Robust genetic interactions in cancer are enriched in protein-protein interaction pairs

Christopher J. Lord[1], 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

Keywords = synthetic lethality; cancer; genetic interactions; protein-protein interactions

Abstract

Genetic interactions, where the mutation of one gene can be associated with altered sensitivity to the inhibition of another gene, can be exploited for the development of targeted therapies in cancer. Many large-scale genetic perturbation screens in tumour cell lines have been used to identify such genetic interactions, including both synthetic lethal and resistance relationships. Despite these efforts, relatively few of the candidate genetic interactions identified have been reproduced across multiple studies. Here, we develop a computational approach to identify genetic interactions that can be reproduced across independent experiments and across non-overlapping cell line panels. We identified 220 such robust genetic interactions and found that they are enriched for gene pairs whose protein products physically interact. This suggests that protein-protein interaction networks may be used not only to understand the mechanistic basis of genetic interaction effects, but also to prioritise robust candidates for further development.

Introduction

Large scale tumour genome sequencing efforts have provided us with lists of driver genes that are recurrently altered in tumours[1]. In a small number of cases, these genetic alterations have been associated with altered sensitivity to targeted therapies in cancer. Examples of such targeted therapies already in clinical use include approaches that exploit oncogene addictions, such as the increased sensitivity of *BRAF* mutant melanomas to BRAF inhibitors[2], and approaches that exploit non-oncogene addiction or synthetic lethality, such as the sensitivity of *BRCA1*/*2* mutant ovarian or breast cancers to PARP inhibitors[3]. An ongoing challenge is to associate the presence of other driver gene alterations with sensitivity to existing therapeutic agents[4,5] or to identify additional proteins whose inhibition may provide therapeutic benefit to patients with specific mutations (i.e. new drug targets). 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)[6–11]. Others have performed screens in 'isogenic' cell line pairs that differ only by the presence of a specific oncogenic alteration[12,13]. Despite these large-scale efforts, very few genetic interactions have been identified in more than one study (recently reviewed[14]). 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[15,16]. This lack of reproducibility may be due to technical issues, e.g. false positives

and false negatives due to inefficient gene targeting reagents[17], and/or real biological issues, such as the context specificity of genetic interactions[14]. 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 220 such 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.

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[6], DEPMAP[7]) and two CRISPR-Cas9 mutagenesis screens (AVANA[8], SCORE[11]). We harmonised the cell line names across all studies, so they could be compared with each other and also with genotypic data[4,5]. 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 (Figure 1A). 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 associations in the validation study (Figure 1B). 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 present in the discovery dataset. 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.

Similar to our previous work[9,18] 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[1] (see methods, Supplementary Table 1). 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, Supplementary Table 2). We analysed each pair of screens in turn and considered a genetic association to be reproducible if it was validated in at least one discovery/validation pair. Using this approach, we identified 229 reproducible genetic associations (Supplementary Table 3, Supplementary Figure 1).

Of these genetic associations 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 indicated oncogene addictions, e.g. *BRAF* mutant cell lines were sensitive to *BRAF* inhibition and *ERBB2* amplified cell lines were sensitive to *ERBB2* inhibition (Figure 1A). Similarly, we also identified robust 'self *vs.* self' associations involving the *CTNNB1* (β-Catenin), *KRAS*, *NRAS*, *EGFR* and *PIK3CA* oncogenes (Supplementary Figure 1B). However, we also identified two examples of 'self *vs.* self' dependencies involving tumour suppressors -*TP53* (aka p53) and *CDKN2A* (aka p16/p14arf) (Supplementary Figure 1B). This type of relationship has previously been reported for *TP53* - *TP53* inhibition appears to offer a growth advantage to *TP53* wild type cells but not to *TP53* mutant cells[19]. Consequently, we observed an association between *TP53* status and sensitivity to TP53 inhibition. Similar effects were seen for *CDKN2A* (Supplementary Figure 1C). 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 (Figure 2A).

Many robust genetic interactions reflect known pathway structure

A number 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*[20]) and resistance relationships, such as an increased resistance of *TP53* mutant cell lines to *MDM2* inhibition (Figure 2A, Figure 2B).

Amongst the set of 220 robust genetic interactions, we identified two previously reported 'paralog lethalities' - synthetic lethal relationships between duplicate genes (paralogs)[21–23] (Figure 2C). We found a robust association between mutation of the tumour suppressor *ARID1A* and sensitivity to inhibition of its paralog *ARID1B*[21] and also an association between mutation of *SMARCA4* and sensitivity to inhibition of its paralog *SMARCA2*[22,23]. Both pairs of genes (*ARID1A/ARID1B* and *SMARCA4/SMARCA2*) encode components of the larger SWI/SNF complex[24].

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* (Figure 3A, 3B)[25]. *BRAF* mutation was also associated with increased sensitivity to inhibition of *PEA15,* presumably a result of the requirement of PEA15 for ERK dimerisation and signalling activity[26,27].

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 (Figure 3C, 3D). 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[28,29]. 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 3C, 3D)*. RB1* loss was also associated with robust sensitivity to *SKP2*, a binding partner of Rb[30] first identified as an *RB1* synthetic lethal partner in retinoblastoma[31] and more recently in as a highly penetrant *RB1* synthetic lethal partner in triple negative breast cancer[32] (Figure 3C, 3D). 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[33].

Robust genetic interactions are enriched in protein-protein interactions

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[34]. 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 (Figure 4A) (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 (Figure 4A) (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 synthetic lethal 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 (Figure 4B);

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 (Supplementary Figure 2A and 2B) and using different sources of protein-protein interaction data (Supplementary Figure 2C and 2D, Supplementary Table 4)[35,36].

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 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 (Figure 3C, OR=6.1 and $p<2x10^{-10}$ when compared to discovered genetic interactions) (Supplementary Table 4). 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 (Figure 3D, OR=8.0 and $p<1x10^{-12}$ when compared to discovered genetic interactions) (Supplementary Table 4). 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.

The robust synthetic lethal interactions identified make more promising starting points for the development of targeted therapeutics than interactions reported in only specific cell line panels or with specific gene inhibition reagents. Three promising synthetic lethalities that are supported by a protein-protein interaction, but have not to our knowledge previously been validated, are highlighted in Figure 5. Mutation or deletion of the tumour suppressor kinase *STK11*, also known as *LKB1*, is associated with increased sensitivity to inhibition of its substrate kinase *MARK2*[37]. We have previously observed this synthetic lethality by analysing an siRNA screen of ovarian cancer cell lines[9]. That the association between *STK11* loss and *MARK2* sensitivity is observed across three distinct gene inhibition modalities (siRNA, shRNA, CRISPR) and across distinct cell line panels suggests that it is a robust synthetic lethality. We also found that loss of the tumour suppressor *SMARCA4* was associated with increased sensitivity to inhibition of the Bromodomain PHD Finger Transcription Factor *BPTF*. Recent work has identified an increased sensitivity of *SMARCA4* mutant tumour models to bromodomain inhibitors, although it is not yet clear which individual bromodomain proteins or combination of proteins is responsible for the observed effects[38]. Finally, we found a robust association between *EP300* mutation and sensitivity to inhibition of its paralog *CREBBP.* To our knowledge this synthetic lethality has not previously been reported. However, work in lung and hematopoietic tumour models has demonstrated the reciprocal synthetic lethality - *CREBBP* mutation is associated with increased sensitivity to inhibition of *EP300*[39]. As these

three interactions were reproduced across multiple screens and also supported by protein-protein interaction evidence, they would appear to be strong candidates for further study.

Prioritising robust synthetic lethal interactions from chemogenetic screens

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[40]. 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[40,41](Figure 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 (Figure 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.

Discussion

Here, we have developed an approach to identify robust genetic interactions that are reproduced across multiple screening libraries and across molecularly distinct cell line panels. We identified a set of 220 robust genetic interactions and 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 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 to cancer driver genes. There are likely to be dependencies associated with altered gene/protein expression of driver genes that will be missed by this approach. There are also likely to be dependencies associated with the alteration of non-driver genes that we have missed (e.g. passenger gene deletions[42]). 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[17].

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[43] or through using isogenic models for discovery and cell line panels for validation[14].

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. One option would be to screen target libraries for specific driver genes based on their known protein interaction partners. Alternatively, for the analysis of unbiased screening libraries, protein-protein interactions could be used as a covariate for hypothesis weighted multiple testing correction approaches[44]. 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.

Previous work has shown that genetic interactions between gene pairs whose protein products physically interact are more highly conserved across species[45–47]. 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[45,47]. 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[48].

Methods

All data analysis was performed using Python 3.7, Pandas 0.24[49] and StatsModels 0.9.0[50].

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[8] 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[7] 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[51] depletion log fold changes for Project SCORE[11] were obtained from the Project SCORE database (https://score.depmap.sanger.ac.uk/). ATARIS[52] scores for the DRIVE dataset[6] 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[5]) using the Cell Model Passports resource where appropriate[53].

Identifying selectively lethal genes

Similar to previous work[6,7], 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[8] - 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[6]. Finally from the combined list we removed genes known to be commonly essential in cancer cell lines[54]. This resulted in a set of 2,470 selectively lethal genes (Supplementary Table 1) which were used for all association analyses.

Identifying driver 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[4,5,55]. We used this to identify likely functional alterations in a panel of cancer driver genes[1] restricting our analysis to those genes that were subject to targeted sequencing as part of the Cancer Cell Line Encyclopedia[5].

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, MYC and MYCN) we considered them to be functionally altered only if they were amplified. For all tumour suppressors we considered loss of function mutations (e.g.premature truncations), recurrent missense mutations, and homozygous deletions to be functional alterations. The matrix of functional alterations is presented in Supplementary Table 2.

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[4,56]. As recent work[11] has highlighted that some dependencies (e.g. *WRN*) can be associated with microsatellite instability rather than individual driver genes, we also incorporated microsatellite instability[57] 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[58].

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 1B. 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 (Supplementary Figure 2).

Protein-protein interactions

Protein-protein interactions were obtained from STRING v10.5[34], BIOGRID 3.5.170[35] and from HIPPIE v2.0[36]. 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 (Supplementary Figure 2).

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.

References

1. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
2. Chapman, P. B. *et al.* Improved survival with vemurafenib in melanoma with BRAF V600E mutation. *N. Engl. J. Med.* **364**, 2507–2516 (2011).
3. Lord, C. J. & Ashworth, A. PARP inhibitors: Synthetic lethality in the clinic. *Science* **355**, 1152–1158 (2017).
4. Iorio, F. *et al.* A Landscape of Pharmacogenomic Interactions in Cancer. *Cell* **166**, 740–754 (2016).
5. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
6. McDonald, E. R., 3rd *et al.* Project DRIVE: A Compendium of Cancer Dependencies and

Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* **170**, 577–592.e10 (2017).
7. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564–576.e16 (2017).
8. Meyers, R. M. *et al.* Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49**, 1779–1784 (2017).
9. Campbell, J. *et al.* Large-Scale Profiling of Kinase Dependencies in Cancer Cell Lines. *Cell Rep.* **14**, 2490–2501 (2016).
10. Marcotte, R. *et al.* Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell* **164**, 293–309 (2016).
11. Behan, F. M. *et al.* Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568**, 511–516 (2019).
12. Steckel, M. *et al.* Determination of synthetic lethal interactions in KRAS oncogene-dependent cancer cells reveals novel therapeutic targeting strategies. *Cell Res.* **22**, 1227–1245 (2012).
13. Martin, T. D. *et al.* A Role for Mitochondrial Translation in Promotion of Viability in K-Ras Mutant Cells. *Cell Rep.* **20**, 427–438 (2017).
14. Ryan, C. J., Bajrami, I. & Lord, C. J. Synthetic Lethality and Cancer - Penetrance as the Major Barrier. *Trends Cancer Res.* **4**, 671–683 (2018).
15. Aguirre, A. J. & Hahn, W. C. Synthetic Lethal Vulnerabilities in KRAS-Mutant Cancers. *Cold Spring Harb. Perspect. Med.* **8**, (2018).
16. Downward, J. RAS Synthetic Lethal Screens Revisited: Still Seeking the Elusive Prize? *Clin. Cancer Res.* **21**, 1802–1809 (2015).
17. Kaelin, W. G., Jr. Molecular biology. Use and abuse of RNAi to study mammalian gene function. *Science* **337**, 421–422 (2012).
18. Bridgett, S., Campbell, J., Lord, C. J. & Ryan, C. J. CancerGD: A Resource for Identifying and Interpreting Genetic Dependencies in Cancer. *Cell Syst* **5**, 82–86.e3 (2017).
19. Giacomelli, A. O. *et al.* Mutational processes shape the landscape of TP53 mutations in human cancer. *Nat. Genet.* **50**, 1381–1387 (2018).
20. Wee, S. *et al.* PTEN-deficient cancers depend on PIK3CB. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 13057–13062 (2008).
21. Helming, K. C. *et al.* ARID1B is a specific vulnerability in ARID1A-mutant cancers. *Nat. Med.* **20**, 251–254 (2014).
22. Hoffman, G. R. *et al.* Functional epigenetics approach identifies BRM/SMARCA2 as a critical synthetic lethal target in BRG1-deficient cancers. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 3128–3133 (2014).
23. Oike, T. *et al.* A synthetic lethality–based strategy to treat cancers harboring a genetic deficiency in the chromatin remodeling factor BRG1. *Cancer Res.* (2013).
24. Wilson, B. G. & Roberts, C. W. M. SWI/SNF nucleosome remodellers and cancer. *Nat. Rev. Cancer* **11**, 481–492 (2011).
25. Hill, K. S. *et al.* PTPN11 Plays Oncogenic Roles and Is a Therapeutic Target for BRAF Wild-Type Melanomas. *Mol. Cancer Res.* **17**, 583–593 (2019).
26. Herrero, A. *et al.* Small Molecule Inhibition of ERK Dimerization Prevents Tumorigenesis by RAS-ERK Pathway Oncogenes. *Cancer Cell* **28**, 170–182 (2015).
27. Formstecher, E. *et al.* PEA-15 Mediates Cytoplasmic Sequestration of ERK MAP Kinase. *Dev. Cell* **1**, 239–250 (2001).
28. Asghar, U., Witkiewicz, A. K., Turner, N. C. & Knudsen, E. S. The history and future of targeting cyclin-dependent kinases in cancer therapy. *Nat. Rev. Drug Discov.* **14**, 130–146 (2015).

29. O'Leary, B. *et al.* The Genetic Landscape and Clonal Evolution of Breast Cancer Resistance to Palbociclib plus Fulvestrant in the PALOMA-3 Trial. *Cancer Discov.* **8**, 1390–1403 (2018).
30. Ji, P. *et al.* 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 (2004).
31. Xu, X. L. *et al.* Rb suppresses human cone-precursor-derived retinoblastoma tumours. *Nature* **514**, 385–388 (2014).
32. Brough, R. *et al.* Identification of highly penetrant Rb-related synthetic lethal interactions in triple negative breast cancer. *Oncogene* (2018).
33. Tadesse, S., Caldon, E., Tilley, W. & Wang, S. Cyclin Dependent Kinase 2 Inhibitors in Cancer Therapy: an Update. *J. Med. Chem.* (2018). doi:10.1021/acs.jmedchem.8b01469
34. Szklarczyk, D. *et al.* STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **43**, D447–52 (2015).
35. Chatr-aryamontri, A. *et al.* The BioGRID interaction database: 2015 update. *Nucleic Acids Res.* (2014). doi:10.1093/nar/gku1204
36. Alanis-Lobato, G., Andrade-Navarro, M. A. & Schaefer, M. H. HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res.* **45**, D408–D414 (2017).
37. Lizcano, J. M. *et al.* LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1. *EMBO J.* **23**, 833–843 (2004).
38. Shorstova, T. *et al.* SWI/SNF-compromised cancers are susceptible to bromodomain inhibitors. *Cancer Res.* (2019). doi:10.1158/0008-5472.CAN-18-1545
39. Ogiwara, H. *et al.* Targeting p300 Addiction in CBP-Deficient Cancers Causes Synthetic Lethality by Apoptotic Cell Death due to Abrogation of MYC Expression. *Cancer Discov.* **6**, 430–445 (2016).
40. Wang, C. *et al.* Genome-wide CRISPR screens reveal synthetic lethality of RNASEH2 deficiency and ATR inhibition. *Oncogene* (2018). doi:10.1038/s41388-018-0606-4
41. Vendetti, F. P. *et al.* 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 (2015).
42. Muller, F. L. *et al.* Passenger deletions generate therapeutic vulnerabilities in cancer. *Nature* **488**, 337–342 (2012).
43. Boehm, J. S. & Golub, T. R. An ecosystem of cancer cell line factories to support a cancer dependency map. *Nat. Rev. Genet.* **16**, 373–374 (2015).
44. Ignatiadis, N., Klaus, B., Zaugg, J. B. & Huber, W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods* **13**, 577–580 (2016).
45. Ryan, C. J. *et al.* Hierarchical modularity and the evolution of genetic interactomes across species. *Mol. Cell* **46**, 691–704 (2012).
46. Roguev, A. *et al.* Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* **322**, 405–410 (2008).
47. Srivas, R. *et al.* A Network of Conserved Synthetic Lethal Interactions for Exploration of Precision Cancer Therapy. *Mol. Cell* **63**, 514–525 (2016).
48. Jerby-Arnon, L. *et al.* Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell* **158**, 1199–1209 (2014).
49. McKinney, W. pandas: a foundational Python library for data analysis and statistics. *Python for High Performance and Scientific Computing* 1–9 (2011).
50. Seabold, S. & Perktold, J. Statsmodels: Econometric and statistical modeling with python.

in *Proceedings of the 9th Python in Science Conference* **57**, 61 (SciPy society Austin, 2010).
51. Iorio, F. *et al.* Unsupervised correction of gene-independent cell responses to CRISPR-Cas9 targeting. *BMC Genomics* **19**, 604 (2018).
52. Shao, D. D. *et al.* ATARiS: computational quantification of gene suppression phenotypes from multisample RNAi screens. *Genome Res.* **23**, 665–678 (2013).
53. van der Meer, D. *et al.* Cell Model Passports—a hub for clinical, genetic and functional datasets of preclinical cancer models. *Nucleic Acids Res.* **47**, D923–D929 (2019).
54. Hart, T. *et al.* Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. *G3* **7**, 2719–2727 (2017).
55. Cancer Cell Line Encyclopedia Consortium & Genomics of Drug Sensitivity in Cancer Consortium. Pharmacogenomic agreement between two cancer cell line data sets. *Nature* **528**, 84–87 (2015).
56. Cokelaer, T. *et al.* GDSCTools for mining pharmacogenomic interactions in cancer. *Bioinformatics* **34**, 1226–1228 (2018).
57. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* (2019). doi:10.1038/s41586-019-1186-3
58. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Series B Stat. Methodol.* **57**, 289–300 (1995).

Figure Legends

Figure 1. Identifying reproducible genetic interactions using partially overlapping loss-of-function screens. A) Venn diagram showing the overlap of cell lines between the four loss-of-function screens analysed in this screen. **B)** Workflow showing how reproducible genetic interactions are identified using discovery and validation sets. **C)** Tukey box plots showing reproducible oncogene addictions identified using the workflow described in B. 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, yellow to indicate resistance.

Figure 2. Identified reproducible genetic interactions. A) Dot plot showing the reproducible genetic interactions identified. Each coloured circle indicates a reproducible 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 reproduced interactions and the total number of reproduced genetic interactions for each driver gene is shown in parentheses. *TP53* has been excluded for clarity. **B)** Examples of known genetic interactions, including an increased sensitivity of *PTEN* mutant cell lines to *PIK3CB* inhibition and increased resistance of *TP53* mutant cancers to *MDM2* inhibition. Top row shows the data used to discover the interactions (DEPMAP) while the bottom row shows the data used to validate the interactions (AVANA less DEPMAP). **C)** Synthetic lethal interactions involving paralog pairs. Top row shows the data used to discover the interactions (DRIVE) while the bottom row shows the data used to validate the interactions (AVANA less DRIVE).

Figure 3. 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

Figure 4. Reproducible 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 Supplementary Table 4. **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.

Figure 5. Novel synthetic lethal relationships between gene pairs whose protein products interact. Boxplots are shown for selected reproducible synthetic lethal interactions between gene pairs whose protein products physically interact. Top row shows discovery dataset, bottom row shows the validation results.

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).

Supplementary Figure Legends

Supplementary Figure 1. 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 interaction (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 interactions 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.

Supplementary Figure 2. Reproducible 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 figure 4B but here the protein-protein interaction pairs are obtained from the HIPPIE database **D)** Similar to main text figure 4B but here the protein-protein interaction pairs are obtained from the BioGRID database.

Supplementary Tables

Supplementary Table 1 – Driver gene alterations
Supplementary Table 2 – Selectively lethal genes
Supplementary Table 3 – Reproducible genetic dependencies
Supplementary Table 4 – Protein-protein interaction enrichment statistics

Figure 1

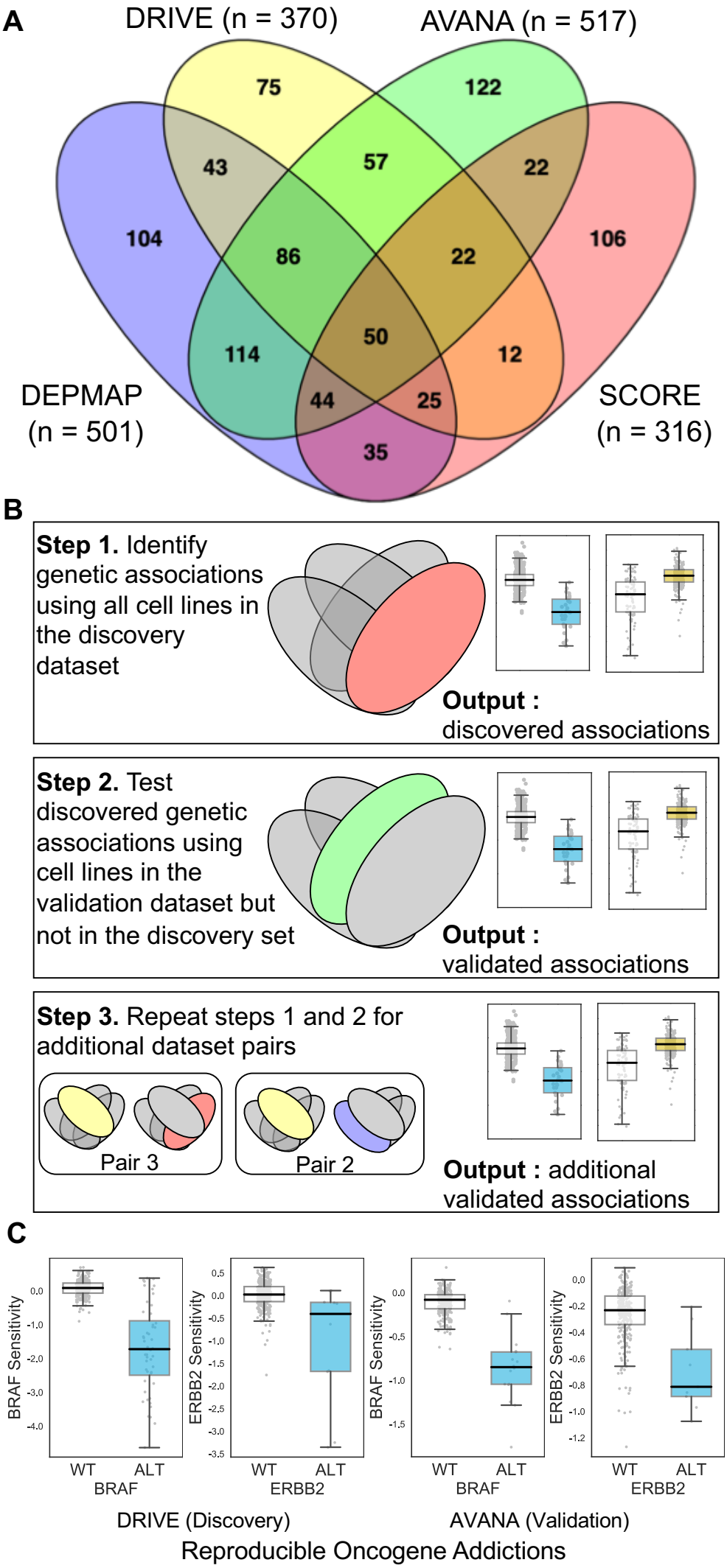

A
DRIVE (n = 370)
AVANA (n = 517)
DEPMAP (n = 501)
SCORE (n = 316)
75
122
43
57
22
104
86
22
106
50
114
12
44
25
35
B
Step 1. Identify genetic associations using all cell lines in the discovery dataset
Output : discovered associations
Step 2. Test discovered genetic associations using cell lines in the validation dataset but not in the discovery set
Output : validated associations
Step 3. Repeat steps 1 and 2 for additional dataset pairs
Pair 3
Pair 2
Output : additional validated associations
C
BRAF Sensitivity
ERBB2 Sensitivity
WT
ALT
BRAF
ERBB2
DRIVE (Discovery)
AVANA (Validation)
Reproducible Oncogene Addictions

Figure 2

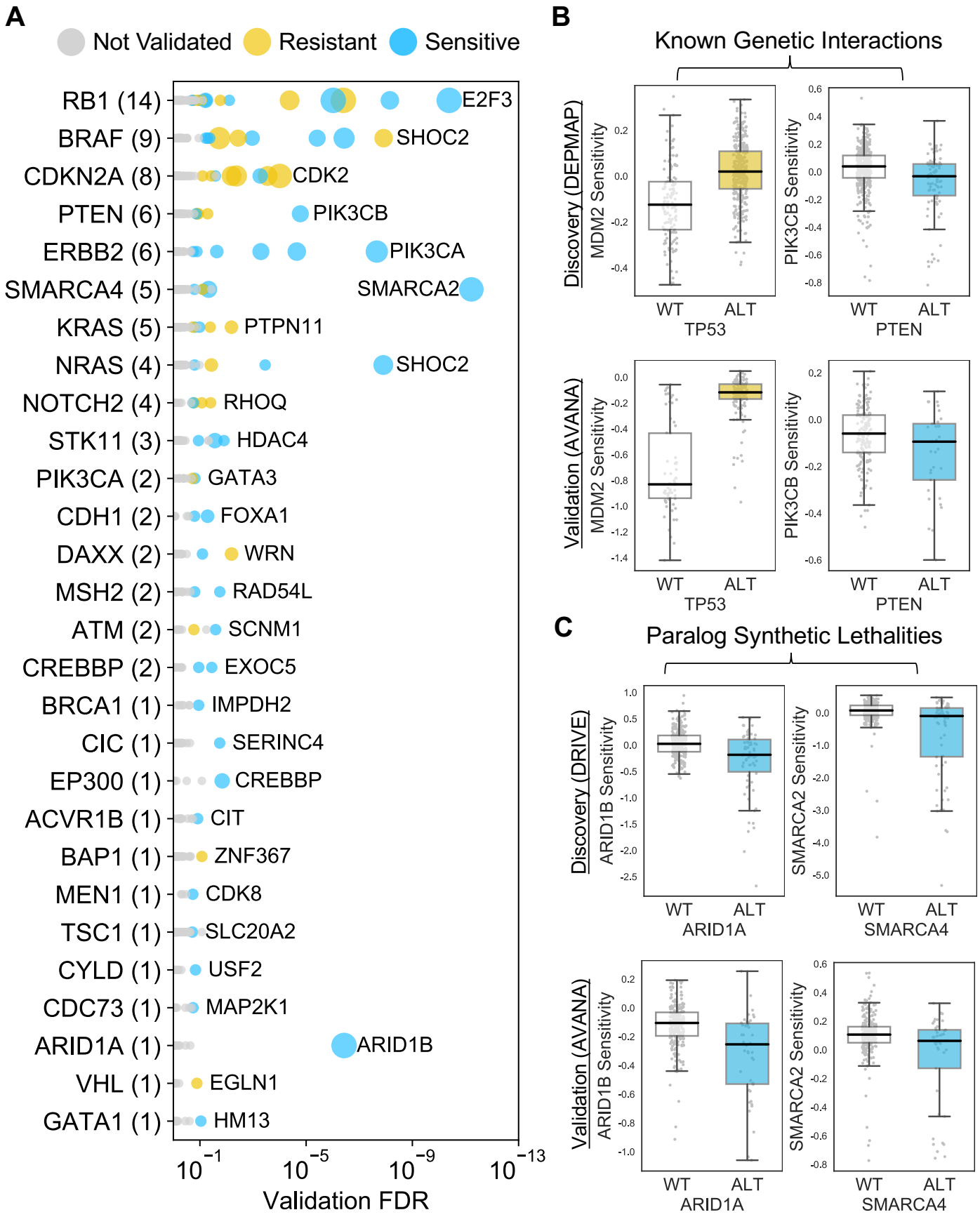

A
Not Validated
Resistant
Sensitive
RB1 (14)
BRAF (9)
CDKN2A (8)
PTEN (6)
ERBB2 (6)
SMARCA4 (5)
KRAS (5)
NRAS (4)
NOTCH2 (4)
STK11 (3)
PIK3CA (2)
CDH1 (2)
DAXX (2)
MSH2 (2)
ATM (2)
CREBBP (2)
BRCA1 (1)
CIC (1)
EP300 (1)
ACVR1B (1)
BAP1 (1)
MEN1 (1)
TSC1 (1)
CYLD (1)
CDC73 (1)
ARID1A (1)
VHL (1)
GATA1 (1)
E2F3
SHOC2
CDK2
PIK3CB
PIK3CA
SMARCA2
PTPN11
SHOC2
RHOQ
HDAC4
GATA3
FOXA1
WRN
RAD54L
SCNM1
EXOC5
IMPDH2
SERINC4
CREBBP
CIT
ZNF367
CDK8
SLC20A2
USF2
MAP2K1
ARID1B
EGLN1
HM13
Validation FDR
B
Known Genetic Interactions
Discovery (DEPMAP)
Validation (AVANA)
MDM2 Sensitivity
PIK3CB Sensitivity
WT
ALT
TP53
PTEN
C
Paralog Synthetic Lethalities
ARID1B Sensitivity
SMARCA2 Sensitivity
ARID1A
SMARCA4

Figure 3

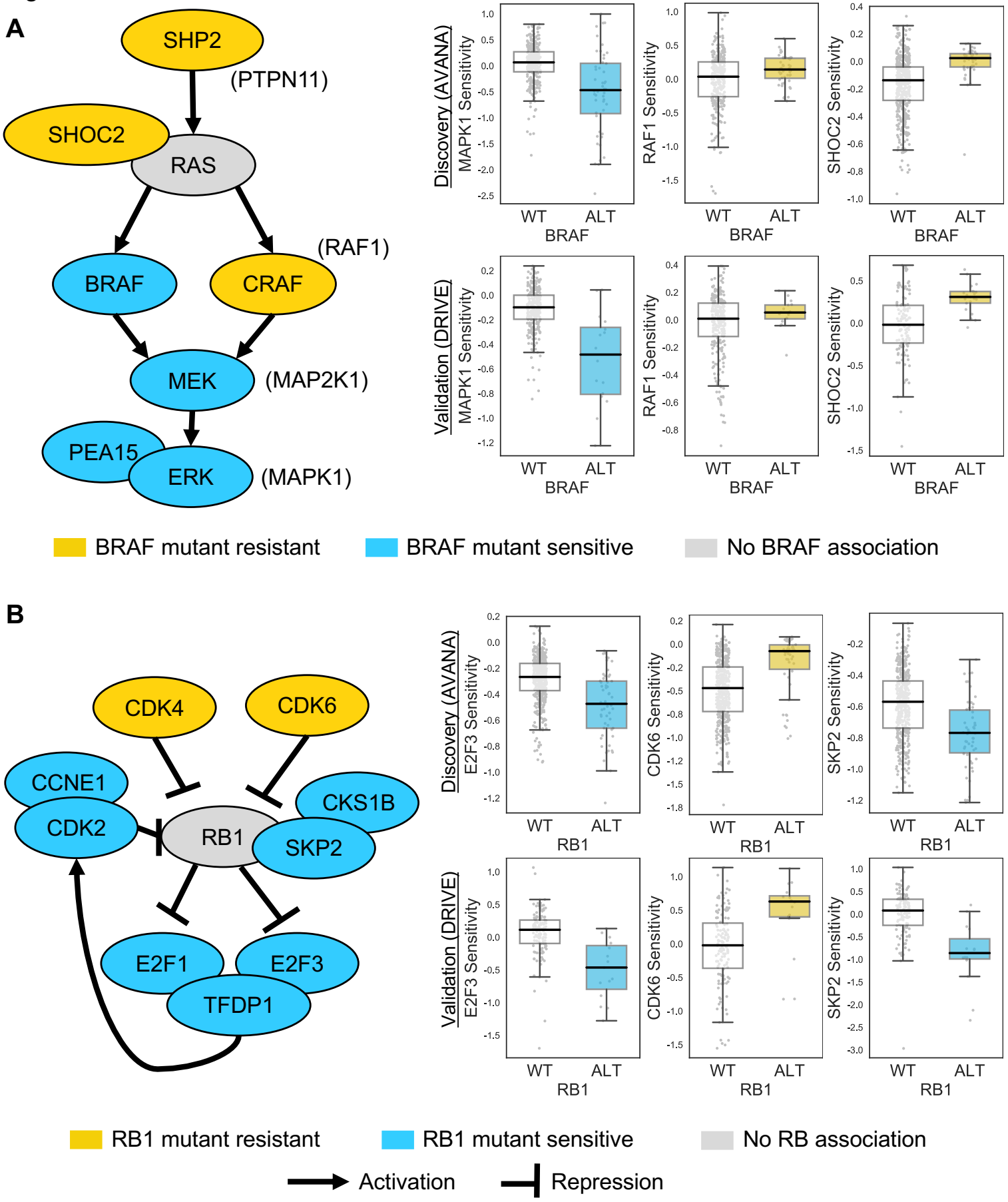

A
SHP2
(PTPN11)
SHOC2
RAS
(RAF1)
BRAF
CRAF
MEK
(MAP2K1)
PEA15
ERK
(MAPK1)
Discovery (AVANA)
Validation (DRIVE)
MAPK1 Sensitivity
RAF1 Sensitivity
SHOC2 Sensitivity
WT
ALT
BRAF
BRAF mutant resistant
BRAF mutant sensitive
No BRAF association
B
CDK4
CDK6
CCNE1
CDK2
CKS1B
RB1
SKP2
E2F1
E2F3
TFDP1
E2F3 Sensitivity
CDK6 Sensitivity
SKP2 Sensitivity
RB1
RB1 mutant resistant
RB1 mutant sensitive
No RB association
Activation
Repression

Figure 4

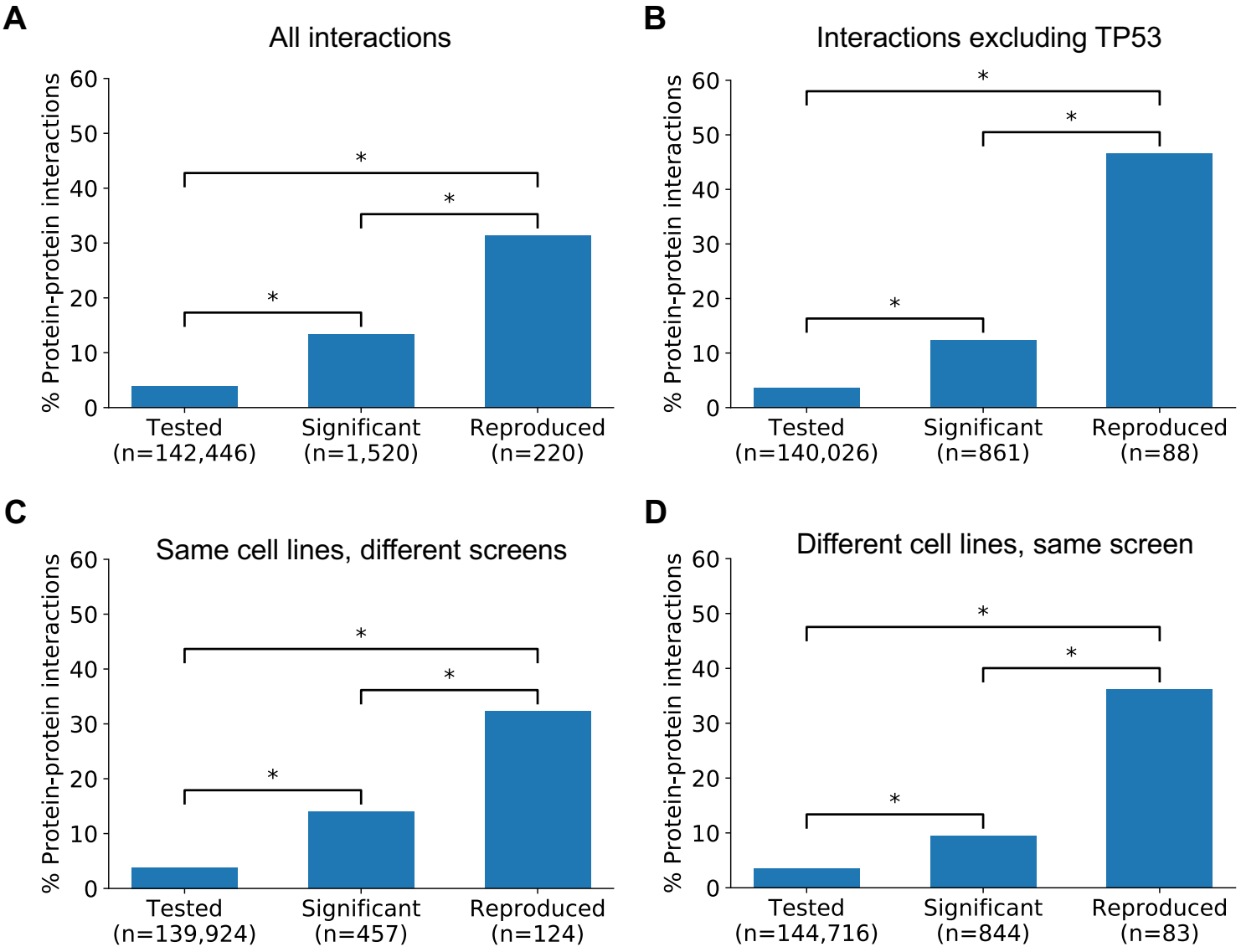

A
All interactions
% Protein-protein interactions
0
10
20
30
40
50
60
*
*
*
Tested
(n=142,446)
Significant
(n=1,520)
Reproduced
(n=220)
B
Interactions excluding TP53
% Protein-protein interactions
0
10
20
30
40
50
60
*
*
*
Tested
(n=140,026)
Significant
(n=861)
Reproduced
(n=88)
C
Same cell lines, different screens
% Protein-protein interactions
0
10
20
30
40
50
60
*
*
*
Tested
(n=139,924)
Significant
(n=457)
Reproduced
(n=124)
D
Different cell lines, same screen
% Protein-protein interactions
0
10
20
30
40
50
60
*
*
*
Tested
(n=144,716)
Significant
(n=844)
Reproduced
(n=83)

Figure 5

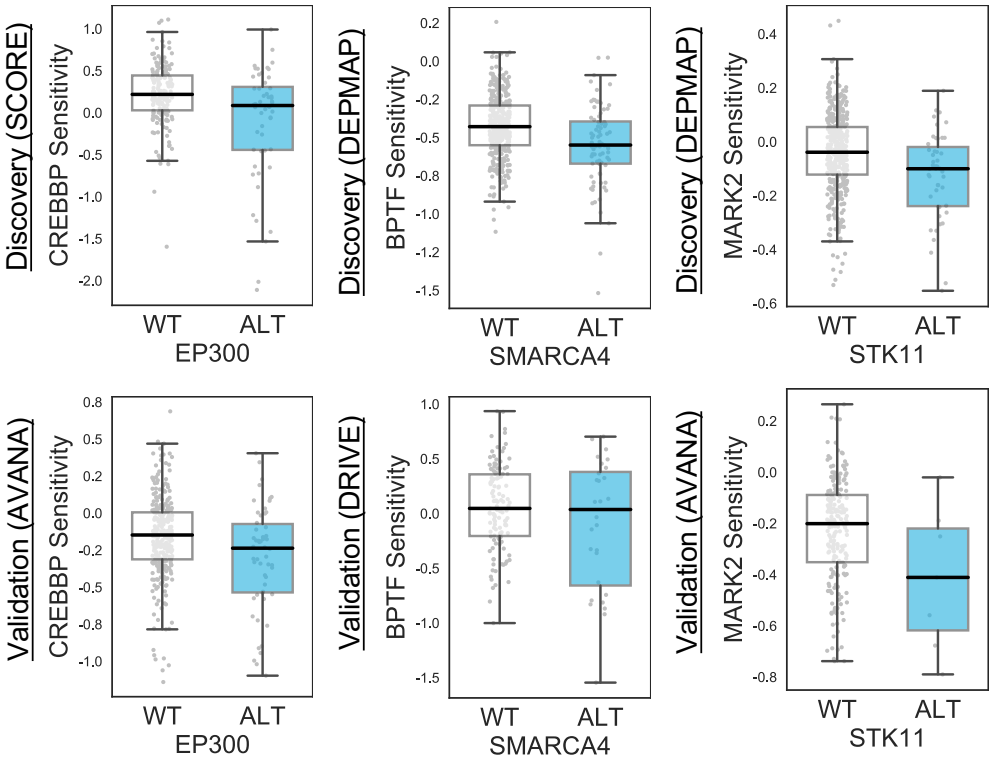
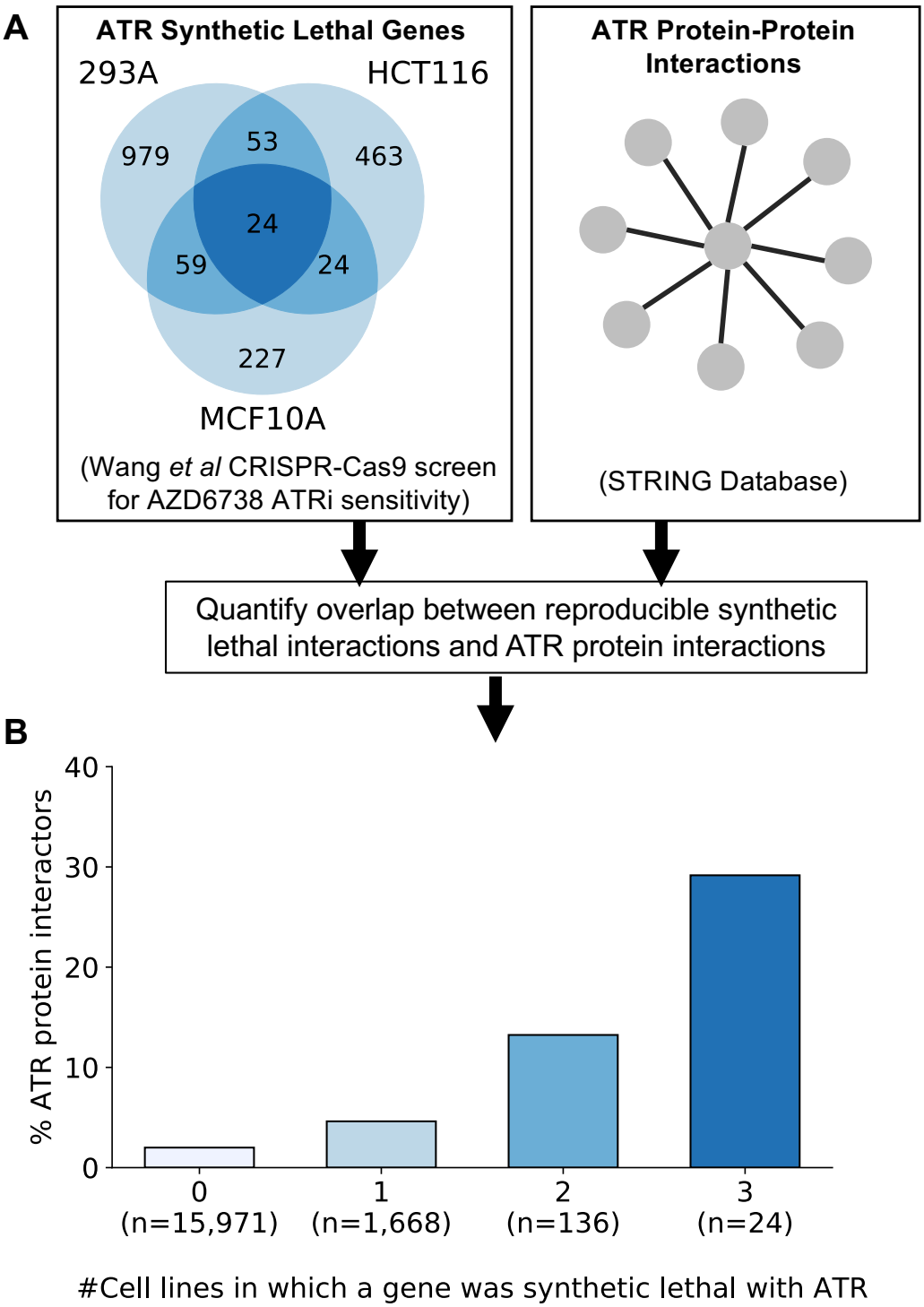
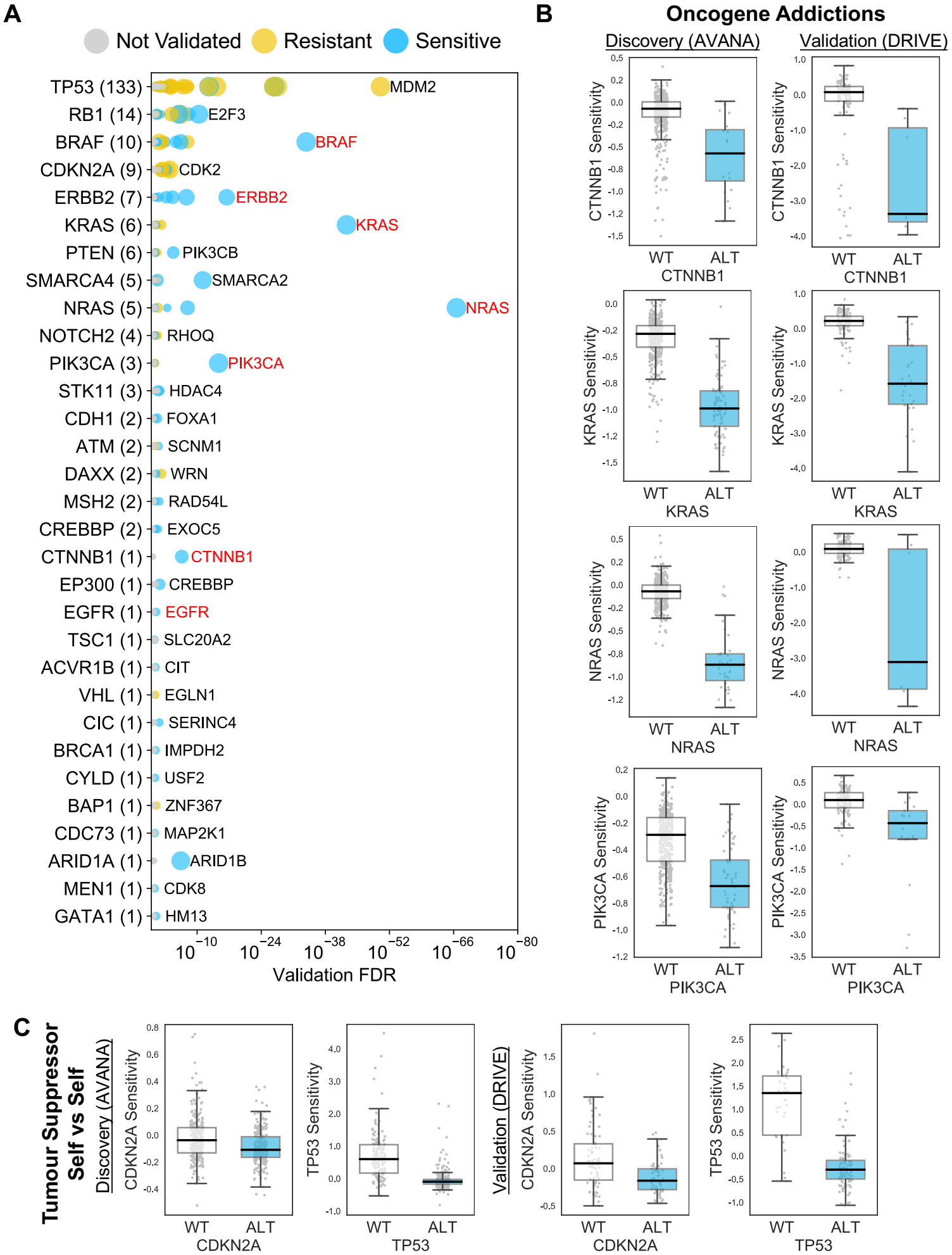
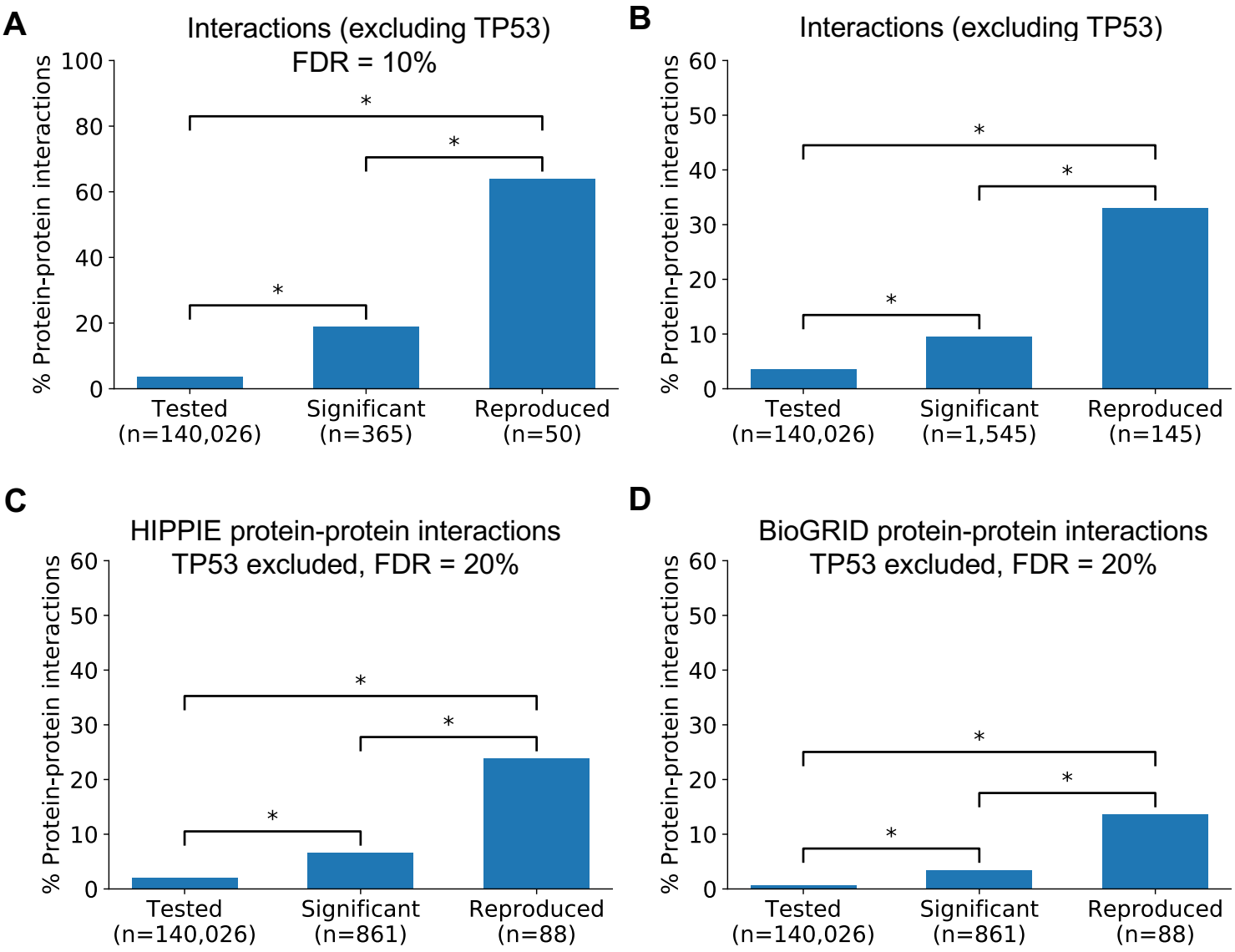

Discovery (SCORE)
CREBBP Sensitivity
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
WT
ALT
EP300
Discovery (DEPMAP)
BPTF Sensitivity
0.2
0.0
-0.2
-0.5
-0.8
-1.0
-1.2
-1.5
WT
ALT
SMARCA4
Discovery (DEPMAP)
MARK2 Sensitivity
0.4
0.2
0.0
-0.2
-0.4
-0.6
WT
ALT
STK11
Validation (AVANA)
CREBBP Sensitivity
0.8
0.5
0.2
0.0
-0.2
-0.5
-0.8
-1.0
WT
ALT
EP300
Validation (DRIVE)
BPTF Sensitivity
1.0
0.5
0.0
-0.5
-1.0
-1.5
WT
ALT
SMARCA4
Validation (AVANA)
MARK2 Sensitivity
0.2
0.0
-0.2
-0.4
-0.6
-0.8
WT
ALT
STK11

Figure 6


A
ATR Synthetic Lethal Genes
293A
HCT116
979
53
463
24
59
24
227
MCF10A
(Wang et al CRISPR-Cas9 screen for AZD6738 ATRi sensitivity)
ATR Protein-Protein Interactions
(STRING Database)
Quantify overlap between reproducible synthetic lethal interactions and ATR protein interactions
B
% ATR protein interactors
40
30
20
10
0
0 (n=15,971)
1 (n=1,668)
2 (n=136)
3 (n=24)
#Cell lines in which a gene was synthetic lethal with ATR

Supplementary Figure 1



Supplementary Figure 2