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]

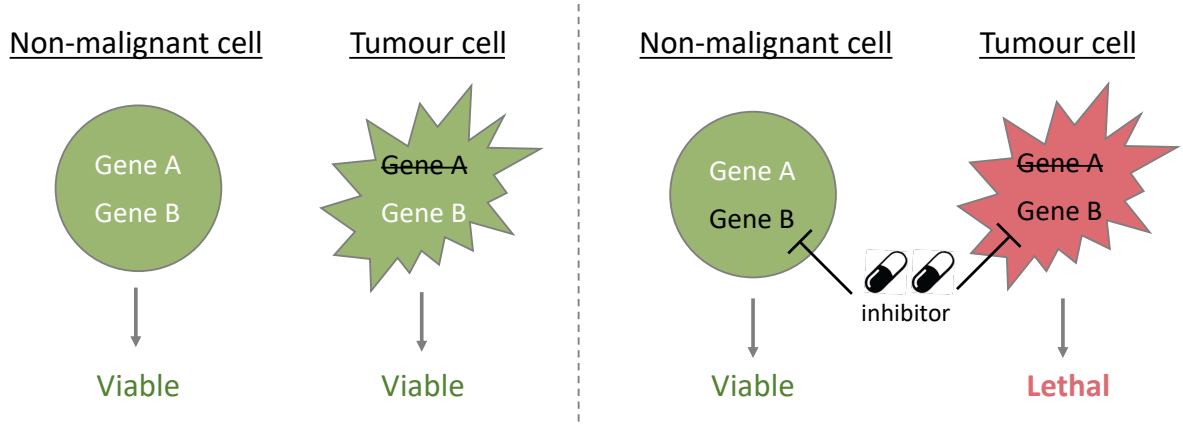

Figure 1.1. Exploiting synthetic lethality for cancer treatment. Tumour cells often lose the function of a TSG (gene A), but loss of this gene alone does not affect viability. If a synthetic lethal partner of gene A is known (gene B), then pharmacological inhibition of this gene should induce death specifically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

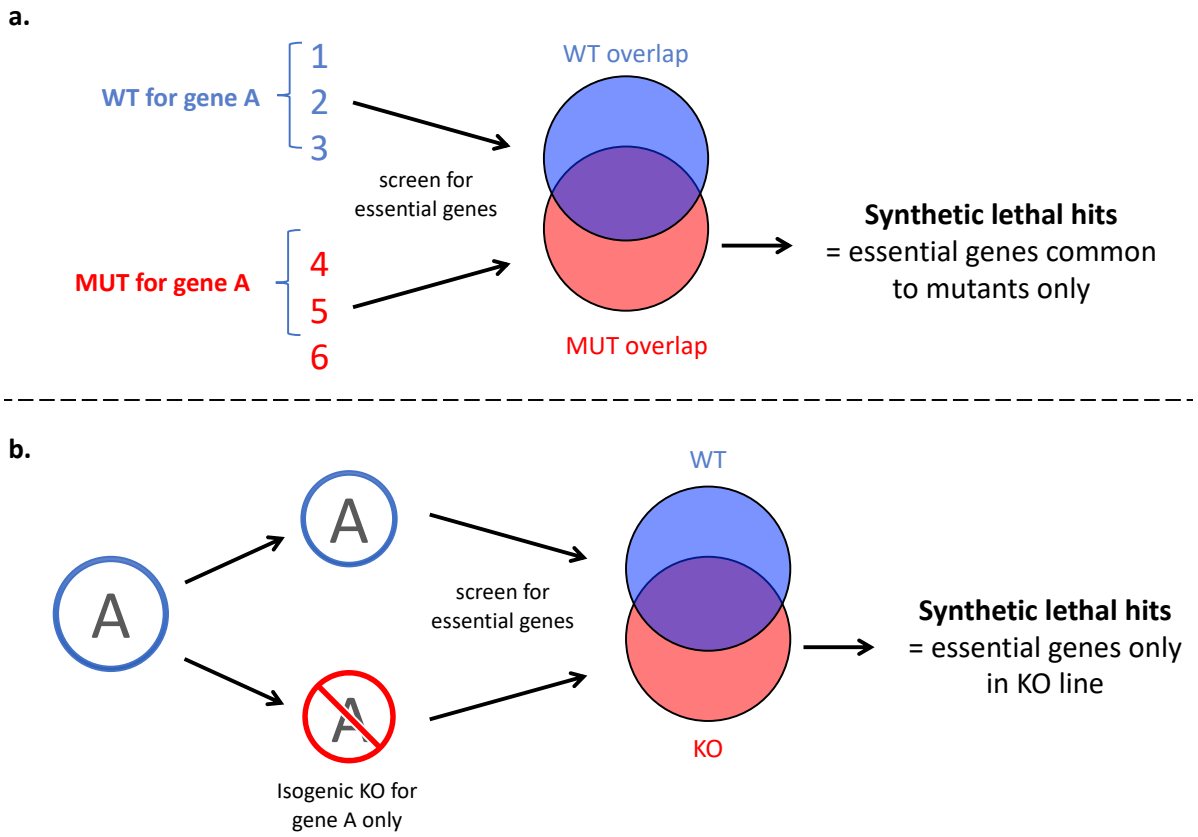

Figure 1.2. Approaches to synthetic lethality screening in human cell lines. a) Cell lines can be grouped into those that are WT or mutant (MUT) for a gene of interest. Screening is performed to identify genes that are essential for cell fitness/survival in each line. The essential genes that overlap 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]

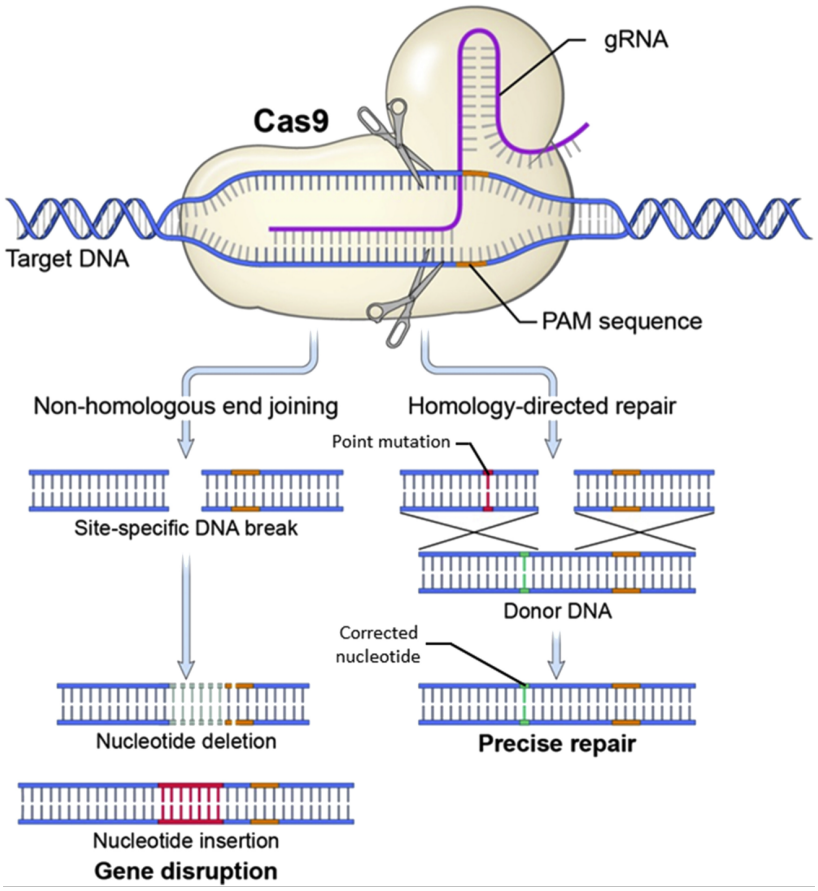

Figure 1.3. CRISPR/Cas9 genome editing in mammalian cells. Cas9 forms a complex with a gRNA and is guided to a specific target DNA region. The gRNA binds to a homologous 20 nucleotide DNA sequence positioned immediately downstream of an 'NGG' PAM sequence. Cas9 cleaves the DNA producing a DSB. The cell 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.

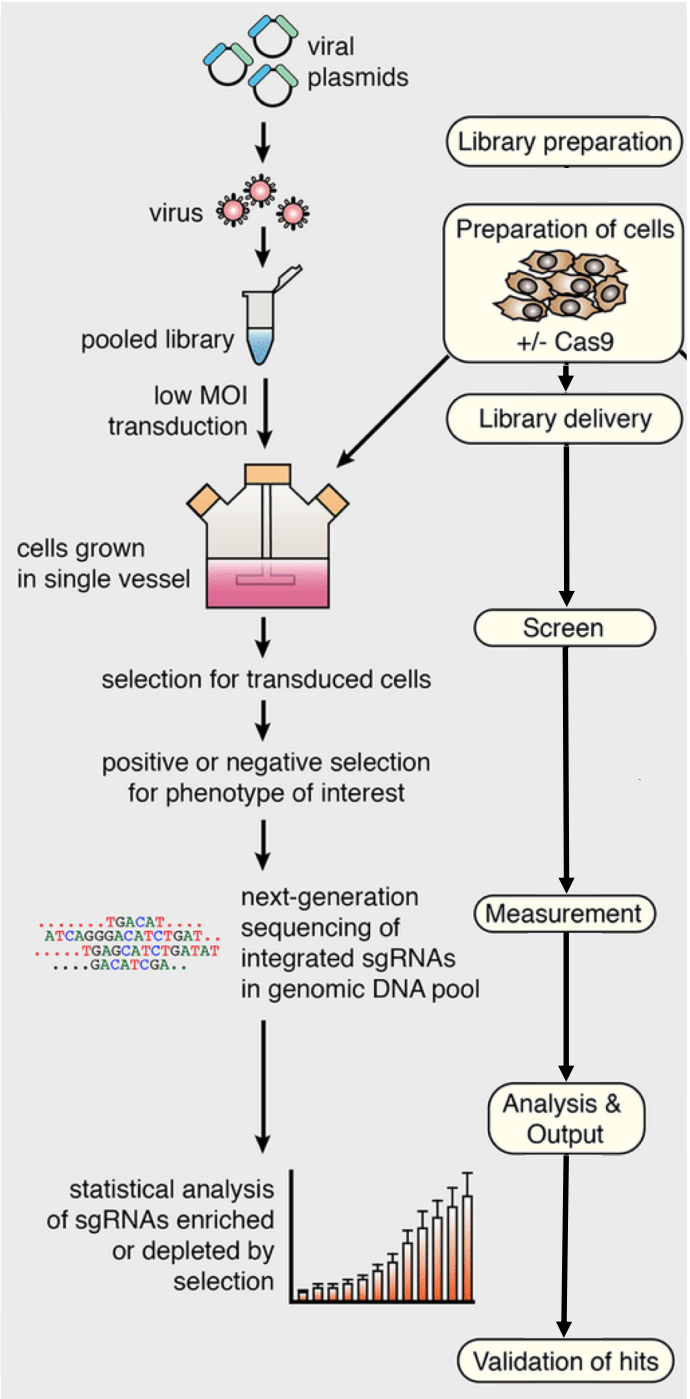

Figure 1.4. Strategy for pooled CRISPR/Cas9 screening. A viral library containing a pool of gRNAs is prepared. If Cas9 is not encoded by the library backbone, cells must be engineered to express Cas9. Cells are transduced with the pooled gRNA virus at a low multiplicity of infection (MOI). Library backbones 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]

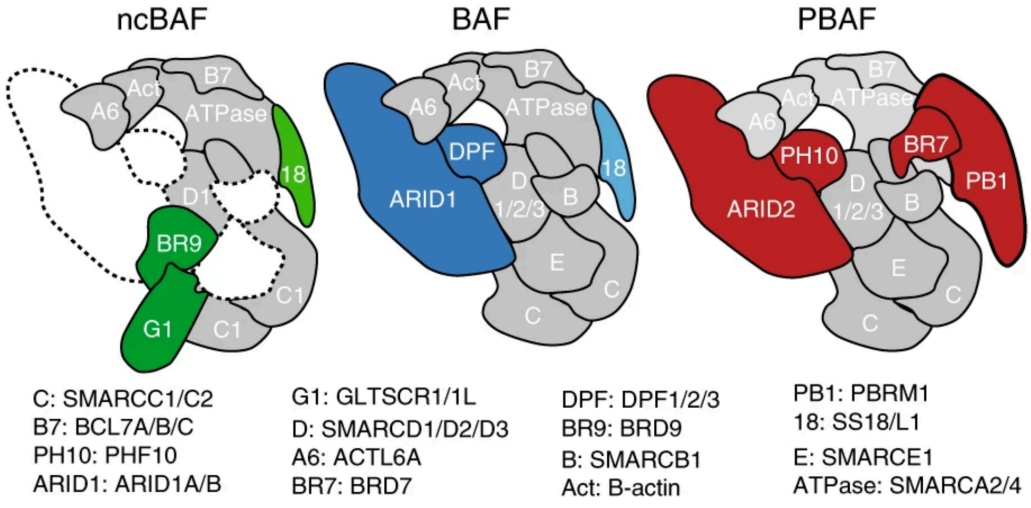

Figure 1.5. Composition of mammalian PBAF/BAF/ncBAF complexes. Schematic showing the subunits present in the mammalian ncBAF, BAF and PBAF complexes. Coloured subunits represent those that are specific to a single complex (ncBAF = green, BAF = blue, PBAF = red). Grey subunits are those that are shared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