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.

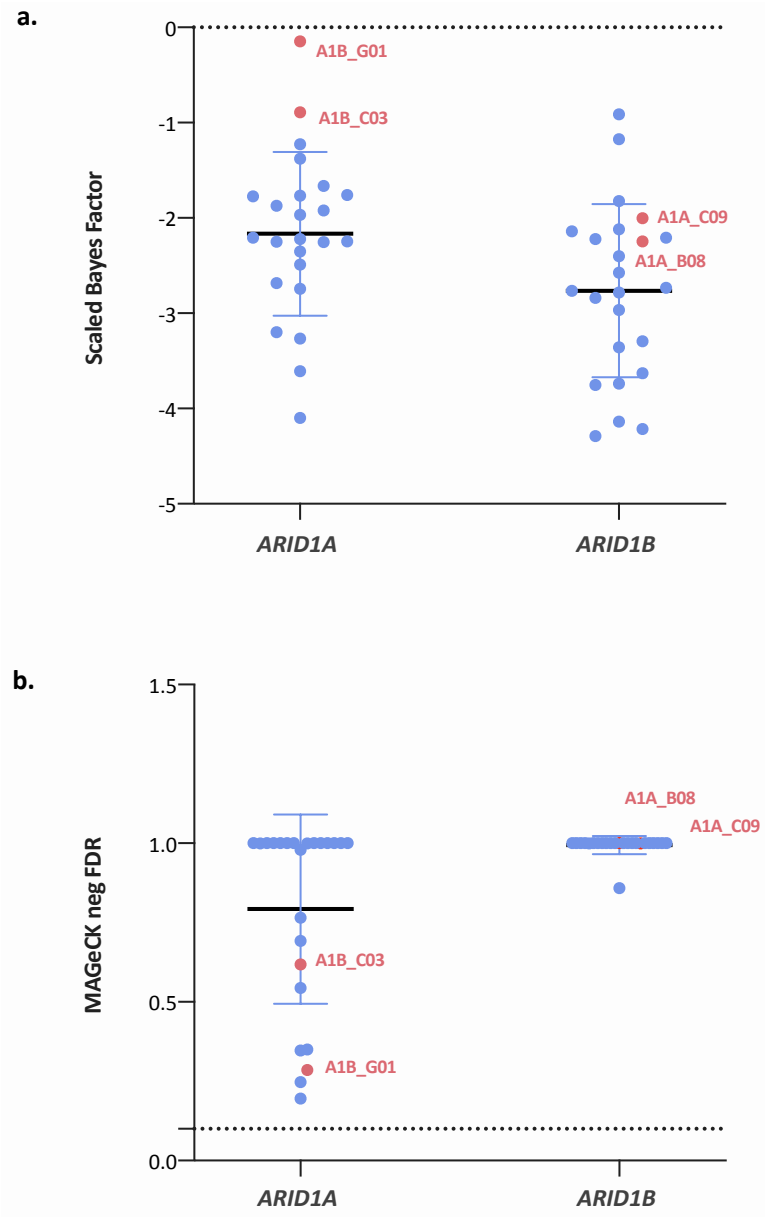


Figure 4.1. Gene essentiality scores for *ARID1A* and *ARID1B* in iPSC screens. Scaled BFs **(a)** and MAGeCK negative FDR values **(b)** are shown for *ARID1A* and *ARID1B* in the 24 iPSC screens (21 cell lines total with 3 replicates of the parental and 2 replicates of the *TP53* KO). The *ARID1A* 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.

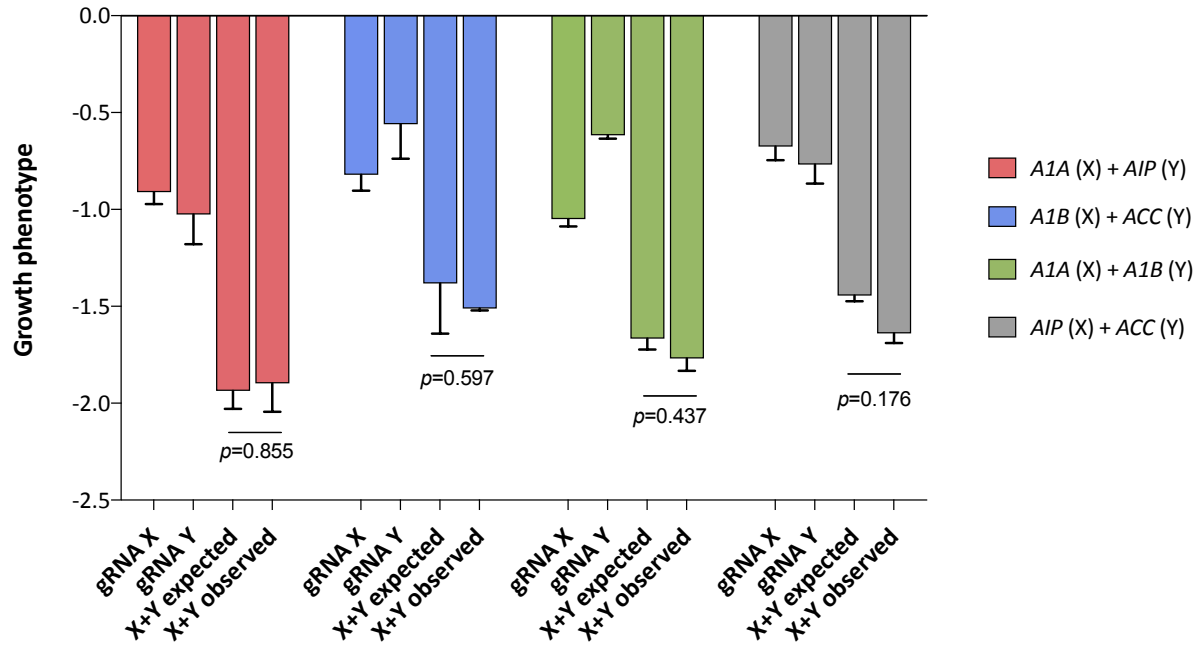


Figure 4.2. Validation of *ARID1A/ARID1B* SLI using a double gRNA strategy. BOB-Cas9 cells were transduced with two gRNA lentiviruses simultaneously (one expressing BFP, one expressing mCherry) to give four populations. Cells were analysed by flow cytometry on day 2 and day 14. To calculate the growth 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).

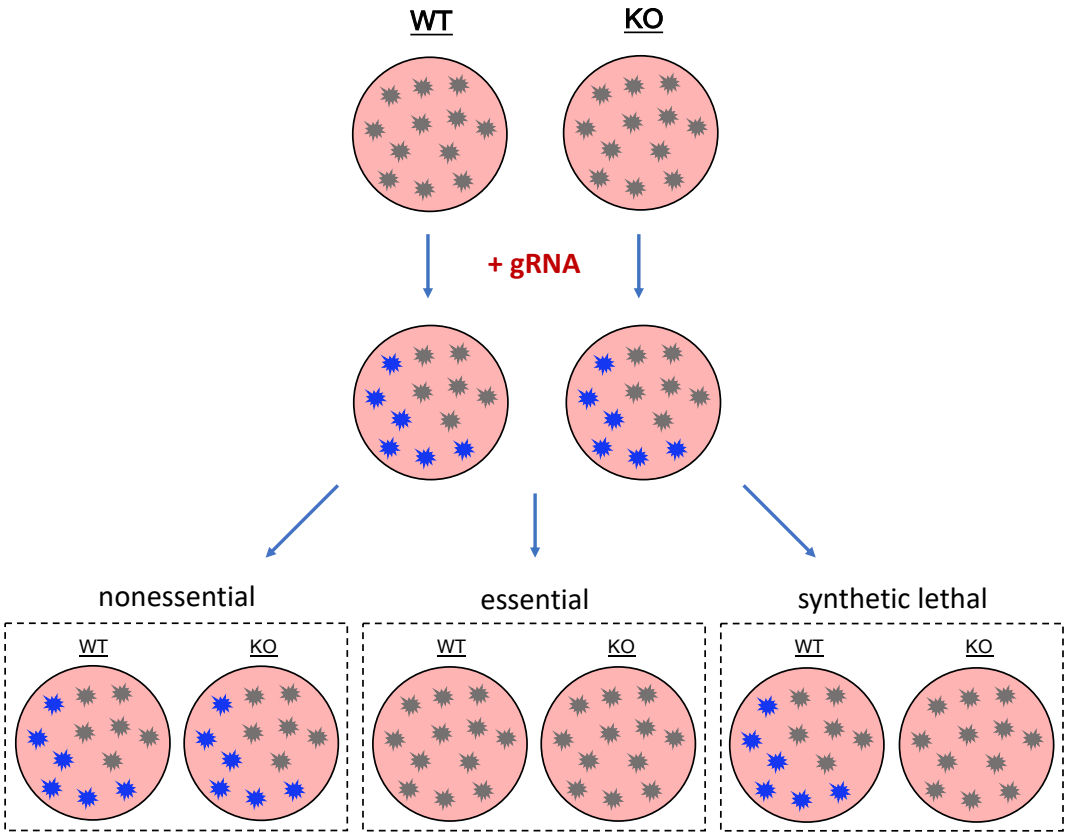


Figure 4.3. Strategy for single gRNA SLI validation. WT and KO stable Cas9 cells were transduced with a lentivirus expressing a gRNA and a BFP or mCherry marker. Expression was measured by flow cytometry on day 2 and day 14. Three outcomes were possible: if the gRNA targeted a nonessential gene, fluorescence 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.

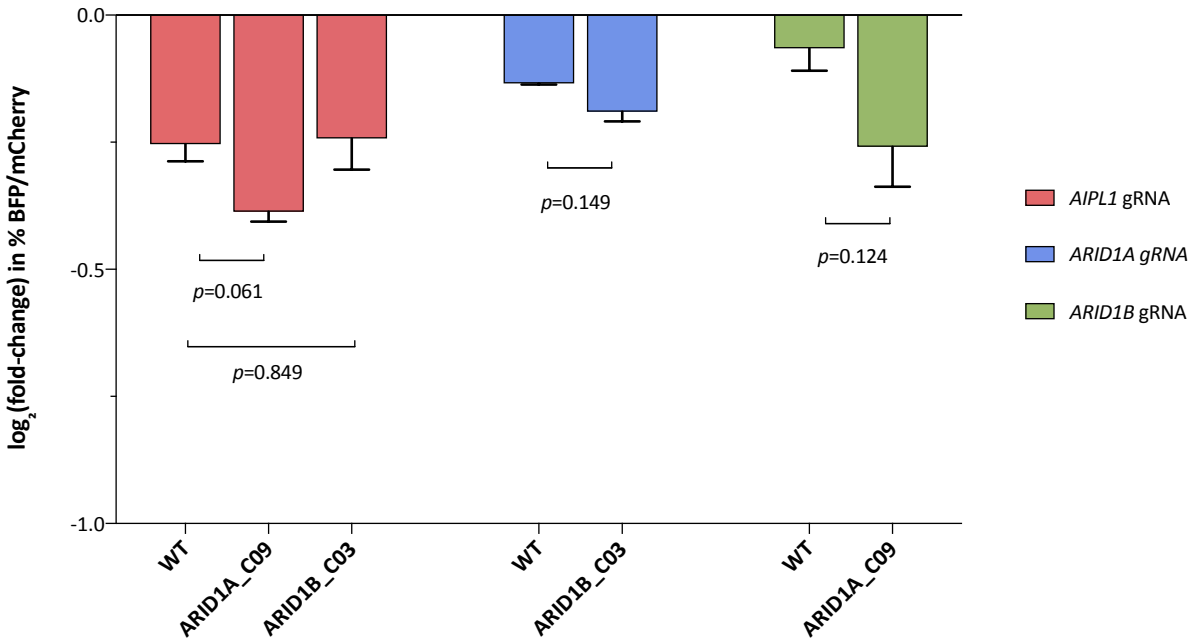


Figure 4.4. Validation of *ARID1A/ARID1B* SLI using a single gRNA strategy. WT BOB-Cas9, ARID1A_C09-Cas9 and ARID1B_C03-Cas9 were transduced with lentiviral gRNAs targeting *AIPL1* (BFP-tagged), *ARID1A* (BFP-tagged) and *ARID1B* (mCherry-tagged). All gRNAs were transduced in separate wells. BFP/mCherry 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.

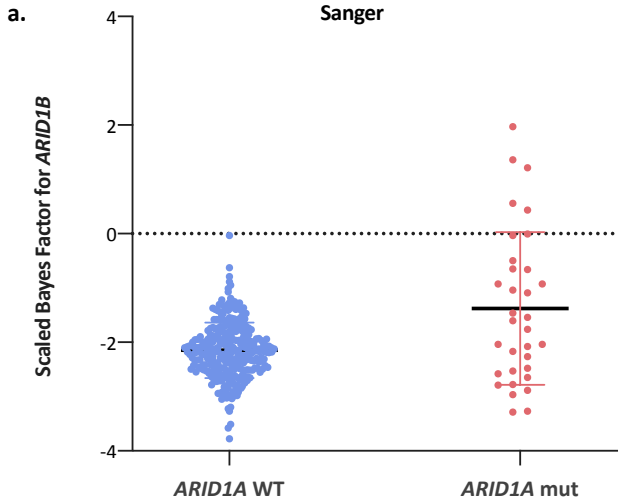
To understand this potential inconsistency in synthetic lethality between *ARID1A* and *ARID1B*, it is important to consider the similarities and differences between the subunits. The exact roles of ARID1A and ARID1B in tumourigenesis are still unclear, as are the reasons for the difference in mutation 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.

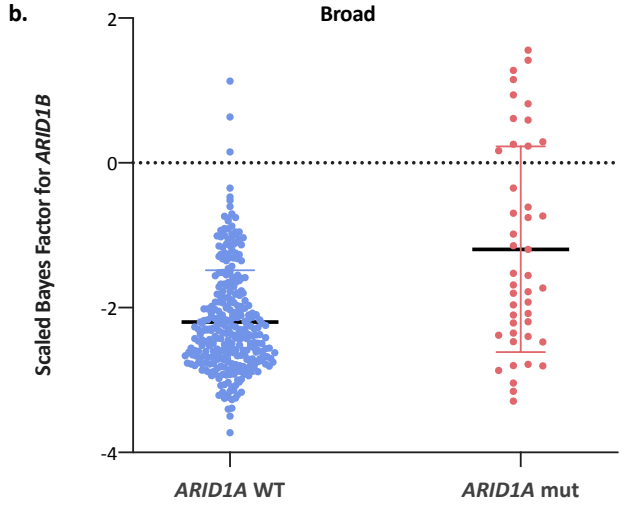
Therefore, all *ARID1A*-mutant cell types may have originally been composed of ARID1A-BAF complexes and *ARID1B*-mutants composed of ARID1B-BAF complexes. If true, these different compositions may cause variation in the functional dynamics between complexes and could explain the potential unidirectional 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**

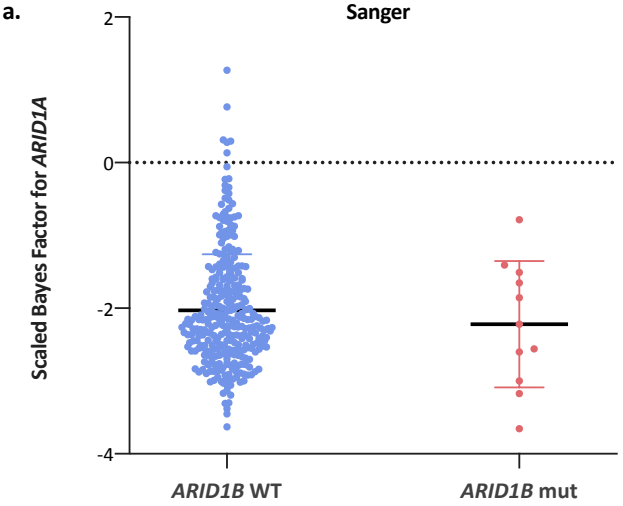


d.

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

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

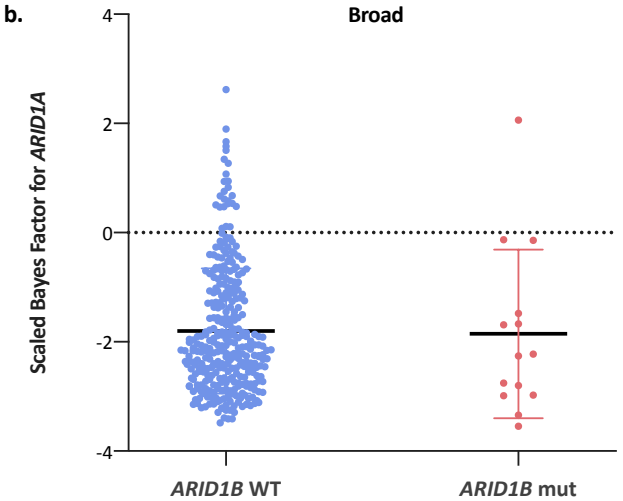
Figure 4.5. *ARID1B* essentiality in *ARID1A*-mutant cancer cell lines. BAGELR was used to process data from 324 cancer cell lines screened by the Sanger Institute and 342 lines screened by the Broad Institute. Cell lines were grouped into *ARID1A* mutant (lines containing a frameshift indel or nonsense 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).

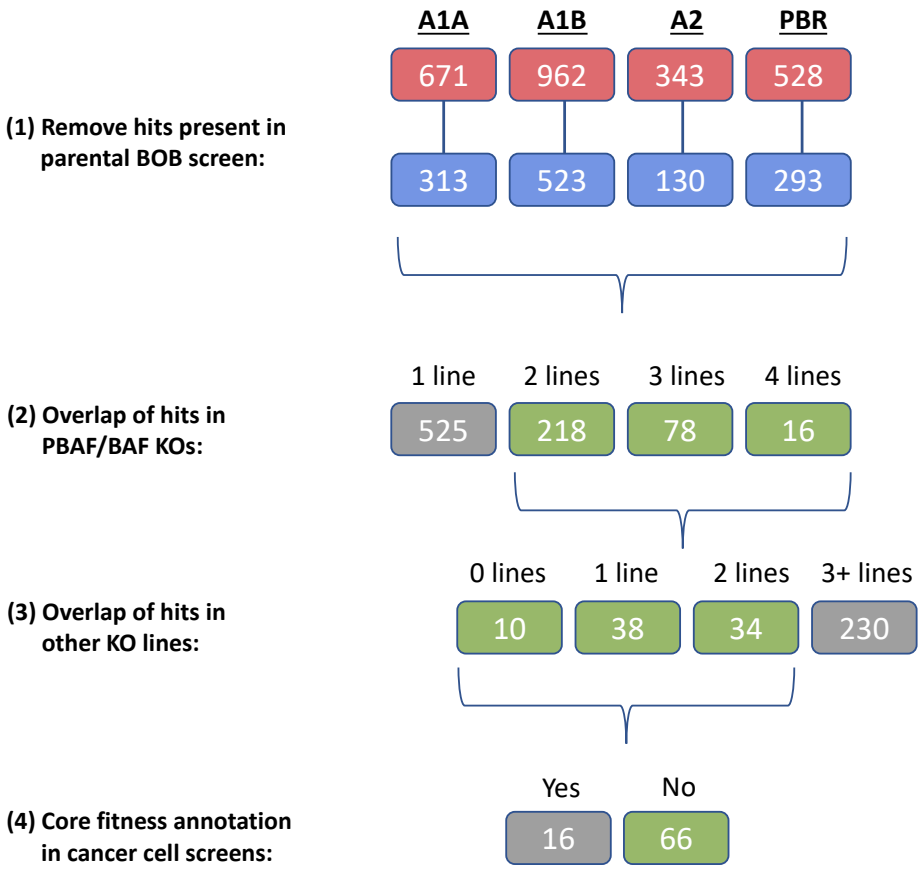


Figure 4.7. Filtering of PBAF/BAF gene SLIs for validation. All genes with a scaled BF > 0 were noted for ARID1A_C09 (A1A), ARID1B_C03 (A1B), ARID2_C11 (A2), and PBRM1_F09 (PBR). These went through four filtering steps. **(1)** Any genes that also had a scaled BF > 0 in the parental BOB screen were removed. **(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).

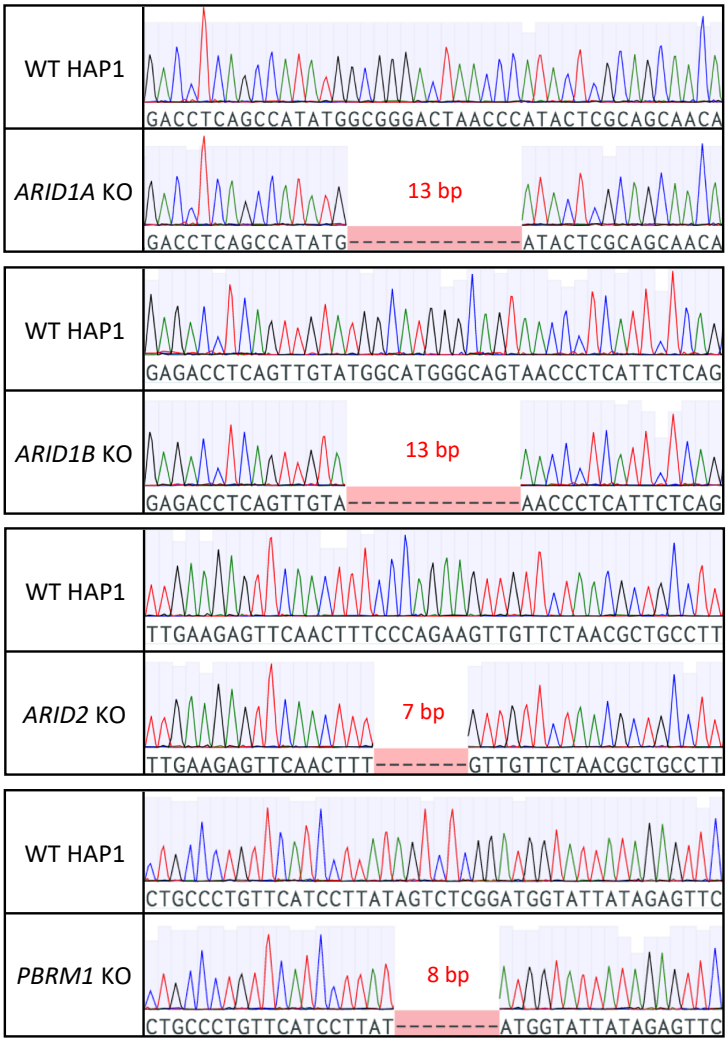
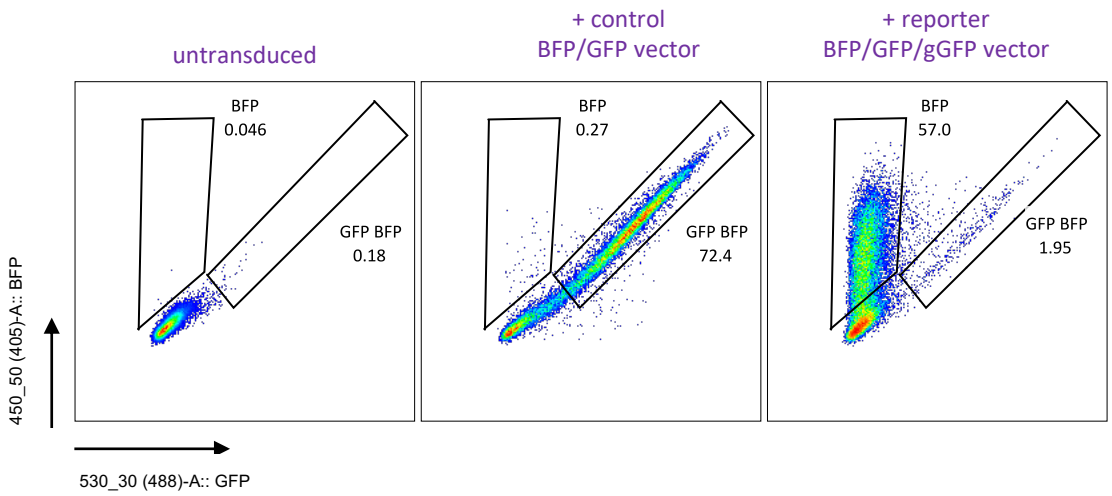


Figure 4.8. Genotypes of PBAF/BAF knockout HAP1 lines. PCR and Sanger sequencing were performed on the edited region in *ARID1A, ARID1B, ARID2* and *PBRM1* KO HAP1 cell lines. For each gene, WT HAP1 cells were also sequenced for comparison. The red areas indicate frameshift deletions. Primers used for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

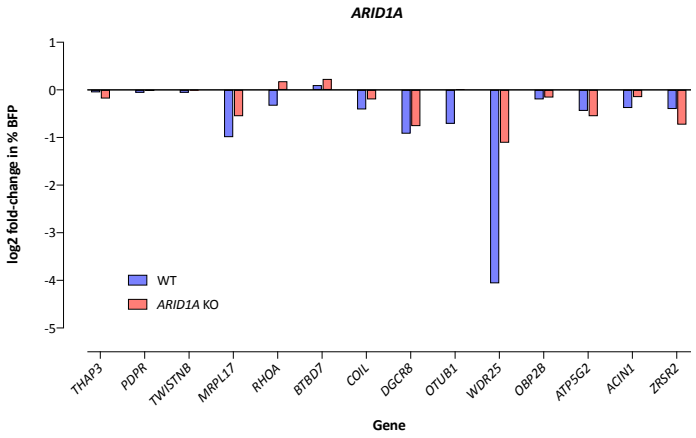
Figure 4.9. Cas9 activity in HAP1 cell lines. After blasticidin selection, Cas9-expressing WT and KO HAP1 cells were transduced with a control vector (BFP/GFP) or a reporter vector (BFP/GFP/gGFP). Fluorescence was measured by flow cytometry at 3 days post-transduction. **(a)** Untransduced cells and 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))**.**

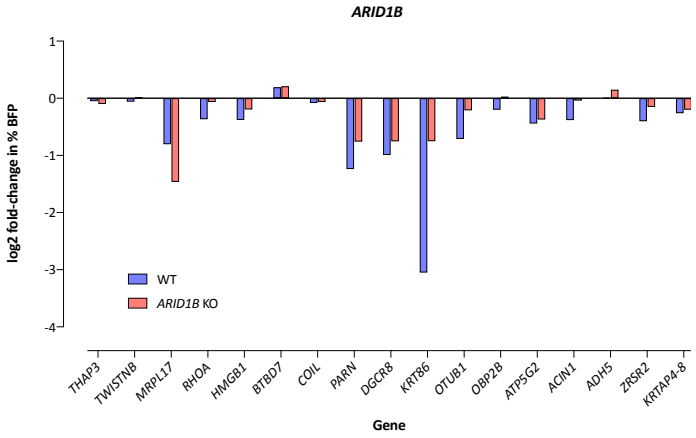
4.5.1 Competition assay in HAP1 cells

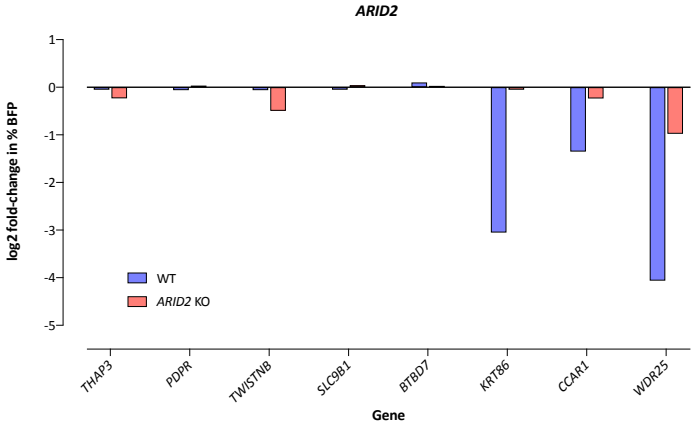
In parallel with validation in iPSCs, the fluorescence-based competition assay was also performed in the HAP1 cell lines using the same gRNAs (as described in Section 4.4 and 7.15.4-5). WT and KO HAP1 cell lines were transduced with a single gRNA per well in a 12-well plate. As with the iPSC validation, 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

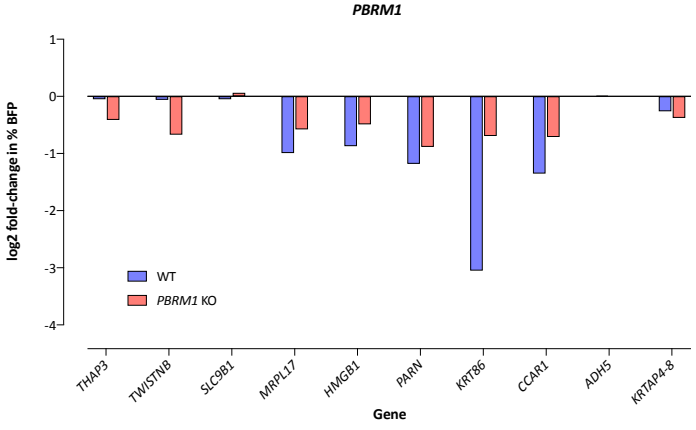

PBRM1
log2 fold-change in % BFP
1
0
-1
-2
-3
-4
WT
PBRM1 KO
THAP3
TWISTNB
SLC9B1
MRPL17
HMGB1
PARN
KRT86
CCAR1
ADH5
KRTAP4-8
Gene

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.

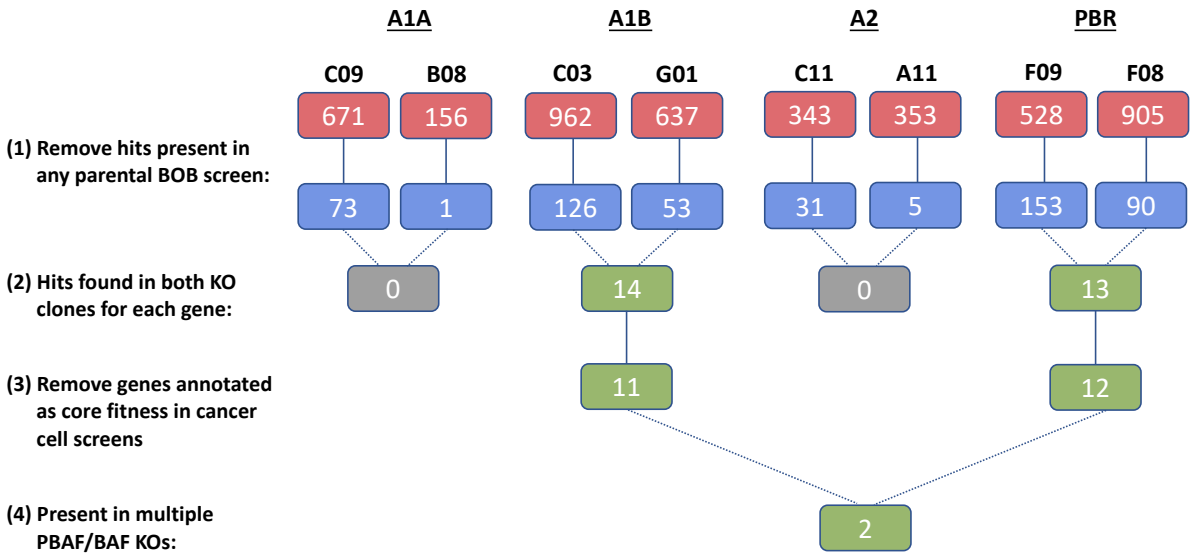


Figure 4.11. Filtering of PBAF/BAF gene SLIs using additional data. All genes with a scaled BF >0 were noted for ARID1A_C09/B08, ARID1B_C03/G01, ARID2_C11/A11, and PBRM1_F09/F08. These went through four filtering steps. **(1)** Genes that also had a scaled BF > 0 in any of the three parental BOB screens 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 = 2.23×10^{-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.