# Drug mechanism-of-action discovery through the integration of pharmacological and CRISPR screens

Emanuel Gonçalves[1], Aldo Segura-Cabrera[2], Clare Pacini[1], Gabriele Picco[1], Fiona M. Behan[1], Patricia Jaaks[1], Elizabeth A. Coker[1], Donny van der Meer[1], Andrew Barthorpe[1], Howard Lightfoot[1], GDSC Screening Team[1], Andrew R. Leach[2], James T. Lynch[3], Ben Sidders[3], Claire Crafter[3], Francesco Iorio[1,5], Stephen Fawell[4], Mathew J. Garnett[1,*]

1. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK
2. European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, UK
3. Research and Early Development, Oncology R&D, AstraZeneca, Cambridge, UK
4. Research and Early Development, Oncology R&D, AstraZeneca, Waltham, MA 02451, USA
5. Human Technopole, Via Cristina Belgioioso, 20157 Milano, Italy

* Corresponding author: mg12@sanger.ac.uk

Keywords: Drug response / CRISPR-Cas9 / Drug mechanism-of-action / Protein Networks

# Abstract

**Low success rates during drug development are due in part to the difficulty of defining drug mechanism-of-action and molecular markers of therapeutic activity. Here, we integrated 199,219 drug sensitivity measurements for 397 unique anti-cancer drugs and genome-wide CRISPR loss-of-function screens in 484 cell lines to systematically investigate *in cellular* drug mechanism-of-action. We observed an enrichment for positive associations between drug sensitivity and knockout of their nominal targets, and by leveraging protein-protein networks we identified pathways that mediate drug response. This revealed an unappreciated role of mitochondrial E3 ubiquitin-protein ligase *MARCH5* in sensitivity to MCL1 inhibitors. We also estimated drug on-target and off-target activity, informing on specificity, potency and toxicity. Linking drug and gene dependency together with genomic datasets uncovered contexts in which molecular networks when perturbed mediate cancer cell loss-of-fitness, and thereby provide independent and orthogonal evidence of biomarkers for drug development. This study illustrates how integrating cell line drug sensitivity with CRISPR loss-of-function screens can elucidate mechanism-of-action to advance drug development.**


# Introduction

Understanding drug mechanism-of-action and evaluating *in cellular* activity is challenging (Santos *et al*, 2017) and widespread target promiscuity contributes to low success rates during drug development (Klaeger *et al*, 2017). For target-based drug development, a detailed understanding of drug mechanism-of-action provides information about specificity and undesirable off-target activity which could lead to toxicity and reduced therapeutic window (Lin *et al*, 2019). Moreover, molecular biomarkers can be used to monitor drug activity and to identify contexts in which drugs are more effective as the basis for patient stratification during clinical development.

The cellular activity of a drug is influenced by multiple factors including the selectivity and affinity of the compound to its target(s) and the penetrance of target engagement on cellular phenotypes. An array of biochemical, biophysical, computational and cellular assays are currently used to investigate drug mechanism-of-action (Schenone *et al*, 2013). For example, protein kinase inhibitors are profiled *in vitro* for their specificity and potency against panels of purified recombinant protein kinases. While informative, this approach fails to recapitulate the native context of the full-length protein in cells which could influence drug activity, it does not identify non-kinase off-target effects, nor is it suitable to evaluate the

selectivity of compounds to non-kinase targets. Existing *in cellular* based approaches include transcriptional profiling following drug treatment of cells, chemical proteomics approaches such as kinobeads to measure drug-protein interactions, and multiplexed imaging or flow-cytometry to measure multiple cellular parameters upon drug treatment (Subramanian *et al*, 2017; Li *et al*, 2017; Reinecke *et al*, 2019). Despite the utility of these different approaches, gaining a full picture of drug mechanism-of-action, particularly in cells, remains a challenge and new approaches would be beneficial.

Pharmacological screens (Barretina *et al*, 2012; Garnett *et al*, 2012; Iorio *et al*, 2016; Subramanian *et al*, 2017; Lee *et al*, 2018) have been used to profile the activity of hundreds of compounds in highly-annotated collections of cancer cell lines with the aim of identifying molecular markers of drug sensitivity to guide clinical development (Cook *et al*, 2014; Nelson *et al*, 2015). More recently, CRISPR-based gene-editing has enabled the evaluation of highly specific and penetrant gene-knockout effects on cell fitness genome-wide in hundreds of cancer cell lines (Jinek *et al*, 2012; Shalem *et al*, 2014; Hart *et al*, 2015; Meyers *et al*, 2017; Behan *et al*, 2019). This has provided rich functional resources to explore cancer vulnerabilities and new potential drug targets (Marcotte *et al*, 2016; Meyers *et al*, 2017; Tsherniak *et al*, 2017; Behan *et al*, 2019). Parallel integration of gene loss-of-function screens with drug response can be used to investigate drug mechanism-of-action (Deans *et al*, 2016; Subramanian *et al*, 2017; Jost & Weissman, 2018).

Here, we integrate recent genome-wide CRISPR-Cas9 loss-of-function screens with pharmacological data for 397 unique anti-cancer compounds in 484 cancer cell lines. We show that CRISPR-Cas9 datasets recapitulate drug targets, can provide insights into drug potency and selectivity, and define cellular networks underpinning drug sensitivity. This approach identified a link between mitochondrial ubiquitin ligase *MARCH5* in MCL1 inhibitors response, and specifically in breast cancer cell lines. Furthermore, we defined robust pharmacogenomic associations, represented by genetic biomarkers independently supported by drug response and gene fitness measurements. These identify genetic contexts associated with drug-pathway dependency and provide a more refined set of biomarkers. Taken together, we present here an approach to leverage pharmacological and CRISPR screening data to inform on drug *in cellular* mechanism-of-action to guide drug development.

# Results

## Cancer cell line drug sensitivity and gene fitness effects

We analysed datasets from a highly-annotated collection of 484 histologically diverse human cancer cell lines (Supplementary Table 1). These have been extensively genetically characterised and utilised for large-scale drug sensitivity testing and CRISPR-Cas9 whole-genome loss-of-function screens (Garnett *et al*, 2012; Iorio *et al*, 2016; Meyers *et al*, 2017; Picco *et al*, 2019; van der Meer *et al*, 2019). We expanded on published single agent drug sensitivity data (Garnett *et al*, 2012; Lynch *et al*, 2016; Iorio *et al*, 2016; Picco *et al*, 2019) to consider 199,219 IC50 values for 397 unique cancer drugs (480 drugs including duplicates, Supplementary Table 2). These encompassed FDA-approved cancer drugs, drugs in clinical development, and investigational compounds with multiple modes of action, including 24 chemotherapeutic agents and 367 small molecule inhibitors. Drugs considered in this study had a response in at least 3 cell lines (IC50 lower than half of the maximum screened concentration) and 86% of all possible drug/cell line IC50 measurements have been evaluated (Supplementary Figure 1a, Supplementary Table 3). Two experimental protocols were used to generate the drug sensitivity measurements, termed here as GDSC1 (Iorio *et al*, 2016) and GDSC2 (Picco *et al*, 2019), chronologically ordered (Supplementary Figure 1b). A principal component analysis (PCA) of IC50 values identified a screen specific batch effect associated with principal component (PC) 2 which explained 2.8% of the total variance (Supplementary Figure 1c). For this reason, despite the fact that compounds screened with both technologies showed good agreement (n=66, mean Pearson's R=0.50), we analysed the measurements of the screens separately. Analysis of the drug response variation across cell lines revealed that PC 1 (28.7% variance captured) was significantly correlated with cell line growth rate (Pearson's R=-0.51, p-value=1.2e-28), particularly for chemotherapy agents and growth inhibitors (Supplementary Figure 1d and1e).

Cell fitness effects for 16,643 gene knockouts have been measured using genome-wide CRISPR-Cas9 screens at the Sanger and Broad Institutes (Meyers *et al*, 2017; Behan *et al*, 2019; DepMap, 2019) (Supplementary Table 4). The first PC across the cell lines (6.8% variance explained) separated the two institutes of origin (Supplementary Figure 2a), consistent with a comparative analysis performed on an overlapping set of cell lines (Dempster *et al*, 2019). Growth rate was less significantly associated with CRISPR knockout response (Supplementary Figure 2b and 2c).

## Gene knockout fitness effects correspond with drug targets

We began by investigating the extent to which drug sensitivity corresponded to CRISPR knock-out of drug targets. We systematically searched for associations between drug sensitivity and gene fitness effects across the 484 cell lines (Figure 1a). We expect this to capture a variety of relationships ranging from direct drug-target interactions to more complex associations arising from interactions with regulators of the drug target(s). We tested a total of 7,988,640 single-feature gene-drug associations using linear mixed regression models. Potential confounding effects such as growth rate, culture conditions, data source and sample structure were considered in the models. We identified 865 significant associations (FDR adjusted p-value < 10%, Supplementary Table 5) between drug response and gene fitness profiles (Figure 1b), termed hereafter as significant drug-gene pairs. For this analysis we were able to manually curate the nominal therapeutic target(s) for 94.7% (n=376) of the anti-cancer drugs (Supplementary Figure 3a and Supplementary Table 1).

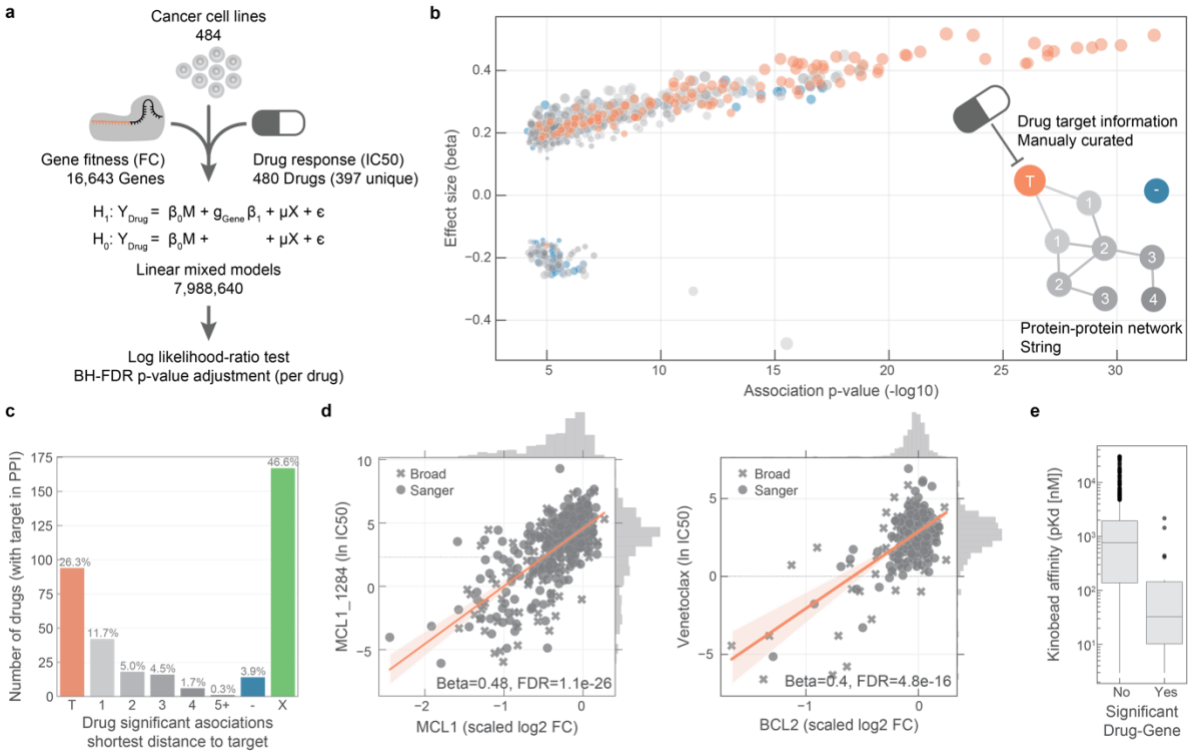



***Figure 1. Integration of drug and gene dependencies in cancer cell lines. a**, we used linear models to integrate drug sensitivity and gene fitness measurements. **b**, volcano plot showing the effect sizes and the p-value for statistically significant associations, Benjamini-Hochberg False Discovery Rate (FDR) adjusted p-value < 10%. Drug-gene associated pairs are coloured according to their shortest distance in a protein-protein interaction network of the gene to any of the nominal target of the drug. **c**, percentage of the 358 drugs with significant associations and their closest distance to the drug nominal targets. T represents drugs that have a significant association with at least 1 of their canonical targets and X are those which have no significant association. **d**, representative examples of the top drug response correlations with target gene fitness. MCL_1284 and Venetoclax are MCL1 and BCL2 selective inhibitors, respectively. Gene fitness log2 fold-changes (FC) are scaled by using*

*previously defined sets of essential (median scaled log2 FC = -1) and non-essential (median scaled log2 FC = 0) genes. Drug response IC50 measurements are represented using the natural log (ln IC50).* ***e****, kinobead affinity is significantly higher (lower pKd) for compounds with a significant association with their target (n = 20, Mann-Whitney p-value=3.1e-07).*

For 26% (n=94) of the 358 drugs with target annotation and for which the target was knocked-out with the CRISPR-Cas9 libraries, we identified significant drug-gene pairs with their putative targets (Figure 1c). For example, there were strong associations between MCL1 and BCL2 inhibitors and their knockouts (Figure 1d). Notably, drug-gene associations with the drug target had a skewed distribution towards positive effect sizes (Mann–Whitney U test p-value < 1.36e-105, Supplementary Figure 3b) and were among the strongest associations (Figure 1b). To investigate this further, we utilised independently acquired kinobead drug-protein affinity measurements for an overlapping 64 protein kinase inhibitors which were profiled for their specificity against 202 kinases (Klaeger *et al*, 2017). Drugs with significant associations with the target also had stronger affinity to their target in the kinobead assay, providing independent evidence that the strongest drug-gene associations are enriched for targets of the drugs (Figure 1e). Overall, we identified the nominal target of approximately one quarter of the drugs tested using orthogonal CRISPR gene fitness screens, and drug targets were amongst the most significant associations.

## Cellular networks underpinning drug response

The remaining 74% (n=264) of drugs were not significantly associated with the CRISPR loss-of-function measurements of their nominal targets (Figure 1c). We reasoned that superimposing the significant drug-gene pairs onto a protein interaction network may shed further insight into drug mechanism-of-action. We used a protein-protein interaction (PPI) network assembled from STRING database (Szklarczyk *et al*, 2017) (10,587 nodes and 205,251 interactions), and for the significant drug-genes pairs calculated the distances between the drug nominal targets and the associated gene-products. For 76 drugs no significant association with their target was identified, but instead had a significant association with their target's first neighbour or a protein closely related in the network (1, 2 or 3 PPI interactions distance from any of the drug targets) (Figure 1b and c). Thus, 47.5% of the annotated compounds (n=170) had an association with either the target or a functionally-related protein.

The strongest drug-gene pair associations were between the drug and the canonical targets rather than components of the PPI network, and significance decreased (along with

the number of associations) as the interaction distance increased (Figure 2a). To exclude that this observation is related to the topology of the network, we calculated the length of all the shortest paths between the drug targets and their associated genes and confirmed the enrichment of first and second neighbours in significant drug-gene associations (Supplementary Figure 3c). In comparison, cell line gene expression profiles are less powered to identify associations with the PPI neighbours of the drug target (Figure 2b; Supplementary Table 6). In particular, the number of drugs significantly associated with their targets substantially decreased (n=17) and significant associations were predominantly found with gene-products further away in the PPI network and close to the average length of all paths ($l_G$ = 3.9). As an example, MIEN1 gene expression is significantly correlated with multiple EGFR and ERBB2 inhibitors which can be explained, not by a functional relationship, but by genomic co-localisation with *ERBB2* on chromosome 17. Hence, CRISPR measurements are more powered than gene expression to identify drug functional interaction networks.

To investigate putative regulatory networks for drugs, we weighted the PPI network edges with the correlation between the fitness profiles of the two connected nodes and integrated the resulting weighted network with drug response associations. *EGFR* inhibitors are the most abundant drug class in our set, and we observed that multiple inhibitors (e.g. cetuximab) showed significant associations with *EGFR* and known pathway members, for example *SHC1* and *GRB2* (Zheng *et al*, 2013; Scaltriti & Baselga, 2006) (Figure 2c). Additionally, the weighted network shows pathway members that have strongly correlated fitness profiles, which are likely functionally related (Pan *et al*, 2018). For EGFR inhibitors these included tyrosine receptor kinases *NTRK3* and *MET*, and the protein phosphatase *PTPN11* (Wang *et al*, 2017; Pan *et al*, 2018) (Figure 2d). Drug-target tailored networks can be used to understand drug mechanism-of-action, and have the potential to identify resistance mechanisms and propose alternative targets in the network.

Despite our finding that we can illuminate drug functional networks, 46.6% (n=167) of the tested drugs had no significant drug-gene associations. This could in part be explained by lower variability in CRISPR fold change measurements for the target of these drugs (Figure 2e). For example, where genetic knockout induces strong uniform loss-of-fitness effects in contrast to incomplete target inhibition by a drug (Supplementary Figure 3d). Additionally, inhibition of a protein is intrinsically different than a knockout, as observed for PARP inhibitors whose activity is mediated through formation of cytotoxic PARP-DNA complexes, whereas as PARP knockout has little or no effect on cells (Gill *et al*, 2015; Murai & Pommier, 2015)(Supplementary Figure 3e). A lack of variability was much less pronounced in the drug sensitivity measurements since we only considered drugs which showed a minimal level of

activity, i.e. IC50 lower than half of the maximum screened concentration (Supplementary Figure 3f). Drugs with no significant association were also approximately 3 times less likely to be associated with a genomic biomarker linked to sensitivity (Supplementary Figure 3g). Thus, the absence of an association between drug sensitivity and CRISPR loss-of-function effects could warrant further investigation into drug mechanism-of-action to understand possible underlying factors, such as low potency, alternative molecular mechanisms, or polypharmacology. Collectively, our network analysis demonstrates that CRISPR screens can provide functional insights into drug *in cellular* activity extending beyond the direct drug target into the associated functional network.

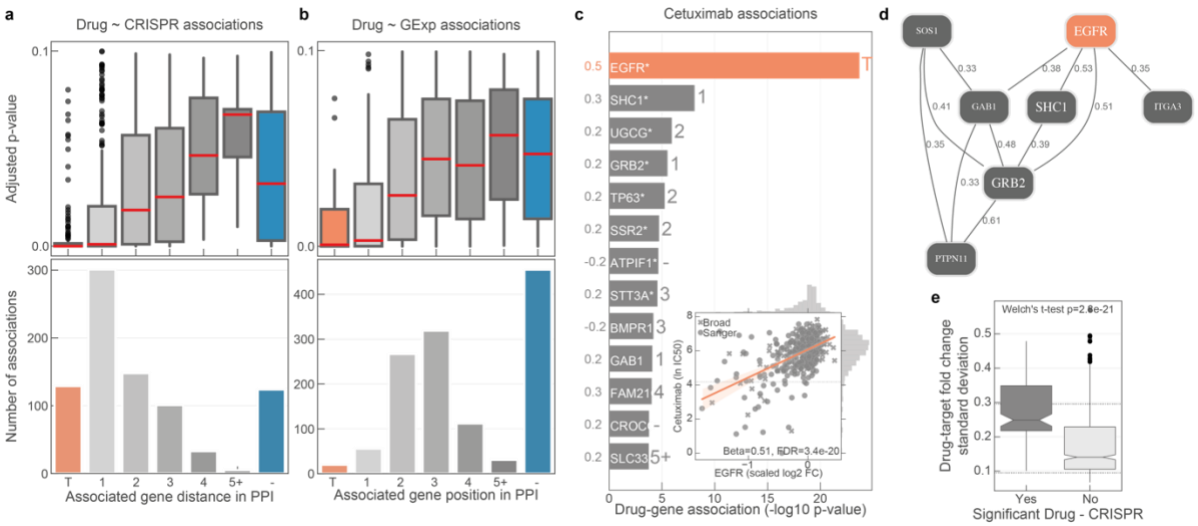



***Figure 2. Drug response protein-protein networks. a****, distribution of the FDR adjusted p-values (top) and count (bottom) of the significant drug-gene (CRISPR) associations according to their distance between the gene and corresponding drug targets in the protein-protein interaction network.* ***b****, similar to* ***a****, but instead gene expression (GExp) was tested to identify associations with drug response.* ***c****, representative example (cetuximab - EGFR inhibitor) of the associations and* ***d****, networks that can be obtained from the integrative analysis. Edges in the network are weighted with the Pearson correlation coefficient obtained between the fitness profiles of interacting nodes.* ***e****, drug-target associations stratified by statistical significance and plotted against the standard deviation of the drug-target CRISPR fold changes. Upper and lower dashed lines represent the standard deviations of essential and non-essential genes, respectively.*

## Cancer drugs mechanism-of-action

Next, we set out to investigate in detail some of the strongest drug sensitivity and gene fitness associations (Supplementary Table 5). Strikingly, 46 of the top 50 strongly associated drugs have significant associations with their nominal target and with known functionally related genes (Figure 3). Some of the strongest associations were between MCL1 inhibitors and their target fitness effects (Figure 1d), including AZD5991 which is currently in clinical trials for treatment of hematologic cancers (Hird *et al*, 2017). Additionally, for several Insulin-

Like Growth Factor 1 Receptor (*IGFR1*) inhibitors the association with the target was recapitulated. Moreover, significant associations with proprotein convertase *furin* were observed, supporting the known genetic association that *IGFR1* is a *furin* substrate, and increased levels of *furin* are associated with increased levels of processed IGFR1 and worse prognosis in several cancers (Thomas, 2002).

Protein kinase inhibitors are an important class of cancer drugs (Santos *et al*, 2017). Because of the conserved structural features of the commonly targeted kinase domain, the clinical development of kinase inhibitors is hampered by poor selectivity, which consequently may lead to undesirable off-target activity (Klaeger *et al*, 2017). Furthermore, some kinases have multiple isoforms with non-redundant roles in tissues, as exemplified by the clinical development of PI3K inhibitors, and this has led to the development of isoform-selective inhibitors to reduce toxicity and increase the therapeutic window (Thorpe *et al*, 2015). Interestingly, several PI3K inhibitors had strong associations with only one gene encoding a single isoform (Figure 3), this together with the increased kinobead binding affinity of significant associations (Figure 1e), suggests these are isoform selective compounds. For example, alpelisib (Figure 3 first row) was associated with *PIK3CA*, consistent with its development as an alpha-isoform selective compound (Thorpe *et al*, 2015), whereas AZD8186 (Figure 3 second row) was only associated with *PIK3CB* confirming its beta-selectivity. Conversely, two *pan*-PI3K inhibitors (buparlisib and omipalisib) displayed no significant association with any *PI3K* isoform (Supplementary Table 5), consistent with less isoform specificity and potential polypharmacology. Interestingly, MTOR and pan-PI3K inhibitor, dactolisib, had significant associations with *RPTOR* and *MTOR* but none with *PI3K* isoforms (Supplementary Table 5), consistent with recently reported greater specificity for inhibition of the MTOR complex (Reinecke *et al*, 2019). Similarly, we observed that selective EGFR inhibitors cetuximab, erlotinib and gefitinib (Figure 3 second and third rows) were associated with *EGFR* but not *ERBB2*, whereas afatinib, poziotinib and sapatinib (AZD8931) (Figure 3 fourth and fifth rows) were all associated with both *EGFR* and *ERBB2*. Furthermore, we observed isoform selectivity of different FGFR inhibitors.

Our analysis can also provide insights into possible off-target activity of drugs. Unsupervised clustering of the drug-gene associations effect sizes (betas) revealed classes of inhibitors with similar targets and mechanism-of-action (Supplementary Figure 3h). Of note, BTK inhibitor, ibrutinib, clustered with EGFR inhibitors and displayed significant associations with EGFR and ERBB2 gene fitness. This is consistent with recent findings that brutinib covalently bind and inhibit EGFR (Lee *et al*, 2018), and is also supported by kinobead measurements (Klaeger *et al*, 2017). Additionally, 24 compounds have significant associations

with genes identified as core fitness (Behan *et al*, 2019) across multiple cancer types, indicating an increased risk of cellular toxicity. Out of these, two compounds, PD0166285 and CCT244747, have significant associations with their nominal target (*PKMYT1* and *CHEK1*/*WEE1*) and the remaining compounds (n=22) are correlated with proteins closely connected in the PPI network.

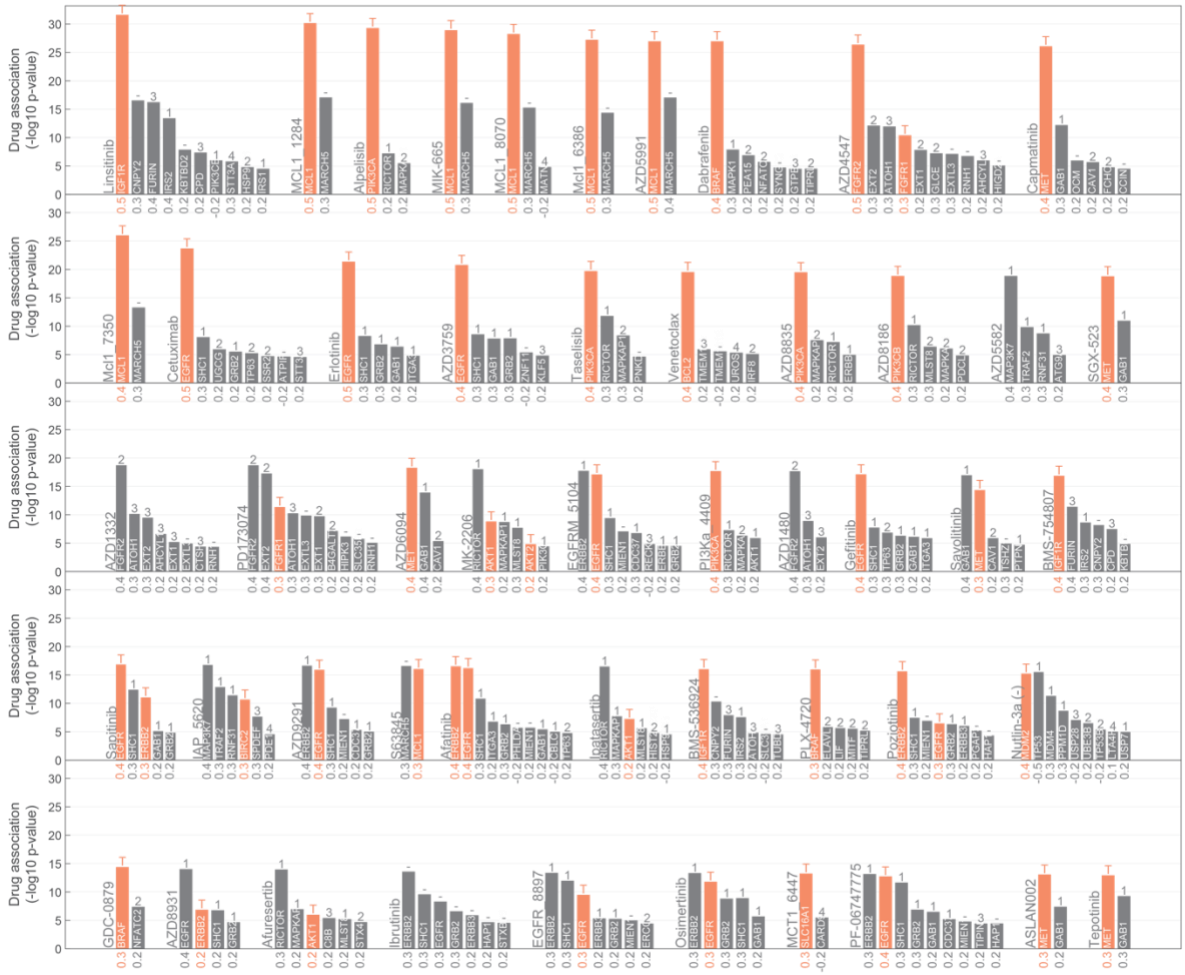



***Figure 3. Top 50 most significantly associated drugs.*** *Each bar plot group represents a unique drug where genes are ranked by statistical significance of their association. Effect sizes of the associations are reported under the bars along the x axis. Shortest distance (number of interactions) in a protein-protein interaction network between the gene and the drug nominal target(s) is represented on the top of the bars, where T and orange bar represent the target and '-' represents no link was found.*

## A functional link between *MARCH5* and MCL1 inhibitor sensitivity

Seven out of nine inhibitors of the anti-apoptotic *BCL2* family member myeloid cell leukemia 1 (*MCL1*) were strongly and nearly exclusively associated with their putative target, suggesting these are potent and specific compounds in cells (Figure 4a). *MCL1* is frequently amplified in human cancers (Beroukhim *et al*, 2010) and associated with chemotherapeutic

resistance and relapse (Wuillème-Toumi *et al*, 2005; Wei *et al*, 2006). MCL1 is a negative regulator of the mitochondrial apoptotic pathway, regulating *BAX*/*BAK1* which co-localise with *Drp1/Fis1* in the mitochondria outer membrane and control mitochondrial fragmentation and cytochrome c release, both of which are important for inducing apoptosis (Youle & Karbowski, 2005; Mojsa *et al*, 2014; Morciano *et al*, 2016). Interestingly, knockout of a key regulator of mitochondrial fission, mitochondrial E3 ubiquitin-protein ligase *MARCH5* (Karbowski et al, 2007), is significantly associated with MCL1 inhibitors sensitivity (Supplementary Figure 4a), and positively correlated with MCL1 gene fitness, suggesting a functional relationship (Figure 4b). A recent study confirmed a synthetic-lethal interaction between *MARCH5* and well know *MCL1* negative regulator *BCL2L1* using dropout screens in isogenic cancer cell lines (DeWeirdt *et al*, 2019). Correlation between *MCL1* and *MARCH5* fitness profiles shows that cell lines dependent on *MARCH5* are also dependent on *MCL1*, while the inverse is not necessarily true with a subgroup of cell lines dependent on *MCL1* but not on *MARCH5*. Cell lines independently dependent on both gene-products have increased sensitivity to MCL1 inhibitors (Supplementary Figure 4b). This is particularly marked in breast carcinoma cancer cell lines, with *MCL1* and *MARCH5* dependent cells having similar sensitivity to hematologic cancer cell lines (acute myeloid leukemia), where MCL1 inhibitors are in clinical development (Figure 4c).

We investigated the potential molecular mechanisms underlying MCL1 inhibitors response. MCL1 copy number and gene expression alone are not a good predictor of MCL1 inhibitors sensitivity (Figure 4d, Supplementary Figure 4c). This is in contrast to BCL2 and BCL2L1 inhibitors, where their target gene expression is significantly correlated with drug sensitivity (Figure 4d). Next, we used multilinear regression models to predict sensitivity to each MCL1 inhibitor using gene fitness and/or gene expression of known regulators of *MCL1* (e.g. *BCL2*, *BCL2L1*, *BAX*) (Czabotar *et al*, 2014) and *MARCH5*. For two MCL1 inhibitors, MIM1 and UMI-77, the trained models performed poorly likely due to lack of *in cellular* activity of these compounds. For the remaining seven MCL1 inhibitors, drug response was well predicted (CRISPR+GEXP mean R-squared=0.55). Models trained with only CRISPR displayed overall better predictions compared to models only trained with gene expression, and models trained with both data types out-performed all others (Figure 4e). As expected, MCL1 fitness-effect was the most predictive feature, followed by BCL2L1 expression and MARCH5 essentiality (Figure 4f). No genomic feature, mutation or copy number alterations correlated significantly with MCL1 inhibitors response, including *MCL1* amplifications (Supplementary Figure 4c), likely a consequence of the strong post-transcriptional regulation and short half-life of MCL1.

Altogether, we highlight a functional link between *MARCH5* and MCL1 inhibitors sensitivity. With further investigation, this could shed light on MCL1 inhibitor mechanism and the development of stratification approaches in solid tumours, such as breast carcinomas.

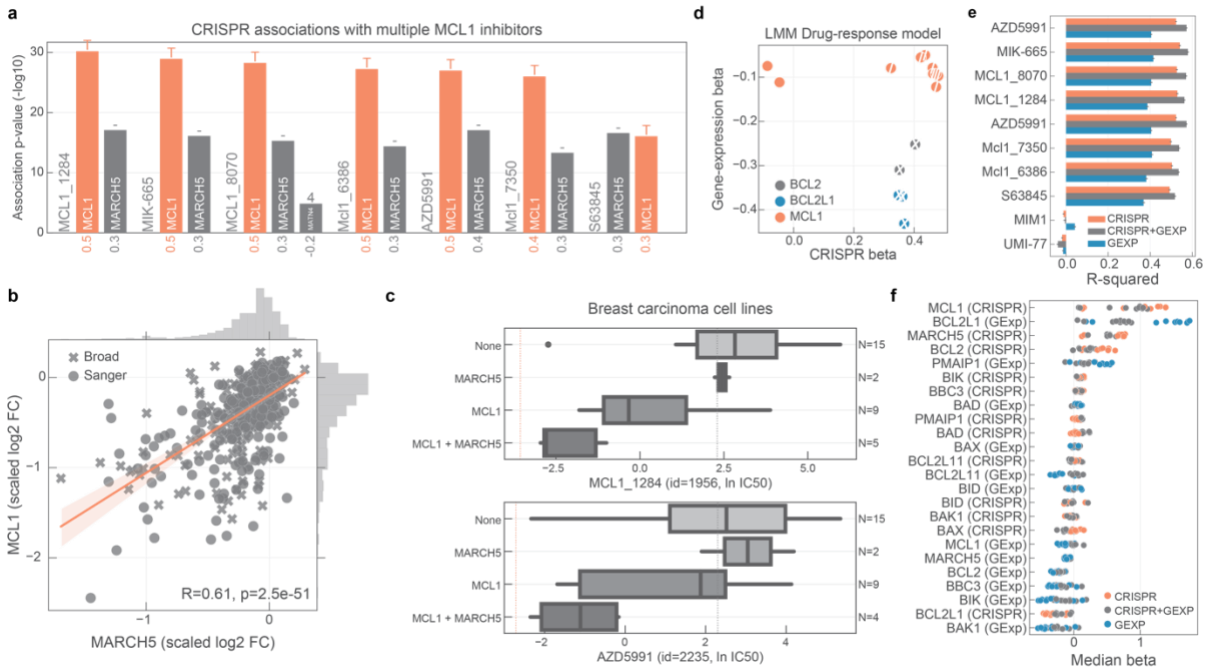



***Figure 4. MCL1 inhibitors associations.*** ***a****, significant associations with MCL1 inhibitors (7 out of 9 included in the screen).* ***b****, association between the gene fitness profiles of MCL1 and MARCH5.* ***c****, stratification of the MCL1 inhibitor sensitivity according to the essentiality profile of MCL1 and MARCH5, where MCL1 + MARCH5 represents a cell line that is independently dependent on both genes. Dashed orange line (left) represents the mean IC50 in acute myeloid leukemia cell lines. Grey dashed line (right) represents the maximum concentration used in the dosage response curve.* ***d****, BCL2, BCL2L1 and MCL1 inhibitors and the respective association with their targets, on the x axis with CRISPR gene fitness and on the y axis with gene expression. The statistical significance of the association is represented with a backward slash for CRISPR and forward slash for GEXP.* ***e****, regularised multilinear regression to predict drug response of all MCL1 inhibitors using gene expression, fitness or both of known regulators of the BCL2 family and MARCH5. Predictive performance is estimated using* $R_2$ *metric represented in the x axis.* ***f****, effect size of each feature used in each MCL1 inhibitor model.*

## Robust molecular markers of drug sensitivity networks

The identification of molecular biomarkers of drug sensitivity is fundamental to guide clinical drug development. We hypothesized that molecular biomarkers independently linked with both drug response and gene fitness would be of particularly high value – termed robust pharmacogenomic biomarkers. To identify these, we used the set of significant drug-gene pairs (n=865) and we searched independently for significant associations between each measurement type in each pair (drug response or gene fitness) and 519 genomic (mutations and copy number alterations) and 15,368 gene expression features (Figure 5a, Supplementary Figure 5a) (Garnett *et al*, 2012; Iorio *et al*, 2016; Garcia-Alonso *et al*, 2018).

This analysis recapitulated established genomic and expression biomarkers of either drug sensitivity or gene fitness effects in cancer cells (Supplementary Figure 5b and c). A total of 224 and 679 robust pharmacogenomic associations were identified with genomic (Supplementary Table 7) and gene expression features (Supplementary Table 8), respectively. Overall, 30.6% (265 of 865) of drug-gene pairs have at least one robust molecular marker that correlated significantly with both drug response and gene fitness (Figure 5b). The number of robust biomarkers was smaller than the number of biomarkers associated with only one type of measurement, likely due to the stringent requirement for an association with both drug sensitivity and gene fitness effects.

From the subset of 129 drug-gene pair associations that were linked by the drug target, 50.4% (n=65) had one or more robust pharmacogenomic associations (Supplementary Figure 5d). Most of these were established dependencies of cancer cells, including: Nutlin-3a sensitivity associated with *TP53* mutation status; *BRAF* and *PIK3CA* mutation induced CRISPR dependency; olaparib sensitivity mediated by the presence of *EWSR1-FLI1* fusion, also recapitulated by *FLI1* essentiality profile; MCL1 inhibitors biomarker association with BCL2L1, and nutlin-3a with BAX expression (Figure 5c and Supplementary Figure 5e and 5f). Similarly, of the 413 significant gene-drug pairs closely related within the PPI network (<=3 interactions from the drug target), we identified robust pharmacogenomic associations for 29.5% (n=122) (Supplementary Figure 5d), enabling the discovery of cellular contexts where drug response networks are important. For example, we identified increased tumour necrosis factor (*TNF*) expression as a robust pharmacogenomic marker for drugs targeting the downstream cellular inhibitor of apoptosis (cIAP) proteins *BIRC2* and *BIRC3* (e.g. IAP_5620), and based on CRISPR dependency data, for multiple members of the cIAP pathway, including *BIRC2*, *MAP3K7* and *RNF31* (Beug *et al*, 2012) (Figure 5d).

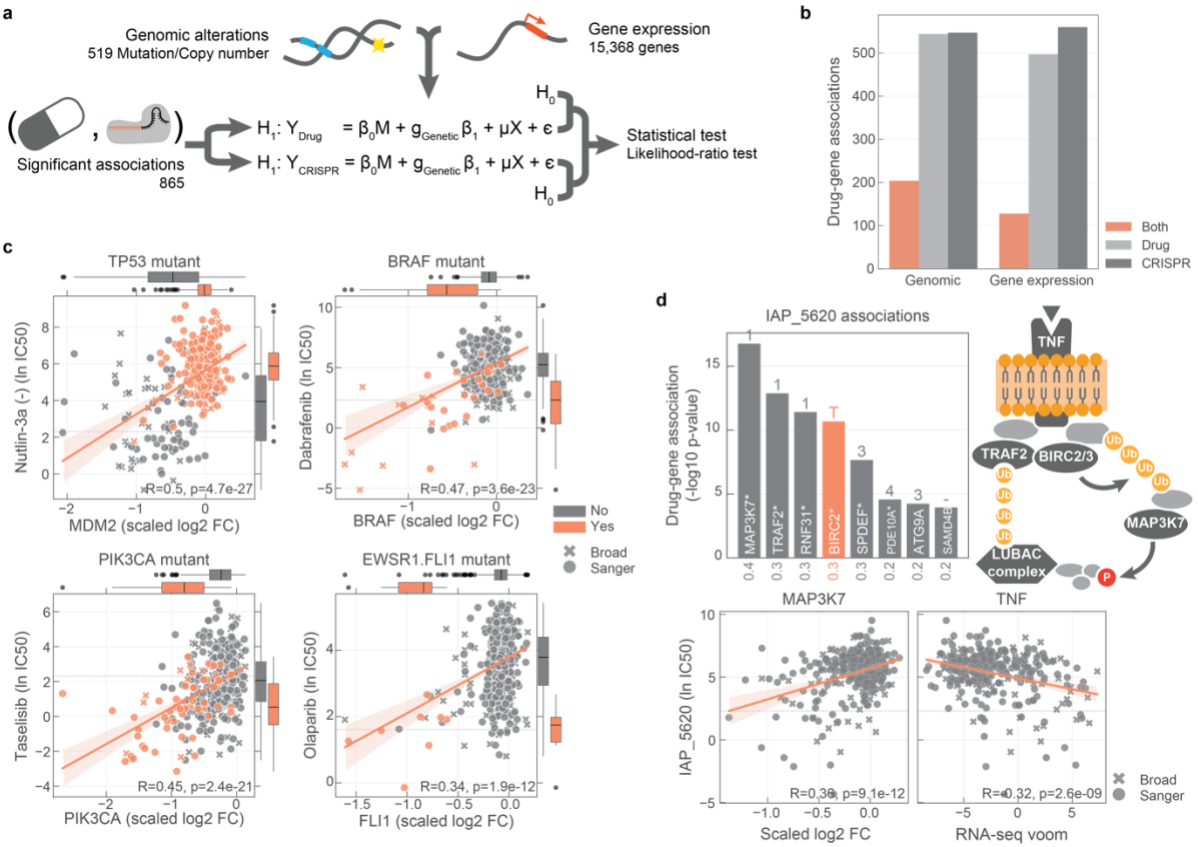



***Figure 5. Robust pharmacological associations.*** ***a****, diagram representing how genomic and gene expression data-sets are integrated to identify significant associations with drug-gene pairs that were previously found to be significantly correlated.* ***b****, number of drug-gene pairs with at least one significant association with drug response, gene fitness or both, considering either genomic or gene expression profiles.* ***c****, canonical examples of robust pharmacological associations.* ***d****, representative example of a BIRC2/BIRC3 inhibitor, IAP_5620, showing the significant associations with CRISPR gene fitness profiles and their location in a representation of the TNF pathway.*

# Discussion

Understanding drug mechanism-of-action and the biological pathways underpinning drug response is an important step in preclinical studies. Here, we demonstrate how the integration of drug sensitivity and CRISPR-Cas9 gene fitness data can be used to inform on multiple aspects of drug mechanisms in cells, including drug specificity and potency. Our analysis recapitulated drug targets for approximately a quarter of drugs tested and for approximately another quarter revealed associations enriched for proteins closely related with the drug target. Critically, the strength of these associations reflects specificity and polypharmacology of the cancer drugs. Furthermore, these associations define networks of protein interactions that are functionally related with drug targets and underpin drug response. This revealed a previously unappreciated interaction between *MARCH5* and MCL1 inhibitors, with potential utility to derive predictive models of MCL1 inhibitor response across multiple

cancer types, and particularly in solid tumours such as breast carcinomas. Robust pharmacogenomic biomarkers leveraged both datasets to provide refined biomarkers that are correlated with both drug response and biological networks. Interestingly, the networks we have defined can provide alternative targets that are functionally related with the drug target and mediate similar effects on cell fitness, potentially providing strategies for combination therapies to limit therapy resistance.

Pre-clinical biomarker development is an important step in drug discovery and is associated with increased success rates during clinical development (Nelson *et al*, 2015). Traditionally this has been performed by building predictive models of drug response using mutation, copy number and gene expression (Iorio *et al*, 2016; Tsherniak *et al*, 2017). Here we extended this approach, and propose what we term as robust pharmacogenomic association - a drug response and gene fitness pair that are significantly correlated and are also both significantly related to the same molecular biomarker. This approach gives greater confidence in molecular biomarkers identified, since they are recapitulated using data from two orthogonal assays and provides markers at the level of the network. In addition, by focusing only on drugs involved in significant gene-drug pairs, we enrich for drugs most likely to have greater specificity, and thereby better enabling biomarker discovery.

Nearly half of the drugs did not have a significant association with gene fitness effects and may warrant further investigation. Possible explanations for this include: (i) drug polypharmacology which is difficult to deconvolute using single gene knockout data; (ii) intrinsic difference between protein inhibition and knockout; (iii) a dosage dependent response leading to incomplete inhibition of the drug target; (iv) functional redundancy between protein isoforms resulting in less penetrant effects with gene knockout; and (v) limitations of the sgRNA efficacy across the cancer cell lines. We expect that some of these issues can be addressed by expanding this analysis to integrate other types of functional genomic screens, such as CRISPR inhibition, which might mimic drug inhibition more closely.

This study extends previous efforts, and utilises new CRISPR loss-of-function datasets, to study drug mechanism-of-action in cells with unparalleled scale and precision. We anticipate this approach to be useful for many compounds, and could become a routine step during drug development. In particular, it is likely to have utility during the hit-to-lead optimisation stage of drug development to select lead chemical series and compounds with optimal potency and selectivity. The utility of this approach is likely to expand as the availability of CRISPR knock-out screening data, and other datasets such as CRISPR activation and inhibition, increases across ever larger collections of highly-annotated cancer cell models. In

conclusion, this study illustrates a new approach for investigating *in cellular* drug mechanism-of-action that can be applied to multiple critical aspects of drug development.

# Materials and Methods

## Table of contents



## 1. Cancer cell lines panel

The 484 cancer cell lines used in this manuscript have been compiled from publicly available repositories as well as private collections and maintained following the supplier guidelines. STR and SNP fingerprints were used to ensure cell lines selected were genetically unique and matched those in public repositories (http://cancer.sanger.ac.uk/cell_lines/download). Detailed cell line model information is available through Cell Model Passports database (https://cellmodelpassports.sanger.ac.uk/) (van der Meer *et al*, 2019). Cell lines growth rate is represented as the ratio between the mean of the untreated negative controls measured at day 1 (time of drug treatment) and the mean of the DMSO treated negative controls at day 4 (72 hours post drug treatment).

## 2. High-throughput drug sensitivity

Experimental details of both GDSC1 and GDSC2 screens can be found in the Genomics of Drug Sensitivity in Cancer (GDSC) project (www.cancerRxgene.org) (Yang *et al*, 2013). Cell viability and dose response curve fitting models were previously described in detail (Iorio *et al*, 2016; Vis *et al*, 2016). Maximum screened drug concentration (µM) are provided

in Supplementary Table 1. Each compound was measured on average across 393 cell lines rendering a nearly complete matrix with only 14.2% missing values. All considered compounds displayed an IC50 lower than half of the maximum screened concentration in at least 3 cell lines. This filter ensures the compounds display an informative profile in at least a small subset of the cell lines. Drug nominal oncology target annotation was manually curated from literature (Supplementary Table 1).

## 3. Genome-wide CRISPR-Cas9 dropout screens

The CRISPR-Cas9 screens for the 484 cancer cell lines considered in this study (Supplementary Table 2) were assembled from two distinct projects, 320 were generated as part of Sanger DepMap Project Score (Behan *et al*, 2019) and 164 from the Broad DepMap version 19Q3 (Meyers *et al*, 2017; DepMap, 2019). Only cell lines that passed quality control filtering similarly to Behan et al. (2019) and with matched drug response measurements were considered. Different CRISPR-Cas9 sgRNA libraries were used in each project (Koike-Yusa *et al*, 2014; Doench *et al*, 2016; Tzelepis *et al*, 2016). Consequently, library-specific effects were present (Dempster *et al*, 2019) (Supplementary Figure 2a) which hampers averaging of cell lines that were screened in both data-sets. Thus, for the overlapping cell lines only data from Sanger DepMap Project Score was used. This also minimises potential cell line specific differences, for example due to genetic drift (Ben-David *et al*, 2018), and thereby increasing concordance with the drug response data-set also generated at the Wellcome Sanger Institute. Fold changes (log2) were estimated comparing samples with the respective control plasmid. As copy number profiles were not available for all of the cell lines, gene-independent deleterious effects induced by copy number amplifications in CRISPR-Cas9 screens (Aguirre *et al*, 2016; Munoz *et al*, 2016; Gonçalves *et al*, 2019) were corrected on a per sample basis using the unsupervised method CRISPRcleanR (Iorio *et al*, 2018). Replicates were mean averaged and gene level fold changes were estimated by taking the mean of all the mapping sgRNAs. Gene level fold changes were quantile normalised per sample and then median scaled using previously defined lists of cancer cell lines essential and non-essential genes (Hart *et al*, 2015), thus essential genes have a median log2 fold change of -1 and non-essential genes a median log2 fold change of 0. Only overlapping genes between the two libraries were considered, thus generating a full matrix of 16,643 genes across the 484 cell lines. A cell line was considered dependent on a gene if the knockout had a log2 fold change of at least 50% of that expected of essential genes (scaled log2 fold change < -0.5).

## 4. PCA of drug sensitivity and gene fitness

Principal component analysis (PCA) was performed using scikit-learn (v0.21.2) (Pedregosa *et al*, 2011) using sklearn.decomposition.PCA with default parameters and the number of components (n_components) set to 10. For the drug response data-set, and only for the PCA analysis, missing values of each drug were imputed using the drug mean IC50 response across the rest of the cell lines. Imputation was not required for the CRISPR-Cas9 data-set since the matrix had no missing values.

## 5. Drug response linear mixed model associations

Associations between drug response and gene fitness scores were performed using an efficient implementation of mixed-effect linear models available in the LIMIX Python module (v3.0.3) (Lippert *et al*, 2014; Casale *et al*, 2017). We considered the following covariates in the model: (i) binary variables indicating the institute of origin of the cell line CRISPR-Cas9 screen; (ii) principal component 1 of the drug response data-set which is a correlative of cell lines growth rate; and (iii) growing conditions (adherent, suspension or semi-adherent) represented as binary variables. Additionally, gene fitness similarity matrix of the samples is considered as random effects in the model to account for potential sample structure. Taken together, we fitted the following mixed linear regression model for each drug-gene pair:

[1] $$d = \beta_0 M + \beta_1 e + \mu X + \varepsilon$$

Where, *d* represents a vector of the drug response IC50 values across the cell lines; *M* is the matrix of covariates and $\beta_0$ is the vector of effect sizes; e is the vector of gene CRISPR-Cas9 log2 fold changes and $\beta_1$ the effect size; *X* the similarity matrix based on the CRISPR-Cas9 gene fitness measurements; $\mu$ is the random effects; $\varepsilon$ is the general noise term. For each drug, cell lines with missing values were dropped from the fit.

We statistically assessed the significance of each association by performing likelihood ratio tests between the alternative model ($\widehat{\theta_1}$) and the null model which excludes the gene CRISPR gene fitness scores vector *e* and its parameter $\beta_1(\widehat{\theta_0})$. The parameter inference is performed using maximum likelihood estimation:

[2] $$\hat{\theta} = argmax\, p(d \mid M, X;\ \theta)$$

And the p-value of the association is defined by:

[3] $\frac{p(d \mid M,X;\widehat{\theta_0})}{p(d \mid M,X;\widehat{\theta_1})}$

We tested all the single-feature pairwise associations between the 480 compounds and the 16,643 genes, making a total of 7,988,640 tested associations. P-value adjustment for multiple testing was performed per drug using the Benjamini-Hochberg False Discovery Rate (FDR). Contrary to performing the adjustment across all tests, per drug correction has the following benefits: (i) associations assembled from the different screening platforms (GDSC1 and GDSC2) are kept separate hence not biasing for measurement type; and (ii) drugs with stronger responses across larger subsets of cancer cell lines, for example Nutlin-3a response across TP53 wild-type cell lines, display stronger associations than most drugs, thus correcting across all drugs would retain more associations from these drugs at a specific error rate, i.e. 10%, compared to the rest.

## 6. Protein-protein interaction network

We assembled from STRING database (Szklarczyk *et al*, 2017) a high confidence undirected protein-protein interaction network. We only consider interactions with a combined confidence score higher than 900. Nodes' STRING identifiers were converted to HUGO gene symbols, nodes not mapping or with multiple mappings were removed. Using *igraph* Python wrapper (Csardi & Nepusz, 2006) the network was simplified by removing unconnected nodes, self-loops and duplicated edges, leaving a total of 10,587 nodes and 205,251 interactions. A weighted version of the network was also assembled by correlating the gene fitness profiles of the connected nodes. Network nodes, and corresponding edges, that were not covered by the CRISPR-Cas9 screens were removed, making a total of 9,595 nodes and 172,584 weighted interactions.

## 7. Robust pharmacogenomic associations

Robust pharmacological associations were estimated similarly to the previous associations, but in this case only drug-gene pairs that are significantly correlated were considered to test associations with the genomic features (binarised copy number and mutation status (Iorio *et al*, 2016)) and gene expression profiles (RNA-seq voom (Law *et al*, 2014) transformed RPKMs (Garcia-Alonso *et al*, 2018)). A robust pharmacogenomic association is defined as: (i) a drug-gene pair whose drug sensitivity and gene fitness is significantly correlated, and (ii) genomic alteration or gene expression profile is significantly correlated with both drug response and gene fitness. Log-ratio test p-values are independently estimated for drug response and gene fitness measurements and corrected per drug-gene.

589 Drug-gene pairs associated to a genomic or gene expression feature with an FDR lower than
590 10% are called robust pharmacogenomic associations (Supplementary Tables 7 and 8).

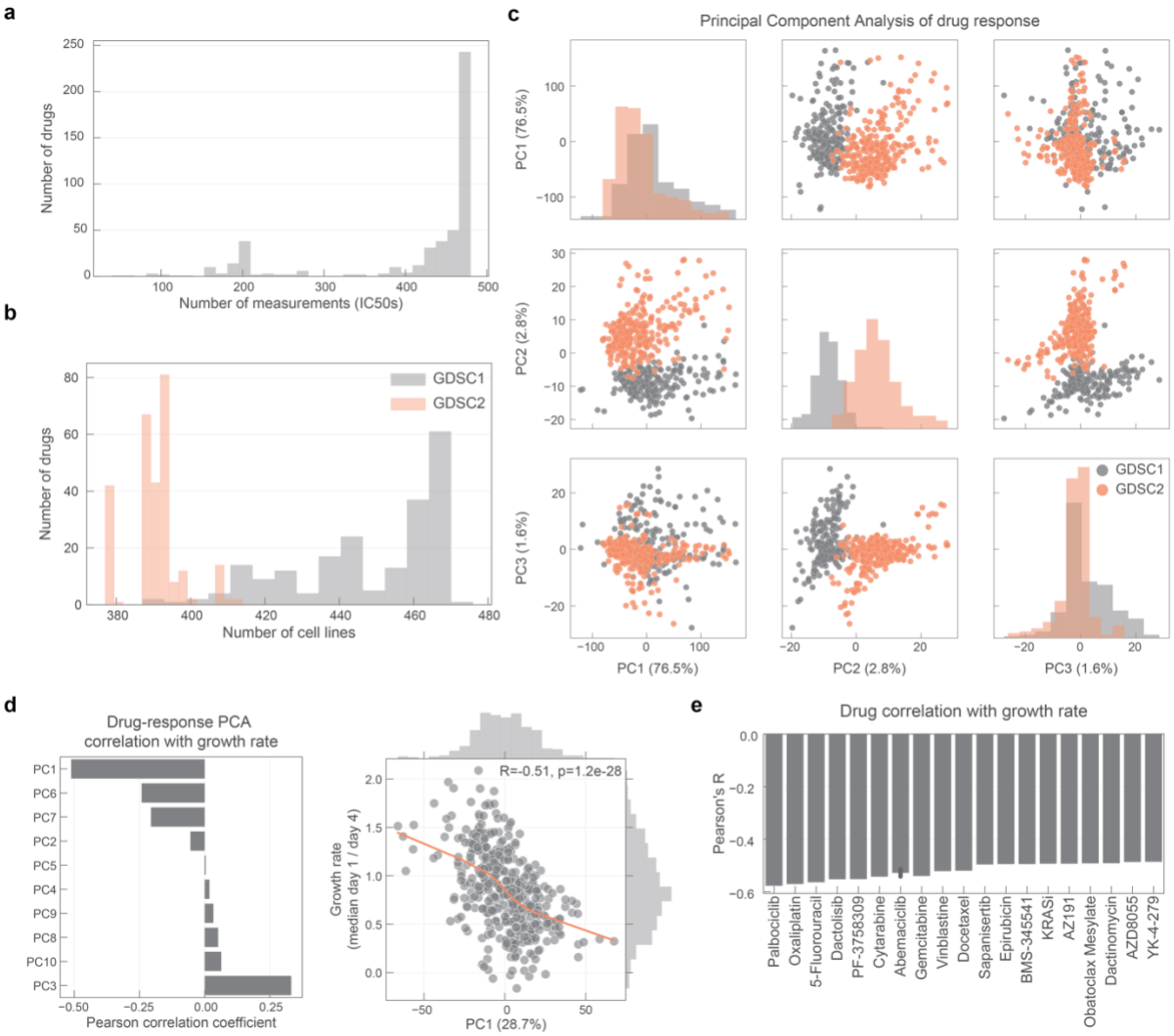
## 591 8. Predictive models of drug response of MCL1 inhibitors

592 L2-regularised linear regression models to predict MCL1 inhibitors drug response were
593 trained using gene fitness, gene expression measurements or both of canonical regulators of
594 *MCL1*, namely *MARCH5*, *MCL1*, *BCL2*, *BCL2L1*, *BCL2L11*, *PMAIP1*, *BAX*, *BAK1*, *BBC3*, *BID*,
595 *BIK*, BAD. For the 9 MCL1 inhibitors considered in this study predictive models of drug
596 response measurements were trained using Ridge regressions with an internal cross-
597 validation optimisation of the regularization parameter, implemented in Sklearn with RidgeCV
598 class (Pedregosa *et al*, 2011). Additionally, drug response measurements are split randomly
599 1,000 times, where 70% of the measurements are for training the model and 30% are left out
600 as a test set. Model's performance is quantified using the $R_2$ metric on the test set, comparing
601 the predicted versus the observed drug response measurements.

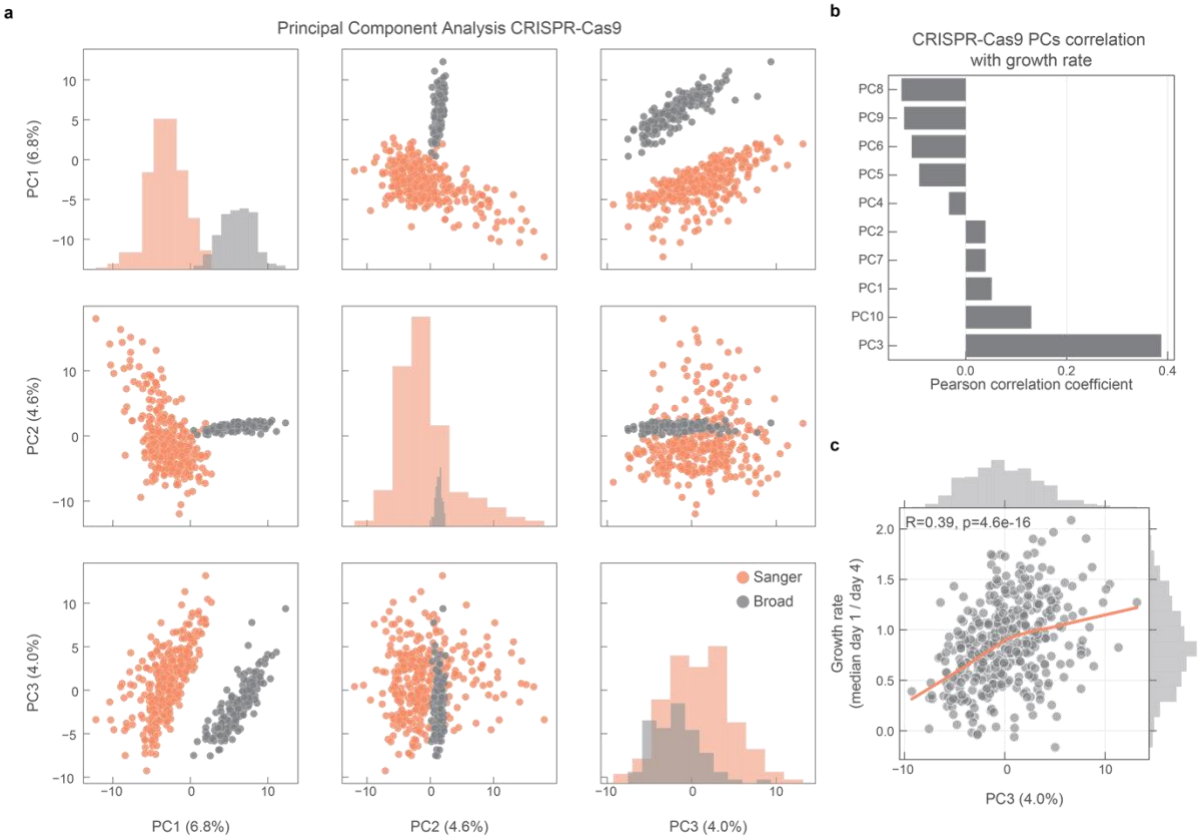
## 602 9. Code and data availability

603 Source code, analysis reports and Jupyter notebooks are publicly available in GitHub
604 project https://github.com/EmanuelGoncalves/dtrace. Drug response and gene fitness
605 CRISPR-Cas9 data-sets used in this analysis are available in the supplementary tables and
606 accessible through figshare on https://doi.org/10.6084/m9.figshare.10338413.v1.
607

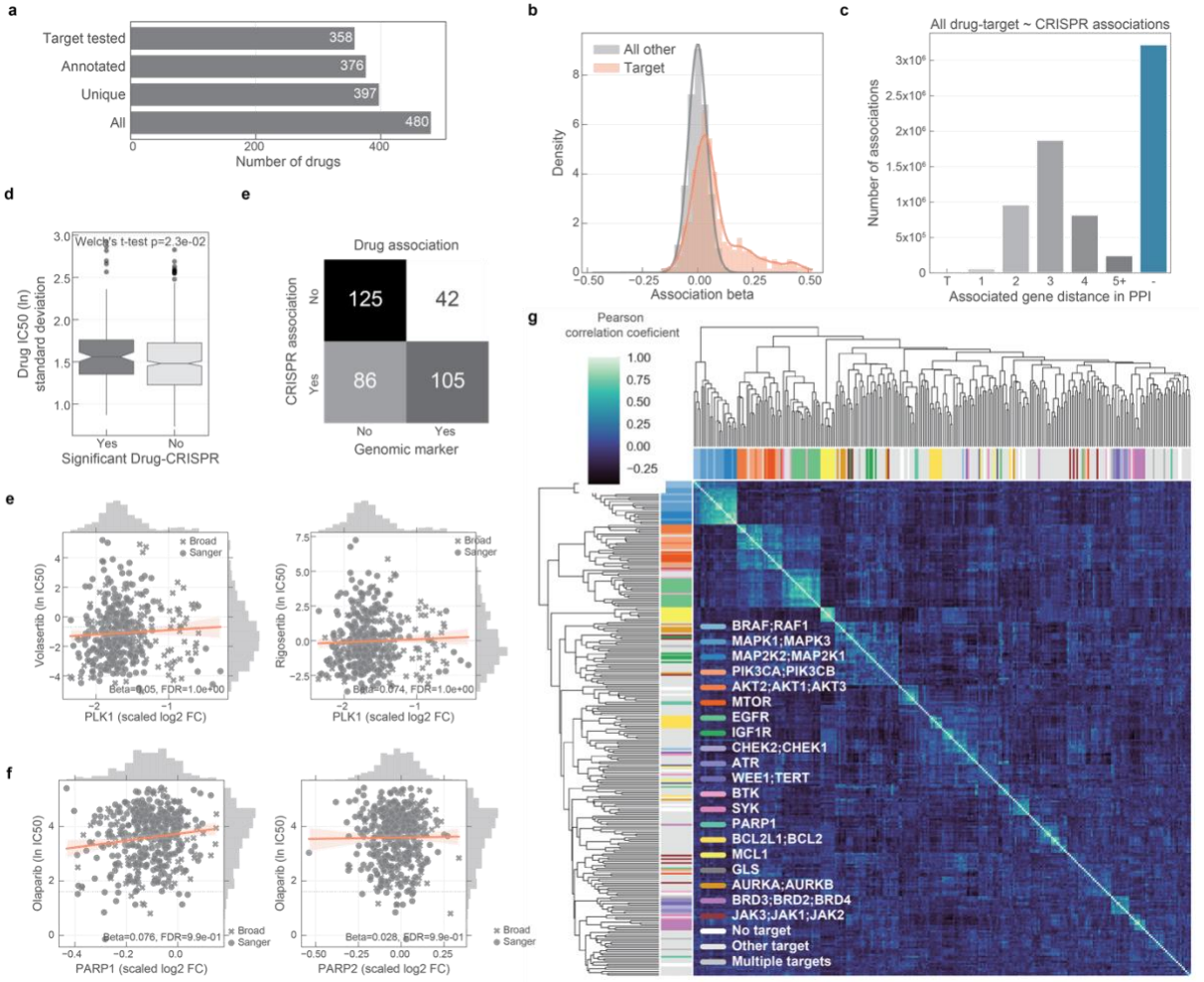
# Supplementary Figures



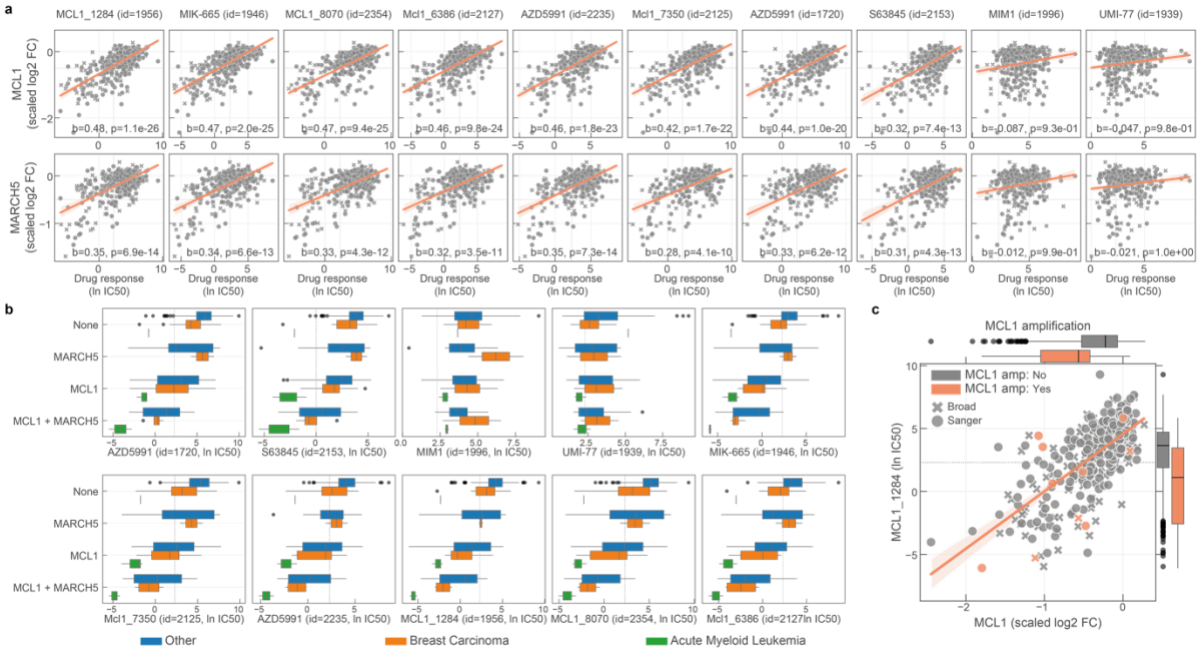

***Supplementary Figure 1. Overview of the drug sensitivity datasets.*** ***a****, histogram of the number of IC50 values measured per drug.* ***b****, number of drugs measured per cell line in each pharmacological dataset.* ***c****, PCA analysis of the drug response measurements separated by the screen type.* ***d****, Pearson correlation coefficient between each principal component (PC) and cell lines growth rate.* ***e****, top absolutely correlated drugs with growth rate.*



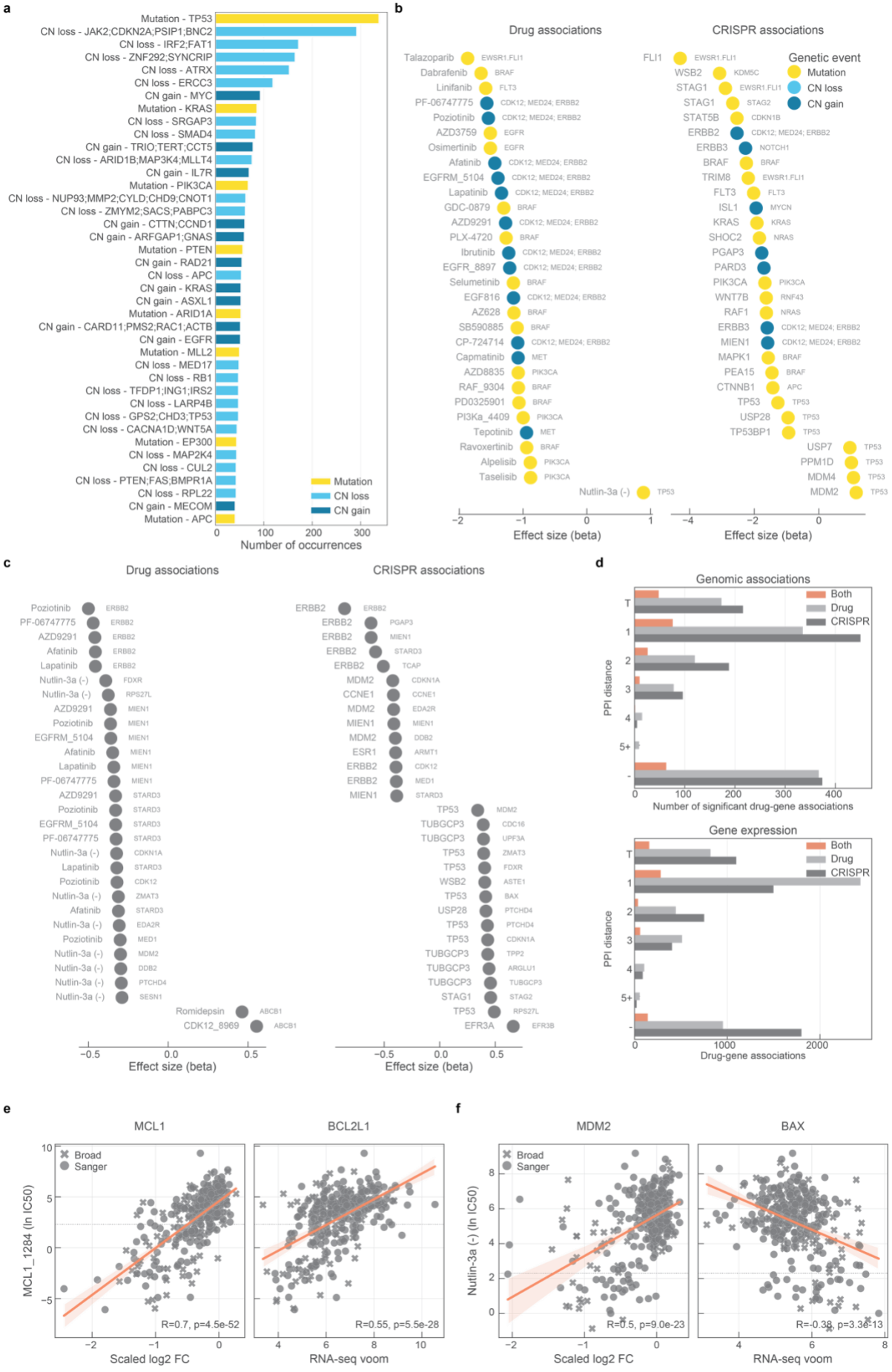

***Supplementary Figure 2. Overview of the CRISPR-Cas9 datasets.*** ***a****, PCA analysis of the samples in the CRISPR-Cas9 screens, samples institute of origin is highlighted.* ***b****, correlation coefficients between all top 10 PCs and growth rate.* ***c****, correlation between cell lines growth rate and PC3 (Pearson correlation coefficient reported in the top left).*




***Supplementary Figure 3. Drug response and gene fitness associations. a****, total number of drugs utilised in the study and the different levels of information available: 'All' represents all the drugs including replicates screened with different technologies (GDSC1 and GDSC2); 'Unique' counts the number of unique drug names; 'Annotated' shows the number of unique drugs with manual annotation of nominal targets; and 'Target tested' represents the number of unique drugs, with target information, for which the target has been knocked-out in the CRISPR-Cas9 screens.* ***b****, histogram of the drug-gene associations effect sizes (beta) highlighting drug-target associations.* ***c****, distribution of the shortest path lengths between all the tested drug-gene pairs. For drugs with multiple targets the smallest shortest path of all the targets was taken.* ***d****, PLK1 inhibitors drug response correlation with PLK1 knockout log2 fold change (FC) gene fitness effects. The dashed grey line indicates the dose response highest drug concentration.* ***e****, similar to d, correlation of olaparib drug response and both targets PARP1 and PARP2 gene fitness effects.* ***f****, drug-target associations split by significance (FDR < 10%) plotted against the standard deviation of the drug IC50 (ln) measurements of the respective pair.* ***g****, contingency matrix of significant drug associations with CRISPR fold changes and binarised event matrix of genomic features, i.e. mutations and copy number gain or loss.* ***h****, correlation heatmap of the drug-gene effect size across all the genes. Drugs are coloured according to their targets.*




***Supplementary Figure 4. MCL1 inhibitors.*** ***a****, correlation of all MCL1 inhibitor IC50 values against MCL1 and MARCH5 gene fitness profiles. Effect sizes (b) and FDR (p) of the association are reported on the bottom.* ***b****, stratification of the MCL1 inhibitors drug response measurements according to the cell line dependency on MARCH5 and/or MCL1. Gene vulnerabilities are independent from each other, meaning knockouts were introduced independently and not at the same time. Responses are then split according to the cancer type of the cell lines. Vulnerable cell lines to MARCH5 and MCL1 knockout were defined as those with a depletion of at least 50% of that visible for essential genes (scaled log2 fold change < -0.5).* ***c****, representative example of a MCL1 inhibitor and their relation with MCL1 gene fitness, with cell lines containing copy number amplification of MCL1 highlighted in orange. Copy number amplified cells were defined taking into consideration their ploidy status, cells with (ploidy <= 2.7 and copy number >= 5) or (ploidy > 2.7 and copy number >= 9) were considered as having MCL1 amplified.*




647

***Supplementary Figure 5. Robust pharmacological associations. a**, most frequent genomic alterations across the cancer cell lines. Most significant associations between **b**, genomic features and **c**, gene expression profiles with drug response and gene fitness. **d**, number of significant drug-gene pairs across the different types of interactions. Drug-gene pairs were categorised considering the shortest path length between the drug targets and the associated gene. **e**, robust pharmacological association between the expression of BCL2L1 and the significantly correlated pair of MCL1_1284 drug and MCL1 gene fitness profile. **f**, similarly to **e**, but instead it represents a robust pharmacological association between BAX and MDM2 and Nutlin-3a.*

# Supplementary Table Legends

**Supplementary Table 1.** Annotated list of the cancer cell lines used in this study.

**Supplementary Table 2.** List of cancer drugs considered in the study.

**Supplementary Table 3.** Drug response matrix (natural log IC50s) for 480 cancer drugs.

**Supplementary Table 4.** Gene fitness CRISPR-Cas9 scaled fold change.

**Supplementary Table 5.** Significant drug-CRISPR associations.

**Supplementary Table 6.** Significant drug-gene expression associations.

**Supplementary Table 7.** Significant robust pharmacogenomic associations with genomic mutation and copy number alterations.

**Supplementary Table 8.** Significant robust pharmacogenomic associations with gene expression.

# Acknowledgements

We would like to thank all members of the Translational Cancer Genomics team who provided useful insight. We are grateful for the diligent contributions of Jon Winter-Holt in terms of compound identification and data clearance and Pedro Beltrao for helpful discussions. Work in M.J.G lab was funded by Wellcome (206194) and AstraZeneca.

# Author contributions

Conceptualization E.G. and M.G.; Formal analysis E.G.; Data curation E.G., C.P., D.v.d.M., A.B., H.L., J.L., B.S., C.C., F.I., S.F. and M.G.; Drug response acquisition and processing D.v.d.M., A.B., H.L. and GDSC Screening Team; Drug annotation E.G., A.S., G.P., F.M.B, P.J., E.C., A.L., C.C. and M.J.G.; Writing original draft preparation E.G. and M.G.; Writing, reviewing and editing all authors; Visualisation E.G.; Supervision: A.L., J.L., B.S., C.C., F.I., S.F. and M.J.G.; Funding acquisition: S.F. and M.J.G.

# Conflict of interest

This work was funded in part by AstraZeneca. M.J.G. receives funding from AstraZeneca. M.J.G. and F.I. receive funding from Open Targets, a public-private initiative involving academia and industry. M.J.G. and F.I. perform consulting for the CRUK-AstraZeneca Functional Genomics Centre. J.T.L., B.S., C.C. and S.F. are current employees of AstraZeneca and hold stock in AstraZeneca.

# References

Aguirre AJ, Meyers RM, Weir BA, Vazquez F, Zhang C-Z, Ben-David U, Cook A, Ha G, Harrington WF, Doshi MB, Kost-Alimova M, Gill S, Xu H, Ali LD, Jiang G, Pantel S, Lee Y, Goodale A, Cherniack AD, Oh C, et al (2016) Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov.* **6:** 914–929

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehár J, Kryukov GV, Sonkin D, Reddy A, Liu M, Murray L, Berger MF, Monahan JE, Morais P, Meltzer J, Korejwa A, Jané-Valbuena J, Mapa FA, et al (2012) The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483:** 603–607

Behan FM, Iorio F, Picco G, Gonçalves E, Beaver CM, Migliardi G, Santos R, Rao Y, Sassi F, Pinnelli M, Ansari R, Harper S, Jackson DA, McRae R, Pooley R, Wilkinson P, van der Meer D, Dow D, Buser-Doepner C, Bertotti A, et al (2019) Prioritization of cancer therapeutic targets using CRISPR-Cas9 screens. *Nature* **568:** 511–516

Ben-David U, Siranosian B, Ha G, Tang H, Oren Y, Hinohara K, Strathdee CA, Dempster J, Lyons NJ, Burns R, Nag A, Kugener G, Cimini B, Tsvetkov P, Maruvka YE, O'Rourke R, Garrity A, Tubelli AA, Bandopadhayay P, Tsherniak A, et al (2018) Genetic and transcriptional evolution alters cancer cell line drug response. *Nature* **560:** 325–330

Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, Barretina J, Boehm JS, Dobson J, Urashima M, Mc Henry KT, Pinchback RM, Ligon AH, Cho Y-J, Haery L, Greulich H, Reich M, Winckler W, Lawrence MS, Weir BA, et al (2010) The landscape of somatic copy-number alteration across human cancers. *Nature* **463:** 899–905

Beug ST, Cheung HH, LaCasse EC & Korneluk RG (2012) Modulation of immune signalling by inhibitors of apoptosis. *Trends Immunol.* **33:** 535–545

Casale FP, Horta D, Rakitsch B & Stegle O (2017) Joint genetic analysis using variant sets reveals polygenic gene-context interactions. *PLoS Genet.* **13:** e1006693

Cook D, Brown D, Alexander R, March R, Morgan P, Satterthwaite G & Pangalos MN (2014) Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework. *Nat. Rev. Drug Discov.* **13:** 419–431

Csardi G & Nepusz T (2006) The igraph software package for complex network research. *InterJournal* **Complex Systems:** Available at: http://igraph.org

Czabotar PE, Lessene G, Strasser A & Adams JM (2014) Control of apoptosis by the BCL-2 protein family: implications for physiology and therapy. *Nat. Rev. Mol. Cell Biol.* **15:** 49–63

Deans RM, Morgens DW, Ökesli A, Pillay S, Horlbeck MA, Kampmann M, Gilbert LA, Li A, Mateo R, Smith M, Glenn JS, Carette JE, Khosla C & Bassik MC (2016) Parallel shRNA and CRISPR-Cas9 screens enable antiviral drug target identification. *Nat. Chem. Biol.* **12:** 361–366

Dempster JM, Pacini C, Pantel S, Behan FM, Green T, Krill-Burger J, Beaver CM, Zhivich V, Najgebauer H, Allen F, Gonçalves E, Shepherd R, Doench JG, Yusa K, Vazquez F, Parts L, Boehm JS, Golub TR, Hahn WC, Root DE, et al (2019) Agreement between two large pan-cancer CRISPR-Cas9 gene dependency datasets. *bioRxiv***:** 604447 Available at: https://www.biorxiv.org/content/10.1101/604447v2 [Accessed May 31, 2019]

DepMap B (2019) DepMap 19Q3 Public. Available at: https://figshare.com/articles/DepMap_19Q3_Public/9201770

DeWeirdt PC, Sanson KR, Hanna RE, Hegde M, Sangree AK, Strand C, Persky NS & Doench JG (2019) Genetic screens in isogenic mammalian cell lines without single cell cloning. *bioRxiv***:** 677385 Available at: https://www.biorxiv.org/content/10.1101/677385v1 [Accessed June 24, 2019]

Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, Smith I, Tothova Z, Wilen C, Orchard R, Virgin HW, Listgarten J & Root DE (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34:** 184–191

Garcia-Alonso L, Iorio F, Matchan A, Fonseca N, Jaaks P, Peat G, Pignatelli M, Falcone F, Benes CH, Dunham I, Bignell G, McDade SS, Garnett MJ & Saez-Rodriguez J (2018) Transcription Factor Activities Enhance Markers of Drug Sensitivity in Cancer. *Cancer Res.* **78:** 769–780

Garnett MJ, Edelman EJ, Heidorn SJ, Greenman CD, Dastur A, Lau KW, Greninger P, Thompson IR, Luo X, Soares J, Liu Q, Iorio F, Surdez D, Chen L, Milano RJ, Bignell GR, Tam AT, Davies H, Stevenson JA, Barthorpe S, et al (2012) Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483:** 570–575

Gill SJ, Travers J, Pshenichnaya I, Kogera FA, Barthorpe S, Mironenko T, Richardson L, Benes CH, Stratton MR, McDermott U & Others (2015) Combinations of PARP inhibitors with temozolomide drive PARP1 trapping and apoptosis in Ewing's sarcoma. *PLoS One* **10:** e0140988

Gonçalves E, Behan FM, Louzada S, Arnol D, Stronach EA, Yang F, Yusa K, Stegle O, Iorio F & Garnett MJ (2019) Structural rearrangements generate cell-specific, gene-independent CRISPR-Cas9 loss of fitness effects. *Genome Biol.* **20:** 27

Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M, Fradet-Turcotte A, Sun S, Mero P, Dirks P, Sidhu S, Roth FP, Rissland OS, Durocher D, Angers S & Moffat J (2015) High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* **163:** 1515–1526

Hird AW, Paul Secrist J, Adam A, Belmonte MA, Gangl E, Gibbons F, Hargreaves D, Johannes JW, Kazmirski SL, Kettle JG, Kurtz SE, Lamb ML, Packer MJ, Peng B,

762 Stewart CR, Tyner JW, Yang W, Ye Q, Zheng X & Clark EA (2017) Abstract DDT01-02:
763 AZD5991: A potent and selective macrocyclic inhibitor of Mcl-1 for treatment of
764 hematologic cancers. *Cancer Res.* **77:** DDT01–02–DDT01–02

765 Iorio F, Behan FM, Gonçalves E, Bhosle SG, Chen E, Shepherd R, Beaver C, Ansari R,
766 Pooley R, Wilkinson P, Harper S, Butler AP, Stronach EA, Saez-Rodriguez J, Yusa K &
767 Garnett MJ (2018) Unsupervised correction of gene-independent cell responses to
768 CRISPR-Cas9 targeting. *BMC Genomics* **19:** 604

769 Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, Aben N, Gonçalves
770 E, Barthorpe S, Lightfoot H, Cokelaer T, Greninger P, van Dyk E, Chang H, de Silva H,
771 Heyn H, Deng X, Egan RK, Liu Q, Mironenko T, et al (2016) A Landscape of
772 Pharmacogenomic Interactions in Cancer. *Cell* **166:** 740–754

773 Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA & Charpentier E (2012) A
774 programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.
775 *Science* **337:** 816–821

776 Jost M & Weissman JS (2018) CRISPR Approaches to Small Molecule Target Identification.
777 *ACS Chem. Biol.* **13:** 366–375

778 Karbowski M, Neutzner A & Youle RJ (2007) The mitochondrial E3 ubiquitin ligase MARCH5
779 is required for Drp1 dependent mitochondrial division. *J. Cell Biol.* **178:** 71–84

780 Klaeger S, Heinzlmeir S, Wilhelm M, Polzer H, Vick B, Koenig P-A, Reinecke M, Ruprecht B,
781 Petzoldt S, Meng C, Zecha J, Reiter K, Qiao H, Helm D, Koch H, Schoof M, Canevari G,
782 Casale E, Depaolini SR, Feuchtinger A, et al (2017) The target landscape of clinical
783 kinase drugs. *Science* **358:** eaan4368

784 Koike-Yusa H, Li Y, Tan E-P, Velasco-Herrera MDC & Yusa K (2014) Genome-wide
785 recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA
786 library. *Nat. Biotechnol.* **32:** 267–273

787 Law CW, Chen Y, Shi W & Smyth GK (2014) voom: Precision weights unlock linear model
788 analysis tools for RNA-seq read counts. *Genome Biol.* **15:** R29

789 Lee J-K, Liu Z, Sa JK, Shin S, Wang J, Bordyuh M, Cho HJ, Elliott O, Chu T, Choi SW,
790 Rosenbloom DIS, Lee I-H, Shin YJ, Kang HJ, Kim D, Kim SY, Sim M-H, Kim J, Lee T,
791 Seo YJ, et al (2018) Pharmacogenomic landscape of patient-derived tumor cells informs
792 precision oncology therapy. *Nat. Genet.* **50:** 1399–1411

793 Li J, Zhao W, Akbani R, Liu W, Ju Z, Ling S, Vellano CP, Roebuck P, Yu Q, Eterovic AK,
794 Byers LA, Davies MA, Deng W, Gopal YNV, Chen G, von Euw EM, Slamon D, Conklin
795 D, Heymach JV, Gazdar AF, et al (2017) Characterization of Human Cancer Cell Lines
796 by Reverse-phase Protein Arrays. *Cancer Cell* **31:** 225–239

797 Lin A, Giuliano CJ, Palladino A, John KM, Abramowicz C, Yuan ML, Sausville EL, Lukow
798 DA, Liu L, Chait AR, Galluzzo ZC, Tucker C & Sheltzer JM (2019) Off-target toxicity is a
799 common mechanism of action of cancer drugs undergoing clinical trials. *Sci. Transl.*
800 *Med.* **11:** Available at: http://dx.doi.org/10.1126/scitranslmed.aaw8412

801 Lippert C, Casale FP, Rakitsch B & Stegle O (2014) LIMIX: genetic analysis of multiple traits.
802 *bioRxiv***:** 003905 Available at: http://www.biorxiv.org/content/early/2014/05/22/003905
803 [Accessed December 14, 2017]

804 Lynch JT, McEwen R, Crafter C, McDermott U, Garnett MJ, Barry ST & Davies BR (2016)

Identification of differential PI3K pathway target dependencies in T-cell acute lymphoblastic leukemia through a large cancer cell panel screen. *Oncotarget* **7:** 22128–22139

Marcotte R, Sayad A, Brown KR, Sanchez-Garcia F, Reimand J, Haider M, Virtanen C, Bradner JE, Bader GD, Mills GB, Pe'er D, Moffat J & Neel BG (2016) Functional Genomic Landscape of Human Breast Cancer Drivers, Vulnerabilities, and Resistance. *Cell* **164:** 293–309

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

Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, Dharia NV, Montgomery PG, Cowley GS, Pantel S, Goodale A, Lee Y, Ali LD, Jiang G, Lubonja R, Harrington WF, Strickland M, Wu T, Hawes DC, Zhivich VA, et al (2017) Computational correction of copy number effect improves specificity of CRISPR-Cas9 essentiality screens in cancer cells. *Nat. Genet.* **49:** 1779–1784

Mojsa B, Lassot I & Desagher S (2014) Mcl-1 ubiquitination: unique regulation of an essential survival protein. *Cells* **3:** 418–437

Morciano G, Giorgi C, Balestra D, Marchi S, Perrone D, Pinotti M & Pinton P (2016) Mcl-1 involvement in mitochondrial dynamics is associated with apoptotic cell death. *Mol. Biol. Cell* **27:** 20–34

Munoz DM, Cassiani PJ, Li L, Billy E, Korn JM, Jones MD, Golji J, Ruddy DA, Yu K, McAllister G, DeWeck A, Abramowski D, Wan J, Shirley MD, Neshat SY, Rakiec D, de Beaumont R, Weber O, Kauffmann A, McDonald ER 3rd, et al (2016) CRISPR Screens Provide a Comprehensive Assessment of Cancer Vulnerabilities but Generate False-Positive Hits for Highly Amplified Genomic Regions. *Cancer Discov.* **6:** 900–913

Murai J & Pommier Y (2015) Classification of PARP Inhibitors Based on PARP Trapping and Catalytic Inhibition, and Rationale for Combinations with Topoisomerase I Inhibitors and Alkylating Agents. In *PARP Inhibitors for Cancer Therapy*, Curtin NJ & Sharma RA (eds) pp 261–274. Cham: Springer International Publishing

Nelson MR, Tipney H, Painter JL, Shen J, Nicoletti P, Shen Y, Floratos A, Sham PC, Li MJ, Wang J, Cardon LR, Whittaker JC & Sanseau P (2015) The support of human genetic evidence for approved drug indications. *Nat. Genet.* **47:** 856–860

Pan J, Meyers RM, Michel BC, Mashtalir N, Sizemore AE, Wells JN, Cassel SH, Vazquez F, Weir BA, Hahn WC, Marsh JA, Tsherniak A & Kadoch C (2018) Interrogation of Mammalian Protein Complex Structure, Function, and Membership Using Genome-Scale Fitness Screens. *Cell Syst* **0:** Available at: http://dx.doi.org/10.1016/j.cels.2018.04.011 [Accessed May 17, 2018]

Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M & Duchesnay É (2011) Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12:** 2825–2830

Picco G, Chen ED, Alonso LG, Behan FM, Gonçalves E, Bignell G, Matchan A, Fu B, Banerjee R, Anderson E, Butler A, Benes CH, McDermott U, Dow D, Iorio F, Stronach E, Yang F, Yusa K, Saez-Rodriguez J & Garnett MJ (2019) Functional linkage of gene fusions to cancer cell fitness assessed by pharmacological and CRISPR-Cas9

screening. *Nat. Commun.* **10:** 2198

Reinecke M, Ruprecht B, Poser S, Wiechmann S, Wilhelm M, Heinzlmeir S, Kuster B & Médard G (2019) Chemoproteomic Selectivity Profiling of PIKK and PI3K Kinase Inhibitors. *ACS Chem. Biol.* **14:** 655–664

Santos R, Ursu O, Gaulton A, Bento AP, Donadi RS, Bologa CG, Karlsson A, Al-Lazikani B, Hersey A, Oprea TI & Overington JP (2017) A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discov.* **16:** 19–34

Scaltriti M & Baselga J (2006) The epidermal growth factor receptor pathway: a model for targeted therapy. *Clin. Cancer Res.* **12:** 5268–5272

Schenone M, Dančík V, Wagner BK & Clemons PA (2013) Target identification and mechanism of action in chemical biology and drug discovery. *Nat. Chem. Biol.* **9:** 232–240

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelson T, Heckl D, Ebert BL, Root DE, Doench JG & Zhang F (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343:** 84–87

Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, Gould J, Davis JF, Tubelli AA, Asiedu JK, Lahr DL, Hirschman JE, Liu Z, Donahue M, Julian B, Khan M, Wadden D, Smith IC, Lam D, Liberzon A, et al (2017) A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171:** 1437–1452.e17

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P, Jensen LJ & von Mering C (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45:** D362–D368

Thomas G (2002) Furin at the cutting edge: from protein traffic to embryogenesis and disease. *Nat. Rev. Mol. Cell Biol.* **3:** 753–766

Thorpe LM, Yuzugullu H & Zhao JJ (2015) PI3K in cancer: divergent roles of isoforms, modes of activation and therapeutic targeting. *Nat. Rev. Cancer* **15:** 7–24

Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, Gill S, Harrington WF, Pantel S, Krill-Burger JM, Meyers RM, Ali L, Goodale A, Lee Y, Jiang G, Hsiao J, Gerath WFJ, Howell S, Merkel E, Ghandi M, et al (2017) Defining a Cancer Dependency Map. *Cell* **170:** 564–576.e16

Tzelepis K, Koike-Yusa H, De Braekeleer E, Li Y, Metzakopian E, Dovey OM, Mupo A, Grinkevich V, Li M, Mazan M, Gozdecka M, Ohnishi S, Cooper J, Patel M, McKerrell T, Chen B, Domingues AF, Gallipoli P, Teichmann S, Ponstingl H, et al (2016) A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. *Cell Rep.* **17:** 1193–1205

Vis DJ, Bombardelli L, Lightfoot H, Iorio F, Garnett MJ & Wessels LF (2016) Multilevel models improve precision and speed of IC50 estimates. *Pharmacogenomics* **17:** 691–700

Wang T, Yu H, Hughes NW, Liu B, Kendirli A, Klein K, Chen WW, Lander ES & Sabatini DM (2017) Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. *Cell* **168:** 890–903.e15

892 Wei G, Twomey D, Lamb J, Schlis K, Agarwal J, Stam RW, Opferman JT, Sallan SE, den
893 Boer ML, Pieters R, Golub TR & Armstrong SA (2006) Gene expression-based chemical
894 genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance.
895 *Cancer Cell* **10:** 331–342

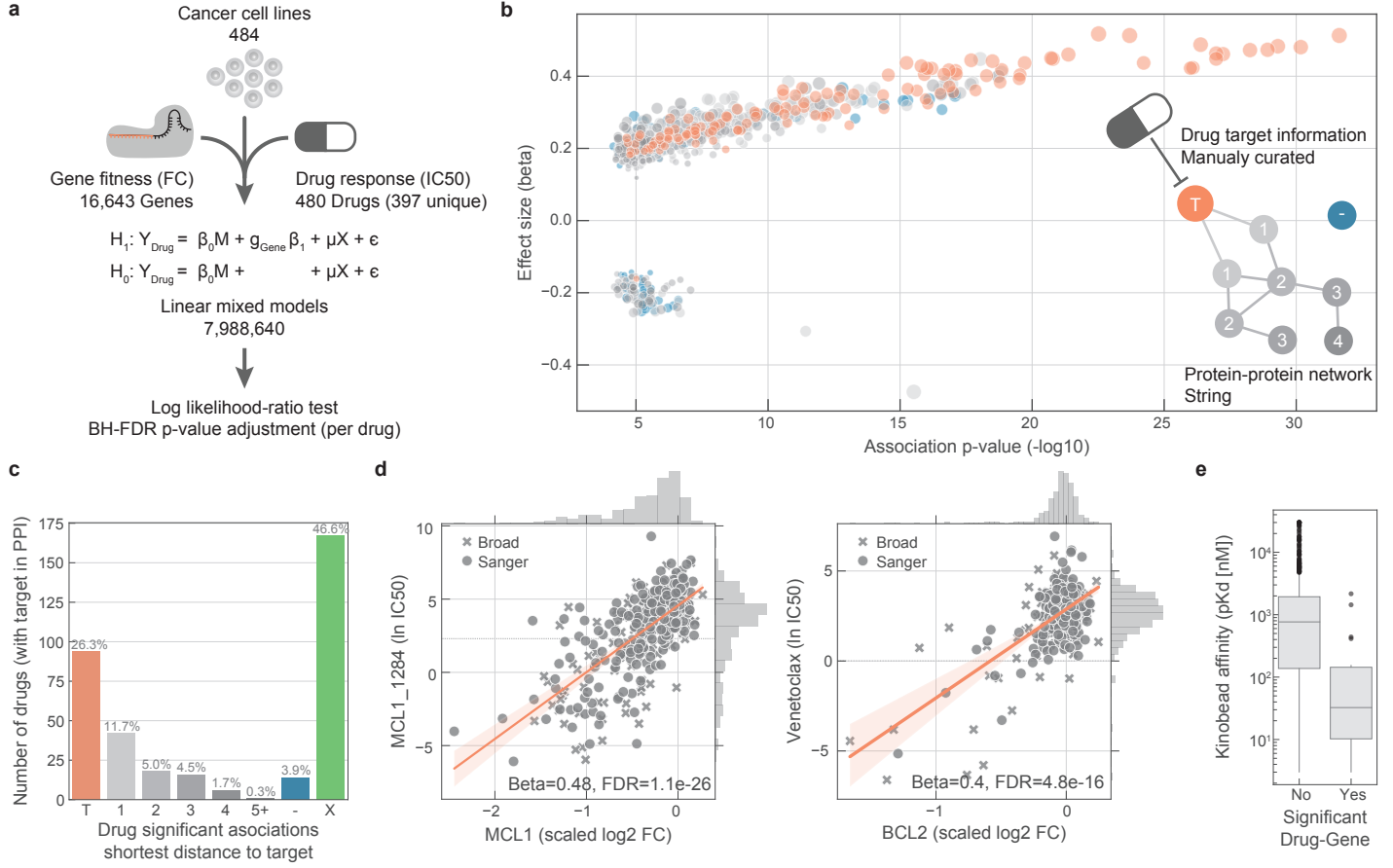
896 Wuillème-Toumi S, Robillard N, Gomez P, Moreau P, Le Gouill S, Avet-Loiseau H,
897 Harousseau J-L, Amiot M & Bataille R (2005) Mcl-1 is overexpressed in multiple
898 myeloma and associated with relapse and shorter survival. *Leukemia* **19:** 1248–1252

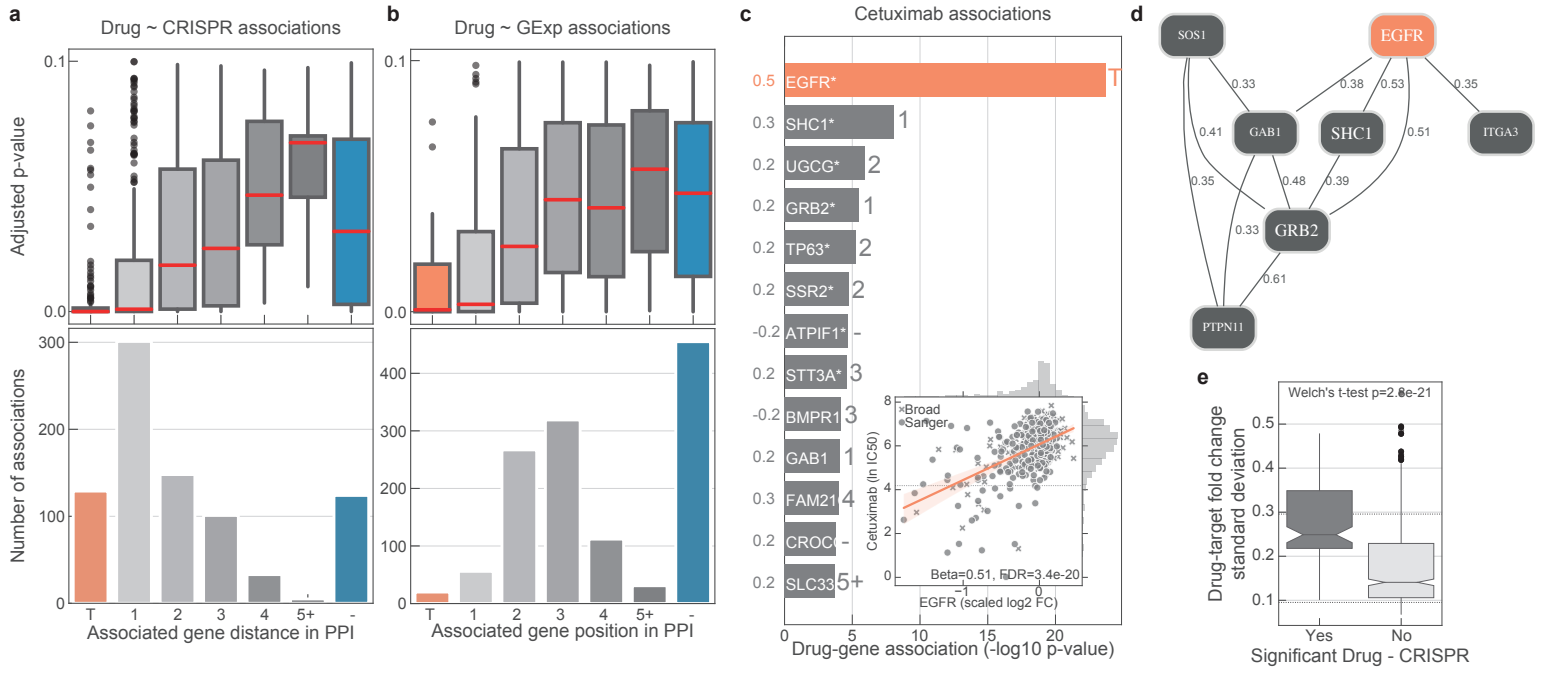
899 Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, Bindal N, Beare D,
900 Smith JA, Thompson IR, Ramaswamy S, Futreal PA, Haber DA, Stratton MR, Benes C,
901 McDermott U & Garnett MJ (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a
902 resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* **41:**
903 D955–61

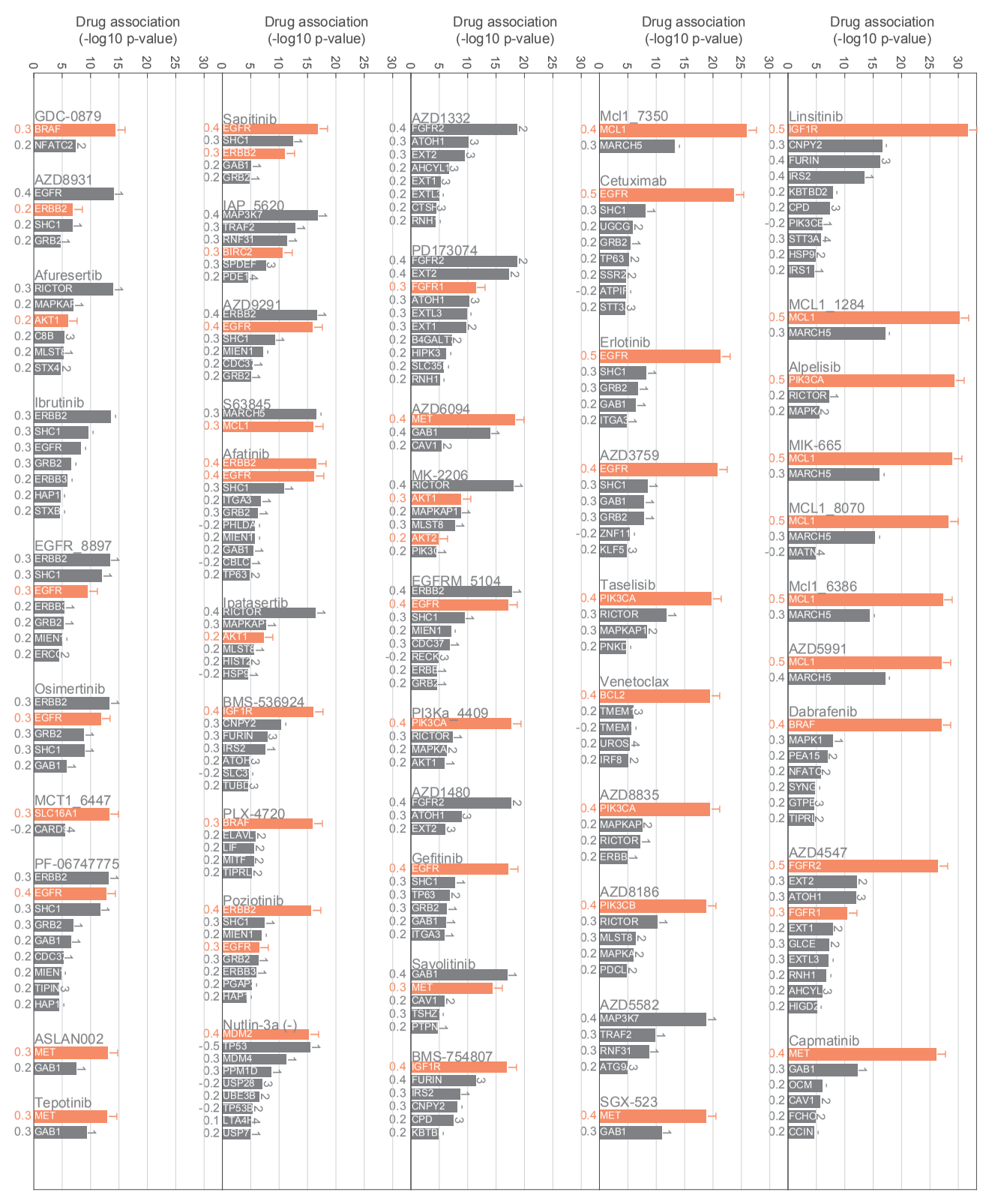
904 Youle RJ & Karbowski M (2005) Mitochondrial fission in apoptosis. *Nat. Rev. Mol. Cell Biol.*
905 **6:** 657

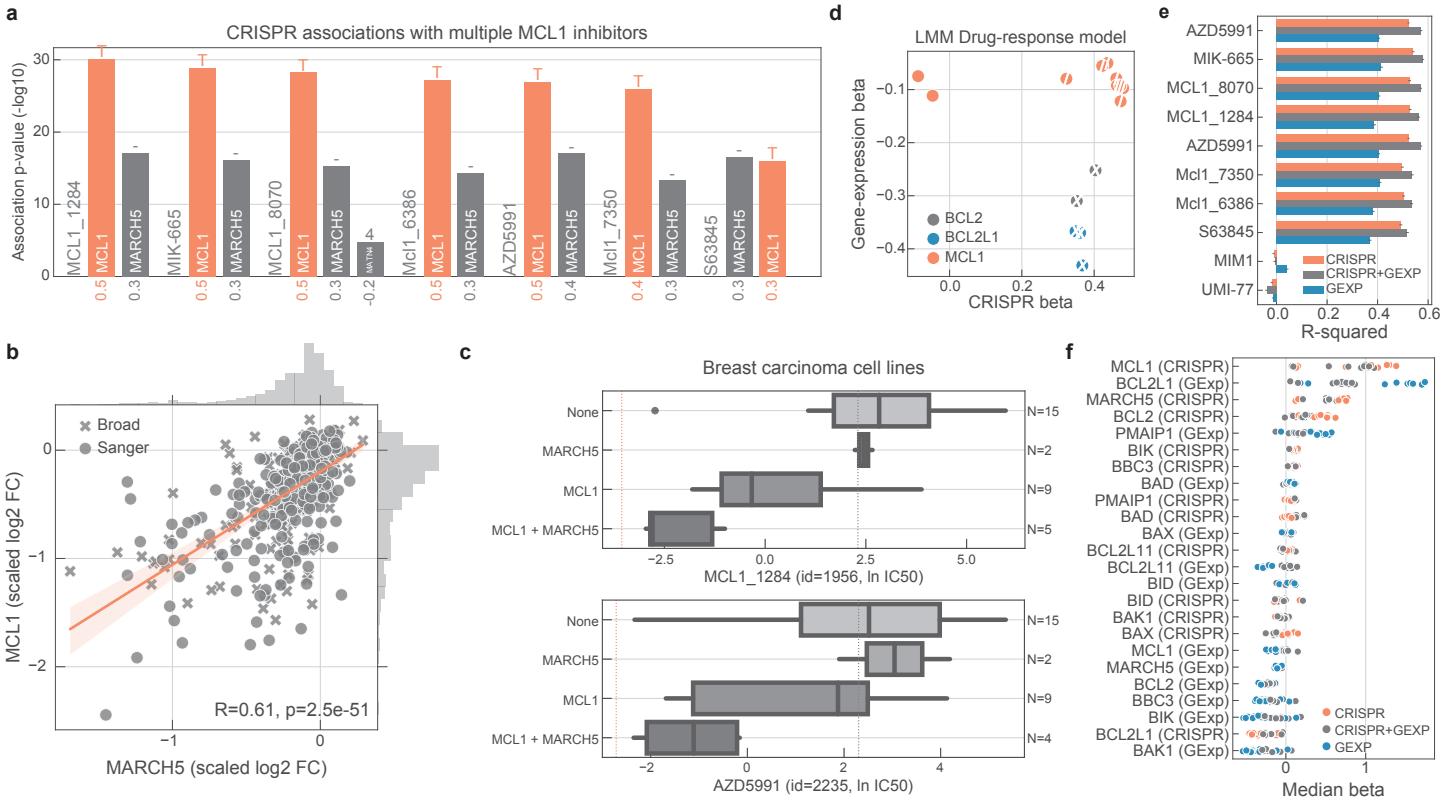
906 Zheng Y, Zhang C, Croucher DR, Soliman MA, St-Denis N, Pasculescu A, Taylor L, Tate
907 SA, Hardy WR, Colwill K, Dai AY, Bagshaw R, Dennis JW, Gingras A-C, Daly RJ &
908 Pawson T (2013) Temporal regulation of EGF signalling networks by the scaffold protein
909 Shc1. *Nature* **499:** 166–171

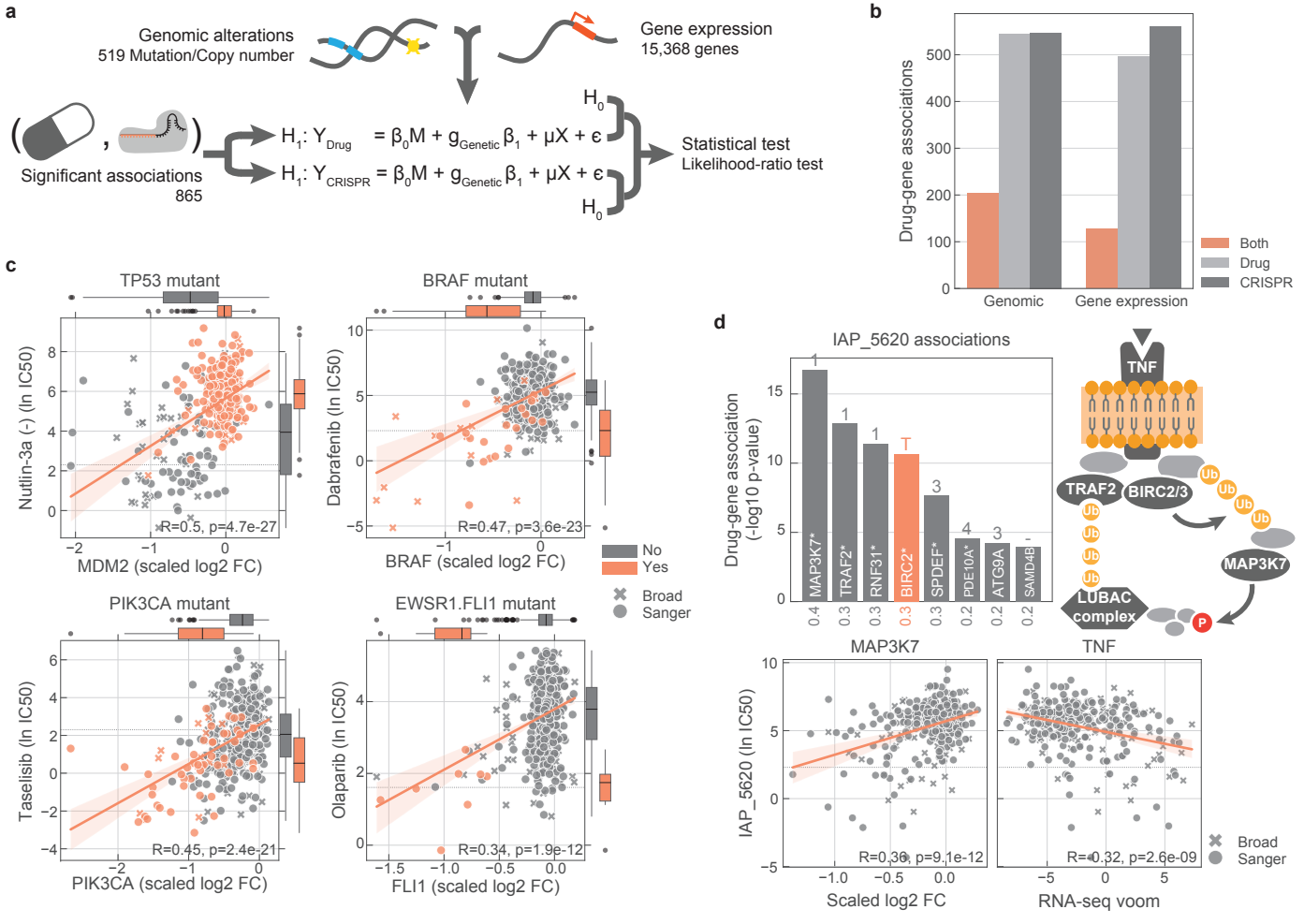
910


a
Cancer cell lines
484
Gene fitness (FC)
16,643 Genes
Drug response (IC50)
480 Drugs (397 unique)
$H_1$: $Y_{Drug} = \beta_0 M + g_{Gene}\beta_1 + \mu X + \epsilon$
$H_0$: $Y_{Drug} = \beta_0 M + \mu X + \epsilon$
Linear mixed models
7,988,640
Log likelihood-ratio test
BH-FDR p-value adjustment (per drug)
b
Effect size (beta)
Association p-value (-log10)
Drug target information
Manualy curated
Protein-protein network
String
c
Number of drugs (with target in PPI)
26.3%
11.7%
5.0%
4.5%
1.7%
0.3%
3.9%
46.6%
T
1
2
3
4
5+
-
X
Drug significant asociations
shortest distance to target
d
Broad
Sanger
MCL1_1284 (ln IC50)
MCL1 (scaled log2 FC)
Beta=0.48, FDR=1.1e-26
Venetoclax (ln IC50)
BCL2 (scaled log2 FC)
Beta=0.4, FDR=4.8e-16
e
Kinobead affinity (pKd [nM])
No
Yes
Significant
Drug-Gene


a
Drug ~ CRISPR associations
b
Drug ~ GExp associations
Adjusted p-value
0.1
0.0
Number of associations
300
200
100
0
T
1
2
3
4
5+
-
Associated gene distance in PPI
Associated gene position in PPI
400
c
Cetuximab associations
EGFR*
SHC1*
UGCG*
GRB2*
TP63*
SSR2*
ATPIF1*
STT3A*
BMPR1
GAB1
FAM21
CROC
SLC33
0.5
0.3
0.2
-0.2
-0.
5+
Broad
Sanger
Cetuximab (ln IC50)
Beta=0.51, FDR=3.4e-20
EGFR (scaled log2 FC)
Drug-gene association (-log10 p-value)
d
SOS1
EGFR
GAB1
SHC1
ITGA3
GRB2
PTPN11
0.33
0.38
0.53
0.35
0.41
0.51
0.35
0.48
0.39
0.33
0.61
e
Welch's t-test p=2.8e-21
Drug-target fold change
standard deviation
Yes
No
Significant Drug - CRISPR


Drug association (-log10 p-value)
GDC-0879
BRAF
NFATC2
AZD8931
EGFR
ERBB2
SHC1
GRB2
Afuresertib
RICTOR
MAPKAP
AKT1
C8B
MLST8
STX4
Ibrutinib
ERBB2
SHC1
EGFR
GRB2
ERBB3
HAP1
STXB
EGFR_8897
ERBB2
SHC1
EGFR
ERBB
GRB2
MIEN
ERC
Osimertinib
ERBB2
EGFR
GRB2
SHC1
GAB1
MCT1_6447
SLC16A1
CARD
PF-06747775
ERBB2
EGFR
SHC1
GRB2
GAB1
CDC3
MIEN
TIPIN
HAP1
ASLAN002
MET
GAB1
Tepotinib
MET
GAB1
Sapitinib
EGFR
SHC1
ERBB2
GAB1
GRB2
IAP_5620
MAP3K7
TRAF2
RNF31
BIRC2
SPDEF
PDE1
AZD9291
ERBB2
EGFR
SHC1
MIEN1
CDC3
GRB2
S63845
MARCH5
MCL1
Afatinib
ERBB2
EGFR
SHC1
ITGA3
GRB2
PHLDA
MIEN1
GAB1
CBLC
TP63
Ipatasertib
RICTOR
MAPKAP
AKT1
MLST8
HIST2
HSP9
BMS-536924
IGF1R
CNPY2
FURIN
IRS2
ATOH
SLC3
TUBD
PLX-4720
BRAF
ELAVL
LIF
MITF
TIPRL
Poziotinib
ERBB2
SHC1
MIEN1
EGFR
GRB2
ERBB3
PGAP
HAP1
Nutlin-3a (-)
MDM2
TP53
MDM4
PPM1D
USP28
UBE3B
TP53B
LTA4H
USP7
AZD1332
FGFR2
ATOH1
EXT2
AHCYL
EXT1
EXTL
CTSH
RNH1
PD173074
FGFR2
EXT2
FGFR1
ATOH1
EXTL3
EXT1
B4GALT
HIPK3
SLC35
RNH1
AZD6094
MET
GAB1
CAV1
MK-2206
RICTOR
AKT1
MAPKAP1
MLST8
AKT2
PIK3
EGFRM_5104
ERBB2
EGFR
SHC1
MIEN1
CDC37
RECK
ERBB
GRB2
PI3Ka_4409
PIK3CA
RICTOR
MAPKA
AKT1
AZD1480
FGFR2
ATOH1
EXT2
Gefitinib
EGFR
SHC1
TP63
GRB2
GAB1
ITGA3
Savolitinib
GAB1
MET
CAV1
TSHZ
PTPN
BMS-754807
IGF1R
FURIN
IRS2
CNPY2
CPD
KBTB
Mcl1_7350
MCL1
MARCH5
Cetuximab
EGFR
SHC1
UGCG
GRB2
TP63
SSR2
ATPIF
STT3
Erlotinib
EGFR
SHC1
GRB2
GAB1
ITGA3
AZD3759
EGFR
SHC1
GAB1
GRB2
ZNF11
KLF5
Taselisib
PIK3CA
RICTOR
MAPKAP1
PNKD
Venetoclax
BCL2
TMEM
TMEM
UROS
IRF8
AZD8835
PIK3CA
MAPKAP
RICTOR
ERBB
AZD8186
PIK3CB
RICTOR
MLST8
MAPKA
PDCL
AZD5582
MAP3K7
TRAF2
RNF31
ATG9
SGX-523
MET
GAB1
Linsitinib
IGF1R
CNPY2
FURIN
IRS2
KBTBD2
CPD
PIK3CB
STT3A
HSP9
IRS1
MCL1_1284
MCL1
MARCH5
Alpelisib
PIK3CA
RICTOR
MAPK
MIK-665
MCL1
MARCH5
MCL1_8070
MCL1
MARCH5
MATN
Mcl1_6386
MCL1
MARCH5
AZD5991
MCL1
MARCH5
Dabrafenib
BRAF
MAPK1
PEA15
NFATC
SYNG
GTPB
TIPRL
AZD4547
FGFR2
EXT2
ATOH1
FGFR1
EXT1
GLCE
EXTL3
RNH1
AHCYL
HIGD2
Capmatinib
MET
GAB1
OCM
CAV1
FCHO
CCIN


a
CRISPR associations with multiple MCL1 inhibitors
Association p-value (-log10)
MCL1_1284
MIK-665
MCL1_8070
Mcl1_6386
AZD5991
Mcl1_7350
S63845
MCL1
MARCH5
MAP1B4
b
Broad
Sanger
R=0.61, p=2.5e-51
MCL1 (scaled log2 FC)
MARCH5 (scaled log2 FC)
c
Breast carcinoma cell lines
None
MARCH5
MCL1
MCL1 + MARCH5
N=15
N=2
N=9
N=5
MCL1_1284 (id=1956, ln IC50)
N=4
AZD5991 (id=2235, ln IC50)
d
LMM Drug-response model
Gene-expression beta
CRISPR beta
BCL2
BCL2L1
MCL1
e
AZD5991
MIK-665
MCL1_8070
MCL1_1284
AZD5991
Mcl1_7350
Mcl1_6386
S63845
MIM1
UMI-77
CRISPR
CRISPR+GEXP
GEXP
R-squared
f
MCL1 (CRISPR)
BCL2L1 (GExp)
MARCH5 (CRISPR)
BCL2 (CRISPR)
PMAIP1 (GExp)
BIK (CRISPR)
BBC3 (CRISPR)
BAD (GExp)
PMAIP1 (CRISPR)
BAD (CRISPR)
BAX (GExp)
BCL2L11 (CRISPR)
BCL2L11 (GExp)
BID (GExp)
BID (CRISPR)
BAK1 (CRISPR)
BAX (CRISPR)
MCL1 (GExp)
MARCH5 (GExp)
BCL2 (GExp)
BBC3 (GExp)
BIK (GExp)
BCL2L1 (CRISPR)
BAK1 (GExp)
CRISPR
CRISPR+GEXP
GEXP
Median beta


a
Genomic alterations
519 Mutation/Copy number
Gene expression
15,368 genes
Significant associations
865
$H_1$: $Y_{Drug} = \beta_0 M + g_{Genetic}\beta_1 + \mu X + \epsilon$
$H_1$: $Y_{CRISPR} = \beta_0 M + g_{Genetic}\beta_1 + \mu X + \epsilon$
$H_0$
$H_0$
Statistical test
Likelihood-ratio test
b
Drug-gene associations
Genomic
Gene expression
Both
Drug
CRISPR
c
TP53 mutant
Nutlin-3a (-) (ln IC50)
MDM2 (scaled log2 FC)
R=0.5, p=4.7e-27
BRAF mutant
Dabrafenib (ln IC50)
BRAF (scaled log2 FC)
R=0.47, p=3.6e-23
No
Yes
Broad
Sanger
PIK3CA mutant
Taselisib (ln IC50)
PIK3CA (scaled log2 FC)
R=0.45, p=2.4e-21
EWSR1.FLI1 mutant
Olaparib (ln IC50)
FLI1 (scaled log2 FC)
R=0.34, p=1.9e-12
d
IAP_5620 associations
Drug-gene association (-log10 p-value)
MAP3K7*
TRAF2*
RNF31*
BIRC2*
SPDEF*
PDE10A*
ATG9A
SAMD4B
TNF
TRAF2
BIRC2/3
Ub
MAP3K7
LUBAC complex
P
MAP3K7
TNF
IAP_5620 (ln IC50)
R=0.39, p=9.1e-12
R=0.32, p=2.6e-09
Scaled log2 FC
RNA-seq voom
Broad
Sanger