Graeme Benstead-Hume[1] / Sarah K. Wooller[1] / Frances M.G. Pearl[1]

Computational Approaches to Identify Genetic Interactions for Cancer Therapeutics

[1] School of Life Sciences, University of Sussex, Brighton, UK, E-mail: f.pearl@sussex.ac.uk

Abstract:
The development of improved cancer therapies is frequently cited as an urgent unmet medical need. Here we describe how genetic interactions are being therapeutically exploited to identify novel targeted treatments for cancer. We discuss the current methodologies that use 'omics data to identify genetic interactions, in particular focusing on synthetic sickness lethality (SSL) and synthetic dosage lethality (SDL). We describe the experimental and computational approaches undertaken both in humans and model organisms to identify these interactions. Finally we discuss some of the identified targets with licensed drugs, inhibitors in clinical trials or with compounds under development.

Keywords: genetic interactions, synthetic sickness lethality, synthetic dosage lethality, oncogene, tumour suppressor

DOI: 10.1515/jib-2017-0027

Received: April 4, 2017; **Revised:** July 28, 2017; **Accepted:** August 10, 2017

1 Introduction

Cancer's genetic origins present a range of challenges for cancer drug discovery including issues related to target selectivity and the development of resistance. In this review we discuss how the search for new targeted therapies has led to the exploitation of genetic interactions, in particular synthetic sickness lethality (SSL) and synthetic dosage lethality (SDL), to successfully identify novel drug targets and therapeutic strategies.

We describe the experimental and computational approaches undertaken to identify, predict and validate genetic interactions. We review the most prominent experimental techniques employed for identifying genetic interactions, the challenges of using these techniques for the larger scale screenings required for studying human interactions, and how *in-silico* models can be used to mitigate some of these challenges. We broadly categorise the computational approaches into techniques that employ biological network data, those that use evolutionary data, and those that take an integrated approach, describing the notable models in each case.

We explore the current landscape of therapeutic SSLs and SDLs that could have utility in the treatment of cancer. We discuss a range of targets that have licensed drugs, inhibitors in clinical trials, or compounds under development. Finally we highlight some of the publicly available genetic interaction data that can be utilised to support the drug discovery process.

2 Targeted Therapies

Cancer is a genetic disease that develops as a result of a number of mutational events caused by endogenous and exogenous processes. The resulting mutations enable a cancer cell to gain a selective advantage over healthy cells, often resulting in uncontrolled proliferation and ultimately metastasis of a cancer [1], [2]. Cancer therapies must by necessity attack the aberrant cells once a tumour is discovered. However, established chemotherapy regimes often affect targets shared by normal and cancer cells and often kill "healthy" but rapidly dividing cells. This leads to significant damage in unintended targets resulting in the trademark side-effects of cancer therapy such as gastrointestinal upset and hair-loss [3].

The therapeutic index (TI) is a comparison of the amount of a therapeutic agent that causes the therapeutic effect to the amount that causes toxicity. Standard chemotherapies often have a low TI due to the challenge presented by selectively targeting cancer cells whilst sparing normal cells [4]. Furthermore, due to cancer cells' predisposition to acquire mutations, a drug that seems effective at the outset of therapy may well be rendered

Frances M.G. Pearl is the corresponding author.

ineffective if even a single cell, and its resulting daughters, gain resistance to that compound [5]. In response to these challenges a number of targeted therapies designed to increase TI are in development or have in some cases been approved [6].

Of the 154 cancer drugs that are licensed by the FDA, 85 are new, targeted therapies, often targeting the genes that directly drive cancer [6]. These driver genes can be broadly classified as oncogenes or as tumour suppressors. When mutated, the protein products of oncogenes show an increase in activity, or a gain or change of function (GOF) that result in tumorigenesis. Conversely in tumour suppressors, mutations (or epigenetic silencing) result in the loss of function (LOF) of the protein product.

Many targeted anticancer drugs work by directly inhibiting activated oncogenes, particularly proteins that are nuclear receptors or those that contain protein kinase domains [7], [8], [9]. Dabrafenib, which has been approved for the treatment of late-stage melanoma, targets the constitutively activated kinase oncogene BRAF V600E. Whilst gefitinib and erlotinib, licensed for the treatment of lung cancer, targets the EGFR tyrosine kinase [10], [11], [12], [13].

A substantively different approach is needed to provide therapies aimed at controlling the damage done by inactivated tumour suppressor genes. It is not usually feasible to repair the protein products of these genes particularly if they are inactivated by truncation, although there are on going attempts to reactivate or restore function to a small subset of p53 missense mutant proteins. These attempts to develop drugs to reactivate p53 have led way to another class of therapy, anti-inhibition. Inhibitors of MDM2, a negative regulator of p53, have shown some promise in restoring function in the p53 pathway including apoptosis and this can lead to tumour regression. A number of compounds related to nutlin-3a, a class of small molecule MDM2 inhibitors, are currently in phase I or II trials [14], [15].

2.1 Genetic Interactions

When the mutation or loss of one gene changes the impact of mutating or losing a second gene, the two genes are said to have a genetic interaction. This phenomenon can reveal functional relationships between genes and pathways [16]. Two types of genetic interaction are of particular interest in the field of cancer drug development; synthetic sensitivity lethality (SSL) and synthetic dosage lethality (SDL) (described below). Here we describe how the identification of these genetic interactions is being used to guide therapeutic strategies for the treatment of cancer.

A natural redundancy of function in our cells allows for a number of otherwise essential pathways to be disrupted by mutations whilst allowing the cell to remain viable. In some cases these disruptions can lead to impaired function of important cell maintenance or regulatory pathways leading to an increased occurrence of mutations or increased proliferation. These mutations are often found in tumour samples as they can often confer an increased fitness over normal cells.

This redundancy gives rise to the possibility of synthetic sickness lethality (SSL), where individuals in a pair (or more) of genes can be disrupted without affecting cell viability whilst disruptions in both genes causes cell sickness or death. Two genes are said to be synthetic lethal when concurrent deleterious mutations or complete deletion of both leads to the death of the host cell whilst a mutation or deletion in either alone leaves the cell viable [17]. Extensions of this concept, 'synthetic sensitive'' or 'synthetic sick' interactions, are similar genetic interactions except that they impair cellular fitness without necessarily killing the cell. Conversely, synthetic dosage lethality (SDL) interactions occur when an over-expression of gene A and a loss of function in gene B results in cell death (see Figure 1). On occasion these terms are used incorrectly in the literature with the term synthetic lethality used to describe a variety of genetic interactions.

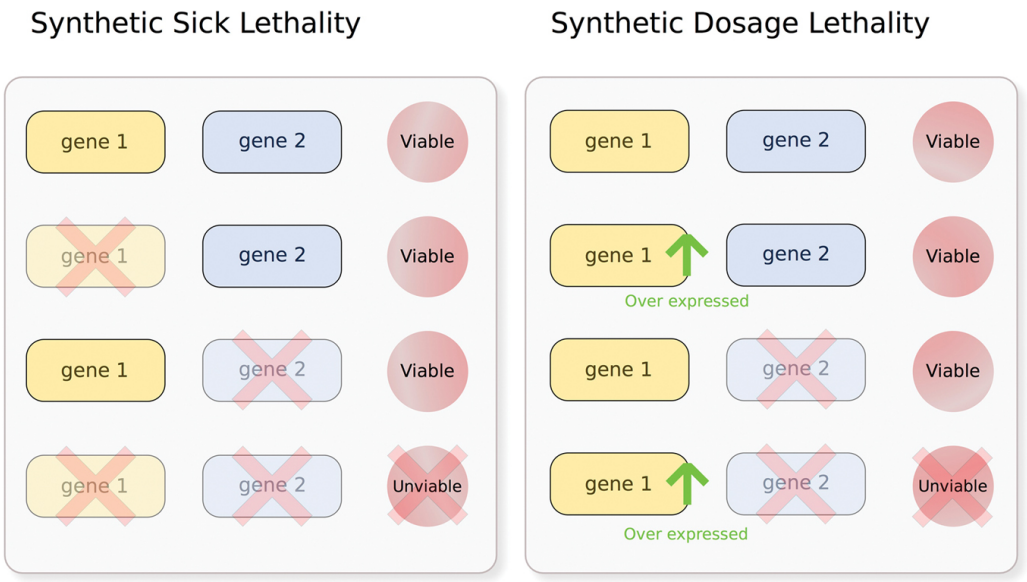

Figure 1: Schematic illustration of synthetic sickness lethality and synthetic dosage lethality.

2.2 Using Genetic Interactions as a Therapeutic Strategy

To exploit genetic interactions therapeutically, the genetic defects in an affected pathway must be combined with a pharmacologically induced defect in a compensating pathway. Synthetic lethality is well suited for targeting deactivated tumour suppressors [18]. SSL causes cell death as a result of one gene being genetically inactivated by mutation (LOF, the tumour suppressor) and another being inactivated by a drug target. While synthetic dosage lethal interaction can be used for targeting cancer cells with over-expressed oncogenes [18]. SDL causes cell death as a result of one gene being genetically activated (GOF, the oncogene) and another being inactivated (LOF, the drug target).

Targeted therapies that exploit these genetic interactions may provide a significantly improved therapeutic index compared to standard chemotherapies [19].

3 Methods that Identify Genetic Interactions

Although there are some insights into where SSL interactions are likely to occur, for example Matteo et al. [20] found an enrichment of SSL interactions between recessive cancer genes and their functional paralogues, identifying SSL interactions is a hard problem. Due to experimental limitations not many SSL interactions in humans have been published, but more is known about those in model organisms.

Approximately 20 % of genes in *Saccharomyces cerevisiae* (*S. cerevisiae*) are essential [21] which leaves the others to have the potential to exhibit genetic interactions (see Figure 2). Systematic double-knockout screens on large subsets of genes is *S. cerevisiae* and *Caenorhabditis elegans* (*C. elegans*) suggest that, on average, 0.5 % of tested gene pairs are synthetic sick or synthetic lethal, and that many SSL interactions involve more than two genes [22]. The result is a combinatorial problem for the traditional screening of all possible interactions. This and our limited data on these molecular networks prevents easy, reliable systematic prediction of SSL interactions [22]. To compound this problem some mutations that occur later in the evolution of cancer may be tolerated due to earlier mutations. This network of interactions may prove extremely complicated though we may find that pathways activated early in tumour progression are likely to make better targets for analysis [23].

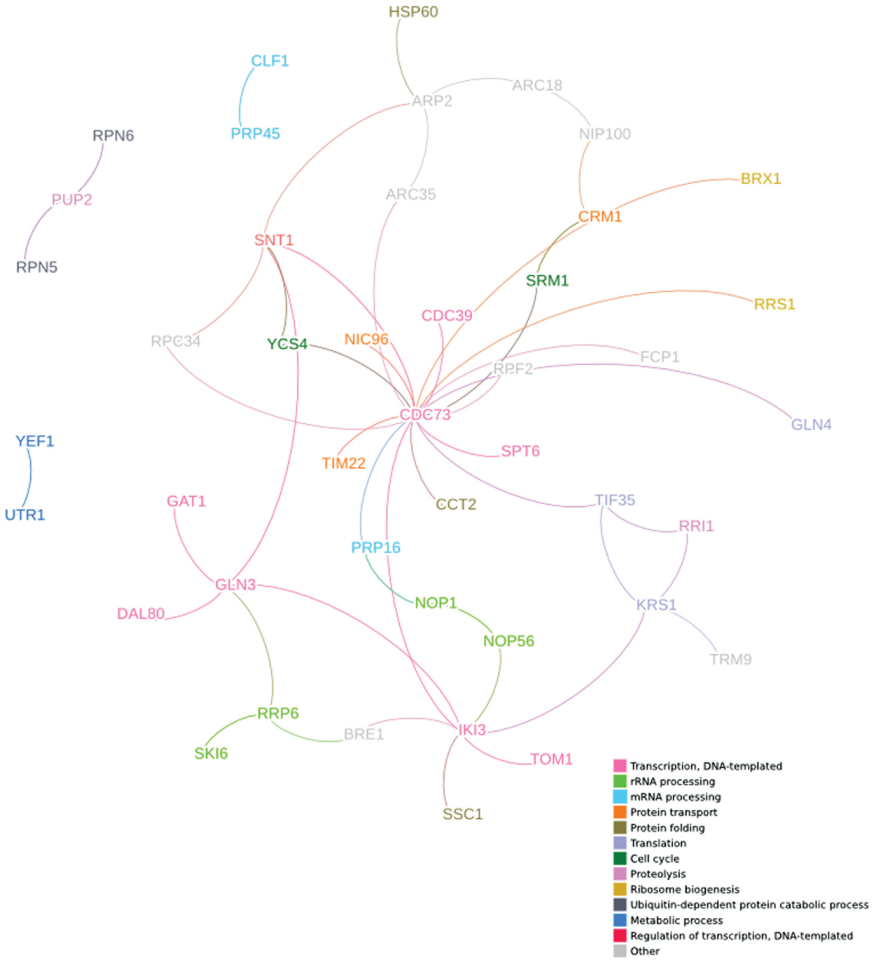

Figure 2: A genetic interaction network of *S. cerevisiae* DDR genes coloured by GO terms using genetic interaction data collected from BioGRID and filtered for orthologues of known human DDR genes. GO terms were sourced from the Gene Ontology Consortium and filtered for functional terms only. The most popular overall GO term was chosen for genes with multiple annotations.

3.1 Experimental Approaches to Identify Genetic Interactions

Synthetic lethality was first described in 1922 by Calvin Bridges in a study crossing *Drosophila melanogaster* and later named by Theodore Dobzhansky, this time crossing *Drosophila pseudoobscura*, in 1946 [24]. Similar to these early experiments contemporary SSL studies have classically focused on crossing eukaryotic model organisms with increasing sophisticated techniques allowing researchers to mutate and mate hybrid genomes and screen using gene silencing techniques such as RNA interference (RNAi).

More recently high throughput approaches to finding genetic interactions in model organisms have been developed based broadly around three distinct platforms; synthetic genetic array (SGA) [25], diploid based synthetic analysis on microarrays (dSLAM) [26] and epistatic miniarray profiles (E-MAP) [27]. Tong et al.'s SGA assay in *S. cerevisiae* uses a yeast strain with a single disabled gene and mates it with an array of yeast strains each with an individual deletion resulting in approximately 4,700 mutation pairs with varying viability. These techniques were further refined in Ooi et al.'s [28] SLAM and again in Pan et al.'s [26] dSLAM. SLAM generates ordered arrays of double yeast knockout mutant sets (YKO) where the query mutation is introduced by integrative transformation rather than mating, and a microarray readout is used to produce a ranked list of candidate genetic interaction genes. In dSLAM, a pool of all heterozygous deletion diploids is transformed en masse with a single query gene disruption construct after which single- and double-mutant haploid pools are derived by sporulation and differential selection. These techniques have been extended from *S. cerevisiae* to *Saccharomyces pombe* (*S. pombe*), *C. elegans* and *Escherichia coli* (*E. coli*) significantly increasing the quantity and quality of genetic data available. Collins et al.'s [27] E-MAP performed SGA on a subset of *S. cerevisiae* genes selected specifically from a pathway or functional grouping.

Although it was at one time hypothesised that SSL pairs could be conserved across species if both species shared orthologues for those respective genes [29], it has since been found that in many cases these SSL interac-

tions are not conserved between lower eukaryotes and their human orthologous equivalents. As such, though SSL data for model organisms can teach us a lot about gene function and pathway interaction, our search has been extended to human cell lines and those of phylogenetically similar organisms.

Classically SSL discovery in humans has been a 'hypothesis driven' process of predicting SSL interactions based on proven associations, often related to loss of particular cell cycle checkpoints or pathways related to those of known tumour suppressors, and subsequent clinical trials. However, with the increasing availability of genetically modified human cell lines and high throughput genetic screening methods that combine RNAi screens with libraries of small molecule inhibitors, an increasing number of human SSLs are being identified [30], [31], [32], [33], [34].

3.2 Computational Techniques Used to Predict Genetic Interactions

A systematic approach to inferring genetic interactions has become increasingly popular in the past decade. The ever-growing amount of screening data available has paved the way for more sophisticated computational techniques employing statistical and machine learning. These *in silico* models have proved significantly cheaper and faster to implement compared to traditional screening methods and have demonstrated impressive levels of accuracy when predicting genetic interactions.

These studies can be broadly classed by the type of parameter, or feature in the context of machine learning, used to train the model (see Table 1). The most prevalent parameter types include biological network data, gene ontology and expression level data, and orthology or evolutionary data. However, a number of studies use a combination of these data. Whilst this review has more emphasis on human SSL interactions we do discuss a number of studies focused on model organisms as much work on human genetic interactions has foundations in early work on lower eukaryotes (see Table 2).

3.2.1 Biological Network Data Approaches

A number of studies have employed a systems approach to predict SSL interactions using network parameters extracted from biological network data. These biological networks include data such as physical interactions and co-expression.

Early attempts to predict genetic interactions such as Wong et al. [35] utilised decision tree classifiers trained on biological network data including a number of topological network features derived from protein to protein interaction graphs, gene co-occurrence data and mRNA co-expression data. This study predicted 740 SSL interactions in 2,356 possible pairs in *S. cerevisiae* with a success rate of 0.31, a vast improvement on the 0.0056 success rates achieved by previous unguided approaches. This approach was extended by Zhong et al. [36] to predict interactions in *C. elegans*, an organism with relatively less available genetic interaction data, through orthology. By training a model using features from the relatively large datasets from yeast and fly models this study was able to predict interactions across species using logistic regression. Further attempts to predict genetic interactions in *S. cerevisiae* using biological networks followed as Paladugu et al. [37] extracted multiple features from protein–protein interaction networks, which were applied to a Support Vector Machine (SVM) classifier to predict new SSL interactions with sensitivity and and specificity exceeding 85%.

By employing random walks and decision tree classifiers on biological networks that include protein–protein interactions, GO interactions and existing known genetic interaction data, Chipman et al. [22] were able to predict synthetic lethal interactions at a true positive rate of 95 % against a false positive rate of 10 % in *S. cerevisiae* and a true positive rate of 95 against a false positive rate of 7 % in *C. elegans*. They noted that including experimentally validated non-interactions into training data significantly improved results for both organisms.

While the majority of preceding studies focused on supervised learning You et al. [38] performed semi-supervised learning on both the functional and topological properties of a functional gene network in *S. cerevisiae*. This network was a result of the integration of protein to protein interaction data along with protein complex and gene expression data and resulted in a maximum accuracy of true positive rate of 92 % against a false positive rate of 9 %.

Attempts to predict SSL interactions using expression data as a primary training parameter led to Bandyopadhyay et al.'s [39] SSLPred used regression on training data that mapped expression levels between genes with single deletion mutations to their corresponding SGA entries to predict SSL interactions in *S. cerevisiae*. Again using expression level data but this time to predict SSL in the context of somatic mutations in TP53 in humans, Wang et al. [40] selected a number of genes which encoded kinases that exhibited significantly higher expression in tumours with functional p53 somatic mutations than in tumours without. These pairs

were treated as potentially druggable synthetic lethal pairings for TP53 and many were confirmed via previous RNAi screenings.

To further improve results through an ensemble machine learning model Zheng et al. [41] developed MetaSL, a model boasting 17 features (11 similarity based features and 6 lethality based features) which was applied to 8 classifiers; random forest, J48 (a type of decision tree), Bayesian logistic regression, Bayesian network, PART (a rule-based classifier), RBFNetwork, bagging (bootstrap aggregating), and classification via regression. The predictions from these classifiers were aggregated yielding ROC AUC scores of 87.1 % on *S. cerevisiae* data. In another novel approach Zhang et al. [42] modelled influence propagation in signalling pathways employing values of phosphorylation levels between signalling proteins in a similar way to that of studies modelling influence across social media platforms. A number of reliable, novel human SSL pairs were predicted along with known interactions using this method.

Building on Zhong et al.'s attempt to predict SSL interactions using training data across species Jacunski et al. [43] developed SINaTRA (Species-INdependent TRAnslation) to compare orthologous gene pairs between *S. cerevisiae* and *S. pombe* along with their respective physical interaction data (including 4 pairwise parameters and 20 ontological features) to calculate what was termed connectivity homology to improve prediction of orthologous interactions. This approach achieved a reported ROC AUC score of 0.86 when predicting SSL interactions between the two studied yeast species. The model trained on yeast data was applied to predict 1,309 human SSL pairs with a reported false positive rate of 0.36 %.

3.2.2 Evolutionary Approaches

Although genetic interactions are not reliably conserved between species, with as little as ~23 % of the interactions conserved between *S. cerevisiae* and *S. pombe* [44], and even less conservation between lower and higher eukaryotes, a number of research groups have managed to use orthological and evolutionary data to infer SSL interactions in humans.

By integrating phylogenetic analysis and data including interactions from BioGRID for interactions, homology from Ensembl and NCBI and GO attributes from Gene Ontology, Conde-Pueyo et al. [45] reconstructed a phylogenetically-inferred SSL gene network for humans. The culmination of this study was to identify a number of genes related to cancer cells (ATM, NF1, FBXW7, MSH2, BUB1, ERCC2, BLM and MSH6) likely to be in therapeutically viable SSL interactions.

In a set of related studies researchers attempted to describe the mechanics of genetic interactions as a function of evolution and conservation across species. VanderSluis et al. [46] attempted to elucidate the evolutionary trajectories of duplicate genes through *S. cerevisiae* genetic interaction data and, as expected, found significant enrichment of genetic interactions between duplicate genes. Koch et al. [47] went on to describe how the rules governing genetic interactions are conserved across species. Using *S. cerevisiae* as a model they predicted the degree of genetic interaction for a number of *S. pombe* genes with high accuracy. That is they predicted how well a gene is connected in a genetic interaction network, or how many other genes a particular gene interacts with. Conserved features used to predict this degree of interaction included a quantitative measurement of single mutant fitness defects of the gene, multi-functionality, degree in a protein to protein interaction network and expression variation of the gene.

Lu et al. [48] also inferred human SSL pairs in human protein complexes by exploiting the evolutionary history of genes in parallel converging pathways in metabolism. This approach predicted around 250 novel SSL interactions 36 of which had a least one cancer related gene.

3.2.3 Integrative Data Approaches

As well as network base systems biology approaches and evolutionary methods a number of studies have also utilised the wealth of functional data such as mutation and copy number profiles, co-expression and functional relationships such as pathway data to predict synthetic lethal interactions.

In an early attempt to use a branch of natural language processing alongside biological data Pesquita et al. [49] focused on the semantic similarity of GO terms, annotations used as a proxy for functional pathways, to successfully compare the functionality of two genes. This technique was later used by Hoehndorf et al. [50] as a method for predicting genetic interactions in a number of model organisms. In 2011 Li et al. [51] attempted to use an expectation-maximisation algorithm on domain genetic interaction data to predict SSL interactions. It was reported that this approach was able to predict 17 novel SSL interaction in *S. cerevisiae* with probability > 0.9. This included the MYO4 – DYN1 pair with a probability of 0.9895. These interactions were further used to predict a number of compensatory pathways.

A number of algorithms have also been introduced that predict pairs of genes that would potentially exhibit genetic interactions using human cancer data directly by identifying and scoring of sets of genetic alterations where the mutations are mutually exclusive, i.e. if gene 1 and gene 2 are synthetically lethal there should be no samples where both these genes are switched off. Initially these models used scoring regimes to prioritise mutual exclusivity with no basis in statistics. However, the approaches have gradually been refined to improve statistical scoring of the results and to integrate different methods of identifying whether or not a gene has essentially been switched off. These include Recurrent Mutually Exclusive aberrations (RME) [52] that uses mutation and copy number variation (CNV) data from 145 glioblastoma samples from The Cancer Genome Atlas (TCGA) [53], and CoMet [54] that used mutation and CNV data from five TCGA studies. It is also worth noting the development of Mutex [55] that uses mutation, RNA expression and copy number variation (CNV) data from 3,299 samples from the TCGA, which also looks at the impact of false negative and positive alterations. The more sophisticated of these (CoMet) looks at small groups of mutually exclusive genes, using a hypergeometric distribution to work out the probability of getting at least as unexpected a result as that seen. Using similar methods Srihari et al. [56] analysed mutual exclusivity in copy number and gene expression data from four cancers to identify 718 genes that potentially share a SSL interaction with at least one of six DDR genes related to those cancers.

Another approach is the DAISY workflow [57] which uses three inference procedures to identify both SSL and SDL pairs using data from cell lines as well as from clinical samples; somatic copy number variation and mutation profiles, shRNA-based functional examination and pairwise gene co-expression. DAISY was applied to VHL, PARP1, MHS2 and KRAS and achieved an AUC score of 0.779 demonstrating a strong propensity (p-value < $1{\times}10^{-4}$) for predicting SSL pairs.

4 Cancer Therapies that Exploit Genetic Interactions

Many studies screening for genetic interactions have naturally focused on known cancer driver genes, specifically tumour suppressors, as promising targets for developing cancer therapies. Genetic interactions have been found in a wide range of cell pathways including cell cycle progression and apoptosis pathways.

Tumour suppressors that make part of DNA damage response pathways are prime candidates for synthetic lethal drug targets [58]. BRCA1 and BRCA2, both important in repair of double strand breaks, have been shown to share a synthetic lethal relationship with PARP [poly(ADP-ribose) polymerase], an important gene in single strand break repair. Cells deficient in either BRCA gene are extremely sensitive to PARP inhibitors presenting therapeutic opportunities [59]. Further studies systematically screening genes for sensitivity to PARP inhibitors identified a number of kinases whose inhibition strongly sensitised the host cell to PARP inhibitor, including cyclin-dependent kinase 5 (CDK5), MAPK12, PLK3, PNKP, STK22c and STK3 [60]. There are number of PARP inhibitors at different phases of trials, a notable example being olaparib (Lynparza–, Astrazeneca) which has already been approved by both the European commission and the US Food and Drug Administration for the treatment of patients with advanced ovarian cancer paired with BRCA mutations [61], [62]. As well as a treatment for ovarian cancer patients Mateo et al. [63] conducted trials for olaparib as a potential therapy for prostate cancer patients identified as having homozygous deletions, deleterious mutations or both in DNA-repair genes including BRCA1 or BRCA2, ATM, Fanconi's anaemia genes, and CHEK2. Of the patients available for evaluation 88 % responded to olaparib including all patients with BRCA loss leading to the conclusion that the drug led to a high response rate in prostate cancer patients with DNA-repair defects who were no longer responding to standard treatments. Recent PARP inhibitor based therapies include rucaparib which has also received FDA approval for patients with advanced ovarian cancer who suffer germline or somatic BRCA1 or BRCA2 mutations [64], [65] and Talazoparib which is showing promise in early trials for early-stage breast cancer patients with BRCA mutations even before any chemotherapy or surgery with all patients exhibiting a reduction in tumour size after 2 months [66].

Other published SSL with possible therapeutic potential include the SSL interaction between TP53 and the PI5P4K gene family where the PI5P4K kinases are essential for growth in the absence of p53 [67]. ARID1A, a chromatin remodeller with a high mutation rate across many cancer types shares a SSL interaction with the EZH2 methyltransferase in *ARID1A*-mutated ovarian cancer cells [68] and *ENO2* which selectively inhibits viability of *ENO1*-deleted glioblastoma cells [69].

PTEN, a gene associated with genomic stability, and APE1, important in DNA base excision repair, have been shown to share a SSL relationship with treatment of APE1 inhibitors in PTEN-deficient cells resulting in the induction of apoptosis [70]. ATR, an important DNA damage response gene has also been identified as potential synthetic lethal pair of ARID1A and a number of ATR inhibitors are in phase I trials as a potential therapy for ARID1A deficient tumours [71], [72].

While much work on genetic interactions as therapy targets has traditionally focused on SSL interactions, research has also been conducted into the SDL interactions of several potent oncogenes such as MYC and KRAS [73]. Members of the RAS superfamily are some of the most commonly activated cancer drivers [74] and showed some promise in early SDL research. These studies described a number of potential SDL pairs including an interaction between KRAS and CDK4 which offers potential opportunities in non-small cell lung carcinoma therapy [75]. Another systematic study of the RAS superfamily found a number of interactions with genes related to the cells mitotic functions including PLK1 [34]. Despite this early promise other therapies related to RAS such as the direct targeting of the RAS protein and immune checkpoint blockade have proved more effective and no promising new therapeutic approaches related to SDL interactions have been discovered to date [74], [76]. Improved screening through CRISPR-cas9-based techniques may provide further potential SDL interactors for mutant RAS genes in future studies [77].

Other genes with potential SDL interactions with KRAS include CDK1, part of the Cyclin-dependent kinase family with CDK4 which is mentioned above [78], TBK1, a serine/threonine kinase important in regulating inflammatory response [33] and GATA2, essential in regulating transcription [79].

5 Discussion

The performance of contemporary models used to predict SSL interactions is difficult to assess due to a lack of a gold standard source of human SSL pairs. This difficulty is compounded by the lack of a single extensive, curated repository of known human SSL pairs. Furthermore the actual number of potential human SSL pairs is so far unavailable proving another challenge when attempting to assess progress in the field.

In studies employing a CRISPR-Cas9-based screen of 18,166 human genes only 1,878 were essential, resulting in 16,288 non-essential genes (or as much as 90.8 % of the whole genome) each potentially part of at least one genetic interaction. Despite this large number of potential synthetic lethal interactions only 503 human gene pairs are classed as synthetic lethal or negative genetic in the BioGRID, a current primary source for curated validated human SSL pairs. There are many more predicted synthetic lethal pairs documented in sources such as SynLethDB which collates 19,952 predicted pairs sourced from *in-silico* predictions from tools such as Daisy (which counts 5,824 SSL pairs), shRNA screening experiments and literature via text-mining though the reliability of many of these observations is very hard to quantify.

It should be noted that identifying and then validating all possible synthetic lethal interactions is unlikely to be feasible. In addition, many genetic interactions may not be valuable in the context of cancer therapeutics. To date, the majority of research has been focused around cancer related genes with many of the studies outlined above focused on a small subset of interactions focused around notable cancer drivers such as BLM [45], TP53 [40], VHL, PARP1, MHS2 [57] amongst others. So far no meta-analysis has been completed on these disparate studies though this might be a good first step towards making SSL interaction data more coherent.

Table 1: List of useful data sources of training data for genetic interaction prediction.

Data type	Source	URL	References
Protein interactions	STRING	http://string-db.org/	[80]
Gene expression	Expression Atlas	https://www.ebi.ac.uk/gxa/home	[81]
	Gene Expression Omnibus	https://www.ncbi.nlm.nih.gov/geo/	[82]
Gene coexpression	CoxpresDB	http://coxpresdb.jp/	[83]
Gene ontology data	Gene Ontology Consortium	http://www.geneontology.org/	[84]
Somatic mutations	COSMIC	http://cancer.sanger.ac.uk/cosmic	[85]
	MOKCa	strubiol.icr.ac.uk/extra/mokca	[91]
Homology	Ensembl – comparative genomics	http://www.ensembl.org/info/genome/-compara/index.html	[86]
Cellular phosphorylation	Networkin	http://networkin.info/	[87]
Integrative – multiplatform data	The Cancer Genome Atlas (TCGA)	cancergenome.nih.gov	[53]
	The International Cancer Genome Consortium (ICGC)	icgc.org	[88]
	cBioPortal	www.cbioportal.org	[89]
	COSMIC	cancer.sanger.ac.uk/cosmic	[90]

[1] This table describes the types of data that have been used to predict SSL and SDL interactions, and highlights some of the internet resources and where these data are available.

Table 2: Synthetic lethal interaction data availability.

Source	Organism	Count of SSL pairs	URL/DOI	References
Biogrid	*H. sapiens*	503	https://thebiogrid.org	[92]
	S. cerevisiae	92,738		
	D. melanogaster	3		
	C. elegans	1237		
	S. Pombe	36,353		
SynLethDB	*H. sapiens*	19,952	http://histone.sce.ntu.edu.sg/SynLethDB	[93]
	S. cerevisiae	13,421		
	D. melanogaster	423		
	M. musculus	366		
	C. elegans	107		
The Cellmap	*S. cerevisiae*	1198	http://thecellmap.org	[94]
Flybase	*D. melanogaster*	9661	http://flybase.org/	[95]
Other studies	*S. cerevisiae*	100	10.1038/nature05649	[96]
	C. elegans	1246	10.1186/jbiol58	[97]

[1] This table describes the resources or studies that contain information about experimentally determined negative genetic and SSL interactions. It contains the name of the resource, the number of reported interactions, the web address and/or a reference for the resource or study.

The majority of standard chemotherapies exhibit a very low therapeutic index (TI). In these therapies the level of treatment that is likely to cause toxicity in a patient is not significantly higher than the level that offers a therapeutic effect. To improve this therapeutic index and, a result, the quality of life and prognosis of our cancer patients, our goal must be to discover targets that can be drugged to selectively affect cancer cells whilst leaving normal cells unharmed. By exploring and exploiting vulnerabilities presented by genetic interactions and, more specifically, SSL interactions in human cancer cells we may find ways to provide personalised care with both an increased therapeutic index and ultimately an improved prognosis for the cancer patient. While SSL interactions may present a unique opportunity in the fields of drug discovery and personalised cancer medicine the genome-wide identification of human SSL interactions comes with its own significant challenges. As well as the difficulty of propagating human cell lines for *in-vitro* screening the combinatorial nature of the problem means that around 200 million pairwise tests would be required to identify all possible pairs, an all but insurmountable experimental burden.

In response to these difficulties studies focussing on model organisms with far fewer genes and no ethical implications have resulted in the identification of a large quantity of SSL interactions. Unfortunately, based on these studies, it has been shown that SSL interactions are often not well conserved between species and even less so between higher and lower eukaryotes such as humans and yeast.

Though a number of unique human SSL interactions have been inferred using orthologous interactions many remain undiscovered and the search for SSL interactions opens to ever increasing quantities of multi-platform genomic data to develop improved systematic approaches for predicting potential SSL interactions utilising *in-silico* models.

Acknowledgments

This work was supported by Medical Research Council studentship [grant number MR/N50189X/1] (to G.B.-H).

Conflict of interest statement: Authors state no conflict of interest. All authors have read the journal's Publication ethics and publication malpractice statement available at the journal's website and hereby confirm that they comply with all its parts applicable to the present scientific work.

References

[1] Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000;100:57–70.

[2] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell. 2011;144:646–74.
[3] Coates A, Abraham S, Kaye SB, Sowerbutts T, Frewin C, Fox RM, et al. On the receiving end—patient perception of the side-effects of cancer chemotherapy. Eur J Cancer Clin Oncol. 1983;19:203–8.
[4] Muller PY, Milton MN. The determination and interpretation of the therapeutic index in drug development. Nat Rev Drug Discov. 2012;11:751–61.
[5] Holohan C, Van Schaeybroeck S, Longley DB, Johnston PG. Cancer drug resistance: an evolving paradigm. Nat Rev Cancer. 2013;13:714–26.
[6] Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, et al. A comprehensive map of molecular drug targets. Nat Rev Drug Discov. 2017;16:19–34.
[7] Iorio F, Knijnenburg T, Vis D, Bignell G, Menden M, Wessels L, et al. Abstract A44: a landscape of pharmacogenomic interactions in cancer. Clin Cancer Res. 2017;23:A44.
[8] Shawver LK, Slamon D, Ullrich A. Smart drugs: tyrosine kinase inhibitors in cancer therapy. Cancer Cell. 2002;1:117–23.
[9] Nguyen D-T, Mathias S, Bologa C, Brunak S, Fernandez N, Gaulton A, et al. Pharos: collating protein information to shed light on the druggable genome. Nucleic Acids Res. 2017;45:D995–1002.
[10] Thatcher N, Chang A, Parikh P, Rodrigues Pereira J, Ciuleanu T, von Pawel J, et al. Gefitinib plus best supportive care in previously treated patients with refractory advanced non-small-cell lung cancer: results from a randomised, placebo-controlled, multicentre study (Iressa Survival Evaluation in Lung Cancer). Lancet. 2005;366:1527–37.
[11] Shepherd FA, Rodrigues Pereira J, Ciuleanu T, Tan EH, Hirsh V, Thongprasert S, et al. Erlotinib in previously treated non-small-cell lung cancer. N Engl J Med. 2005;353:123–32.
[12] Stinchcombe TE, Socinski MA. Gefitinib in advanced non-small cell lung cancer: does it deserve a second chance?. Oncologist. 2008;13:933–44.
[13] Lindeman NI, Cagle PT, Beasley MB, Chitale DA, Dacic S, Giaccone G, et al. Molecular testing guideline for selection of lung cancer patients for EGFR and ALK tyrosine kinase inhibitors: guideline from the College of American Pathologists, International Association for the Study of Lung Cancer, and Association for Molecular Pathology. J Mol Diagn. 2013;15:415–53.
[14] Khoo KH, Verma CS, Lane DP. Drugging the p53 pathway: understanding the route to clinical efficacy. Nat Rev Drug Discov. 2014;13:314.
[15] Burgess A, Chia KM, Haupt S, Thomas D, Haupt Y, Lim E. Clinical overview of MDM2/X-targeted therapies. Front Oncol. 2016;6:7.
[16] Krause SA, Gray JV. The functional relationships underlying a synthetic genetic network. Commun Integr Biol. 2009;2:4–6.
[17] Hartwell LH, Szankasi P, Roberts CJ, Murray AW, Friend SH. Integrating genetic approaches into the discovery of anticancer drugs. Science. 1997;278:1064–8.
[18] Megchelenbrink W, Katzir R, Lu X, Ruppin E, Notebaart RA. Synthetic dosage lethality in the human metabolic network is highly predictive of tumor growth and cancer patient survival. Proc Natl Acad Sci USA. 2015;112:12217–22.
[19] McLornan DP, List A, Mufti GJ. Applying synthetic lethality for the selective targeting of cancer. N Engl J Med. 2014;371:1725–35.
[20] D'Antonio M, Guerra RF, Cereda M, Marchesi S, Montani F, Nicassio F, et al. Recessive cancer genes engage in negative genetic interactions with their functional paralogs. Cell Rep. 2013;5:1519–26.
[21] Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, et al. Global mapping of the yeast genetic interaction network. Science. 2004;303:808–13.
[22] Chipman KC, Singh AK. Predicting genetic interactions with random walks on biological networks. BMC Bioinform. 2009;10:17.
[23] Kaelin WG. The concept of synthetic lethality in the context of anticancer therapy. Nat Rev Cancer. 2005;5:689–98.
[24] Nijman SM. Synthetic lethality: general principles, utility and detection using genetic screens in human cells. FEBS Lett. 2011;585:1–6.
[25] Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Pagé N, et al. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science. 2001;294:2364–8.
[26] Pan X, Yuan DS, Ooi S-L, Wang X, Sookhai-Mahadeo S, Meluh P, et al. dSLAM analysis of genome-wide genetic interactions in Saccharomyces cerevisiae. Methods. 2007;41:206–21.
[27] Collins SR, Roguev A, Krogan NJ. Quantitative genetic interaction mapping using the E-MAP approach. Methods Enzymol. 2010;470:205–31.
[28] Ooi SL, Shoemaker DD, Boeke JD. DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray. Nat Genet. 2003;35:277–86.
[29] Wu M, Min W, Xuejuan L, Fan Z, Xiaoli L, Chee-Keong K, et al.. Meta-analysis of genomic and proteomic features to predict synthetic lethality of yeast and human cancer. Proceedings of the international conference on Bioinformatics, Computational Biology and Biomedical Informatics – BCB'13, 2007.
[30] Iorns E, Lord CJ, Turner N, Ashworth A. Utilizing RNA interference to enhance cancer drug discovery. Nat Rev Drug Discov. 2007;6:556–68.
[31] Scholl C, Fröhling S, Dunn IF, Schinzel AC, Barbie DA, Kim SY, et al. Synthetic lethal interaction between oncogenic KRAS dependency and STK33 suppression in human cancer cells. Cell. 2009;137:821–34.
[32] Berns K, Hijmans EM, Mullenders J, Brummelkamp TR, Velds A, Heimerikx M, et al. A large-scale RNAi screen in human cells identifies new components of the p53 pathway. Nature. 2004;428:431–7.
[33] Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature. 2009;462:108–12.
[34] Luo J, Emanuele MJ, Li D, Creighton CJ, Schlabach MR, Westbrook TF, et al. A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. Cell. 2009;137:835–48.
[35] Wong SL, Zhang LV, Tong AHY, Li Z, Goldberg DS, King OD, et al. Combining biological networks to predict genetic interactions. Proc Natl Acad Sci USA. 2004;101:15682–7.
[36] Zhong W. Genome-wide prediction of C. elegans genetic interactions. Science. 2006;311:1481–4.
[37] Paladugu SR, Zhao S, Ray A, Raval A. Mining protein networks for synthetic genetic interactions. BMC Bioinform. 2008;9:426.
[38] You Z-H, Yin Z, Han K, Huang D-S, Zhou X. A semi-supervised learning approach to predict synthetic genetic interactions by combining functional and topological properties of functional gene network. BMC Bioinform. 2010;11:343.

[39] Bandyopadhyay N, Ranka S, Kahveci T. SSLPred: predicting synthetic sickness lethality. Pac Symp Biocomput. 2012;7–18 http://www.worldscientific.com/worldscibooks/10.1142/8254#t=toc.
[40] Wang X, Simon R. Identification of potential synthetic lethal genes to p53 using a computational biology approach. BMC Med Genomics. 2013;6:30.
[41] Zheng J, Wu M, Li X, Zhang F, Li X, Kwoh C-K, et al. In silico prediction of synthetic lethality by meta-analysis of genetic interactions, functions, and pathways in yeast and human cancer. Cancer Inform. 2014;13:71.
[42] Zhang F, Fan Z, Min W, Xue-Juan L, Xiao-Li L, Kwoh CK, et al. Predicting essential genes and synthetic lethality via influence propagation in signaling pathways of cancer cell fates. J Bioinform Comput Biol. 2015;13:1541002.
[43] Jacunski A, Dixon SJ, Tatonetti NP. Connectivity homology enables inter-species network models of synthetic lethality. PLoS Comput Biol. 2015;11:e1004506.
[44] Dixon SJ, Andrews BJ, Boone C. Exploring the conservation of synthetic lethal genetic interaction networks. Commun Integr Biol. 2009;2:78–81.
[45] Conde-Pueyo N, Munteanu A, Solé RV, Rodríguez-Caso C. Human synthetic lethal inference as potential anti-cancer target gene detection. BMC Syst Biol. 2009;3:116.
[46] VanderSluis B, Bellay J, Musso G, Costanzo M, Papp B, Vizeacoumar FJ, et al. Genetic interactions reveal the evolutionary trajectories of duplicate genes. Mol Syst Biol. 2010;6:429.
[47] Koch EN, Costanzo M, Bellay J, Deshpande R, Chatfield-Reed K, Chua G, et al. Conserved rules govern genetic interaction degree across species. Genome Biol. 2012;13:R57.
[48] Lu X, Kensche PR, Huynen MA, Notebaart RA. Genome evolution predicts genetic interactions in protein complexes and reveals cancer drug targets. Nat Commun. 2013;4:2124.
[49] Pesquita C, Faria D, Falcão AO, Lord P, Couto FM. Semantic similarity in biomedical ontologies. PLoS Comput Biol. 2009;5:e1000443.
[50] Hoehndorf R, Hardy NW, Osumi-Sutherland D, Tweedie S, Schofield PN, Gkoutos GV. Systematic analysis of experimental phenotype data reveals gene functions. PLoS One. 2013;8:e60847.
[51] Li X-L. Biological data mining in protein interaction networks. Hershey, PA: IGI Global, 2009.
[52] Miller CA, Settle SH, Sulman EP, Aldape KD, Milosavljevic A. Discovering functional modules by identifying recurrent and mutually exclusive mutational patterns in tumors. BMC Med Genomics. 2011;4:34.
[53] Tomczak K, Czerwińska P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. Contemp Oncol. 2015;19:A68–77.
[54] Leiserson MD, Wu H-T, Vandin F, Raphael BJ. CoMEt: a statistical approach to identify combinations of mutually exclusive alterations in cancer. Lect Notes Comput Sci. 2015;202–4.
[55] Babur Ö, Gönen M, Aksoy BA, Schultz N, Ciriello G, Sander C, et al. Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. Genome Biol. 2015;16:45.
[56] Srihari S, Singla J, Wong L, Ragan MA. Inferring synthetic lethal interactions from mutual exclusivity of genetic events in cancer. Biol Direct. 2015;10:57.
[57] Jerby-Arnon L, Pfetzer N, Waldman YY, McGarry L, James D, Shanks E, et al. Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell. 2014;158:1199–209.
[58] Pearl LH, Schierz AC, Ward SE, Al-Lazikani B, Pearl FM. Therapeutic opportunities within the DNA damage response. Nat Rev Cancer. 2015;15:166–80.
[59] Bryant HE, Niklas S, Thomas HD, Parker KM, Dan F, Elena L, et al. Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase. Nature. 2005;434:913–7.
[60] Turner NC, Lord CJ, Iorns E, Brough R, Swift S, Elliott R, et al. A synthetic lethal siRNA screen identifying genes mediating sensitivity to a PARP inhibitor. EMBO J. 2008;27:1368–77.
[61] Liu JF, Konstantinopoulos PA, Matulonis UA. PARP inhibitors in ovarian cancer: current status and future promise. Gynecol Oncol. 2014;133:362–9.
[62] Tangutoori S, Baldwin P, Sridhar S. PARP inhibitors: a new era of targeted therapy. Maturitas. 2015;81:5–9.
[63] Mateo J, Carreira S, Sandhu S, Miranda S, Mossop H, Perez-Lopez R, et al. DNA-repair defects and olaparib in metastatic prostate cancer. N Engl J Med. 2015;373:1697–708.
[64] Syed YY. Rucaparib: first global approval. Drugs. 2017;77:585–92.
[65] Rucaparib approved for ovarian cancer. Cancer Discov. 2017;7:120–1.
[66] Litton JK, Scoggins M, Ramirez DL, Murthy RK, Whitman GJ, Hess KR, et al. A pilot study of neoadjuvant talazoparib for early-stage breast cancer patients with a BRCA mutation. Ann Oncol. 2016;27:43–67.
[67] Emerling BM, Hurov JB, Poulogiannis G, Tsukazawa KS, Choo-Wing R, Wulf GM, et al. Depletion of a putatively druggable class of phosphatidylinositol kinases inhibits growth of p53-null tumors. Cell. 2013;155:844–57.
[68] Bitler BG, Aird KM, Garipov A, Li H, Amatangelo M, Kossenkov AV, et al. Synthetic lethality by targeting EZH2 methyltransferase activity in ARID1A-mutated cancers. Nat Med. 2015;21:231–8.
[69] Muller FL, Colla S, Aquilanti E, Manzo VE, Genovese G, Lee J, et al. Passenger deletions generate therapeutic vulnerabilities in cancer. Nature. 2012;488:337–42.
[70] Abbotts R, Jewell R, Nsengimana J, Maloney DJ, Simeonov A, Seedhouse C, et al. Targeting human apurinic/apyrimidinic endonuclease 1 (APE1) in phosphatase and tensin homolog (PTEN) deficient melanoma cells for personalized therapy. Oncotarget. 2014;5:3273–86.
[71] Williamson CT, Miller R, Pemberton HN, Jones SE, Campbell J, Konde A, et al. ATR inhibitors as a synthetic lethal therapy for tumours deficient in ARID1A. Nat Commun. 2016;7:13837.
[72] Karnitz LM, Zou L. Molecular pathways: targeting ATR in cancer therapy. Clin Cancer Res. 2015;21:4780–5.
[73] Workman P, Al-Lazikani B, Clarke PA. Genome-based cancer therapeutics: targets, kinase drug resistance and future strategies for precision oncology. Curr Opin Pharmacol. 2013;13:486–96.
[74] Pylayeva-Gupta Y, Grabocka E, Bar-Sagi D. RAS oncogenes: weaving a tumorigenic web. Nat Rev Cancer. 2011;11:761–74.

[75] Puyol M, Martín A, Dubus P, Mulero F, Pizcueta P, Khan G, et al. A synthetic lethal interaction between K-Ras oncogenes and Cdk4 unveils a therapeutic strategy for non-small cell lung carcinoma. Cancer Cell. 2010;18:63–73.
[76] Downward J. RAS synthetic lethal screens revisited: still seeking the elusive prize?. Clin Cancer Res. 2015;21:1802–9.
[77] Papke B, Der CJ. Drugging RAS: know the enemy. Science. 2017;355:1158–63.
[78] Costa-Cabral S, Brough R, Konde A, Aarts M, Campbell J, Marinari E, et al. CDK1 is a synthetic lethal target for KRAS mutant tumours. PLoS One. 2016;11:e0149099.
[79] Kumar MS, Hancock DC, Molina-Arcas M, Steckel M, East P, Diefenbacher M, et al. The GATA2 transcriptional network is requisite for RAS oncogene-driven non-small cell lung cancer. Cell. 2012;149:642–55.
[80] Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45:D362–8.
[81] Petryszak R, Keays M, Tang YA, Fonseca NA, Barrera E, Burdett T, et al. Expression Atlas update—an integrated database of gene and protein expression in humans, animals and plants. Nucleic Acids Res. 2016;44:D746–52.
[82] Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002;30:207–10.
[83] Okamura Y, Aoki Y, Obayashi T, Tadaka S, Ito S, Narise T, et al. COXPRESdb in 2015: coexpression database for animal species by DNA-microarray and RNAseq-based expression data with multiple quality assessment systems. Nucleic Acids Res. 2015;43:D82–6.
[84] Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9.
[85] Forbes SA, Beare D, Boutselakis H, Bamford S, Bindal N, Tate J, et al. COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. 2017;45:D777–83.
[86] Aken BL, Achuthan P, Akanni W, Ridwan Amode M, Bernsdorff F, Bhai J, et al. Ensembl 2017. Nucleic Acids Res. 2016;45(D1):D635–D642.
[87] Linding R, Jensen LJ, Pasculescu A, Olhovsky M, Colwill K, Bork P, et al. NetworKIN: a resource for exploring cellular phosphorylation networks. Nucleic Acids Res. 2008;36:D695–9.
[88] Zhang J, Baran J, Cros A, Guberman JM, Haider S, Hsu J, et al. International Cancer Genome Consortium Data Portal—a one-stop shop for cancer genomics data. Database. 2011;2011:bar026.
[89] Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6:l1.
[90] Forbes SA, Beare D, Bindal N, Bamford S, Ward S, Cole CG, et al. COSMIC: high-resolution cancer genetics using the catalogue of somatic mutations in cancer. Curr Protoc Hum Genet. 2016;91:10.11.1–37.
[91] Richardson CJ, Gao Q, Mitsopoulous C, Zvelebil M, Pearl LH, Pearl FM. MoKCa database—mutations of kinases in cancer. Nucleic Acids Res. 2009;37:D824–31.
[92] Stark C. BioGRID: a general repository for interaction datasets. Nucleic Acids Res. 2006;34:D535–9.
[93] Guo J, Liu H, Zheng J. SynLethDB: synthetic lethality database toward discovery of selective and sensitive anticancer drug targets. Nucleic Acids Res. 2016;44:D1011–7.
[94] Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, et al. A global genetic interaction network maps a wiring diagram of cellular function. Science. 2016;pii: aaf1420353.
[95] Kondrashov AS. Progress and prospects in evolutionary biology: the Drosophila model. Jeffrey R. Powell. Q Rev Biol. 1998;73:349–50.
[96] Collins SR, Miller KM, Maas NL, Roguev A, Fillingham J, Chu CS, et al. Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. Nature. 2007;446:806–10.
[97] Byrne AB, Weirauch MT, Victoria W, Martina K, Dixon SJ, Stuart JM, et al. A global analysis of genetic interactions in Caenorhabditis elegans. J Biol. 2007;6:8.