\log_2$(fold-change) is then calculated for each gRNA, comparing the abundance at the screen endpoint to that in the library plasmid. The average $\log_2$(fold-change) for each gRNA is calculated across screen replicates. BAGEL then uses the fold-change distribution of all gRNAs targeting the essential and nonessential genes (Fig. 3.6) to calculate the likelihood that a given gRNA belongs to either set, based on the observed fold-change. The output of this probability calculation is a value termed the Bayes Factor (BF); every gRNA is assigned a BF. A recently published R implementation (BAGELR) calculates the gene-level BFs by taking the average of the gRNA values.[109] The original Python version of BAGEL calculated a sum rather than an average value.[114] A positive BF indicates that the gene is likely to be essential for cell fitness.

Bayes Factors were computed using BAGELR for all of our screens (Appendix A.5). To determine statistical significance, a threshold of 5% False Discovery Rate (FDR, 1- Precision) was defined for each screen. Genes were assigned a scaled BF (Appendix A.6), which was calculated by subtracting the BF at the 5% FDR threshold for that screen from the original BF. Any gene with a scaled BF > 0 was considered to be significantly depleted.

### 3.3.2 Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout

Model-based Analysis of Genome-wide CRISPR/Cas9 Knockout (MAGeCK) is an algorithm which, unlike BAGEL, identifies significant differences in gRNA abundance using no prior knowledge.[116] Median-normalisation is performed on the read counts to account for differences in sequencing depth across replicates and conditions, and the gRNA variance is estimated. Replicates can be analysed together, with a mean read count calculated for every gRNA. Using a negative binomial model, MAGeCK then determines whether the abundance of a gRNA is significantly different between the control and treatment; in our analysis this was the library plasmid vs the screen endpoint. A robust ranking aggregation (RRA) algorithm is used to rank gRNAs by the p-value obtained from the negative binomial model (Fig. 3.4).

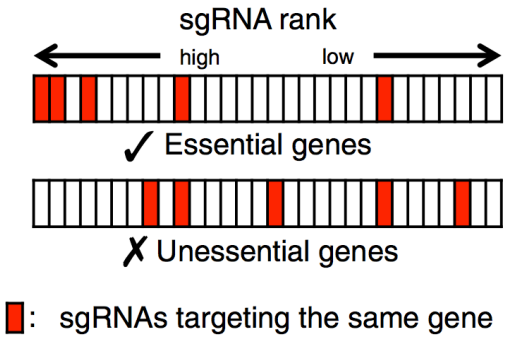



**Figure 3.4. Robust rank aggregation.** A negative binomial model is used to determine the significance of any change in gRNA abundance compared to control. gRNAs are then ranked based on their p-value. If a gene is essential, gRNAs targeting it should be ranked highly more often than expected. If a gene is nonessential, gRNAs targeting it should be uniformly distributed. Figure taken from [116].

If a gene has no effect on fitness, the assumption is that gRNAs targeting this gene will be evenly distributed in the rankings. If several gRNAs targeting a gene are ranked higher than expected, this gene would be considered significant. Each gene is assigned a p-value and an FDR is computed using the Benjamini-Hochberg method. MAGeCK can be used for bi-directional analysis; from one screen it can identify genes whose knockout impairs cell fitness (negative selection) and genes whose knockout induces cell proliferation (positive selection). Genes under negative selection would have a significant depletion of gRNAs compared to the control. Those under positive selection would have significant enrichment of gRNAs compared to the control. MAGeCK analysis was performed on all of our screens to calculate depletion values for every gene (Appendix A.7). A threshold of negative FDR 0.1 was applied to identify significant hits.

# 3.4 Assessing screen performance

## 3.4.1 Receiver Operating Characteristic & Precision-Recall curves

As a measure of the sensitivity and specificity of the screens, receiver operating characteristic (ROC) and precision-recall (PrRc) curves were computed (Fig. 3.5). This was done using the gene-level count fold-changes, with the average taken across technical replicates for each screen. Using the pre-defined sets of BAGEL essential and nonessential genes, these analyses can indicate how well a screen performed. ROC curves plot sensitivity (i.e. true positive rate) against specificity (1 - false positive rate). The area under the curve (AUC) is a measure of how accurately the essential and nonessential genes were identified as distinct groups. If these genes cannot be separated, the AUC would be 0.5. A screen with 100% specificity and sensitivity would have an AUC of 1. PrRc curves plot recall (i.e. true positive rate, the same as sensitivity) against precision (i.e. positive predictive value). Precision and specificity are slightly different: precision measures how many of the predicted positives are actually true positives; specificity measures how many of the expected negatives are called as negative. Similar to the ROC curve, a high AUC for the PrRc curve indicates good performance with high precision and recall. Based on these models, performance across all of the screens was fairly consistent. The median area under the ROC curve was 0.91 (Fig. 3.5a) and area under the PrRc curve was 0.87 (Fig. 3.5b). These results were similar to those obtained in the Behan *et al.* (2019) study of cancer cell lines (0.92 and 0.9, respectively).[109]

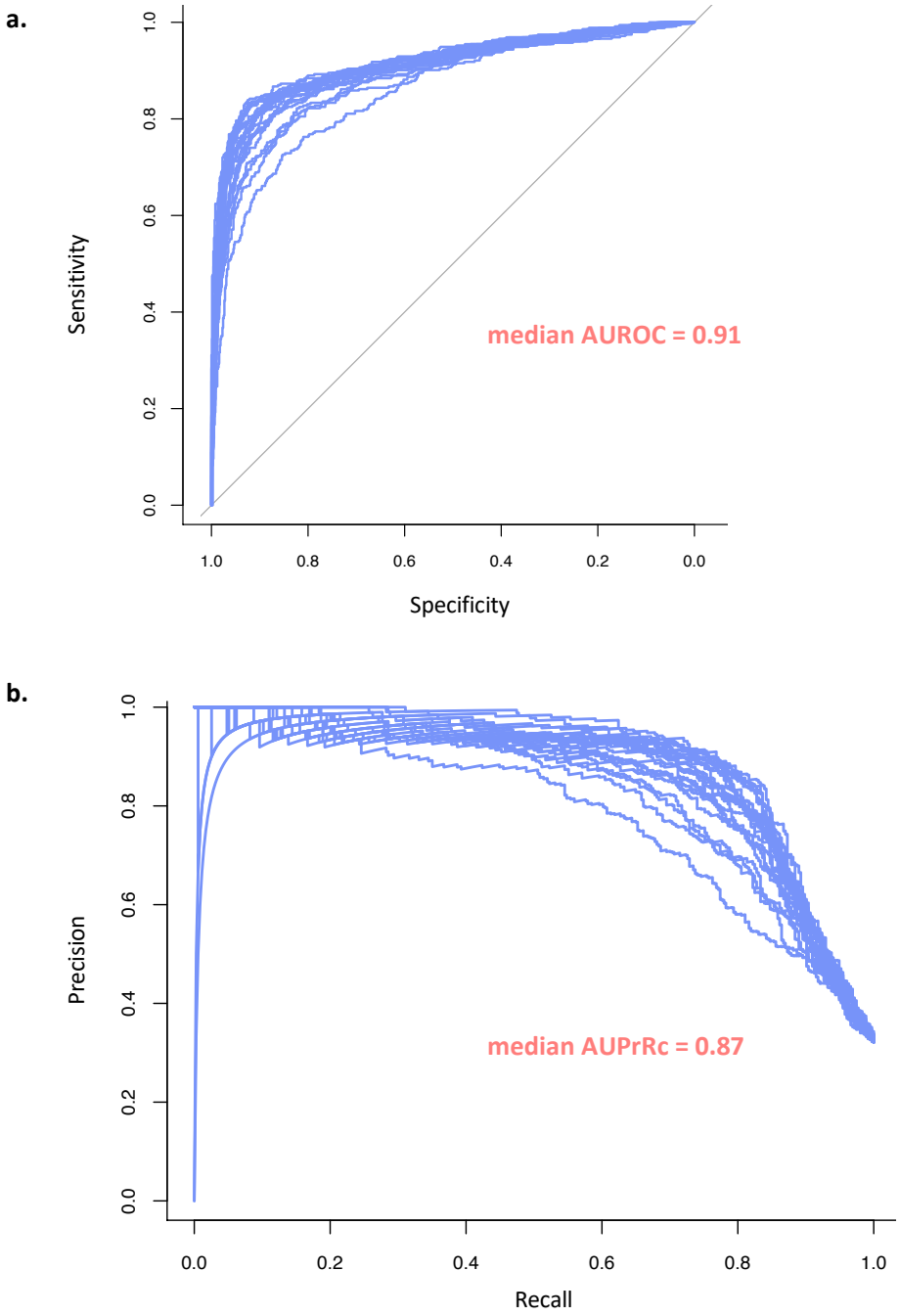



**Figure 3.5. Assessment of screen performance.** ROC **(a)** and PrRc **(b)** curves were plotted for all 24 screens. These were obtained by classifying the pre-defined BAGEL essential (n=354) and nonessential (n=747) genes using the gene-level fold-changes. The median AUC across all screens is shown for both. These were computed using the ROC and PrRc functions in the CRISPRcleanR package.[255]

### 3.4.2 Distributions of BAGEL essential and nonessential genes

Another indicator of screen performance is the separation between the results for the BAGEL essential and nonessential genes. We analysed the distribution of gene-level fold-changes (Fig. 3.6) and scaled BFs (Fig. 3.7) for each of these sets across all screens. There was a slight separation between the essential and nonessential genes but the overlap was high, although this was improved in the scaled BF distributions. It may be the case that some of these genes were not essential in this iPSC line, as these genes were identified in immortalised cell lines. We also analysed the fold-change distribution for genes encoding ribosomal proteins, which we would expect to be vital for cell function regardless of the cell type (Fig. 3.6). These were slightly more depleted and separated from the nonessential genes compared to the essentials, but in some cases (e.g. PBRM1_F09) there was still high overlap. It appeared that depletion was simply not large enough to clearly separate the sets, suggesting this was most likely an issue with screen performance. For reference, the fold-changes (Fig. 3.6) and scaled BFs (Fig. 3.7) for an ovarian cancer cell line, A2780ADR, are shown. This cell line was screened by Behan *et al.* (2019) using the same library and the data was processed in the same way.[109] This cell line passed all quality control tests and had high AUC values for the ROC (0.93) and PrRc (0.93) analyses, thus we considered it to be a good representation of a high-quality screen. For this cell line, the fold-changes for essential and ribosomal genes spread further and were more distinct from the nonessential population than in our iPSC screens. There was also a greater separation between the scaled BFs for the essential and nonessential genes in the A270ADR screen, with the majority of essential genes being correctly called as essential. Of all the lines we screened, the results for the *TP53* knockout line were most similar to this cancer cell line, indicating that this was the best performing screen.

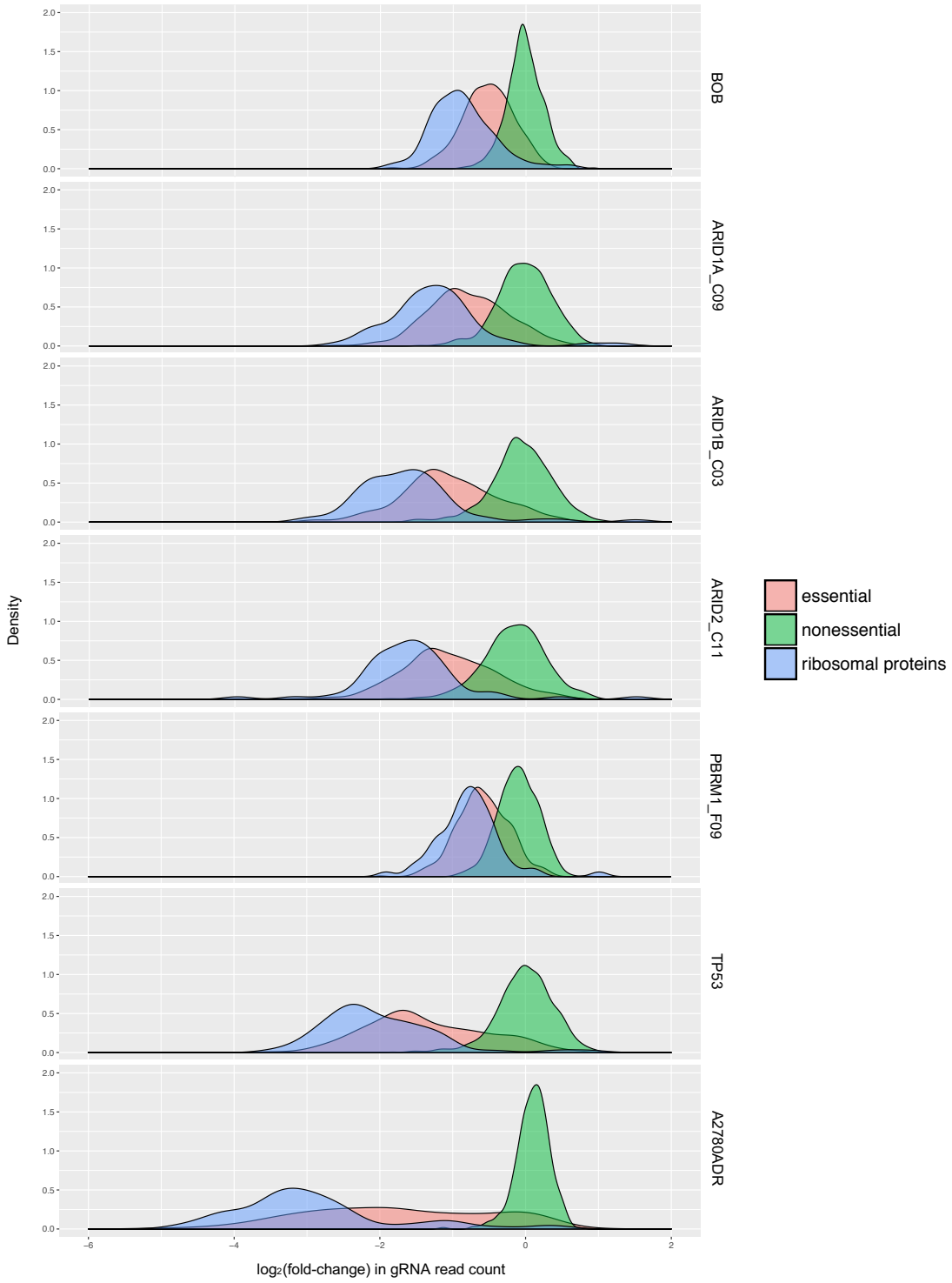



**Figure 3.6. Distribution of fold-changes for BAGEL essential and nonessential genes.** The fold-change of each gRNA was calculated for every screen replicate relative to the library plasmid. The average of the replicates was calculated for each cell line, and then a gene-level fold-change was calculated by taking average of the values for all of the gRNAs targeting each gene. The distribution of $\log_2$(fold-changes) of the BAGEL essential (n=354) and nonessential (n=747) genes are plotted for the parental BOB, ARID1A_C09, ARID1B_C03, ARID2_C11, PBRM1_F09 and TP53 screens. For comparison, results are also shown for the A2780ADR ovarian cancer cell line, screened by Behan *et al.* using the same library.[109] Distributions are also shown for the genes that encode ribosomal proteins.

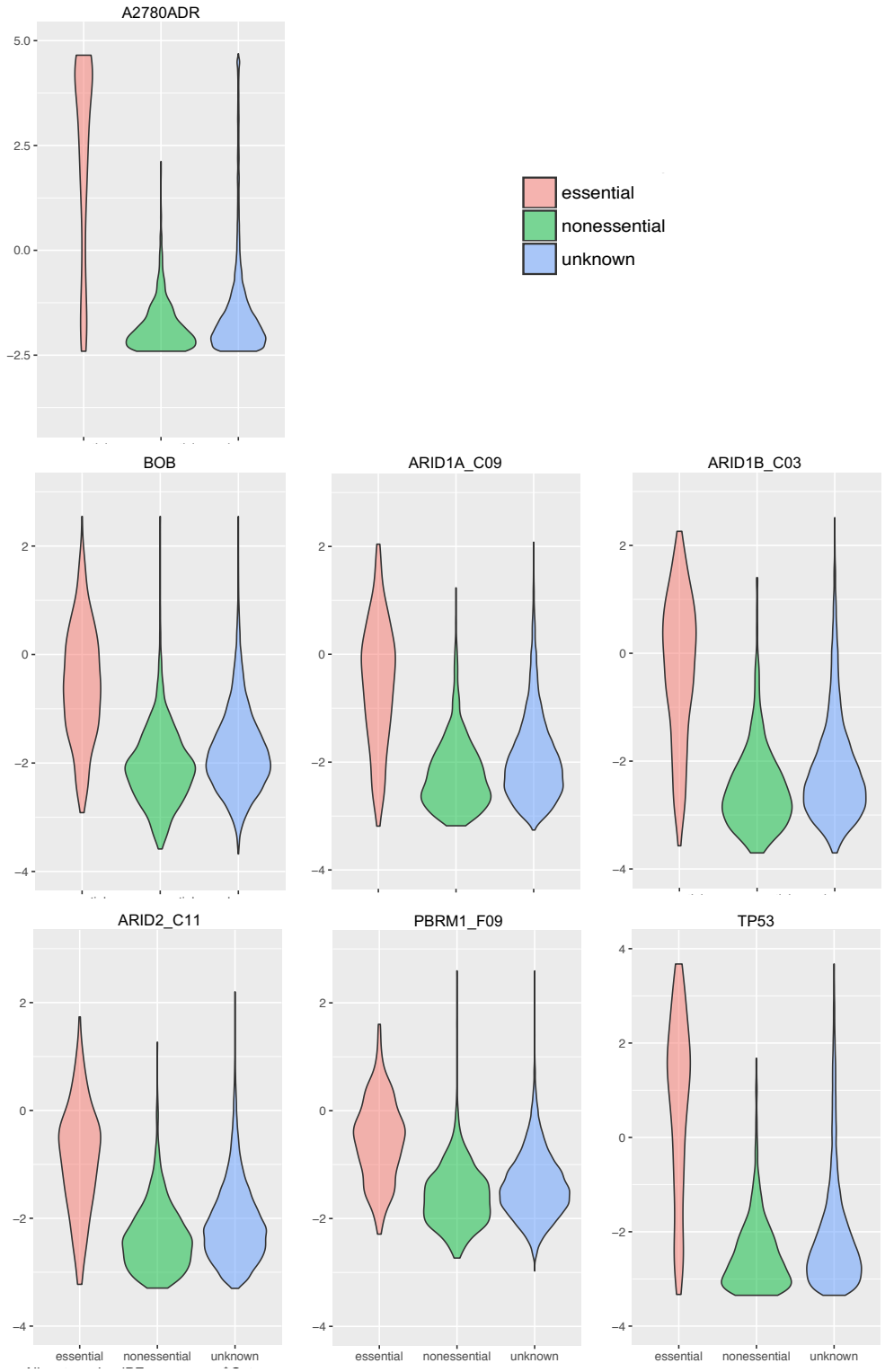



**Figure 3.7. Distribution of scaled Bayes Factors.** BAGEL was applied to compute BFs for all genes in every screen, calculating an average across replicates. The values were scaled using a 5% FDR threshold, with a value > 0 representing a significant hit. Plots show the distributions of the scaled BFs for the BAGEL essential (n=354) and nonessential (n=747) genes, and all other genes (unknown, n=16,906). Results are shown for the parental BOB, ARID1A_C09, ARID1B_C03, ARID2_C11, PBRM1_F09 and TP53 screens. For comparison, results are also shown for the A2780ADR ovarian cancer cell line, screened by Behan *et al.* using the same library[109].

### 3.4.3 Recall of known fitness genes

Whilst the ROC and PrRc curve models performed well, the distribution analyses indicated that we may have had issues with detecting known fitness genes and that performance was variable. In the Behan *et al.* (2019) pan-cancer study, a set of 552 pan-cancer core fitness genes were identified[109], providing an additional reference set for comparison. Using both the BAGEL and MAGeCK analyses outputs, we assessed exactly how many of the BAGEL essential genes and pan-cancer core fitness genes were called as hits in our screens (Table 3.1). Although these gene sets were identified in cancer cell lines, the pan essentiality across many cell types suggests that they are likely required for general cell fitness and survival, regardless of tissue type or tumourigenicity. Considering this, it was expected that many of these should also be essential in our iPSCs. As our KO lines were all derived from the same parental, we also expected that there should be a large overlap in the essential genes identified. Consistently more of the pan-cancer core fitness genes were detected than the BAGEL essential genes (Table 3.1). This may be because the BAGEL gene list was derived from shRNA screen data rather than CRISPR/Cas9 screen data. However, in general the number of essential genes identified in the parental and KO iPSC screens was highly variable. In several screens, including one of the parental screens (BOB_2), there was a high recall of core fitness genes. This suggested that many of the core genes identified in cancer cell lines were also essential in this iPSC line, but the ability to consistently detect them was impeded by variable screen performance.

**Table 3.1. Recall of pre-defined essential genes in iPSC screens.** MAGeCK and BAGEL were applied to identify significantly depleted genes in each screen. The % of pan-cancer core fitness genes (n=552) and BAGEL essential genes (n=354) that were called as hits are shown, based on the results of both analyses. Increasing colour intensity reflects increasing % recall.

| | MAGeCK | | BAGEL | |
|---|---|---|---|---|
| **Cell line** | **Core fitness** | **BAGEL essential** | **Core fitness** | **BAGEL essential** |
| APC | 63% | 48% | 63% | 47% |
| ARID1A_B08 | 36% | 25% | 15% | 12% |
| ARID1A_C09 | 28% | 21% | 47% | 35% |
| ARID1B_C03 | 44% | 34% | 65% | 48% |
| ARID1B_G01 | 42% | 31% | 49% | 38% |
| ARID2_A11 | 57% | 42% | 30% | 22% |
| ARID2_C11 | 15% | 10% | 26% | 17% |
| ATM_A12 | 67% | 52% | 66% | 50% |
| ATM_B11 | 52% | 39% | 10% | 9% |
| B2M | 72% | 54% | 91% | 66% |
| BOB | 18% | 13% | 41% | 32% |
| BOB_2 | 47% | 36% | 80% | 60% |
| BOB_3 | 47% | 35% | 13% | 12% |
| CUX1 | 60% | 45% | 83% | 62% |
| FAT1 | 29% | 21% | 51% | 39% |
| FBXW7 | 39% | 28% | 39% | 28% |
| MAP2K4 | 53% | 40% | 58% | 43% |
| PBRM1_F08 | 43% | 33% | 61% | 46% |
| PBRM1_F09 | 6% | 5% | 30% | 24% |
| PIK3R1 | 40% | 29% | 33% | 21% |
| RASA1 | 61% | 47% | 80% | 60% |
| RB1 | 68% | 52% | 53% | 40% |
| TP53 | 70% | 53% | 95% | 68% |
| TP53_2 | 69% | 51% | 91% | 66% |

## 3.5 Comparison of MAGeCK and BAGEL

The inconsistent detection of established core fitness genes indicated that there may be high variability in the overall essentiality profiles of these cell lines. As described previously, both MAGeCK and BAGEL were used to identify genes that were significantly depleted in the parental and KO lines. We calculated the total number of significant hits in each screen and compared the results from both analyses (Table 3.2). For many lines, the number of hits called by MAGeCK and BAGEL varied considerably. However, the overlap of the genes that were identified was generally high. In screens where BAGEL detected less than MAGeCK, the majority of the hits detected by BAGEL were also identified by MAGeCK, and vice versa. It is not surprising that these analyses identified different hits and it was reassuring to see that many of these were shared. However, it is not clear why there was no trend in the variability: MAGeCK detected more hits in some screens but BAGEL detected more in others. This was also reflected in the detection of *a priori* known essentials discussed earlier, with inconsistent variability between both analyses. The results of these analyses are dependent on the chosen significance threshold. Here, a threshold of FDR 0.1 was used for MAGeCK and FDR 0.05 was used for BAGEL. These can be adjusted to alter the stringency of the analysis; increasing the stringency too far will result in identification of very few hits and decreasing it may introduce noise and cause a high false positive rate. The most robust hits are likely to be those that were identified by both analyses, although this may lead to an increased false negative rate. For subsequent analysis, we considered the outputs from both MAGeCK and BAGEL rather than excluding data.

**Table 3.2. Number of significantly depleted genes identified in iPSC screens.** MAGeCK and BAGEL were applied to identify significantly depleted genes in each screen. The number of genes called as hits by MAGeCK using an FDR of 0.1 and by BAGEL using an FDR of 0.05 are shown for each cell line. The overlap of hits that were identified by both analyses is also shown.

| Screen | MAGeCK hits | BAGEL hits | Overlap |
|---|---|---|---|
| APC | 1068 | 863 | 713 |
| ARID1A_B08 | 490 | 156 | 116 |
| ARID1A_C09 | 413 | 671 | 343 |
| ARID1B_C03 | 715 | 962 | 589 |
| ARID1B_G01 | 654 | 637 | 435 |
| ARID2_A11 | 1009 | 353 | 319 |
| ARID2_C11 | 232 | 343 | 144 |
| ATM_A12 | 1235 | 995 | 848 |
| ATM_B11 | 872 | 95 | 86 |
| B2M | 1244 | 1506 | 1109 |
| BOB | 264 | 625 | 228 |
| BOB_2 | 758 | 1379 | 714 |
| BOB_3 | 820 | 151 | 128 |
| CUX1 | 981 | 1384 | 875 |
| FAT1 | 475 | 832 | 419 |
| FBXW7 | 671 | 541 | 383 |
| MAP2K4 | 930 | 885 | 689 |
| PBRM1_F08 | 675 | 905 | 565 |
| PBRM1_F09 | 86 | 528 | 75 |
| PIK3R1 | 680 | 429 | 326 |
| RASA1 | 1029 | 1244 | 868 |
| RB1 | 1258 | 683 | 613 |
| TP53 | 1178 | 1656 | 1100 |
| TP53_2 | 1079 | 1468 | 991 |

# 3.6 Assessing screen reproducibility

## 3.6.1 Comparison of biological replicates

All screens were carried out in technical triplicate (or duplicate for the *FAT1* KO) with cells split into three populations at the passage prior to setting up the screen, and then transduced and maintained separately throughout. However, we considered that biological replicates may be more informative with regards to reproducibility. For the parental BOB line and the *TP53* KO line, screening was repeated weeks apart using cells thawed from different vials. The data was analysed as described previously, and the overlap of the results was assessed (gene lists are provided in Appendix A.8).

I carried out the initial parental BOB screen and two further biological replicates were performed by CGaP using their adapted protocol. Using MAGeCK, only 185 genes were significantly depleted in all replicates (Fig. 3.8). A further 314 hits were detected in both the second and third screens, but not in the first. Using BAGEL, fewer genes were identified in BOB_3 but almost all of them overlapped with BOB_2 (Fig. 3.8). Similarly, the majority of hits from BOB were also found in BOB_2. The detection of core fitness and BAGEL essential genes was higher in BOB_2 than in the others (Table 3.1). Thus, the incomplete overlap may be due to poorer performance in the BOB and BOB_3 screens.

The replicates of the *TP53* KO line (referred to as *TP53* and *TP53*_2) had a greater correlation, with a higher overlap between the hits identified using both BAGEL and MAGeCK (Fig. 3.9). BAGEL detected more significantly depleted genes in both replicates, but these included almost all of the genes detected by MAGeCK. When the overlap of both replicates from both analyses were compared, 847 genes were found to be significantly depleted, in comparison to only 62 in the parental overlap.

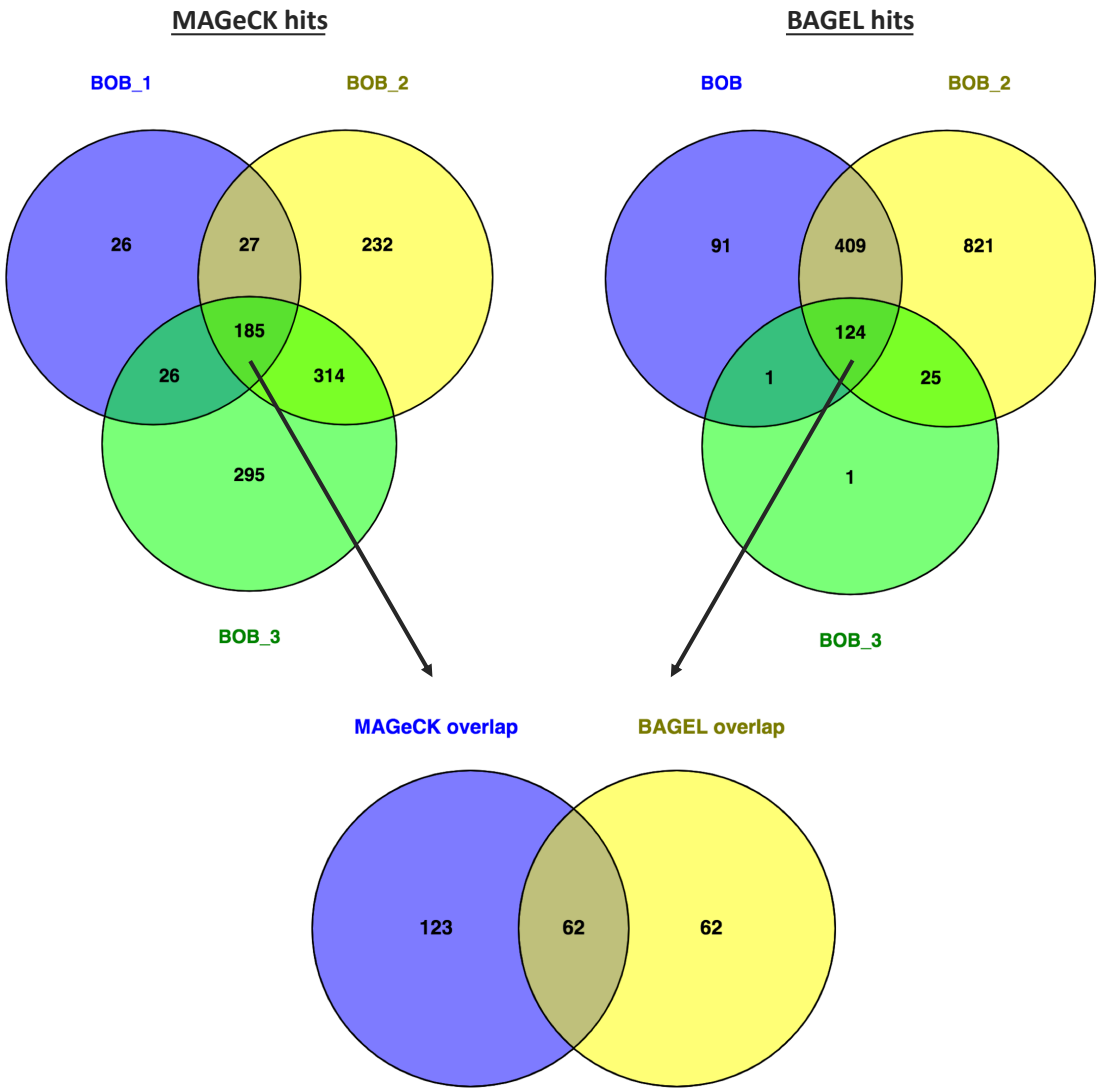



**Figure 3.8. Overlapping hits in biological replicates of the parental BOB screen.** The parental BOB cell line was screened three times, with technical triplicate in each case. MAGeCK and BAGEL were applied to identify genes that were significantly depleted compared to the library plasmid. The outputs for all three screens were compared to find common hits. The overlapping MAGeCK hits were compared with the overlapping BAGEL hits to assess the correlation of the two analyses. Diagram created using Venny.[256]

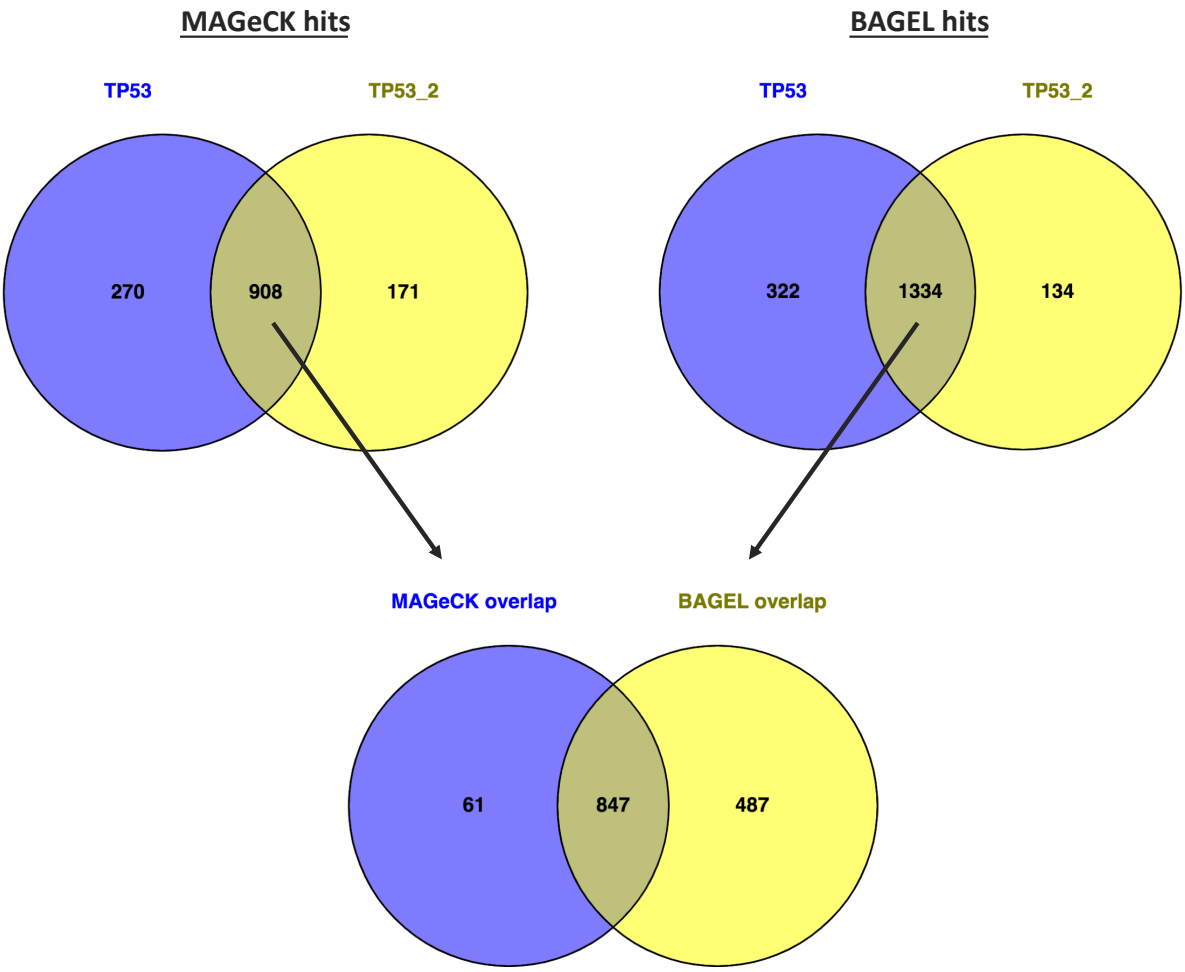



**Figure 3.9. Overlapping hits in biological replicates of *TP53* KO line screen.** A *TP53* KO derivative of BOB was screened twice, with technical triplicate in both experiments. MAGeCK and BAGEL were applied to identify genes that were significantly depleted compared to the library plasmid. The outputs for both screens were compared to find common hits. The overlapping MAGeCK hits were compared with the overlapping BAGEL hits to assess the correlation of the two analyses. Diagram created using Venny.[256]

### 3.6.2 Comparison across all screens

As all of the cell lines differed by only a single genetic change, we considered that results from screening different lines could also act in some way as biological replicates. Thus, we compared the results across all screens as another measure of reproducibility and to further define core fitness genes in the BOB iPSC line. We anticipated that the majority of the hits would be shared, however very few were identified in every screen (25 using BAGEL, 17 using MAGeCK) and 18% of genes were called only once (Fig. 3.10). We analysed all of the genes that were significant in 20-23 (out of 24) screens to determine whether specific screens consistently failed to identify common hits (Fig. 3.11). Using the BAGEL output, 5 screens in particular (ARID2_C11, PBRM1_F09, BOB_3, ARID1A_B08 and ATM_B11) consistently failed to detect hits that were identified by the majority of the other screens. The results were slightly different using the MAGeCK output, with the ARID2_C11, PBRM1_F09, BOB and ARID1A_C09 screens accounting for the majority of missed hits. Whilst BAGEL and MAGeCK differed, the results correlated well with their respective detection of known essentials/fitness genes. The screens that failed to detect the highest number of common hits also had the poorest recall (Table 3.1). In line with the previous data, this indicated that the screens were not highly reproducible. This limited our ability to accurately define core essential genes for the BOB iPSC line.

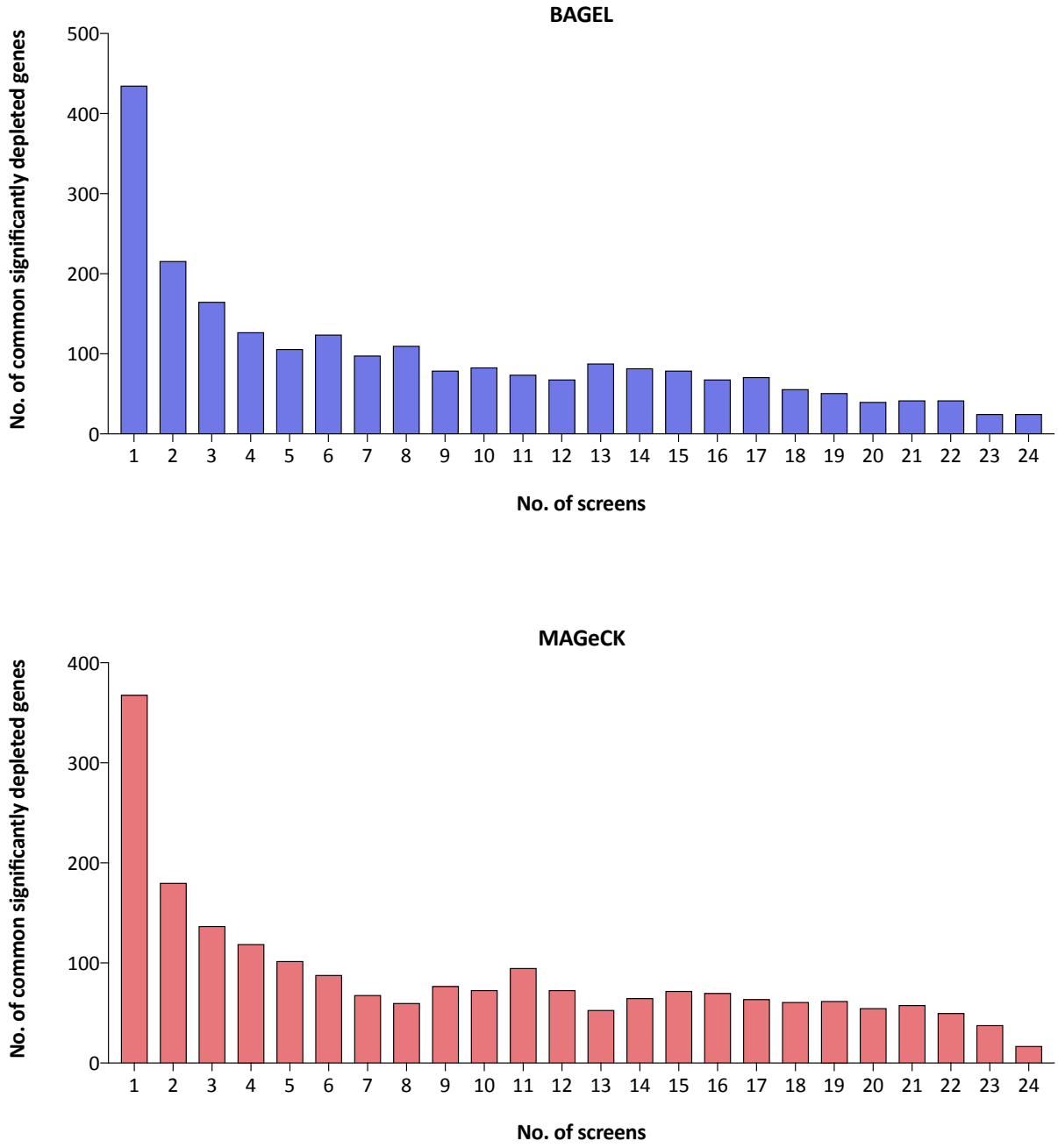



**Figure 3.10. Frequency of significantly depleted genes across all iPSC screens.** Across all screens, a total of 2371 genes were identified as significantly depleted by BAGEL, and 2105 by MAGeCK. Some of these hits were specific to one screen, but many were identified in multiple screens. These plots show the frequency with which genes were identified by one or more screens.

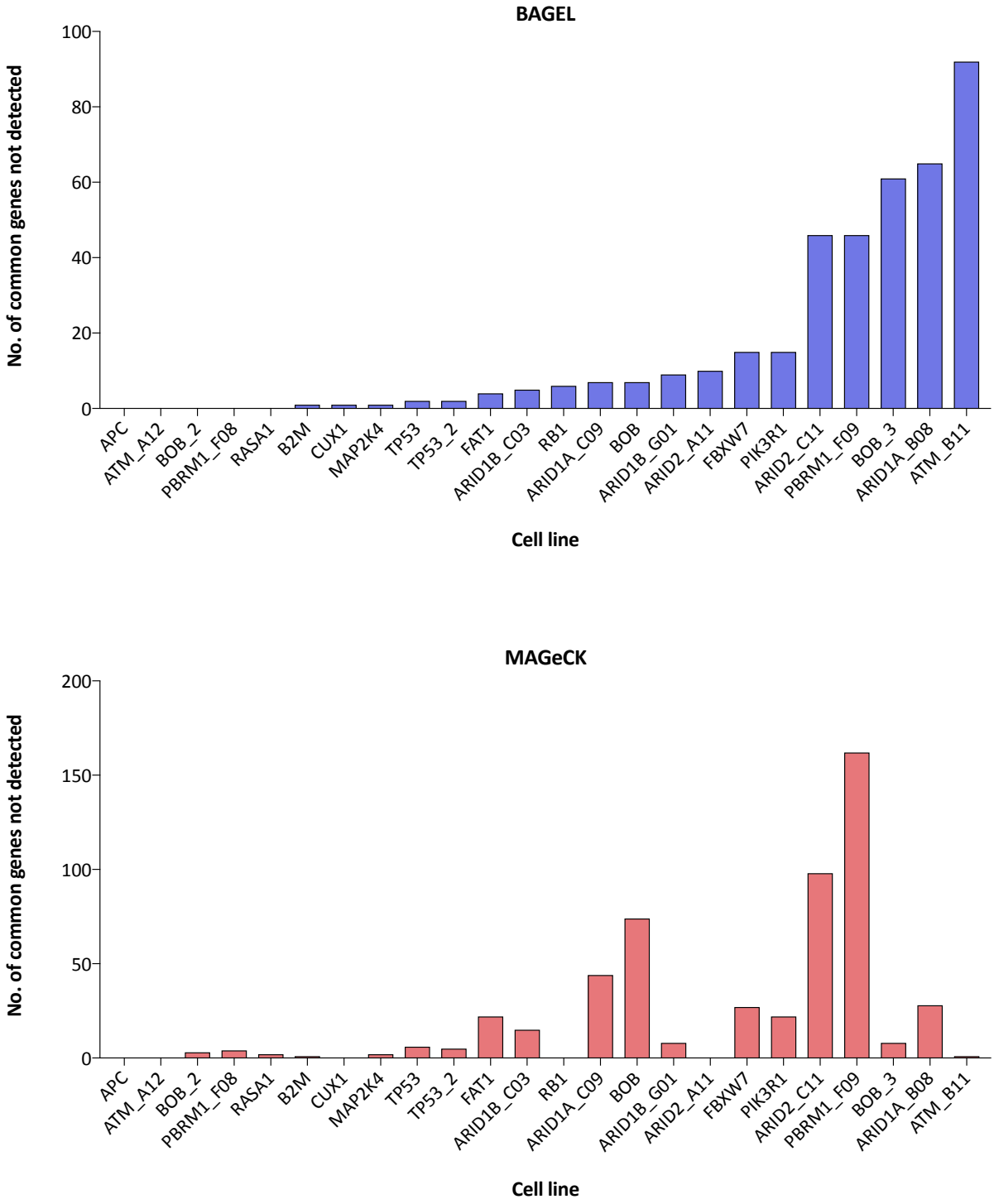



**Figure 3.11. Detection of common gene hits in iPSC screens.** Considering only genes that were significantly depleted in 20-23 of the iPSC screens, this plot shows the number of these genes that were not detected by each screen. Results are shown for both BAGEL and MAGeCK outputs.

## 3.7 Filtering for KO-specific dependencies

Despite the variability in the data, the results were not completely inconsistent and there were indications that known essentials could be detected. Therefore, we decided to continue using these data for our primary aim of identifying SLIs. To do so, we were interested in finding genes that were specifically essential in the TSG KO lines but not in the parental line. To identify KO-specific hits, we compiled lists of genes that were significantly depleted in each KO line and removed any genes that were also significant in the parental. For simplicity, I will discuss filtering using only the BAGEL outputs, but the same could be performed on the MAGeCK outputs or the overlap of both. As data was obtained for 3 biological replicates of the parental, various strategies were possible. One approach was to remove genes that were hits in every parental screen, ensuring that only high confidence hits in the parental were discarded. Another option was to exclude all genes that came up in any of the parental screens, accounting for the fact that detection of some genes may have been affected by some replicates performing poorly. A final strategy was to filter based only on the hits from BOB_2, which appeared to be the highest quality screen. Table 3.3 indicates the number of KO-specific genes identified in each screen using all of these filtering approaches. These gene lists are provided in Appendix A.9.

The screen which had the highest recall of established core fitness genes was ‘TP53’, closely followed by the biological replicate ‘TP53_2’ (Table 3.1). With this in mind, I have selected these screens to provide an example of the KO-specific gene lists. Table 3.4 shows the scaled BFs for the 20 top-ranking genes in the ‘TP53’ screen, excluding genes that were hits in any of the parental BOB screens and removing established core fitness genes. Of these, 19/20 genes were also significantly depleted in the ‘TP53_2’ replicate screen. It has been previously shown that one of these genes, *ATR,* is synthetically lethal with *TP53* (as reviewed by Qiu *et al.*, 2018).[257] This warrants further validation of the other genes, particularly those with a higher ranking, to identify novel SLIs with *TP53*.

In Chapter 4 I will discuss more advanced filtering of results from screens in the PBAF/BAF gene KO lines, and subsequent experimental validation of these genes.

**Table 3.3. Number of KO-specific screen hits.** The scaled BFs computed by BAGELR analysis of all screens were used to identify significantly depleted genes (scaled BF > 0). Genes that were significantly depleted in all BOB screens OR in at least one BOB screen OR in the BOB_2 screen, were removed from the list of significant hits in each KO line screen. The number of remaining genes are shown for each screen, based on each filtering strategy.

| Screen | Not in every BOB screen | Not in any BOB screen | Not in BOB_2 |
|---|---|---|---|
| TP53 | 1536 | 453 | 496 |
| TP53_2 | 1346 | 329 | 367 |
| ARID1A_C09 | 560 | 73 | 87 |
| ARID1A_B08 | 86 | 1 | 5 |
| ARID1B_C03 | 848 | 126 | 150 |
| ARID1B_G01 | 528 | 53 | 72 |
| ARID2_A11 | 238 | 5 | 9 |
| ARID2_C11 | 262 | 31 | 42 |
| PBRM1_F09 | 448 | 153 | 168 |
| PBRM1_F08 | 781 | 90 | 107 |
| FAT1 | 715 | 106 | 126 |
| APC | 741 | 60 | 77 |
| FBXW7 | 436 | 66 | 81 |
| ATM_A12 | 872 | 93 | 120 |
| ATM_B11 | 41 | 0 | 1 |
| MAP2K4 | 763 | 86 | 111 |
| PIK3R1 | 330 | 27 | 38 |
| RB1 | 564 | 32 | 43 |
| CUX1 | 1260 | 272 | 308 |
| RASA1 | 1121 | 208 | 239 |
| B2M | 1384 | 327 | 366 |

**Table 3.4. Candidate synthetic lethal partners of *TP53*.** The scaled BFs obtained by BAGELR analysis of the first *TP53* KO line screen were ranked from highest to lowest. The top 20 genes are shown, with scaled BFs noted for both biological replicates of this line.

| Gene | TP53 | TP53_2 |
|---|---|---|
| *SBNO1* | 2.12 | 0.51 |
| *HIST2H3A* | 2.11 | 2.62 |
| *SNAP23* | 2.06 | 1.57 |
| *HSD17B7* | 1.95 | 1.40 |
| *RINT1* | 1.90 | 1.53 |
| *ALDOA* | 1.88 | 1.25 |
| *HIRA* | 1.88 | 0.68 |
| *MED14* | 1.84 | 1.90 |
| *DR1* | 1.74 | 0.56 |
| *SOX2* | 1.73 | 0.54 |
| *MRPS12* | 1.71 | 1.11 |
| *ATR* | 1.67 | 0.36 |
| *ALG10* | 1.65 | -0.47 |
| *RPP21* | 1.65 | 0.90 |
| *MRPL23* | 1.62 | 0.34 |
| *PRIM1* | 1.58 | 0.87 |
| *HNRNPA1* | 1.54 | 0.36 |
| *PRR13* | 1.53 | 1.67 |

## 3.8 Gene enrichment in iPSC screens

As described previously, MAGeCK can also be applied to identify significantly enriched genes. Considering that we observed an enrichment of NTC gRNAs in our initial data analysis (Section 3.2.2), we were interested to see whether any targeted genes were also enriched. Eleven genes were recurrently significantly enriched in at least 50% of the screens (Fig. 3.12a) (all MAGeCK enrichment values are provided in Appendix A.10). Five of these encoded for proteins that are involved in activation of apoptotic signalling in response to DNA damage, including TP53. Thus, it is logical that knockout of these genes would provide a proliferative advantage by preventing an apoptotic response to Cas9-induced DSBs.

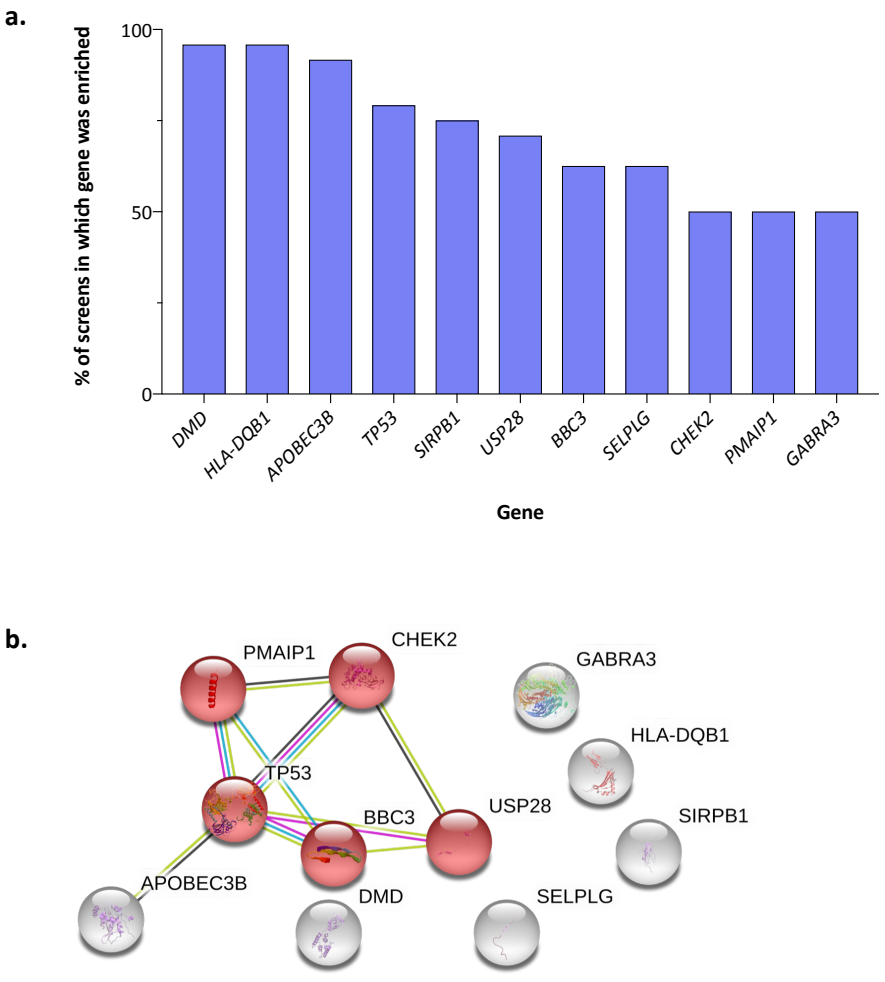



**Figure 3.12. Enriched genes in iPSC screens. a)** MAGeCK was applied for all iPSC screens and enrichment scores (positive FDR values) were computed for each gene. Genes with a positive FDR value < 0.1 were compared across all screens. Eleven genes were significantly enriched in at least 50% of the screens, **b)** The proteins encoded by these genes were analysed using STRING[258] to identify any interactions. Lines between the nodes indicate known/predicted interactions between proteins. Nodes highlighted in red are involved in apoptotic signalling pathways.

## 3.9 Summary

We performed CRISPR/Cas9 KO screens in 21 iPSC lines, with 3 biological replicates of the parental BOB line, 2 biological replicates of the *TP53* KO line and a single screen in 19 other KO lines. Due to an unexpected enrichment of non-targeting controls, which we attributed to Cas9 toxicity in the presence of targeting gRNAs, we had to remove these controls from our data. Initial quality control and screen performance tests produced results similar to published screens in cancer cell lines. However, further analysis indicated that the iPSC screens were highly variable and this made it difficult to confidently deduce which genes were essential for cell fitness. In some screens there was high recall of previously established core fitness genes, indicating that true positives could be identified but there was evidently a high risk of false negatives. Our aim was to identify genes that were specifically essential in the KO lines, and hence could be potential synthetic lethal partners. This variability made it challenging as there was a high possibility that genes identified in the KOs may be universally essential but were missed due to screen performance in the parental. Equally, low performance in the KO line screens may have led to false negatives and missed interactions. Despite this, we computed lists of KO-specific genes identified in each screen and hence have identified candidate SLIs. Results from the *TP53* KO line screen included a known SLI, which provides more confidence to the findings. However, validation is critical for any conclusions to be drawn from these datasets. As an aside, we also identified genes that were recurrently enriched in the screens. These may be informative with regards to iPSC biology and more specifically, their response to the CRISPR/Cas9 screening process.

# Chapter 4

# Validation of candidate synthetic lethal interactions with PBAF/BAF genes

## 4.1 Introduction

By screening parental and KO derivatives of an iPSC line, BOB, we identified a number of candidate synthetic lethal partners for four PBAF/BAF complex genes: *ARID1A, ARID1B, ARID2* and *PBRM1*. Many of these were specific to one gene, but some overlapped in two or more KO lines. Several synthetic lethal partners of *ARID1A* have been published but the other genes have not been well studied (as discussed in Section 1.4.4). We initially sought to validate an interaction between *ARID1A* and *ARID1B*, which has been previously reported in cancer cell lines. We then chose a panel of genes to validate from our screens, focusing on those that were significant hits in at least two KO lines, as these would be more widely applicable. We tested these interactions in iPSCs using a competitive growth assay. Validation was also carried out in a cancer cell line (HAP1), as we aimed to find hits that were relevant in cancer and not specific to iPSCs. Further analyses were performed to identify candidate interactions using published cancer cell line CRISPR/Cas9 screen datasets.

### 4.1.1 Aims of this chapter

- To investigate whether an SLI exists between *ARID1A* and *ARID1B* in BOB iPSCs.
- To select a gene panel for validation of candidate SLIs with PBAF/BAF genes.
- To validate the selected panel of genes in iPSCs using a competitive growth assay.
- To validate the selected panel of genes in HAP1 cells using a competitive growth assay.
- To analyse cancer cell screen data for dependencies associated with PBAF/BAF mutations.

## 4.2 *ARID1A/ARID1B* synthetic lethality

ARID1A and ARID1B are mutually exclusive subunits of the BAF complex and it has been reported that they have antagonistic functions.[163] They are the only known DNA-binding proteins in the complex. It has been shown that *ARID1A*-deficient cancer cell lines are specifically vulnerable to loss of *ARID1B*.[74,172,259] Helming *et al.* (2014) analysed genome-wide loss-of-function shRNA screen data from Project Achilles to identify dependencies caused by *ARID1A* mutations in cancer cell lines.[172] They identified *ARID1B* as one of the top genes essential for cell growth in mutant lines. Using ovarian cancer cell lines, they confirmed that depletion of ARID1B impaired proliferation and colony formation in *ARID1A*-mutant cells but not in WT cells. In the Project DRIVE study, deep shRNA screening of ~8000 genes was performed in 398 cancer cell lines and *ARID1B* was identified as a specific dependency in lines with inactivating *ARID1A* mutations.[74] Some cell lines and primary tumours harbour mutations in both genes, but in these cases at least one allele of either gene is retained.[172]

These findings suggest that in the absence of *ARID1A*, cells are dependent on *ARID1B* to maintain a functional BAF complex. In our iPSC screens, *ARID1B* was not essential in either *ARID1A* KO line (Fig. 4.1). Equally, *ARID1A* was not essential in either *ARID1B* KO line, although the values for *ARID1A* were closer to the significance thresholds than in most of the other screens. This indicated that there was no SLI between *ARID1A* and *ARID1B* in the BOB iPSC line.

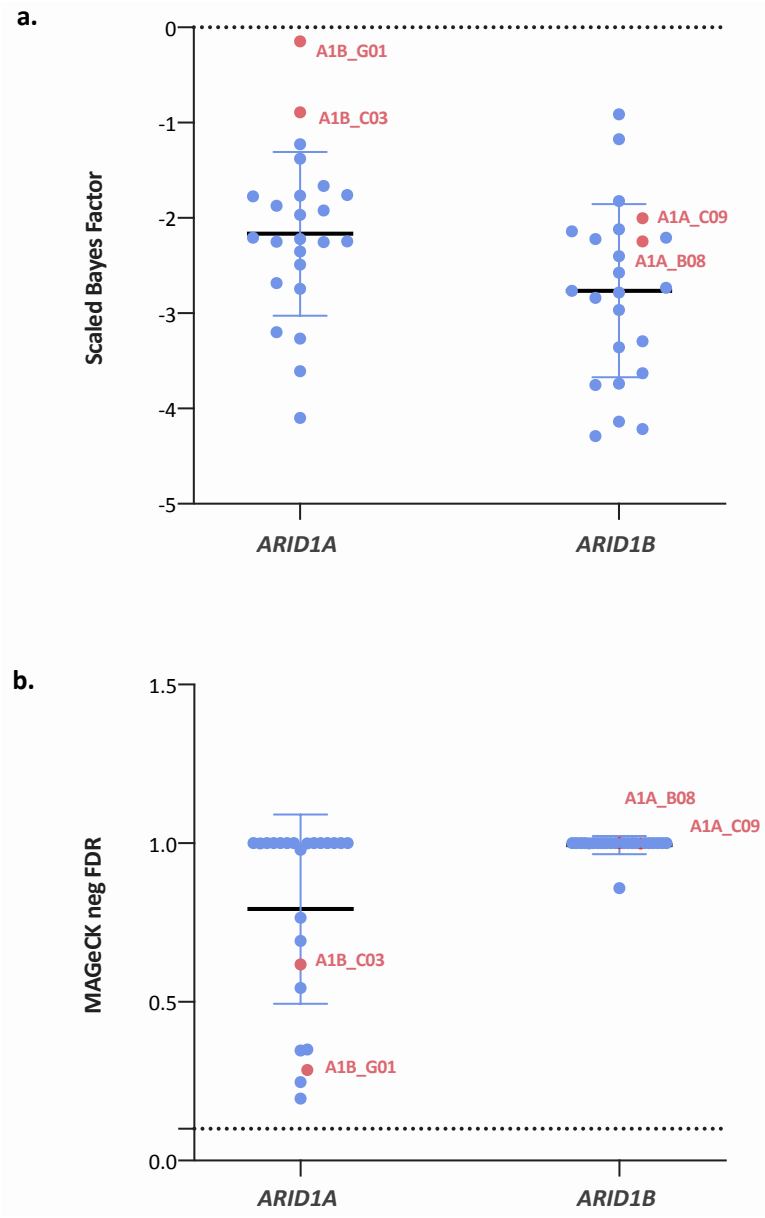



**Figure 4.1. Gene essentiality scores for *ARID1A* and *ARID1B* in iPSC screens.** Scaled BFs **(a)** and MAGeCK negative FDR values **(b)** are shown for *ARID1A* and *ARID1B* in the 24 iPSC screens (21 cell lines total with 3 replicates of the parental and 2 replicates of the *TP53* KO). The *ARID1A* values in the *ARID1B* KO lines (A1B_C03/A1B_G01), and the *ARID1B* values in the *ARID1A* KO lines (A1A_C09/A1A_B08), are highlighted in red. The dotted lines represent the thresholds for significance: using BAGELR, BFs were scaled based on an FDR of 0.05 so any value > 0 was considered significant; using MAGeCK, any negative FDR value < 0.1 was considered significant. Scaled BFs are detailed in Appendix A.6, and MAGeCK depletion values are in Appendix A.7.

### 4.2.1 Experimental validation of *ARID1A/ARID1B* SLI in iPSCs

To confirm the lack of interaction observed in the screens, a fluorescence-based competition assay was performed. gRNAs were selected from the neoR-IRES library used for the screens, cloned into the Yusa v1.1 library backbone and packaged into lentiviruses (as described in Section 7.16.2). BOB-Cas9 cells were transduced simultaneously with two lentiviruses containing a gRNA targeting *ARID1A* and a gRNA targeting *ARID1B* (as described in Section 7.16.3.1). As a control, each of these targeting gRNAs was also transduced alongside a gRNA targeting *AIPL1* or *ACCSL*. These genes were chosen as neither had a significant effect in the genome-wide screens. *ARID1A* and *AIPL1* gRNAs were cloned into a version of the backbone expressing BFP; *ARID1B* and *ACCSL* were cloned into in a backbone expressing mCherry. Cells were transduced at an MOI that generated four populations: untransduced, BFP positive, mCherry positive, and double positive. By measuring the abundance of each population at day 2 and day 14 post-transduction, any growth effects caused by the gRNAs could be assessed. The relative abundance of each population was calculated by normalising against the untransduced population. To assign a value for the growth phenotype, the $\log_2$(fold-change) of relative abundance was calculated between day 14 and day 2. The expected growth phenotype of knocking out two genes that do not interact was calculated as the sum of the phenotype of both single knockouts (based on the principles of the Bliss independence model[260]). If the genes were synthetic lethal, the double KO should have had a more negative growth phenotype than this expected value. All single gRNAs caused a negative growth phenotype, likely due to cell toxicity as a result of Cas9-induced double-strand breaks (Fig. 4.2). There was no significant difference between the observed and expected phenotypes for any of the control double KOs (*ARID1A+AIPL1, ARID1B+ACCSL, AIPL1+ACCSL*). However, there was also no significant effect when the *ARID1A* and *ARID1B* targeting gRNAs were combined, supporting the results of the screens.

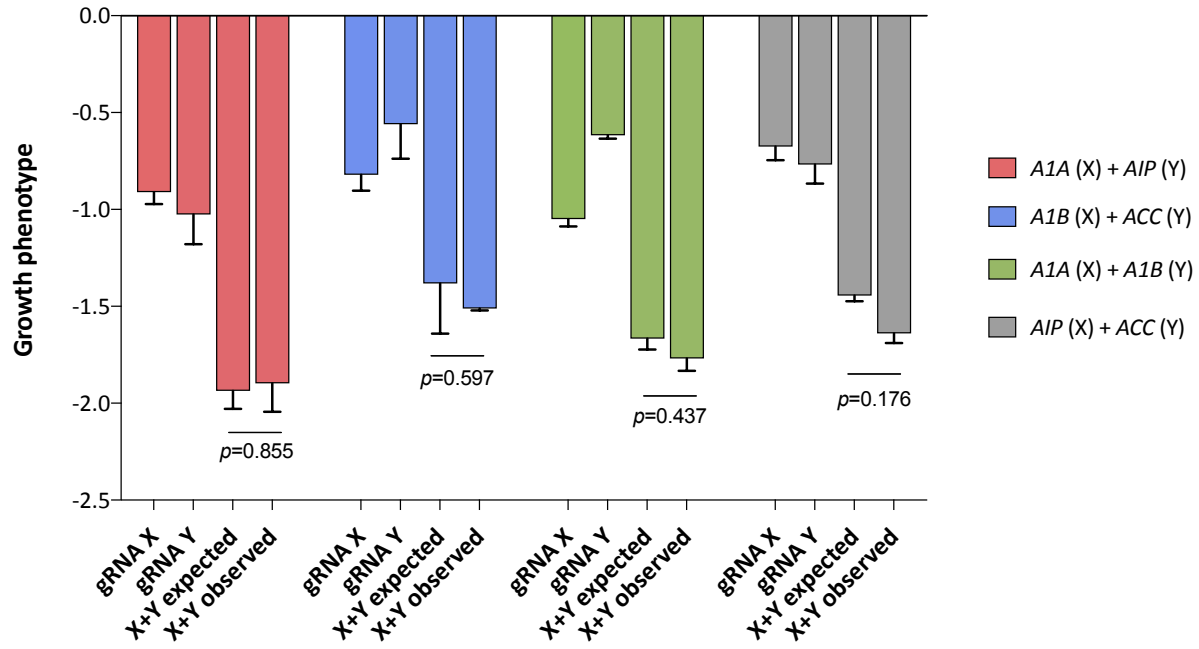



**Figure 4.2. Validation of *ARID1A/ARID1B* SLI using a double gRNA strategy.** BOB-Cas9 cells were transduced with two gRNA lentiviruses simultaneously (one expressing BFP, one expressing mCherry) to give four populations. Cells were analysed by flow cytometry on day 2 and day 14. To calculate the growth phenotype, the % of the single and double positive populations were normalised against the untransduced population (relative abundance), and the $\log_2$(fold-change) in relative abundance was calculated between day 14 and day 2. The growth phenotypes of gRNA X, gRNA Y and gRNA X+Y (expected (sum of X + Y phenotypes) and observed) are shown. This assay was performed in technical duplicate. Error bars show standard deviation. P-values were calculated using a two-tailed paired t-test. *ARID1A = A1A, ARID1B = A1B, AIPL1 = AIP, ACCSL = ACC.*

We considered the possibility that this SLI may depend on the cells adapting to either *ARID1A* or *ARID1B* depletion over a longer period, rather than loss of both genes being induced simultaneously. Thus, we performed a similar assay using the parental BOB-Cas9 line (WT), one of the *ARID1A* KO iPSC lines (ARID1A_C09-Cas9) and one of the *ARID1B* KO iPSC lines (ARID1B_C03-Cas9) (as described in Section 7.16.3.2). The BFP- and mCherry-expressing gRNA lentiviruses prepared in the previous assay were used. Rather than transducing two gRNAs simultaneously, the WT and KO lines were transduced in parallel, with a single control or targeting gRNA (Fig. 4.3).

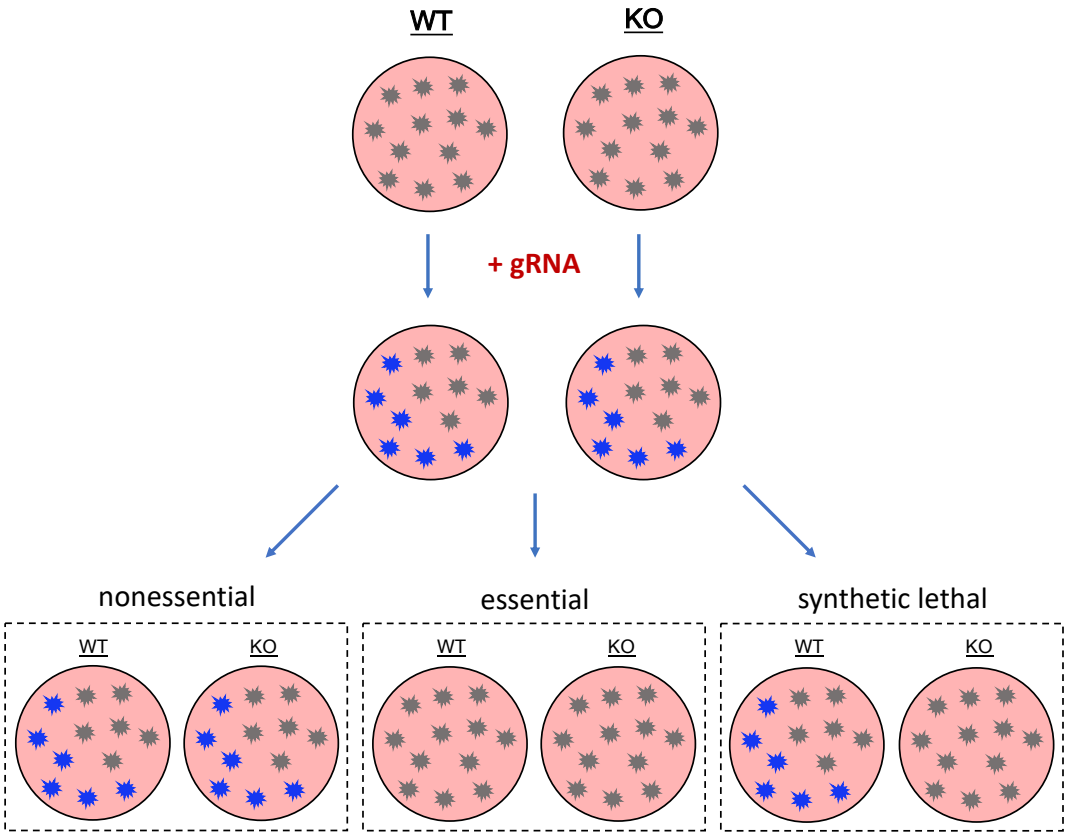



**Figure 4.3. Strategy for single gRNA SLI validation.** WT and KO stable Cas9 cells were transduced with a lentivirus expressing a gRNA and a BFP or mCherry marker. Expression was measured by flow cytometry on day 2 and day 14. Three outcomes were possible: if the gRNA targeted a nonessential gene, fluorescence would remain stable in the WT and KO; if the gRNA targeted an essential gene, cells would die and fluorescence would drop in the WT and KO; if the gRNA targeted a synthetic lethal partner of the KO gene, fluorescence would drop in the KO line but remain stable in the WT.

BOB-Cas9 cells were transduced with the *ARID1A, ARID1B* and *AIPL1* gRNAs. ARID1A_C09-Cas9 cells were transduced with the *ARID1B* and *AIPL1* gRNAs, and ARID1B_C03-Cas9 cells with the *ARID1A* and *AIPL1* gRNAs. BFP/mCherry expression was measured on day 2 and day 14. The $\log_2$(fold-change) in expression was calculated between the two timepoints. Similar to the previous assay, a decrease in expression was observed with each gRNA in all lines (Fig. 4.4). The KO lines had a larger decrease than WT but the differences were not statistically significant, and a difference was also observed with the control *AIPL1* gRNA. We did not test the expression of *ARID1A* or *ARID1B* in these validation assays to confirm that the knockout was functional. The data obtained here suggests that loss of *ARID1A* and *ARID1B* is not synthetic lethal in BOB iPSCs, but confirmation of protein loss would be required to confirm this. Further experiments could also be carried out using different gRNAs, alternative assays or technologies such as si/shRNA.

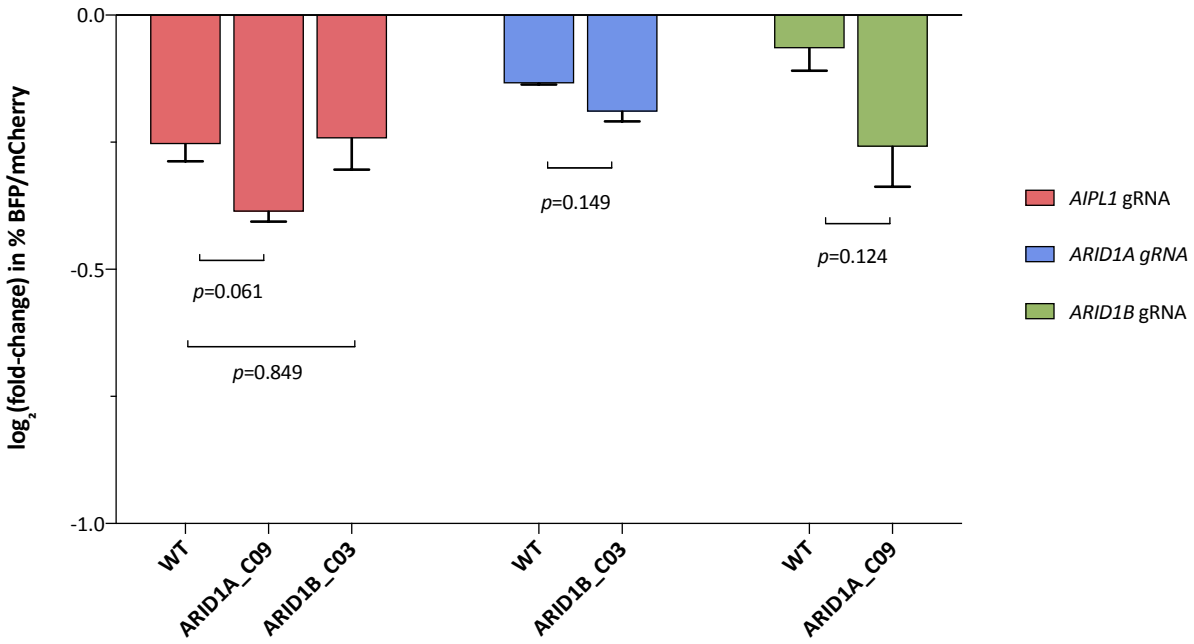



**Figure 4.4. Validation of *ARID1A/ARID1B* SLI using a single gRNA strategy.** WT BOB-Cas9, ARID1A_C09-Cas9 and ARID1B_C03-Cas9 were transduced with lentiviral gRNAs targeting *AIPL1* (BFP-tagged), *ARID1A* (BFP-tagged) and *ARID1B* (mCherry-tagged). All gRNAs were transduced in separate wells. BFP/mCherry expression was measured on day 2 and day 14. The $\log_2$(fold-change) in expression between the two timepoints was calculated. Error bars show standard deviation. P-values were calculated using a two-tailed unpaired t-test using Welch's correction.

### 4.2.2 *ARID1A/ARID1B* interaction in cancer cell lines

The *ARID1A/ARID1B* interaction was originally identified through analysis of shRNA screen data.[172] With a vast amount of CRISPR/Cas9 genome-wide screen data now available, we performed similar analyses to determine whether the interaction could be identified in cancer cell line datasets. Colleagues at the Sanger Institute screened 324 cancer cell lines using the Yusa v1.1 gRNA library.[109] The same pipeline (CRISPRcleanR and BAGELR) was used to process raw data from 342 cancer cell lines screened at the Broad Institute using the Avana gRNA library.[261] This processing was performed by Clare Pacini, a postdoctoral fellow in Francesco Iorio's group at WSI. These datasets included cell lines carrying LOF mutations (frameshift indel or nonsense) in *ARID1A*, *ARID1B, ARID2* and *PBRM1* (Table 4.1). The scaled BFs for *ARID1B* across all screens in each dataset were separated into two groups: *ARID1A* WT and *ARID1A* mutant cell lines (Fig. 4.5a-b). If the scaled BF was > 0, *ARID1B* was considered to be essential; the outcome for each cell line was categorised as 'essential' or 'nonessential'.

**Table 4.1. PBAF/BAF mutant cell lines screened by the Sanger/Broad Institute.** The number of screened cancer cell lines containing frameshift indels or nonsense mutations in *ARID1A, ARID1B, PBRM1* and *ARID2* is indicated. The number of cell lines that were screened by both institutes is indicated. The names of all lines are detailed in Appendix A.11.

| | ***ARID1A* mut** | ***ARID1B* mut** | ***PBRM1* mut** | ***ARID2* mut** |
|---|---|---|---|---|
| **Sanger Institute** | 33 | 11 | 8 | 14 |
| **Broad Institute** | 44 | 14 | 14 | 8 |
| **Overlapping lines** | 17 | 7 | 3 | 2 |

A Fisher's exact test was then applied to determine whether there was an enrichment for *ARID1B* essentiality in the *ARID1A*-mutant cell lines (Fig. 4.5c-d). Benjamini-Hochberg correction was applied to correct for multiple testing. The same process was repeated for *ARID1A* essentiality in *ARID1B*-mutant lines (Fig. 4.6). *ARID1B* was essential in a statistically significant number of *ARID1A*-mutant lines in both datasets (Sanger adjusted p-value = 0.0486, Broad adjusted p-value 0.000029). Conversely, *ARID1B*-mutant lines were not significantly enriched for *ARID1A* essentiality in either dataset (adjusted p-values = 1).

This inconsistency could be due to both groups screening approximately 3-fold more *ARID1A*-mutant cell lines. However, even without statistical significance, there appears to be no trend towards enrichment in the *ARID1B* mutants. Another possibility is that the *ARID1A*-targeting gRNAs in both libraries had low efficacy and the gene was not depleted, but further investigation would be needed to confirm this. LOF mutations in both *ARID1A* and *ARID1B* were present in 6 of the cell lines screened by Sanger and 9 lines screened by Broad. We considered the possibility that these double mutants may have adapted to loss of both proteins and so targeting of either gene would have no effect i.e. the genes would not be synthetic lethal. If true, this would leave only 5 *ARID1B* lines in each dataset that could be reliably analysed for *ARID1A* dependency. However, some double mutants (2/6 in Sanger, 3/9 in Broad) were sensitive to *ARID1B* depletion, suggesting that synthetic lethality can still occur in these lines. In support of this, it is interesting to note that no other studies have demonstrated a dependency on *ARID1A* in *ARID1B*-mutant cells.

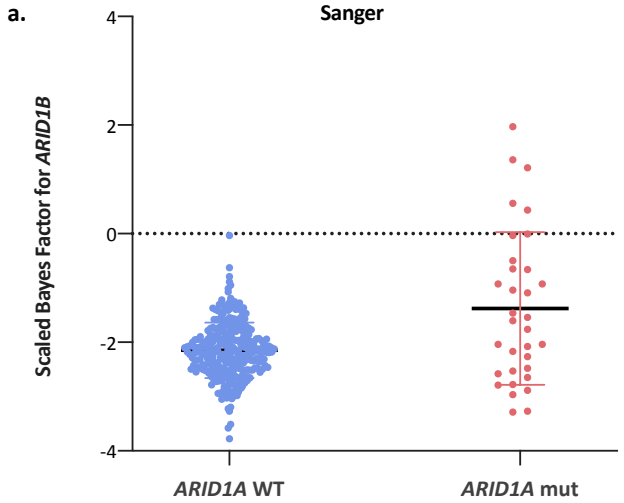
To understand this potential inconsistency in synthetic lethality between *ARID1A* and *ARID1B*, it is important to consider the similarities and differences between the subunits. The exact roles of ARID1A and ARID1B in tumourigenesis are still unclear, as are the reasons for the difference in mutation rate of these genes and the cancer types that they are associated with. ARID1A and ARID1B have been shown to have similar DNA binding affinities[262] and both bind in a non-sequence-specific manner.[162] ARID1A and ARID1B expression vary during cell

cycle progression, with accumulation of ARID1A during the G0/G1 phase but constant expression of ARID1B throughout.[263] Various studies have investigated the effects of these subunits on transcription, although more focus has generally been placed on ARID1A.

A recent study found that knockout of *ARID1A* in HCT116 colorectal cancer cells had a large impact on chromatin state across the genome, with increased or decreased accessibility at thousands of sites.[264] Interestingly, knockdown of *ARID1B* had no effect in wildtype cells but caused changes at hundreds of sites in cells that had also lost *ARID1A*. *ARID1A* was more abundant than *ARID1B* in HCT116 cells which might explain the difference, as wildtype cells may have more BAF complexes containing ARID1A. No change in *ARID1B* expression levels was observed in the *ARID1A* knockout cells, suggesting that the effect was not due to compensatory upregulation of *ARID1B*. Decreased accessibility after ARID1A and ARID1B loss was more common, implying that these proteins predominantly function to maintain open chromatin. Accessible sites that appeared to be ARID1A/1B-dependent were primarily located in enhancers rather than promoters. A study in OCCC cells found that loss of ARID1A causes repression of RNA polymerase II transcription as a result of impaired polymerase pausing.[265] This effect appeared to be greater than the impact on chromatin accessibility in these cells. Upregulation of *ARID1B* occurred to compensate for this, but transcription of some genes was specifically dependent on ARID1A and could not be rescued. Many of the genes that were dependent on ARID1A were also targets of p53. Raab *et al.* (2015) mapped the localisation of complexes containing ARID1A, ARID1B and ARID2 in HepG2 cells.[266] There was a high level of overlap between the regulatory sites bound by each of these subunits. This study also investigated the interactions between the subunits. Hundreds of genes were found to be cooperatively repressed by both ARID1B and ARID2. Competitive interactions were also identified, with ARID1A activating genes that are repressed by ARID2/ARID1B.

It is evident that these subunits have both overlapping and independent roles in regulating transcription. It is likely that the effects of loss of either subunit will be largely dependent on the predominant subunit composition in a given cell type. The functional studies have largely been performed in a single cell type; repetition across multiple lines would be valuable to determine whether the function of these subunits varies with cell type. Gaining more insight into the functional relationships between different PBAF/BAF complexes is vital to understanding the synthetic lethal interactions between subunits. It is difficult to speculate why *ARID1A* mutants may be more dependent on *ARID1B* than *ARID1B* mutants are on *ARID1A*. It may be logical to assume that if a cell has BAF complexes predominantly composed of ARID1A, a mutation in *ARID1B* would not be tumourigenic, and vice versa.

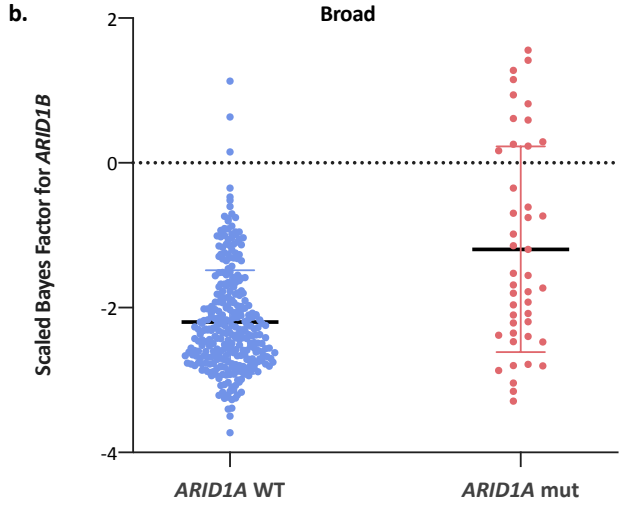
Therefore, all *ARID1A*-mutant cell types may have originally been composed of ARID1A-BAF complexes and *ARID1B*-mutants composed of ARID1B-BAF complexes. If true, these different compositions may cause variation in the functional dynamics between complexes and could explain the potential unidirectional synthetic lethality. Given the cooperation identified between ARID1B and ARID2[266], it is possible that cells originally driven by ARID1B-BAF complexes compensate for ARID1B loss by upregulating PBAF complexes containing ARID2. Conversely, cell types that predominantly carried ARID1A-BAF complexes may primarily default to using ARID1B for compensation. Investigating the subunit composition in lines that have existing mutations in each gene would be the first step in understanding these differences. Analysis of additional data such as RNAseq and proteomics could also be used to elucidate any common alterations in *ARID1B*-mutant lines that differ from *ARID1A* mutants.




c.

| | *ARID1B* essential | *ARID1B* non-essential |
|---|---|---|
| *ARID1A* WT | 0 | 292 |
| *ARID1A* mut | 5 | 28 |

Fisher's test adjusted **p-value = 0.04859411**

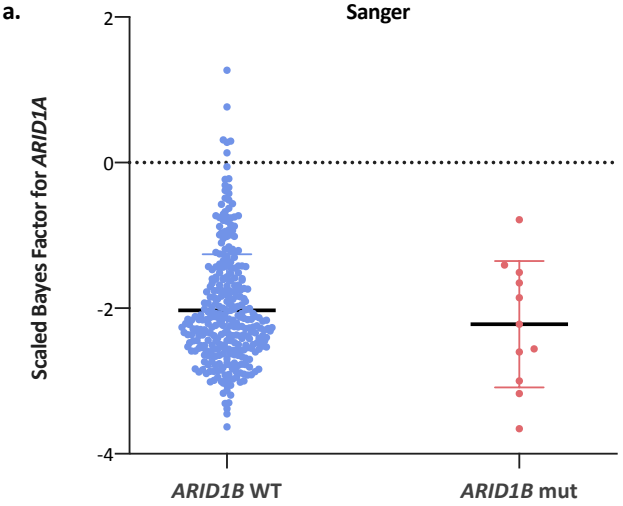



d.

| | *ARID1B* essential | *ARID1B* non-essential |
|---|---|---|
| *ARID1A* WT | 3 | 295 |
| *ARID1A* mut | 12 | 32 |

Fisher's test adjusted **p-value = 0.000029**

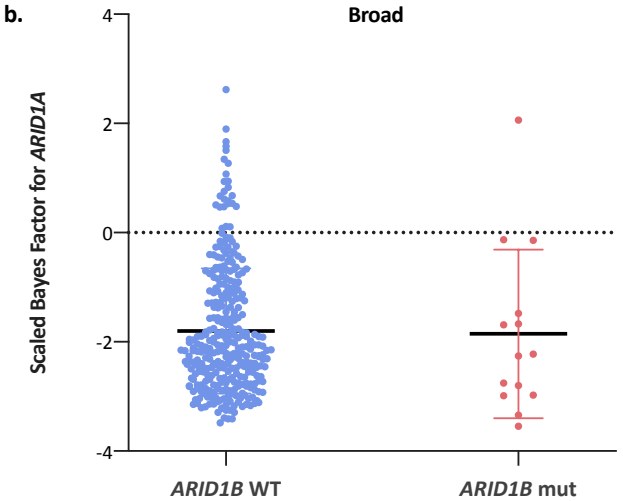
**Figure 4.5. *ARID1B* essentiality in *ARID1A*-mutant cancer cell lines**. BAGELR was used to process data from 324 cancer cell lines screened by the Sanger Institute and 342 lines screened by the Broad Institute. Cell lines were grouped into *ARID1A* mutant (lines containing a frameshift indel or nonsense mutation in *ARID1A*) or *ARID1A* WT (all other lines). Scaled BFs for *ARID1B* were calculated in both the Sanger (a) and Broad (b) datasets. A Fisher's test was applied to compare the number of WT and mutant lines where *ARID1B* was essential or nonessential in the Sanger (c) and Broad (d) datasets. Adjusted p-values were calculated using Benjamini-Hochberg correction for multiple testing. Scaled BFs for the Sanger screens are detailed in Appendix A.12, and for the Broad screens in Appendix A.13.




c.

| | *ARID1A* essential | *ARID1A* non-essential |
|---|---|---|
| *ARID1B* WT | 6 | 308 |
| *ARID1B* mut | 0 | 11 |

Fisher's test adjusted **p-value = 1**




d.

| | *ARID1A* essential | *ARID1A* non-essential |
|---|---|---|
| *ARID1B* WT | 25 | 303 |
| *ARID1B* mut | 1 | 13 |

Fisher's test adjusted **p-value = 1**

**Figure 4.6. *ARID1A* essentiality in *ARID1B*-mutant cancer cell lines**. BAGELR was used to process data from 324 cancer cell lines screened by the Sanger Institute and 342 lines screened by the Broad Institute. Cell lines were grouped into *ARID1B* mutant (lines containing a frameshift indel or nonsense mutation in *ARID1B*) or *ARID1B* WT (all other lines). Scaled BFs for *ARID1A* were calculated in both the Sanger (a) and Broad (b) datasets. A Fisher's test was applied to compare the number of WT and mutant lines where *ARID1A* was essential or nonessential in the Sanger (c) and Broad (d) datasets. Adjusted p-values were calculated using Benjamini-Hochberg correction for multiple testing. Scaled BFs for the Sanger screens are detailed in Appendix A.12, and for the Broad screens in Appendix A.13.

## 4.3 Selection of candidate SLIs for validation

Although the interaction between *ARID1A* and *ARID1B* was not observed in our screens, there were other genes that appeared to be specifically depleted in the PBAF/BAF gene KO lines. For each subunit, we selected several candidate synthetic lethal partners to validate. At the point of validation, we had prioritised analysis using BAGELR and candidate genes were chosen using this dataset. Four criteria were set for filtering hits:

1. The gene must not be essential in the parental BOB screen.
2. The gene must be essential in at least two of the KO screens.
3. The gene must not be significantly essential in more than two of the six unrelated KO screens (*APC, ATM, TP53, RB1, FAT1, FBXW7* knockout BOB lines).
4. The gene must not be a core fitness gene, as annotated by Behan *et al.* (2019) using cancer cell line screen data.

At this stage we did not have replicate screen data for the parental line. Therefore, criteria (3) was used to filter out genes that may be essential in this cell line background but were missed in the first parental BOB screen. This was based on the assumption that it was unlikely for a gene to also have a synthetic lethal interaction with 3+ other genes that are not involved in the PBAF/BAF complexes. These KO lines were chosen because, at the time of validation, these were the only screens that we had data for. This filtering strategy produced a candidate list of 66 genes (Fig. 4.7, Appendix A.14).

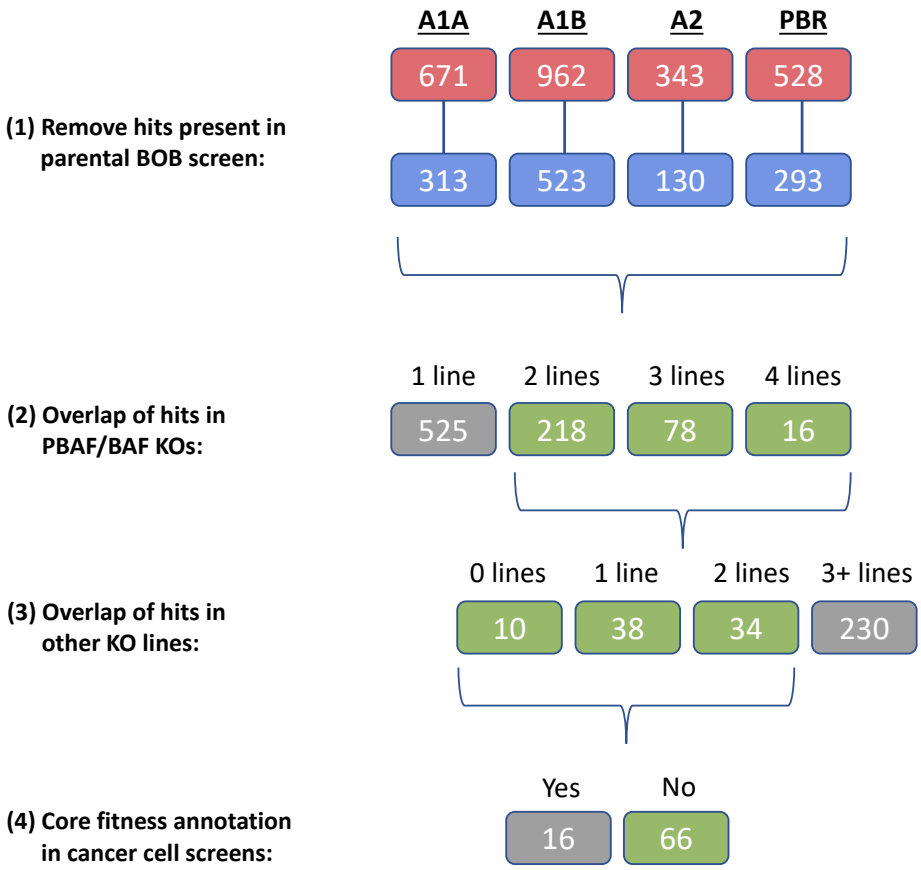



**Figure 4.7. Filtering of PBAF/BAF gene SLIs for validation.** All genes with a scaled BF > 0 were noted for ARID1A_C09 (A1A), ARID1B_C03 (A1B), ARID2_C11 (A2), and PBRM1_F09 (PBR). These went through four filtering steps. **(1)** Any genes that also had a scaled BF > 0 in the parental BOB screen were removed. **(2)** The lists were overlapped and any genes that were present in only one of the four lines were removed. **(3)** Any genes with a scaled BF > 0 in more than 2 of the *APC, ATM, TP53, RB1, FAT1, FBXW7* KO line screens were removed. **(4)** Any gene that was called as a core fitness gene in the Behan *et al.* study[109] was removed. The final gene list is provided in Appendix A.14.

### 4.3.1 Gene enrichment analysis of candidate genes

To explore whether the list of 66 candidate genes shared common features, a gene enrichment analysis was performed based on various parameters using Enrichr.[267,268] We tested for enrichment of ontologies (GO Biological Process and GO Molecular Function, Table 4.2) and pathways (WikiPathways, Reactome, Table 4.3). For ontologies, there was one significant enrichment observed and this was for genes that were annotated to have a role in RNA binding (Table 4.2). Several pathway annotations were significantly enriched in the gene set, mainly related to DNA damage response and roles in the cell cycle (Table 4.3), but few genes contributed to these enrichments.

**Table 4.2. Gene set enrichment analysis: ontologies.** Top 15 results, ranked by p-value, based on 'GO Biological Process 2018' and 'GO Molecular Function 2018' annotations. Analysis performed using Enrichr.[267,268]

| Term | P-value | Adjusted P-value | Genes |
|---|---|---|---|
| **GO Biologcal Process 2018** | | | |
| mitochondrial translational termination (GO:0070126) | 2.23E-04 | 0.284874957 | *MRPS24;MRPL17;MRPL28;MRPL44* |
| primary miRNA processing (GO:0031053) | 4.74E-04 | 0.30264963 | *DGCR8;MRPL44* |
| translational termination (GO:0006415) | 2.98E-04 | 0.303737829 | *MRPS24;MRPL17;MRPL28;MRPL44* |
| mitochondrial translation (GO:0032543) | 4.48E-04 | 0.326581939 | *MRPS24;MRPL17;MRPL28;MRPL44* |
| protein deubiquitination (GO:0016579) | 2.07E-04 | 0.351622993 | *INO80B;ENY2;PSMD4;MDM2;RHOA;OTUB1* |
| translational elongation (GO:0006414) | 4.17E-04 | 0.354931388 | *MRPS24;MRPL17;MRPL28;MRPL44* |
| mitochondrial translational elongation (GO:0070125) | 2.05E-04 | 0.522533528 | *MRPS24;MRPL17;MRPL28;MRPL44* |
| negative regulation of DNA replication (GO:0008156) | 0.001249177 | 0.579504755 | *RAD17;ATR* |
| regulation of signal transduction by p53 class mediator (GO:1901796) | 0.00115331 | 0.588534038 | *MDM2;RAD17;BRD7;ATR* |
| regulation of DNA replication (GO:0006275) | 0.001129585 | 0.640474947 | *DBF4;RAD17;ATR* |
| cellular response to gamma radiation (GO:0071480) | 0.001585937 | 0.674419809 | *MDM2;ATR* |
| protein modification by small protein removal (GO:0070646) | 0.001730675 | 0.679356514 | *INO80B;PSMD4;MDM2;RHOA;OTUB1* |
| mitochondrial ribosome assembly (GO:0061668) | 1.60E-04 | 0.813974607 | *DDX28;DHX30* |
| gene expression (GO:0010467) | 0.002361877 | 0.860904195 | *DHX30;MRPS24;FIP1L1;NUP50;MRPL28;MRPL44* |
| apoptotic nuclear changes (GO:0030262) | 0.002594763 | 0.882738454 | *ACIN1;HMGB1* |
| **GO Molecular Function 2018** | | | |
| RNA binding (GO:0003723) | 1.76E-06 | 0.002024337 | *DDX28;MRPS24;CCDC137;PARN;MRPL28;HMGB1;AATF;SAFB;CCAR1;MRPL44;DHX30;PSMD4;FIP1L1;DGCR8;ACIN1;TARDBP;EIF3C* |
| double-stranded RNA binding (GO:0003725) | 0.001026875 | 0.590966809 | *DHX30;DGCR8;MRPL44* |
| endoribonuclease activity, producing 5'-phosphomonoesters (GO:001689 | 0.003063747 | 1 | *DGCR8;MRPL44* |
| double-stranded DNA binding (GO:0003690) | 0.004418151 | 1 | *HMGB1;SAFB;TARDBP* |
| NADH dehydrogenase (quinone) activity (GO:0050136) | 0.009279665 | 1 | *NDUFB1;NDUFC2* |
| NADH dehydrogenase (ubiquinone) activity (GO:0008137) | 0.009279665 | 1 | *NDUFB1;NDUFC2* |
| rRNA binding (GO:0019843) | 0.009279665 | 1 | *DDX28;MDM2* |
| mRNA 3'-UTR binding (GO:0003730) | 0.018958545 | 1 | *PARN;TARDBP* |
| histone methyltransferase binding (GO:1990226) | 0.019639652 | 1 | *CBX1* |
| NEDD8 transferase activity (GO:0019788) | 0.019639652 | 1 | *MDM2* |
| ligand-dependent nuclear receptor transcription coactivator activity (GO:0 | 0.020086948 | 1 | *ENY2;CCAR1* |
| RNA helicase activity (GO:0003724) | 0.021242881 | 1 | *DDX28;DHX30* |
| ATP-dependent RNA helicase activity (GO:0004004) | 0.021242881 | 1 | *DDX28;DHX30* |
| RNA-dependent ATPase activity (GO:0008186) | 0.021831047 | 1 | *DDX28;DHX30* |
| protein-DNA loading ATPase activity (GO:0033170) | 0.022875799 | 1 | *RAD17* |

**Table 4.3. Gene set enrichment analysis: pathways.** Top 15 results, ranked by p-value, based on 'WikiPathways 2019' and 'Reactome 2016' annotations. Analysis performed using Enrichr.[267,268]

| Term | P-value | Adjusted P-value | Genes |
|---|---|---|---|
| **WikiPathways 2019** | | | |
| DNA Damage Response WP707 | 7.51E-05 | 0.035468925 | *MDM2;RAD17;RAD1;ATR* |
| miRNA Regulation of DNA Damage Response WP1530 | 8.90E-05 | 0.021002442 | *MDM2;RAD17;RAD1;ATR* |
| DNA IR-damage and cellular response via ATR WP4016 | 1.42E-04 | 0.02228976 | *MDM2;RAD17;RAD1;ATR* |
| ATR Signaling WP3875 | 3.80E-04 | 0.044885074 | *RAD1;ATR* |
| DNA IR-Double Strand Breaks (DSBs) and cellular response via ATM WP3959 | 7.97E-04 | 0.075195782 | *MDM2;RAD17;ATR* |
| DNA Damage Response (only ATM dependent) WP710 | 0.005760736 | 0.453177875 | *MDM2;HMGB1;RHOA* |
| Cell Cycle WP179 | 0.007323424 | 0.493808035 | *DBF4;MDM2;ATR* |
| Integrated Cancer Pathway WP1971 | 0.009279665 | 0.54750025 | *MDM2;ATR* |
| Mitochondrial complex I assembly model OXPHOS system WP4324 | 0.014728393 | 0.772422369 | *NDUFC2;NDUFB1* |
| Oxidative phosphorylation WP623 | 0.016785955 | 0.792297083 | *NDUFC2;NDUFB1* |
| Genotoxicity pathway WP4286 | 0.018404794 | 0.789732994 | *COIL;MDM2* |
| exRNA mechanism of action and biogenesis WP2805 | 0.019639652 | 0.772492975 | *DGCR8* |
| miRNA Biogenesis WP2338 | 0.019639652 | 0.713070438 | *DGCR8* |
| Hypothetical Craniofacial Development Pathway WP3655 | 0.022875799 | 0.771241217 | *RHOA* |
| DDX1 as a regulatory component of the Drosha microprocessor WP2942 | 0.022875799 | 0.719825136 | *DGCR8* |
| **Reactome 2016** | | | |
| Activation of ATR in response to replication stress_Homo sapiens_R-HSA-176187 | 6.58E-06 | 0.010067724 | *DBF4;RAD17;RAD1;ATR* |
| Cell Cycle Checkpoints_Homo sapiens_R-HSA-69620 | 3.02E-05 | 0.023108141 | *DBF4;PSMD4;MDM2;RAD17;RAD1;ATR* |
| G2/M Checkpoints_Homo sapiens_R-HSA-69481 | 1.37E-04 | 0.069999104 | *DBF4;PSMD4;RAD17;RAD1;ATR* |
| Regulation of TP53 Activity_Homo sapiens_R-HSA-5633007 | 1.42E-04 | 0.054159898 | *MDM2;RAD17;RAD1;BRD7;ATR* |
| Mitochondrial translation initiation_Homo sapiens_R-HSA-5368286 | 1.71E-04 | 0.052368228 | *MRPS24;MRPL17;MRPL28;MRPL44* |
| Mitochondrial translation termination_Homo sapiens_R-HSA-5419276 | 1.71E-04 | 0.04364019 | *MRPS24;MRPL17;MRPL28;MRPL44* |
| Mitochondrial translation elongation_Homo sapiens_R-HSA-5389840 | 1.71E-04 | 0.037405877 | *MRPS24;MRPL17;MRPL28;MRPL44* |
| Regulation of TP53 Activity through Phosphorylation_Homo sapiens_R-HSA-6804756 | 2.14E-04 | 0.040908972 | *MDM2;RAD17;RAD1;ATR* |
| Mitochondrial translation_Homo sapiens_R-HSA-5368287 | 2.23E-04 | 0.037960998 | *MRPS24;MRPL17;MRPL28;MRPL44* |
| HDR through Single Strand Annealing (SSA)_Homo sapiens_R-HSA-5685938 | 2.46E-04 | 0.037657161 | *RAD17;RAD1;ATR* |
| Presynaptic phase of homologous DNA pairing and strand exchange_Homo sapiens_R-HSA-5693616 | 2.88E-04 | 0.040076224 | *RAD17;RAD1;ATR* |
| Homologous DNA Pairing and Strand Exchange_Homo sapiens_R-HSA-5693579 | 3.59E-04 | 0.045821963 | *RAD17;RAD1;ATR* |
| Cell Cycle_Homo sapiens_R-HSA-1640170 | 5.31E-04 | 0.062502192 | *NUDC;DBF4;PSMD4;NUP50;MDM2;RAD17;RAD1;ATR* |
| HDR through Homologous Recombination (HRR)_Homo sapiens_R-HSA-5685942 | 0.001238565 | 0.135357469 | *RAD17;RAD1;ATR* |
| The citric acid (TCA) cycle and respiratory electron transport_Homo sapiens_R-HSA-1428517 | 0.001642254 | 0.167509902 | *PDPR;NDUFC2;NDUFB1;ATP5G2* |

## 4.4 Validation of screen hits in iPS cells

From this filtered list of 66 genes, we selected 20 genes to experimentally validate in the screened iPSC lines. The previously described single gRNA competitive growth assay was used for validation (Fig. 4.3). For each candidate gene, a gRNA from an independent library (i.e. different to those used in the screen) was selected (Appendix A.1, Section 7.16.4). All gRNAs were obtained from the Sanger Human Whole Genome CRISPR arrayed library, which uses a backbone that expresses a BFP marker.[103] A gRNA targeting *THAP3* was chosen as a negative control as this gene had a low scaled BF across all screens. A gRNA targeting *TWISTNB* was chosen as a positive control as this gene was consistently significant across the screens. gRNAs were packaged into lentiviruses and cells were transduced with each gRNA individually in 6-well plates (as described in Section 7.16.5). Parental BOB (WT) cells were transduced with all gRNAs; KO lines were transduced only with gRNAs targeting genes that were hits in the respective screen (Table 4.4).

Cells were passaged on day 2 and some cells were fixed for analysis of BFP expression by flow cytometry. Transduction was successful, with a high % of BFP (average ~85%) in all conditions. Some cell death was observed which was likely due to initial toxicity induced by the Cas9 cutting. After 5 days, cells were passaged again. Further cell death had occurred in some conditions, suggesting that gRNAs were having an effect, so the split ratio was altered for each well accordingly. However, very few cells survived this passage across all conditions. This assay was repeated several times, varying density and trying to limit the stress to the cells during passaging, but the same issue occurred. The level of cell death appeared to be variable, with no clear trend to suggest that KO lines were more susceptible than the WT. Cells carrying the negative control gRNA were also affected. Due to unforeseen issues with the iPSC medium and time restrictions, we were unable to progress further with this validation.

The transduction levels were high so the first step would be to repeat this assay with less lentivirus in an attempt to improve cell survival. The backbone used in this assay contained a piggyBac transposon, so transfection of the plasmids with a transposase could be trialled as an alternative to lentiviral transduction. The toxicity may have been caused by Cas9-induced DSBs; this effect has been observed previously but not to the extent seen here. It was surprising that toxicity related to the lentivirus or Cas9-cutting occurred at this later stage. It would be expected that these effects would occur in the days immediately after transduction. However, the cells that survived the first passage may have had impaired fitness but were able to continue

proliferating until the stress of the next passage. An alternative targeting method such as si/shRNA could be used to validate the candidates, removing the complication of DSB toxicity.

We must consider the possibility that all of the genes chosen were in fact essential in the WT and KO lines, hence the universal cell death. In support of this, 37/66 of the filtered genes were significantly depleted in a repeat screen of the parental BOB line (data obtained after validation experiments). These included 12 of the 20 genes chosen for validation (indicated by an * in Table 4.4). Although this does not explain the toxicity observed for the other 8 genes and the negative control, it does indicate that there was a lot of noise in the screen data which may affect validation rates. Repeating the assay in the ways discussed here should elucidate whether the issue was technical or due to false positives.

**Table 4.4. Genes selected for SLI validation.** Each gene selected for validation was assigned an ID number. The KO lines in which they were tested are indicated. * indicates genes that were found to be significantly depleted in the second screen of parental BOB cells.

| Gene | gRNA ID | KO lines |
|---|---|---|
| *THAP3* | 1 | Positive control, all lines |
| *PDPR** | 2 | *ARID1A + ARID2* |
| *TWISTNB* | 3 | Negative control, all lines |
| *SLC9B1* | 4 | *ARID2 + PBRM1* |
| *MRPL17* | 5 | *ARID1A + ARID1B + PBRM1* |
| *RHOA* | 6 | *ARID1A + ARID1B* |
| *AK4** | 7 | *ARID1B + ARID2* |
| *HMGB1** | 8 | *ARID1B + PBRM1* |
| *BTBD7** | 9 | *ARID1A + ARID1B + ARID2* |
| *COIL** | 10 | *ARID1A + ARID1B* |
| *PARN* | 11 | *ARID1B + PBRM1* |
| *DGCR8** | 12 | *ARID1A + ARID1B* |
| *KRT86* | 13 | *ARID1B+ARID2+PBRM1* |
| *CCAR1** | 14 | *ARID2 + PBRM1* |
| *OTUB1** | 15 | *ARID1A + ARID1B* |
| *WDR25** | 16 | *ARID1A + ARID2* |
| *OBP2B** | 17 | *ARID1A + ARID1B* |
| *ATP5G2* | 18 | *ARID1A + ARID1B* |
| *ACIN1** | 19 | *ARID1A + ARID1B* |
| *ADH5** | 20 | *ARID1B + PBRM1* |
| *ZRSR2* | 21 | *ARID1A + ARID1B* |
| *KRTAP4-8* | 22 | *ARID1B + PBRM1* |

## 4.5 Validation of screen hits in HAP1 cells

The aim of this project was to find SLIs that could be exploited to treat cancer and so it was important to confirm that any interactions we identified were not specific to iPSCs. Therefore, in parallel to validation in BOB iPSCs, we also performed validation in a cancer cell line. HAP1 is a near-haploid adherent human cell line derived from KBM7, a male chronic myelogenous leukaemia line.[269] The advantage of using haploid cells for gene editing experiments is that they only have one copy of each gene, and so LOF mutations can be introduced more efficiently. Also, isogenic KO derivatives of HAP1 are commercially available. Considering these factors, we chose to validate our iPSC screen hits in HAP1 cells, using parental and *ARID1A/ARID1B/ARID2/PBRM1* KO derivatives. HAP1 cells were purchased from Horizon; PCR and Sanger sequencing were performed to confirm that the correct mutations were present in each line (Fig. 4.8) (as described in Section 7.16.1 and 7.4). Stable Cas9 lines were engineered and a Cas9 activity test was performed after blasticidin selection, confirming that all lines had high activity (Fig. 4.9) (as described in Section 7.10).

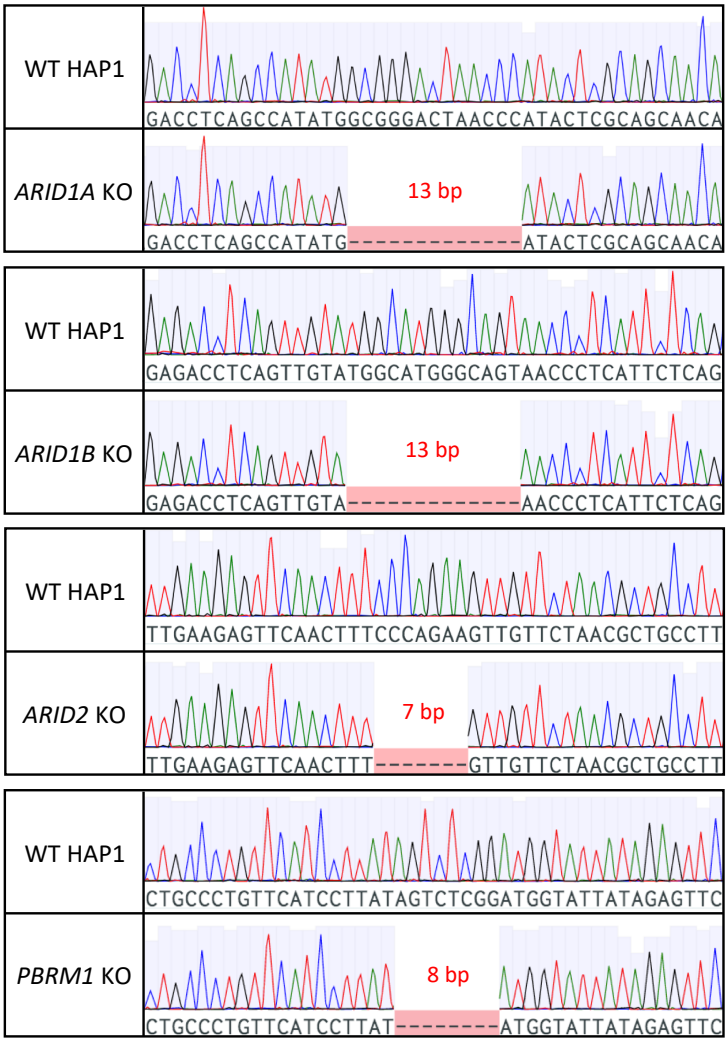
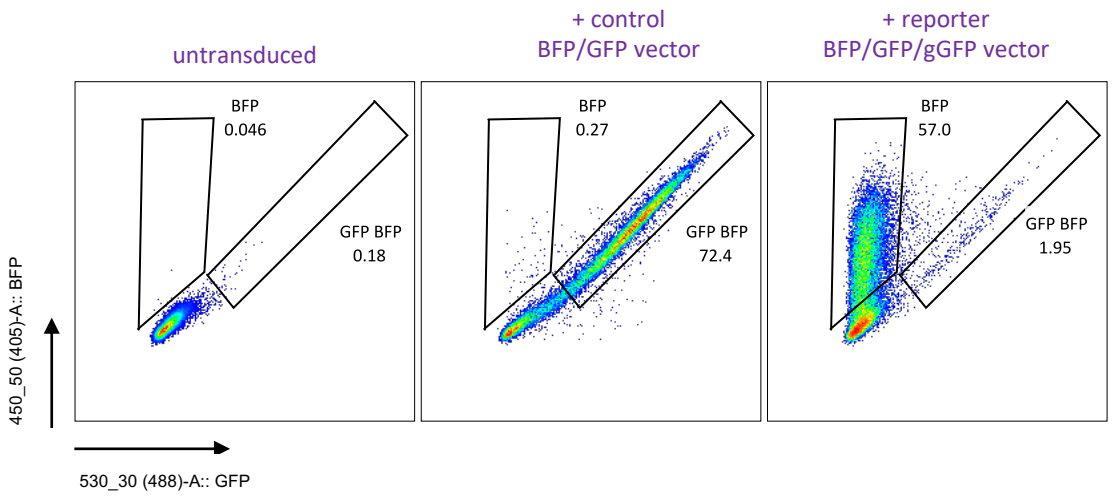



**Figure 4.8. Genotypes of PBAF/BAF knockout HAP1 lines.** PCR and Sanger sequencing were performed on the edited region in *ARID1A, ARID1B, ARID2* and *PBRM1* KO HAP1 cell lines. For each gene, WT HAP1 cells were also sequenced for comparison. The red areas indicate frameshift deletions. Primers used for genotyping are detailed in Appendix A.1.

**a.**




**b.**

| HAP1 cell line | Cas9 activity |
|---|---|
| WT | 97% |
| *ARID1A* KO | 94% |
| *ARID1B* KO | 97% |
| *ARID2* KO | 89% |
| *PBRM1* KO | 94% |

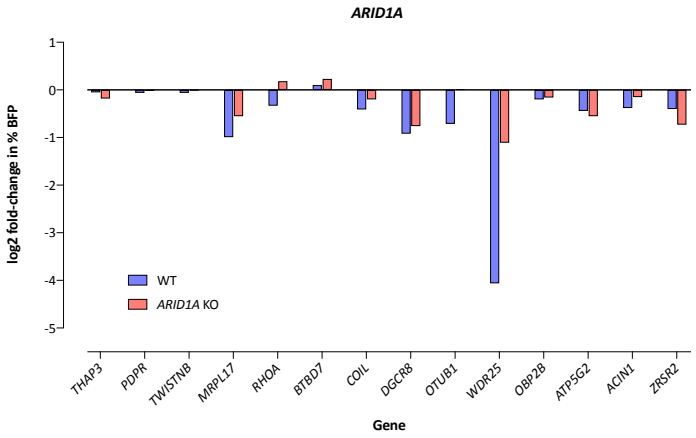
**Figure 4.9. Cas9 activity in HAP1 cell lines.** After blasticidin selection, Cas9-expressing WT and KO HAP1 cells were transduced with a control vector (BFP/GFP) or a reporter vector (BFP/GFP/gGFP). Fluorescence was measured by flow cytometry at 3 days post-transduction. **(a)** Untransduced cells and cells transduced with the control vector were used to gate for expression of BFP and GFP. Plots are shown for the WT HAP1-Cas9 line, **(b)** Cas9 activity was calculated as the percentage of BFP positive cells divided by the percentage of total cells transduced with the reporter vector (i.e. BFP+(BFP/GFP))**.**

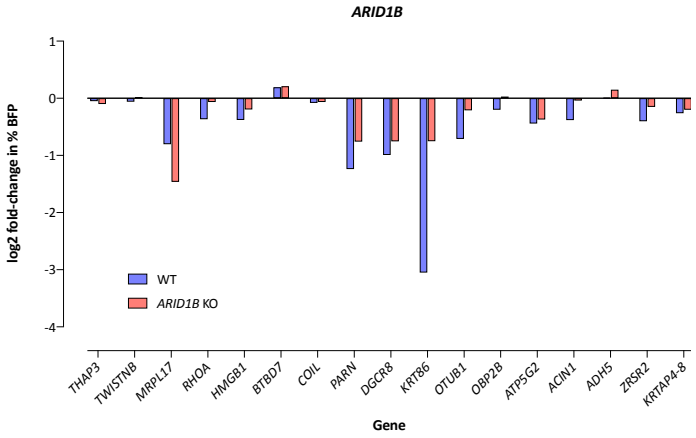
### 4.5.1 Competition assay in HAP1 cells

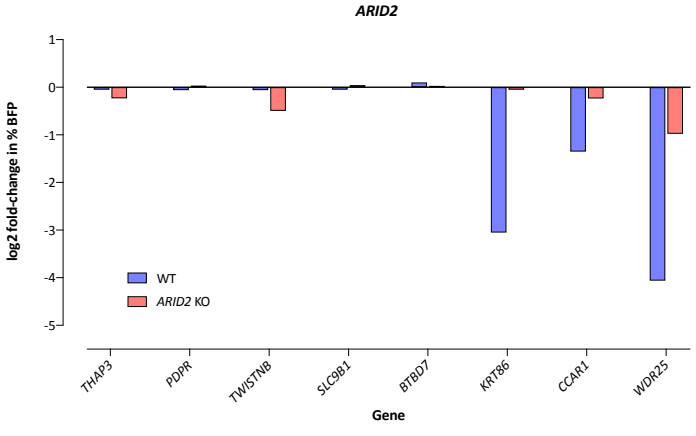
In parallel with validation in iPSCs, the fluorescence-based competition assay was also performed in the HAP1 cell lines using the same gRNAs (as described in Section 4.4 and 7.15.4-5). WT and KO HAP1 cell lines were transduced with a single gRNA per well in a 12-well plate. As with the iPSC validation, WT cells were transduced with all gRNAs and each KO line was transduced only with the genes that were hits in the respective KO iPSC screen (Table 4.4). Passaging and analysis of BFP expression were performed on day 2. Unlike the iPSCs, HAP1 cells survived after further passaging and were maintained until day 14 post-transduction. Some cell death was observed but this was minimal in comparison to the iPSCs,

indicating that HAP1 cells were not as sensitive to Cas9-induced DSBs. BFP expression was analysed again on day 14 and the $\log_2$(fold-change) in expression between both timepoints was calculated (Fig. 4.10). Many of the gRNAs had no effect in either the WT or KO lines. Some conditions had a greater loss of BFP expression in the KO compared to WT (*ATP5G2* and *ZRSR2* in *ARID1A* KO; *MRPL17* in *ARID1B* KO; *THAP3* and *TWISTNB* in *ARID2* KO; *THAP3*, *TWISTNB* and *KRTAP4*-8 in *PBRM1* KO). Further replicates are needed to determine any statistical significance for these differences. In many conditions, a larger reduction in expression was observed in the WT compared to the KOs. Some of these results were substantial and were observed in multiple KOs (particularly gRNAs 13 and 16).

This was unexpected and there are a number of ways that this could be interpreted. Loss of PBAF/BAF subunits may make HAP1 cells less dependent on these genes, although in most cases a negative phenotype was also observed in the KO, albeit much less prominent. Genome editing may be less efficient in the KO lines, resulting in fewer cells losing expression of the targeted gene. Further investigation and repetition of the assay would be needed before any conclusions can be made regarding interactions between these genes. If it was confirmed that the WT cells are more susceptible to gRNA activity in general, then the differences between effects in WT and KO may be more substantial than they appeared in this assay. Initial focus for follow-up should be placed on those genes that induced a more negative phenotype in the KO line. It would also be essential to repeat this with different gRNAs, or use an alternative strategy such as shRNA, to confirm that the phenotype was caused by depletion of a given gene and not due to an off-target effect.


ARID1A
log2 fold-change in % BFP
1
0
-1
-2
-3
-4
-5
WT
ARID1A KO
THAP3
PDPR
TWISTNB
MRPL17
RHOA
BTBD7
COIL
DGCR8
OTUB1
WDR25
OBP2B
ATP5G2
ACIN1
ZRSR2
Gene


ARID1B
log2 fold-change in % BFP
1
0
-1
-2
-3
-4
WT
ARID1B KO
THAP3
TWISTNB
MRPL17
RHOA
HMGB1
BTBD7
COIL
PARN
DGCR8
KRT86
OTUB1
OBP2B
ATP5G2
ACIN1
ADH5
ZRSR2
KRTAP4-8
Gene


ARID2
log2 fold-change in % BFP
1
0
-1
-2
-3
-4
-5
WT
ARID2 KO
THAP3
PDPR
TWISTNB
SLC9B1
BTBD7
KRT86
CCAR1
WDR25
Gene

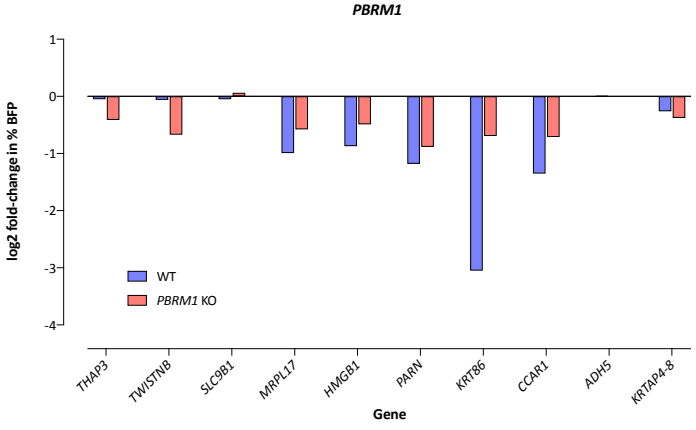

PBRM1
log2 fold-change in % BFP
1
0
-1
-2
-3
-4
WT
PBRM1 KO
THAP3
TWISTNB
SLC9B1
MRPL17
HMGB1
PARN
KRT86
CCAR1
ADH5
KRTAP4-8
Gene

**Figure 4.10. Validation of candidate SLIs in HAP1 cells.** WT and KO HAP1 Cas9 cells were transduced with a single BFP-expressing gRNA lentivirus per well in a 12-well plate. BFP expression was analysed by flow cytometry on day 2 and day 14 post-transduction. The $\log_2$(fold-change) in expression between the two timepoints was calculated. Plots show the $\log_2$(fold-changes) measured in each KO line, with results for each gRNA tested in that line; the corresponding results in the WT line are plotted for comparison. Due to technical issues, no data was obtained for gRNAs 7 or 8 in any lines, or gRNA 17 in the *ARID2* KO.

# 4.6 Further analyses to identify candidate SLIs

## 4.6.1 Re-analysis of PBAF/BAF SLIs using additional iPSC screen data

As stated previously, the genes for validation were chosen before all of the iPSC screens were completed, including two additional replicates of the parental line and an independent KO line for each PBAF/BAF gene. We performed further analysis to determine how the results changed with the inclusion of this new data (Fig. 4.11). We initially removed any gene that had a scaled BF > 0 in at least one of the three BOB screens. One of these screens, BOB_2, had an improved performance compared to the other two and resulted in exclusion of many more genes. As discussed in Section 3.7, there are various other ways that the parental screen data could be used to filter for KO-specific dependencies.

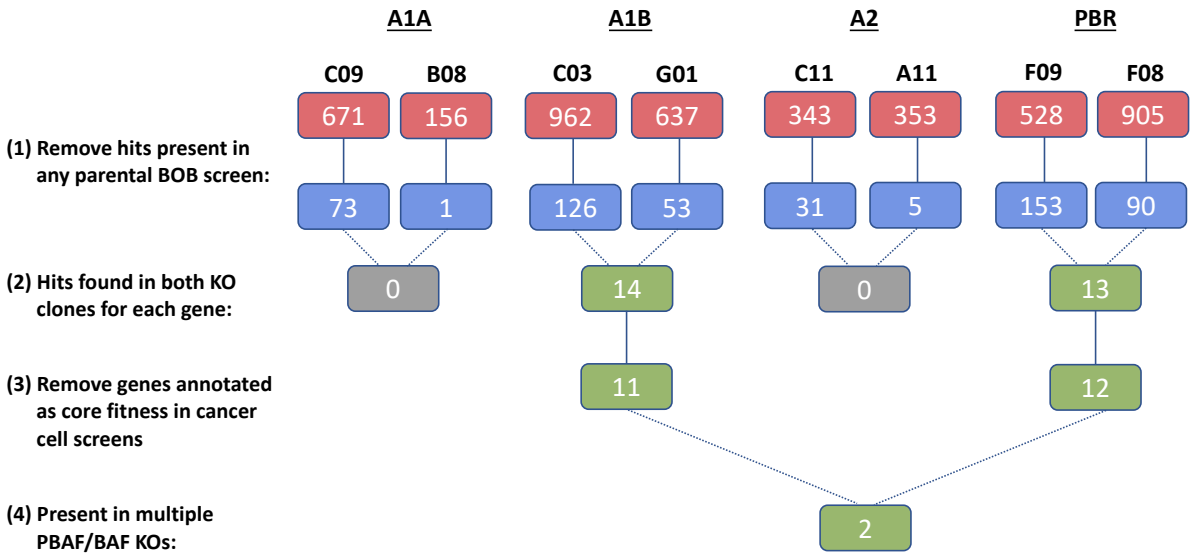



**Figure 4.11. Filtering of PBAF/BAF gene SLIs using additional data.** All genes with a scaled BF >0 were noted for ARID1A_C09/B08, ARID1B_C03/G01, ARID2_C11/A11, and PBRM1_F09/F08. These went through four filtering steps. **(1)** Genes that also had a scaled BF > 0 in any of the three parental BOB screens were removed. **(2)** Genes that were a hit only in one of the KO clones for each subunit were removed. **(3)** Any gene that was considered to be a core fitness gene in the Behan *et al.* study[109] was removed. **(4)** The remaining genes were cross-referenced to identify hits common to more than one subunit.

375 genes remained that were significant in at least one PBAF/BAF line. The majority (297) of these were specific to one subunit. We then filtered for genes that were hits in both knockout clones of a PBAF/BAF subunit, removing genes previously annotated as core fitness: 11 genes were significant in both *ARID1B* clones; 12 were significant in both *PBRM1* clones; none were significant in both *ARID1A* or *ARID2* clones. Only 2 genes (*KRT86* and *KCMF1*) were significant in both knockouts of more than one PBAF/BAF subunit. These data can be filtered in many ways by altering various parameters, but one would assume that the most reliable hits are those that are supported by more than one knockout clone. Experimental validation of candidates from different filtering methods can be used to determine the most robust way of analysing the data. *KRT86* was a hit in both *ARID1B* and *PBRM1* knockout clones, and in one knockout clone of *ARID1A* and *ARID2*. KRT86 encodes a type II keratin protein, involved in the formation of hair and nails. There are no known alternative roles for this protein and so it is unclear why loss of this gene would be synthetic lethal with the PBAF/BAF genes. *KCMF1* was also a hit in both *ARID1B* and *PBRM1* knockout clones, and in one of the *ARID2* knockout clones. Potassium channel modulatory factor 1 (KCMF1) is a poorly characterised E3 ubiquitin ligase. Upregulation of the protein in gastric cancer has been linked to fibroblast growth factor signalling pathways.[270] *KCMF1* has also been reported to play a pro-oncogenic role in pancreatic cancer[271] and knockdown was associated with reduced cell proliferation and colony formation in colon cancer stem cells.[272] The lack of literature regarding its function makes it difficult to speculate on the mechanism of synthetic lethality with the PBAF/BAF genes but given its association with several cancers, *KCMF1* may be a more interesting candidate to investigate further.

### 4.6.2 Dependencies associated with PBAF/BAF mutation in cancer cells

As discussed in Section 1.4.4, few synthetic lethal partners of *ARID1A, ARID1B, ARID2* and *PBRM1* have been robustly established in cancer cells. We performed further analyses of the Sanger and Broad CRISPR/Cas9 screen datasets to determine whether any additional interactions could be identified, and to cross-reference these with the iPSC screen data. As described in Section 4.2.2, lines were separated into WT and mutants, based on LOF mutation status in *ARID1A, ARID1B, ARID2* or *PBRM1*. Genes were categorised as ‘essential’ or ‘nonessential’ based on the scaled BF. For each subunit, a Fisher’s test was applied for all genes to identify any that were essential in a significantly greater number of mutant lines

compared to WT. Benjamini-Hochberg correction was used to correct for multiple testing, and each gene was assigned an adjusted p-value.

No genes were enriched in the *ARID1B, ARID2* or *PBRM1* mutant lines in either dataset. In addition to the previously identified enrichment for *ARID1B* essentiality in the *ARID1A* mutants (Section 4.2.2), *WRN* was also significantly enriched in both datasets (Sanger adjusted p-value = $2.23 \times 10^{-9}$, Broad adjusted p-value = 0.0025). *RPL22L1* was significant in the Sanger screens (adjusted p-value = 0.0015) but not in the Broad screens, after the multiple testing correction (p-value = 0.0006, adjusted p-value = 1). Neither of these genes were identified in the iPSC screens.

It has been shown that dependency on *WRN* is highly associated with microsatellite instability (MSI) status[109,125]. In the Sanger dataset, 27/324 lines were MSI-high; 18 of these were *ARID1A* mutants and 15/18 had a *WRN* dependency. Considering this, the association between *ARID1A* and *WRN* could be an artefact caused by the enrichment of MSI-high status in *ARID1A* mutants. Indeed, when MSI-high *ARID1A* mutants were removed, there was no significant association between *ARID1A* mutation status and *WRN* essentiality.

We cannot assume from these analyses that no SLIs occur with the PBAF/BAF genes, especially considering that other dependencies in *ARID1A* mutants have been published. The small sample size available for these mutants is likely a limiting factor. It is also possible that many dependencies may be specific to a few cell lines or subtypes, and so cannot be identified in such a broad dataset.

## 4.7 Summary

Based on data from screening BOB iPSCs that were knockout for *ARID1A, ARID1B, ARID2* or *PBRM1*, we attempted to validate genes that were specifically essential in the KO lines but not in the parental line. We initially sought to validate an interaction between *ARID1A/ARID1B* that was previously identified in cancer cell lines. Our preliminary data suggests that loss of these genes is not synthetic lethal in BOB cells. Analysis of CRISPR/Cas9 screen data from two large cancer cell line studies also indicated that the interaction may be uni-directional, with an apparent lack of dependency on *ARID1A* in *ARID1B*-mutant lines. Further investigation would be required to confirm and characterise this.

We selected a list of candidate SLIs to validate in both iPSCs and a cancer cell line, HAP1. We were unable to complete validation in the iPSCs; this may be because the candidates were false positive hits, but technical issues made it difficult to make any conclusions so further experiments are required. Validation assays in the HAP1 cells produced unexpected results, with the WT line appearing to be more dependent on many genes than the PBAF/BAF gene KO lines. Several candidates did appear to be more essential in the KO lines and warrant further investigation. HAP1 cells were chosen as a model to ensure that any SLIs that we identified were not specific to iPSCs. However, it may be more informative to validate SLIs in a large panel of cancer cell lines, with particular focus on those that have existing mutations in the PBAF/BAF genes as these are most clinically relevant. As demonstrated by the *ARID1A/ARID1B* data, SLIs are not consistently observed in every mutant cell line, thus finding the right context for validation may be challenging.

# Chapter 5

# Discussion and future directions

## 5.1 Summary of findings

In this project, we performed genome-wide CRISPR/Cas9 knockout screens in a panel of isogenic iPSCs with the aim of identifying novel synthetic lethal partners of 15 tumour suppressor genes. We engineered a panel of iPSC lines that each had LOF in a single TSG, prepared a genome-wide gRNA library suitable for screening in these cells and optimised a protocol to do so (discussed in Chapter 2). We screened the parental iPSC line and 21 KO lines. Analysis of the data revealed that whilst known fitness genes could be identified, the outputs were highly variable, even in replicates of the same cell line (discussed in Chapter 3). It was evident that this variability was likely caused by Cas9-induced toxicity, as a result of DSB formation. Despite this, we performed further analysis to identify candidate synthetic lethal partners of each TSG. Focusing on four TSGs that were members of the PBAF/BAF complex, we then attempted to experimentally validate hits from the screens (discussed in Chapter 4). Due to experimental issues, we were unable to complete validation in iPSCs but preliminary results in an independent cancer cell line highlighted several candidates for further study. We also analysed published cancer cell line screen datasets to identify potential SLIs involving the PBAF/BAF genes but found no novel significant associations.

# 5.2 CRISPR/Cas9 screening in iPSCs

## 5.2.1 Comparison with published stem cell screens

In the last 18 months, three groups have published genome-wide screens in hESCs,[238-240] but no iPSC screens have been published. Due to the lack of available literature at the start of this project, we did not anticipate the challenges that arose during iPSC screening. However, it is evident from these recent publications that the issues we faced were not specific to this project. In each of these studies, screens were performed with different lentiviral genome-wide gRNA libraries.[238-240] There were various other experimental differences which may influence the comparability of these studies, such as the substrate used to culture the cells, the Cas9 system and the library coverage (Table 5.1). Mair *et al.* (2019) repeated screening in the same cells using two different culture systems (mouse embryonic fibroblast feeder cells (MEFs) and laminin) and found that the results differed considerably. This highlights the fact that culture conditions can influence the identification of genes required for cell fitness, emphasising the need for cross-validation of results from CRISPR/Cas9 screens.

**Table 5.1. Comparison of stem cell screens.**

| Study | Cell line | Culture substrate | Cas9 system | Target library coverage |
|---|---|---|---|---|
| This project | BOB, human iPSC line | Vitronectin | Constitutive Cas9 | ~150x |
| Yilmaz *et al.* (2018) | h-pES10, haploid hESC line | MEFs | Delivered in same vector as gRNAs | ~700x |
| Mair *et al.* (2019) | H1, hESC line | MEFs and laminin | Inducible Cas9 | ~400 |
| Ihry *et al.* (2019) | H1, hESC line | Vitronectin | Inducible Cas9 | ~100x |

The outputs of these screens were analysed using similar but not directly comparable systems. To allow for comparison, Mair *et al.* (2019) re-processed the results from the other studies to match their analysis and calculated Bayes Factors for each gene. They assessed the overlap across all screens. Additionally, they considered the overlap with 1580 core fitness genes previously established by Hart *et al.* (2015) using CRISPR/Cas9 knockout screens in immortalised cell lines. Excluding these core fitness genes, a total of 36 genes were found to be essential for cell fitness in all of the stem cell screens. By cross-referencing these with genes that were significantly depleted in any of our three BOB screen replicates (based on BAGELR

analysis), 17 genes were common to our iPSC line and the hESC lines (Table B.1, Appendix B). As discussed in Section 3.3.1, BAGELR varies slightly from the original BAGEL and so re-processing of either our data or the published datasets may allow a more direct comparison and could increase this overlap. It is reassuring that we identified overlapping hits and our data may help in establishing a list of core stem cell fitness genes, which could be a useful reference for future studies. Colleagues at the WSI are performing genome-wide knockout screening in iPSCs from different human donors. They are using an alternative library using fewer gRNAs/gene and screening at a larger coverage than we did. It will be interesting to compare with their results, particularly as they may have significant variability between donors. This may also allow us to identify hits that are common to iPSCs but differ from ESCs.

### 5.2.2 Technical issues with screening stem cells

The largest issue we faced in our screens was Cas9-induced toxicity, which manifested in several ways. Vast cell death occurred in the days immediately following transduction, and there was a significant enrichment of NTC gRNAs (Section 3.2.2) and gRNAs targeting genes such as *TP53*, which play a role in the DNA damage response (Section 3.8). Another group recently published similar findings regarding the toxicity of Cas9 in a hESC line and explored this in more depth.[241] Using 47 gRNAs targeting 16 genes, they found that editing was highly efficient but the majority of cells did not survive. They selected a gRNA targeting a gene that was not expressed in the cell line and compared this with an NTC gRNA. The targeting gRNA caused a decrease in confluence over time, whereas cells expressing the NTC gRNA had increased confluence. They considered the possibility that this effect could be due to off-target activity, but found no evidence of editing at the top 6 predicted off-target sites. They also trialled transient Cas9 exposure and use of a Cas9 with enhanced specificity to reduce off-target activity, but toxicity was still observed.

In the same study, a high coverage (1000x) screen was performed using a library with 13,000 gRNAs, both in the presence and absence of Cas9 activity. In cells where Cas9 was not induced, the fold-changes for most gRNAs were evenly distributed, but they observed a 1.3-1.4 enrichment of NTC gRNAs when Cas9 was active. This suggested that there was a global depletion of targeting gRNAs. Mair *et al.* (2019) also observed toxicity when Cas9 was induced during screening, and noted an enrichment of non-targeting gRNAs. These data correlate well with the enrichment of NTC gRNAs observed in our screens (Fig. 3.2). Ihry *et*

*al.* (2018) analysed the fold-change of NTC gRNAs in screens of immortalised or tumour cell lines and found little enrichment, indicating that this toxicity effect is heightened in stem cells.

Ihry *et al*. (2018) demonstrated that CRISPR/Cas9 knockout did not affect expression of pluripotency markers, which supports data produced in this project (Section 2.8) and in a previous study.[273] Using RNA-seq analysis, they compared cells expressing a targeting gRNA vs an NTC gRNA and found that many genes involved in apoptosis were up-regulated in the DSB-induced cells. Interactome analysis indicated that p53 was linked to many of the expression changes associated with DSBs. They demonstrated that *TP53*-mutant cells increased in confluency despite DSB induction. However, the growth was ~50% less than in cells where Cas9 was not induced, suggesting that there was still an effect on cell fitness. In our screens, the NTC gRNAs were also enriched in a *TP53* KO line (Fig. 3.2), indicating that the toxicity effect was not solely dependent on the p53-mediated DNA damage response. However, we did find that the *TP53* KO screens were more reproducible than the parental (Section 3.6.1), had the highest recall of known core fitness genes (Section 3.4.3) and showed greater separation of established essential and nonessential genes (Section 3.4.2).

Similar to our findings, all of the published genome-wide screen studies in hESCs observed enrichment of *TP53*-targeting gRNAs.[238-240] Other genes were also identified, including *PMAIP1* and *CHEK2*, which we found to be recurrently enriched in our iPSC screens (Fig. 3.12). Ihry *et al.* (2019) performed further analysis of *PMAIP1*, a pro-apoptotic regulator downstream of p53, and found that this enrichment appeared to be specific to stem cells. Experimental validation confirmed that mutation of *PMAIP1* reduced Cas9-induced toxicity.[238]

### 5.2.3 Potential improvements to screening in iPSCs

Our findings and those of the few other related studies have been informative with regards to the ways in which stem cell screening protocols could be improved in future. Data regarding gene enrichment provides a better insight into the toxic response observed in stem cells and could provide strategies for improving the application of CRISPR/Cas9 in these cells. For example, transient inhibition of these DNA damage response pathways could help improve survival. However, even temporary interference with these genes would likely alter the results of any screen and would need to be considered when interpreting any results. Another key revelation is that NTC gRNAs are not an ideal control, which may also be relevant for other cell types. As an alternative, safe-targeting gRNAs could be used. These target genomic regions which have no functional impact, but still induce DSBs and so would act as a better control.

One option to improve the quality of our iPSC data would be to increase the coverage (i.e. transduce and maintain more cells per gRNA), potentially making it easier to differentiate between general depletion due to Cas9 toxicity and specific depletion due to gene essentiality. As discussed in Section 5.2.1, our initial coverage was much lower than others have used. The scale of our screening protocol was largely dictated by G418 selection, as the antibiotic did not work effectively when transduced cells were seeded densely. Untransduced iPSCs were highly sensitive to G418 selection regardless of density, but we found that the presence of resistant cells in the population impaired the selection capability. Cells were seeded sparsely to allow selection to occur and for cells to proliferate for more than a couple of days post-transduction. Due to the issues with measuring BFP expression (Section 2.7.3), it was difficult to confirm that selection was complete and if so, when this was achieved. However, there appeared to be no significant cell death after 1 week. In comparison to puromycin, G418 was much less efficient. Previous work in our lab (not described here) has shown that puromycin can effectively select these cells within 48 hours, at a density 4-fold higher than that used in our screens. Thus, it is not surprising that the majority of available gRNA library backbones have a puromycin selectable marker.

We were unable to use a puromycin-resistant library backbone because the KO iPSC lines were engineered to carry a puromycin cassette (Section 2.3). The cells were also resistant to blasticidin, which was used as a selectable marker for Cas9 expression (Section 2.6). In addition to G418, hygromycin B and zeocin antibiotics can also be used for selection of mammalian cells. Based on past experience, colleagues informed us that zeocin was difficult to use. Clara Alsinet had previously tested hygromycin B in the BOB iPSC line (data not shown) and found that G418 worked better, hence our decision to use this antibiotic. The cost of screening in iPSCs was significant due to the requirement for vitronectin coating on all culture dishes and daily medium changes. Performing screens at this scale also required a lot of time and incubator space. Thus, increasing the scale of the screen to improve library coverage was not feasible. However, if we had used puromycin selection, approximately 4-fold greater library coverage could have been achieved with more cells seeded per dish. In hindsight, using an alternative selectable marker for the KO iPSCs would have been preferable but engineering of these lines began before the screening aspect of this project was planned.

Another option to address the toxicity issue could be to screen with CRISPRi technology, which inhibits gene expression by fusing repression domains to a catalytically inactive Cas9. This would avoid the complication of DSB-induced toxicity as Cas9 would not cut the DNA. In theory, this should improve the quality of screen data by removing noise associated with

Cas9 toxicity. A higher library coverage could be achieved without the necessity to scale up to account for cell death, saving both time and money. One potential limitation of CRISPRi is that transcriptional repression may be less effective at reducing protein expression than a loss-of-function mutation. Mandegar *et al.* (2016) compared CRISPRi with CRISPR knockout in iPSCs using gRNAs targeting *OCT4* and *NANOG*. They found that CRISPRi was in fact more efficient, with almost complete loss of protein expression, whereas approximately a third of cells maintained expression using CRISPR knockout. Sequencing analysis highlighted that 30-50% of these cells had in-frame indels, which would explain the lack of protein knockout. Various others have shown successful application of CRISPRi in iPSCs and iPS-derived cells.[274-279] However, these studies have involved either targeted genes or more focused screens, not on a genome-scale. In future, a genome-wide CRISPRi screen could be performed in the parental BOB line to compare the results with our knockout screens.

Despite the challenges we faced, in some of our screens we were able to detect known essential genes and identify genes that were identified as being specific to stem cells in published studies. However, considering our aim was to identify specific genetic changes between isogenic cell lines, the variability we observed was a significant limitation. The potential improvements discussed here may improve the reproducibility of iPSC screens and allow more reliable identification of genetic interactions.

## 5.3 Screening for PBAF/BAF dependencies

To date, research into targeting PBAF/BAF mutant cancers has been largely focused on the most commonly mutated subunit, *ARID1A*. None of the previously reported dependencies associated with *ARID1A* loss were identified in our iPSC screens. This could be due to the quality of the screen data, as the *ARID1A* KO screens did not perform as well as others (Section 3.4). However, it could also be due to context-specificity of the previously identified synthetic lethal interactions. The most widely validated synthetic lethal partner of *ARID1A* is *ARID1B.* As we did not detect this in our screens, we performed further experiments to test this but found no evidence of an interaction (Section 4.2.1). More extensive validation could be performed, using more gRNAs or possibly using si/shRNA technology, however it is possible that this interaction does not occur in the BOB cell line. Many of the other *ARID1A* dependencies were found in OCCC cell lines and were not broadly validated in other tissue types, therefore it is possible that they are also specific to certain cell lineages. No studies have demonstrated systematic screening for synthetic lethality in *ARID1B, ARID2* or *PBRM1* mutants so we had little reference for comparing the output of our screens. We analysed large-scale pan-cancer CRISPR/Cas9 screen datasets but found no dependencies associated with LOF mutation in any of these 3 genes, although this may be due to the small number of mutants (Section 4.6.2).

In our screens, we identified candidate SLIs for all 4 PBAF/BAF subunits and reassuringly, some hits were identified in two independent KO clones for the same gene (Section 4.6.1). Due to the variability of our data and incomplete validation, it is difficult to draw any conclusions regarding these. Further experiments are required to confirm that these interactions occur in the iPSCs, and then these must be assessed in cancer cell lines to confirm they are not stem cell-specific. We chose to validate in a haploid CML line (HAP1) and had potentially promising preliminary data for some hits (Section 4.5) but repetition of this assay and further validation is needed. For clinical relevance, it would be ideal to instead focus on cell lines that represent the tissues most commonly affected by loss of these genes e.g. ovarian cancer for *ARID1A* and renal cancer for *PBRM1*. One possible strategy would be to harness the available CRISPR/Cas9 screen data to select cell lines for validation. Hits could be cross-referenced with screen data from PBAF/BAF mutant cancer lines. This may give an indication of whether a hit would validate without having to test a large panel of cancer cell lines. Although the small sample size could be an issue here, and it must be kept in mind that false negatives can occur in these screens so some hits may be missed.

## 5.4 Isogenic models for SLI screening

Choosing the right context to identify and validate candidate SLIs is one of the biggest challenges of not only this project, but synthetic lethality studies in general. In this project we chose iPSCs as a model as they have a clean genetic background and should therefore have limited confounding factors. Considering the limitations discussed in Section 5.2, further validation experiments are needed to assess how effective they were as a model. However, I expect that although technical improvements may be required, screening in iPSCs can be used to identify novel genetic interactions with higher confidence than in a cancer cell line with many genetic aberrations. The main barrier may not be the model, but rather that screening in any single cellular context is not sufficient to identify broadly applicable SLIs.

Analysis we performed on pan-cancer CRISPR/Cas9 screen datasets supported previous findings that *ARID1A* mutant cells are dependent on *ARID1B* (Section 4.2.2). However, it is clear from these analyses that, even though this was the strongest association identified in *ARID1A* mutants, this interaction is not fully penetrant. In this context, penetrance refers to the fraction of tumour cell lines harbouring a genetic mutation that are sensitive to inhibition of a synthetic lethal partner.[62] In the Sanger screen dataset, only 15% of *ARID1A* mutant lines were dependent on *ARID1B*; this was slightly higher at 30% in the Broad dataset (Fig. 4.5). This is not a unique observation; many SLIs have incomplete penetrance, and this is likely one of the main reasons that only one interaction has progressed to the clinic.[62] For example, several large-scale screens have been performed to identify synthetic lethal partners of *KRAS,* but few hits have replicated across multiple studies (as reviewed by Downward, 2015).[280] It is possible that the lack of reproducibility is due to technical issues, but even studies using the same screening method in different cell lines have shown a high degree of cell line-specificity for SLIs.[129] It is likely that many SLIs are either cell-type specific or specific to a certain genetic context, with other genes interfering with the interaction.

Screening across a panel of cancer cell lines and associating results with genetic mutations has evidently been useful in identifying SLIs such as the *ARID1A/ARID1B* pairing. However, this is limited by the number of cell lines available with existing mutations. The large genetic heterogeneity also makes it difficult to confidently deduce which, if any, single genetic change is responsible for the effect observed across all lines. The results in an isogenic system can be interpreted more clearly. Going forward, I would propose that an alternative strategy could be to combine both strategies by screening across multiple isogenic pairs. I would select cell lines that have naturally occurring LOF mutations in the PBAF/BAF genes but are

otherwise genetically heterogenous. CRISPR/Cas9 could then be used to correct the mutation and engineer a wildtype derivative. This strategy would allow robust identification of SLIs that are not specific to one cell line or tissue type, which is vital for any potential targets to be widely applicable in the clinic.

## 5.5 Future perspectives

In this project we produced a large amount of iPS cell screen data; screening in this cell type has not been shown in the literature and there are very few datasets from other pluripotent cells. Our data has given an insight into the significant variability that occurs when screening this cell type. This has been valuable in emphasising how important it is to replicate and validate screens thoroughly. In Section 5.2.3 I proposed several ways to address the issues we faced and I would recommend that increasing library coverage and/or using a technology such as CRISPRi would be the best approaches to deal with the issue of cell death. It was also evident that non-targeting gRNAs were poor controls as they did not induce a DNA damage response and were positively selected. For future screens, safe-targeting gRNAs which cut DNA may be a better control. We tested the pluripotency of the cells post-screen but in hindsight, it would have been valuable to measure expression of pluripotency markers during the screen whilst the cells were stressed and undergoing morphological changes. Our aim was to use this as a model rather than to understand the biology of the stem cells, but validating the state of the cells would be particularly pertinent for researchers interested in using screen data to study iPSCs. We did not have the opportunity to screen another iPSC line for comparison, but another group at WSI is screening multiple lines using a higher coverage and an independent library. We hope that integration of these datasets will provide a starting point for identifying core and patient-specific genes that are essential for iPS cell fitness.

Further validation is needed to determine the utility of our data for identifying synthetic lethal interactions. Whilst there may be true interactions present in the results, the variability makes the data difficult to interpret with confidence. It would have been valuable to do additional biological replicates of the knockout lines, as we did with the parental line. We did not detect known interactions with the PBAF/BAF genes but as these were not tested independently in our cell line, we could not conclude that this was a fault of the screen. We performed experiments to test the *ARID1A/ARID1B* interaction but going forward I would repeat this and test other published interactions using other methods e.g. siRNA and chemical inhibitors. Confirming the presence or absence of these interactions in the BOB cell line would

be a good indicator of screen performance. For future validation, I think it would be most valuable to test our candidate hits in cancer cell lines that have existing mutations in the gene of interest e.g. *ARID1A*-mutant ovarian cancer cell lines. Although HAP1 has technical advantages, it does not reflect the clinical setting that these interactions occur in. As a haploid cell line, it also removed our ability to identify interactions with haploinsufficient genes.

When it became clear that CRISPR/Cas9 KO screening in iPSCs would be challenging and the initial data was variable, I also became involved in another project in the Adams' lab. In doing so, I gained experience using a different screening technology in a different model system. I performed these experiments in parallel with my own project, whilst I analysed my screen data and carried out validation assays. In Chapter 6 I have detailed the background to this additional project, the results we obtained and the subsequent work that is ongoing.

# Chapter 6

# *In vivo* CRISPR screening to identify tumour cell intrinsic regulators of metastatic colonisation

## 6.1 Introduction

### 6.1.1 What is metastasis?

It is estimated that 90% of cancer-associated deaths are caused by metastasis.[281] Metastasis is a process whereby cancer cells spread from the primary tumour to a secondary organ. Multiple stages are involved: invasion of surrounding tissue, intravasation and survival in the circulatory system, extravasation and proliferation in a distant tissue (in a process known as 'colonisation') (Fig. 6.1). Metastasis is controlled by both tumour cell intrinsic factors and extrinsic factors in the tumour microenvironment (reviewed by Liu, *et al.,* 2017).[282] Many studies have been carried out to identify such factors but we still have a lot to learn about this process, particularly the colonisation step.

### 6.1.2 Metastasis models

Due to the biologically complex nature of this process, animal models are critical for studying metastasis. In particular, the role of the immune system in metastasis makes it difficult to accurately study this process *in vitro*. Research has shown that the immune system can act both to inhibit and promote metastasis at various stages of the cascade. Cytotoxic CD8+ T cells, NK cells and M1-like macrophages have inhibitory roles and will recognise and kill tumour cells.[283] However, tumour cells can develop mechanisms to evade the immune response, for example by downregulating the expression of MHC class I molecules which are recognised by CD8+ T cells.[284] T regulatory cells and M2-like macrophages can also help to promote metastasis by inhibiting CD8+ T cells and NK cells.[283] Immune cells play various roles throughout the metastatic cascade (as reviewed by Blomberg, *et al*., 2018).[285] In the early stages of

dissemination from the primary tumour, immune cells regulate the extracellular matrix, formation of blood and lymph vessels, and can promote migration. After intravasation, circulating tumour cells are under further attack by NK cells and T cells and must adapt to evade these responses. However, some immune cells can aid survival and assist with extravasation to distant sites. For example, interaction with platelets in the circulation can shield tumour cells from NK cells[286,287] and can also promote an invasive phenotype in the tumour cells.[288] If cells survive these pressures, they must then establish another immunosuppressive environment at the secondary site to allow further growth. The interaction between tumour cells and the host environment, particularly the immune system is pivotal in determining the success of metastasis. Identifying proteins that are up- or down-regulated in tumour cells to allow them to metastasise is vital for developing therapies to prevent this.

One of the most common *in vivo* strategies is to transplant cancer cell lines or tissues into mice.[289] This can be done using cells/tissues derived from an animal with the same genetic background (syngeneic) or using human cells/tissues in immunocompromised mice (xenograft). The interaction between the tumour and the host is critical in metastasis. The syngeneic model is ideal for studying this as the tumour cells are from the same species as the host, whereas human xenograft models may face species-specificity issues.

Two experimental approaches are used with transplantable models: experimental metastasis assays (injection of tumour cells directly into the circulation) and spontaneous metastasis assays (injection of tumour cells into a tissue from which they then have to enter the circulation). Experimental metastasis assays allow analysis of the later stages of metastasis (extravasation and colonisation), whereas spontaneous assays follow all stages from the primary site. The spontaneous model is more clinically relevant but the experimental model is quicker and avoids issues with the primary tumour mass growing too large (and the mouse having to be sacrificed) before the cells have had time to enter the circulation and metastasise.

One of the most commonly used experimental models involves intravenous injection of B16 mouse melanoma cells into the tail vein, typically leading to pulmonary metastases.[290] B16 tumour cells were derived from a spontaneous melanoma that occurred in the C57BL/6 mouse strain. Several derivatives of this line have been generated with varying metastatic potential and sites of colonisation.[291,292] This model (using B16-F10 cells; a highly metastatic derivative[290]) has been used extensively in the Adams' lab.[293]

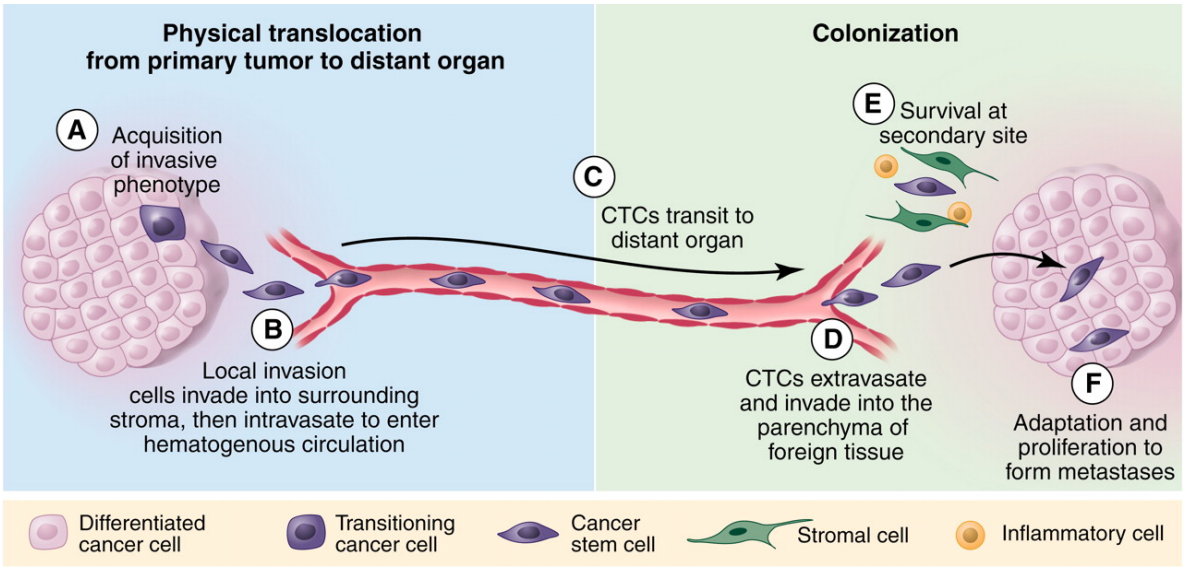



**Figure 6.1. Metastatic cascade.** Tumour cells acquire an invasive phenotype **(a)**, begin to invade surrounding tissue and then intravasate into the vasculature **(b)**. Circulating tumour cells (CTCs) travel through the circulatory system **(c)** until they extravasate into a distant tissue **(d)**. Here, they must survive challenges such as the innate immune system **(e)** and adapt to the new microenvironment, allowing them proliferate and form metastases **(f)**. Figure taken from [281].

### 6.1.3 Screening for regulators of metastasis

By applying high-throughput screening strategies to established *in vivo* models, we can interrogate the metastatic process in a systematic and unbiased manner. A recent study identified tumour extrinsic ('host') factors that regulate metastasis by carrying out an experimental metastasis assay in 810 genetically-modified mouse lines.[293] Tumour intrinsic factors have also been identified by applying functional genomic approaches, such as shRNA and cDNA libraries, to cells *in vitro* then observing the phenotypic effect in *in vivo* metastasis models.[294-296] More recently, CRISPR/Cas9 technology has been used for various loss-of-function screens *in vivo,*[297-299] including the study of regulators of tumour growth and metastasis in mice.[300]

### 6.1.4 CRISPR activation

Originally a tool developed to allow for genome editing[87,88], CRISPR/Cas9 has been repurposed for many applications, including gene regulation. A catalytically inactive version of Cas9 (dCas9) was engineered by introducing point mutations into the RuvC1 and HNH nuclease domains.[96] dCas9 is unable to cleave DNA but can still be targeted to a genomic region by a gRNA. By fusing transcriptional regulation domains to dCas9, the expression of

targeted endogenous genes can be controlled.[96,97,301] Thus, CRISPR/Cas9 can be used to investigate phenotypes associated with gene activation as well as silencing (Fig. 6.2). In comparison to KO screens, so-called CRISPR activation (CRISPRa) is still in its infancy, with few published *in vivo* studies.[302-304] When combined with an *in vivo* experimental model, CRISPRa technology offers a novel approach to screen for regulators of metastasis.

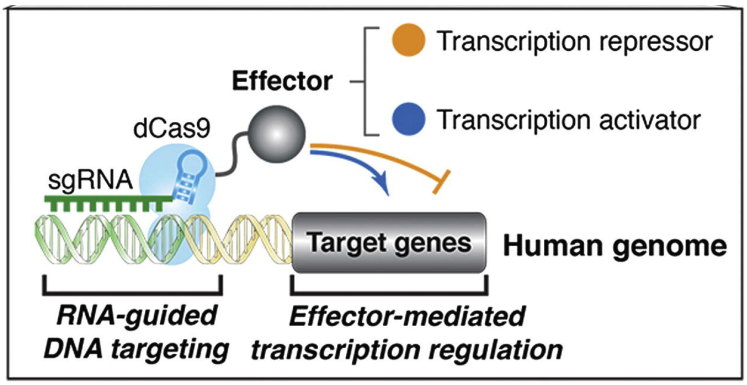



**Figure 6.2. CRISPR-mediated transcriptional regulation.** Catalytically inactive dCas9 can be fused to an effector domain and when bound to a gRNA, the protein is directed to a specific genomic region. Depending on which effector domain is used, transcriptional activators or repressors will be recruited, leading to regulation in expression of an endogenous gene. [Figure adapted from Qi et al.] [96]

### 6.1.5 Aims of this project

The genes and signalling pathways that must be initiated in order for a melanoma cell to survive and proliferate at a secondary site are not well understood. In this project, we aim to identify and characterise the genes that are upregulated in a melanoma cell in order for it to effectively colonise the lung. To do this we will:

- Use a CRISPRa library of gRNAs against mouse cell surface genes. We have chosen to specifically target cell surface proteins as they represent clinically tractable targets for immunotherapy, similar to the rationale for Trastuzumab (Herceptin), in which the monoclonal antibody binds to the HER2 receptors on breast cancer cells and both inhibits the normal function of the HER2 receptor and induces immune-mediated killing.[10]
- Use the B16-F0 mouse melanoma cell line. We chose this cell line as it is weakly metastatic, thus only having a low level of ability to effectively colonise the lung after tail vein dosing, and will allow us to identify genes that can enhance this ability.
- Perform a CRISPRa screen *in vivo* using the experimental metastasis assay. We chose this assay as using an *in vivo* protocol will ensure key regulators of metastatic colonisation are

present (the immune system, stroma, etc.) and the experimental metastasis assay avoids the gRNAs also having an effect on primary tumour growth (which would have been an issue with the spontaneous metastasis assay).

- Develop a suitable analysis pipeline for *in vivo* CRISPRa screen data.
- Validate candidate enhancers of metastatic colonisation identified by the screen.

This project was led by Dr Louise van der Weyden (LvdW), a Senior Staff Scientist in the Adams' lab. I carried out all lentiviral production, transductions, cell culture and flow cytometry during the *in vitro* phase of the screen and subsequent validation experiments. All other work was performed by LvdW, who has kindly allowed me to include this unpublished information.

# 6.2 Results

## 6.2.1 Preparation for genome-wide CRISPRa screen

The Weissman lab designed a genome-scale gRNA activation library targeting the mouse protein-coding transcriptome, designated mCRISPRa-v2.[305] This was designed using a comprehensive model to predict gRNAs with high activity, based on features such as nucleosome occupancy and TSS position. They also established subpools of this CRISPRa library which target specific functional gene groups. As we were interested in cell surface genes, for this project we used the 'membrane proteins (m6)' subpooled library; this consists of 10,975 gRNAs targeting 2,104 genes that encode for membrane proteins, and 250 non-targeting control gRNAs.

The m6 library was acquired from Addgene and expanded as described in Section 7.17.2 (backbone shown in Fig. 6.3). Plasmid DNA was extracted and packaged into a lentiviral vector. LvdW transfected B16-F0 mouse melanoma cells with an activator construct that contained two VP64 activation domains fused to dCas9 (Addgene #113341).[306] Blasticidin was used to select for and maintain the dCas9-expressing population (termed 'B16-F0-dCas9' cells). A titration was performed by transducing B16-F0-dCas9 cells with various volumes of the m6 library lentivirus (as described in Section 7.17.3). After 48 hours, BFP expression was measured by flow cytometry. The aim was to achieve an MOI of 0.2-0.5, to ensure that each cell carried only a single gRNA. The volume of lentivirus required to obtain an MOI of 0.3 (30% BFP-positive cells) was calculated and scaled up for the screen.

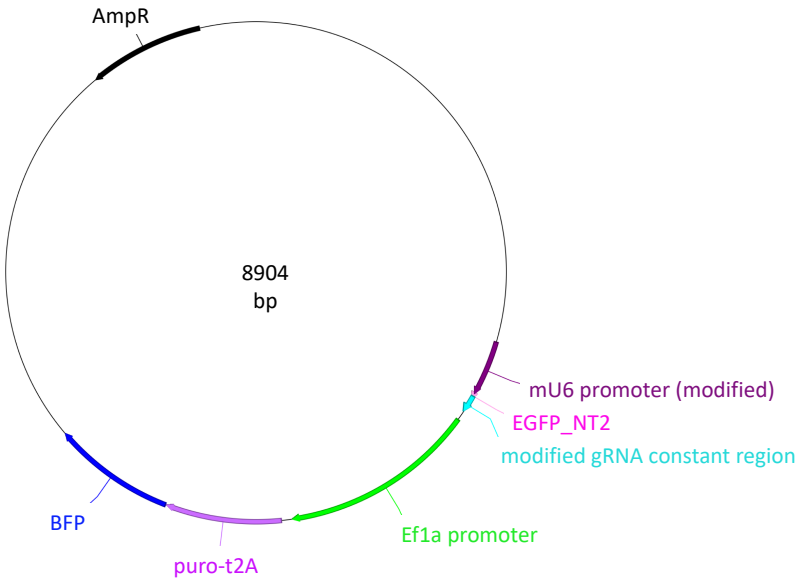



**Figure 6.3. pCRISPRia-v2 plasmid map.** The key features of the Weissman lab library backbone plasmid are shown. gRNAs are under the control of a mouse U6 (mU6) promoter. An Ef1a promoter drives expression of puromycin resistance (puro-t2A) and BFP.

### 6.2.2 Screening in an experimental metastasis model

The highest library coverage we could achieve per mouse was 50x ($5.5x10^5$ cells), as tail vein administration of a higher amount of this cell line could result in a tumour embolism that would be lethal to the mouse. We aimed for a higher *in vitro* coverage to minimise loss of coverage prior to moving *in vivo*. The initial target was ~300x and so $12x10^6$ cells were transduced at an MOI of 0.3. Details of the full screening protocol are provided in Section 7.17.4. After 48 hours, cells were passaged and puromycin was added to select for successfully transduced cells. A small sample was taken for flow cytometry; a transduction rate of 20% was achieved i.e. 200x library coverage (Fig. 6.4). Although lower than expected, this was still largely in excess of the required 50x and so we proceeded with the screen. Cells were passaged after 4 days in culture with puromycin and flow cytometry confirmed that selection was complete, with an almost 100% BFP-positive population (Fig. 6.4). Cells were collected for dosing at 9 days post-transduction, allowing enough time for the lentivirus particles to be cleared and for gene activation to occur.

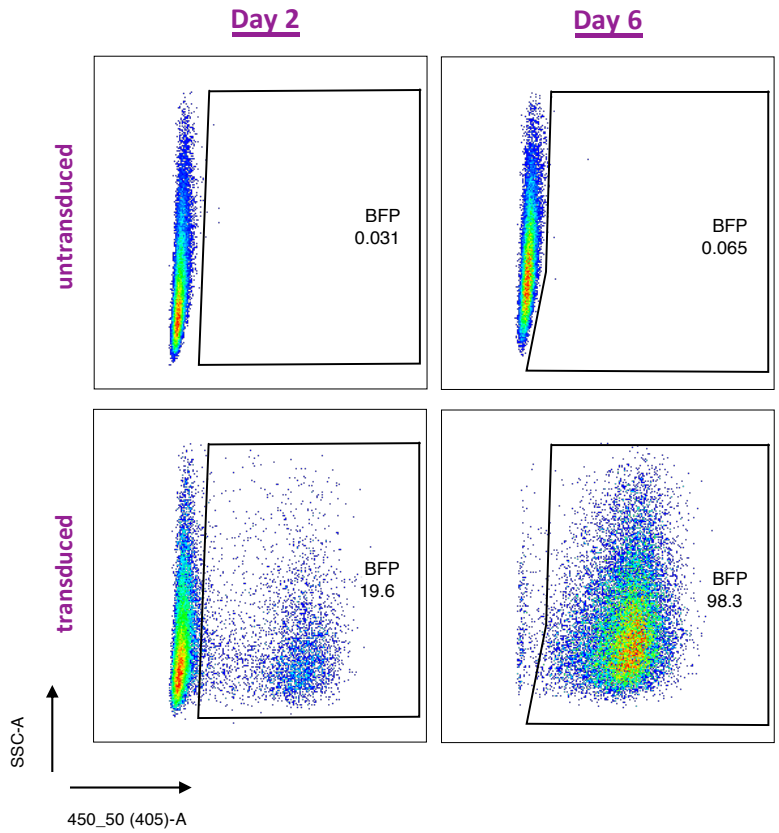



**Figure 6.4. Expression of CRISPRa library.** B16-F0-dCas9 cells were transduced with the m6 library lentivirus and analysed by flow cytometry on day 2 post-transduction. Puromycin was added and cells were analysed again on day 6 post-transduction. BFP expression (450_50 (405)-A) was measured, using untransduced cells as a control to gate the positive population.

After dilution in phosphate-buffered saline (PBS), $5.5x10^5$ cells were intravenously administered into the tail veins of 70 mice. As a control, some cells were also pelleted and frozen so that the baseline gRNA abundance could be measured. Activation of some genes may have altered cell proliferation/survival *in vitro*, therefore it was important that these effects were accounted for and did not interfere with the signal from gRNAs that impacted metastatic ability *in vivo*. Lungs were collected from mice at 2 timepoints: 35 mice were collected at 4 hours post-dosing (to determine what proportion of gRNAs successfully made it into the lungs) and 35 mice were collected at day 19 post-dosing (by this point culling was necessary as mice began to show symptoms associated with pulmonary tumour burden, such as rapid breathing). The entire lungs were collected and homogenised, and a portion of the homogenate taken for genomic DNA (gDNA) extraction. PCR was performed on the gDNA to amplify the gRNAs present in each sample (lung). PCR was also performed on gDNA from the cells that were collected on the day of dosing, and on the library plasmid DNA. The gRNAs present in each sample were identified by sequencing on a HiSeq 2500.

### 6.2.3 Analysis of CRISPRa screen data

As there is no well-established method for analysing *in vivo* CRISPRa screens, Vivek Iyer (a senior bioinformatician in the Adams' lab) trialled three independent analysis strategies to identify potential hits. The results from these methods were variable with little overlap overall. Design of gRNAs for CRISPR activation is challenging as transcription start sites are not well annotated for many genes. A pilot experiment performed by LvdW demonstrated that only 1 in 5 of the gRNAs in the m6 library that targeted the *CD8a* gene caused up-regulation of the CD8a protein (Table 6.1). Considering this, combined with the additional noise of doing a low coverage screen in a difficult setting (*in vivo*), two of the three bioinformatic analyses focused on the individual gRNAs rather than combining gRNAs to get a 'per gene' result.

**Table 6.1. Up-regulation of *CD8a* using m6 library gRNAs.** B16-F0-dCas9 cells were transfected with gRNAs targeting *CD8a* or non-targeting controls (NT_1-3). Expression of BFP and CD8a were measured by flow cytometry, using untransfected cells as a control for gating. BFP was used as a measure of the transfection rate, acting as a proxy for gRNA expression. The % of cells that were positive for each protein is shown. This experiment was performed by LvdW.

| gRNA | % BFP+ | % CD8a+ |
|---|---|---|
| untransfected | 0.11 | 0 |
| NT_1 | 77.4 | 0.13 |
| NT_2 | 31.6 | 0 |
| NT_3 | 86.1 | 0.14 |
| CD8a_1 | 86.4 | 63.8 |
| CD8a_2 | 71.2 | 0.2 |
| CD8a_3 | 78.6 | 0.19 |
| CD8a_4 | 75.8 | 0.51 |
| CD8a_5 | 51.2 | 0.7 |

#### 6.2.3.1 '98$^{th}$ percentile' method

For each day 19 mouse sample, gRNAs were ranked by their relative abundance (based on read count) and those in the 98$^{th}$ percentile (top 2%) were recorded. Those gRNAs that ranked highly in multiple mice were of most interest. Any gRNA which was present in < 50% of the day 19 samples, regardless of ranking, was disregarded. Additionally, a Z-score was calculated for each gRNA, comparing read counts from day 19 to the cells collected prior to dosing (day 0):

$$\text{Z-score} = \frac{\text{(average read count at day 19) - (average read count at day 0)}}{\text{Standard Error of the Mean for day 19 samples}}$$

To identify potential hits, gRNAs were first ranked by the number of mice in which they were present in the 98$^{th}$ percentile, and then by their Z-score (Table 6.2).

**Table 6.2 Screen analysis: '98$^{th}$ percentile' method.** Top 10 hits from analysis using the 98$^{th}$ percentile method.

| gRNA | Present in 98% percentile (no. of mice) | Present in day 19 sample (no. of mice) | Z-score |
|---|---|---|---|
| Lrrn4cl_+_8850757.23 | 4 | 17 | 3.19 |
| Slc4a3_+_75546398.23 | 3 | 20 | 3.92 |
| Tmem194b_+_52630698.23 | 2 | 23 | 2.97 |
| Zmynd12_+_119422735.23 | 2 | 19 | 2.09 |
| Fut2_-_45666416.23 | 2 | 14 | 2.09 |
| Smco3_+_136835494.23 | 2 | 22 | 2.05 |
| Tango6_+_106683055.23 | 2 | 14 | 2.01 |
| Gpr27_+_99692177.23 | 2 | 26 | 1.78 |
| Olfr1360_+_21675377.23 | 2 | 18 | 1.67 |
| Gpr75_-_30885367.23 | 2 | 21 | 1.67 |

### 6.2.3.2 JACKS method

A second analysis using the JACKS package[112] produced a different list of potential hits, with results obtained for genes rather than single gRNAs. A Z-score was calculated as described for the '98$^{th}$ percentile' method, but the average read count in the lung samples at 4 hours post-dosing was subtracted rather than using the read count in the pre-dosing cell samples from day 0 (Table 6.3).

**Table 6.3. Screen analysis: JACKS method.** Top 10 hits from analysis using the JACKS method.

| Gene | Z-score |
|---|---|
| *Lrrn4cl* | 0.76 |
| *Slc22a30* | 0.21 |
| *Tm4sf19* | 0.60 |
| *Kcnd1* | 0.77 |
| *Olfr323* | 0.76 |
| *Tmem54* | 0.48 |
| *Kcna5* | 0.47 |
| *Rhbdd2* | 0.39 |
| *Rxfp2* | 0.49 |
| *Scn8a* | 0.43 |

#### 6.2.3.3 'Appearance vs enrichment' method

A third analysis developed by Felicity Allen (a colleague in the Parts' lab, WSI) produced another list of potential hits. This analysis independently considered two factors: 'appearance' and 'enrichment'. 'Appearance' describes how many lung samples at day 19 carried a given gRNA; the presence of NTC gRNAs was used as a control to determine whether a targeting gRNA appeared significantly more than expected by chance. 'Enrichment' describes the abundance of a gRNA in a lung sample at day 19 relative to the abundance in the cells prior to administration to the mice on day 0. An 'enrichment' score was calculated by considering the difference in abundance and the variability of the data for this gRNA across all samples. For 'enrichment', the p-values for each gRNA were combined across all samples that the gRNA was present in. If a gRNA was significantly abundant in the lungs, or was simply present in a significant number of lungs, this could suggest that activation of that gene was advantageous for metastatic colonisation. The ranking of gRNAs varied depending on the weight given to each factor. Some gRNAs had high 'enrichment' but low 'appearance', and vice versa (Fig. 6.5). Others had moderately high scores for both factors. As this was a novel analysis strategy, we were unsure which ranking method was appropriate so we considered all possibilities.

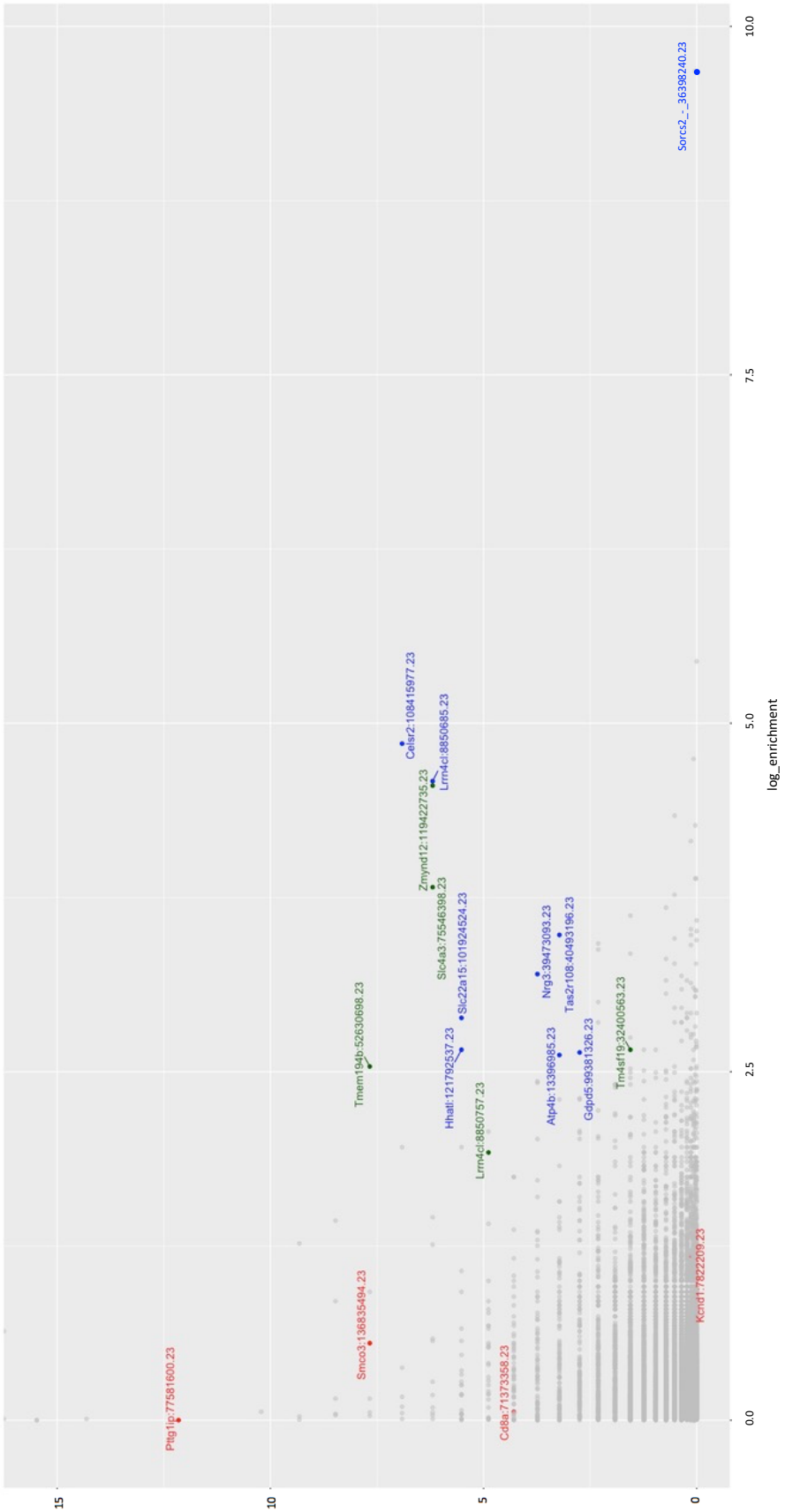



**Figure 6.5. Screen analysis: 'appearance vs enrichment' method.** Log-transformed 'enrichment' and 'appearance' scores are plotted for all gRNAs in the library. Several gRNAs of interest are highlighted and annotated.

### 6.2.4 Potential regulators of metastasis

We selected candidate genes from each of the three analyses for further validation (Table 6.4). From the '98$^{th}$ percentile' analysis, we chose to validate the top four scoring gRNAs which targeted *Lrrn4cl, Slc4a3, Tmem194b* and *Zmynd12. Lrrn4cl* was also the top hit in the JACKS analysis, and from this we additionally selected the third and fourth ranked genes, *Tm4sf19* and *Kcnd1*. The second ranked gene was *Slc22a30,* which appears to be a mouse-specific gene that arose from a recent duplication event[307]; we decided not to pursue this due to lack of human-relevance. From the third analysis we chose to validate gRNAs targeting *Pttg1ip* (high appearance), *Sorcs2* (high enrichment), *Celsr2* and *Lrrn4cl* (highest combined appearance and enrichment). The *Lrrn4cl* gRNA was different to the one identified by the other analyses. The *Slc4a3, Tmem194b* and *Zmynd12* gRNAs selected from the first analysis also had moderately high scores for both appearance and enrichment.

**Table 6.4. Candidates selected for validation.** The gRNAs chosen for validation are listed and the analysis method used to select each of these is indicated.

| Gene | gRNA ID | Analysis method |
|---|---|---|
| *Lrrn4cl* | Lrrn4cl_+_8850757.23-P1P2 | 98$^{th}$ percentile |
| *Slc4a3* | Slc4a3_+_75546398.23-P1P2 | 98$^{th}$ percentile |
| *Tmem194b* | Tmem194b_+_52630698.23-P1P2 | 98$^{th}$ percentile |
| *Zmynd12* | Zmynd12_+_119422735.23-P1P2 | 98$^{th}$ percentile |
| *Tm4sf19* | Tm4sf19_-_32400563.23-P1P2 | JACKS |
| *Kcnd1* | Kcnd1_-_7822209.23-P1P2 | JACKS |
| *Pttg1ip* | Pttg1ip_+_77581600.23-P1P2 | Appearance vs enrichment |
| *Sorcs2* | Sorcs2_-_36398240.23-P1P2 | Appearance vs enrichment |
| *Celsr2* | Celsr2_-_108415977.23-P1P2 | Appearance vs enrichment |
| *Lrrn4cl* | Lrrn4cl_+_8850685.23-P1P2 | Appearance vs enrichment |

#### 6.2.4.1 Biology of candidate genes

All of the genes selected for validation have a human homologue but the functions of many of the proteins are not well understood. Leucine Rich Repeat Neuronal 4 C-terminal-like (LRRN4CL) is a single-pass type I membrane protein. *TMEM194B*, also known as *NEMP2*, encodes for nuclear envelope integral membrane protein 2. *ZMYND12* encodes for Zinc finger MYND domain-containing protein 12. *TM4SF19* encodes for Transmembrane 4 L Six Family Member 19. Currently there is no literature regarding the biological function of these four

proteins, but all had consistently low expression in RNA-seq analysis of 27 normal human tissues.[308] Solute Carrier Family 4 Anion Exchanger Member 3 (SLC4A3) is a plasma membrane protein involved in exchange of chloride and bicarbonate.[309] It is most abundantly expressed in the brain and heart, and has been associated with idiopathic generalised epilepsy[310] and regulation of cardiac function.[311] Potassium Voltage-Gated Channel Subfamily D Member 1 (KCND1) is a multipass membrane protein. It is a pore-forming subunit of voltage-gated A-type potassium channels, which create currents in neurons and cardiac myocytes.[312] PTTG1IP is a single-pass type I membrane protein that binds to pituitary tumour-transforming 1 protein (PTTG1) and promotes its translocation to the nucleus.[313] It has been identified as an oncogene in endocrine cancers such as thyroid cancer,[298] and overexpression has also been shown in cancers such as breast[314] and colorectal.[300] It is has been to shown to decrease p53 stability by ubiquitination.[315,316] One study demonstrated that PTTG1IP interacts with CTTN to induce cellular invasion and migration in thyroid and breast cancer cells[317]. *SORCS2* encodes for vacuolar protein sorting 10 domain-containing receptor protein. It has roles in neuronal function and viability,[318,319] and has been linked to diseases such as bipolar disorder[320] and schizophrenia.[321] Cadherin EGF LAG seven-pass G-type receptor 2 (CELSR2) has been shown to have roles in ciliogenesis and neuronal development, including neuron migration and axon guidance.[322] Mutations in *CELSR2* have been associated with poor prognosis in endometrial cancer but there is no functional evidence for this.[323] *PTTG1IP* is the only candidate that has a clear functional association with metastasis, but several of the genes are not well characterised and may have interesting functions that have not been identified yet.

#### 6.2.4.2 Validation of candidate genes

LvdW cloned a gRNA targeting each candidate gene into the library backbone and packaged into lentiviral vectors (as described in Section 7.17.7.1). Three random NTC gRNAs from the library were also cloned, pooled and packaged into a single lentiviral vector for use as a negative control. All lentiviruses were titred in B16-F0-dCas9 cells (as described in Section 7.17.3). Each gRNA was transduced independently and cells were processed as in the screen: BFP was measured and puromycin was added after 48 hours, BFP was measured again on day 6 to confirm complete selection and on day 10 cells were collected for intravenous administration (as described in Section 7.17.7.2). Each mouse was dosed with $2x10^5$ cells, with 9-10 mice/condition. Lungs were collected after 10 days and melanocytic metastases were

counted by eye (as described in Section 7.17.7.3). Validation experiments were performed in batches with different mice cohorts, but the NTC was repeated in every batch.

Mice dosed with cells carrying the *Lrrn4cl, Slc4a3, Tmem194b* and *Zmynd12* gRNAs developed significantly more pulmonary metastases compared to the NTC (Fig. 6.6a). Similarly, cells expressing the *Tm4sf19* gRNA formed significantly higher numbers of metastases compared; the *Kcnd1* gRNA had a significant effect in one cohort ($p < 0.0001$), but not in another ($p = 0.5401$) (Fig. 6.6b&c). Mice dosed with cells carrying the second gRNA targeting *Lrrn4cl* or the *Sorcs2* gRNA had statistically significantly more metastases than the NTC, but a gRNA targeting *Celsr2* had no significant effect (Fig. 6.6d). Finally, there was no significant difference in the number of metastases formed by cells carrying a *Pttg1ip* gRNA compared to the NTC (the first *Lrrn4cl* gRNA was repeated in parallel as a positive control, Fig. 6.6e).

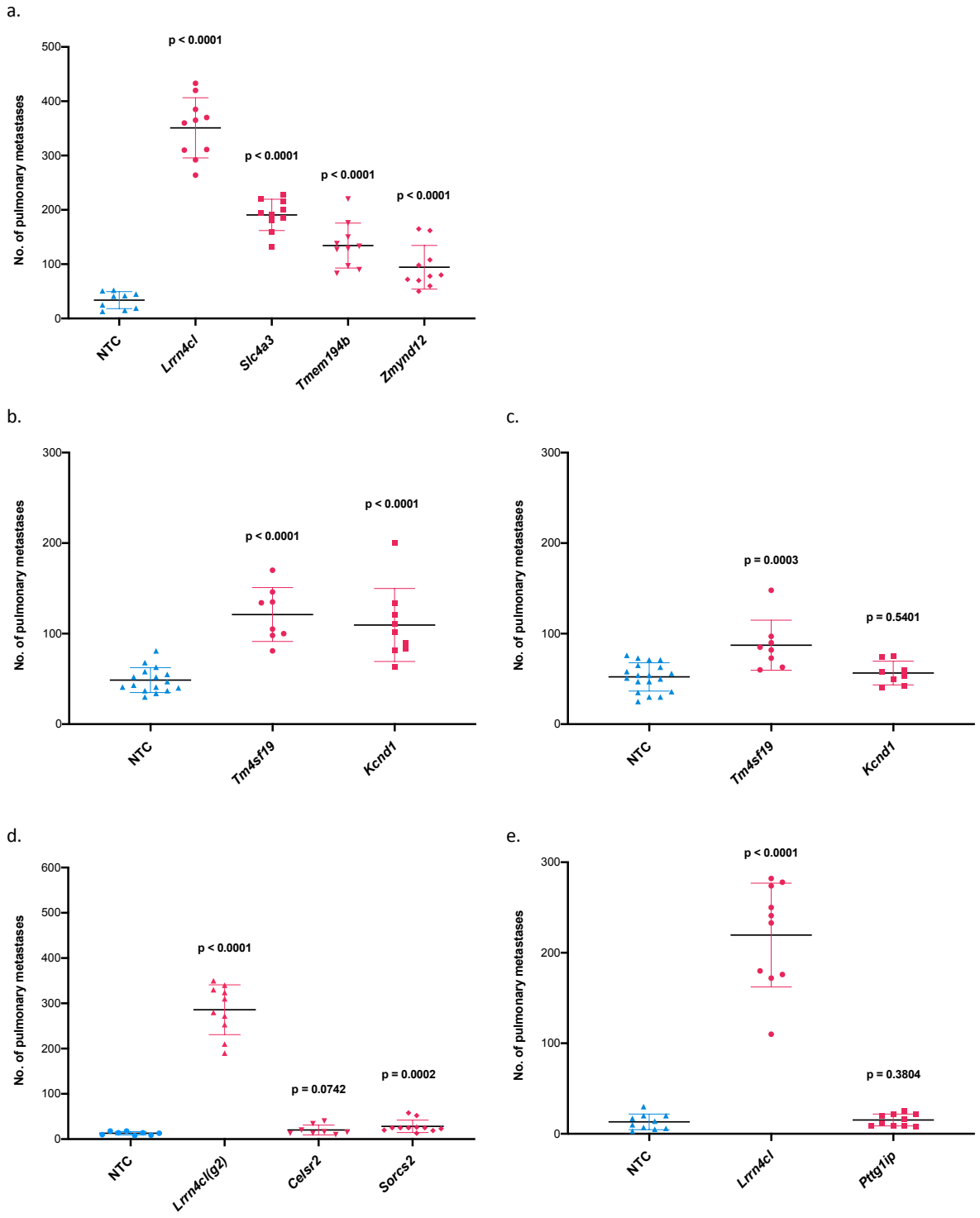



**Figure 6.6. *In vivo* validation of top scoring gRNAs from CRISPRa screen.** B16-F0-dCas9 cells were transduced with lentiviruses containing single targeting gRNAs, or a pool of NTC gRNAs. Transduced cells were intravenously administered to at least 8 mice per condition. On day 10 post-dosing, lungs were collected from the mice and metastases were counted by eye. The number of pulmonary metastases present in mice dosed with each gRNA are plotted; the mean is indicated by a line and error bars show the standard deviation. Each graph represents a single cohort. A Mann-Whitney test was performed to compare each gRNA with the NTC (p-values shown). *Lrrn4cl* and *Tmem194b* validated in 2 additional cohorts. *Slc4a3* validated in one independent cohort. *Zymnd12* was repeated in a further 4 cohorts, and validated in 3 of these. Data is not shown for these extra cohorts.

## 6.3 Discussion and future directions

In this project, we established a protocol for carrying out a pooled CRISPRa screen in B16-F0 melanoma cells. We extended this to an *in vivo* setting using an experimental metastasis model, allowing us to screen for genes that enhance metastatic pulmonary colonisation. The transfer from *in vitro* to *in vivo* was successful and using various analyses, we could identify gRNAs that were enriched in the lungs of mice. Despite little overlap in the results of the analysis strategies, several hits from all of them did validate in the experimental metastasis assay. As more data is produced, analysis of CRISPRa screens will undoubtedly become more accurate and refined but the approaches applied here were sufficient to identify true hits. Improvements to CRISPRa gRNA design should also increase the reliability of these data; variability in efficacy currently means that hits are often not supported by data from multiple gRNAs. Results from the 'appearance vs enrichment' method were the least robust, but it was unclear how best to interpret these data prior to validation. The strongest candidate gene was *Lrrn4cl*, which was a hit in all analyses; cells carrying gRNAs targeting this gene consistently formed more pulmonary metastases than any of the others.

Our data suggest that overexpression of *Lrrn4cl* enhances the ability of B16-F0 melanoma cells to colonise the lung. LvdW repeated the experimental metastasis assay using a *Lrrn4cl* cDNA (and empty vector as a control) as an alternative over-expression strategy and the result was replicated. This assay was also performed in immunodeficient mice lacking an adaptive immune system (B cells, T cells, NK cells) and a significant phenotype was still observed. LvdW repeated the cDNA assay in three additional melanoma cell lines, plus a breast, a colorectal and a bladder cancer cell line. Pulmonary metastases were increased in all lines, suggesting that the effect of *Lrrn4cl* activation is not specific to B16-F0 cells. The *Lrrn4cl* gene has a human ortholog with the same name (*LRRN4CL)* which encodes for a single-pass type I membrane protein. LvdW confirmed that this effect was not mouse-specific: enhanced pulmonary colonisation was observed in immunodeficient mice dosed with two human melanoma cell lines expressing a human *LRRN4CL* cDNA.

These findings suggest that LRRN4CL could be a potential target to reduce the development of pulmonary metastases in human cancer. RNA expression data from primary tumours indicates that *LRRN4CL* is highly enriched in melanoma (Fig. 6.7). Thus, it could have clinical relevance for the treatment of melanoma patients. Its location on the cell membrane also makes it an ideal drug target. To explore this, LvdW and Victoria Harle (a postdoc in the

Adams' lab) are investigating whether knocking out *LRRN4CL* in human melanoma cell lines can reduce pulmonary colonisation in mice.

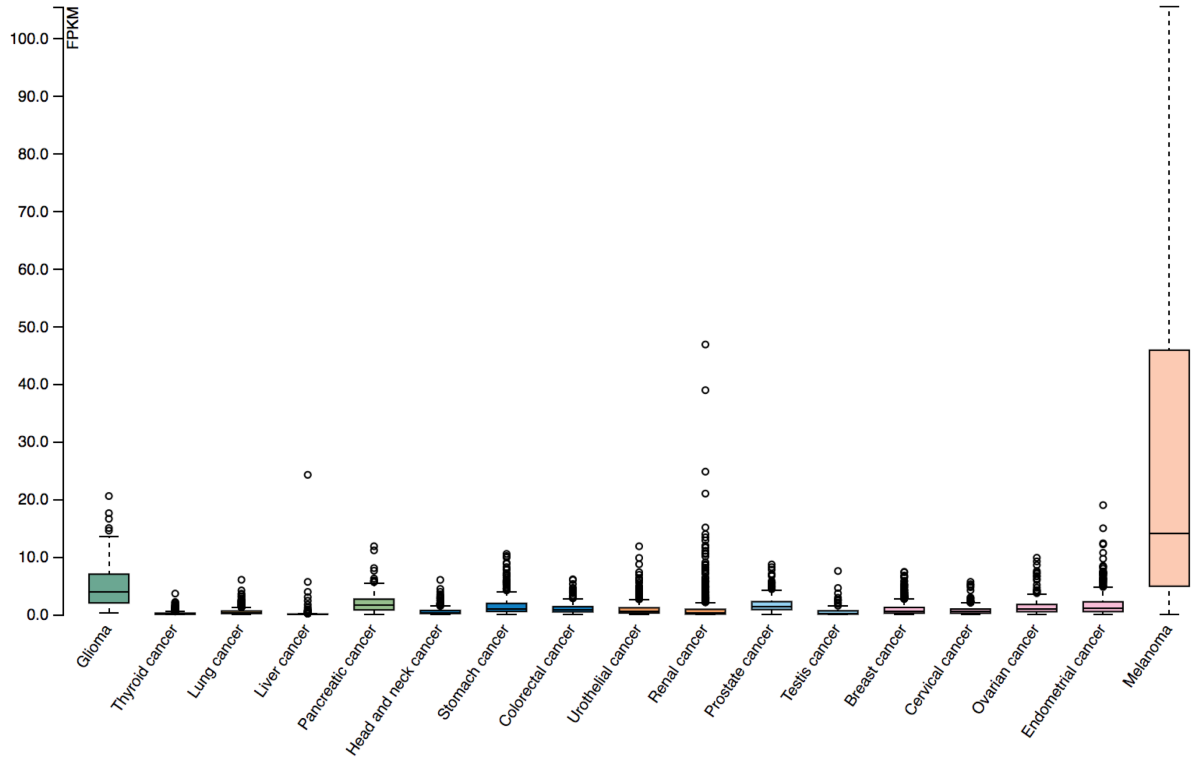



**Figure 6.7. Expression of *LRRN4CL* RNA in cancer.** RNA-seq data from 17 cancer types, reported as the median FPKM (number of Fragments Per Kilobase of exon per Million reads), was generated by TCGA. Graph taken from the Human Protein Atlas (available at proteinatlas.org/ENSG00000177363-LRRN4CL/pathology).

Limited information is available regarding this gene/protein in both mouse and human. In normal human tissues, the gene appears to be expressed at a low level (Fig. 6.8). As its normal cellular function is unclear, it is difficult to infer the mechanism by which activation of this gene could cause increased metastatic colonisation. Elucidation of any proteins that interact with LRRN4CL could reveal a wider network of genes with a similar role in metastasis. LvdW observed the *Lrrn4cl* phenotype in immunodeficient mice suggesting that, at least in mouse, increased colonisation is not due to an interaction with the adaptive immune system. Further investigation is required to understand the role of LRRN4CL in regulating metastatic colonisation and its potential as a therapeutic target.

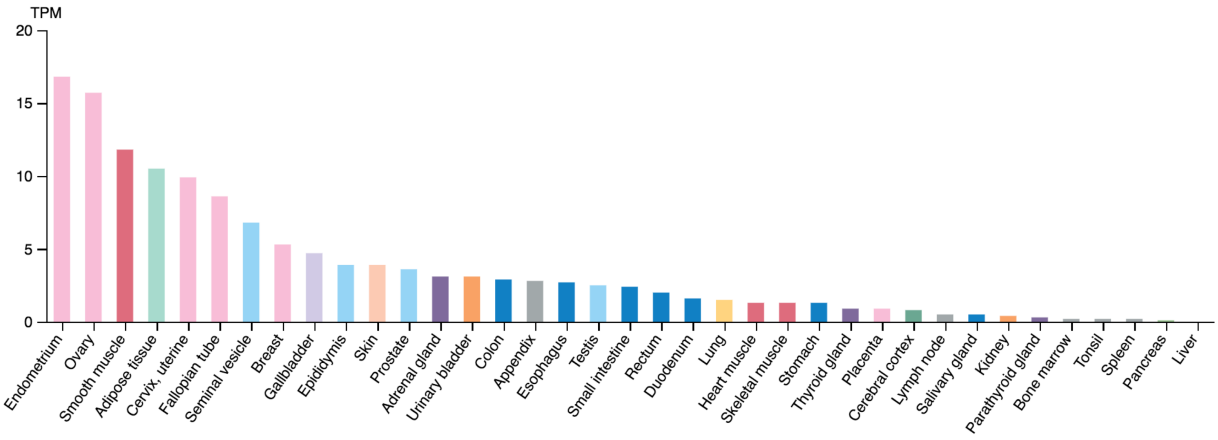



**Figure 6.8. Expression of *LRRN4CL* RNA in normal tissue.** HPA RNA-seq data from 172 normal tissue samples. Expression is reported as mean pTPM (protein-coding transcripts per million), corresponding to mean values of the individual samples for each tissue type. Graph taken from the Human Protein Atlas (available at https://www.proteinatlas.org/ENSG00000177363-LRRN4CL/tissue).

## 6.4 Final conclusion

CRISPR/Cas9 screening is proving to be an invaluable tool for functional genetic studies. Whilst its application in identifying fitness genes in cancer cell lines is fairly well-established, it has potential for many other uses that are still in their infancy. In this thesis, I have discussed two applications of CRISPR/Cas9 which have not been widely explored: knockout screening in iPSCs and activation screening in an *in vivo* model. Both of these proved to be technically challenging and produced lower quality data than more established strategies. We gained insights into the issues associated with screening in iPSCs and with further work, the data quality could be improved and these cells could be a useful model for studying genetic interactions. Analysis of *in vivo* screens and CRISPRa screens is not well-established but will undoubtedly become more refined as more data is produced. Despite this, our results were very promising and we successfully validated several hits that could further our understanding of metastatic colonisation. As the CRISPR toolkit improves and expands, so too will our ability to explore areas of cancer biology in a way that was not possible with other technologies.

# Chapter 7

# Materials and Methods

## 7.1 General molecular biology

### 7.1.1 Agarose gel electrophoresis

Agarose powder was dissolved in TAE (generally at 1% i.e. 1 g of agarose/100 ml buffer). Ethidium bromide was added (5 µl/100 ml) by heating and the gel was poured into a mould with a gel comb. When set, the gel was transferred to an electrophoresis tank and covered with TAE. Samples were mixed with loading dye and loaded into each well, with a molecular weight marker run in parallel (HyperLadder 1 kb (#BIO-33053, Bioline)). Samples were run at ~120V for 20-45 minutes and visualised using a Gel Doc XR+ System (Bio-Rad).

For instances when the DNA was to be extracted from the gel, ethidium bromide was not used. Post-electrophoresis staining was carried out using SYBR™ Green I Nucleic Acid Gel Stain (#S7563, Invitrogen) diluted in TAE; gels were incubated at room temperature for 30 minutes in darkness. An LED-based trans-illuminator (Nippon Genetics) was used to visualise the gel and the desired band was excised using a scalpel.

### 7.1.2 Digest/PCR product purification

#### 7.1.2.1 Column purification

For column purification, DNA/PCR products were purified using either the QIAquick PCR Purification Kit (#28104, QIAGEN) or the Monarch® PCR & DNA Cleanup Kit (5 µg) (#T1030S, NEB), following the manufacturer's protocol.

#### 7.1.2.2 Gel extraction

DNA/PCR products were extracted from agarose gel using the QIAquick Gel Extraction Kit (#28704, QIAGEN) or Monarch® DNA Gel Extraction Kit (#T1020S, NEB), following the manufacturer's protocol.

## 7.1.3 Plasmid amplification and purification

### 7.1.3.1 Chemical transformation

Plasmid DNA was transformed into OneShot™ TOP10 Chemically Competent *E. coli* (#C404003, Invitrogen, ThermoFisher) following the manufacturer's rapid chemical transformation procedure and plated on Luria broth (LB) agar plates containing 100 µg/ml ampicillin. After overnight incubation at 37°C, single colonies were inoculated into liquid LB with 100 µg/ml ampicillin and incubated overnight in a shaking incubator at 37°C, 225 rpm.

### 7.1.3.2 Glycerol stocks

After overnight incubation, 500 µl of bacterial culture was mixed with 500 µl of 50% glycerol, snap frozen in dry ice and stored at -80°C.

### 7.1.3.3 Purification

Plasmid DNA was extracted from bacterial cultures using QIAGEN Plasmid *Plus* DNA purification kits (Mini, Midi, Maxi or Mega) following the manufacturer's protocol.

## 7.1.4 Gibson Assembly

Reactions were set up on ice with a total volume of 20 µl (components specified in the relevant sections) using the Gibson Assembly® MasterMix (#E2611S, NEB). A control reaction was always set up in parallel with water replacing the insert, to give an indication of how successful the cloning had been. Reactions were incubated at 50°C for 1 hour then diluted 1:4 in nuclease-free water and 2 µl of each was chemically transformed into 50 µl of OneShot™ TOP10 Chemically Competent *E. coli* (Section 7.1.3.1). At least 4 colonies from the transformation plate were inoculated for minipreps and plasmid DNA was purified.

## 7.2 iPSC culture

BOB is a male human iPSC line re-programmed from fibroblasts using Sendaiviral vectors. The cells originally carried a single point mutation in the *A1AT* gene, which was corrected to wildtype using zinc finger nuclease technology[245]. A reciprocal translocation, t(6;8)( p21.1;8q24.1), is present but the karyotype is stable. This line was established at the WSI and it has been used and characterised extensively within the institute. All BOB cell lines (parental and KO lines) were kindly provided by the Gene Editing facility at WSI. These were grown in feeder-free conditions in TeSR™-E8™ medium (#05990, STEMCELL) at 37$^{o}$C, 5% $CO_2$ and medium was changed every day. Details provided here are relevant to all iPSC culture discussed in subsequent sections.

### 7.2.1 Coating culture dishes

Cells were maintained in culture dishes coated in either Synthemax®-II-SC (#CLS3535, Corning) or Vitronectin-XF™ (#07180, STEMCELL). Synthemax®-II-SC powder was resuspended in sterile cell culture water to 1 mg/ml. From this stock, it was further diluted to 25 µg/ml in sterile cell culture water, added to the dish (10 ml/10 cm$^2$ dish) and incubated at room temperature for 2 hours. Vitronectin-XF™ was diluted to 10 µg/ml in PBS or sterile water, added to the dish (6 ml/10 cm$^2$ dish) and incubated at room temperature for 1 hour. In both cases, coated dishes were used on the day or stored at 4$^{o}$C in parafilm for up to 3 days.

### 7.2.2 Thawing BOB lines

Cryovials were briefly warmed in a 37$^{o}$C water bath to thaw the cell suspension. Cells were transferred to a Falcon containing a 4-fold excess of TeSR-E8 medium. Centrifugation was carried out at 300 g for 3 minutes and the supernatant was removed. Cells were resuspended in 4 ml of medium plus 10 µM ROCK inhibitor (Y-27632, #72304, STEMCELL). Coating solution was aspirated from a 6 cm$^2$ dish and the cell suspension was transferred to the dish.

### 7.2.3 Passaging BOB lines as single cells

All values given are for a 6 cm$^2$ dish; these were scaled based on surface area as required. For general maintenance, cells were passaged every 6-7 days once dense colonies had formed. Medium was aspirated and cells were washed once with 4 ml of PBS. Cells were incubated

with 2 ml of accutase at 37ºC for 5-10 minutes until cells detached. An equal volume of TeSR-E8 medium was added, cells were pipetted 5-10 times to make a single cell suspension and transferred to a Falcon tube. A small aliquot of cells was mixed 1:1 with Trypan blue. A live cell count was calculated using a Countess™ II Automated Cell Counter (ThermoFisher). Cells were centrifuged at 300 g for 3 minutes and the supernatant was removed. Cells were resuspended in 2 ml of fresh medium. Coating solution was aspirated from the culture dish and 10 ml of medium was added with 10 µM ROCK inhibitor. The desired number of cells were transferred to the culture dish and incubated at 37ºC, 5% $CO_2$.

### 7.2.4 Freezing BOB lines

Cells were processed as described in Section 7.2.3. After resuspension, the desired number of cells (generally $2x10^5$ cells/vial) were transferred to a Falcon tube containing KnockOut™ Serum Replacement (KOSR, #10828028, Gibco, ThermoFisher) to a total volume of 500 µl/vial. Another Falcon tube was prepared with 500 µl/vial of KOSR + 20% DMSO. 500 µl of both solutions was added to each vial and inverted several times to mix. Cryovials were stored in a polystyrene rack at -80ºC for 48 hours before transfer to liquid nitrogen.

## 7.3 Engineering knockout iPSC lines

All KO BOB lines were produced by the Gene Editing facility at WSI. An asymmetrical exon in the gene of interest was substituted with a puromycin cassette and a frameshift indel was introduced into the other allele. A template vector with an EF1a-puromycin cassette was cloned for each gene, with two 1.5 kb homology arms designed to match the sequence around the targeted exon. Two gRNAs targeting each exon were designed (Appendix A.1). The template vector (2 µg), both gRNA vectors (3 µg) and hSpCas9 (4 µg) were transfected into $2x10^6$ cells using the Human Stem Cell Nucleofector® Kit 2 (#VPH-5022, Lonza). Cells were plated in 10 $cm^2$ dishes and after 72 hours, cells were selected with 3 µg/ml puromycin. Single cells were expanded and genotyped for the presence of a frameshift indel by Sanger sequencing. I repeated genotyping for all lines detailed in this thesis (described in Section 7.4). KO lines were maintained in 1 µg/ml puromycin (#ant-pr-1, InvivoGen).

# 7.4 Genotyping of knockout lines

## 7.4.1 DNA extraction

For genotyping, gDNA was extracted from cell pellets using the Gentra Puregene Core Kit A (QIAGEN) following the manufacturer's protocol. DNA was quantified using a NanoDrop® ND-1000 UV-Vis Spectrophotometer or a Thermo Scientific™ Multiskan™ GO Microplate Spectrophotometer.

## 7.4.2 Genotyping PCR

For each KO gene (in iPSCs and HAP1 cells), primers were designed to cover the edited region (Appendix A.1). PCR was performed on gDNA extracted from all edited lines using Platinum™ *Taq* DNA Polymerase High Fidelity (#11304011, Invitrogen, ThermoFisher) (Table 7.1). PCR products were visualised on an agarose gel (Section 7.1.1) to confirm the correct band was present. These were then column purified (Section 7.1.2.1) and sent for Sanger sequencing (Eurofins) (Appendix A.1).

**Table 7.1. Genotyping PCR.** Top - reaction composition, bottom - thermocycler conditions.

| Reagent | Volume/amount |
|---|---|
| 10X buffer | 2.5 µl |
| $MgSO_4$ | 1 µl |
| 10 mM dNTP | 0.5 µl |
| 10 µM forward primer | 0.5 µl |
| 10 µM reverse primer | 0.5 µl |
| *Taq* polymerase enzyme | 0.1 µl |
| DNA | ~500 ng |
| Nuclease-free water | to 25 µl total |

| Cycle Number | Denature | Annealing | Extension |
|---|---|---|---|
| 1 | 94˚C, 30 seconds | | |
| 2-31 (30 cycles total) | 94˚C, 10 seconds | Varied with primer $T_m$, 30 seconds | 68˚C, 1 min/kb |

## 7.5 Protein-level knockout confirmation

### 7.5.1 Protein lysis

Cells were washed twice with ice-cold PBS then lysis buffer was added (RIPA buffer (#20-188, Merck) diluted in distilled water, plus 1:100 protease/phosphatase inhibitor (#5872, Cell Signaling Technology)); the volume was dependent on cell number. Cells were incubated on ice for 5-10 minutes then scraped and transferred to Eppendorfs. Samples were centrifuged at 14,000 g for 15 minutes at 4$^{o}$C. Supernatant was transferred to fresh Eppendorfs and stored at -80$^{o}$C.

### 7.5.2 Protein quantification

Protein lysates were quantified using the Pierce™ BCA Protein Assay Kit (#23227, Thermo Scientific), following the manufacturer's microplate protocol. Absorbance was measured using a Thermo Scientific™ Multiskan™ GO Microplate Spectrophotometer.

### 7.5.3 Protein separation

SDS-PAGE was performed to separate proteins. NuPAGE™ LDS Sample Buffer (#NP0007, ThermoFisher Scientific) and NuPAGE™ Sample Reducing Agent (#NP0009, ThermoFisher Scientific) were mixed with protein lysate. Samples were heated to 70$^{o}$C for 10 minutes then placed on ice prior to loading into a pre-cast gel (NuPAGE™ Novex 4-12% Bis-Tris Gel, 1.5 mm x 15 well, #NP0336BOX, ThermoFisher). A protein molecular weight marker was also loaded. The gel was set up in an XCell SureLock™ Mini-Cell and covered with 1X NuPAGE™ MOPS SDS Running Buffer (#NP0001, ThermoFisher) supplemented with NuPAGE™ Antioxidant (500 µl/200 ml, #NP0005, ThermoFisher). Samples were run at 150 V until sufficient separation was achieved.

### 7.5.4 Western blotting

NuPAGE™ Transfer Buffer (#NP0006, ThermoFisher) was diluted in deionised water and supplemented with NuPAGE™ Antioxidant (1%) and methanol (10%). PVDF membrane was briefly soaked in methanol and placed on top of the gel, assembled with filter paper and blotting pads in an XCell II™ Blot Module. The module was filled with the prepared transfer buffer

and run at 30 V for 1 h. The membrane was then incubated in blocking buffer (5% non-fat milk) dissolved in Tris-Buffered Saline supplemented with 0.1% Tween-20 (TBS-T). Incubation was carried out for 20 minutes - 2 hours at room temperature on a shaker. The membranes were then incubated with primary antibody (diluted in blocking buffer, Table 7.2) overnight at 4$^{o}$C on a shaker. Five washes were carried out with TBS-T for 5 minutes at room temperature. The membrane was then incubated with secondary antibody (diluted in blocking buffer, Table 7.2) at room temperature for 1 hour on a shaker. Finally, the membrane was washed again 5 times with TBS-T for 5 minutes. To visualise, the membrane was sprayed with ECL reagent (#K-12049-D50, Advansta) and incubated for 1 minute before reading on an ImageQuant™ LAS 4000 using chemiluminescence.

**Table 7.2. Antibodies for Western blotting.**

| Antibody | Supplier | Product # | Dilution (in 5% milk) | Secondary antibody |
|---|---|---|---|---|
| **Primary antibodies** | | | | |
| Arid1b | Abcam | ab57461 | 1:1000 | Goat anti-mouse |
| Arid2 | SantaCruz | sc-166117 | 1:1000 | Goat anti-mouse |
| Arid1a (PSG3) | SantaCruz | sc-32761 | 1:1000 | Goat anti-mouse |
| β-Actin (8H10D10) | Cell Signalling Technology | #3700 | 1:1000 | Goat anti-mouse |
| **Secondary antibody** | | | | |
| Goat anti-mouse IgG H&L (HRP) | Abcam | ab97023 | 1:10,000 | - |

## 7.6 Proteomics

Liquid chromatography-mass spectrometry was performed on the parental BOB line and KO derivatives, by Clara Alsinet (previous postdoctoral fellow in the Adams' lab) in collaboration with Jyoti Choudhary and Theodoros Roumeliotis (previously in the Proteomics facility at WSI). Sample preparation and processing were carried out as described in.[324] Data for the parental BOB line and *ARID1A/ARID2*/*PBRM1* KO clones was published in that manuscript as part of a separate project.

# 7.7 Cloning of a neomycin-resistant gRNA library

## 7.7.1 Insertion of neomycin resistance gene

Empty Yusa v1.1 library backbone DNA (Addgene #67974) was kindly provided by Fiona Behan (postdoctoral fellow in the Garnett lab, WSI). The backbone was digested with Kpn2I at 37°C for 2 hours (Table 7.3), visualised on an agarose gel to confirm digestion (Section 7.1.1) and then column purified (Section 7.1.2.1). A 901 bp DNA fragment (fragment A, Appendix A.1) (GeneArt Strings, ThermoFisher) was designed, containing a neomycin resistance gene coding region and sequence complementary to the backbone. An additional 120 bp DNA oligonucleotide (fragment B, Appendix A.1) (Ultramer®, Integrated DNA Technologies) was also designed to extend the overlap upstream of the initial Kpn2I site in the PGK promoter. Both fragments were cloned into the digested backbone by Gibson Assembly (Table 7.3) (Section 7.1.4). DNA from four colonies was sent for Sanger sequencing (Eurofins) with six primers which together provided sequence covering the full edited region (Appendix A.1). This backbone is referred to hereafter as 'neoR'.

**Table 7.3. Cloning to insert neomycin resistance gene.** Top - restriction digest, bottom - Gibson Assembly.

| Reagent | Volume/amount |
|---|---|
| NEB 3.1 buffer | 15 µl |
| Kpn2I (BspEI) (#R0540S, NEB) | 10 µl |
| Yusa v1.1 backbone | 10 µg |
| Nuclease-free water | to 150 µl total |

| Reagent | Volume/amount |
|---|---|
| Kpn2I-digested backbone | 195.8 ng |
| Fragment A | 111.4 ng |
| Fragment B | 14.84 ng |
| 2X Gibson Assembly MM | 10 µl |
| Nuclease-free water | to 20 µl total |

### 7.7.2 Testing the neoR library backbone

The neoR library backbone was packaged into a lentivirus and titred (Section 7.8). BOB-Cas9 cells were detached and counted. Cells were transduced with the lentivirus at an MOI of 0.3, and 200,000 cells/well were seeded in a 6-well plate with 2 ml total volume. Untransduced cells were also seeded in parallel. Fresh medium was added the following morning. After 48 hours, cells were passaged and a sample was fixed and analysed by flow cytometry to measure BFP expression (Section 7.9). For a clonogenic assay, 10,000 transduced cells were seeded per well in 5 wells of a 6-well plate. One well was seeded with untransduced cells. On the day after seeding, 0, 0.1, 0.3, 0.6 or 1 mg/ml G418 was added to the transduced cells. The same experiment was repeated with 10,000 cells/well in a 12-well plate to allow for further flow cytometry analysis. After 5 days in G418, media was aspirated from the 6-well plate and cells were washed twice with ice-cold PBS. Cells were incubated in 4% paraformaldehyde (PFA) for 20 minutes at room temperature and then washed twice with PBS. Crystal violet (0.5% in methanol) was then added to each well, incubated briefly and then wells were washed with water. Images were taken after plates had dried. After 5 days in G418, cells in the 12-well plate were fixed and BFP expression was measured by flow cytometry (Section 7.9).

### 7.7.3 Insertion of AjuI restriction sites

The neoR backbone was digested with BbsI at 37ºC for 16 hours (Table 7.4), visualised on an agarose gel to confirm digestion (Section 7.1.1) and the linearised band was gel extracted (Section 7.1.2.2). A 94 base DNA oligonucleotide (fragment C, Appendix A.1) (Ultramer®, Integrated DNA Technologies) was designed to contain the same sequence as the backbone but replacing the BbsI recognition site with AjuI restriction sites. The fragment was inserted into the digested backbone by Gibson Assembly (Table 7.4) (Section 7.1.4). DNA from six colonies was sent for Sanger sequencing (Eurofins) with primers covering the edited region (Appendix A.1).

**Table 7.4. Cloning to insert AjuI sites.** Top - restriction digest, bottom - Gibson Assembly.

| Reagent | Volume/amount |
|---|---|
| NEB 2.1 buffer | 10 µl |
| BbsI (#R0539S, NEB) | 4 µl |
| neoT2A backbone | 10 µg |
| Nuclease-free water | to 100 µl total |

| Reagent | Volume/amount |
|---|---|
| BbsI-digested backbone | 435 ng |
| Fragment C | 11.62 ng |
| 2X Gibson Assembly MM | 10 µl |
| Nuclease-free water | to 20 µl total |

### 7.7.4 Insertion of IRES element

PCR was performed to amplify a region containing part of the neomycin resistance gene and an IRES sequence (Table 7.5, primers detailed in Appendix A.1) from an unpublished plasmid kindly provided by Luca Crepaldi (Parts' lab, WSI). 5 µl of the reaction was visualised on an agarose gel (Section 7.1.1) to confirm the correct product size was amplified and then the remainder was column purified (Section 7.1.2.1).

**Table 7.5. Amplification of neomycin resistance gene and IRES sequence.** Top - reaction composition, bottom - thermocycler conditions.

| Reagent | Volume/amount |
|---|---|
| KAPA HF HS 2X MM (#KK2601, Kapa Biosystems) | 12.5 µl |
| 5 µM forward primer | 1.5 µl |
| 5 µM reverse primer | 1.5 µl |
| Plasmid DNA | 1 ng |
| Nuclease-free water | to 25 µl total |

| Cycle Number | Denature | Annealing | Extension |
|---|---|---|---|
| 1 | 95˚C, 3 minutes | | |
| 2-31 (30 cycles total) | 98˚C, 20 seconds | 60˚C, 15 seconds | 72˚C, 15 seconds |
| 32 | | | 94˚C, 1 minute |

An edited clone containing the AjuI sites (Section 7.7.3) was then digested with RsrII and BsrGI at 37°C for 16 hours (Table 7.6). The backbone was visualised on an agarose gel (Section 7.1.1) to confirm digestion and then column purified (Section 7.1.2.1). A 123 bp DNA oligonucleotide (fragment E, Appendix A.1) (Ultramer®, Integrated DNA Technologies) was designed with an overlap between the IRES sequence and the BFP coding region in the backbone. Gibson Assembly was performed to insert the neo-IRES PCR amplicon and the DNA oligonucleotide into the digested backbone (Table 7.6) (Section 7.1.4). DNA from twelve colonies was sent for Sanger sequencing (Eurofins) with primers covering the edited region (Appendix A.1). One successfully edited clone was taken forward, referred to hereafter as 'neoR-IRES'.

**Table 7.6. Cloning to insert IRES element.** Top - restriction digest, bottom - Gibson Assembly.

| Reagent | Volume/amount |
|---|---|
| 10X NEB CutSmart buffer | 10 µl |
| RsrII (#R0501S, NEB) | 2 µl |
| BsrGI-HF (#R3575S, NEB) | 2 µl |
| neoR(AjuI) backbone | 5 µg |
| Nuclease-free water | to 100 µl total |

| Reagent | Volume/amount |
|---|---|
| RsrII/BsrGI-digested backbone | 106.2 ng |
| Fragment D | 29.5 ng |
| Fragment E | 7.6 ng |
| 2X Gibson Assembly MM | 10 µl |
| Nuclease-free water | to 20 µl total |

### 7.7.5 AjuI digestion of backbone

Various protocols were trialled to optimise digestion of the neoR-IRES backbone with AjuI. Both AjuI (#ER1951, ThermoFisher) and FastDigest AjuI (#FD1954, ThermoFisher) were tested with varying amounts of DNA, DNA:enzyme unit ratios, total reaction volumes and incubation times (from 2-16 hours). An example reaction setup is shown in Table 7.7 (reactions were always incubated at 37°C). Digests were visualised on an agarose gel (Section 7.1.1) then the digested product was gel extracted and purified (Section 7.1.2.2). 10 ng of digested and

undigested plasmid were transformed into OneShotTOP10 cells (Section 7.1.3.1). Colonies were counted after overnight incubation and the % of background was calculated as the number of colonies present with digested plasmid relative to undigested plasmid.

**Table 7.7. Restriction digest of neoIRES backbone with AjuI *(or FastDigest AjuI).***

| Reagent | Volume/amount |
|---|---|
| 10X Buffer R *(10X FastDigest Buffer)* | 5 µl |
| 50X SAM *(20X SAM)* | to 1X |
| AjuI *(FastDigest AjuI)* | 1 µl |
| neoR-IRES backbone | 1 µg |
| Nuclease-free water | to 20 µl total |

To confirm that AjuI was cutting at both restriction sites, PCR was carried out to amplify a 194 bp covering the region (Table 7.8) (Appendix A.1). The product was column purified (Section 7.1.2.1) and digested with FastDigest AjuI (Table 7.7). Digested and undigested PCR products were visualised an agarose gel (Section 7.1.1).

**Table 7.8. Amplification of AjuI sites.** Top - reaction composition, bottom - thermocycler conditions.

| Reagent | Volume/amount |
|---|---|
| Q5 HS HF 2X MM | 12.5 µl |
| 10 µM forward primer | 1.25 µl |
| 10 µM reverse primer | 1.25 µl |
| neoIRES backbone | 1 µg |
| Nuclease-free water | to 25 µl total |

| Cycle Number | Denature | Annealing | Extension |
|---|---|---|---|
| 1 | 98˚C, 30 seconds | | |
| 2-31 (30 cycles total) | 98˚C, 10 seconds | 67˚C, 30 seconds | 72˚C, 30 seconds |
| 17 | | | 72˚C, 2 minutes |

## 7.7.6 Testing the neoR-IRES library backbone

The neoR-IRES library backbone was tested as described in Section 7.7.2. The neoR library backbone was also tested again in parallel for comparison, using the 6-well clonogenic assay.

### 7.7.7 MluI/AgeI digestion of backbone

The neoR-IRES library backbone was digested with MluI and AgeI to prepare for transfer of gRNAs (Table 7.9). Digestion was carried out at 37ºC for 8 hours, then the reaction was visualised on an agarose gel (Section 7.1.1) and the digested product was gel extracted and purified (Section 7.1.2.2). To confirm efficient digestion, 10 ng of digested and undigested plasmid were transformed into OneShotTOP10 cells (Section 7.1.3.1). Colonies were counted after overnight incubation and the % of background was calculated as the number of colonies present with digested plasmid relative to undigested plasmid.

**Table 7.9. Restriction digest of neoIRES backbone with MluI/AgeI.**

| Reagent | Volume/amount |
|---|---|
| 10X NEB CutSmart Buffer | 30 µl |
| AgeI-HF (#R3552S, NEB) | 5 µl |
| MluI-HF (#R3198S, NEB) | 7.5 µl |
| neoIRES backbone | 10 µg |
| Nuclease-free water | to 300 µl total |

### 7.7.8 PCR of existing genome-wide gRNA library

DNA for the Yusa v1.1 library was kindly provided by Fiona Behan (Garnett lab, WSI). To amplify gRNAs from the library, PCR was performed using primers that covered the gRNA and extended across the MluI/AgeI restriction sites (Table 7.10, primers detailed in Appendix A.1). An initial optimisation experiment was carried out to identify the best annealing temperature and primer set. Two primer sets were tested as described in Table 7.10, with six annealing temperatures between 66-71ºC. 5 µl of each PCR product was visualised on an agarose gel (Section 7.1.1) and the condition with the strongest band was chosen (67ºC annealing temperature with primer set 2). Using these optimised conditions, 11 PCR reactions were performed with 100 ng of library DNA input per reaction. All reactions were pooled and visualised on an agarose gel (Section 7.1.1). The 599 bp band was then gel extracted and purified (Section 7.1.2.2).

**Table 7.10. Amplification of gRNAs from Yusa v1.1 library.** Top - reaction composition, bottom - thermocycler conditions.

| Reagent | Volume/amount |
|---|---|
| Q5 HS HF 2X MM (#M0494S, NEB) | 12.5 µl |
| 10 µM forward primer | 1.25 µl |
| 10 µM reverse primer | 1.25 µl |
| Yusa v1.1 library DNA | 100 ng |
| Nuclease-free water | to 25 µl total |

| Cycle Number | Denature | Annealing | Extension |
|---|---|---|---|
| 1 | 98˚C, 30 seconds | | |
| 2-16 (15 cycles total) | 98˚C, 10 seconds | 67˚C, 30 seconds | 72˚C, 30 seconds |
| 17 | | | 72˚C, 2 minutes |

### 7.7.9 Gibson Assembly to clone gRNA library

Twelve Gibson Assembly reactions were performed to insert the gRNA fragment into the MluI/AgeI digested backbone (Table 7.11, Section 7.1.4). All reactions were pooled and column purified using the Monarch kit (Section 7.1.2.1), with a 12 µl final elution volume.

**Table 7.11. Gibson Assembly reaction to clone gRNA library.**

| Reagent | Volume/amount |
|---|---|
| Digested backbone | 439.2 ng |
| gRNA PCR product | 88.84 ng |
| 2X Gibson Assembly MM | 10 µl |
| Nuclease-free water | to 20 µl total |

### 7.7.10 Bacterial amplification of gRNA library

Endura ElectroCompetent cells (#60242-1, Lucigen) were thawed on ice and 25 µl of cells was added to 2 µl of purified Gibson Assembly product in a chilled Eppendorf. This was transferred to a chilled 1.0 mm electroporation cuvette, ensuring even distribution and no bubbles. Electroporation was performed with the following parameters: 10 µF, 600 Ohms, 1800 Volts. Immediately after the pulse, 975 µl of recovery medium was added and cells were transferred

to a 15 ml BD Falcon polystyrene tube with snap-cap. Cells were incubated at 37ºC, 225 rpm for 1 hour. This was repeated 9 times and all cultures were pooled after incubation.

For colony counting to estimate transformation efficiency, dilutions of $10^{-3}$, $10^{-4}$, $10^{-5}$ and $10^{-6}$ were prepared and 50/100/150 µl of each were plated on LB agar plates containing 100 µg/ml ampicillin. These were incubated for 16 hours at 37ºC. The rest of the culture was divided between 6 flasks, each containing 500 ml of liquid LB plus 100 µg/ml ampicillin. These were incubated for 16 hours at 37ºC, 225 rpm.

Colonies were counted on each dilution plate and an average value was calculated and extrapolated to the total culture volume, giving an estimate of the library coverage achieved. Bacterial stabs for 94 colonies were made in a 96-well plate containing LB agar with 100 µg/ml ampicillin. The plate was incubated for 16 hours at 37ºC and sent for plasmid purification and Sanger sequencing (Eurofins) with primers covering the gRNA region (Appendix A.1).
The liquid cultures were pooled and split into 12x 250 ml centrifuge bottles. Centrifugation was performed at 6000 g for 15 minutes at 4ºC. The pellets were resuspended, pooled and transferred to 15 ml Falcons then centrifugation was repeated. Supernatant was removed and plasmid purification was performed using Megaprep and Maxiprep kits (Section 7.1.3.3). All DNA was pooled and quantified using a NanoDrop® ND-1000 UV-Vis Spectrophotometer.

Details of PCR amplification, purification and sequencing of the gRNAs in the resulting neoR-IRES library are provided in Section 7.12.

# 7.8 Lentivirus production and titration

## 7.8.1 Lentivirus production

All values given are for lentivirus production in a T150 flask (Table 7.12); these were scaled according to surface area for different culture dishes. HEK293T cells were cultured in IMDM (#12-722F, Lonza) supplemented with 10% Fetal Bovine Serum (FBS) (#A3160401, Gibco, ThermoFisher) and incubated at 37°C, 5% $CO_2$. Cells were seeded at a density of $18x10^6$ cells/T150 24 hours prior to transfection. Immediately before transfection, medium was aspirated and replaced with 12.5 ml Opti-MEM™ (#31985062, Gibco, ThermoFisher). Transfer vector and packaging plasmids pMD2.G (Addgene #12259) and psPAX2 (Addgene #12260) were diluted in Opti-MEM™. PLUS™ reagent (#15338100, ThermoFisher) was then added and the mixture was inverted 10-15x and incubated for 5 minutes at room temperature. Lipofectamine™ LTX (#15338100, ThermoFisher) was added, the solution was inverted 10-15x and incubated for 30 minutes at room temperature. The lipofectamine/DNA complex was then carefully added to the cells. After 16 hours, virus medium was aspirated and replaced with 24 ml complete IMDM. Approximately 30 hours later, virus supernatant was collected and filtered with a 0.45 µm low protein-binding filter (#SLHP033RS, Merck). Aliquots were stored at -80°C and were frozen for at least 2 hours before being used in experiments.

**Table 7.12. Composition of lentiviral transfection complex.**

| Reagent | Volume/amount |
|---|---|
| Opti-MEM™ | 7.5 ml |
| Transfer vector | 7.5 µg |
| psPAX2 | 18.5 µg |
| pMD2.G | 4 µg |
| PLUS™ reagent | 30 µl |
| Lipofectamine™ LTX | 90 µl |

## 7.8.2 Lentivirus titration

All lentiviruses described in this thesis were titrated by measuring fluorescent marker expression using flow cytometry (Section 7.9). Cells were detached and diluted to the concentration that would be required for the assay in which the lentiviruses would be used. For HAP1 cells, polybrene was added to a final concentration of 8 µg/ml; for iPSCs no polybrene

was used. Cell suspension was then added to a 6-well plate and varying volumes of virus were added to each well. As a control, one well was cultured without virus. After 48 hours, cells were detached and prepared for flow cytometry (Section 7.9). Untransduced cells were used to set a gate for fluorescent marker expression. The % of positive cells was used as an indicator of the MOI at each dose, and a line of best fit was plotted to allow calculation of the virus volume required for any given MOI. The amount of virus, cell number and total volume were then scaled based on the culture surface area used in further experimental transductions.

## 7.9 Analysis of cells by flow cytometry

### 7.9.1 Cell fixation

iPSCs or HAP1 cells were detached, transferred to an Eppendorf and centrifuged at 300 g for 3 minutes. Supernatant was removed and cells were washed with 500 µl PBS to remove residual medium. Centrifugation was repeated (300 g, 3 minutes), PBS was removed and cells were resuspended in 100 µl of 4% PFA (#F8775, Sigma). Fixation was carried out at room temperature for 15 - 20 minutes, then 500 µl PBS was added. Cells were centrifuged (300 g, 3 minutes) and the supernatant was discarded. Cells were resuspended in 500 µl PBS, centrifuged (300 g, 3 minutes) and supernatant was removed. Finally, cells were resuspended in PBS (volume dependent on pellet size, generally ~ 500 µl) and transferred to a Falcon® 5 ml Round Bottom Polystyrene Tube (#352054, Corning).

### 7.9.2 Flow cytometry analysis

Cells were analysed using a BD LSRFortessa™ (or occasionally BD LSRII) with BD FACSDiva™ software. Expression of BFP (405 (450/50)) and GFP (488 (530/30)) were detected on the Fortessa or LSRII, and mCherry (561 (610/20)) was detected only on the Fortessa. Results were then further analysed and plotted using FlowJo.

## 7.10 Engineering stable Cas9 lines

### 7.10.1 Cas9 transduction

A plasmid expressing Cas9 and blasticidin resistance (Addgene #68343) was packaged into a lentivirus (Section 7.8.1). Cells were detached and transduced in suspension. For iPSC lines, $7.5x10^5$ cells were transduced with 450 µl of lentivirus and seeded in a 6 $cm^2$ dish. For HAP1 cells, $5x10^5$ cells were transduced with 100 µl of lentivirus plus 8 µg/ml polybrene and seeded in a T25 flask. In both cases, control cells containing no lentivirus were seeded in parallel. The following morning, fresh complete medium was added. At 48 hours post-transduction, 10 µg/ml blasticidin was added (#antbl-1, InvivoGen). Complete death in the control flask confirmed successful blasticidin selection. All iPSC and HAP1 Cas9-expressing cell lines were continuously cultured in blasticidin, except during screens, to ensure high Cas9 activity was maintained. After ~2 weeks of selection, Cas9 activity was assessed (Section 7.10.2) to confirm that lines were functional before freezing stocks and using in further assays.

### 7.10.2 Quantification of Cas9 activity

Two vectors were used to assess Cas9 activity; one expressing BFP and GFP (control, Addgene #67979) and one expressing BFP, GFP and a gRNA targeting GFP (reporter, Addgene #67980). Both vectors were packaged into lentiviruses (Section 7.8.1). Cells were detached and transduced in suspension; one condition with no lentivirus, one with the control and one with the reporter. For iPSC lines, $2x10^5$ cells/well were seeded in a 12-well plate with 50 µl of lentivirus. For HAP1 lines, $2x10^5$ cells/well were seeded in a 6-well plate with 100 µl of lentivirus and 8 µg/ml polybrene. Fresh complete medium was added the following morning and after 48 hours cells were fixed and analysed by flow cytometry (Section 7.9). BFP and GFP expression were measured, using untransduced cells and cells transduced with the control vector to set gates for both markers.

# 7.11 NeoR-IRES library screen in iPSCs

## 7.11.1 Library transduction

The neoR-IRES gRNA library was packaged into a lentivirus and titred in all stable Cas9 iPSC lines (parental and KOs) (Section 7.8). For each screened line, cells were split into three separate populations at the passage prior to transduction. Each population was then detached and counted as single cells, and three separate transduction replicates were set up. For each replicate, a total of $54x10^6$ cells were transduced with the library lentivirus at an MOI of 0.3 in 594 ml total ($50x10^6$ cells were seeded but extra cells were transduced to allow for volume lost in pipetting). Cells were then seeded at $2x10^6$ cells/dish in a 15 cm dish (25 dishes total, for screens I performed) or $3.4x10^6$ cells in a T225 (17 flasks total, for screens CGaP performed). A small sample (3 ml of transduced cell suspension) was seeded in a 6 $cm^2$ dish for each replicate, and untransduced cells were also seeded in parallel at the same density. After 48 hours, 1 mg/ml G418 was added to all transduced cells. The small samples of transduced and untransduced cells were fixed and BFP expression was analysed by flow cytometry to measure transduction rate (Section 7.9). Untransduced cells were re-seeded and maintained in parallel with the screens.

## 7.11.2 Passage of cells during screen

For the screens that I performed, cells were passaged as clumps on day 8 post-transduction. Medium was aspirated and cells were washed with PBS. Gentle Dissociation Reagent (#07174, STEMCELL) was added (10 ml/15 cm dish) and cells were incubated at 37ºC for 10 minutes. An equal volume of TeSR-E8 was then added and the suspension was mixed no more than 5 times with a stripette. All cells were combined in a single bottle for each replicate and mixed gently by inversion.

To give a more accurate estimate of cell count, 3 ml of cell suspension was transferred to 1.5 ml Eppendorfs (1 ml/Eppendorf) and mixed further to dissociate clumps to single cells. Cells were centrifuged for 3 minutes at 300 g, media was aspirated and 500 µl of accutase was added. Cells were incubated in the Eppendorfs for 5-10 minutes at 37ºC. An equal volume of TeSR-E8 was added and 10 µl from each Eppendorf was mixed with Trypan blue and counted using a Countess™ II Automated Cell Counter. Based on the average cell count, cells were

diluted in TeSR-E8 and $50x10^6$ cells total were maintained per replicate. Cells were seeded at the same density of $2x10^6$ cells/152 $cm^2$, but in 5x 5-layer flasks (#11597421, Corning).

For screens that CGaP performed, the same protocol was followed but cells were passaged as clumps on day 6 post-transduction, maintaining $100x10^6$ cells total per replicate in 10x T225 flasks. Cells were then passaged as clumps again on day 8/9 post-transduction, maintaining $50x10^6$ cells total per replicate in 5x 5-layer flasks (#11597421, Corning).

At the point of passage, a small sample of cells from each replicate was fixed and BFP expression was analysed by flow cytometry (Section 7.9). Untransduced cells that had been maintained in parallel were also fixed and analysed as a control.

### 7.11.3 Harvesting of cells at screen endpoint

Cells were maintained in 1 mg/ml G418 until day 13 post-transduction. Cells were then detached as single cells using accutase, counted and $60x10^6$ cells were transferred to a 50 ml Falcon. A small sample of cells from each replicate was also fixed and BFP expression was analysed by flow cytometry (Section 7.9). Untransduced cells that had been maintained in parallel were also fixed and analysed as a control. Cells were centrifuged at 300 g for 3 minutes and the supernatant was aspirated. Pellets were resuspended in PBS, transferred to a 15 ml Falcon and centrifuged again. PBS was aspirated and the pellets were stored at -80$^o$C until DNA extraction.

# 7.12 Sequencing of neoR-IRES library gRNAs

## 7.12.1 DNA extraction

For each screen replicate, gDNA was extracted from pellets of $60x10^6$ cells using a Blood & Cell Culture DNA Maxi Kit, with each pellet split over two columns (#13362, QIAGEN).

## 7.12.2 First round PCR & purification

PCR was performed to amplify the gRNA region of the gDNA in each screen sample and the plasmid DNA from the neoR-IRES library (Table 7.13, primers detailed in Appendix A.1). A total of 72 PCR reactions with 2 µg gDNA/reaction were performed for every screen sample to ensure high representation of the library. For sequencing of the library from plasmid DNA, the same PCR was performed but with 10 reactions and 15 ng plasmid DNA/reaction.

**Table 7.13. First round gRNA PCR.** Top - reaction composition, bottom - thermocycler conditions.

| Reagent | Volume/amount per rxn |
|---|---|
| Genomic DNA | 2 µg |
| Q5 HS HF 2X MM | 25 µl |
| Primer mix (10 µM of both forward and reverse) | 1 µl |
| Nuclease-free water | to 50 µl total |

| Cycle Number | Denature | Annealing | Extension |
|---|---|---|---|
| 1 | 98˚C, 30 seconds | | |
| 2-29 (28 cycles in total) | 98˚C, 10 seconds | 61˚C, 15 seconds | 72˚C, 20 seconds |
| 30 | | | 72˚C, 2 minutes |

To confirm successful amplification, 5 µl of each reaction was visualised on an agarose gel (Section 7.1.1). For each sample, 5 µl of all PCR reactions was pooled and purified using a QIAquick PCR Purification Kit (#28104, QIAGEN), following the manufacturer's protocol. The PCR products were then quantified using a High Sensitivity Qubit and diluted to 40 pg/µl.

### 7.12.3 Second round PCR & purification

A second PCR was performed to add sequencing tags to the initial product (Table 7.14, primers detailed in Appendix A.1).

**Table 7.14. Second round gRNA PCR.** Top - reaction composition, bottom - thermocycler conditions.

| Reagent | Volume |
|---|---|
| 1st PCR Product (40 pg/µl dilution) | 5 µl (200 pg) |
| Primer mix (5 µM of both forward and reverse) | 2 µl |
| Nuclease-free water | 18 µl |
| KAPA HF HS 2X MM (#KK2601, Kapa Biosystems) | 25 µl |

| Cycle Number | Denature | Annealing | Extension |
|---|---|---|---|
| 1 | 98˚C, 30 seconds | | |
| 2-9 (8 cycles in total) | 98˚C, 10 seconds | 66˚C, 15 seconds | 72˚C, 20 seconds |
| 10 | | | 72˚C, 5 minutes |

For each sample, 1 µl of PCR product was run on a Bioanalyser to confirm that the correct band (~300 bp) was present at ~ 2 ng/µl. AMPure SPRI beads were added to the PCR products at a ratio of 0.8:1, mixed by pipetting and incubated at room temperature for 5 minutes. The Eppendorfs were placed on a magnetic rack for 5 minutes to allow the beads to be pulled out of the solution, and the supernatant was discarded. Samples were then washed twice with 200 µl of 80% ethanol for 30 seconds. Ethanol was discarded, the tube was centrifuged briefly to remove residual ethanol and then left to air dry for 5 minutes. The Eppendorfs were removed from the magnet and 35 µl of nuclease-free water was added; samples were left for 3 minutes to allow the DNA to elute. Tubes were placed back on the magnet for 3 minutes and the DNA solution was transferred to a clean Eppendorf. The Bioanalyser was used to confirm that the clean-up was successful and samples were stored at -20ºC prior to sequencing.

Single-end sequencing was performed on an Illumina HiSeq 2500 (primer detailed in Appendix A.1). Generally, six screen samples (replicates from two screens) were multiplexed and run together in two lanes. The neoR-IRES library was sequenced alone in two lanes.

### 7.12.4 Sequencing data analysis

After sequencing, reads were aligned to the gRNA sequences by the internal CASM pipeline. I then performed all analysis of gRNA sequencing data, with guidance and scripts kindly provided by Vivek Iyer (previous senior bioinformatician, Adams' lab) and Francesco Iorio (group leader, WSI). Details of the data analysis are provided in the relevant results chapters.

## 7.13 Assays to test library toxicity and expression issues

### 7.13.1 Comparison of BFP expression in BOB and BOB-Cas9 cells

BOB and BOB-Cas9 cells were detached as single cells, counted and transduced in suspension with the neoR-IRES library lentivirus. Both lines were seeded at $1.8x10^5$ cells/well in a 6-well plate with 0-90 µl of lentivirus and 2 ml total volume. After 72 h, cells were fixed and BFP expression was analysed by flow cytometry (Section 7.9).

### 7.13.2 Assay to measure BFP expression in single colonies

After completion of genome-wide screening with the neoR-IRES library, a small fraction of BOB-Cas9 cells were maintained in culture. Cells were seeded at a low density and after several days in culture, medium was aspirated and PBS was added. Under the microscope, 18 single colonies were scraped from the dish and transferred to a 24-well plate. Additionally, 2 colonies were picked from a population of untransduced BOB-Cas9 cells cultured for the same period of time These were cultured for 2 weeks and at the point of passage, a sample of cells were fixed and analysed by flow cytometry to measure BFP expression (Section 7.9).

### 7.13.3 Assay to test cross-resistance to G418

A BOB-Cas9 clonal line with ~100% BFP expression was mixed with an untransduced BOB-Cas9 clonal line (from the assay described in Section 7.13.2). Cells were detached as single cells, counted and combined in various ratios: 0:100, 30:70, 50:50, 75:25, 100:0 (BFP positive:untransduced). From these mixtures, 24,000 cells total/well were seeded in 12-well plates +/- 1 mg/ml G418. The remaining cells for each condition were fixed and analysed by flow cytometry to measure BFP expression (Section 7.9). After 5 days in culture, this analysis was repeated.

## 7.14 Karyotyping of iPSCs

iPSCs were harvested for karyotyping at 50-70% confluency. First, medium was aspirated and fresh TeSR-E8 medium with 10 μM ROCK inhibitor was added to the cells. Cells were then transferred to Sandra Louzada Gomes Pereira (FISH facility at WSI) to perform the rest of the assay. Nocodazole was added to a final concentration of 25-100 ng/ml and incubated for 2-3 hours at 37°C, 5% $CO_2$. Medium was transferred to a Falcon and cells were washed with PBS pre-warmed to 37°C. 1 ml of 37°C accutase was added and cells were incubated at 37°C, 5% $CO_2$ for 3-5 minutes. Cells were resuspended using the medium removed previously and centrifuged at 300 g for 5 minutes. Supernatant was aspirated and cells were resuspended gently in 8-10 ml of hypotonic buffer (0.4% KCl in 10 mM HEPES, pH 7.4). Cells were incubated for 10-20 minutes at 37°C, 5% $CO_2$. 1 ml of fixative (4:1 methanol:glacial acetic acid) was added and the tube was gently inverted. Cells were centrifuged at 300 g for 5 minutes, supernatant was removed and cells were resuspended. 1 ml of fixative was slowly added, mixed by pipetting and hypotonic buffer was added to a total volume of 10 ml. Cells were centrifuged at 300 g for 5 minutes and supernatant was removed. Cells were resuspended in 1 ml of fixative and 10-12 μl was dropped on to a clean microscope slide. Preparation of human 24-colour M-FISH probes and slide treatments were performed as described in [325,326]. Ten metaphases were analysed per sample.

## 7.15 Staining for pluripotency markers

This assay was performed with the assistance of Mary Goodwin (CGaP, WSI).

### 7.15.1 Seeding and fixation

BOB-Cas9 cells (untransduced and cells that had been screened with the neoR-IRES library) were detached as single cells, counted and seeded at 10,000 cells/well in a CELLSTAR® black 96-well plate (#655090, Greiner Bio-One). Cells were fixed after 72 hours: medium was aspirated and 50 μl of 4% PFA (#F8775, Sigma) was added to each well. Fixation was carried out at 4°C for 20 minutes. PFA was aspirated and cells were washed twice with 100 μl/well of PBS for 5 minutes at room temperature. Prior to staining, 100 μl/well of PBS was added, the plate was sealed in parafilm and stored at 4°C.

### 7.15.2 Staining

PBS was aspirated and 100 µl/well of 1% blocking solution (1g BSA/100 ml of 0.1% Triton X-100 in PBS) was added. Blocking solution was aspirated and 50 µl/well of diluted primary antibody (Table 7.15) was added (four wells/antibody for each cell line). As a control, 50 µl/well of 1% blocking solution was added (two wells/antibody for each cell line). Plates were sealed with parafilm and incubated overnight at 4ºC. DAPI solution (10 mg/ml) was added to each secondary antibody dilution to a final dilution of 1:1000, and these were kept on ice and protected from light. Primary antibody was aspirated and cells were washed three times with 100 µl/well PBS for 5 minutes. 50 µl/well of diluted secondary antibody plus DAPI was added, the plate was wrapped in foil and incubated at room temperature for 1 hour. Secondary antibody was aspirated and cells were washed 3 times with 100 µl/well PBS for 5 minutes. After the final wash, 100 µl/well of PBS was added and plates were sealed with a foil cover and stored at 4ºC prior to analysis.

Cells were analysed using an ArrayScan VTI HCS Reader. For each well, 10,000-12,000 DAPI positive cells were analysed for the presence of OCT4, NANOG or SOX2. The mean % of positive cells was calculated across the 4 technical replicates for each condition and the % of positive cells in the relevant control wells was subtracted to remove any background.

**Table 7.15. Antibodies for pluripotency test.**

| Antibody | Supplier | Product # | Dilution (in 1% blocking solution) | Secondary antibody |
|---|---|---|---|---|
| **Primary antibodies** | | | | |
| Oct4 | Santa Cruz Biotech | SC-5279 | 1:100 | Donkey anti-mouse |
| Sox2 | R&D Systems | AF2018 | 1:100 | Donkey anti-goat |
| Nanog | R&D Systems | AF1997 | 1:100 | Donkey anti-goat |
| **Secondary antibodies** | | | | |
| Donkey anti-mouse (AF647) | Invitrogen | A31571 | 1:1000 | - |
| Donkey anti-goat (AF488) | Invitrogen | A11055 | 1:1000 | - |

# 7.16 Validation assays

## 7.16.1 HAP1 cell culture

HAP1 cells (parental and knockout derivatives, Table 7.16) were purchased from Horizon Discovery (Cambridge, UK). All lines were cultured in IMDM (#12440053, Gibco, ThermoFisher) supplemented with 10% FBS at 37°C, 5% $CO_2$. Knockout lines were genotyped to confirm their mutational status as described in Section 7.4.

**Table 7.16. Details of HAP1 cell lines.**

| Cell line | Mutation | Product code |
|---|---|---|
| HAP1 parental control | N/A | C631 |
| HAP1 ARID1A knockout | 13 bp deletion in exon 2 | HZGHC000618c010 |
| HAP1 ARID1B knockout | 13 bp deletion in exon 2 | HZGHC000582c007 |
| HAP1 ARID2 knockout | 7 bp deletion in exon 3 | HZGHC000907c009 |
| HAP1 PBRM1 knockout | 8 bp deletion in exon 3 | HZGHC001135c010 |

### 7.16.1.1 Thawing HAP1 cells

Cryovials were heated in a 37°C water bath to thaw the cell suspension. A 4-fold volume of medium was added and centrifuged at 300 g for 3 minutes. The supernatant was aspirated, cells were resuspended and transferred to a flask containing fresh medium.

### 7.16.1.2 Passaging HAP1 cells

(All volumes given are for a T75 flask.) Medium was aspirated and cells were washed with 10 ml PBS. To detach cells, 2 ml of trypsin-EDTA (#25200056, Gibco, ThermoFisher) was added and incubated at 37°C for 3-5 minutes. An equal volume of medium was added to neutralise the trypsin and cells were transferred to a new flask, diluted to 1:15 for general maintenance. If an exact density was required, 10 µl of cell suspension was mixed 1:1 with Trypan Blue and live cell count was then calculated using a Countess™ II Automated Cell Counter.

### 7.16.1.3 Freezing HAP1 cells

Cells were detached as described above, and centrifuged at 300 g for 3 minutes. The supernatant was aspirated and cells were re-suspended in complete medium supplemented with

20% FBS and 10% DMSO. Cells were transferred to a cryovial and stored in a polystyrene rack at -80°C for 48 hours before being transferred to liquid nitrogen.

### 7.16.2 Cloning single gRNAs for validation

For validation of the *ARID1A/ARID1B* synthetic lethal interaction, gRNAs were selected from the neoR-IRES genome-wide library. For each gRNA, a pair of single-stranded DNA oligonucleotides containing the gRNA sequence (sense and anti-sense) were ordered, with overlaps to allow for cloning (Appendix A.1). Each oligonucleotide pair was phosphorylated and annealed (Table 7.17) in a thermocycler at 37°C for 30 minutes followed by 95°C for 5 minutes, then left to cool to room temperature. Annealed oligonucleotides were diluted to 7.1 fmol/µl in EB buffer (Qiagen).

**Table 7.17. Cloning single gRNA plasmids.** Top - annealing gRNA oligonucleotides, bottom left - restriction digest with BbsI, bottom right - ligation of gRNAs into backbone.

| Reagent | Volume |
|---|---|
| 100 µM top strand oligo | 1 µl |
| 100 µM bottom strand oligo | 1 µl |
| 10X T4 ligase buffer | 1 µl |
| T4 PNK enzyme (#M0201, NEB) | 0.5 µl |
| Nuclease-free water | 6.5 µl |

| Reagent | Volume/amount |
|---|---|
| Plasmid | 5 µg |
| 10X CutSmart Buffer | 5 µl |
| BbsI (#R0539S, NEB) | 1 µl |
| Nuclease-free water | to 50 µl total |

| Reagent | Volume |
|---|---|
| 20 ng/µl digested backbone | 1 µl |
| 7.1 fmol/µl annealed oligo | 2 µl |
| 10X ligase buffer | 1 µl |
| T4 ligase enzyme (#M0202S, NEB) | 1 µl |
| Nuclease-free water | 5 µl |

gRNA backbones containing either mCherry (Addgene #67977) or BFP (Addgene #67974) were digested with BbsI for 3 hours at 37°C then heat inactivated at 65°C for 20 minutes (Table 7.17). Digestion was confirmed on an agarose gel (Section 7.1.1) and both vectors were column purified (Section 7.1.2.1) then diluted to 20 ng/µl in nuclease-free water (QIAGEN).

Ligations were performed to insert each gRNA into either the mCherry or BFP backbone; these were incubated at 16ºC for 2 hours (Table 7.17). For each ligation, 2.5 µl was chemically transformed into 25 µl of OneShot TOP10 cells (Section 7.1.3.1). Three colonies per ligation were then inoculated in liquid LB broth + 100 µg/ml ampicillin and plasmid DNA was purified (Section 7.1.3.3). Successful cloning was confirmed by Sanger sequencing (Eurofins) of the gRNA region (Appendix A.1). One clone for each gRNA was packaged into a lentivirus and titred in BOB-Cas9 cells (Section 7.8). An mCherry backbone containing a gRNA targeting *ACCSL* and a BFP backbone containing a gRNA targeting *AIP1L* (the same backbones used for *ARID1A/ARID1B* gRNA cloning) were kindly provided by Nicola Thompson (PhD student in the Adams' lab).

## 7.16.3 Competitive growth assays for *ARID1A/ARID1B* SLI validation

### 7.16.3.1 Double gRNA assay

Cells were detached as single cells, counted and transduced in suspension. BOB-Cas9 cells were seeded at a density of 180,000 cells/well in a 6-well plate with 2 ml total volume. Each well was transduced with two lentiviruses, both containing a single gRNA plasmid, aiming for an MOI of ~0.6 for each. The following combinations of gRNA plasmids were transduced: *ARID1B*/*ACCSL*, *ARID1A/AIPL1, ACCSL/AIPL1.* A control well with no lentivirus was also included. Fresh complete medium was added the following morning. After 48 hours, cells were passaged to a new 6-well plate and a sample from each condition was fixed for flow cytometry analysis (Section 7.9). BFP and mCherry expression were measured to establish a baseline level; the % of cells in each population (two single positives and a double positive) was normalised against the untransduced population to calculate relative abundance. Cells were maintained for 2 weeks post-transduction, then were fixed for flow cytometry analysis (Section 7.9). BFP and mCherry expression were measured again and the % of cells in each population was normalised against the untransduced population to calculate relative abundance. The $\log_2$(fold-change) in relative abundance between day 14 and day 2 was calculated, as a measure of change in expression. The expected growth phenotype of the double positive population was calculated as the sum of the changes in both single positive populations.

#### 7.16.3.2 Single gRNA assay

BOB-Cas9, ARID1A_C09-Cas9 and ARID1B_C03-Cas9 lines were seeded and assayed in the same way as described in Section 7.16.3.1, but each line was transduced with only a single gRNA lentivirus per well. BOB-Cas9 cells were transduced with *AIPL1, ARID1A* and *ARID1B* gRNA plasmids. ARID1A_C09-Cas9 cells were transduced with *AIPL1* and *ARID1B* gRNA plasmids. ARID1B_C03-Cas9 cells were transduced with *AIPL1* and *ARID1A* gRNA plasmids. The % of BFP positive cells on day 14 was calculated as a $\log_2$(fold-change) relative to day 2.

### 7.16.4 Sanger arrayed library gRNAs

An internal facility at WSI hold glycerol stocks for all gRNAs in the Sanger Human Whole Genome CRISPR arrayed library. I provided a 96-well block containing liquid LB broth + 100 µg/ml ampicillin and the facility inoculated each well with a single gRNA plasmid (Appendix A.1). I cultured these overnight in a shaking incubator at 37ºC, 225 rpm, prepared glycerol stocks and purified plasmid DNA (Section 7.1.3). Each plasmid was sent for Sanger sequencing (Eurofins) with a primer complementary to the U6 promoter (Appendix A.1) to confirm the correct gRNA was present. These were then packaged into lentiviruses (Section 7.8.1).

### 7.16.5 Competitive growth assays for validation of screen hits

Cells were detached, counted and transduced in suspension (as described in Sections 7.2.3 and 7.15.1.2). HAP1-Cas9 cells (parental and KOs) were seeded at a density of 90,000 cells/well in a 12-well plate with 1 ml total volume. BOB-Cas9, ARID1A_C09-Cas9, ARID1B_C03-Cas9, ARID2_C11-Cas9 and PBRM1_F09-Cas9 cells were seeded at 100,000 cells/well in a 12-well plate or 200,000 cells/well in a 6-well plate with 1 ml or 2ml total volume, respectively. Polybrene was added to a final concentration of 8 µg/ml (for HAP1-Cas9 lines only) and each well was transduced with a lentivirus containing a single gRNA plasmid, aiming for an MOI of 0.5-0.8. A control well with no lentivirus was always included. Fresh complete medium was added the following morning. After 48 hours, cells were passaged to a new plate and a sample of each condition was fixed for flow cytometry analysis (Section 7.9). BFP expression was measured to establish a baseline level. Cells were maintained for 2 weeks post-transduction, then cells were fixed for flow cytometry analysis (Section 7.9) (iPSCs did not survive past 1 week). Expression of BFP was measured again on day 14 for HAP1-Cas9 cells and a $\log_2$(fold-change) was calculated relative to day 2, to assess change over time.

# 7.17 Methods for *in vivo* CRISPRa project

## 7.17.1 Cell culture

B16-F0-dCas9 cells were cultured in DMEM-high glucose medium (#11965092, Gibco, ThermoFisher), supplemented with 10% FBS, 1% penicillin/streptomycin/glutamine and 15 µg/ml blasticidin. Cells were incubated at 37°C, 5% $CO_2$. For passaging, medium was aspirated and the cells were washed once with 10 ml PBS (values given for a T150 flask). Cells were then incubated in 5 ml of trypsin-EDTA at 37°C for 5 minutes until they detached. An equal volume of medium was added, cells were re-suspended and seeded in a new flask with fresh medium, generally at a dilution of 1:15-1:30. Cells were counted by mixing 1:1 with trypan blue and analysing live cell number using a Countess™ II Automated Cell Counter (ThermoFisher).

## 7.17.2 Preparation of the CRISPRa library

The mCRISPRa-v2 m6 subpooled library was obtained from Addgene (#84003), transformed into ElectroMAX™ DH5α-E™ Competent Cells (#11319019, Invitrogen, ThermoFisher) and re-amplified in liquid culture following the Weismann lab protocol (available on the Addgene page). DNA was extracted using a QIAGEN EndoFree Plasmid Mega Kit (#12381, QIAGEN).

## 7.17.3 Titration of CRISPRa lentiviruses

Lentiviruses were prepared as described in Section 7.8.1. To measure the titre of the lentiviruses, cells were detached and diluted to $9x10^4$ cells/ml. Polybrene was added to a final concentration of 8 µg/ml. 2 ml of cell suspension was then added to a 6-well plate and different volumes of virus were added to each well. As a control, one well was cultured without virus. After 48-72 hours, cells were detached and prepared for flow cytometry (Section 7.9) Untransduced cells were used to set a gate for BFP expression. The % of BFP positive cells was used as an indicator of the MOI at each dose, and a line of best fit was plotted to allow calculation of the volume required for any given MOI. The amount of virus, cell number and total volume were scaled based on the culture surface area used in further transductions.

## 7.17.4 CRISPRa screen

### 7.17.4.1 *In vitro* phase

A total of $12x10^6$ cells were transduced with the m6 library lentivirus at an MOI of 0.3 in 180 ml total volume with 8 μg/ml polybrene. The cell suspension was split into 6x T150 flasks. A culture of untransduced cells was maintained in parallel as a control. Medium was changed the following day. Cells were passaged after 48 hours and $16x10^6$ cells were re-seeded in 5 μg/ml of puromycin was added. A small sample of transduced and untransduced cells were analysed by flow cytometry to measure BFP expression and confirm successful transduction (Section 7.9). After a further 4 days, cells were passaged again, maintaining $12x10^6$ cells. As before, some cells were taken for analysis by flow cytometry (Section 7.9). On day 9, cells were detached, counted, centrifuged at 300g for 5 minutes then diluted in PBS. In parallel, aliquots of $5.5x10^6$ (500x) cells were pelleted and frozen at -80°C.

### 7.17.4.2 *In vivo* phase

Seventy wildtype (C57BL6/NTac) female mice aged 6-8 weeks were intravenously administered (via the tail vein) 100 μl of cells in PBS at a concentration of $5.5x10^6$ cells/ml. Mice were then Schedule 1 sacrificed at two timepoints: 4 hours post-dosing and day 19 post-dosing (35 mice per timepoint), and saline cardiac perfused (to remove any blood from the lungs, thus removing circulating tumour cells). The lungs were then collected and snap-frozen.

## 7.17.5 Processing of CRISPRa screen samples

gDNA was extracted from both the lung samples and the ‘day 0’ B16-F0-dCas9 cells using the Puregene kit (Qiagen) according to the manufacturer’s protocol. PCR reactions were performed with 500 ng of gDNA per reaction, using the Phusion® High-Fidelity PCR Master Mix with HF Buffer (#M0531L, NEB) to amplify the gRNAs. The forward primer contained an 8mer barcode, 5’ Illumina adapters and homology to the CRISPRia-v2 plasmid. The reverse primer contained 3’ Illumina adapters and homology to the CRISPRia-v2 plasmid. For each lung, 16 PCR reactions were performed and all products were pooled. A portion of this was purified to select for only the ~280 bp product, using a Select-a-Size DNA Clean & Concentrator kit according to the manufacturer’s protocol (#D4080, Zymo). PCR was also performed on gDNA from the B16-F0-dCas9 cells collected on the day of dosing and the library plasmid DNA. Purity of all PCR samples was confirmed by analysis on a Bioanalyser. The samples were then

combined in two pools (each containing 35 lung samples + 2 cell samples + 1 plasmid sample) and sequenced on a HiSeq2500. Two sequencing primers were used: a bespoke primer and a standard Illumina primer.

## 7.17.6 Analysis of sequencing data

All sequencing results were processed and further analysed by Vivek Iyer using the methods described in Section 6.2.3.

## 7.17.7 Validation of screen hits

### 7.17.7.1 Preparation of single gRNA plasmids

Single gRNAs (detailed in Section 6.2.4) were individually cloned into the CRISPRia_v2 backbone, following the Weissman lab protocol (this can be accessed at https://weissmanlab.ucsf.edu/CRISPR/CRISPR.html). Plasmids were sent for Sanger sequencing (Eurofins) to confirm successful insertion of the gRNAs. The gRNA plasmids were packaged into lentiviruses (Section 7.8.1) and titred in B16-F0-dCas9 cells (Section 7.17.3).

### 7.17.7.2 *In vitro* phase

A total of $5x10^5$ cells were transduced with each lentivirus at an MOI of 0.3 and 8 μg/ml polybrene, in a total volume of 5.5 ml. Cells were seeded in a T25 flask. Untransduced cells were maintained in parallel as a control. Cells were processed, selected and analysed by flow cytometry in the same way as described for the screen (Section 7.17.4.1).

### 7.17.7.3 *In vivo* phase

On day 10 post-transduction, cells were detached, counted and diluted to $2x10^6$ cells/ml in PBS. Intravenous administration was performed via the tail vein with 100 μl of cell suspension ($2x10^5$ cells per mouse). In every cohort, 8-10 wildtype (C57BL/6NTac) mice aged 6-8 weeks were dosed per gRNA (both sexes were used for different cohorts, but same sex used within a cohort). On day 10 post-dosing, lungs were collected from all mice and metastases were counted by eye. The number of metastases present in mice carrying cells transduced with each gRNA was compared to those carrying the control gRNAs (NTC pool) in the same cohort. A Mann-Whitney test was performed to determine the significance of any differences observed.

# Bibliography


1 UK, C. R. *Cancer risk statistics*, <https://www.cancerresearchuk.org/health-professional/cancer-statistics/risk>

2 UK, C. R. *Cancer survival statistics*, <https://www.cancerresearchuk.org/health-professional/cancer-statistics/survival>

3 Siddik, Z. H. Cisplatin: mode of cytotoxic action and molecular basis of resistance. *Oncogene* **22**, 7265-7279, doi:10.1038/sj.onc.1206933 (2003).

4 Pommier, Y. Topoisomerase I inhibitors: camptothecins and beyond. *Nature Reviews Cancer* **6**, 789-802, doi:10.1038/nrc1977 (2006).

5 Longley, D. B., Harkin, D. P. & Johnston, P. G. 5-Fluorouracil: mechanisms of action and clinical strategies. *Nature Reviews Cancer* **3**, 330-338, doi:10.1038/nrc1074 (2003).

6 Chabner, B. A. & Roberts, T. G. Chemotherapy and the war on cancer. *Nature Reviews Cancer* **5**, 65-72, doi:10.1038/nrc1529 (2005).

7 Imai, K. & Takaoka, A. Comparing antibody and small-molecule therapies for cancer. *Nature Reviews Cancer* **6**, 714-727, doi:10.1038/nrc1913 (2006).

8 Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674, doi:10.1016/j.cell.2011.02.013 (2011).

9 Baselga, J. The EGFR as a target for anticancer therapy--focus on cetuximab. *European journal of cancer (Oxford, England : 1990)* **37 Suppl 4**, S16-22, doi:10.1016/s0959-8049(01)00233-7 (2001).

10 Carter, P. *et al.* Humanization of an anti-p185HER2 antibody for human cancer therapy. *Proceedings of the National Academy of Sciences* **89**, 4285, doi:10.1073/pnas.89.10.4285 (1992).

11 Souers, A. J. *et al.* ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. *Nature medicine* **19**, 202, doi:10.1038/nm.3048 (2013).

12 Ferrara, N., Hillan, K. J., Gerber, H. P. & Novotny, W. Discovery and development of bevacizumab, an anti-VEGF antibody for treating cancer. *Nat Rev Drug Discov* **3**, 391-400, doi:10.1038/nrd1381 (2004).

13 Duvic, M. & Vu, J. Vorinostat: a new oral histone deacetylase inhibitor approved for cutaneous T-cell lymphoma. *Expert opinion on investigational drugs* **16**, 1111-1120, doi:10.1517/13543784.16.7.1111 (2007).

14 Hodi, F. S. *et al.* Biologic activity of cytotoxic T lymphocyte-associated antigen 4 antibody blockade in previously vaccinated metastatic melanoma and ovarian carcinoma patients. *Proc Natl Acad Sci U S A* **100**, 4712-4717, doi:10.1073/pnas.0830997100 (2003).

15 Phan, G. Q. *et al.* Cancer regression and autoimmunity induced by cytotoxic T lymphocyte-associated antigen 4 blockade in patients with metastatic melanoma. *Proc Natl Acad Sci U S A* **100**, 8372-8377, doi:10.1073/pnas.1533209100 (2003).

16 Mullard, A. FDA approves landmark tissue-agnostic cancer drug. *Nature Reviews Drug Discovery* **18**, 7, doi:10.1038/nrd.2018.226 (2018).

17 Chial, H. Proto-oncogenes to oncogenes to cancer. *Nature Education* **1**, 33 (2008).

18 Davies, H. *et al.* Mutations of the BRAF gene in human cancer. *Nature* **417**, 949-954, doi:10.1038/nature00766 (2002).

19 Shimizu, K. *et al.* Three human transforming genes are related to the viral ras oncogenes. *Proc Natl Acad Sci U S A* **80**, 2112-2116, doi:10.1073/pnas.80.8.2112 (1983).

20 Chial, H. Tumor suppressor (TS) genes and the two-hit hypothesis. *Nature Education* **1**, 177 (2008).
21 Oren, M. p53: the ultimate tumor suppressor gene? *FASEB journal : official publication of the Federation of American Societies for Experimental Biology* **6**, 3169-3176, doi:10.1096/fasebj.6.13.1397838 (1992).
22 Murphree, A. L. & Benedict, W. F. Retinoblastoma: clues to human oncogenesis. *Science* **223**, 1028-1033, doi:10.1126/science.6320372 (1984).
23 Morris, L. G. & Chan, T. A. Therapeutic targeting of tumor suppressor genes. *Cancer* **121**, 1357-1368, doi:10.1002/cncr.29140 (2015).
24 Guo, X. E., Ngo, B., Modrek, A. S. & Lee, W.-H. Targeting tumor suppressor networks for cancer therapeutics. *Curr Drug Targets* **15**, 2-16 (2014).
25 Zheng, L. & Lee, W.-H. The Retinoblastoma Gene: A Prototypic and Multifunctional Tumor Suppressor. *Experimental Cell Research* **264**, 2-18, doi:10.1006/excr.2000.5129 (2001).
26 Du, W. & Searle, J. S. The rb pathway and cancer therapeutics. *Curr Drug Targets* **10**, 581-589 (2009).
27 Yamaguchi, T. *et al.* Identification of JTP-70902, a p15INK4b-inductive compound, as a novel MEK1/2 inhibitor. *Cancer Science* **98**, 1809-1816, doi:10.1111/j.1349-7006.2007.00604.x (2007).
28 Saidi, S. A., Holland, C. M., Charnock-Jones, D. S. & Smith, S. K. In vitro and in vivo effects of the PPAR-alpha agonists fenofibrate and retinoic acid in endometrial cancer. *Mol Cancer* **5**, 13-13, doi:10.1186/1476-4598-5-13 (2006).
29 Das, P. M. & Singal, R. DNA methylation and cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **22**, 4632-4642, doi:10.1200/jco.2004.07.151 (2004).
30 Bolden, J. E., Peart, M. J. & Johnstone, R. W. Anticancer activities of histone deacetylase inhibitors. *Nat Rev Drug Discov* **5**, 769-784, doi:10.1038/nrd2133 (2006).
31 Morris, L. G. T. & Chan, T. A. Therapeutic targeting of tumor suppressor genes. *Cancer* **121**, 1357-1368, doi:10.1002/cncr.29140 (2015).
32 Hartman, J. L., Garvik, B. & Hartwell, L. Principles for the Buffering of Genetic Variation. *Science* **291**, 1001, doi:10.1126/science.1056072 (2001).
33 Hartwell, L. H., Szankasi, P., Roberts, C. J., Murray, A. W. & Friend, S. H. Integrating genetic approaches into the discovery of anticancer drugs. *Science* **278**, 1064-1068, doi:10.1126/science.278.5340.1064 (1997).
34 Ashworth, A. & Lord, C. J. Synthetic lethal therapies for cancer: what's next after PARP inhibitors? *Nature Reviews Clinical Oncology* **15**, 564-576, doi:10.1038/s41571-018-0055-6 (2018).
35 Kaelin, W. G. The Concept of Synthetic Lethality in the Context of Anticancer Therapy. *Nature Reviews Cancer* **5**, 689-698, doi:10.1038/nrc1691 (2005).
36 Kroll, E. S., Hyland, K. M., Hieter, P. & Li, J. J. Establishing Genetic Interactions by a Synthetic Dosage Lethality Phenotype. *Genetics* **143**, 95 (1996).
37 Cox, A. D., Fesik, S. W., Kimmelman, A. C., Luo, J. & Der, C. J. Drugging the undruggable RAS: Mission possible? *Nat Rev Drug Discov* **13**, 828-851, doi:10.1038/nrd4389 (2014).
38 Muller, F. L. *et al.* Passenger deletions generate therapeutic vulnerabilities in cancer. *Nature* **488**, 337, doi:10.1038/nature11331 (2012).
39 Muller, F. L., Aquilanti, E. A. & DePinho, R. A. Collateral Lethality: A new therapeutic strategy in oncology. *Trends Cancer* **1**, 161-173, doi:10.1016/j.trecan.2015.10.002 (2015).

40 Jackson, S. P. & Bartek, J. The DNA-damage response in human biology and disease. *Nature* **461**, 1071-1078, doi:10.1038/nature08467 (2009).
41 Li, G.-M. Mechanisms and functions of DNA mismatch repair. *Cell Research* **18**, 85-98, doi:10.1038/cr.2007.115 (2008).
42 Krokan, H. E. & Bjørås, M. Base excision repair. *Cold Spring Harb Perspect Biol* **5**, a012583-a012583, doi:10.1101/cshperspect.a012583 (2013).
43 de Laat, W. L., Jaspers, N. G. & Hoeijmakers, J. H. Molecular mechanism of nucleotide excision repair. *Genes & development* **13**, 768-785, doi:10.1101/gad.13.7.768 (1999).
44 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* **79**, 181-211, doi:10.1146/annurev.biochem.052308.093131 (2010).
45 San Filippo, J., Sung, P. & Klein, H. Mechanism of Eukaryotic Homologous Recombination. *Annu Rev Biochem* **77**, 229-257, doi:10.1146/annurev.biochem.77.061306.125255 (2008).
46 Bryant, H. E. *et al.* Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. *Nature* **434**, 913-917, doi:10.1038/nature03443 (2005).
47 Farmer, H. *et al.* Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**, 917-921, doi:10.1038/nature03445 (2005).
48 Administration, U. S. F. D. *Niraparib (ZEJULA)*, <https://www.fda.gov/drugs/resources-information-approved-drugs/niraparib-zejula> (2017).
49 Administration, U. S. F. D. *Rucaparib*, <https://www.fda.gov/drugs/resources-information-approved-drugs/rucaparib> (2016).
50 Administration, U. S. F. D. *FDA approves olaparib for germline BRCA-mutated metastatic breast cancer*, <https://www.fda.gov/drugs/resources-information-approved-drugs/fda-approves-olaparib-germline-brca-mutated-metastatic-breast-cancer> (2018).
51 Administration, U. S. F. D. *FDA approves talazoparib for gBRCAm HER2-negative locally advanced or metastatic breast cancer*, <https://www.fda.gov/drugs/drug-approvals-and-databases/fda-approves-talazoparib-gbrcam-her2-negative-locally-advanced-or-metastatic-breast-cancer> (2018).
52 AstraZeneca. *LYNPARZA™ approved by the US Food and Drug Administration for the treatment of advanced ovarian cancer in patients with germline BRCA-mutations*, <https://www.astrazeneca-us.com/media/press-releases/2014/lynparza-approved-by-the-us-fda-20141219.html#> (2014).
53 Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563-547, doi:10.1126/science.274.5287.546 (1996).
54 Hanson, P. K. Saccharomyces cerevisiae: A Unicellular Model Genetic Organism of Enduring Importance. *Current Protocols Essential Laboratory Techniques* **16**, e21, doi:10.1002/cpet.21 (2018).
55 Yan Tong, A. H. & Boone, C. in *Yeast Protocol* (ed Wei Xiao) 171-191 (Humana Press, 2006).
56 Pan, X. *et al.* A DNA Integrity Network in the Yeast Saccharomyces cerevisiae. *Cell* **124**, 1069-1081, doi:10.1016/j.cell.2005.12.036 (2006).
57 Pan, X. *et al.* A Robust Toolkit for Functional Profiling of the Yeast Genome. *Molecular Cell* **16**, 487-496, doi:10.1016/j.molcel.2004.09.035 (2004).
58 Decourty, L. *et al.* Linking functionally related genes by sensitive and quantitative characterization of genetic interaction profiles. *Proceedings of the National Academy of Sciences* **105**, 5821, doi:10.1073/pnas.0710533105 (2008).

59 Tong, A. H. Y. *et al.* Global Mapping of the Yeast Genetic Interaction Network. *Science* **303**, 808, doi:10.1126/science.1091317 (2004).
60 Kachroo, A. H. *et al.* Evolution. Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science (New York, N.Y.)* **348**, 921-925, doi:10.1126/science.aaa0769 (2015).
61 Srivas, R. *et al.* A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. *Molecular cell* **63**, 514-525, doi:10.1016/j.molcel.2016.06.022 (2016).
62 Ryan, C. J., Bajrami, I. & Lord, C. J. Synthetic Lethality and Cancer - Penetrance as the Major Barrier. *Trends Cancer* **4**, 671-683, doi:10.1016/j.trecan.2018.08.003 (2018).
63 Dolma, S., Lessnick, S. L., Hahn, W. C. & Stockwell, B. R. Identification of genotype-selective antitumor agents using synthetic lethal chemical screening in engineered human tumor cells. *Cancer cell* **3**, 285-296 (2003).
64 Torrance, C. J., Agrawal, V., Vogelstein, B. & Kinzler, K. W. Use of isogenic human cancer cells for high-throughput screening and drug discovery. *Nat Biotechnol* **19**, 940-945, doi:10.1038/nbt1001-940 (2001).
65 Turcotte, S. *et al.* A molecule targeting VHL-deficient renal cell carcinoma that induces autophagy. *Cancer cell* **14**, 90-102, doi:10.1016/j.ccr.2008.06.004 (2008).
66 Ward, T. A., McHugh, P. J. & Durant, S. T. Small molecule inhibitors uncover synthetic genetic interactions of human flap endonuclease 1 (FEN1) with DNA damage response genes. *PloS one* **12**, e0179278-e0179278, doi:10.1371/journal.pone.0179278 (2017).
67 Reddy, A. S. & Zhang, S. Polypharmacology: drug discovery for the future. *Expert Rev Clin Pharmacol* **6**, 41-47, doi:10.1586/ecp.12.74 (2013).
68 MacDonald, M. L. *et al.* Identifying off-target effects and hidden phenotypes of drugs in human cells. *Nature Chemical Biology* **2**, 329-337, doi:10.1038/nchembio790 (2006).
69 Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806-811, doi:10.1038/35888 (1998).
70 Silva, J., Chang, K., Hannon, G. J. & Rivas, F. V. RNA-interference-based functional genomics in mammalian cells: reverse genetics coming of age. *Oncogene* **23**, 8401-8409, doi:10.1038/sj.onc.1208176 (2004).
71 Paddison, P. J. & Hannon, G. J. RNA interference: the new somatic cell genetics? *Cancer cell* **2**, 17-23, doi:10.1016/S1535-6108(02)00092-2 (2002).
72 Campeau, E. & Gobeil, S. RNA interference in mammals: behind the screen. *Briefings in Functional Genomics* **10**, 215-226, doi:10.1093/bfgp/elr018 (2011).
73 Boettcher, M. & Hoheisel, J. D. Pooled RNAi Screens - Technical and Biological Aspects. *Curr Genomics* **11**, 162-167, doi:10.2174/138920210791110988 (2010).
74 McDonald, E. R., 3rd *et al.* Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* **170**, 577-592.e510, doi:10.1016/j.cell.2017.07.005 (2017).
75 Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564-576.e516, doi:10.1016/j.cell.2017.06.010 (2017).
76 Williams, S. P., Barthorpe, A. S., Lightfoot, H., Garnett, M. J. & McDermott, U. High-throughput RNAi screen for essential genes and drug synergistic combinations in colorectal cancer. *Scientific Data* **4**, 170139, doi:10.1038/sdata.2017.139 (2017).
77 Semizarov, D. *et al.* Specificity of short interfering RNA determined through gene expression signatures. *Proc Natl Acad Sci U S A* **100**, 6347-6352, doi:10.1073/pnas.1131959100 (2003).
78 Jackson, A. L. *et al.* Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* **21**, 635-637, doi:10.1038/nbt831 (2003).

79 Mohr, S. E., Smith, J. A., Shamu, C. E., Neumüller, R. A. & Perrimon, N. RNAi screening comes of age: improved techniques and complementary approaches. *Nat Rev Mol Cell Biol* **15**, 591-600, doi:10.1038/nrm3860 (2014).
80 Boettcher, M. & McManus, M. T. Choosing the Right Tool for the Job: RNAi, TALEN, or CRISPR. *Molecular cell* **58**, 575-585, doi:10.1016/j.molcel.2015.04.028 (2015).
81 Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *Journal of Bacteriology* **169**, 5429, doi:10.1128/jb.169.12.5429-5433.1987 (1987).
82 Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. & Soria, E. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution* **60**, 174-182, doi:10.1007/s00239-004-0046-3 (2005).
83 Pourcel, C., Salvignol, G. & Vergnaud, G. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading, England)* **151**, 653-663, doi:10.1099/mic.0.27437-0 (2005).
84 Barrangou, R. *et al.* CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712, doi:10.1126/science.1138140 (2007).
85 Makarova, K. S., Grishin, N. V., Shabalina, S. A., Wolf, Y. I. & Koonin, E. V. A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biology direct* **1**, 7, doi:10.1186/1745-6150-1-7 (2006).
86 Lander, E. S. The Heroes of CRISPR. *Cell* **164**, 18-28, doi:10.1016/j.cell.2015.12.041 (2016).
87 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
88 Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826, doi:10.1126/science.1232033 (2013).
89 Le Rhun, A., Escalera-Maurer, A., Bratovič, M. & Charpentier, E. CRISPR-Cas in Streptococcus pyogenes. *RNA Biol* **16**, 380-389, doi:10.1080/15476286.2019.1582974 (2019).
90 Brouns, S. J. *et al.* Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964, doi:10.1126/science.1159689 (2008).
91 Deltcheva, E. *et al.* CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602, doi:10.1038/nature09886 (2011).
92 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
93 Garneau, J. E. *et al.* The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67, doi:10.1038/nature09523 (2010).
94 Chapman, J. R., Taylor, Martin R. G. & Boulton, Simon J. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Molecular Cell* **47**, 497-510, doi:10.1016/j.molcel.2012.07.029 (2012).
95 Barrangou, R. & Doudna, J. A. Applications of CRISPR technologies in research and beyond. *Nature Biotechnology* **34**, 933, doi:10.1038/nbt.3659 (2016).
96 Qi, L. S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-1183, doi:10.1016/j.cell.2013.02.022 (2013).
97 Gilbert, L. A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451, doi:10.1016/j.cell.2013.06.044 (2013).

98 Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57, doi:10.1038/nature26155 (2018).
99 Ma, D. *et al.* Engineer chimeric Cas9 to expand PAM recognition based on evolutionary information. *Nature Communications* **10**, 560, doi:10.1038/s41467-019-08395-8 (2019).
100 Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR–Cas technologies and applications. *Nature Reviews Molecular Cell Biology* **20**, 490-507, doi:10.1038/s41580-019-0131-5 (2019).
101 Guitart, J. R., Johnson, J. L. & Chien, W. W. Research Techniques Made Simple: The Application of CRISPR-Cas9 and Genome Editing in Investigative Dermatology. *Journal of Investigative Dermatology* **136**, e87-e93, doi:10.1016/j.jid.2016.06.007 (2016).
102 Kampmann, M. CRISPRi and CRISPRa Screens in Mammalian Cells for Precision Biology and Medicine. *ACS Chem Biol* **13**, 406-416, doi:10.1021/acschembio.7b00657 (2018).
103 Metzakopian, E. *et al.* Enhancing the genome editing toolbox: genome wide CRISPR arrayed libraries. *Scientific reports* **7**, 2244, doi:10.1038/s41598-017-01766-5 (2017).
104 Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84, doi:10.1126/science.1247005 (2014).
105 Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* **343**, 80, doi:10.1126/science.1246981 (2014).
106 Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3: Genes|Genomes|Genetics* **7**, 2719, doi:10.1534/g3.117.041277 (2017).
107 Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784, doi:10.1038/nmeth.3047 (2014).
108 Park, R. J. *et al.* A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nature genetics* **49**, 193-203, doi:10.1038/ng.3741 (2017).
109 Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511-516, doi:10.1038/s41586-019-1103-9 (2019).
110 Tzelepis, K. *et al.* A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep* **17**, 1193-1205, doi:10.1016/j.celrep.2016.09.079 (2016).
111 Morgens, D. W. *et al.* Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nature communications* **8**, 15178-15178, doi:10.1038/ncomms15178 (2017).
112 Allen, F. *et al.* JACKS: joint analysis of CRISPR/Cas9 knockout screens. *Genome Research*, doi:10.1101/gr.238923.118 (2019).
113 Diaz, A. A., Qin, H., Ramalho-Santos, M. & Song, J. S. HiTSelect: a comprehensive tool for high-complexity-pooled screen analysis. *Nucleic Acids Res* **43**, e16-e16, doi:10.1093/nar/gku1197 (2015).
114 Hart, T. & Moffat, J. BAGEL: a computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics* **17**, 164, doi:10.1186/s12859-016-1015-8 (2016).
115 König, R. *et al.* A probability-based approach for the analysis of large-scale RNAi screens. *Nature Methods* **4**, 847, doi:10.1038/nmeth1089 (2007).
116 Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biology* **15**, 554, doi:10.1186/s13059-014-0554-4 (2014).

117 Luo, J. *et al.* A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835-848, doi:10.1016/j.cell.2009.05.006 (2009).
118 Yu, J., Silva, J. & Califano, A. ScreenBEAM: a novel meta-analysis algorithm for functional genomics screens via Bayesian hierarchical modeling. *Bioinformatics (Oxford, England)* **32**, 260-267, doi:10.1093/bioinformatics/btv556 (2016).
119 Agrotis, A. & Ketteler, R. A new age in functional genomics using CRISPR/Cas9 in arrayed library screening. *Frontiers in genetics* **6**, 300, doi:10.3389/fgene.2015.00300 (2015).
120 Cui, Y., Xu, J., Cheng, M., Liao, X. & Peng, S. Review of CRISPR/Cas9 sgRNA Design Tools. *Interdisciplinary sciences, computational life sciences* **10**, 455-465, doi:10.1007/s12539-018-0298-z (2018).
121 Wilson, L. O. W., O'Brien, A. R. & Bauer, D. C. The Current State and Future of CRISPR-Cas9 gRNA Design Tools. *Front Pharmacol* **9**, 749-749, doi:10.3389/fphar.2018.00749 (2018).
122 Yau, E. H. *et al.* Genome-Wide CRISPR Screen for Essential Cell Growth Mediators in Mutant KRAS Colorectal Cancers. *Cancer Res* **77**, 6330-6339, doi:10.1158/0008-5472.Can-17-2043 (2017).
123 Sun, N. *et al.* <em>VHL</em> Synthetic Lethality Signatures Uncovered by Genotype-specific CRISPR-Cas9 Screens. *bioRxiv*, 588707, doi:10.1101/588707 (2019).
124 Wang, T. *et al.* Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* **168**, 890-903.e815, doi:10.1016/j.cell.2017.01.013 (2017).
125 Chan, E. M. *et al.* WRN helicase is a synthetic lethal target in microsatellite unstable cancers. *Nature* **568**, 551-556, doi:10.1038/s41586-019-1102-x (2019).
126 Wang, C. *et al.* Genome-wide CRISPR screens reveal synthetic lethality of RNASEH2 deficiency and ATR inhibition. *Oncogene* **38**, 2451-2463, doi:10.1038/s41388-018-0606-4 (2019).
127 Adikusuma, F., Pfitzner, C. & Thomas, P. Q. Versatile single-step-assembly CRISPR/Cas9 vectors for dual gRNA expression. *PloS one* **12**, e0187236-e0187236, doi:10.1371/journal.pone.0187236 (2017).
128 Han, K. *et al.* Synergistic drug combinations for cancer identified in a CRISPR screen for pairwise genetic interactions. *Nat Biotechnol* **35**, 463-474, doi:10.1038/nbt.3834 (2017).
129 Shen, J. P. *et al.* Combinatorial CRISPR-Cas9 screens for de novo mapping of genetic interactions. *Nat Methods* **14**, 573-576, doi:10.1038/nmeth.4225 (2017).
130 Vidigal, J. A. & Ventura, A. Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. *Nature Communications* **6**, 8083, doi:10.1038/ncomms9083 (2015).
131 Horlbeck, M. A. *et al.* Mapping the Genetic Landscape of Human Cells. *Cell* **174**, 953-967.e922, doi:10.1016/j.cell.2018.06.010 (2018).
132 Jacunski, A., Dixon, S. J. & Tatonetti, N. P. Connectivity Homology Enables Inter-Species Network Models of Synthetic Lethality. *PLOS Computational Biology* **11**, e1004506, doi:10.1371/journal.pcbi.1004506 (2015).
133 Wu, M. *et al.* In Silico Prediction of Synthetic Lethality by Meta-Analysis of Genetic Interactions, Functions, and Pathways in Yeast and Human Cancer. *Cancer Informatics* **13s3**, CIN.S14026, doi:10.4137/CIN.S14026 (2014).

134 Conde-Pueyo, N., Munteanu, A., Solé, R. V. & Rodríguez-Caso, C. Human synthetic lethal inference as potential anti-cancer target gene detection. *BMC Syst Biol* **3**, 116-116, doi:10.1186/1752-0509-3-116 (2009).
135 Deshpande, R. *et al.* A comparative genomic approach for identifying synthetic lethal interactions in human cancer. *Cancer Res* **73**, 6128-6136, doi:10.1158/0008-5472.Can-12-3956 (2013).
136 Cancer Genome Atlas Research, N. *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nature genetics* **45**, 1113-1120, doi:10.1038/ng.2764 (2013).
137 Ciriello, G., Cerami, E., Sander, C. & Schultz, N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome Research* **22**, 398-406 (2012).
138 Etemadmoghadam, D. *et al.* Synthetic lethality between CCNE1 amplification and loss of BRCA1. *Proc Natl Acad Sci U S A* **110**, 19489-19494, doi:10.1073/pnas.1314302110 (2013).
139 Jerby-Arnon, L. *et al.* Predicting Cancer-Specific Vulnerability via Data-Driven Detection of Synthetic Lethality. *Cell* **158**, 1199-1209, doi:10.1016/j.cell.2014.07.027 (2014).
140 Wang, W. *et al.* Purification and biochemical heterogeneity of the mammalian SWI-SNF complex. *EMBO J* **15**, 5370-5382 (1996).
141 Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E. & Green, M. R. Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* **370**, 477-481, doi:10.1038/370477a0 (1994).
142 Pan, J. *et al.* Interrogation of Mammalian Protein Complex Structure, Function, and Membership Using Genome-Scale Fitness Screens. *Cell systems* **6**, 555-568.e557, doi:10.1016/j.cels.2018.04.011 (2018).
143 Michel, B. C. *et al.* A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. *Nature cell biology* **20**, 1410-1420, doi:10.1038/s41556-018-0221-1 (2018).
144 Alpsoy, A. & Dykhuizen, E. C. Glioma tumor suppressor candidate region gene 1 (GLTSCR1) and its paralog GLTSCR1-like form SWI/SNF chromatin remodeling subcomplexes. *The Journal of biological chemistry* **293**, 3892-3903, doi:10.1074/jbc.RA117.001065 (2018).
145 Yoo, A. S., Staahl, B. T., Chen, L. & Crabtree, G. R. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature* **460**, 642-646, doi:10.1038/nature08139 (2009).
146 Lessard, J. *et al.* An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* **55**, 201-215, doi:10.1016/j.neuron.2007.06.019 (2007).
147 Lickert, H. *et al.* Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432**, 107-112, doi:10.1038/nature03071 (2004).
148 Tuoc, T. C. *et al.* Chromatin regulation by BAF170 controls cerebral cortical size and thickness. *Developmental cell* **25**, 256-269, doi:10.1016/j.devcel.2013.04.005 (2013).
149 Ho, L. *et al.* An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency. *Proc Natl Acad Sci U S A* **106**, 5181-5186, doi:10.1073/pnas.0812889106 (2009).
150 Li, S. *et al.* Genome-wide coactivation analysis of PGC-1alpha identifies BAF60a as a regulator of hepatic lipid metabolism. *Cell metabolism* **8**, 105-117, doi:10.1016/j.cmet.2008.06.013 (2008).

151 Meng, Z. X. *et al.* Baf60c drives glycolytic metabolism in the muscle and improves systemic glucose homeostasis through Deptor-mediated Akt activation. *Nature medicine* **19**, 640-645, doi:10.1038/nm.3144 (2013).
152 Hodges, C., Kirkland, J. G. & Crabtree, G. R. The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer. *Cold Spring Harb Perspect Med* **6**, a026930, doi:10.1101/cshperspect.a026930 (2016).
153 Kadoch, C. *et al.* Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet* **45**, 592-601, doi:10.1038/ng.2628 (2013).
154 Shain, A. H. & Pollack, J. R. The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS One* **8**, e55119, doi:10.1371/journal.pone.0055119 (2013).
155 Sandhya, S., Maulik, A., Giri, M. & Singh, M. Domain architecture of BAF250a reveals the ARID and ARM-repeat domains with implication in function and assembly of the BAF remodeling complex. *PLoS One* **13**, e0205267, doi:10.1371/journal.pone.0205267 (2018).
156 Li, X. S., Trojer, P., Matsumura, T., Treisman, J. E. & Tanese, N. Mammalian SWI/SNF--a subunit BAF250/ARID1 is an E3 ubiquitin ligase that targets histone H2B. *Molecular and cellular biology* **30**, 1673-1688, doi:10.1128/mcb.00540-09 (2010).
157 Jones, S. *et al.* Frequent mutations of chromatin remodeling gene ARID1A in ovarian clear cell carcinoma. *Science* **330**, 228-231, doi:10.1126/science.1196333 (2010).
158 Wiegand, K. C. *et al.* ARID1A mutations in endometriosis-associated ovarian carcinomas. *The New England journal of medicine* **363**, 1532-1543, doi:10.1056/NEJMoa1008433 (2010).
159 Kandoth, C. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67-73, doi:10.1038/nature12113 (2013).
160 Wang, K. *et al.* Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. *Nat Genet* **43**, 1219-1223, doi:10.1038/ng.982 (2011).
161 Gui, Y. *et al.* Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet* **43**, 875-878, doi:10.1038/ng.907 (2011).
162 Wang, X. *et al.* Two related ARID family proteins are alternative subunits of human SWI/SNF complexes. *Biochem J* **383**, 319-325, doi:10.1042/BJ20040524 (2004).
163 Nagl, N. G., Jr., Wang, X., Patsialou, A., Van Scoy, M. & Moran, E. Distinct mammalian SWI/SNF chromatin remodeling complexes with opposing roles in cell-cycle control. *EMBO J* **26**, 752-763, doi:10.1038/sj.emboj.7601541 (2007).
164 Cancer Genome Atlas Research, N. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43-49, doi:10.1038/nature12222 (2013).
165 Brownlee, P. M., Chambers, A. L., Cloney, R., Bianchi, A. & Downs, J. A. BAF180 promotes cohesion and prevents genome instability and aneuploidy. *Cell Rep* **6**, 973-981, doi:10.1016/j.celrep.2014.02.012 (2014).
166 Niimi, A., Chambers, A. L., Downs, J. A. & Lehmann, A. R. A role for chromatin remodellers in replication of damaged DNA. *Nucleic Acids Res* **40**, 7393-7403, doi:10.1093/nar/gks453 (2012).
167 Hodis, E. *et al.* A landscape of driver mutations in melanoma. *Cell* **150**, 251-263, doi:10.1016/j.cell.2012.06.024 (2012).
168 Ding, L. *et al.* Clonal architectures and driver mutations in metastatic melanomas. *PLoS One* **9**, e111153, doi:10.1371/journal.pone.0111153 (2014).

169 Lee, J. J. *et al.* Targeted next-generation sequencing reveals high frequency of mutations in epigenetic regulators across treatment-naive patient melanomas. *Clinical epigenetics* **7**, 59, doi:10.1186/s13148-015-0091-3 (2015).
170 Manceau, G. *et al.* Recurrent inactivating mutations of ARID2 in non-small cell lung carcinoma. *International journal of cancer* **132**, 2217-2221, doi:10.1002/ijc.27900 (2013).
171 Li, M. *et al.* Inactivating mutations of the chromatin remodeling gene ARID2 in hepatocellular carcinoma. *Nat Genet* **43**, 828-829, doi:10.1038/ng.903 (2011).
172 Helming, K. C. *et al.* ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nature medicine* **20**, 251-254, doi:10.1038/nm.3480 (2014).
173 Ogiwara, H. *et al.* Targeting the Vulnerability of Glutathione Metabolism in ARID1A-Deficient Cancers. *Cancer cell* **35**, 177-190.e178, doi:10.1016/j.ccell.2018.12.009 (2019).
174 Miller, R. E. *et al.* Synthetic Lethal Targeting of <em>ARID1A</em>-Mutant Ovarian Clear Cell Tumors with Dasatinib. *Molecular Cancer Therapeutics* **15**, 1472, doi:10.1158/1535-7163.MCT-15-0554 (2016).
175 Xiao, W., Awadallah, A. & Xin, W. Loss of ARID1A/BAF250a expression in ovarian endometriosis and clear cell carcinoma. *International journal of clinical and experimental pathology* **5**, 642-650 (2012).
176 Kwan, S. Y. *et al.* Loss of ARID1A expression leads to sensitivity to ROS-inducing agent elesclomol in gynecologic cancer cells. *Oncotarget* **7**, 56933-56943, doi:10.18632/oncotarget.10921 (2016).
177 Caumanns, J. J. *et al.* Integrative Kinome Profiling Identifies mTORC1/2 Inhibition as Treatment Strategy in Ovarian Clear Cell Carcinoma. *Clinical cancer research : an official journal of the American Association for Cancer Research* **24**, 3928-3940, doi:10.1158/1078-0432.Ccr-17-3060 (2018).
178 Shen, J. *et al.* ARID1A Deficiency Impairs the DNA Damage Checkpoint and Sensitizes Cells to PARP Inhibitors. *Cancer discovery* **5**, 752-767, doi:10.1158/2159-8290.CD-14-0849 (2015).
179 Williamson, C. T. *et al.* ATR inhibitors as a synthetic lethal therapy for tumours deficient in ARID1A. *Nature Communications* **7**, 13837, doi:10.1038/ncomms13837 (2016).
180 Bitler, B. G. *et al.* Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. *Nature medicine* **21**, 231, doi:10.1038/nm.3799 (2015).
181 Kim, K. H. *et al.* SWI/SNF-mutant cancers depend on catalytic and non-catalytic activity of EZH2. *Nature medicine* **21**, 1491-1496, doi:10.1038/nm.3968 (2015).
182 Januario, T. *et al.* PRC2-mediated repression of SMARCA2 predicts EZH2 inhibitor activity in SWI/SNF mutant tumors. *Proc Natl Acad Sci U S A* **114**, 12249-12254, doi:10.1073/pnas.1703966114 (2017).
183 Fukumoto, T. *et al.* Repurposing Pan-HDAC Inhibitors for ARID1A-Mutated Ovarian Cancer. *Cell Reports* **22**, 3393-3400, doi:10.1016/j.celrep.2018.03.019 (2018).
184 van der Vlag, J. & Otte, A. P. Transcriptional repression mediated by the human polycomb-group protein EED involves histone deacetylation. *Nat Genet* **23**, 474-478, doi:10.1038/70602 (1999).
185 Bitler, B. G. *et al.* ARID1A-mutated ovarian cancers depend on HDAC6 activity. *Nature cell biology* **19**, 962, doi:10.1038/ncb3582 (2017).
186 Hopkins, S. R., McGregor, G. A., Murray, J. M., Downs, J. A. & Savic, V. Novel synthetic lethality screening method identifies TIP60-dependent radiation sensitivity in the absence of BAF180. *DNA Repair* **46**, 47-54, doi:10.1016/j.dnarep.2016.05.030 (2016).

187 Xue, Y. *et al.* SMARCA4 loss is synthetic lethal with CDK4/6 inhibition in non-small cell lung cancer. *Nat Commun* **10**, 557, doi:10.1038/s41467-019-08380-1 (2019).
188 Xue, Y. *et al.* CDK4/6 inhibitors target SMARCA4-determined cyclin D1 deficiency in hypercalcemic small cell carcinoma of the ovary. *Nat Commun* **10**, 558, doi:10.1038/s41467-018-06958-9 (2019).
189 Romero, O. A. *et al.* <em>MAX</em> Inactivation in Small Cell Lung Cancer Disrupts MYC–SWI/SNF Programs and Is Synthetic Lethal with BRG1. *Cancer discovery* **4**, 292, doi:10.1158/2159-8290.CD-13-0799 (2014).
190 Lissanu Deribe, Y. *et al.* Mutations in the SWI/SNF complex induce a targetable dependence on oxidative phosphorylation in lung cancer. *Nature medicine* **24**, 1047-1057, doi:10.1038/s41591-018-0019-5 (2018).
191 Oike, T. *et al.* A Synthetic Lethality–Based Strategy to Treat Cancers Harboring a Genetic Deficiency in the Chromatin Remodeling Factor BRG1. *Cancer Research* **73**, 5508, doi:10.1158/0008-5472.CAN-12-4593 (2013).
192 Hoffman, G. R. *et al.* Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc Natl Acad Sci U S A* **111**, 3128-3133, doi:10.1073/pnas.1316793111 (2014).
193 Wang, X. *et al.* Oncogenesis caused by loss of the SNF5 tumor suppressor is dependent on activity of BRG1, the ATPase of the SWI/SNF chromatin remodeling complex. *Cancer research* **69**, 8094-8101, doi:10.1158/0008-5472.CAN-09-0733 (2009).
194 Jones, S. E. *et al.* ATR Is a Therapeutic Target in Synovial Sarcoma. *Cancer research* **77**, 7014-7026, doi:10.1158/0008-5472.CAN-17-2056 (2017).
195 Alimova, I. *et al.* Inhibition of EZH2 suppresses self-renewal and induces radiation sensitivity in atypical rhabdoid teratoid tumor cells. *Neuro Oncol* **15**, 149-160, doi:10.1093/neuonc/nos285 (2013).
196 Knutson, S. K. *et al.* Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc Natl Acad Sci U S A* **110**, 7922-7927, doi:10.1073/pnas.1303800110 (2013).
197 Wilson, B. G. *et al.* Epigenetic antagonism between polycomb and SWI/SNF complexes during oncogenic transformation. *Cancer cell* **18**, 316-328, doi:10.1016/j.ccr.2010.09.006 (2010).
198 Sun, Y., Jiang, X., Chen, S., Fernandes, N. & Price, B. D. A role for the Tip60 histone acetyltransferase in the acetylation and activation of ATM. *Proc Natl Acad Sci U S A* **102**, 13182-13187, doi:10.1073/pnas.0504211102 (2005).
199 Fenech, M. *et al.* Molecular mechanisms of micronucleus, nucleoplasmic bridge and nuclear bud formation in mammalian and human cells. *Mutagenesis* **26**, 125-132, doi:10.1093/mutage/geq052 (2011).
200 Clark, J. *et al.* Identification of novel genes, SYT and SSX, involved in the t(X;18)(p11.2;q11.2) translocation found in human synovial sarcoma. *Nature Genetics* **7**, 502-508, doi:10.1038/ng0894-502 (1994).
201 Kadoch, C. & Crabtree, G. R. Reversible disruption of mSWI/SNF (BAF) complexes by the SS18-SSX oncogenic fusion in synovial sarcoma. *Cell* **153**, 71-85, doi:10.1016/j.cell.2013.02.036 (2013).
202 Versteege, I. *et al.* Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer. *Nature* **394**, 203-206, doi:10.1038/28212 (1998).
203 Eaton, K. W., Tooke, L. S., Wainwright, L. M., Judkins, A. R. & Biegel, J. A. Spectrum of SMARCB1/INI1 mutations in familial and sporadic rhabdoid tumors. *Pediatric Blood & Cancer* **56**, 7-15, doi:10.1002/pbc.22831 (2011).
204 Biegel, J. A. *et al.* Germ-line and acquired mutations of INI1 in atypical teratoid and rhabdoid tumors. *Cancer Res* **59**, 74-79 (1999).

205 Hohmann, A. F. *et al.* Sensitivity and engineered resistance of myeloid leukemia cells to BRD9 inhibition. *Nature Chemical Biology* **12**, 672, doi:10.1038/nchembio.2115 (2016).
206 Martin, L. J. *et al.* Structure-Based Design of an in Vivo Active Selective BRD9 Inhibitor. *Journal of Medicinal Chemistry* **59**, 4462-4475, doi:10.1021/acs.jmedchem.5b01865 (2016).
207 Gulati, N., Beguelin, W. & Giulino-Roth, L. Enhancer of zeste homolog 2 (EZH2) inhibitors. *Leukemia & lymphoma* **59**, 1574-1585, doi:10.1080/10428194.2018.1430795 (2018).
208 Yee, A. J. *et al.* Ricolinostat plus lenalidomide, and dexamethasone in relapsed or refractory multiple myeloma: a multicentre phase 1b trial. *The Lancet Oncology* **17**, 1569-1578, doi:10.1016/S1470-2045(16)30375-8 (2016).
209 Caumanns, J. J., Wisman, G. B. A., Berns, K., van der Zee, A. G. J. & de Jong, S. ARID1A mutant ovarian clear cell carcinoma: A clear target for synthetic lethal strategies. *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* **1870**, 176-184, doi:10.1016/j.bbcan.2018.07.005 (2018).
210 Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:10.1016/j.cell.2006.07.024 (2006).
211 Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872, doi:10.1016/j.cell.2007.11.019 (2007).
212 Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science* **318**, 1917, doi:10.1126/science.1151526 (2007).
213 Cyranoski, D. 'Reprogrammed' stem cells approved to mend human hearts for the first time. *Nature* **557**, 619-620, doi:10.1038/d41586-018-05278-8 (2018).
214 Cyranoski, D. ‘Reprogrammed’ stem cells implanted into patient with Parkinson’s disease. *Nature*, doi:doi: 10.1038/d41586-018-07407-9 (2018).
215 Cyranoski, D. ‘Reprogrammed’ stem cells to treat spinal-cord injuries for the first time. *Nature*, doi:doi: 10.1038/d41586-019-00656-2 (2019).
216 Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature biotechnology* **27**, 851-857, doi:10.1038/nbt.1562 (2009).
217 Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nature biotechnology* **29**, 731-734, doi:10.1038/nbt.1927 (2011).
218 Santos, D. P., Kiskinis, E., Eggan, K. & Merkle, F. T. Comprehensive Protocols for CRISPR/Cas9-based Gene Editing in Human Pluripotent Stem Cells. *Curr Protoc Stem Cell Biol* **38**, 5B.6.1-5B.6.60, doi:10.1002/cpsc.15 (2016).
219 Alateeq, S. *et al.* Identification of on-target mutagenesis during correction of a beta-thalassemia splice mutation in iPS cells with optimised CRISPR/Cas9-double nickase reveals potential safety concerns. *APL Bioengineering* **2**, 046103, doi:10.1063/1.5048625 (2018).
220 Shinkuma, S., Guo, Z. & Christiano, A. M. Site-specific genome editing for correction of induced pluripotent stem cells derived from dominant dystrophic epidermolysis bullosa. *Proceedings of the National Academy of Sciences* **113**, 5676, doi:10.1073/pnas.1512028113 (2016).
221 Smith, C. *et al.* Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* **15**, 12-13, doi:10.1016/j.stem.2014.06.011 (2014).

222 Takayama, K. *et al.* Highly efficient biallelic genome editing of human ES/iPS cells using a CRISPR/Cas9 or TALEN system. *Nucleic Acids Res* **45**, 5198-5207, doi:10.1093/nar/gkx130 (2017).
223 Tan, E. P., Li, Y., Del Castillo Velasco-Herrera, M., Yusa, K. & Bradley, A. Off-target assessment of CRISPR-Cas9 guiding RNAs in human iPS and mouse ES cells. *genesis* **53**, 225-236, doi:10.1002/dvg.22835 (2015).
224 Park, C. Y. *et al.* Functional Correction of Large Factor VIII Gene Chromosomal Inversions in Hemophilia A Patient-Derived iPSCs Using CRISPR-Cas9. *Cell Stem Cell* **17**, 213-220, doi:10.1016/j.stem.2015.07.001 (2015).
225 Chang, C. W. *et al.* Modeling Human Severe Combined Immunodeficiency and Correction by CRISPR/Cas9-Enhanced Gene Targeting. *Cell Rep* **12**, 1668-1677, doi:10.1016/j.celrep.2015.08.013 (2015).
226 Li, C. *et al.* Novel HDAd/EBV Reprogramming Vector and Highly Efficient Ad/CRISPR-Cas Sickle Cell Disease Gene Correction. *Scientific reports* **6**, 30422-30422, doi:10.1038/srep30422 (2016).
227 Veres, A. *et al.* Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell* **15**, 27-30, doi:10.1016/j.stem.2014.04.020 (2014).
228 Ben Jehuda, R., Shemer, Y. & Binah, O. Genome Editing in Induced Pluripotent Stem Cells using CRISPR/Cas9. *Stem Cell Reviews and Reports* **14**, 323-336, doi:10.1007/s12015-018-9811-3 (2018).
229 Horii, T., Tamura, D., Morita, S., Kimura, M. & Hatada, I. Generation of an ICF syndrome model by efficient genome editing of human induced pluripotent stem cells using the CRISPR system. *International journal of molecular sciences* **14**, 19774-19781, doi:10.3390/ijms141019774 (2013).
230 Flynn, R. *et al.* CRISPR-mediated genotypic and phenotypic correction of a chronic granulomatous disease mutation in human iPS cells. *Experimental hematology* **43**, 838-848.e833, doi:10.1016/j.exphem.2015.06.002 (2015).
231 Wang, L. *et al.* CRISPR/Cas9-mediated targeted gene correction in amyotrophic lateral sclerosis patient iPSCs. *Protein & Cell* **8**, 365-378, doi:10.1007/s13238-017-0397-3 (2017).
232 Shi, Y., Inoue, H., Wu, J. C. & Yamanaka, S. Induced pluripotent stem cell technology: a decade of progress. *Nat Rev Drug Discov* **16**, 115-130, doi:10.1038/nrd.2016.245 (2017).
233 He, X. *et al.* Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. *Nucleic Acids Res* **44**, e85, doi:10.1093/nar/gkw064 (2016).
234 Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832, doi:10.1038/nbt.2647 (2013).
235 Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766, doi:10.7554/eLife.04766 (2014).
236 Zhang, Y. *et al.* Functional genomic screen of human stem cell differentiation reveals pathways involved in neurodevelopment and neurodegeneration. *Proceedings of the National Academy of Sciences* **110**, 12361, doi:10.1073/pnas.1309725110 (2013).
237 Chia, N.-Y. *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316, doi:10.1038/nature09531 (2010).
238 Ihry, R. J. *et al.* Genome-Scale CRISPR Screens Identify Human Pluripotency-Specific Genes. *Cell Reports* **27**, 616-630.e616, doi:10.1016/j.celrep.2019.03.043 (2019).

239 Mair, B. *et al.* Essential Gene Profiles for Human Pluripotent Stem Cells Identify Uncharacterized Genes and Substrate Dependencies. *Cell Reports* **27**, 599-615.e512, doi:10.1016/j.celrep.2019.02.041 (2019).
240 Yilmaz, A., Peretz, M., Aharony, A., Sagi, I. & Benvenisty, N. Defining essential genes for human pluripotent stem cells by CRISPR–Cas9 screening in haploid cells. *Nature cell biology* **20**, 610-619, doi:10.1038/s41556-018-0088-1 (2018).
241 Ihry, R. J. *et al.* p53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nature medicine* **24**, 939-946, doi:10.1038/s41591-018-0050-6 (2018).
242 Wong, C. C. *et al.* Inactivating CUX1 mutations promote tumorigenesis. *Nat Genet* **46**, 33-38, doi:10.1038/ng.2846 (2014).
243 Rashid, S. T. *et al.* Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *The Journal of clinical investigation* **120**, 3127-3136, doi:10.1172/jci43122 (2010).
244 Perlmutter, D. H. Autophagic disposal of the aggregation-prone protein that causes liver inflammation and carcinogenesis in α-1-antitrypsin deficiency. *Cell Death & Differentiation* **16**, 39-45, doi:10.1038/cdd.2008.103 (2009).
245 Yusa, K. *et al.* Targeted gene correction of α1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**, 391-394, doi:10.1038/nature10424 (2011).
246 Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191, doi:10.1038/nbt.3437 (2016).
247 Hart, T. *et al.* High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163**, 1515-1526, doi:10.1016/j.cell.2015.11.015 (2015).
248 Paludan, K., Duch, M., Jorgensen, P., Kjeldgaard, N. O. & Pedersen, F. S. Graduated resistance to G418 leads to differential selection of cultured mammalian cells expressing the neo gene. *Gene* **85**, 421-426, doi:10.1016/0378-1119(89)90435-6 (1989).
249 Kim, J. H. *et al.* High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. *PLoS One* **6**, e18556, doi:10.1371/journal.pone.0018556 (2011).
250 Jang, S. K. *et al.* A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. *J Virol* **62**, 2636-2643 (1988).
251 Pelletier, J. & Sonenberg, N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* **334**, 320-325, doi:10.1038/334320a0 (1988).
252 Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E. & Hayakawa, T. IRES-dependent second gene expression is significantly lower than cap-dependent first gene expression in a bicistronic vector. *Molecular therapy : the journal of the American Society of Gene Therapy* **1**, 376-382, doi:10.1006/mthe.2000.0050 (2000).
253 Sumarriva Lezama, L. *et al.* An analysis of blastic plasmacytoid dendritic cell neoplasm with translocations involving the MYC locus identifies t(6;8)(p21;q24) as a recurrent cytogenetic abnormality. *Histopathology* **73**, 767-776, doi:10.1111/his.13668 (2018).
254 Bejjani, B. A. & Shaffer, L. G. Application of array-based comparative genomic hybridization to clinical diagnostics. *J Mol Diagn* **8**, 528-533, doi:10.2353/jmoldx.2006.060029 (2006).
255 Iorio, F. *et al.* Unsupervised correction of gene-independent cell responses to CRISPR-Cas9 targeting. *BMC genomics* **19**, 604, doi:10.1186/s12864-018-4989-y (2018).

256 Oliveros, J. C. *Venny. An interactive tool for comparing lists with Venn's diagrams.* , <https://bioinfogp.cnb.csic.es/tools/venny/index.html> (2007-2015).
257 Qiu, Z., Oleinick, N. L. & Zhang, J. ATR/CHK1 inhibitors and cancer therapy. *Radiother Oncol* **126**, 450-464, doi:10.1016/j.radonc.2017.09.043 (2018).
258 Szklarczyk, D. *et al.* STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res* **47**, D607-d613, doi:10.1093/nar/gky1131 (2019).
259 Sato, E. *et al.* ARID1B as a Potential Therapeutic Target for ARID1A-Mutant Ovarian Clear Cell Carcinoma. *International journal of molecular sciences* **19**, doi:10.3390/ijms19061710 (2018).
260 Foucquier, J. & Guedj, M. Analysis of drug combinations: current methodological landscape. *Pharmacol Res Perspect* **3**, e00149-e00149, doi:10.1002/prp2.149 (2015).
261 Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nature genetics* **49**, 1779-1784, doi:10.1038/ng.3984 (2017).
262 Wilsker, D. *et al.* The DNA-binding properties of the ARID-containing subunits of yeast and mammalian SWI/SNF complexes. *Nucleic Acids Res* **32**, 1345-1353, doi:10.1093/nar/gkh277 (2004).
263 Flores-Alcantar, A., Gonzalez-Sandoval, A., Escalante-Alcalde, D. & Lomeli, H. Dynamics of expression of ARID1A and ARID1B subunits in mouse embryos and in cells during the cell cycle. *Cell and tissue research* **345**, 137-148, doi:10.1007/s00441-011-1182-x (2011).
264 Kelso, T. W. R. *et al.* Chromatin accessibility underlies synthetic lethality of SWI/SNF subunits in ARID1A-mutant cancers. *eLife* **6**, doi:10.7554/eLife.30506 (2017).
265 Trizzino, M. *et al.* The Tumor Suppressor ARID1A Controls Global Transcription via Pausing of RNA Polymerase II. *Cell reports* **23**, 3933-3945, doi:10.1016/j.celrep.2018.05.097 (2018).
266 Raab, J. R., Resnick, S. & Magnuson, T. Genome-Wide Transcriptional Regulation Mediated by Biochemically Distinct SWI/SNF Complexes. *PLoS genetics* **11**, e1005748, doi:10.1371/journal.pgen.1005748 (2015).
267 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).
268 Kuleshov, M. V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* **44**, W90-97, doi:10.1093/nar/gkw377 (2016).
269 Carette, J. E. *et al.* Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477**, 340-343, doi:10.1038/nature10348 (2011).
270 Jang, J.-H. FIGC, a novel FGF-induced ubiquitin-protein ligase in gastric cancers. *FEBS Lett* **578**, 21-25, doi:10.1016/j.febslet.2004.10.071 (2004).
271 Beilke, S. *et al.* The zinc-finger protein KCMF1 is overexpressed during pancreatic cancer development and downregulation of KCMF1 inhibits pancreatic cancer development in mice. *Oncogene* **29**, 4058-4067, doi:10.1038/onc.2010.156 (2010).
272 Zou, J., Mi, L., Yu, X. F. & Dong, J. Interaction of 14-3-3sigma with KCMF1 suppresses the proliferation and colony formation of human colon cancer stem cells. *World journal of gastroenterology* **19**, 3770-3780, doi:10.3748/wjg.v19.i24.3770 (2013).
273 Mandegar, M. A. *et al.* CRISPR Interference Efficiently Induces Specific and Reversible Gene Silencing in Human iPSCs. *Cell Stem Cell* **18**, 541-553, doi:10.1016/j.stem.2016.01.022 (2016).

274 Eskildsen, T. V. *et al.* MESP1 knock-down in human iPSC attenuates early vascular progenitor cell differentiation after completed primitive streak specification. *Developmental biology* **445**, 1-7, doi:10.1016/j.ydbio.2018.10.020 (2019).
275 Genga, R. M. J. *et al.* Single-Cell RNA-Sequencing-Based CRISPRi Screening Resolves Molecular Drivers of Early Human Endoderm Development. *Cell reports* **27**, 708-718.e710, doi:10.1016/j.celrep.2019.03.076 (2019).
276 Heman-Ackah, S. M., Bassett, A. R. & Wood, M. J. A. Precision Modulation of Neurodegenerative Disease-Related Gene Expression in Human iPSC-Derived Neurons. *Scientific reports* **6**, 28420-28420, doi:10.1038/srep28420 (2016).
277 Liu, S. J. *et al.* CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. *Science (New York, N.Y.)* **355**, aah7111, doi:10.1126/science.aah7111 (2017).
278 Luo, Y. *et al.* Targeted Inhibition of the miR-199a/214 Cluster by CRISPR Interference Augments the Tumor Tropism of Human Induced Pluripotent Stem Cell-Derived Neural Stem Cells under Hypoxic Condition. *Stem cells international* **2016**, 3598542, doi:10.1155/2016/3598542 (2016).
279 Tian, R. *et al.* CRISPR Interference-Based Platform for Multimodal Genetic Screens in Human iPSC-Derived Neurons. *Neuron*, doi:10.1016/j.neuron.2019.07.014 (2019).
280 Downward, J. RAS Synthetic Lethal Screens Revisited: Still Seeking the Elusive Prize? *Clinical cancer research : an official journal of the American Association for Cancer Research* **21**, 1802-1809, doi:10.1158/1078-0432.CCR-14-2180 (2015).
281 Chaffer, C. L. & Weinberg, R. A. A Perspective on Cancer Cell Metastasis. *Science* **331**, 1559-1564 (2011).
282 Liu, Q. *et al.* Factors involved in cancer metastasis: a better understanding to "seed and soil" hypothesis. *Mol Cancer* **16**, 176-176, doi:10.1186/s12943-017-0742-4 (2017).
283 Janssen, L. M. E., Ramsay, E. E., Logsdon, C. D. & Overwijk, W. W. The immune system in cancer metastasis: friend or foe? *Journal for ImmunoTherapy of Cancer* **5**, 79, doi:10.1186/s40425-017-0283-9 (2017).
284 Bubenik, J. Tumour MHC class I downregulation and immunotherapy (Review). *Oncology reports* **10**, 2005-2008 (2003).
285 Blomberg, O. S., Spagnuolo, L. & de Visser, K. E. Immune regulation of metastasis: mechanistic insights and therapeutic opportunities. *Disease models & mechanisms* **11**, doi:10.1242/dmm.036236 (2018).
286 Camerer, E. *et al.* Platelets, protease-activated receptors, and fibrinogen in hematogenous metastasis. *Blood* **104**, 397-401, doi:10.1182/blood-2004-02-0434 (2004).
287 Palumbo, J. S. *et al.* Platelets and fibrin(ogen) increase metastatic potential by impeding natural killer cell-mediated elimination of tumor cells. *Blood* **105**, 178-185, doi:10.1182/blood-2004-06-2272 (2005).
288 Labelle, M., Begum, S. & Hynes, R. O. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. *Cancer cell* **20**, 576-590, doi:10.1016/j.ccr.2011.09.009 (2011).
289 Khanna, C. & Hunter, K. Modeling metastasis in vivo. *Carcinogenesis* **26**, 513-523, doi:10.1093/carcin/bgh261 (2005).
290 Fidler, I. J. Metastasis: Quantitative Analysis of Distribution and Fate of Tumor Emboli Labeled With 125I-5-Iodo-2′ -deoxyuridine23. *JNCI: Journal of the National Cancer Institute* **45**, 773-782, doi:10.1093/jnci/45.4.773 (1970).
291 Fidler, I. J. Selection of Successive Tumour Lines for Metastasis. *Nature New Biology* **242**, 148-149, doi:10.1038/newbio242148a0 (1973).

292 Fidler, I. J. Biological Behavior of Malignant Melanoma Cells Correlated to Their Survival in Vivo. *Cancer Research* **35**, 218-224 (1975).
293 van der Weyden, L. *et al.* Genome-wide in vivo screen identifies novel host regulators of metastatic colonization. *Nature* **541**, 233, doi:10.1038/nature20792 (2017).
294 Gumireddy, K. *et al.* KLF17 is a negative regulator of epithelial-mesenchymal transition and metastasis in breast cancer. *Nature cell biology* **11**, 1297-1304, doi:10.1038/ncb1974 (2009).
295 Gumireddy, K. *et al.* In vivo selection for metastasis promoting genes in the mouse. *Proc Natl Acad Sci U S A* **104**, 6696-6701, doi:10.1073/pnas.0701145104 (2007).
296 Sasaki, K. *et al.* Genome-wide in vivo RNAi screen identifies ITIH5 as a metastasis suppressor in pancreatic cancer. *Clinical & experimental metastasis* **34**, 229-239, doi:10.1007/s10585-017-9840-3 (2017).
297 Katigbak, A. *et al.* A CRISPR/Cas9 Functional Screen Identifies Rare Tumor Suppressors. *Scientific reports* **6**, 38968, doi:10.1038/srep38968 (2016).
298 Kodama, M. *et al.* In vivo loss-of-function screens identify KPNB1 as a new druggable oncogene in epithelial ovarian cancer. *Proc Natl Acad Sci U S A* **114**, E7301-E7310, doi:10.1073/pnas.1705441114 (2017).
299 Manguso, R. T. *et al.* In vivo CRISPR screening identifies Ptpn2 as a cancer immunotherapy target. *Nature* **547**, 413-418, doi:10.1038/nature23270 (2017).
300 Chen, S. *et al.* Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246-1260, doi:10.1016/j.cell.2015.02.038 (2015).
301 Maeder, M. L. *et al.* CRISPR RNA–guided activation of endogenous human genes. *Nature Methods* **10**, 977, doi:10.1038/nmeth.2598 (2013).
302 Braun, C. J. *et al.* Versatile in vivo regulation of tumor phenotypes by dCas9-mediated transcriptional perturbation. *Proc Natl Acad Sci U S A* **113**, E3892-3900, doi:10.1073/pnas.1600582113 (2016).
303 Wangensteen, K. J. *et al.* Combinatorial genetics in liver repopulation and carcinogenesis with a in vivo CRISPR activation platform. *Hepatology (Baltimore, Md.)* **68**, 663-676, doi:10.1002/hep.29626 (2018).
304 Zhou, H. *et al.* In vivo simultaneous transcriptional activation of multiple genes in the brain using CRISPR-dCas9-activator transgenic mice. *Nature neuroscience* **21**, 440-446, doi:10.1038/s41593-017-0060-6 (2018).
305 Horlbeck, M. A. *et al.* Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *eLife* **5**, doi:10.7554/eLife.19760 (2016).
306 Chong, Z. S., Ohnishi, S., Yusa, K. & Wright, G. J. Pooled extracellular receptor-ligand interaction screening using CRISPR activation. *Genome Biol* **19**, 205, doi:10.1186/s13059-018-1581-3 (2018).
307 Wu, W., Baker, M. E., Eraly, S. A., Bush, K. T. & Nigam, S. K. Analysis of a large cluster of SLC22 transporter genes, including novel USTs, reveals species-specific amplification of subsets of family members. *Physiol Genomics* **38**, 116-124, doi:10.1152/physiolgenomics.90309.2008 (2009).
308 Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & cellular proteomics : MCP* **13**, 397-406, doi:10.1074/mcp.M113.035600 (2014).
309 Alper, S. L. Molecular physiology of SLC4 anion exchangers. *Experimental physiology* **91**, 153-161, doi:10.1113/expphysiol.2005.031765 (2006).
310 Sander, T. *et al.* Association of the 867Asp variant of the human anion exchanger 3 gene with common subtypes of idiopathic generalized epilepsy. *Epilepsy research* **51**, 249-255, doi:10.1016/s0920-1211(02)00152-3 (2002).

311 Wang, H. S., Chen, Y., Vairamani, K. & Shull, G. E. Critical role of bicarbonate and bicarbonate transporters in cardiac function. *World journal of biological chemistry* **5**, 334-345, doi:10.4331/wjbc.v5.i3.334 (2014).
312 Birnbaum, S. G. *et al.* Structure and function of Kv4-family transient potassium channels. *Physiol Rev* **84**, 803-833, doi:10.1152/physrev.00039.2003 (2004).
313 Chien, W. & Pei, L. A novel binding factor facilitates nuclear translocation and transcriptional activation function of the pituitary tumor-transforming gene product. *The Journal of biological chemistry* **275**, 19422-19427, doi:10.1074/jbc.M910105199 (2000).
314 Watkins, R. J. *et al.* Pituitary tumor transforming gene binding factor: a new gene in breast cancer. *Cancer Res* **70**, 3739-3749, doi:10.1158/0008-5472.Can-09-3531 (2010).
315 Read, M. L. *et al.* The PTTG1-binding factor (PBF/PTTG1IP) regulates p53 activity in thyroid cells. *Endocrinology* **155**, 1222-1234, doi:10.1210/en.2013-1646 (2014).
316 Read, M. L. *et al.* The proto-oncogene PBF binds p53 and is associated with prognostic features in colorectal cancer. *Molecular carcinogenesis* **55**, 15-26, doi:10.1002/mc.22254 (2016).
317 Watkins, R. J. *et al.* Pro-invasive Effect of Proto-oncogene PBF Is Modulated by an Interaction with Cortactin. *The Journal of clinical endocrinology and metabolism* **101**, 4551-4563, doi:10.1210/jc.2016-1932 (2016).
318 Glerup, S. *et al.* SorCS2 is required for BDNF-dependent plasticity in the hippocampus. *Molecular Psychiatry* **21**, 1740-1751, doi:10.1038/mp.2016.108 (2016).
319 Malik, A. R. *et al.* SorCS2 Controls Functional Expression of Amino Acid Transporter EAAT3 and Protects Neurons from Oxidative Stress and Epilepsy-Induced Pathology. *Cell reports* **26**, 2792-2804.e2796, doi:10.1016/j.celrep.2019.02.027 (2019).
320 Baum, A. E. *et al.* A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. *Mol Psychiatry* **13**, 197-207, doi:10.1038/sj.mp.4002012 (2008).
321 Christoforou, A. *et al.* Convergence of linkage, association and GWAS findings for a candidate region for bipolar disorder and schizophrenia on chromosome 4p. *Molecular Psychiatry* **16**, 240-242, doi:10.1038/mp.2010.25 (2011).
322 Goffinet, A. M. & Tissir, F. Seven pass Cadherins CELSR1-3. *Seminars in cell & developmental biology* **69**, 102-110, doi:10.1016/j.semcdb.2017.07.014 (2017).
323 Qiao, Z. *et al.* Mutations in KIAA1109, CACNA1C, BSN, AKAP13, CELSR2, and HELZ2 Are Associated With the Prognosis in Endometrial Cancer. *Front Genet* **10**, 909, doi:10.3389/fgene.2019.00909 (2019).
324 Roumeliotis, T. I. *et al.* Genomic Determinants of Protein Abundance Variation in Colorectal Cancer Cells. *Cell Rep* **20**, 2201-2214, doi:10.1016/j.celrep.2017.08.010 (2017).
325 Agu, C. A. *et al.* Successful Generation of Human Induced Pluripotent Stem Cell Lines from Blood Samples Held at Room Temperature for up to 48 hr. *Stem cell reports* **5**, 660-671, doi:10.1016/j.stemcr.2015.08.012 (2015).
326 Gonçalves, E. *et al.* Structural rearrangements generate cell-specific, gene-independent CRISPR-Cas9 loss of fitness effects. *Genome Biology* **20**, 27, doi:10.1186/s13059-019-1637-z (2019).

# Appendix A

Datasets that were too large to be included in the main thesis have been provided in a digital format and have been referenced throughout. The contents of each file are detailed below.

## A.1 Oligonucleotide sequences

Sequences and details are provided for all oligonucleotides, primers, and gRNAs used for the work described in this thesis. They are grouped into different sheets based on the section of work that they relate to.

## A.2 Mass spectrometry data

The $\log_2$(fold-change) in protein abundance for each KO line (compared to the parental BOB line) is shown for all targeted proteins.

## A.3 NeoR-IRES library backbone

The plasmid map and full sequence of the final neoR-IRES library backbone is provided.

## A.4 gRNA library read counts

The raw gRNA read counts from HiSeq2500 analysis of the neoR-IRES, Yusa v1, and Yusa v1.1 gRNA libraries are provided.

## A.5 Bayes Factors for iPSC screens

The gene-level BFs obtained from BAGELR analysis of each screen are shown.

## A.6 Scaled Bayes Factors for iPSC screens

Gene-level BFs were scaled based on an FDR of 0.05, with any value greater than 0 representing a significant hit. Scaled BFs for all screens are shown.

## A.7 MAGeCK depletion values for iPSC screens

The gene-level depletion (negative FDR) values obtained from MAGeCK analysis of each screen are shown.

## A.8 Biological replicate overlaps

Genes that were significantly depleted in the biological replicates of the parental BOB and *TP53* KO line screens are shown.

## A.9 KO-specific depleted genes

This file shows genes that were significantly depleted in each KO line (based on BAGELR analysis) but not in the parental BOB line screens. Results are provided for three filtering strategies. Genes with a scaled BF > 0 (highlighted in red) were significantly depleted.

## A.10 MAGeCK enrichment values for iPSC screens

The gene-level enrichment (positive FDR) values obtained from MAGeCK analysis of each screen are shown.

## A.11 PBAF/BAF mutant cancer cell lines

Of the cell lines screened by the Sanger[109] and Broad,[261] those lines that harbour LOF (nonsense or frameshift indel) mutations in *ARID1A, ARID1B, ARID2* or *PBRM1* are listed. These were used for the association analyses described in Chapter 4.

## A.12 Sanger Institute screen data

Scaled BFs (FDR 0.05) are provided for all cancer cell line screens screened by Sanger, as published by Behan *et al.* (2019).[109]

## A.13 Broad Institute screen data

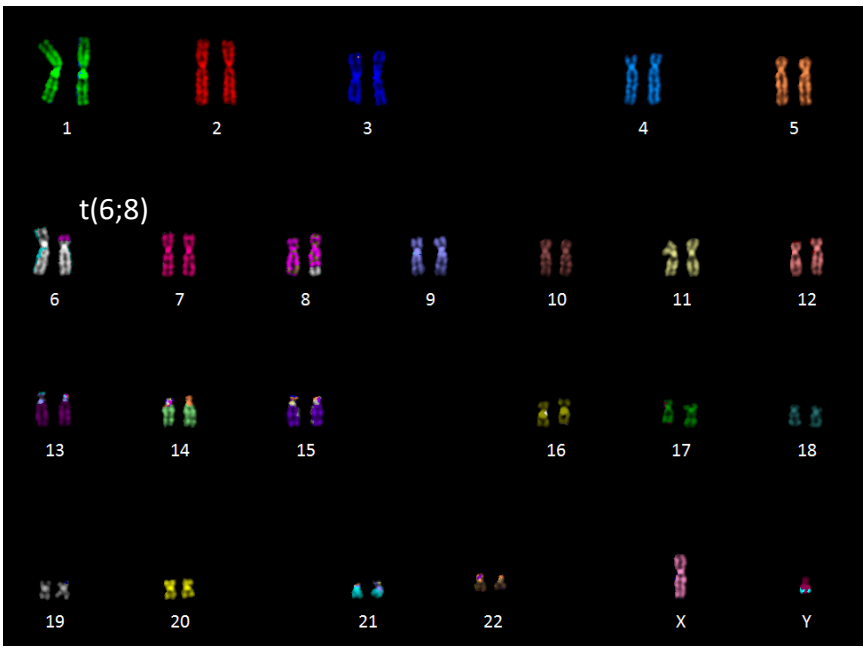
Raw gRNA counts from the Broad cancer cell line screens[261] were re-processed using the pipeline applied by Behan *et al.*[109] (by Clare Pacini, a postdoctoral fellow in Francesco Iorio's lab). Scaled BFs (FDR 0.05) obtained from BAGELR analysis are shown for all lines.

## A.14 Candidate PBAF-BAF gene SLIs

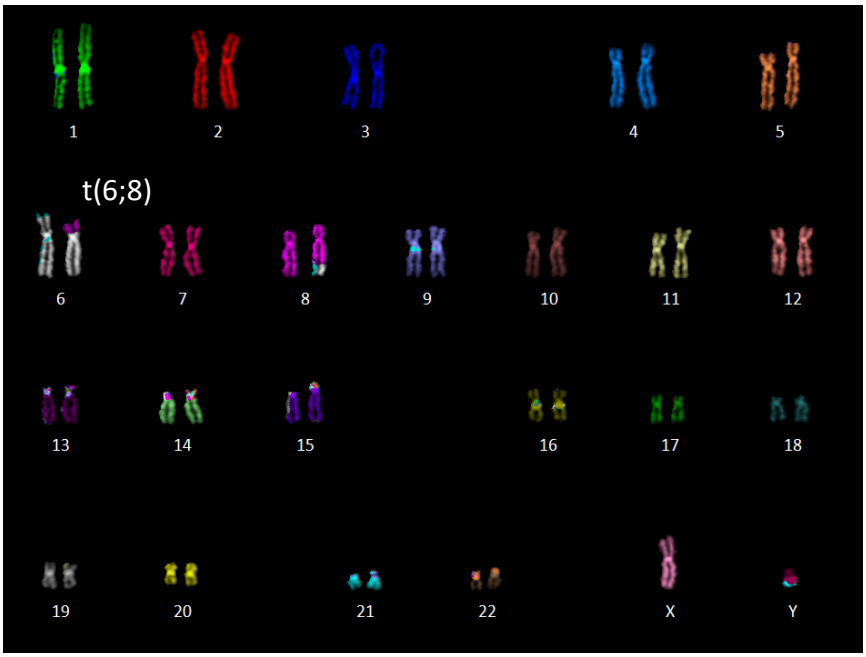
Genes that were identified after filtering for hits specific to PBAF/BAF gene KO lines (as described in Section 4.3) are provided, with the scaled BFs shown for all screens that were used in the filtering. Their status at each stage of the filtering is indicated in the final columns.

# Appendix B

**a.**




**b.**



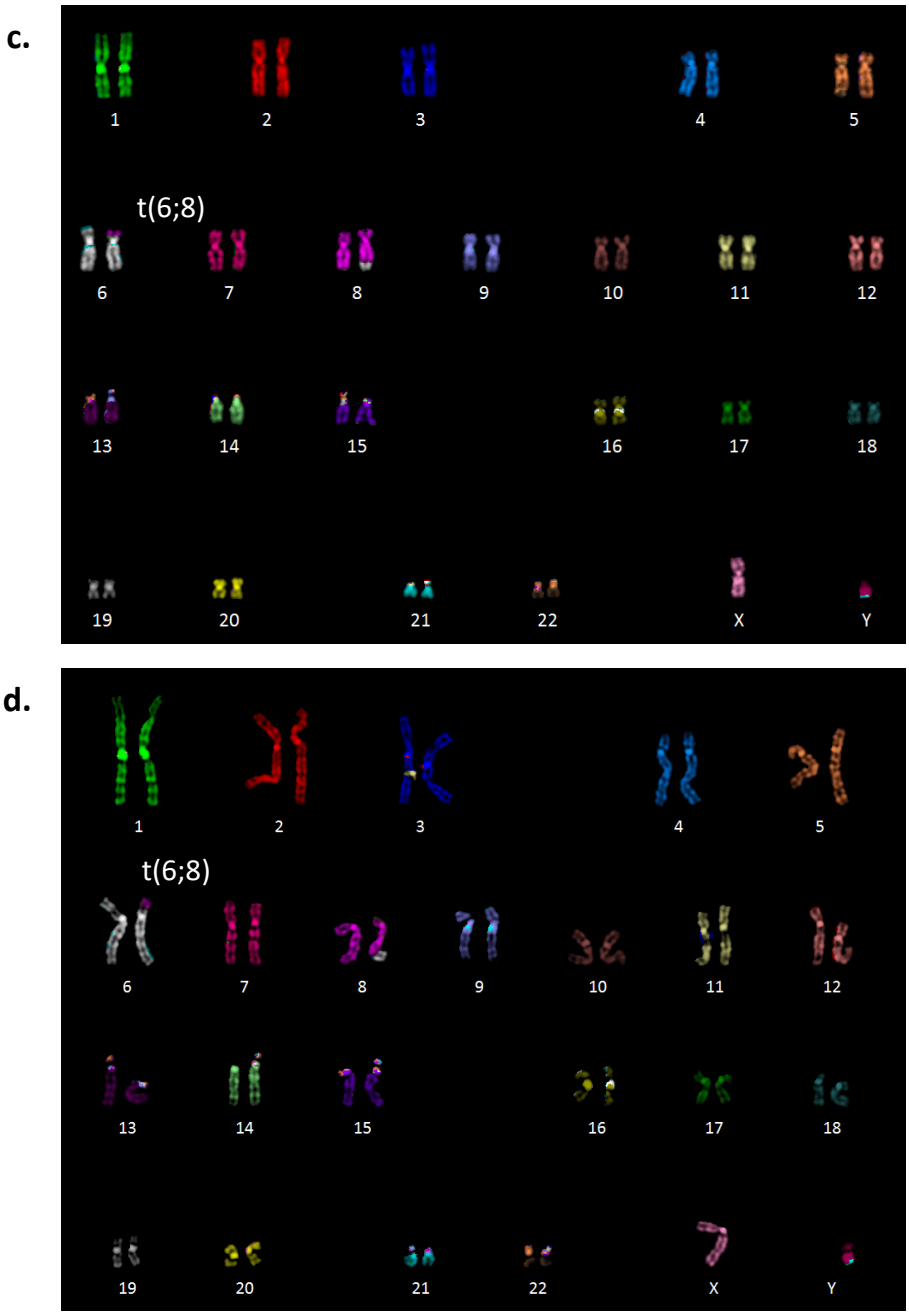



**Figure B.1. Karyotype of KO BOB-Cas9 cells.** M-FISH of ARID1A_C09-Cas9 **(a)**, ARID1B_C03-as9 **(b)**, ARID2_C11-Cas9 **(c)**, and PBRM1_F09-Cas9 **(d)** cell lines. Ten randomly selected metaphases were analysed for each line; a representative karyotype is shown for each. A balanced translocation between chromosomes 6 and 8 is indicated.

**Table B.1. Essential genes identified in parental BOB screens and published hESC screens.**

| Gene |
|---|
| *CDIPT* |
| *CYC1* |
| *FANCA* |
| *FDXR* |
| *H2AFZ* |
| *MBTPS1* |
| *MRPL50* |
| *RPL32* |
| *RPL34* |
| *RPS25* |
| *SCAP* |
| *SDE2* |
| *SETD1B* |
| *SNRNP40* |
| *TBCB* |
| *WDR1* |
| *CTDSPL2* |