Integrative analysis of large-scale loss-of-function screens identifies robust cancer-associated genetic interactions

Christopher J. Lord[1], Niall Quinn[2], Colm J. Ryan[2*]

1. Breast Cancer Now Toby Robins Research Centre and Cancer Research UK Gene Function Laboratory, Institute of Cancer Research, London SW3 6JB, UK
2. School of Computer Science and Systems Biology Ireland, University College Dublin, Dublin 4, Ireland
* Correspondence to colm.ryan@ucd.ie

Running title: Identifying robust genetic interactions in cancer
Keywords = synthetic lethality; cancer; genetic interactions; protein-protein interactions; genetic heterogeneity

Abstract
Genetic interactions, such as synthetic lethal effects, can now be systematically identified in cancer cell lines using high-throughput genetic perturbation screens. Despite this advance, few genetic interactions have been reproduced across multiple studies and many appear highly context-specific. Understanding which genetic interactions are robust in the face of the molecular heterogeneity observed in tumours and what factors influence this robustness could streamline the identification of therapeutic targets. Here, we develop a computational approach to identify robust genetic interactions that can be reproduced across independent experiments and across non-overlapping cell line panels. We used this approach to evaluate >140,000 potential genetic interactions involving cancer driver genes and identified 1,520 that are significant in at least one study but only 220 that reproduce across multiple studies. Analysis of these interactions demonstrated that: (i) oncogene addiction effects are more robust than oncogene-related synthetic lethal effects; and (ii) robust genetic interactions in cancer are enriched for gene pairs whose protein products physically interact. This suggests that protein-protein interactions can be used not only to understand the mechanistic basis of genetic interaction effects, but also to prioritise robust targets for further development. To explore the utility of this approach, we used a protein-protein interaction network to guide the search for robust synthetic lethal interactions associated with passenger gene alterations and validated two novel robust synthetic lethalities.

Introduction
Large-scale tumour genome sequencing efforts have provided us with a compendium of driver genes that are recurrently altered in human cancers(Vogelstein et al., 2013). In some cases, these genetic alterations have been associated with altered sensitivity to targeted therapies. Examples of targeted therapies already in clinical use include approaches that exploit oncogene addictions, such as the increased sensitivity of *BRAF* mutant melanomas to BRAF inhibitors(Chapman et al., 2011), and approaches that exploit non-oncogene addiction/synthetic lethality, such as the sensitivity of *BRCA1*/*2* mutant ovarian or breast cancers to PARP inhibitors(Lord and Ashworth, 2017). An ongoing challenge is to associate the presence of other driver gene alterations with sensitivity to existing therapeutic agents(Barretina et al., 2012; Iorio et al., 2016) or to identify candidate therapeutic targets whose inhibition may provide therapeutic benefit to patients with specific mutations. Towards this end, multiple groups have performed large-scale loss-of-function genetic perturbation screens in panels of tumour cell lines to identify vulnerabilities that are associated with the presence or absence of specific driver gene mutations (i.e. genetic interactions)(Behan et

al., 2019; Campbell et al., 2016; Marcotte et al., 2016; McDonald et al., 2017; Meyers et al., 2017; Tsherniak et al., 2017). Others have performed screens in 'isogenic' cell line pairs that differ only by the presence of a specific oncogenic alteration(Martin et al., 2017; Steckel et al., 2012). Despite these large-scale efforts, very few genetic interactions have been identified in more than one study (recently reviewed(Ryan et al., 2018)). Even in the case of cancer driver genes subjected to multiple screens, such as *KRAS*, few genetic interactions have been identified in more than one screen(Downward, 2015). This lack of reproducibility may be due to technical issues, e.g. false positives and false negatives due to inefficient gene targeting reagents(Kaelin, 2012), and/or real biological issues, such as the context specificity of genetic interactions(Henkel et al., 2019; Ryan et al., 2018). We refer to those genetic interactions that can be reproduced across multiple screens and across distinct cell line contexts as *robust genetic interactions*. Given that tumours exhibit considerable molecular heterogeneity both within and between patients there is a real need to: (i) identify robust genetic interactions that can be reproduced across heterogeneous cell line panels, reasoning that these reproducible effects will be more likely to be robust in the face of the molecular heterogeneity seen in human cancers; (ii) prioritise these robust genetic interactions for further therapeutic development; and (iii) understand the characteristics of robust genetic interactions in cancer as a means to predict new therapeutic targets.

To achieve this, we developed, and describe here, a computational approach that leverages large-scale cell line panel screens to identify those genetic interactions that can be reproducibly identified across multiple independent experiments. We found that for all oncogenes studied, the most significant reproducible dependency identified was an oncogene-addiction rather than a synthetic lethal effect. Excluding oncogene addictions, we found 220 reproducible genetic interactions. In investigating the nature of these robust genetic interactions, we found that they are significantly enriched among gene pairs whose protein products physically interact. This suggests that incorporating prior knowledge of protein-protein interactions may be a useful approach to guide the selection of reproducible "hits" from genetic screens as candidates worth considering as therapeutic targets in cancer. We demonstrate the utility of the approach in identifying robust synthetic lethal interactions from chemogenetic screens and in identifying synthetic lethal interactions associated with 'passenger' gene alterations.

Results

A "discovery and validation" approach to the analysis of loss-of-function screens identifies reproducible genetic dependencies

We first wished to identify genetic interactions that could be independently reproduced across multiple distinct loss-of-function screens. To do this, we obtained gene sensitivity scores from four large-scale loss-of-function screens in panels of tumour cell lines, including two shRNA screens (DRIVE(McDonald et al., 2017), DEPMAP(Tsherniak et al., 2017)) and two CRISPR-Cas9 mutagenesis screens (AVANA(Meyers et al., 2017), SCORE(Behan et al., 2019)). We harmonised the cell line names across all studies, so they could be compared with each other and also with genotypic data (Barretina et al., 2012; Iorio et al., 2016)(Fig. 1a). In total, 917 tumour cell lines were screened in at least one loss-of-function study. Only 50 of these cell lines were common to all four studies while 407 cell lines were included in only a single study (Fig. 1b). It is the partially overlapping nature of the screens that motivated the subsequent approach we took for our analysis. We used a 'discovery set' and 'validation set' approach to identifying genetic interactions across multiple screens - first identifying associations between driver gene alterations and gene inhibition sensitivity in the discovery study and then testing the discovered dependency in the validation study (Fig. 1c). However, to ensure that any reproducibility observed was not merely due to cell lines common to both datasets, we first removed cell lines from the validation dataset if they were

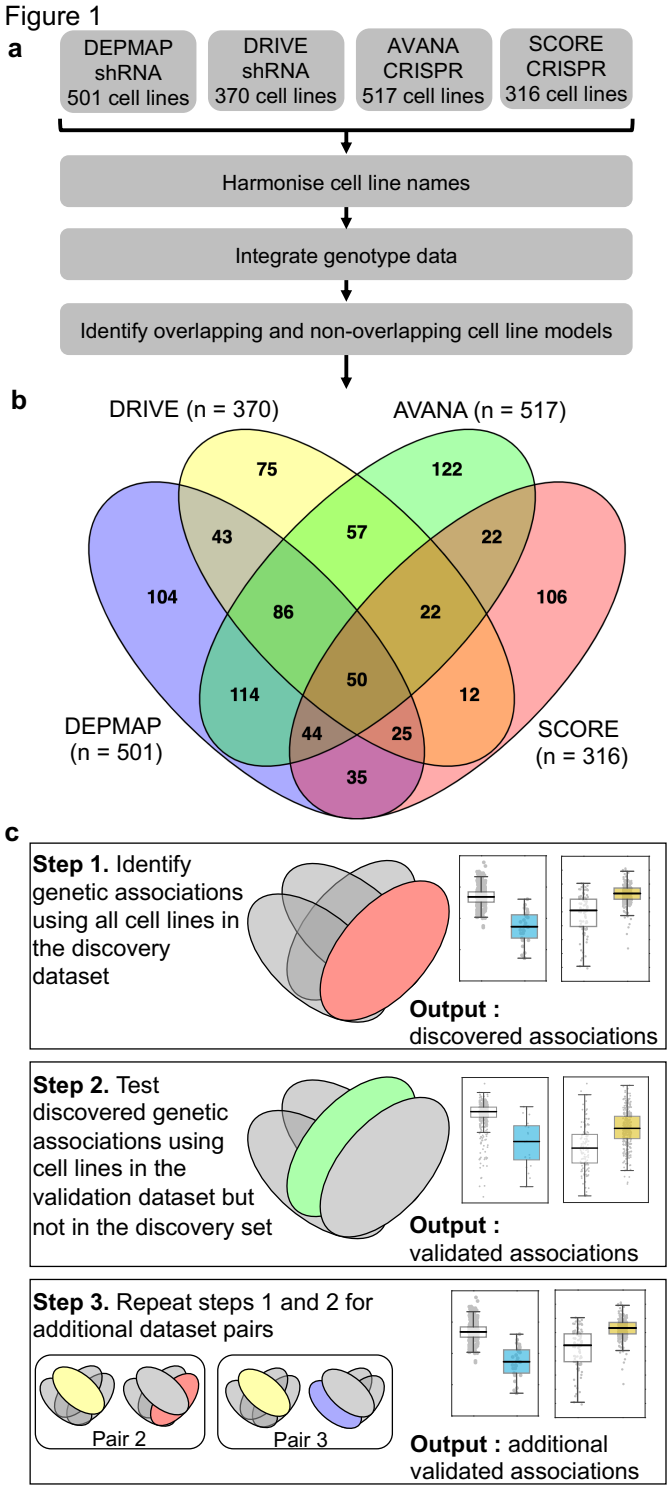


Figure 1. Identifying robust genetic interactions using partially overlapping loss-of-function screens. a) Workflow showing the integration of four different loss-of-function screen datasets. **b)** Venn diagram showing the overlap of cell lines between the four datasets analysed in this study. **c)** Workflow showing how robust genetic interactions are identified using discovery and validation sets.

present in the discovery dataset (Fig. 1c). For example, when using DEPMAP as the discovery dataset and AVANA as the validation dataset, we performed the validation analysis on the subset of cell lines that were present in AVANA but not in DEPMAP. In doing

so, we ensured that any genetic interactions discovered were reproducible across different screening platforms (either distinct gene inhibition approaches, i.e. shRNA vs CRISPR, or distinct shRNA/CRISPR libraries) and also robust to the molecular heterogeneity seen across different cell line panels.

Figure 2

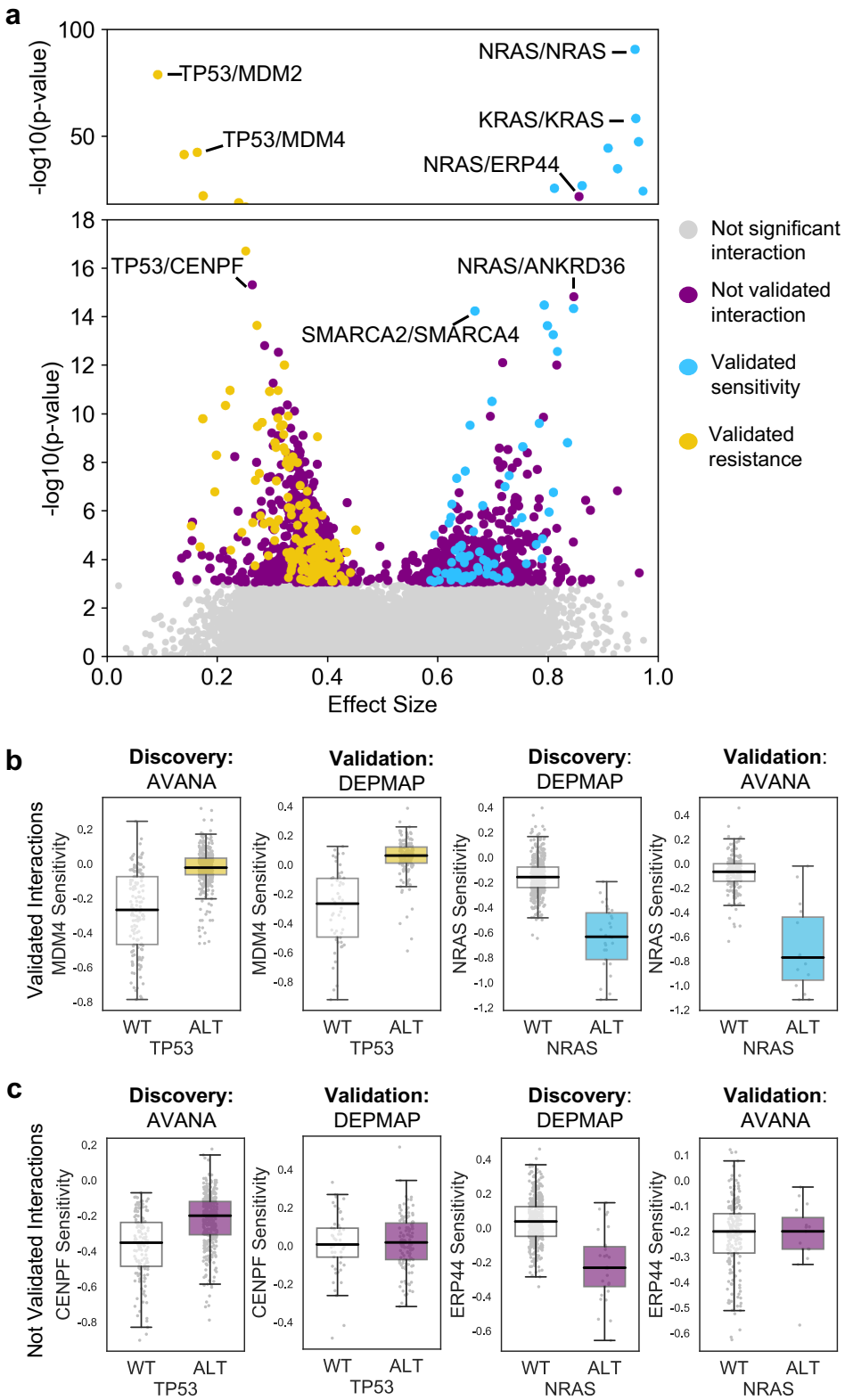


Figure 2. Discovered and validated genetic dependencies. a) Scatterplot showing the genetic dependencies identified across all datasets. Each individual point represents a gene pair, the x-axis shows the common language effect size, and the y-axis shows the -log10 p-value from the discovery dataset. Selected gene pairs are highlighted – the driver gene is listed first, followed by the associated dependency. Each gene pair may have been tested in multiple discovery studies, only the interaction with the most significant discovery p-value is shown. Scatterplots for individual studies are presented in Supplemental Fig. S1. **b)** Tukey boxplots showing examples of robust genetic dependencies, including an increased resistance of *TP53* mutant tumour cell lines to *MDM4* inhibition and increased sensitivity of *NRAS* mutant tumour cell lines to *NRAS* inhibition. In each box plot the top and bottom of the box represents the third and first quartiles and the box band represents the median; whiskers extend to 1.5 times the interquartile distance from the box. WT = wild type, ALT = altered. Throughout blue is used to indicate increased sensitivity (synthetic lethality or oncogene addiction), yellow to indicate resistance to inhibition of the target gene. **c)** Boxplots showing examples of genetic dependencies discovered but not validated, including an increased resistance of *TP53* mutant cancers to *CENPF* inhibition and increased sensitivity of *NRAS* mutant tumour cell lines to *ERP44* inhibition.

Similar to our previous work(Bridgett et al., 2017; Campbell et al., 2016) we integrated copy number profiles and exome sequencing data to annotate all cell lines according to whether or not they featured likely functional alterations in any one of a panel of cancer driver genes(Vogelstein et al., 2013) (see Methods, Supplemental Table S1). We then identified associations between driver gene alterations and sensitivity to the inhibition of specific genes using a multiple regression model that included tissue type as a covariate to reduce the possibility of confounding by tissue type (see Methods). We focused this analysis on 'selectively lethal' genes - i.e. those genes whose inhibition killed some, but not all cell lines (Methods, Supplemental Table S1). We analysed each pair of screens in turn and considered a genetic dependency to be reproducible if it was validated in at least one discovery/validation pair. Using this approach, we tested 142,477 potential genetic dependencies and identified 1,530 dependencies that were significant in at least one screen (Fig. 2a, Supplemental Fig. S1). Of these 1,530 dependencies, 229 could be validated in a second screen (Supplemental Table S3, Fig. 2a). For example, in the AVANA dataset *TP53* mutation was associated with resistance to inhibition of both *MDM4* and *CENPF*, but only the association with *MDM4* could be validated in a second dataset (Fig. 2b, 2c). Similarly, in the DEPMAP dataset *NRAS* mutation was associated with increased sensitivity to the inhibition of both *NRAS* itself and *ERP44*, but only the sensitivity to inhibition of *NRAS* could be validated in a second dataset (Fig. 2b, 2c).

Of the reproducible genetic dependencies nine were 'self *vs.* self' associations, where the alteration of a gene was associated with sensitivity to its own inhibition. The majority of these 'self *vs.* self' associations were oncogene addiction effects, such as the increased sensitivity of *NRAS* mutant cell lines to *NRAS* inhibition (Fig. 2b). Similarly, we identified robust oncogene addictions involving the *CTNNB1* (β-Catenin), *KRAS*, *EGFR*, *BRAF*, *ERBB2* and *PIK3CA* oncogenes (Fig. 3a, 3b, Supplemental Fig. S2b). For *EGFR* and *CTNNB1*, the only identified robust dependency was an oncogene addiction effect. For all other oncogenes there were additional robust dependencies identified, but in all cases the most significant reproducible dependency was an oncogene addiction (Supplemental Fig. S2a). These observations suggest that for most oncogenes the oncogene addiction effect might be more robust than any oncogene-related synthetic lethal effects.

We also identified two examples of 'self *vs.* self' dependencies involving tumour suppressors -*TP53* (aka p53) and *CDKN2A* (aka p16/p14arf) (Supplemental Fig. S2c). This type of relationship has previously been reported for *TP53*: *TP53* inhibition appears to

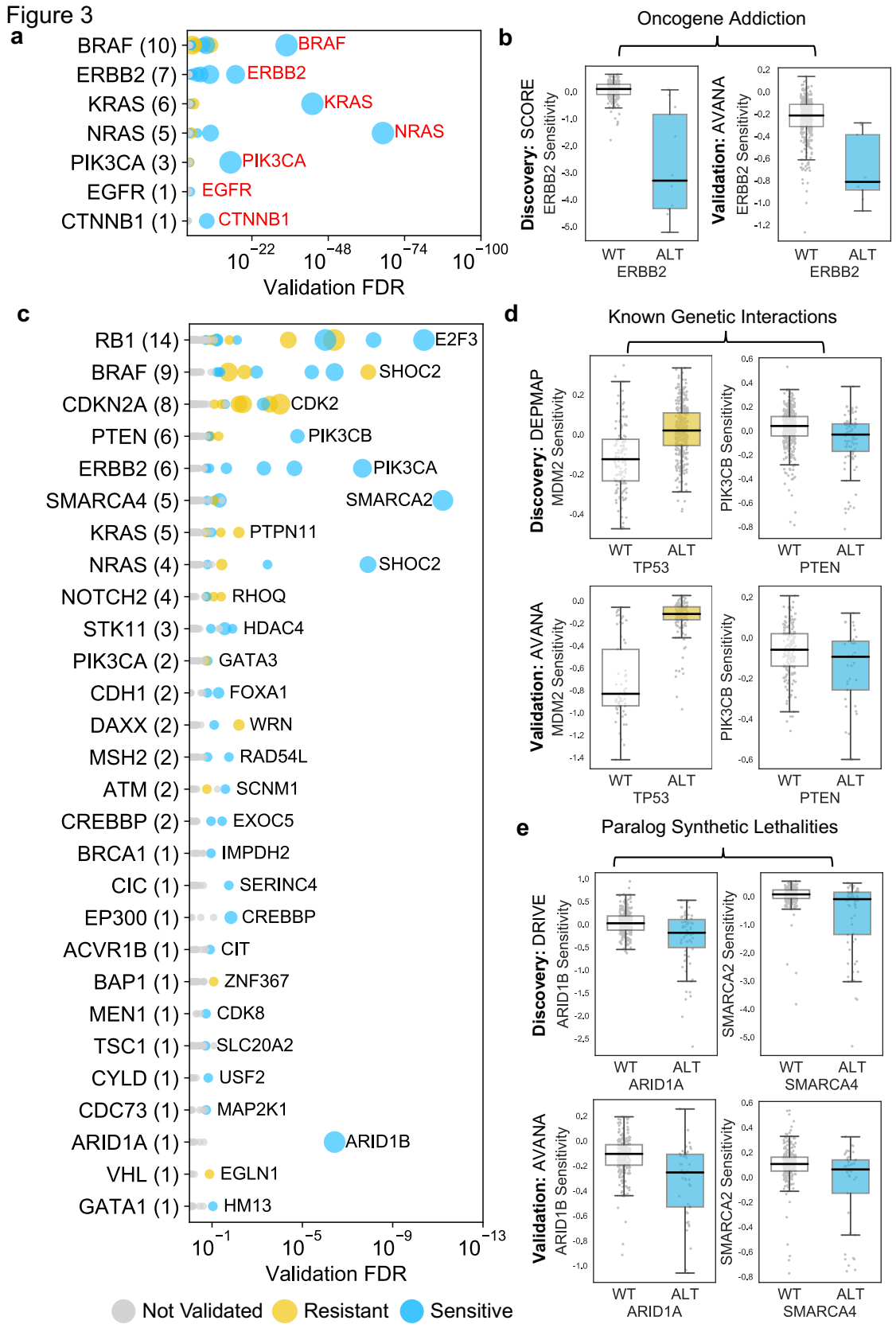


Figure 3. Identified robust genetic interactions. a) Dot plot showing the robust genetic dependencies identified for oncogenes. Each coloured circle indicates a robust genetic

dependency, scaled according to the number of dataset pairs it was validated in. The most significant genetic dependency (lowest FDR in a validation set) for each driver gene is labelled. Oncogenes are sorted by the number of robust dependencies and the total number of robust genetic dependencies for each driver gene is shown in parentheses. **b**) Example of a validated oncogene addiction – *ERBB2* amplified cells are sensitive to *ERBB2* inhibition. Left shows the discovery dataset (SCORE) and right shows the validation dataset (AVANA). **c)** Dot plot showing the robust genetic interactions identified for all driver genes. Each coloured circle indicates a robust genetic interaction, scaled according to the number of dataset pairs it was validated in. The most significant genetic interaction (lowest FDR in a validation set) for each driver gene is labelled. Drivers are sorted by the number of robust interactions and the total number of robust genetic interactions for each driver gene is shown in parentheses. *TP53* (132 robust genetic interactions) has been excluded for clarity, as have all self-self dependencies. **d)** Examples of known genetic interactions identified from the integrated analysis, including an increased sensitivity of *PTEN* mutant tumour cell lines to *PIK3CB* inhibition and increased resistance of *TP53* mutant tumour cell lines to *MDM2* inhibition. Top row shows the data used to discover the interactions (DEPMAP dataset) while the bottom row shows the data used to validate the interactions (AVANA dataset with cell lines from DEPMAP excluded). **e)** Synthetic lethal interactions involving paralog pairs. Top row shows the data used to discover the interactions (DRIVE dataset) while the bottom row shows the data used to validate the interactions (AVANA dataset with cell lines from DRIVE excluded).

offer a growth advantage to *TP53* wild type cells but not to *TP53* mutant cells(Giacomelli et al., 2018). Consequently, we observed an association between *TP53* status and sensitivity to TP53 inhibition. Similar effects were seen for *CDKN2A* (Supplemental Fig. S2c). These 'self *vs.* self' dependencies, in particular the oncogene addictions, serve as evidence that our approach could identify well characterised genetic associations. However, as our primary interest was in genetic interactions between different genes, we excluded 'self *vs.* self' interactions from further analysis, leaving us with 220 robust genetic interactions (Fig. 3c).

Many robust genetic interactions reflect known pathway structure

Many of the reproducible genetic interactions we identified have been previously reported, including both sensitivity relationships, such as increased sensitivity of *PTEN* mutant cell lines to inhibition of the phosphoinositide 3-kinase-coding gene *PIK3CB*(Wee et al., 2008), and resistance relationships, such as an increased resistance of *TP53* mutant cell lines to *MDM2* inhibition (Fig. 3d).

Amongst the set of 220 robust genetic interactions, we identified two previously reported 'paralog lethalities' – synthetic lethal relationships between duplicate genes (paralogs) (Helming et al., 2014; Hoffman et al., 2014; Oike et al., 2013) (Fig. 3e). We found a robust association between mutation of the tumour suppressor *ARID1A* and sensitivity to inhibition of its paralog *ARID1B* (Helming et al., 2014) and also an association between mutation of *SMARCA4* and sensitivity to inhibition of its paralog *SMARCA2* (Hoffman et al., 2014; Oike et al., 2013). Both pairs of genes (*ARID1A/ARID1B* and *SMARCA4/SMARCA2*) encode components of the larger SWI/SNF complex(Wilson and Roberts, 2011).

Some of the robust genetic dependencies could be readily interpreted using known pathway structures. For instance, many of the robust dependencies associated with the oncogene *BRAF* could be interpreted in terms of BRAF's role in the MAPK pathway. *BRAF* mutation was associated with increased sensitivity to inhibition of its downstream effectors MEK (*MAP2K1)* and ERK *(MAPK1)*, and increased resistance to inhibition of the alternative RAF isoform gene CRAF (*RAF1*) and the MAPK regulator *PTPN11* (Fig. 4a, 4b)(Hill et al., 2019). *BRAF* mutation was also associated with increased sensitivity to inhibition of *PEA15,*

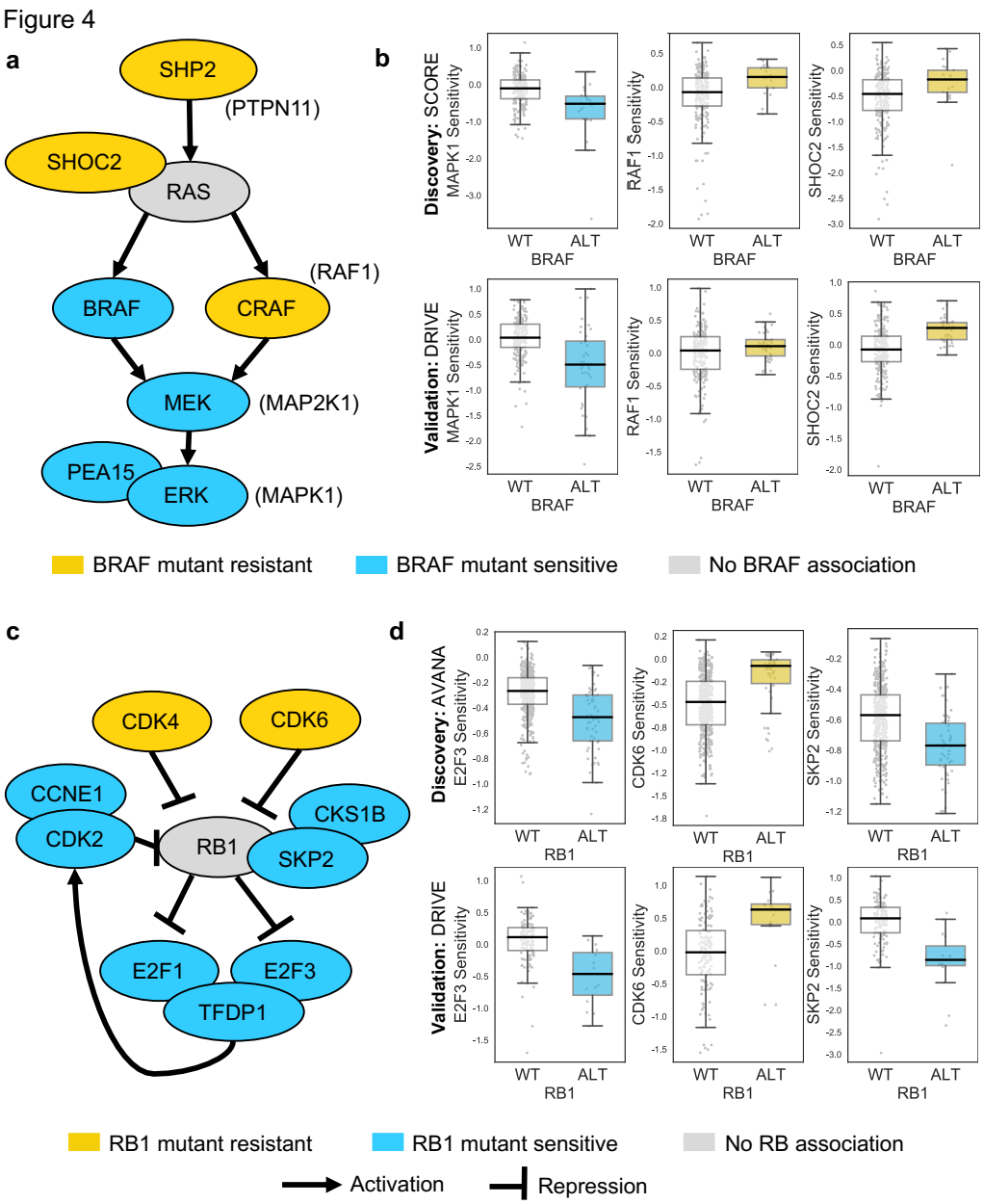


Figure 4. Robust genetic interactions involving *RB1* and *BRAF* recapitulate pathway relationships. a) Simplified RAS/RAF/MEK/ERK pathway diagram. Protein names (e.g. MEK) are shown inside nodes, while associated gene names are shown adjacent (e.g. *MAP2K1*). Nodes are coloured according to their association with BRAF mutation - blue indicates increased sensitivity of *BRAF* mutant cell lines, yellow indicates increased resistance. Selected robust genetic interactions are shown in boxplots below. **b)** Boxplots showing selected genetic interactions associated with *BRAF* mutation. **c)** Simplified Rb pathway diagram, highlighting robust genetic interactions involved in the Rb pathway. **d)** Boxplots showing selected genetic interactions associated with *RB1* alteration

presumably a result of the requirement of PEA15 for ERK dimerisation and signalling activity (Formstecher et al., 2001; Herrero et al., 2015).

Mutation or deletion of the tumour suppressor *RB1* (Retinoblastoma 1, Rb) was associated with increased sensitivity or resistance to inhibition of multiple Rb pathway members (Fig. 4c, 4d). We found that *RB1* loss was reproducibly associated with resistance to inhibition of

its negative regulators *CDK4* and *CDK6*, consistent both with the known Rb pathway structure and with preclinical data suggesting that *RB1* mutation confers resistance to CDK4/6 inhibitors(Asghar et al., 2015; O'Leary et al., 2018). Rb is a negative regulator of multiple E2F transcription factors, and we found that *RB1* loss was reproducibly associated with increased sensitivity to both *E2F1* and *E2F3* inhibition (Figure 4c, 4d). *RB1* loss was also associated with robust sensitivity to *SKP2*, a binding partner of Rb(Ji et al., 2004) first identified as an *RB1* synthetic lethal partner in retinoblastoma(Xu et al., 2014) and more recently as a highly penetrant *RB1* synthetic lethal partner in triple negative breast cancer(Brough et al., 2018) (Fig. 4c, 4d). Finally, *RB1* loss was reproducibly associated with increased sensitivity to inhibition of Cyclin Dependent Kinase 2 (*CDK2)*, suggesting that it may be a useful biomarker for CDK2-specific inhibitors(Tadesse et al., 2018).

Robust genetic interactions are enriched in protein-protein interaction pairs

In seeking to understand what particular characteristics robust genetic interactions might have, we noted that many of the robust genetic interactions we identified involved gene pairs whose protein products operate in the same pathway (e.g. the Rb pathway) or protein complex (e.g. SWI/SNF) suggesting that genetic interactions between gene pairs whose protein products physically interact may be more robust than other genetic interactions. To test this hypothesis, we compared the robust genetic interactions we identified with protein-protein interactions from the STRING protein-protein interaction database(Szklarczyk et al., 2015). We found that, when considering the set of all gene pairs tested, gene pairs whose protein products physically interact were more likely to be identified as significant genetic interactors in at least one dataset (Fig. 5a) (Odds Ratio (OR) = 4.0, $p<2x10^{-16}$, Fisher's Exact test). Furthermore, of the genetic interactions identified as significant in at least one dataset, those that are supported by a protein-protein interaction were significantly more likely to be reproduced in a second dataset (Fig. 5a) (OR = 3.9, $p<1x10^{-13}$). We therefore concluded that protein-protein interaction pairs are more likely to be significant hits in one dataset and even more likely to be reproduced across multiple datasets, suggesting this might be a feature of robust genetic interaction effects.

We noted that a large number (n = 132) of robust genetic interactions involved *TP53*, presumably as a result of the high number of *TP53* mutant tumour cell lines in the datasets (and its high mutation frequency in human cancer) and the associated increased statistical power to detect *TP53*-related genetic interactions. We therefore considered whether the significant number of *TP53*-related genetic interactions in our dataset could confound our analyses, especially as *TP53* is also associated with a disproportionately high number of protein interactions (>1700 medium confidence interactors in the STRING database alone, compared to a median of 37 medium confidence interactions across all proteins). However, even after excluding genetic interactions involving *TP53,* the observation that robust genetic interactions were enriched in protein-protein interaction pairs was still evident (Fig. 5b); known protein interaction pairs were more likely to be identified as significant genetic interactions in one screen (OR = 3.8, $p<2x10^{-16}$) and among the significant genetic interactions discovered in one screen, those involving protein-protein interaction pairs were more likely to be reproduced in a validation screen (OR = 9.3, $p<2x10^{-16}$). The same effects were observed when considering genetic interactions observed at different false-discovery rate (FDR) thresholds (Supplemental Fig. S3a and 3b) and using different sources of protein-protein interaction data (Supplemental Fig. S3c and 3d, Supplemental Table S4)(Alanis-Lobato et al., 2017; Chatr-aryamontri et al., 2014).

The increased reproducibility of genetic interactions associated with protein-protein interactions across different genetic perturbation screen datasets could have two distinct causes - increased reproducibility across distinct technologies or libraries (e.g. CRISPR/shRNA) or increased reproducibility/robustness of genetic interactions in cell line

Figure 5

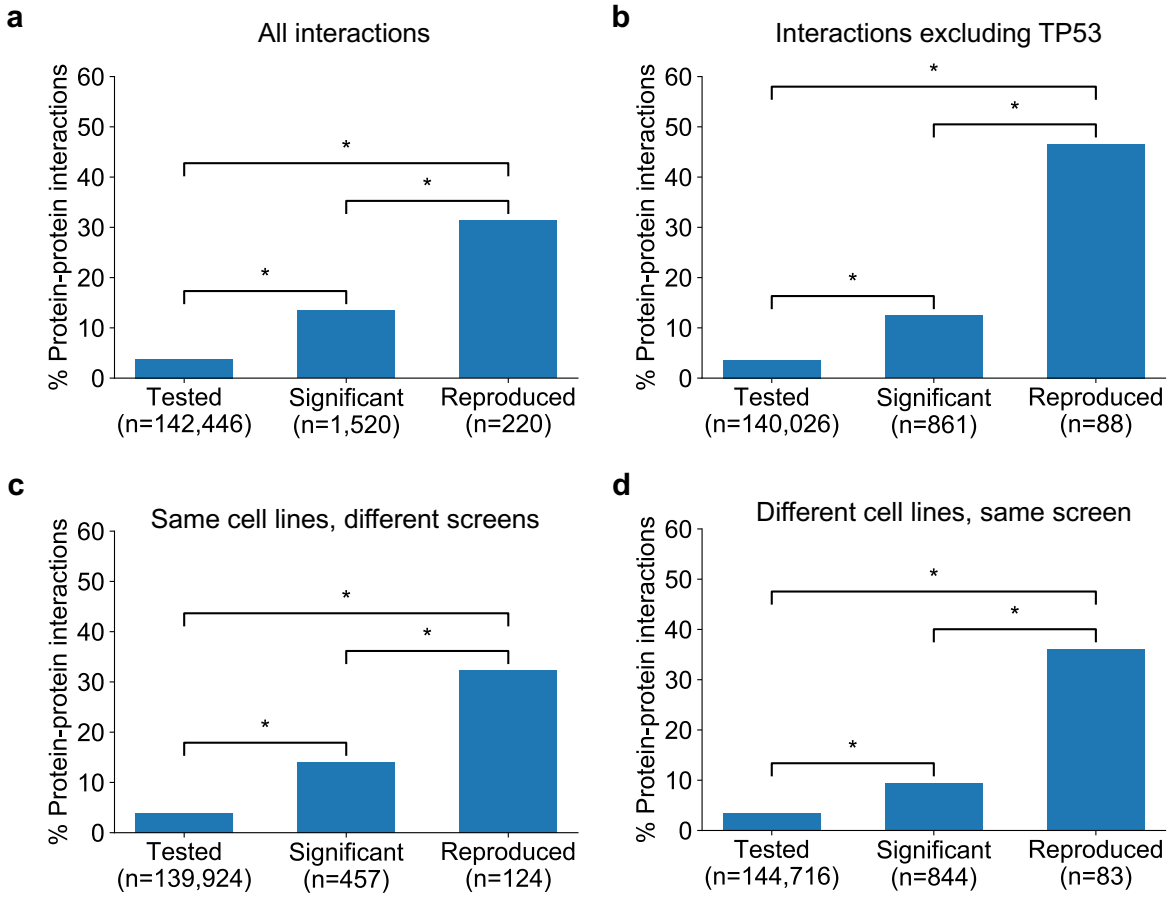


Figure 5. Robust genetic interactions are enriched in protein-protein interaction pairs. **a)** Barchart showing the percentage of protein-protein interacting pairs observed among different groups of gene pairs. The groups represent all gene pairs tested, gene pairs found to be significantly interacting in at least one screen (FDR < 20%), and gene pairs found to reproducibly interact across multiple screens (i.e. a discovery and validation screen). Stars (*) indicate significant differences between groups, all significant at P<0.001 using Fisher's Exact Test. Odds ratios and p-values are provided in Supplemental Table S4. **b)** As a but with interactions associated with *TP53* removed. **c)** As b but here the discovery and validation sets contain the same cell lines screened in different studies (e.g. 'AVANA ∩ DEPMAP' as discovery and 'DEPMAP ∩ AVANA' as validation). Consequently, reproducibility here means 'technical reproducibility' using different screening platforms. **d)** Similar to b but here the discovery and validation sets contain single datasets partitioned into non-overlapping cell line sets (e.g. 'AVANA \ DEPMAP' as discovery and 'AVANA ∩ DEPMAP' as validation). Consequently, reproducibility here means 'genetic robustness' - the same association between gene pairs is observed across distinct genetic backgrounds.

panels with distinct molecular backgrounds. To test the former possibility, we repeated our discovery/validation approach but focused on the set of cell lines that were common to different genetic perturbation screen datasets. Using this approach, the molecular backgrounds (i.e. cell lines) tested were the same, but the screening approach or library used differed. Upon doing this, we found that genetic interactions between gene pairs whose protein products physically interact were significantly more reproducible across studies (Fig. 5c, OR=6.1 and $p<2x10^{-10}$ when compared to discovered genetic interactions) (Supplemental Table S4). To test reproducibility using the same screening approach across molecularly distinct cell lines, we artificially split individual datasets into non-overlapping discovery and

validation sets of cell lines. Again, we found that genetic interactions between gene pairs whose protein products physically interact were more reproducible across distinct cell line panels (Fig. 5d, OR=8.0 and $p<1x10^{-12}$ when compared to discovered genetic interactions) (Supplemental Table S4). We therefore concluded that genetic interactions supported by protein-protein interactions were more reproducible across different screening approaches and across distinct cell line contexts, suggesting that these interactions are, overall, more robust.

Prioritising robust synthetic lethal interactions from chemogenetic screens

Figure 6

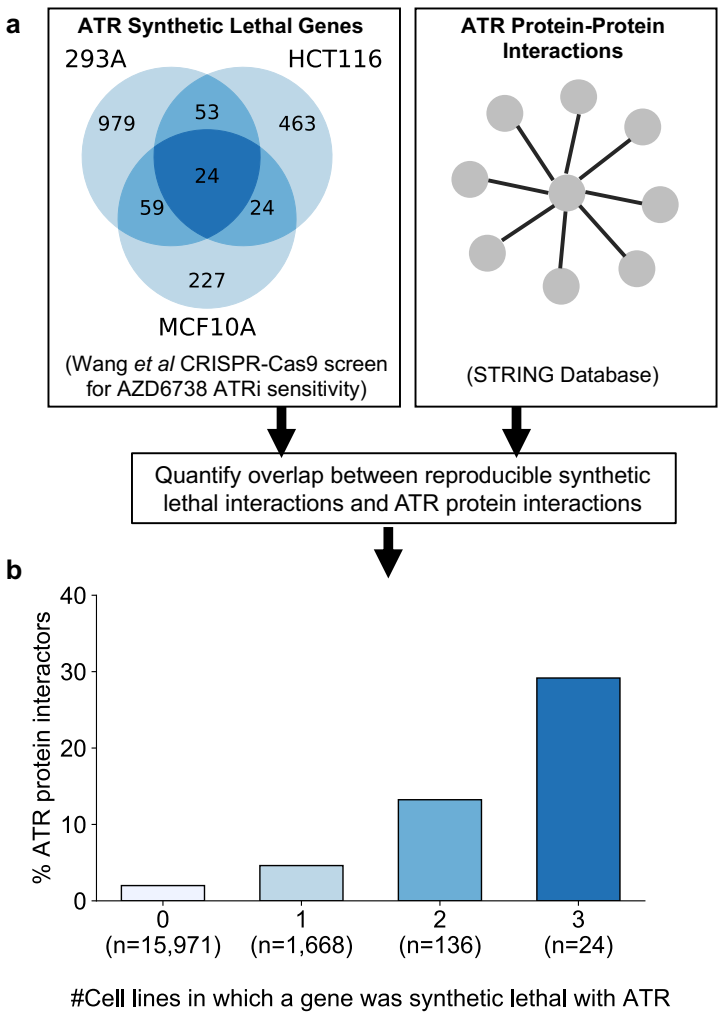


Figure 6. Reproducible ATR synthetic lethal interactions are enriched in ATR protein-protein interaction partners. a) Workflow - synthetic lethal interactions from CRISPR-Cas9 screens in three cell lines (Wang *et al,* 2018) were compared to identify reproducible synthetic lethal partners. These genes were then compared with known ATR protein-protein interaction partners from the STRING database. **b)** Bar chart showing the percentage of ATR protein interaction partners observed in different groups of genes. Genes are grouped according to whether they were identified as an ATR synthetic lethal partner in 0, 1, 2, or 3 cell line screens. Comparisons between all pairs of groups are significant at P<0.001 (Fisher's exact test) except for the comparison between genes that were hits in 2 and 3 cell lines (P=0.06).

As an alternative to genetic perturbation screening in large cell line panels, genetic interactions can also be identified using chemogenetic screens, where loss-of-function screens are performed in the presence and absence of specific small molecule inhibitors whose targets are relatively well defined. Based on the observations made earlier, we hypothesised that genetic interactions identified in chemogenetic screens that involved genes whose protein products physically interact with the target of the inhibitor should both be more likely to be identified as genetic interaction partners in one screen and also more likely to be reproduced across multiple screens (i.e. to be more robust). To test this hypothesis, we analysed the results of a recent chemogenetic screen performed to identify genes whose loss is synthetic lethal with ATR inhibition(Wang et al., 2018). In this study, genome-wide CRISPR-Cas9 screens in three cell lines from different histologies (breast, kidney, colon) were used to identify genes whose inhibition is selectively essential in the presence of a small molecule ATR kinase inhibitor(Vendetti et al., 2015; Wang et al., 2018)(Fig. 6a).

As predicted, we found that protein interaction partners of ATR are more likely than random genes to be identified as a significant synthetic lethal interactor of ATR in at least one cell line (Figure 6b). Furthermore, we found that among the synthetic lethal interactions identified in at least one cell line, those involving known ATR protein interaction partners were significantly more likely to be reproduced in a second or even third cell line (Fig. 6b). This suggests that, of the candidate genes identified in one screen, those that encode protein-protein interaction partners of ATR are significantly more likely to validate in additional contexts than genes with no known functional relationship to ATR.

Prioritising robust synthetic lethal interactions involving passenger gene alterations

In addition to alterations of cancer driver genes, tumour cells typically harbour genetic alterations of large numbers of ‘passenger’ genes. Although these genes may not facilitate tumourigenesis or promote cancer cell growth, their alteration may still impart genetic vulnerabilities upon tumour cells. Indeed, multiple synthetic lethal interactions have been identified involving passenger genes that exhibit recurrent copy number loss in cancer cells due to their chromosomal proximity to tumour suppressor genes lost via loss-of-heterozygosity(Kryukov et al., 2016; Marjon et al., 2016; Mavrakis et al., 2016; Muller et al., 2012, 2015). The space of genetic interactions to test involving passenger gene alterations is much larger than that involving driver genes, as nearly every gene in the genome is either mutated or deleted in some cancer context. In addition, passenger genes are typically altered at frequencies lower than for driver genes and therefore the statistical power to identify genetic interactions associated with their alteration is somewhat reduced. With these issues in mind, we reasoned that protein-protein interaction maps might help narrow the search space considerably and thus reduce the burden of multiple hypothesis testing. For all passenger genes that were recurrently lost in at least ten tumour cell lines, either through homozygous deletion or loss-of-function mutation, we searched for genetic interactions with their protein-protein interaction partners using the same discovery and validation approach previously used for driver genes. In total we tested 47,781 interactions involving 2,972 passenger genes and 2,149 selectively lethal genes. To perform an all-against-all test without filtering based on protein-protein interactions would have required more than six million tests, significantly increasing the burden of multiple-hypothesis testing. At an FDR of 20% we found 11 robust genetic interactions involving passenger gene alterations (Supplemental Table S6). Three of these interactions involve genes frequently deleted with the tumour suppressor *CDKN2A (CDKN2B* and *MTAP*) and mirror known associations with *CDKN2A*. A further two genetic interactions involve a single chromosomal

Figure 7

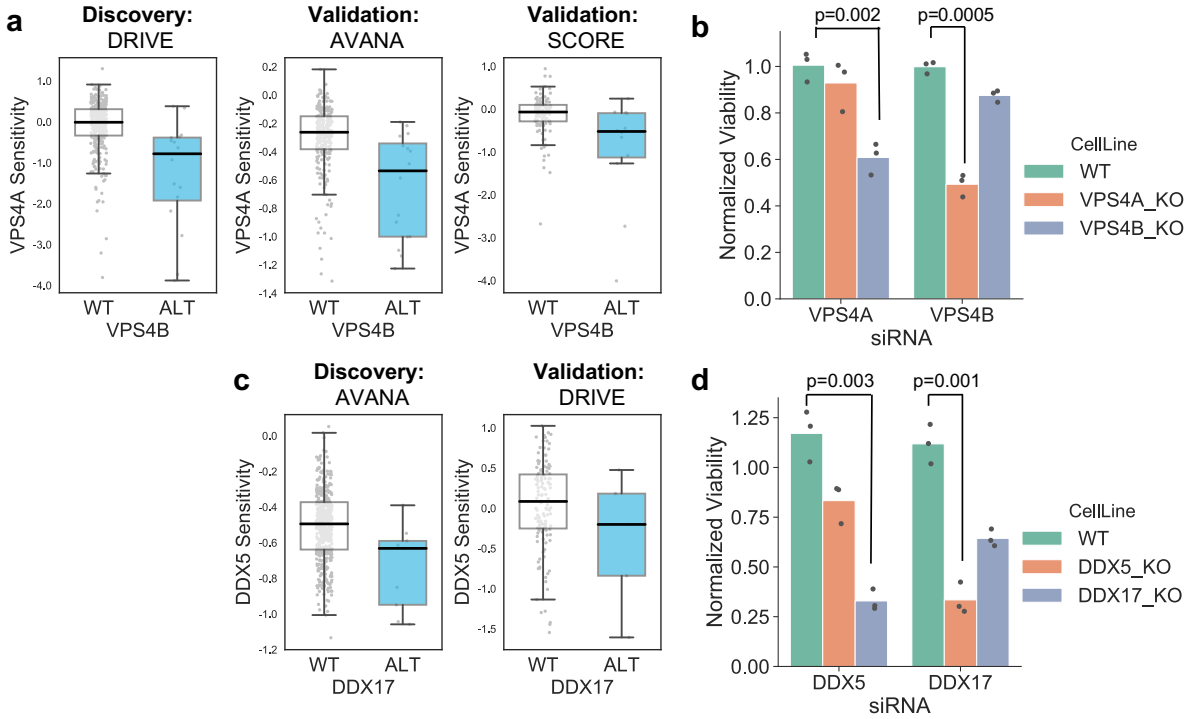


Figure 7. Robust synthetic lethalities associated with passenger gene loss a) Boxplots showing the association between *VPS4B* loss and *VPS4A* sensitivity in the discovery dataset (DRIVE) and two validation datasets (AVANA and SCORE). **b)** Mean viability of HAP1 cells treated with siRNA smartpools targeting *VPS4A* or *VPS4B*. Individual data points are shown as black dots. Data are normalized within each cell line such that the mean viability of cells treated with a negative control (non-targeting scrambled siRNA) is equal to 1 and the mean viability treated with a positive control (siRNA smartpool targeting the broadly essential *PLK1* gene) is equal to 0. P-values from two-sided heteroscedastic T-tests. **c)** Boxplots showing the association between *DDX17* loss and *DDX5* sensitivity in the discovery dataset (AVANA) and the validation dataset (DRIVE). **d)** Mean viability of HAP1 cells treated with siRNA smartpools targeting *DDX5* or *DDX17*, normalization and statistics as per **c**

region (19p21.3) containing two interferon genes (*IFNB1* and *IFNW1*) which are frequently deleted together and consequently these two interactions really represent a single association (an increased sensitivity to thrombopoietin receptor *MPL*). Of the six remaining genetic interactions identified, four represent examples of paralog lethalities – loss of one member of a paralog pair is associated with increased sensitivity to the inhibition of the other member. *RPL22* loss was associated with increased sensitivity to its paralog *RPL22L1*, *TIMM17B* with its paralog *TIMM17A*, *DDX17* with its paralog *DDX5*, and *VPS4B* with its paralog *VPS4A*. We selected two of these robust synthetic lethal interactions for further validation – *VPS4B/VPS4A* and *DDX17/DDX5*. VPS4A and VPS4B are highly sequence similar whole genome duplicates with protein sequence identity of 81%. Both proteins can form a complex with the Vacuolar protein sorting-associated protein VTA1(Huttlin et al., 2017) and are involved in endosomal trafficking. *VPS4B* is located at 18q21 and is frequently deleted with the tumour suppressor *SMAD4*, explaining the relatively high frequency of loss of *VPS4B* in cancer. Previous analysis of the DRIVE shRNA dataset identified an association between *VPS4B* copy number loss and *VPS4A* sensitivity(McDonald et al., 2017). Here we find evidence of this association in two additional datasets – AVANA and

SCORE (Fig. 7a). Although this association is robust, it does not establish a causal link between *VPS4B* loss and *VPS4A* sensitivity. Indeed, there are 39 protein-coding genes on chromosome 18 located between *SMAD4* and *VPS4B*, any one of which could cause sensitivity to *VPS4A* inhibition. To verify that *VPS4B* is the cause of *VPS4A* sensitivity we transfected isogenic knockouts of either *VPS4A* and *VPS4B* with siRNA smartpools targeting either *VPS4A* or *VPS4B* and found that, consistent with a negative genetic interaction between the two genes, compared to wildtype parental cells *VPS4A* knockout cells were sensitive to siRNA targeting *VPS4B* and *VPS4B* knockout cells were sensitive to siRNA targeting *VPS4A* (Fig 7b, Supplemental Table S7).

Like VPS4A and VPS4B, DDX5 and DDX17 are widely conserved highly sequence similar whole genome duplicates (protein sequence identity 69%). They are DEAD box family RNA helicases that have multiple roles in both transcription and splicing; they act as coregulators for multiple transcription factors and also function as components of the spliceosome (Dardenne et al., 2014; Fuller-Pace, 2013). A direct protein-protein interaction between the two genes has also been reported (Hegele et al., 2012; Huttlin et al., 2017). *DDX17* is located at 22q12 in close proximity to the tumour suppressor *MYH9*, potentially explaining its recurrent deletion in tumour cell lines. We identified an association between *DDX17* loss and *DDX5* sensitivity in the AVANA CRISPR dataset and validated this association in the DRIVE shRNA dataset (Fig 7c). As with VPS4A/VPS4B, to verify that *DDX17 loss* is the cause of *DDX5* sensitivity we transfected isogenic knockouts of either *DDX17* and *DDX5* with siRNA smartpools targeting either *DDX5* or *DDX17*. We found that, consistent with a negative genetic interaction between the two genes, compared to wildtype parental cells *DDX17* knockout cells were sensitive to siRNA targeting *DDX5* and *DDX5* knockout cells were sensitive to siRNA targeting *DDX17* (Fig 7d, Supplemental Table S7).

Discussion

While the reproducibility of pharmacogenomic screens in cancer cell lines has been much discussed(Cancer Cell Line Encyclopedia Consortium and Genomics of Drug Sensitivity in Cancer Consortium, 2015; Haibe-Kains et al., 2013; Niepel et al., 2019), relatively little attention has been paid to the reproducibility of results from large-scale genetic screens in cell lines. Analyses of the pharmacogenomic screen datasets have primarily focused on reproducibility in a very strict sense - i.e. quantifying the extent to which the same drug elicits the same response in the same cell line when assayed across different sites(Niepel et al., 2019). In some cases these analyses have been extended to quantify the extent to which the same associations between biomarkers and drugs can be observed across the same cell line panels assayed in different experiments(Cancer Cell Line Encyclopedia Consortium and Genomics of Drug Sensitivity in Cancer Consortium, 2015). Here we were interested in reproducibility in a much broader sense and sought to identify genetic interactions that could be reproduced both across distinct experiments and across distinct cell line panels, i.e. interactions that are robust to genetic and molecular heterogeneity. We developed an approach to identify these robust genetic interactions and used it to identify a set of 220 robust genetic interactions associated with cancer driver genes. We found that these robust genetic interactions are enriched among gene pairs whose protein products physically interact, suggesting a means by which we might prioritise the most promising candidates from screens for follow on studies.

We do not claim that our set of robust genetic interactions is comprehensive, as there are many reasons that real robust genetic interactions may not be identified by our approach. There are many driver genes that we have not included in our analysis because they are infrequently mutated in the datasets studied. Consequently, we can report no interactions for these genes. We have also focussed only on identifying interactions associated with

mutation or copy number changes. There are likely to be dependencies associated with altered gene/protein expression that will be missed by this approach. Furthermore, for the genes that we do analyse, it is likely that some real interactions are not detected due to a lack of statistical power. Finally, of the dependencies identified in a discovery screen but absent in a validation screen, false negatives due to reagents with poor gene targeting ability likely play a significant role(Kaelin, 2012).

We have exclusively focussed on identifying dependencies that are evident across panels of cell lines from multiple cancer types ('pan-cancer dependencies'). It is likely that there are robust dependencies only evident *within* specific cancer types, but it is difficult to use our approach to identify them due to the restricted number of cell lines available for each cancer type. Even with a relatively common mutation (e.g. *KRAS* mutation in non-small cell lung cancer) it is challenging to partition the available cell lines into distinct discovery and validation sets while maintaining statistical power to identify potential dependencies. This issue may be alleviated by efforts to create large numbers of new tumour cell lines(Boehm and Golub, 2015) or through using isogenic models for discovery and cell line panels for validation(Ryan et al., 2018).

Many published synthetic lethal screens have focussed on identifying new drug targets for 'undruggable' oncogenes such as *MYC* and *RAS* (reviewed in (Cermelli et al., 2014) and (Downward, 2015) respectively). The rationale for such studies is that the oncogene addiction itself cannot be exploited directly and consequently a synthetic lethal approach is needed. However, here we found that for all oncogenes studied the most significant reproducible dependency identified was an oncogene addiction (Fig. 3a). This suggests that any synthetic lethal interactions that are identified for oncogenes will likely be of a smaller effect size or operate in a more restricted context than the oncogene addiction itself. Previous work has suggested this to be true of *KRAS*(McDonald et al., 2017) but here we find that it appears to be a general property of all oncogenes studied. This implies that wherever possible, direct targeting of oncogenes might be more therapeutically effective than exploiting oncogene-related synthetic lethal effects

Our approach to identify robust genetic dependencies involving cancer driver genes is unbiased in the sense that we did not incorporate prior knowledge of functional relationships to identify candidate gene pairs to test. Nonetheless, many of the robust synthetic lethalities identified reflect known biology. In particular, for each of the well-studied tumour suppressors *ARID1A, SMARCA4* and *PTEN* the most significant robust synthetic lethal interaction we identified has previously been reported in the literature. For *ARID1A*, its known synthetic lethal partner *ARID1B* was the only robust candidate interaction identified while for *PTEN* and *SMARCA4* their established synthetic lethal partners (*PIK3CB* and *SMARCA2* respectively) are the most significant robust hits by a large margin (Fig. 3c). As with oncogenes, this suggests that if novel single gene vulnerabilities for these drivers are to be discovered, they may have a smaller effect size or operate in a more restricted setting.

Previous work has shown that genetic interactions between gene pairs whose protein products physically interact are more highly conserved across species(Roguev et al., 2008; Ryan et al., 2012; Srivas et al., 2016). Our analysis here suggests that the same principles may be used to identify genetic interactions conserved across genetically heterogeneous tumour cell lines. Although we have not tested them here, other features predictive of between-species conservation may also be predictive of robustness to genetic heterogeneity(Ryan et al., 2012; Srivas et al., 2016). Our set of robust genetic interactions may serve as the starting point for such analyses and may also serve as a training set for computational approaches to predict synthetic lethality(Jerby-Arnon et al., 2014).

Our results suggest that knowledge of protein-protein interactions could be used to improve the design and analysis of loss-of-function screens for synthetic lethal interactions. We have demonstrated the utility of incorporating such prior knowledge for identifying robust synthetic lethal interactions from genome-wide chemogenetic screens. We have also demonstrated that protein-protein interactions can aid the identification of genetic interactions associated with passenger gene alterations, where statistical power is limited due to relatively infrequent alterations and the number of potential interactions to test is enormous. An alternative to these approaches, where knowledge of protein-protein interactions is used after the screen has already been performed, would be to screen target libraries for specific driver genes based on their known protein interaction partners. Regardless of the approach used to identify candidate synthetic lethal interactions in a large-scale screen, our results suggest that candidates supported by a protein-protein interaction should be prioritised for follow on study as they are more likely to be robust to the genetic heterogeneity observed in tumour cells.

Methods

All data analysis was performed using Python 3.7, Pandas 0.24(McKinney, 2011) and StatsModels 0.9.0(Seabold and Perktold, 2010).

Loss of function screens

Different scoring systems have been developed for calculating 'gene level' dependency scores from loss-of-function screens performed with multiple gene targeting reagents per gene (i.e. shRNAs or gRNAs). For the analysis of all loss-of-function screens we used the original authors' own preferred approaches. CERES dependency scores(Meyers et al., 2017) for AVANA (release 18Q4) were obtained from the DepMap portal (https://depmap.org/portal/download/), while DEMETER v2 gene dependency scores for the DEPMAP shRNA screen(Tsherniak et al., 2017) were obtained from the same resource. For the DEPMAP screen, some genes were only screened in a subset of cell lines and these were excluded from all analyses. Quantile normalized CRISPRcleaned(Iorio et al., 2018) depletion log fold changes for Project SCORE(Behan et al., 2019) were obtained from the Project SCORE database (https://score.depmap.sanger.ac.uk/). ATARIS(Shao et al., 2013) scores for the DRIVE dataset(McDonald et al., 2017) were obtained from the authors. 28 of the 398 cell lines screened in DRIVE had missing gene scores for ~25% of genes screened and these cell lines were excluded from further analysis. All screens were mapped to a common cell line name format (that followed by the Cancer Cell Line Encyclopaedia(Barretina et al., 2012)) using the Cell Model Passports resource where appropriate(van der Meer et al., 2019).

Identifying selectively lethal genes

Similar to previous work(McDonald et al., 2017; Tsherniak et al., 2017), to reduce the burden of multiple hypothesis testing we focused our analysis on genes whose inhibition appeared to cause growth defects in subsets of the cancer cell lines screened. That is, rather than testing for associations with genes whose inhibition was always lethal or never lethal, we focused our analyses on genes that could be associated with distinct sensitive and resistant cell line cohorts. We first identified a set of 'selectively lethal' genes using the Avana dataset(Meyers et al., 2017) - those with a gene dependency score <-0.6 in at least 10 cell lines but no more than 259 cell lines (half of the screened cell lines). We augmented this with a list of 65 'outlier genes' identified by the authors of the DRIVE study as having a skewed distribution suggesting distinct sensitive and resistant cohorts(McDonald et al., 2017). Finally from the combined list we removed genes known to be commonly essential in cancer cell lines(Hart et al., 2017). This resulted in a set of 2,470 selectively lethal genes (Supplemental Table S1) which were used for all association analyses.

Identifying gene alterations from copy number and exome profiling

For all cell lines we obtained sequencing data (CCLE_DepMap_18q3_maf_20180718.txt) and copy number profiles (public_18Q3_gene_cn_v2.csv) from the DepMap portal. These datasets contain integrated genotyping data from both the Cancer Cell Line Encyclopedia and GDSC resources(Barretina et al., 2012; Cancer Cell Line Encyclopedia Consortium and Genomics of Drug Sensitivity in Cancer Consortium, 2015; Iorio et al., 2016). We used this to identify likely functional alterations in a panel of cancer driver genes(Vogelstein et al., 2013) restricting our analysis to those genes that were subject to targeted sequencing as part of the Cancer Cell Line Encyclopedia(Barretina et al., 2012).

For most oncogenes we considered the gene to be functionally altered if it contained a protein altering mutation at a residue that is recurrently altered in either the COSMIC database or the Cancer Genome Atlas. For a small number of oncogenes (*ERBB2, CCND1,*

MDM2, MDM4) we considered them to be functionally altered only if they were amplified. For all tumour suppressors we considered all protein-coding mutations and homozygous deletions to be functional alterations. The matrix of functional alterations to driver genes is presented in Supplemental Table S2.

To identify loss-of-function alterations to passenger genes, a similar pipeline was used. However the driver genes from Vogelstein *et al*(Vogelstein et al., 2013) were excluded from analysis and only clear loss-of-function alterations were considered to be functional. The matrix of gene loss in passenger genes is presented in Supplemental Table S5.

Identifying genetic dependencies in individual datasets

We wished to identify associations between driver gene mutations and gene sensitivity scores that were not confounded by tissue specific gene sensitivity effects (e.g. SOX10 sensitivity scores can be naively associated with *BRAF* mutational status because SOX10 is essential in melanoma cell lines and *BRAF* mutation is common in melanoma). Thus, we wished to model gene sensitivity after first accounting for tissue type. To this end, associations between individual driver genes and gene sensitivity scores were identified using an ANOVA model that incorporated both tissue type and mutational status as covariates, similar to the method previously developed for identifying pharmacogenomic interactions in cancer cell line panels(Cokelaer et al., 2018; Iorio et al., 2016). As recent work(Behan et al., 2019) has highlighted that some dependencies (e.g. *WRN*) can be associated with microsatellite instability rather than individual driver genes, we also incorporated microsatellite instability(Ghandi et al., 2019) as a covariate in our model. The model had the form 'gene_X_sensitivity ~ MSI_status + C(Tissue) + driver_gene_Y_status' and was used to test the association between each recurrently mutated driver gene Y and all gene sensitivity scores X assayed in a given dataset. Driver genes were included in this analysis if they were functionally altered in at least five cell lines in the dataset being analysed. Correction for multiple hypothesis testing was performed using the Benjamini and Hochberg false discovery rate(Benjamini and Hochberg, 1995).

Identifying genetic dependencies common to multiple datasets

When comparing a pair of datasets, we used one dataset as a discovery dataset and a second as a validation set, as outlined in Figure 1C. The discovery analysis was limited to the set of interactions that could be tested in both datasets, i.e. associations between the set of sensitivity scores for genes screened in both studies and the set of driver genes recurrently altered in both studies. An initial set of genetic interactions was identified in the discovery dataset at a specific FDR threshold and these associations were then tested in the validation set. We considered interactions to be reproduced in the validation dataset if: (1) the FDR was less than the threshold; (2) the uncorrected p-value was < 0.05 and; (3) the sign of the association (sensitivity / resistance) was the same in both discovery and validation set. A FDR of 0.2 was used for all analysis presented in the main text but additional FDR thresholds (0.1, 0.3) were tested to ensure that all findings were robust to the exact choice of FDR (Supplemental Fig. S2).

Protein-protein interactions

Protein-protein interactions were obtained from STRING v10.5 (Szklarczyk et al., 2015), BIOGRID 3.5.170 (Chatr-aryamontri et al., 2014) and from HIPPIE v2.0(Alanis-Lobato et al., 2017). Results in the main text make use of medium confidence STRING interactions (STRING integrated score >0.4). However, to ensure robustness to the thresholds shown, all analyses were repeated for the full set of HIPPIE interactions and the full set of BioGRID interactions (Supplemental Fig. S2).

siRNA experiments

HAP1 cell lines were obtained from Horizon Discovery: HAP1_WT (C631), HAP1_VPS4A_KO (HZGHC004623c005), HAP1_VPS4B_KO (HZGHC006889c011), HAP1_DDX5_KO (HZGHC006136c012) and HAP1_DDX17_KO(HZGHC007221c009). Cells were cultured in IMDM (10-016-CV; Corning) supplemented with 10% FBS (10270-106; Thermo Fisher Scientific). ON-TARGETplus siRNA SMARTpools targeting *VPS4A* (L-013092-00-0005), *VPS4B* (L-013119-00-0005), *DDX5* (L-003774-00-0005), *DDX17* (L-013450-01-0005), *PLK1* (L-003290-00-0005) and a non-targeting scramble control (D-001810-10-20) were obtained from Dharmacon. HAP1 cells were seeded to a density of 5000 cells per well of a 96-well plate and 5nM siRNA was transfected with Lipofectamine 3000 (L3000015; Thermo Fisher Scientific) in Opti-MEM I Reduced Serum Medium (31985070; Thermo Fisher Scientific). Cell viability was measured 72 hours after siRNA transfection using CellTiter-Glo Luminescent Cell Viability Assay (G7570; Promega). The 96 well plates were read using a SpectraMax® M3 Microplate Reader (Molecular devices). Viability effects for each siRNA targeting each gene X in each cell line y were normalised using the following formula:

$$\text{NormalisedViability}\ (\text{siRNA_X}_y) = 1 - (\text{siCTRL}_y - \text{siRNA_X}_y) / (\text{siCTRL}_y - \text{siPLK1}_y)$$

where siCTRL_y is the average of 3 measurements of non-targeting scramble control in cell line y and siPLK1_y is the average of 3 measurements of an siRNA smartpool targeting *PLK1* in cell line y. Raw and normalised viability data are in Supplemental Table S7.

Acknowledgements

CJL's work on this study was funded by Breast Cancer Now as part of Programme Funding to the Breast Cancer Now Toby Robins Research Centre (CTR-Q4-Y2) and via a Cancer Research UK Programme Grant (CRUK/A14276). CJR's work on this study was funded by the HRB/SFI/Wellcome Trust partnership (grant number 103049/Z/13/Z) and by the Irish Research Council Laureate Awards 2017/2018. We thank members of the Ryan and Lord labs for helpful comments and discussions.

References

Alanis-Lobato, G., Andrade-Navarro, M.A., and Schaefer, M.H. (2017). HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. Nucleic Acids Res. *45*, D408–D414.

Asghar, U., Witkiewicz, A.K., Turner, N.C., and Knudsen, E.S. (2015). The history and future of targeting cyclin-dependent kinases in cancer therapy. Nat. Rev. Drug Discov. *14*, 130–146.

Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature *483*, 603–607.

Behan, F.M., Iorio, F., Picco, G., Gonçalves, E., Beaver, C.M., Migliardi, G., Santos, R., Rao, Y., Sassi, F., Pinnelli, M., et al. (2019). Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. Nature *568*, 511–516.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J. R. Stat. Soc. Series B Stat. Methodol. *57*, 289–300.

Boehm, J.S., and Golub, T.R. (2015). An ecosystem of cancer cell line factories to support a cancer dependency map. Nat. Rev. Genet. *16*, 373–374.

Bridgett, S., Campbell, J., Lord, C.J., and Ryan, C.J. (2017). CancerGD: A Resource for Identifying and Interpreting Genetic Dependencies in Cancer. Cell Syst *5*, 82-86.e3.

Brough, R., Gulati, A., Haider, S., Kumar, R., Campbell, J., Knudsen, E., Pettitt, S., Ryan, C.J., and Lord, C.J. (2018). Identification of highly penetrant Rb-related synthetic lethal interactions in triple negative breast cancer. Oncogene *37*, 5701–5718.

Campbell, J., Ryan, C.J., Brough, R., Bajrami, I., Pemberton, H.N., Chong, I.Y., Costa-Cabral, S., Frankum, J., Gulati, A., Holme, H., et al. (2016). Large-Scale Profiling of Kinase Dependencies in Cancer Cell Lines. Cell Rep. *14*, 2490–2501.

Cancer Cell Line Encyclopedia Consortium, and Genomics of Drug Sensitivity in Cancer Consortium (2015). Pharmacogenomic agreement between two cancer cell line data sets. Nature *528*, 84–87.

Cermelli, S., Jang, I.S., Bernard, B., and Grandori, C. (2014). Synthetic lethal screens as a means to understand and treat MYC-driven cancers. Cold Spring Harb. Perspect. Med. *4*.

Chapman, P.B., Hauschild, A., Robert, C., Haanen, J.B., Ascierto, P., Larkin, J., Dummer, R., Garbe, C., Testori, A., Maio, M., et al. (2011). Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N. Engl. J. Med. *364*, 2507–2516.

Chatr-aryamontri, A., Breitkreutz, B.-J., Oughtred, R., Boucher, L., Heinicke, S., Chen, D., Stark, C., Breitkreutz, A., Kolas, N., O'Donnell, L., et al. (2014). The BioGRID interaction database: 2015 update. Nucleic Acids Res.

Cokelaer, T., Chen, E., Iorio, F., Menden, M.P., Lightfoot, H., Saez-Rodriguez, J., and Garnett, M.J. (2018). GDSCTools for mining pharmacogenomic interactions in cancer. Bioinformatics *34*, 1226–1228.

Dardenne, E., Polay Espinoza, M., Fattet, L., Germann, S., Lambert, M.-P., Neil, H., Zonta, E., Mortada, H., Gratadou, L., Deygas, M., et al. (2014). RNA helicases DDX5 and DDX17 dynamically orchestrate transcription, miRNA, and splicing programs in cell differentiation. Cell Rep. *7*, 1900–1913.

Downward, J. (2015). RAS Synthetic Lethal Screens Revisited: Still Seeking the Elusive Prize? Clin. Cancer Res. *21*, 1802–1809.

Formstecher, E., Ramos, J.W., Fauquet, M., Calderwood, D.A., Hsieh, J.-C., Canton, B., Nguyen, X.-T., Barnier, J.-V., Camonis, J., Ginsberg, M.H., et al. (2001). PEA-15 Mediates Cytoplasmic Sequestration of ERK MAP Kinase. Dev. Cell *1*, 239–250.

Fuller-Pace, F.V. (2013). DEAD box RNA helicase functions in cancer. RNA Biol. *10*, 121–132.

Ghandi, M., Huang, F.W., Jané-Valbuena, J., Kryukov, G.V., Lo, C.C., McDonald, E.R., 3rd, Barretina, J., Gelfand, E.T., Bielski, C.M., Li, H., et al. (2019). Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature.

Giacomelli, A.O., Yang, X., Lintner, R.E., McFarland, J.M., Duby, M., Kim, J., Howard, T.P., Takeda, D.Y., Ly, S.H., Kim, E., et al. (2018). Mutational processes shape the landscape of TP53 mutations in human cancer. Nat. Genet. *50*, 1381–1387.

Haibe-Kains, B., El-Hachem, N., Birkbak, N.J., Jin, A.C., Beck, A.H., Aerts, H.J.W.L., and Quackenbush, J. (2013). Inconsistency in large pharmacogenomic studies. Nature *504*, 389–393.

Hart, T., Tong, A.H.Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. G3 *7*, 2719–2727.

Hegele, A., Kamburov, A., Grossmann, A., Sourlis, C., Wowro, S., Weimann, M., Will, C.L., Pena, V., Lührmann, R., and Stelzl, U. (2012). Dynamic protein-protein interaction wiring of the human spliceosome. Mol. Cell *45*, 567–580.

Helming, K.C., Wang, X., Wilson, B.G., Vazquez, F., Haswell, J.R., Manchester, H.E., Kim, Y., Kryukov, G.V., Ghandi, M., Aguirre, A.J., et al. (2014). ARID1B is a specific vulnerability in ARID1A-mutant cancers. Nat. Med. *20*, 251–254.

Henkel, L., Rauscher, B., and Boutros, M. (2019). Context-dependent genetic interactions in cancer. Curr. Opin. Genet. Dev. *54*, 73–82.

Herrero, A., Pinto, A., Colón-Bolea, P., Casar, B., Jones, M., Agudo-Ibáñez, L., Vidal, R., Tenbaum, S.P., Nuciforo, P., Valdizán, E.M., et al. (2015). Small Molecule Inhibition of ERK Dimerization Prevents Tumorigenesis by RAS-ERK Pathway Oncogenes. Cancer Cell *28*, 170–182.

Hill, K.S., Roberts, E.R., Wang, X., Marin, E., Park, T.D., Son, S., Ren, Y., Fang, B., Yoder, S., Kim, S., et al. (2019). PTPN11 Plays Oncogenic Roles and Is a Therapeutic Target for BRAF Wild-Type Melanomas. Mol. Cancer Res. *17*, 583–593.

Hoffman, G.R., Rahal, R., Buxton, F., Xiang, K., McAllister, G., Frias, E., Bagdasarian, L., Huber, J., Lindeman, A., Chen, D., et al. (2014). Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. Proc. Natl. Acad. Sci. U. S. A. *111*, 3128–3133.

Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. Nature *545*, 505–509.

Iorio, F., Knijnenburg, T.A., Vis, D.J., Bignell, G.R., Menden, M.P., Schubert, M., Aben, N., Gonçalves, E., Barthorpe, S., Lightfoot, H., et al. (2016). A Landscape of Pharmacogenomic Interactions in Cancer. Cell *166*, 740–754.

Iorio, F., Behan, F.M., Gonçalves, E., Bhosle, S.G., Chen, E., Shepherd, R., Beaver, C., Ansari, R., Pooley, R., Wilkinson, P., et al. (2018). Unsupervised correction of gene-independent cell responses to CRISPR-Cas9 targeting. BMC Genomics *19*, 604.

Jerby-Arnon, L., Pfetzer, N., Waldman, Y.Y., McGarry, L., James, D., Shanks, E., Seashore-Ludlow, B., Weinstock, A., Geiger, T., Clemons, P.A., et al. (2014). Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. Cell *158*, 1199–1209.

Ji, P., Jiang, H., Rekhtman, K., Bloom, J., Ichetovkin, M., Pagano, M., and Zhu, L. (2004). An Rb-Skp2-p27 pathway mediates acute cell cycle inhibition by Rb and is retained in a partial-penetrance Rb mutant. Mol. Cell *16*, 47–58.

Kaelin, W.G., Jr (2012). Molecular biology. Use and abuse of RNAi to study mammalian gene function. Science *337*, 421–422.

Kryukov, G.V., Wilson, F.H., Ruth, J.R., Paulk, J., Tsherniak, A., Marlow, S.E., Vazquez, F., Weir, B.A., Fitzgerald, M.E., Tanaka, M., et al. (2016). MTAP deletion confers enhanced dependency on the PRMT5 arginine methyltransferase in cancer cells. Science *351*, 1214–1218.

Lord, C.J., and Ashworth, A. (2017). PARP inhibitors: Synthetic lethality in the clinic. Science *355*, 1152–1158.

Marcotte, R., Sayad, A., Brown, K.R., Sanchez-Garcia, F., Reimand, J., Haider, M., Virtanen, C., Bradner, J.E., Bader, G.D., Mills, G.B., et al. (2016). Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. Cell *164*, 293–309.

Marjon, K., Cameron, M.J., Quang, P., Clasquin, M.F., Mandley, E., Kunii, K., McVay, M., Choe, S., Kernytsky, A., Gross, S., et al. (2016). MTAP Deletions in Cancer Create Vulnerability to Targeting of the MAT2A/PRMT5/RIOK1 Axis. Cell Rep. *15*, 574–587.

Martin, T.D., Cook, D.R., Choi, M.Y., Li, M.Z., Haigis, K.M., and Elledge, S.J. (2017). A Role for Mitochondrial Translation in Promotion of Viability in K-Ras Mutant Cells. Cell Rep. *20*, 427–438.

Mavrakis, K.J., McDonald, E.R., 3rd, Schlabach, M.R., Billy, E., Hoffman, G.R., deWeck, A., Ruddy, D.A., Venkatesan, K., Yu, J., McAllister, G., et al. (2016). Disordered methionine metabolism in MTAP/CDKN2A-deleted cancers leads to dependence on PRMT5. Science *351*, 1208–1213.

McDonald, E.R., 3rd, de Weck, A., Schlabach, M.R., Billy, E., Mavrakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., et al. (2017). Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. Cell *170*, 577-592.e10.

McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. Python for High Performance and Scientific Computing 1–9.

van der Meer, D., Barthorpe, S., Yang, W., Lightfoot, H., Hall, C., Gilbert, J., Francies, H.E., and Garnett, M.J. (2019). Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. Nucleic Acids Res. *47*, D923–D929.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. Nat. Genet. *49*, 1779–1784.

Muller, F.L., Colla, S., Aquilanti, E., Manzo, V.E., Genovese, G., Lee, J., Eisenson, D., Narurkar, R., Deng, P., Nezi, L., et al. (2012). Passenger deletions generate therapeutic vulnerabilities in cancer. Nature *488*, 337–342.

Muller, F.L., Aquilanti, E.A., and DePinho, R.A. (2015). Collateral Lethality: A new therapeutic strategy in oncology. Trends Cancer Res. *1*, 161–173.

Niepel, M., Hafner, M., Mills, C.E., Subramanian, K., Williams, E.H., Chung, M., Gaudio, B., Barrette, A.M., Stern, A.D., Hu, B., et al. (2019). A Multi-center Study on the Reproducibility of Drug-Response Assays in Mammalian Cell Lines. Cell Syst *9*, 35-48.e5.

Oike, T., Ogiwara, H., Tominaga, Y., Ito, K., Ando, O., and Tsuta, K. (2013). A synthetic lethality–based strategy to treat cancers harboring a genetic deficiency in the chromatin remodeling factor BRG1. Cancer Res.

O'Leary, B., Cutts, R.J., Liu, Y., Hrebien, S., Huang, X., Fenwick, K., André, F., Loibl, S., Loi, S., Garcia-Murillas, I., et al. (2018). The Genetic Landscape and Clonal Evolution of Breast Cancer Resistance to Palbociclib plus Fulvestrant in the PALOMA-3 Trial. Cancer Discov. *8*, 1390–1403.

Roguev, A., Bandyopadhyay, S., Zofall, M., Zhang, K., Fischer, T., Collins, S.R., Qu, H., Shales, M., Park, H.-O., Hayles, J., et al. (2008). Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. Science *322*, 405–410.

Ryan, C.J., Roguev, A., Patrick, K., Xu, J., Jahari, H., Tong, Z., Beltrao, P., Shales, M., Qu, H., Collins, S.R., et al. (2012). Hierarchical modularity and the evolution of genetic interactomes across species. Mol. Cell *46*, 691–704.

Ryan, C.J., Bajrami, I., and Lord, C.J. (2018). Synthetic Lethality and Cancer - Penetrance as the Major Barrier. Trends Cancer Res. *4*, 671–683.

Seabold, S., and Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. In Proceedings of the 9th Python in Science Conference, (SciPy society Austin), p. 61.

Shao, D.D., Tsherniak, A., Gopal, S., Weir, B.A., Tamayo, P., Stransky, N., Schumacher, S.E., Zack, T.I., Beroukhim, R., Garraway, L.A., et al. (2013). ATARiS: computational quantification of gene suppression phenotypes from multisample RNAi screens. Genome Res. *23*, 665–678.

Srivas, R., Shen, J.P., Yang, C.C., Sun, S.M., Li, J., Gross, A.M., Jensen, J., Licon, K., Bojorquez-Gomez, A., Klepper, K., et al. (2016). A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. Mol. Cell *63*, 514–525.

Steckel, M., Molina-Arcas, M., Weigelt, B., Marani, M., Warne, P.H., Kuznetsov, H., Kelly, G., Saunders, B., Howell, M., Downward, J., et al. (2012). Determination of synthetic lethal interactions in KRAS oncogene-dependent cancer cells reveals novel therapeutic targeting strategies. Cell Res. *22*, 1227–1245.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., Simonovic, M., Roth, A., Santos, A., Tsafou, K.P., et al. (2015). STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. *43*, D447-52.

Tadesse, S., Caldon, E., Tilley, W., and Wang, S. (2018). Cyclin Dependent Kinase 2 Inhibitors in Cancer Therapy: an Update. J. Med. Chem.

Tsherniak, A., Vazquez, F., Montgomery, P.G., Weir, B.A., Kryukov, G., Cowley, G.S., Gill, S., Harrington, W.F., Pantel, S., Krill-Burger, J.M., et al. (2017). Defining a Cancer Dependency Map. Cell *170*, 564-576.e16.

Vendetti, F.P., Lau, A., Schamus, S., Conrads, T.P., O'Connor, M.J., and Bakkenist, C.J. (2015). The orally active and bioavailable ATR kinase inhibitor AZD6738 potentiates the anti-

tumor effects of cisplatin to resolve ATM-deficient non-small cell lung cancer in vivo. Oncotarget *6*, 44289–44305.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr, and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

Wang, C., Wang, G., Feng, X., Shepherd, P., Zhang, J., Tang, M., Chen, Z., Srivastava, M., McLaughlin, M.E., Navone, N.M., et al. (2018). Genome-wide CRISPR screens reveal synthetic lethality of RNASEH2 deficiency and ATR inhibition. Oncogene.

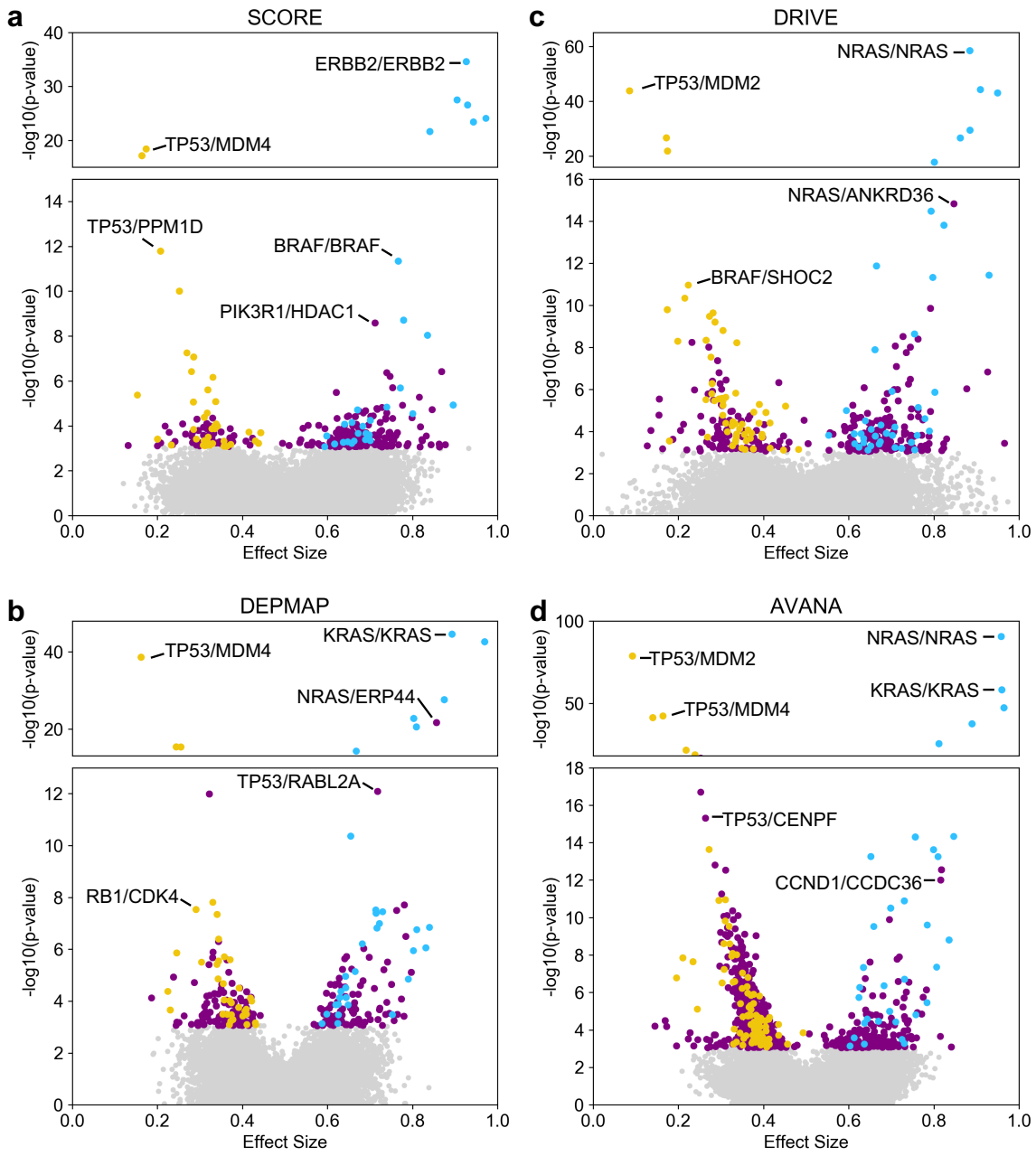
Wee, S., Wiederschain, D., Maira, S.-M., Loo, A., Miller, C., deBeaumont, R., Stegmeier, F., Yao, Y.-M., and Lengauer, C. (2008). PTEN-deficient cancers depend on PIK3CB. Proc. Natl. Acad. Sci. U. S. A. *105*, 13057–13062.

Wilson, B.G., and Roberts, C.W.M. (2011). SWI/SNF nucleosome remodellers and cancer. Nat. Rev. Cancer *11*, 481–492.

Xu, X.L., Singh, H.P., Wang, L., Qi, D.-L., Poulos, B.K., Abramson, D.H., Jhanwar, S.C., and Cobrinik, D. (2014). Rb suppresses human cone-precursor-derived retinoblastoma tumours. Nature *514*, 385–388.

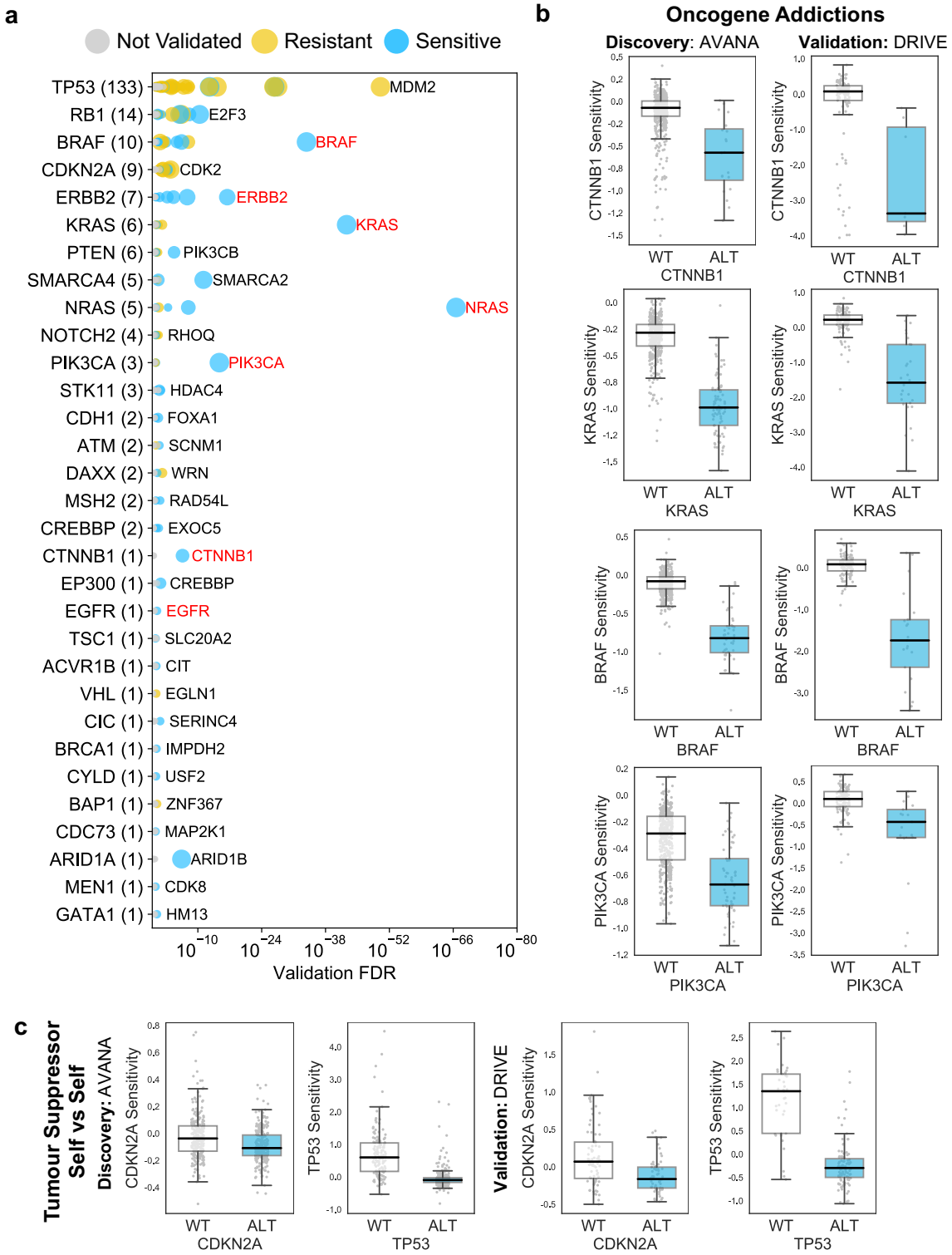
Supplemental Figures

Supplemental Figure S1



Supplemental Figure S1. Discovered and validated genetic dependencies for individual datasets. a) Scatterplot showing the genetic dependencies identified in the SCORE dataset. Each individual point represents a gene pair, the x-axis shows the common language effect size, and the y-axis shows the -log10 p-value from the discovery dataset. **b)** Scatterplot for DEPMAP dataset. **c)** Scatterplot for DRIVE dataset. **d)** Scatterplot for AVANA dataset.

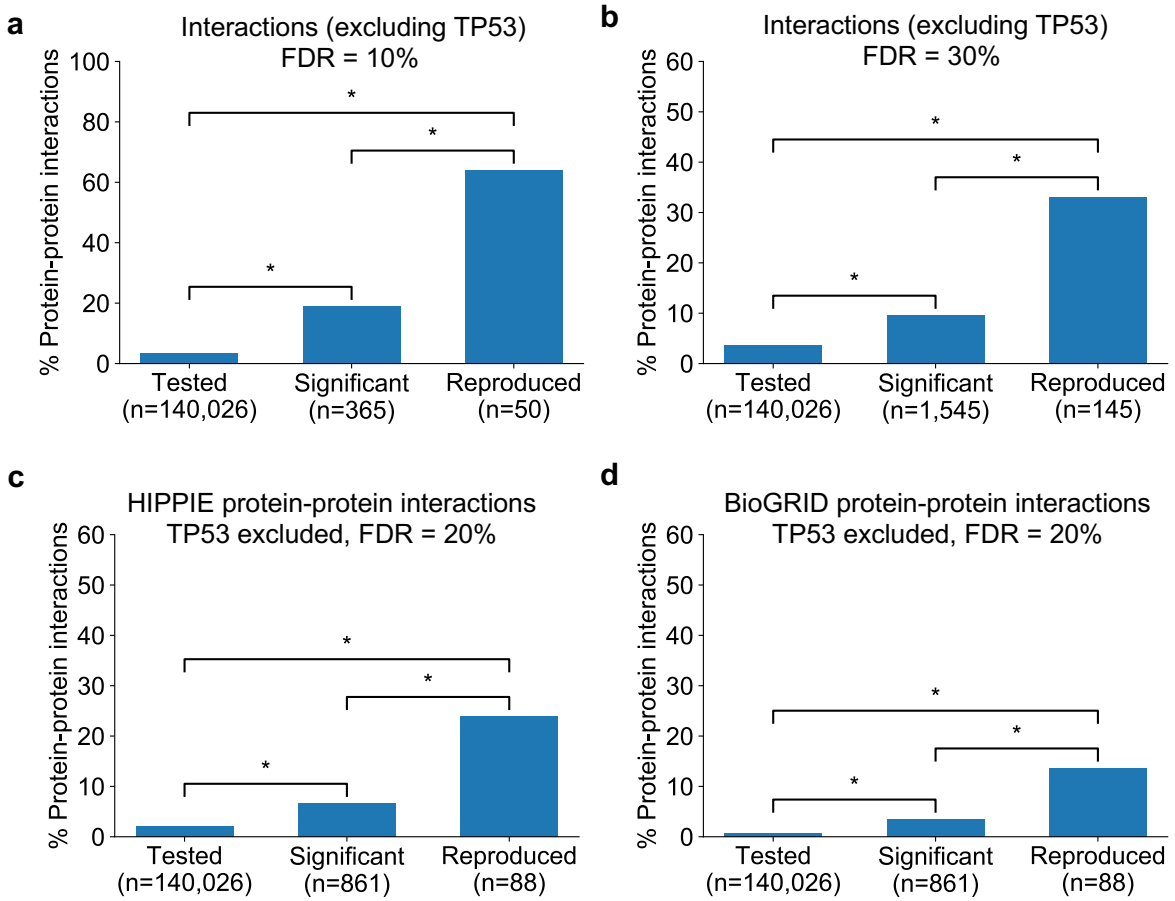
Supplemental Figure S2



Supplemental Figure S2. Reproducible genetic dependencies include oncogene addictions. a) Dot plot showing the reproducible genetic dependencies identified. Each coloured circle indicates a reproducible genetic dependency, scaled according to the number of dataset pairs it was validated in. The most significant genetic dependency (lowest FDR in a validation set) for each driver gene is labelled. Instances where the most significant dependency is a 'self *vs* self' dependency are highlighted in red. Drivers are sorted by the

number of validated dependencies and the total number of reproduced genetic dependencies for each driver gene is shown in parentheses. **b)** Boxplots showing oncogene addictions, where the alteration of an oncogene is associated with increased sensitivity to its inhibition. **c)** Boxplots showing tumour suppressor genes whose inhibition provides a growth advantage to cells that have no genetic alteration of those genes.

Supplemental Figure S3



Supplemental Figure S3. Robust genetic interactions are enriched in protein-protein interaction pairs at different thresholds and using different databases a) Barchart showing the percentage of protein-protein interacting pairs observed among different groups of gene pairs. The groups represent all gene pairs tested, gene pairs found to be significantly interacting in at least one screen (FDR < 10%), and gene pairs found to reproducibly interact across multiple screens (i.e. a discovery and validation screen). Stars (*) indicate significant differences between groups, all significant at P<0.001 using Fisher's Exact Test. Due to the high percentage of protein-protein interaction pairs among the reproducible hits at this FDR, the y-axis uses a different maximum value to all other charts. **b)** Same as a but with interactions identified at an FDR of 30% **c)** Similar to main text Fig. 4b but here the protein-protein interaction pairs are obtained from the HIPPIE database **d)** Similar to main text Fig. 4b but here the protein-protein interaction pairs are obtained from the BioGRID database.

Supplemental Tables

Supplemental Table S1 – Selectively lethal genes
Supplemental Table S2 – Driver gene alterations
Supplemental Table S3 – Reproducible genetic dependencies
Supplemental Table S4 – Protein-protein interaction enrichment
Supplemental Table S5 – Passenger gene loss alterations
Supplemental Table S6 – Passenger gene dependencies
Supplemental Table S7 – VPS4A_VPS4B and DDX5_DDX17 viability data