Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany for the degree of

Doctor of Natural Sciences

Presented by

Dina Silvia Cramer, M.Sc.

born in: Bergisch Gladbach

Oral examination: May 20, 2019

An interaction-based modeling approach to predict response to cancer drugs

Referees:

Prof. Dr. Roland Eils

Prof. Dr. Benedikt Brors

Summary

In oncology, predictive biomarkers define patient subgroups that are likely to benefit from a specific cancer treatment. Since clinical studies entail high costs and low success rates, pre-clinical model systems like cancer cell lines are needed to generate biomarker hypotheses. Existing computational methods to predict drug response have several limitations. First, models often include large numbers of altered genes which contrasts with clinical predictive biomarkers that mostly include single altered genes. Second, models often assume that the effects of individual alterations are independent, although many biological processes rely on the interplay of multiple molecular components.

We developed an analytical framework to investigate the role of interactions in drug response based on linear regression models. Using data from two large cancer cell line panels, we conducted an exhaustive analysis of models with up to three genomic alterations. To increase model size, we constructed mutation interaction networks and applied module search algorithms to select subsets of mutations for drug response prediction models. We summarized important covariates as background models that served as a reference to evaluate the performance of models with genomic alterations.

We observed that including interactions increased the performance and robustness of drug response prediction models. Moreover, we identified several candidate interactions with consistent association patterns in two large cancer cell line panels. For example, we observed that cancer cell lines with BRAF and TP53 mutations showed worse response to BRAF inhibitors than cell lines with only BRAF mutations. Clinical data supports the resistance interaction between BRAF and TP53 mutations since patients with BRAF and TP53 mutations respond worse to the BRAF inhibitor Vemurafenib than patients with only BRAF mutations. This suggests that inhibition of the oncoprotein BRAF and reactivation of the tumor suppressor protein TP53 could be a promising combination therapy. Our analytical framework moreover allows to distinguish tissue-specific mutation associations from associations that are generalizable across tissues. In addition, we identified synthetic lethal triplets where the simultaneous mutation of two genes sensitizes cells to a drug. Our network-based approach outperformed a standard method for drug response prediction, the

regularized regression algorithm elastic net. Based on 14 million models of different size, seven mutations were determined as the optimal model size.

In summary, we show that considering interactions in drug response prediction models unlocks a large predictive potential. Our interaction-based modeling approach contributes to a system-level understanding of the factors that mediate drug response.

Zusammenfassung

In der Onkologie definieren prädiktive Biomarker Untergruppen von Patienten, die mit hoher Wahrscheinlichkeit von einer bestimmten Krebsbehandlung profitieren. Da klinische Studien mit hohen Kosten und niedrigen Erfolgsraten einhergehen, sind präklinische Modellsysteme wie Krebszelllinien erforderlich, um Biomarker-Hypothesen aufzustellen. Bestehende rechnergestützte Methoden zur Vorhersage der Wirkstoffantwort weisen mehrere Einschränkungen auf. Erstens beinhalten Modelle oft eine große Anzahl von veränderten Genen, was im Gegensatz zu klinischen prädiktiven Biomarkern, die meist einzelne veränderte Gene beinhalten, steht. Zweitens folgen Modelle oft der Annahme, dass die Auswirkungen einzelner Änderungen voneinander unabhängig sind, obwohl viele biologische Prozesse auf das Zusammenspiel mehrerer molekularer Komponenten angewiesen sind.

Wir entwickelten einen Analyserahmen, um die Rolle von Interaktionen in der Wirkstoffantwort auf Grundlage von linearen Regressionsmodellen zu untersuchen. Anhand von Daten aus zwei großen Krebszelllinienbanken führten wir eine umfassende Analyse von Modellen mit bis zu drei genomischen Veränderungen durch. Um die Modellgröße zu erhöhen, konstruierten wir Mutationsinteraktionsnetzwerke und verwendeten Modulsuchalgorithmen, um Mutationsgruppen für die Vorhersage der Wirkstoffantwort auszuwählen. Wir fassten wichtige Kovariablen als Hintergrundmodelle, die als Referenz für die Bewertung von Modellen mit genomischen Veränderungen dienen, zusammen.

Wir beobachteten, dass das Berücksichtigen von Interaktionen die Güte und Robustheit der Modelle zur Vorhersage der Wirkstoffantwort erhöhte. Darüber hinaus identifizierten wir mehrere Kandidateninteraktionen mit konsistenten Assoziationsmustern in zwei großen Krebszelllinienbanken. Zum Beispiel beobachteten wir, dass Krebszelllinien mit BRAF- und TP53-Mutationen schlechter auf BRAF-Inhibitoren ansprechen als Zelllinien mit nur BRAF-Mutationen. Klinische Daten unterstützen die Resistenzinteraktion zwischen BRAF- und TP53-Mutationen, da Melanompatienten mit BRAF- und TP53-Mutationen schlechter auf den BRAF-Inhibitor Vemurafenib ansprechen als Patienten mit nur BRAF-Mutationen. Dies legt nahe, dass die Hemmung des Onkoproteins BRAF und die Reaktivierung des Tumorsuppressorproteins TP53 eine vielversprechende Kombinationstherapie sein

könnte. Unser Analyserahmen erlaubt es zudem, gewebespezifische Mutationsassoziationen von Assoziationen zu unterscheiden, die man über Gewebe hinweg verallgemeinern kann. Darüber hinaus identifizierten wir synthetisch letale Tripletts, bei denen die gleichzeitige Mutation zweier Gene Zellen für ein Medikament sensibilisiert. Unser netzwerkbasierter Ansatz übertraf eine Standardmethode für die Vorhersage von Wirkstoffantworten, den regularisierten Regressionsalgorithmus Elastic Net. Auf Grundlage von 14 Millionen Modellen unterschiedlicher Größe wurden sieben Mutationen als optimale Modellgröße ermittelt.

Zusammenfassend zeigen wir, dass die Berücksichtigung von Interaktionen in Modellen zur Vorhersage der Wirkstoffantwort ein großes prädiktives Potenzial bietet. Unser interaktionsbasierter Modellierungsansatz trägt zu einem Verständnis der Faktoren, die die Wirkstoffantwort vermitteln, auf Systemebene bei.

Table of contents

Introduction

1. Concepts of cancer development

Cancer arises when normal cells acquire alterations that confer a selective growth advantage (Vogelstein et al., 2013). Genomic instability fosters the multistep process of cancer development where successive alterations trigger the outgrowth of cell clones (Hanahan and Weinberg, 2011). Only a subset of alterations, termed cancer driver alterations, promote tumorigenesis. The remaining alterations, termed passenger alterations, accumulate over time as a by-product of defective genome maintenance, but do not confer a selective growth advantage.

Genomic alterations can affect the sequence or the copy number of a genomic region (Vogelstein et al., 2013). Mutations, which alter the DNA sequence, comprise single-base substitutions and insertions or deletions of one or a few DNA bases. Since mutations change the protein product of the affected gene, it is possible to distinguish activating and inactivating mutations. A prominent example of a gene that is frequently targeted by activating mutations is BRAF. Compared to the wild-type protein kinase BRAF, the exchange of the amino acid valine at position 600 by glutamic acid increases the enzymatic activity of the oncogenic kinase 500-fold (Wan et al., 2004).

Inactivating mutations often target tumor suppressor genes that counteract tumorigenesis in healthy cells (Vogelstein et al., 2013). A frequently mutated tumor suppressor is the transcription factor TP53, termed the guardian of the genome (Lane, 1992). While intact TP53 triggers cell death in response to DNA damage, tumor cells with mutated TP53 tolerate the accumulation of DNA damage.

Besides mutations, cancer driver genes can be targeted by copy number alterations (CNAs; Vogelstein et al., 2013). Since amplified or deleted genomic regions often span several genes, identifying the gene that provides the cancer cell with a selective growth advantage can be challenging (Vogelstein et al., 2013). Genomic copy number changes mostly translate into changes in transcript and protein abundances (Tang and Amon, 2013). In contrast to mutations, the amino acid sequence of the wild-type protein remains unchanged.

2. Pharmacogenomics

Despite the growing availability of cancer drugs, their effectiveness is often confined to small patient subgroups which limits the overall success of cancer therapies (Garraway, 2013). Precision medicine approaches account for the large heterogeneity of tumors by selecting a tailored therapy for each individual patient based on their tumor's genetic alterations. Molecular markers that characterize the populations of likely responders and non-responders enable informed therapy decisions at an early stage. To identify these predictive biomarkers, genomic and pharmacological data must be linked. Since clinical trials are time-consuming, cost-intensive and often unsuccessful (Prasad and Mailankody, 2017), extensive preclinical efforts to identify candidate biomarkers are needed.

Genomic and pharmacological screens of cancer cell lines enable the identification of molecular markers that are associated with drug sensitivity or drug resistance. The large number of potential genomic markers, the large heterogeneity of tumors and the low frequency of individual alterations call for cancer cell line panels of extensive size. Each of the pharmacogenomic screening projects Genomics of Drug Sensitivity in Cancer (GDSC; Garnett et al., 2012; Iorio et al., 2016), Cancer Cell Line Encyclopedia (CCLE; Barretina et al., 2012) and Cancer Therapeutic Response Portal (CTRP; Basu et al., 2013; Seashore-Ludlow et al., 2015) contains publicly available data for up to 1001 cancer cell lines and up to 481 drugs. Screened drugs include cytotoxic chemotherapeutics, which mainly affect rapidly dividing cells like tumor cells, and targeted drugs, which inhibit specific molecules that are critical for tumor growth. Drug response is measured by exposing cancer cell lines to different concentrations of a given drug and recording the percent viability. Plotting the viability values against the drug concentration generates dose-response curves (Figure 1).

Despite the discordance of drug response measurements between duplicate screens (Safikhani et al., 2016) and across datasets (Haibe-Kains et al., 2013), concordant candidate biomarkers and known markers can be found (Haverty et al., 2016). The successful prediction of drug response in patients by models based on cell line data (Geeleher et al., 2014) suggests clinical translatability. This implies that pharmacogenomic *in vitro* drug screens represent a valuable model system.

3. Drug response prediction models

A variety of computational methods to predict drug response has been developed. A standard algorithm (Barretina et al., 2012; Basu et al., 2013; Garnett et al., 2012; Iorio et al., 2016) is the regularized linear regression algorithm elastic net that extends least squares regression by two penalty terms on the regression coefficients (Zou and Hastie, 2005). Additional approaches include non-linear methods like support vector machines (Dong et al., 2015), random forests (Nguyen et al., 2016) and neural networks (Menden et al., 2013). Attempts to identify the best-performing algorithm for drug response prediction yielded controversial results. While one study selected regularized linear regression as the winning algorithm (Jang et al., 2014), another study reported that non-linear models perform best (Costello et al., 2014). A third study concluded that no algorithm consistently outperformed the other, suggesting that the algorithm choice plays a minor role (Bayer et al., 2013).

An advantage of linear models is their good interpretability (Azuaje, 2017). Linear models represent the association between input features like cancer alterations and an output feature like drug response in compact formulas. In the simplest case, alterations are encoded by binary variables that represent the presence or absence of alterations. Regression coefficients can then be interpreted as the strength of the association between each alteration and drug response. The predicted drug response value for a given sample equals the sum of the regression coefficients that correspond to the alterations it harbors.

Several studies on drug response prediction concluded that expression data is the most predictive data type, while models based on genomic data perform poorly (Costello et al., 2014; Iorio et al., 2016; Jang et al., 2014). However, genomic biomarkers are better suited to translate to clinical biomarkers (Nass and Moses, 2007; Paziewska et al., 2014). For instance, DNA is more stable than RNA, which facilitates its utilization by liquid biopsies, a non-invasive technique for disease diagnostics and monitoring (Crowley et al., 2013). Since existing drug response prediction models mostly focus on expression data (Azuaje, 2017), the availability of models utilizing genomic data is limited.

In the clinical setting, patients receiving a given treatment are often selected based on single-gene predictive biomarkers (Goossens et al., 2015). For example, patients that receive a BRAF inhibitor must harbor the $BRAF^{V600E}$ mutation. However, a one-to-one mapping between the drug target and the predictive biomarker does not always suffice. Instead of single-gene models, computational algorithms often output large multi-gene models for the prediction of drug response. On the one hand, the predictive performance of multi-gene models is better than the performance of single-gene models (Knijnenburg et al., 2016; Nguyen et al., 2016). On the other hand, model complexity limits biological interpretability and validation in independent datasets. To bridge the gap between model sizes in clinical practice and computational methods, novel approaches are needed.

Models that predict drug response based on multiple altered genes usually assume that the effects of individual alterations are independent (Jiang et al., 2018). However, many cellular processes involve the joint action of multiple molecular components, which implies that the effects of individual alterations depend on each other. To quantify interdependencies, models that predict drug response based on multiple alterations can discriminate the joint effect from the individual alteration effects. For two alterations, the joint effect is additive if the combined phenotype in cells with both alterations corresponds to the sum of the phenotypes in cells with either of the alterations. Deviations from additivity reflect interdependencies between alterations and are defined as genetic interactions (Bateson et al., 1905).

In model organisms like yeast, pairwise genetic interactions were extensively studied by introducing two mutations at once and comparing the resulting phenotype with the phenotype in presence of either mutation (Jasnos and Korona, 2007; Tong et al., 2001, 2004). These studies showed that interactions are common. More recently, methods that consider interactions were proposed in the pharmacogenomic field. Proposed drug response prediction algorithms rely on logic combinations of up to four genomic alterations (Knijnenburg et al., 2016), pairs of alterations where one mediates drug sensitivity while the other suppresses this effect (Liu et al., 2016) and a genome-wide score that is calculated based on interactions between alteration pairs (Jiang et al., 2018). However, these methods are restricted to logic models (Knijnenburg et al., 2016; Liu et al., 2016) or pairwise interactions between genomic and transcriptomic

alterations (Jiang et al., 2018; Liu et al., 2016). Systematic and quantitative studies of interactions between two genomic alterations in drug response are rare.

Analogous to genetic interactions, the association between an alteration and drug response can depend on the tissue of origin. To date, it remains controversial whether drug response in cancer cell lines is tissue-specific or not (Garnett et al., 2012; Iorio et al., 2016; Jaeger et al., 2015). Deciding whether drug response should be predicted by tissue-specific or pan-cancer models is regarded as a critical challenge in the development of computational drug response prediction methods (Azuaje, 2017).

4. Aim of this thesis

The main objective of this thesis is to evaluate the role of genetic interactions and tissue specificity in drug response prediction models. To assess whether the poor predictive potential of genomic data (Iorio et al., 2016) can be improved by incorporating interactions, we developed an analytical framework (Figure 1) and a network method based on linear regression models. We used the large-scale drug screening dataset GDSC (Iorio et al., 2016) for identification of candidate interactions and the CTRP dataset (Seashore-Ludlow et al., 2015) for validation.

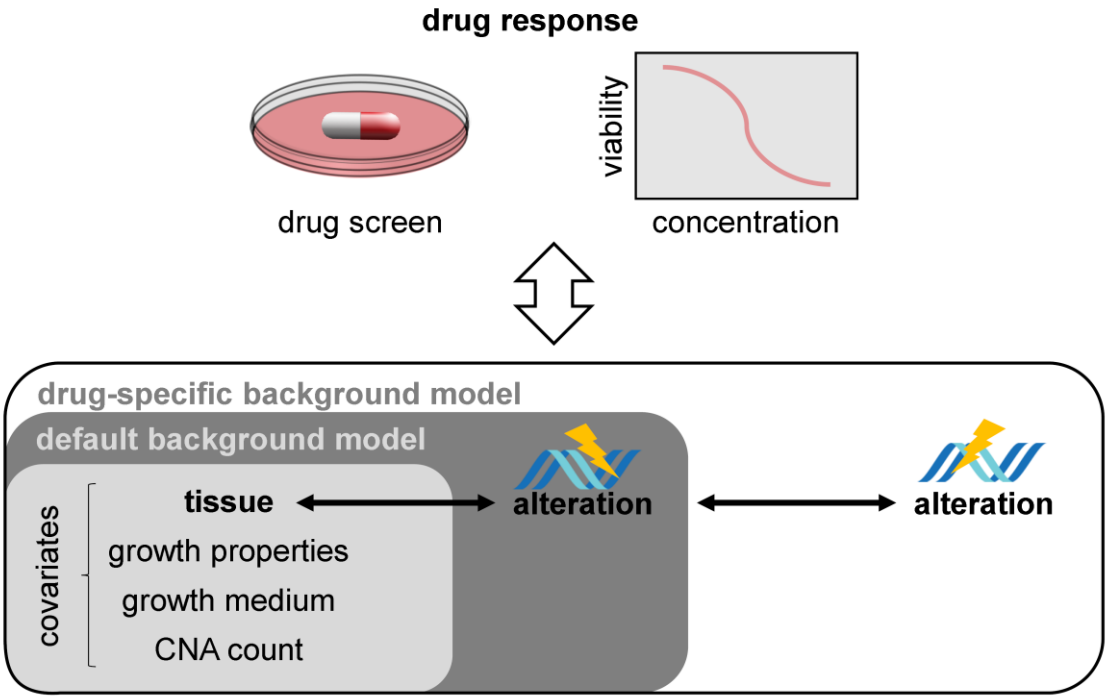

Figure 1: Schematic overview of the analytical framework. Building blocks for drug response prediction models comprise covariates, alterations, alteration-tissue interactions and alteration-alteration interactions. Filled boxes represent background models, horizontal arrows represent interactions and the vertical arrow represents the association with drug response.

In Chapter 1, we fitted drug response prediction models including up to three mutations or CNAs. We evaluated and compared models based on different performance metrics while focusing on the relevance of genetic interactions.

In Chapter 2, we tested whether a mutation has different effects depending on the tissue in which it occurs. We used mutation-tissue interactions to investigate which associations of mutations with drug response are tissue-specific.

In Chapter 3, we combined the findings of Chapter 1 and 2. In models that account for tissue-specific mutation associations, we searched for examples of mutation pairs that interact with respect to drug response. Furthermore, we identified synthetic lethal relationships where the simultaneous mutation of two genes confers drug sensitivity.

In Chapter 4, we represented interactions between mutations with respect to drug response in a network of mutations. We applied module search algorithms to identify

sets of mutations that strongly interact and used these sets of mutations to predict drug response.

Results

1. Genetic interactions

We aimed at analyzing the association of genomic alterations with drug response in small, easily interpretable models. However, factors other than alterations can affect drug response. To separate these effects from alteration effects, we summarized these factors as covariates (Figure 1).

1.1. Covariate selection

To select covariates, we assessed experimental conditions and global cell line properties as potential confounders in the GDSC dataset (Iorio et al., 2016). We analyzed tissue of origin, growth properties, growth medium, microsatellite instability status, total number of mutations and total number of CNAs. We additionally tested 30 mutational signatures as reported by Jarvis and colleagues (Jarvis et al., 2018). Mutational signatures are mutation patterns that can be attributed to specific biological processes (Alexandrov et al., 2013).

To identify a consensus model that includes the most predictive confounders across drugs, we assessed the association between potential confounders and drug response in three model settings. First, a univariate model with a single confounder as the predictive variable was fitted. Second, to estimate the effect of a confounder in a multivariate model with all other confounders, we compared models including all confounders to models that excluded exactly one confounder at a time. Third, we preselected the most predictive confounders and tested if including any additional confounder significantly improved model performance. Confounders showing high predictive performance in all three model settings were selected as covariates.

In the first model setting, we fitted univariate models relating individual confounders to drug response. To estimate the variation in drug response that a model explains, we computed the adjusted coefficient of determination (adj. R^2). Figure 2 shows the adj. R^2 for all tested confounders and all drugs. To summarize the predictive performance

of a given confounder, we used the median adj. R^2 across drugs. Confounders with a positive median adj. R^2 were defined as predictive of drug response. With the exception of mutation count, all confounders other than mutational signatures were predictive of drug response. In addition, eleven mutational signatures showed a predictive potential. In total 16 out of 36 potential confounders were classified as predictive, which implies that many factors other than alterations affect drug response.

When determining the number of drugs for which a given confounder showed a significant association with drug response ($p < 0.05$, F-test), tissue (97% of the drugs), growth properties (73%), growth medium (53%) and CNA count (51%) were associated with response for more than 50% of the drugs. These four confounders also explained the highest proportion of variation in drug response (Figure 2). The tissue of origin was the most predictive confounder (maximal adj. $R^2 = 0.44$), which is in line with previous research (Iorio et al., 2016).

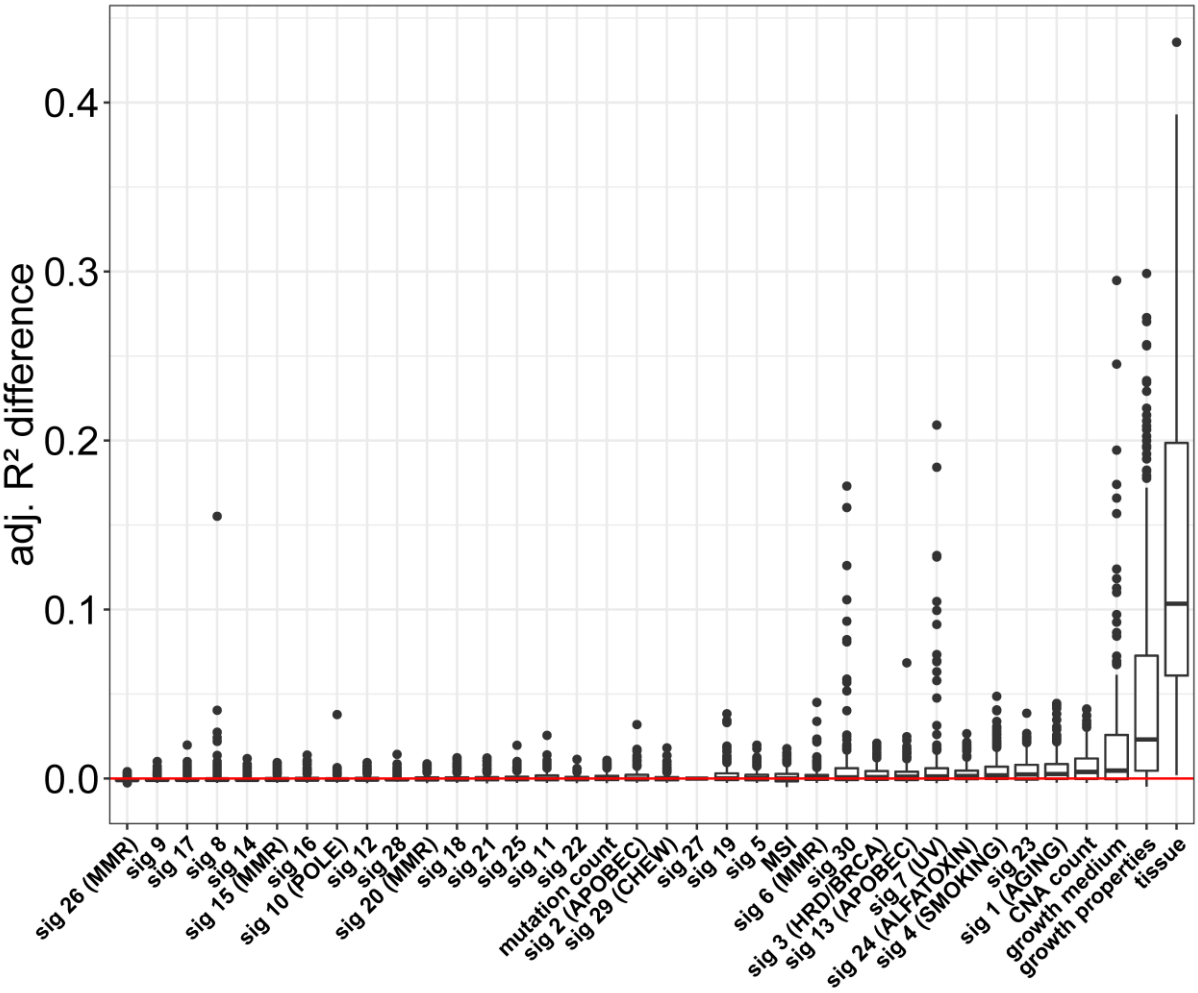

Figure 2: Explained variation in drug response by single confounders. The adj R^2 for univariate models relating single confounders to drug response is shown. For each confounder, models were fitted for all 265 drugs. Confounders are sorted by median adj. R^2; values for signature 27 concur with the zero line (in red).

In the second model setting, we assessed the predictive potential of individual confounders in presence of all other confounders. To this end, we set up multivariate models including all 36 confounders, which we termed full models. Figure 3 shows the performance differences between the full model and the model excluding one confounder based on the adj. R^2. Confounders with a positive median adj. R^2 difference across drugs were defined as predictive in the full model. CNA count, growth properties, growth medium, and tissue were the only predictive confounders, which confirms their outstanding predictive potential in univariate models (Figure 2).

When computing the percentage of drugs for which the full model performed significantly better than a model with all except one confounder ($p < 0.05$, F-test), the most essential confounders across drugs were tissue (87% of the drugs), growth medium (38%), growth properties (25%), and CNA count (21%). Together with our findings based on univariate models, this suggests that these four confounders play an important role in drug response prediction models. Therefore, we preselected tissue, growth medium, growth properties and CNA count as covariates.

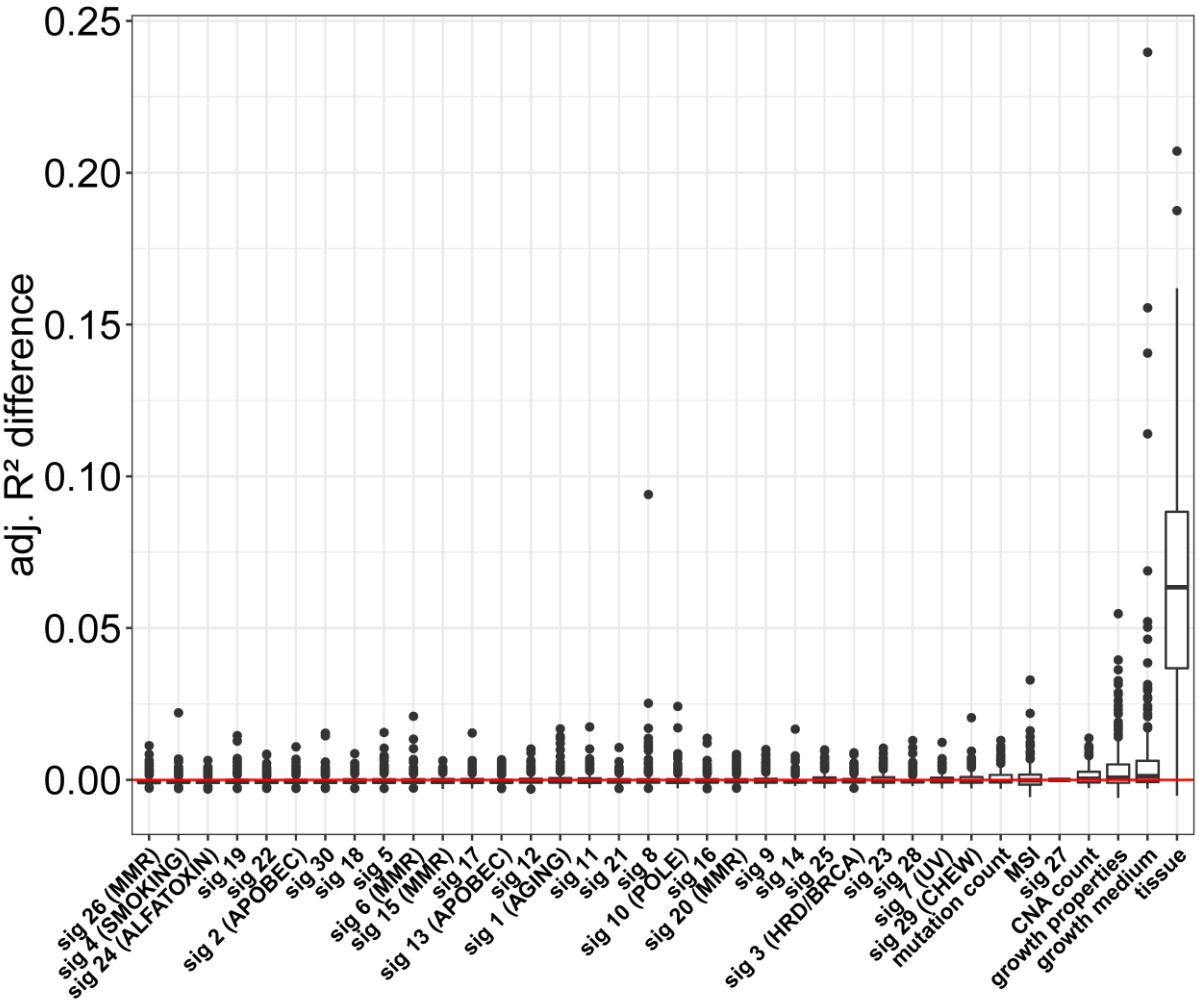

Figure 3: Variation in drug response explained by all but one confounder. In each model, all but the indicated confounder were included as predictors. The effect of excluding one confounder is estimated by comparing this model to a model including all predictor confounders. Each value represents the performance difference for one drug. Confounders are sorted by median adj. R^2 difference; values for signature 27 concur with the zero line (in red).

Third, to confirm the preselection of covariates, we used a model including tissue, growth medium, growth properties, and CNA count and tested whether the inclusion of any additional confounder significantly improved the model ($p < 0.05$, F-test). For all confounders, an improvement was achieved for less than 20% of the drugs, which supports the preselection of covariates.

To analyze the association of these four covariates with drug response in the CTRP dataset (Seashore-Ludlow et al., 2015), we used the set of overlapping drugs between the GDSC and the CTRP dataset. In univariate models, tissue showed significant associations with drug response for 91%, growth medium for 36% and CNA count for 18% of the drugs ($p < 0.05$, F-test). We did not analyze the growth properties covariate since all cell lines in the CTRP dataset are adherent. Although the number of drugs showing significant associations with drug response in the CTRP dataset was lower than in the GDSC dataset, we maintained the selection of covariates.

Based on the analysis of potential confounders, we defined a model with the four covariates tissue, growth medium, growth properties and CNA count as the default background model. Subsequent models that include genomic alterations contain the covariates of the default background model. By comparing models that include genomic alterations to the default background model, alteration effects can be separated from covariate-specific effects.

Having assessed the association of covariates with drug response, we estimated the predictive performance of genomic data. To this end, we fitted univariate models relating mutations and CNAs that are likely to promote tumorigenesis (see Methods) to drug response. For most drugs, the most predictive genomic variable explained less than 10% of the variation in drug response (Figure 4). This confirms previous reports about the poor predictive potential of genomic data (Iorio et al., 2016) and supports the need for using a background model based on covariates.

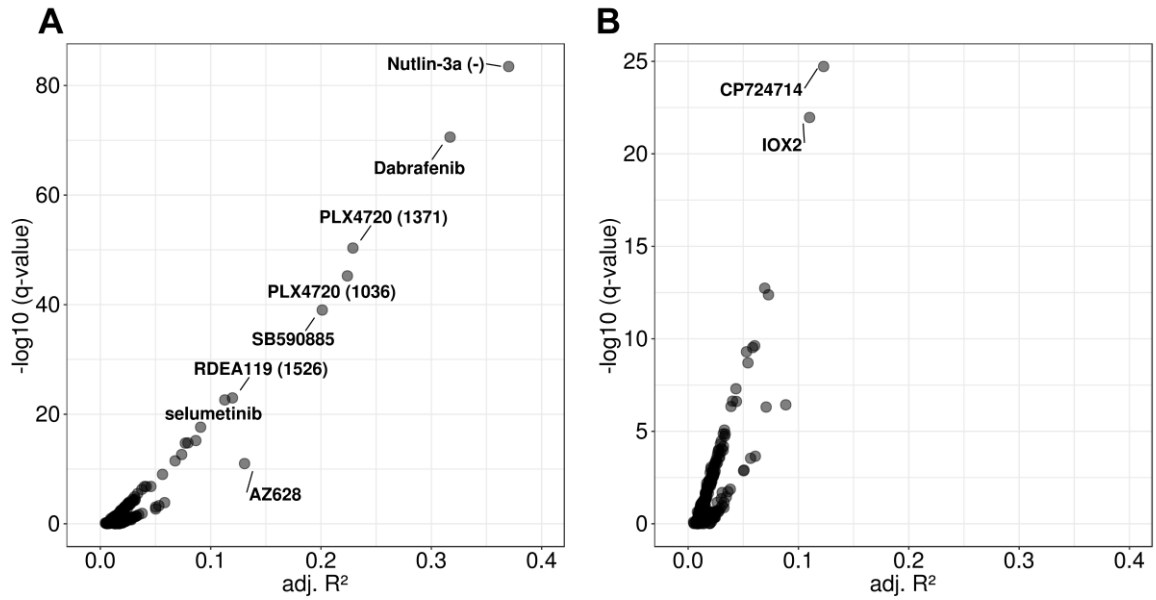

Figure 4: Performance of univariate A) mutation and B) CNA models. For each drug, the maximal adj. R^2 is shown together with the model significance (adjusted p-value (Benjamini-Hochberg correction); F-test). Drugs for which the best model achieves an adj. $R^2 > 0.1$ are specified. Drug IDs are indicated in brackets where applicable.

Table 1: Best univariate mutation and CNA models (cf. Figure 4).

Drug name (ID)	Drug target	Altered gene/ region	Alteration type
AZ628	BRAF	BRAF	mutation
PLX4720 (1036)	BRAF	BRAF	mutation
Nutlin-3a (-)	MDM2	TP53	mutation
SB590885	BRAF	BRAF	mutation
PLX4720 (1371)	BRAF	BRAF	mutation
Dabrafenib	BRAF	BRAF	mutation
selumetinib	MEK1, MEK2	BRAF	mutation
RDEA119 (1526)	MEK1, MEK2	BRAF	mutation
CP724714	ERBB2	7q12 (includes ERBB2)	amplification
IOX2	EGLN1	8q24.23	deletion

1.2. Alteration distribution

Before analyzing the association of genomic alterations with drug response, we studied the distribution of alterations across cell lines. To this end, we used a list of mutations and CNAs that are likely to represent functional events in cancer development (see Methods). We assessed the frequency of single genomic alterations and the co-occurrence of two or three genomic alterations in cancer cell lines. For alteration pairs, we additionally assessed whether they tend to co-occur or to be mutually exclusive.

1.2.1. Alteration frequency

Since the frequency of alterations limits the detection of statistical relationships with drug response, we determined the number of cell lines harboring an alteration or a combination of alterations. Figure 5 shows the average number of cell lines with a specific alteration or combination of alterations across drug screens. As expected, single alterations were more frequent than alteration pairs and alteration pairs were more frequent than alteration triplets.

For a given number of alterations, single mutations were more frequent than single CNAs ($p = 0.008$, Mann-Whitney-Wilcoxon test), whereas mutation pairs were less frequent than CNA pairs ($p < 10^{-15}$, Mann-Whitney-Wilcoxon test). Pairs of one mutation and one CNA occurred more frequently than pure mutation or CNA pairs ($p < 10^{-15}$ for both comparisons, Mann-Whitney-Wilcoxon test). This suggests that the frequency of an alteration or a combination of alterations depends on the alteration type.

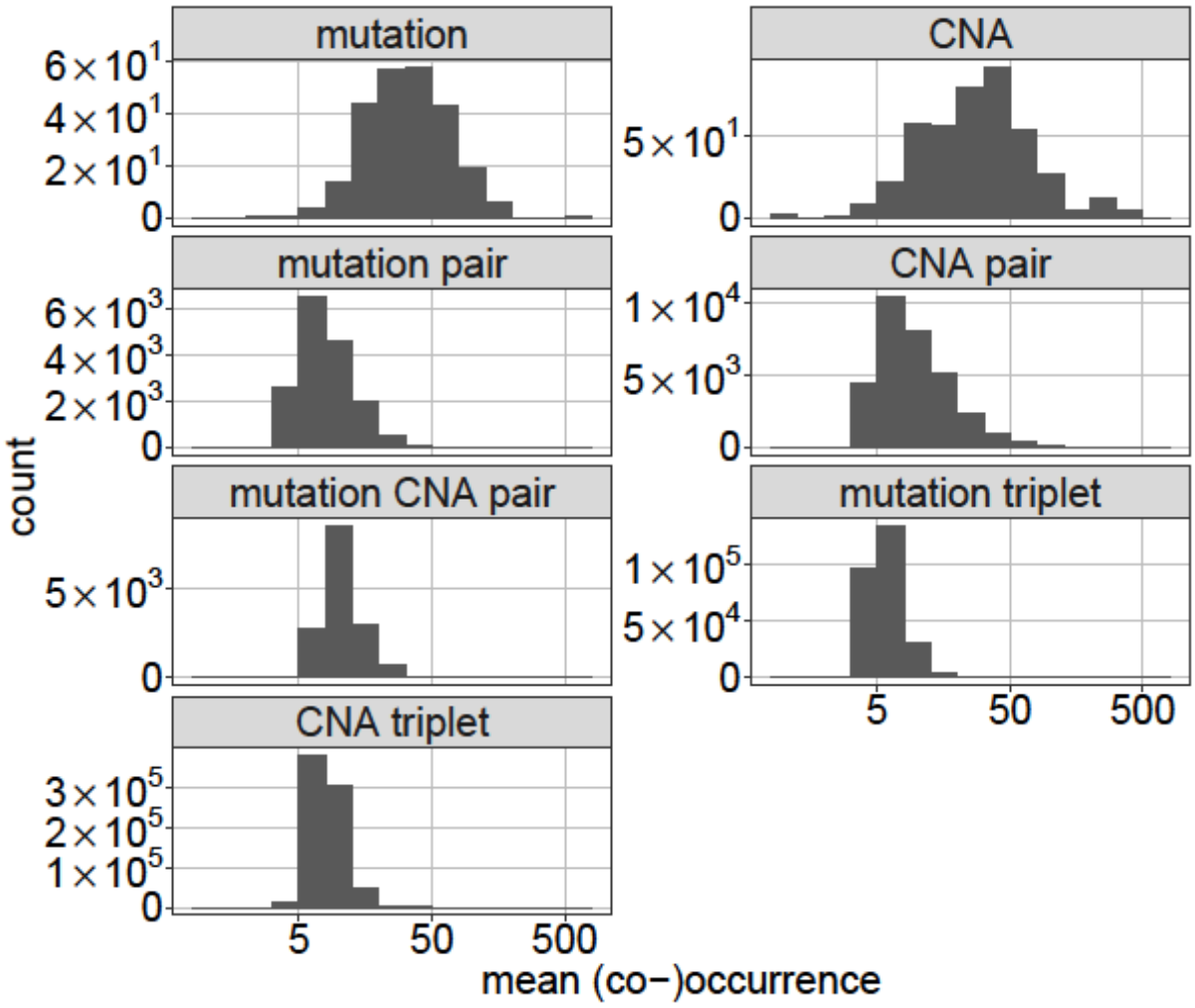

Figure 5: Alteration frequency. The number of cell lines with available drug response data that harbors a specific alteration or combination of alterations is averaged across drug screens.

1.2.2. Correlation

To test whether a given alteration is more frequent or less frequent in presence of a second alteration, we computed the absolute correlation (phi coefficient) across cancer cell lines for alteration pairs. The maximal correlation for mutation pairs and mutation-CNA pairs was 0.34 and 0.21, respectively. This suggests a weak to moderate relationship between the occurrence of two mutations or one mutation and a CNA. In comparison, we observed a stronger relationship between the occurrence of two CNAs. In total, 1429 CNA pairs showed a correlation above 0.4 which could explain why CNA pairs are more frequent than mutation pairs, although single CNAs

are less frequent than single mutations. We identified 20 CNA pairs with a perfect correlation of 1. As a result, models including different CNAs can have identical performance.

1.3. Systematic analysis of model complexities

To assess drug response prediction models with different genomic alterations and different model sizes, we tested mutations and CNAs in models with one, two or three alterations with and without interaction. For each drug, models for 248 single mutations, 425 single CNAs, 16766 mutation pairs, 32608 CNA pairs, 14855 mutation-CNA pairs, 275656 mutation triplets, and 779502 CNA triplets were fitted using the GDSC dataset. Due to missing drug response values, the final number of models varied across drugs and can be substantially smaller for individual drugs. All models contained the covariates of the default background model in addition to the alterations.

1.3.1. Role of interactions

To illustrate the relevance of interactions, we investigated mutation pair models with interaction. Interactions can be neglected if the joint effect of two mutations on drug response is additive. In this case, the joint effect can be correctly estimated by adding up the individual effects of both mutations. Interactions should be considered if the joint effect deviates from additivity, meaning that the estimated effect for cell lines with both mutations is smaller or larger than the sum of the individual effects. We defined joint effects that were smaller than additive as antagonistic and joint effects that were larger than additive as synergistic.

Applying this rationale to all mutation pair models we fitted, we classified mutation-mutation interactions as antagonistic or synergistic. To evaluate the relevance of non-additivity, we compared the frequency and strength of antagonistic and synergistic interactions for each drug. When assessing the proportion of antagonistic and synergistic interactions per drug, synergistic interactions were more

frequent than antagonistic interactions for 88% of the drugs and less frequent than antagonistic interactions for 12% of the drugs. Compared to antagonistic interactions, synergistic interactions showed stronger associations with drug response for 81% of the drugs, weaker associations for 9% of the drugs ($p < 0.05$, Mann-Whitney-Wilcoxon test) and no significant differences for the remaining drugs ($p \geq 0.05$, Mann-Whitney-Wilcoxon test). Since synergistic interactions are more frequent and stronger than antagonistic interactions, models that neglect interactions tend to underestimate the joint effect of a mutation pair.

1.3.2. Best model per complexity

To determine the maximal model performance that can be achieved for a given model complexity, we assessed the performance of models that were fitted to the entire GDSC dataset by two metrics, the adj. R^2 and the Bayesian information criterion (BIC). We computed these metrics for all fitted models. We then selected the best model per drug and model complexity, resulting in twelve models per drug and 265 models per complexity (Figure 6). The overlap between the best models selected by adj. R^2 and BIC was 98%, which implies a high level of concordance.

When evaluating the performance of the best models per model complexity, we found that models with up to three alterations explained up to 60% of the variation in drug response (Figure 6). In general, model performance increased with increasing model complexity. However, alteration pair models with interaction tended to outperform alteration triplet models without interaction ($p = 0.1$ for mutations and CNAs, Mann-Whitney-Wilcoxon test). This implies that adding an interaction term instead of an additional alteration to an alteration pair model can result in better model performance.

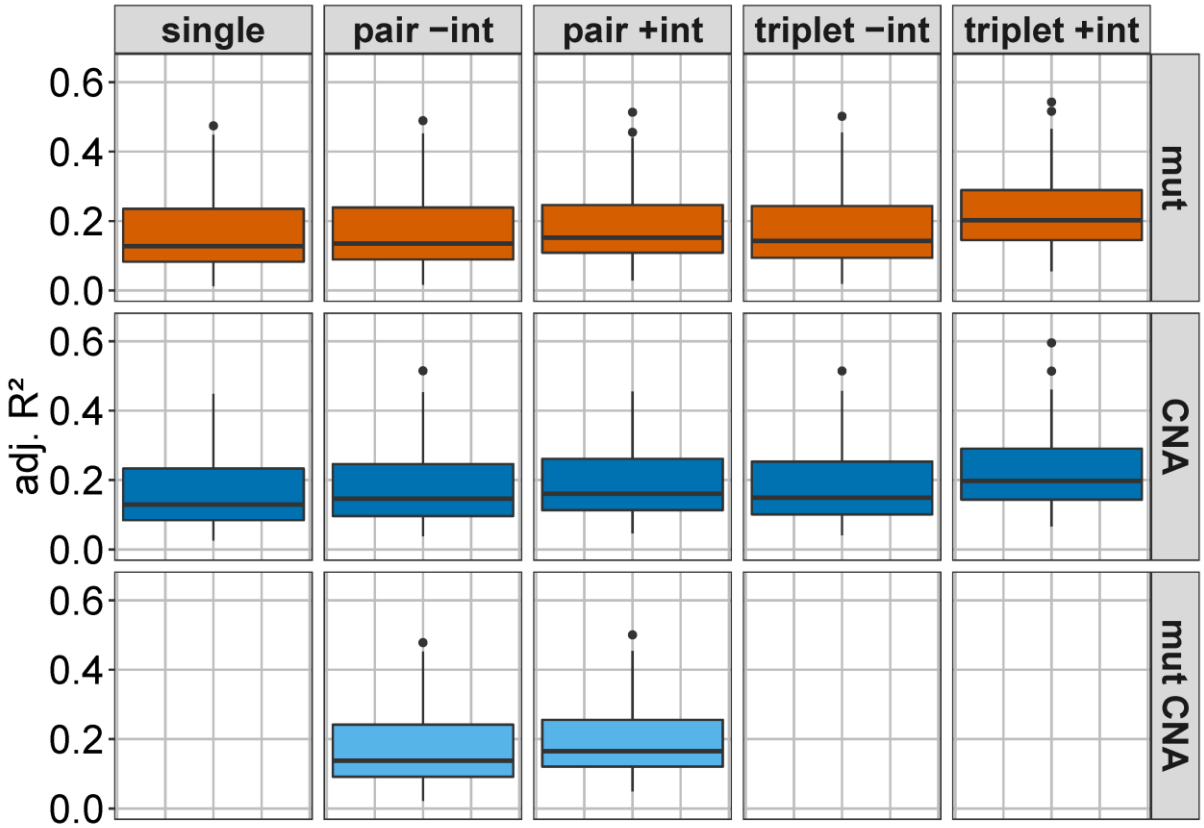

Figure 6: Best model per complexity. Each panel shows the model with the highest adj. R^2 per drug for the respective model complexity.

1.3.2.1. Nested models

To estimate the contribution of interactions to model performance, we tested whether the best models per complexity were nested, meaning that more complex models include the variables of less complex models. For alteration pair and triplet models, we assessed whether the best model with interaction included the variables of the best model without interaction. If models with interaction include less complex, purely additive models, this implies that additive effects predominate non-additive effects.

Figure 7 shows the percentage of best models with interaction that are nested. Across all model complexities, 9% of the best models with interaction were nested. The low percentage of nested models suggests that for most drugs, additive contributions are not predominant. As a result, the variables of the best alteration pair (or triplet) model

with interaction cannot be foreseen by knowing the variables of the best alteration pair (or triplet) model without interaction.

For individual model complexities, the percentage of nested models varied between 3% for CNA triplet models with interaction and 15% for CNA pair models with interaction (Figure 7). Compared to alteration triplet models, alteration pair models showed higher percentages of nested models. This is expected since alteration triplet models include more than one interaction term (see Methods) which entails greater differences in model complexity and potentially model performance. For mutation-CNA pair models, the percentage of nested models was lower than for mutation pair models and CNA pair models. This suggests that for most drugs, including interactions in mutation-CNA pair models unlocks a high predictive potential.

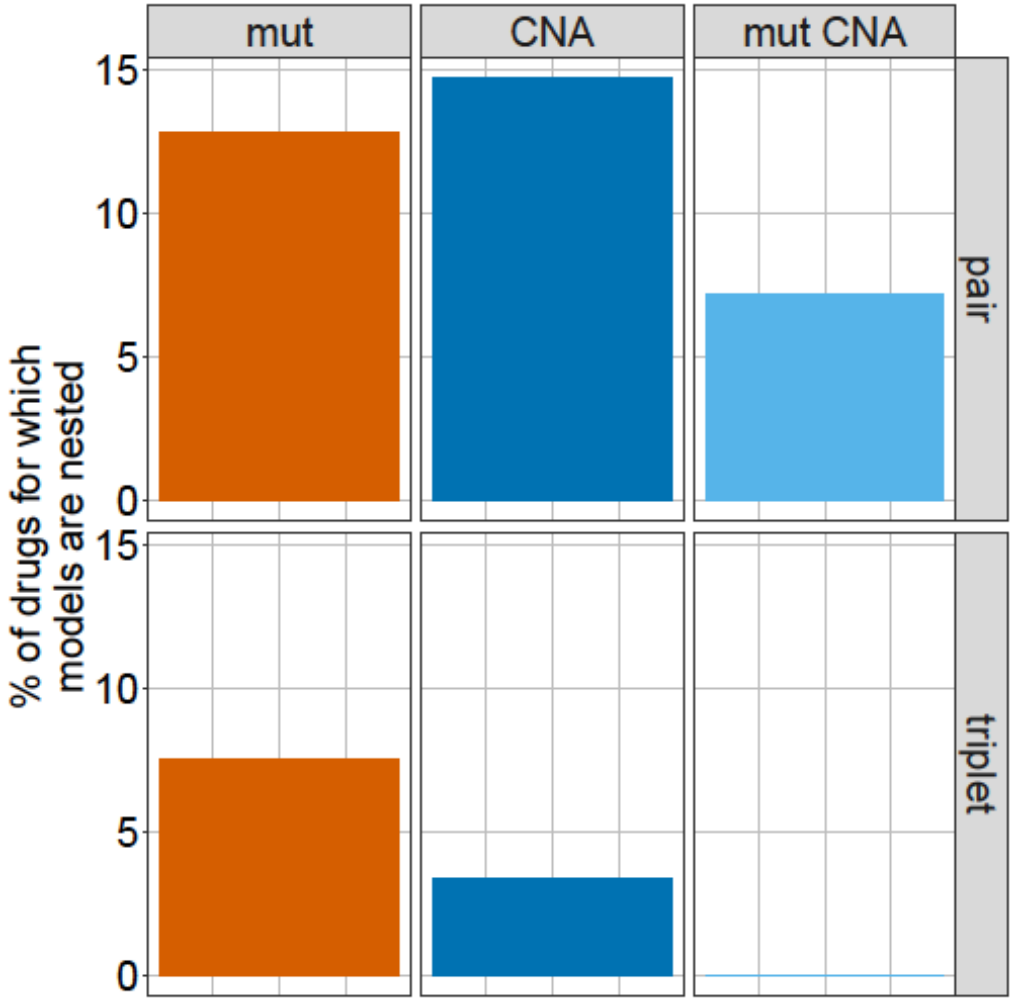

Figure 7: Proportion of nested models. The percentage of drugs for which the best model with interaction contains the variables of the best model without interaction is shown. Models are grouped by alteration type (horizontal axis) and model size (vertical axis). Mut: mutation.

1.3.2.2. Model performance with respect to the default background model

To further assess the performance of best models per complexity, we used the default background model as a lower performance boundary. We evaluated model performance based on model comparison tests and based on the BIC. Models outperforming the default background model were defined as useful models.

First, we used model comparison tests to assess for how many drugs the best model selected by adj. R^2 performed better than the default background model. Figure 8

shows the percentage of drugs for which the best model per complexity outperformed the default background model ($q < 0.1$, Benjamini-Hochberg correction per drug and complexity, F-test).

Of all best models, 40% outperformed the default background model. For individual model complexities, the percentage of drugs for which the default background model was outperformed varied between 22% for mutation triplet models without interaction and 67% for mutation triplet models with interaction. In general, the number of drugs for which the background model was outperformed was substantially greater for models with interaction compared to models without interaction. This suggests that including interactions improves model performance. For 85% of the drugs, the default background model was outperformed for at least one model complexity, indicating that the model sizes we studied are sufficiently complex for most drugs.

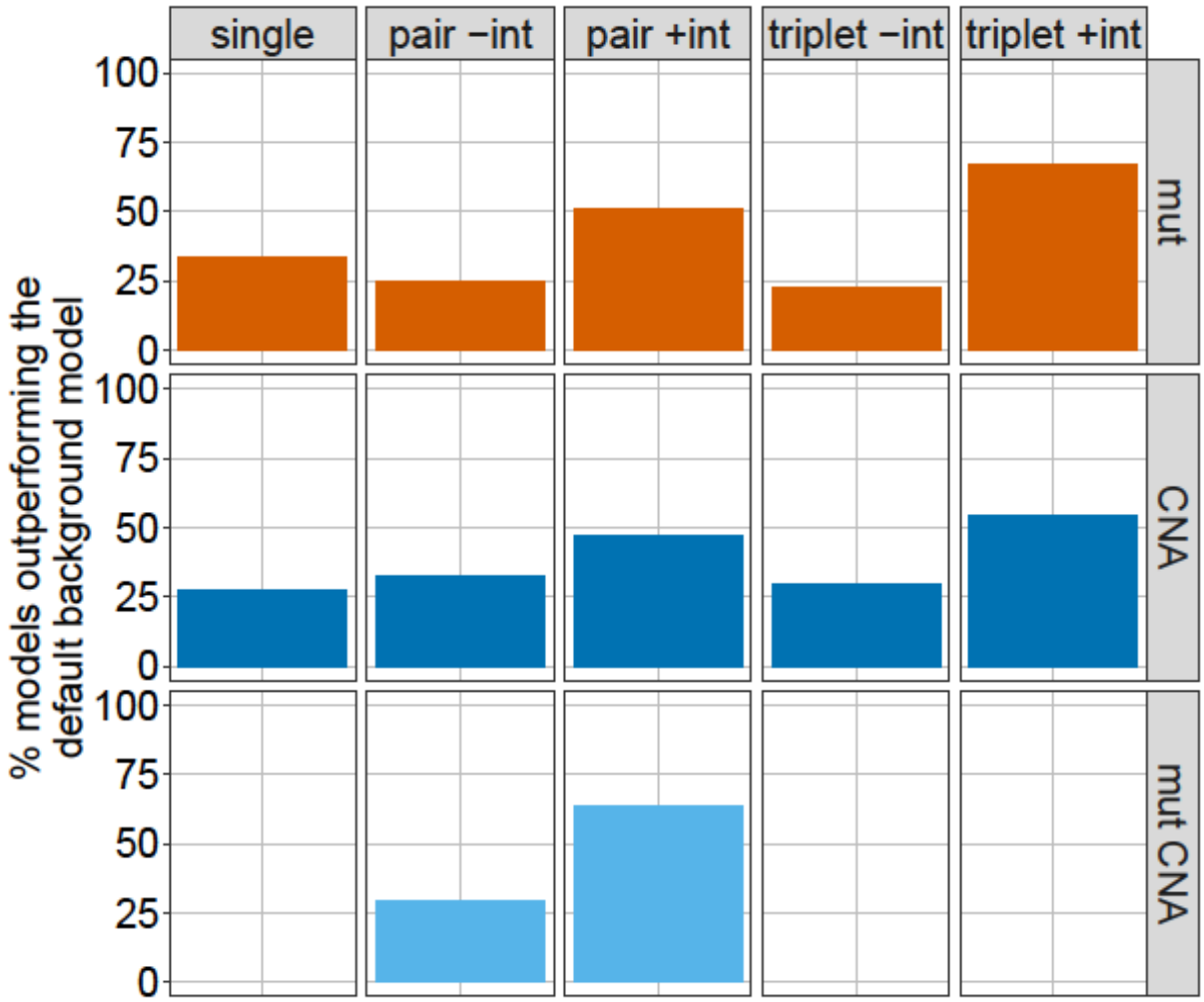

Figure 8: Proportion of drugs for which the best model per complexity outperforms the default background model based on a model comparison test ($p < 0.05$, F-test). Models are grouped by model size (horizontal facets) and alteration type (vertical facets). Percentages refer to the best models selected by adj. R^2. Mut: mutation.

To complement the selection of useful models based on model comparison tests, we evaluated model performance based on the BIC. As a reference value, we computed the BIC of the default background model. For each model complexity, we extracted the model with the lowest BIC. To determine the percentage of drugs for which a model of a given complexity performed better, we compared the BIC values of the model with alterations and the default background model (Figure 9).

Of all best models, 81% had lower BIC values than the default background model. For individual model complexities, the percentage varied between 95% for mutation-CNA

pair models with interaction and 48% for CNA triplet models with interaction. While single alterations and alteration pairs showed consistently high percentages, we observed considerably lower fractions for alteration triplet models. This suggests that alteration triplets constitute an upper complexity boundary for some drugs. For all drugs, there was at least one model with a lower BIC than the default background model, which implies that models with up to three alterations represent useful drug prediction models.

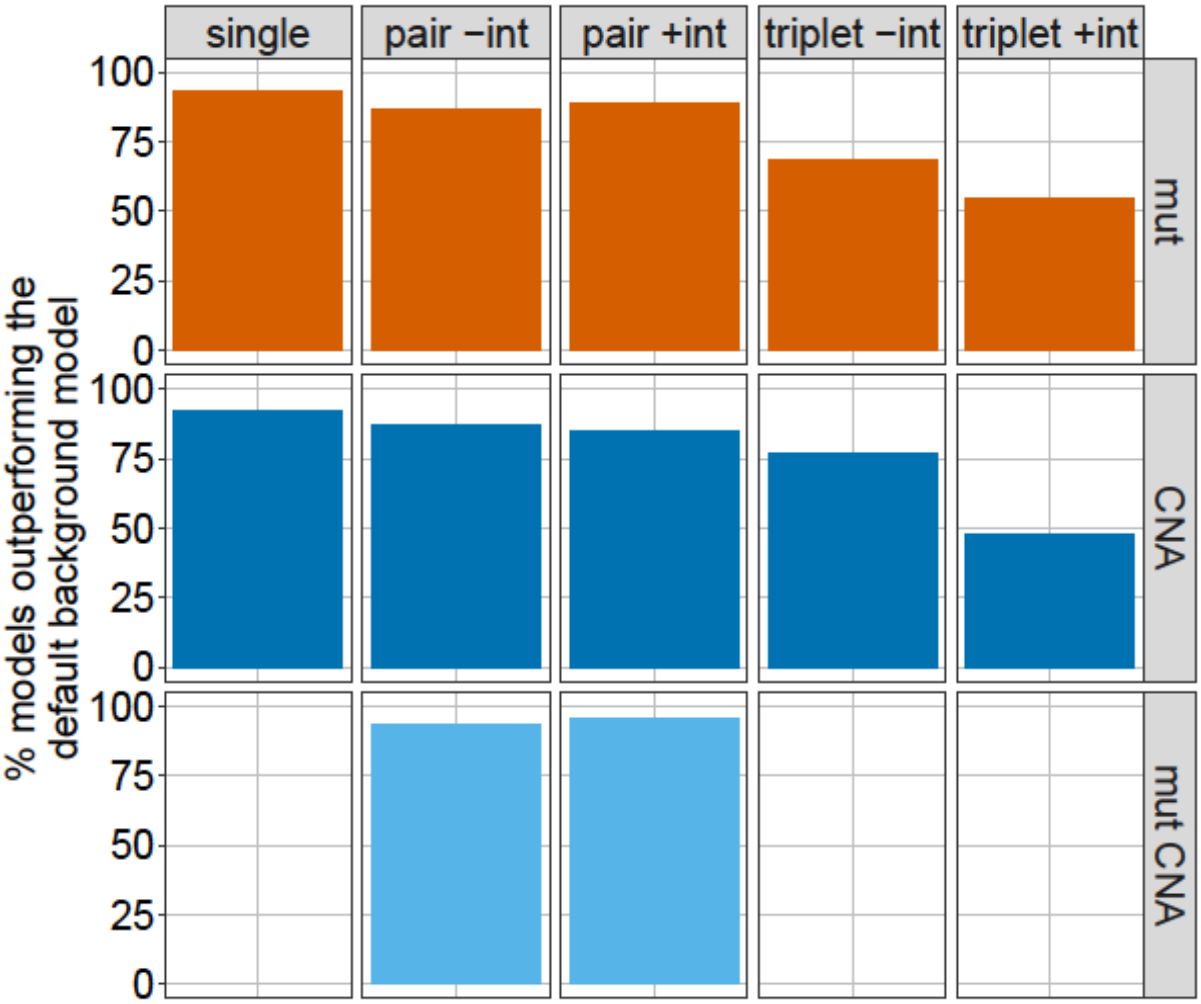

Figure 9: Proportion of drugs for which the best model per complexity outperforms the default background model based on the Bayesian information criterion (BIC). Models are grouped by model size (horizontal facets) and alteration type (vertical facets). Percentages refer to the best models selected by BIC. Mut: mutation.

1.3.3. Best model per drug

Having chosen the best models per drug and complexity, we selected the best model per drug across complexities, resulting in one model per drug. As described above, we used the adj. R^2 and the BIC as model performance metrics.

In summary, the adj. R^2 clearly favored alteration triplet models with interaction over all other model complexities (Figure 10). Mutation triplet models accounted for more than 50% of all best models per drug. Based on the BIC, the best-performing models were more evenly distributed across model complexities. In general, models with interaction made up a greater proportion of best models than models without interaction. Mutation-CNA pair models with interaction represented the best model complexity for the greatest number of drugs. In summary, the selection of best models per drug suggests that models with interaction outperform models without interaction.

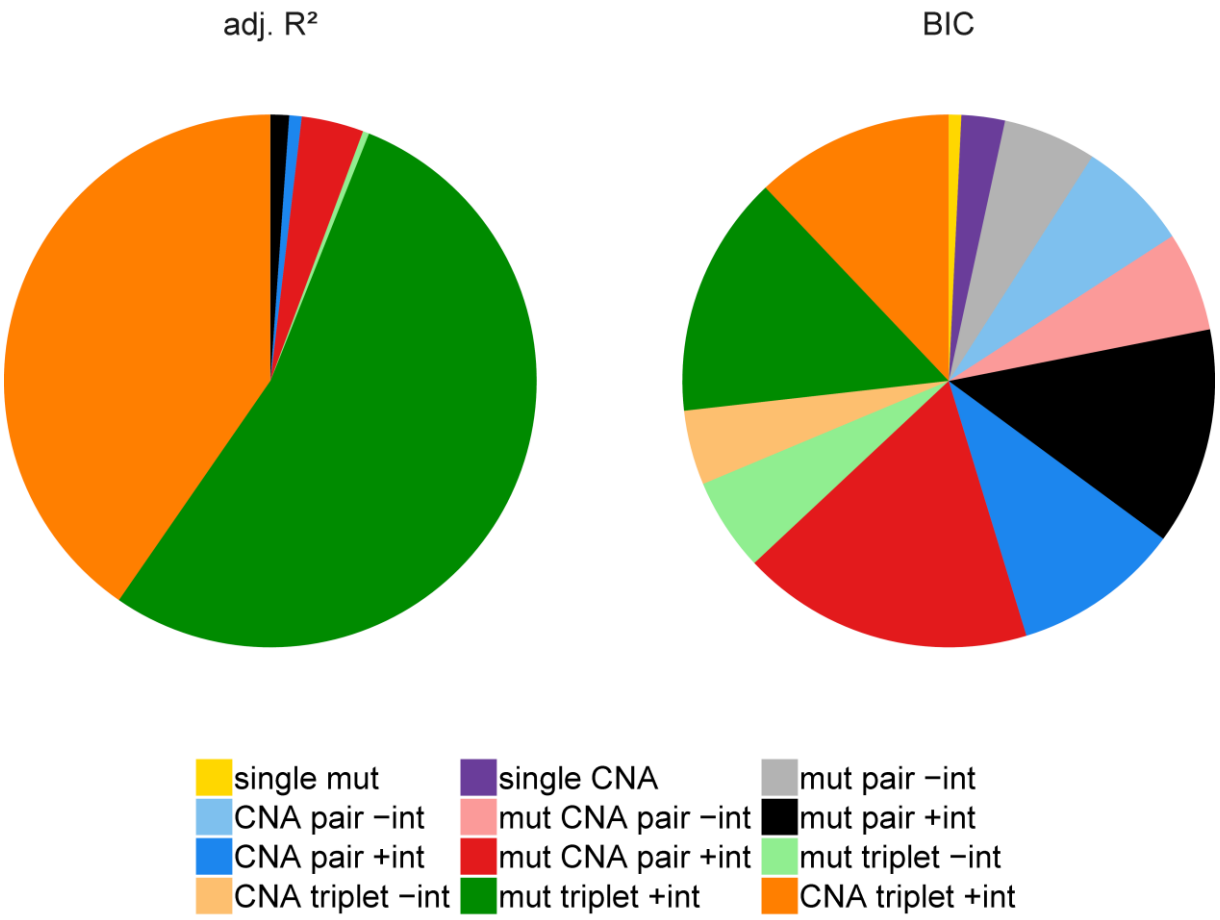

Figure 10: Best model per drug across model complexities. The proportion of models for which a given model complexity represents the best model across model complexities based on adj. R^2 (left) or BIC (right) is shown. Mut: mutation, -int: without interaction, +int: with interaction.

1.3.4. Model robustness

To estimate model robustness, we performed a cross-validation analysis. Due to the processing and memory intensity in a cross-validation setting, alteration triplet models were excluded from this analysis part. In each cross-validation instance (see Methods), the data was split into training and test set. We fitted models of different complexity to the training data and evaluated model performance using the prediction error of the test set. The model with the lowest test error was chosen as the best model so that a different model can be selected in each cross-validation instance. We used the frequency of the most frequently selected model as a proxy for model robustness.

When comparing the robustness of models including the same number, but different types of alterations, we found that single-CNA models were less robust than single-mutation models ($p < 10^{-7}$, Mann-Whitney-Wilcoxon test; Figure 11). Similarly, CNA pair models with interaction showed lower robustness than mutation pair models with interaction ($p = 0.02$, Mann-Whitney-Wilcoxon test) and mutation-CNA pair models with interaction ($p = 10^{-4}$, Mann-Whitney-Wilcoxon test). This implies that alteration pair models that include only CNAs are more sensitive to changes in the training data than alteration pair models that include at least one mutation.

When comparing model robustness within a given alteration type, but across model complexities, single-alteration models were more robust than alteration pair models ($p < 10^{-15}$ for all pairwise comparisons, Mann-Whitney-Wilcoxon test). However, alteration pair models with interaction were more robust than alteration pair models without interaction ($p < 10^{-5}$ for mutation pair models, $p = 0.02$ for CNA pair models and $p = 10^{-5}$ for mutation-CNA pair models). This suggests that including interactions increases model robustness to changes in the training data.

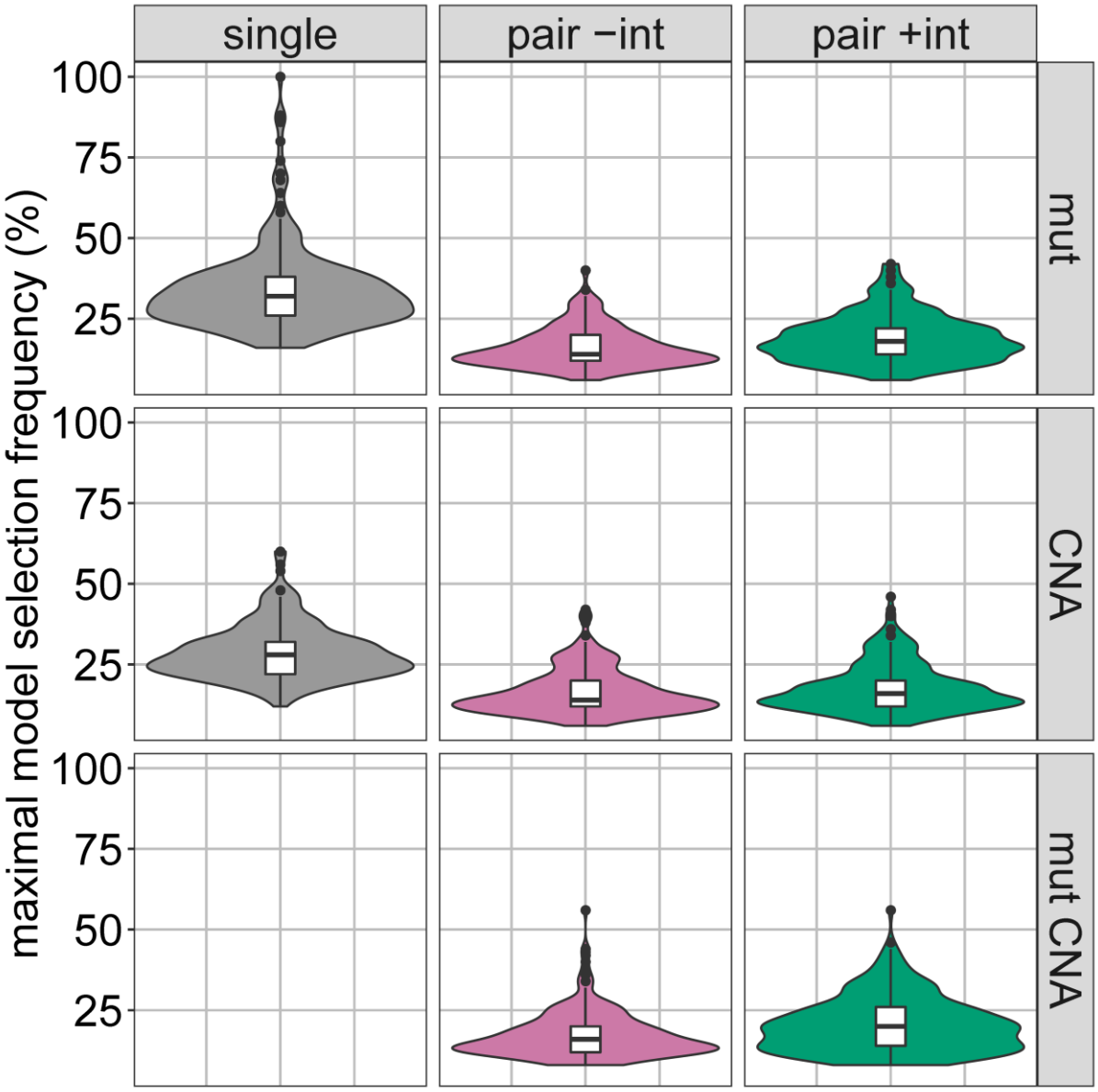

Figure 11: Model robustness. Each data point represents the most frequent model across cross-validation instances for one drug. Models are grouped by model size (horizontal facets) and alteration type (vertical facets).

1.3.5. Performance differences between model complexities

To complement our analyses based on adj. R^2 and BIC, we evaluated model performance based on test errors. As lower performance boundaries, we included a null model predicting the mean drug sensitivity across cell lines and the default background model containing only covariates. For each drug, we compared all model

complexities in pairs. Figure 12 shows the percentage of drugs for which a specific model complexity predicts drug response better or worse than another model complexity ($p < 0.05$, Mann-Whitney-Wilcoxon test).

Since the covariates of the default background model are included in all models with alterations and explain large proportions of the variation in drug response (Figure 2), we used the default background model as a reference to evaluate performance differences. We found that models with interaction outperformed the default background model for more drugs than models without interaction. Including interactions increased the percentage of models outperforming the default background model by 21% (48% - 27%) for mutation pair models, 19% (42% - 23%) for CNA pair models, and 26% (54% - 28%) for mutation-CNA pair models (Figure 12). Models with two instead of one alteration showed comparable increases of 21% (27% - 6%) for mutation pair models, 17% (23% - 6%) for CNA pair models, and 22% (28% - 6%) for mutation-CNA pair models. Accordingly, adding an interaction term to an alteration pair model can be as beneficial to performance as adding a second alteration to a single-alteration model. In contrast to CNA pair models, mutation pair models and mutation-CNA pair models with interaction were never outperformed by any other model complexity we tested ($p \geq 0.05$, Mann-Whitney-Wilcoxon test), which suggests that these model complexities are optimal in our setting.

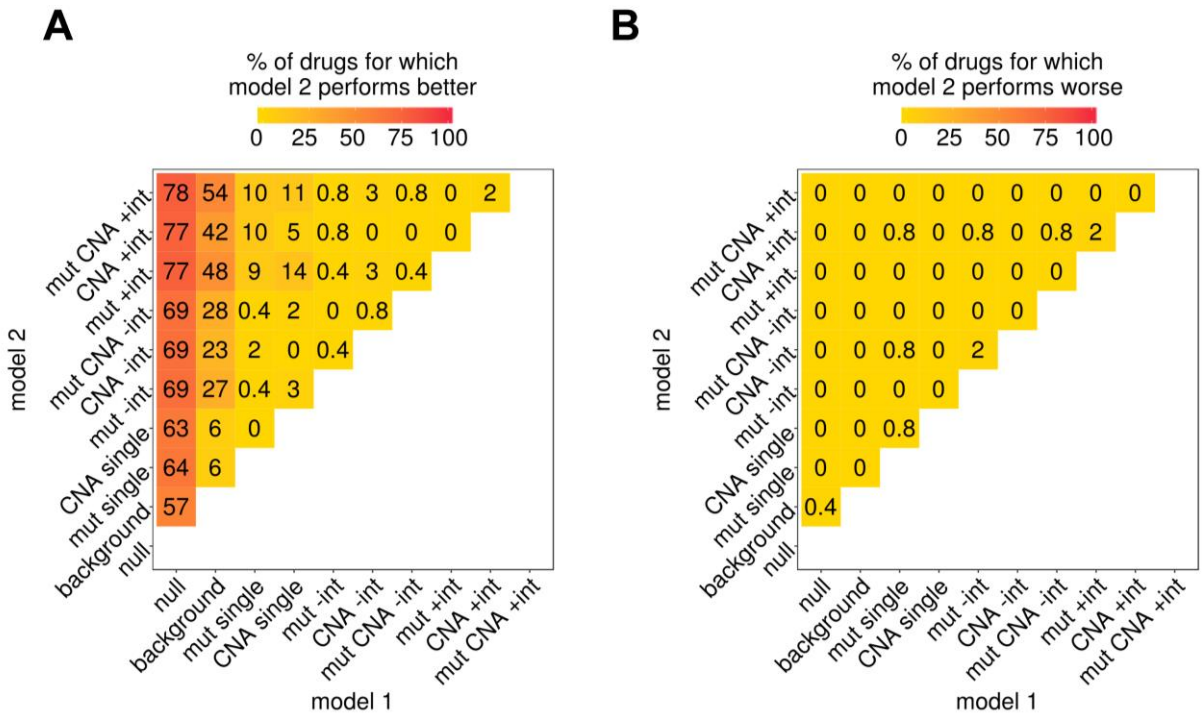

Figure 12: Comparison of model complexities based on test errors. The distribution of test errors across 50 cross-validation instances for model 1 and 2 are compared. The percentage of drugs for which model 2 had A) better and B) worse test errors than model 1 ($p < 0.05$, Mann-Whitney-Wilcoxon test) is shown. Models are denoted as null (null model), background (default background model), single (single-alteration model), -int (alteration pair model without interaction) and +int (alteration pair model with interaction). Mut: mutation.

1.3.6. Interaction examples

To stringently select interaction examples, we focused on alteration pair models with interaction that significantly outperformed alteration pair models without interaction based on test errors ($p < 0.05$, Mann-Whitney-Wilcoxon test; Figure 12A).

1.3.6.1. Mutation pair models

We found that mutation pair models with interaction outperformed mutation pair models without interaction for one drug, the BRAF inhibitor Dabrafenib ($p = 0.04$, Mann-

Whitney-Wilcoxon test). The most frequent model across cross-validation instances (frequency: 34%) was BRAF*TP53. This model was also selected as the best model based on adj. R^2 and BIC. The BRAF*TP53 model predicts that cell lines with BRAF and TP53 mutations respond worse to Dabrafenib than cell lines with only BRAF mutations ($p < 10^{-10}$, t-test; Figure 13A).

Similarly, the BRAF*TP53 model was the best mutation pair model with interaction based on adj. R^2 and BIC for the BRAF inhibitor PLX4720 (drug ID 1371). We observed that the BRAF-TP53 interaction was associated with resistance to PLX4720 ($p < 10^{-7}$, t-test; Figure 13B). For both Dabrafenib and PLX4720, we validated the interaction effect between BRAF and TP53 in the CTRP dataset ($p = 0.005$ (Dabrafenib) and $p < 10^{-4}$ (PLX4720), t-test; Figure 13). This suggests that the resistance association we observed is robust across different BRAF inhibitors and datasets.

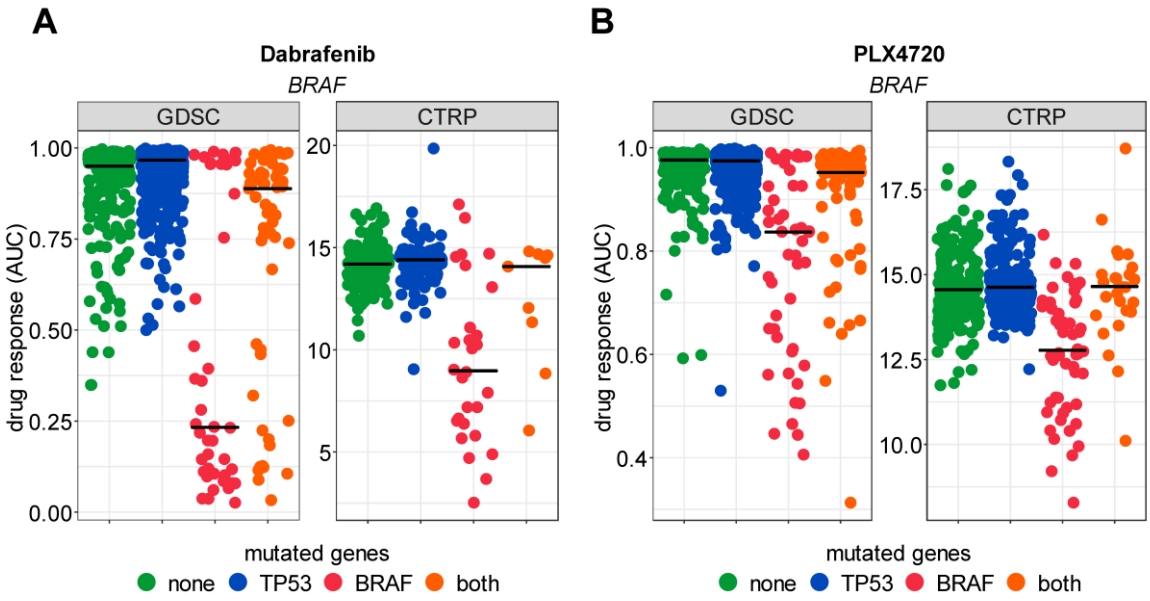

Figure 13: Simultaneous mutation of BRAF and TP53 mediates resistance to the BRAF inhibitors A) Dabrafenib and B) PLX4720 in the GDSC and the CTRP dataset. Each point represents a single cell line. Cell lines are grouped by mutation status of TP53 and BRAF and by dataset. The median drug response per group is specified by horizontal lines. Drugs and drug targets are depicted in bold and italics, respectively.

Both Dabrafenib and PLX4720 are selective inhibitors of BRAF kinases harboring the gain-of-function V600E mutation (Rheault et al., 2013; Tsai et al., 2008). Therefore, we reassessed the BRAF*TP53 model by restricting BRAF mutations to $BRAF^{V600E}$ mutations. We confirmed the BRAF-TP53 interaction for Dabrafenib in the GDSC dataset ($p < 10^{-9}$, t-test) and for PLX4720 in both datasets ($p < 10^{-4}$ (GDSC) and $p = 0.048$ (CTRP), t-test). We could not confirm the BRAF-TP53 interaction for Dabrafenib in the CTRP dataset since too few cell lines with simultaneous $BRAF^{V600E}$ and TP53 mutations were screened (see Methods).

To test for translatability of our *in vitro* results to *in vivo* data, we used a clinical trial dataset including 31 melanoma patients with $BRAF^{V600}$-mutated tumors that were treated with the BRAF inhibitor Vemurafenib (Van Allen et al., 2014). We extracted the mutation status of BRAF and TP53 and clinical response to the BRAF inhibitor Vemurafenib, which is a PLX4720 analog (Michaelis et al., 2014). We found that patients with BRAF, but without TP53 mutations tend to respond better to Vemurafenib treatment than patients with both BRAF and TP53 mutations (disease control rate: 50% and 85%, relative risk: 3.25 [90% confidence interval 1.06-9.94], $p = 0.1$; see Methods). The response differences in patients with BRAF-mutated tumors with or without additional TP53 mutations suggest clinical relevance.

1.3.6.2. Mutation-CNA pair models

For two drugs, a mutation-CNA pair model with interaction outperformed a mutation-CNA pair model without interaction (Figure 12A). For example, mutation-CNA models with interaction performed better for the BRAF inhibitor SB590885 ($p = 0.04$, Mann-Whitney-Wilcoxon test). The most frequent model across cross-validation instances (frequency: 42%) contained a BRAF mutation and an amplification of the genomic region 3p14.1. This model was also selected as the best model based on adj. R^2 and BIC. The amplified 3p14.1 region includes the genes encoding the transcription factors FOXP1 and MITF. The model predicts that simultaneous mutation of BRAF and amplification of 3p14.1 is associated with drug sensitivity ($p < 10^{-19}$, t-test; Figure 14).

We could not validate the model in the CTRP dataset since SB590885 was not screened.

Since SB590885 selectively inhibits activated $BRAF^{V600E}$ kinases (King et al., 2006), we excluded cell lines with other BRAF mutations than $BRAF^{V600E}$ from the BRAF-mutated group. Retesting our model confirmed the sensitizing interaction effect of a BRAF mutation and a 3p14.1 amplification ($p < 10^{-18}$, t-test).

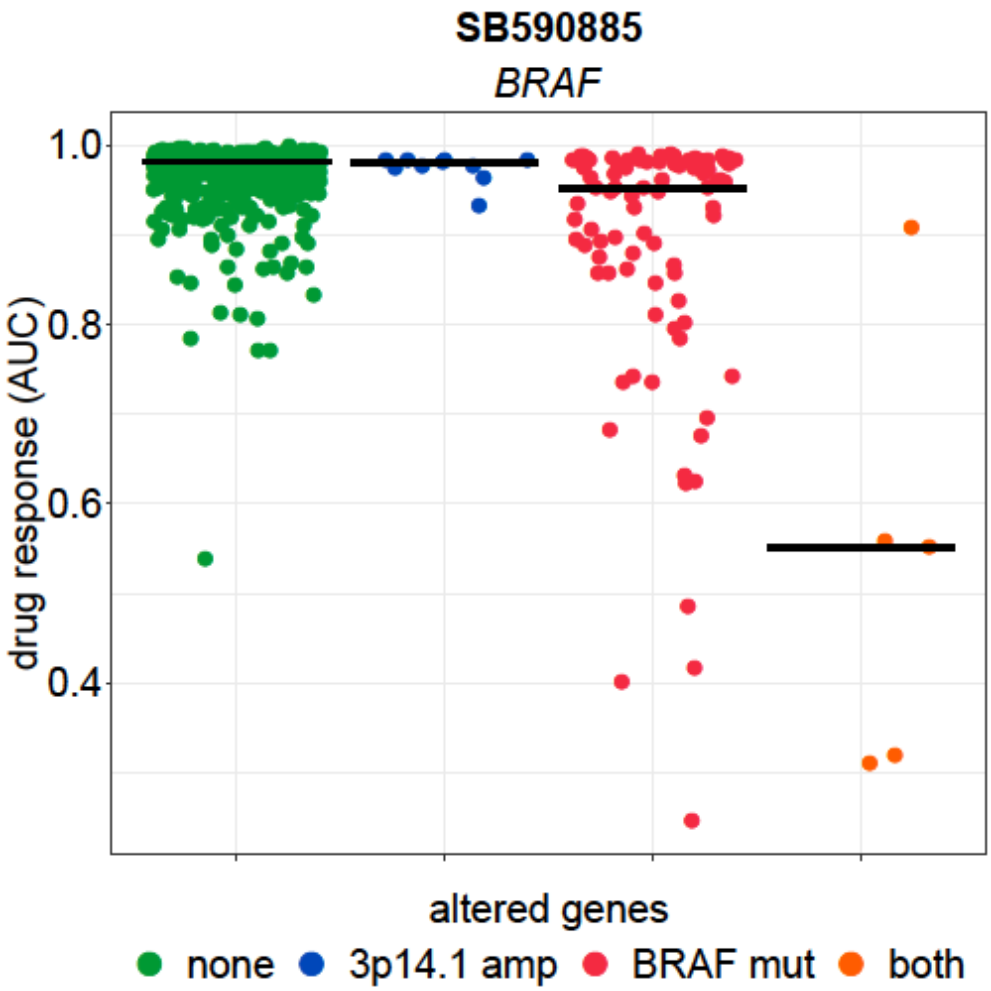

Figure 14: Simultaneous mutation of BRAF and amplification of 3p14.1 mediates sensitivity to the BRAF inhibitor SB590885 in the GDSC dataset. Each point represents a single cell line. Cell lines are grouped by alterations (amp: amplification, mut: mutation) and the median drug response per group is indicated by horizontal lines. The drug target is depicted in italics.

The second drug for which mutation-CNA models with interaction outperformed mutation-CNA models without interaction was PLX4720 (drug ID 1036, $p = 0.03$, Mann-Whitney-Wilcoxon test). For PLX4720, the most frequent mutation-CNA pair model with interaction included a BRAF mutation and a 3q26.1 amplification (frequency: 56%). This model was also selected as the best model based on adj. R^2 and BIC. The sensitizing contribution of the interaction term between a BRAF mutation and a 3q26.1 amplification was consistent in the GDSC ($p < 10^{-19}$, t-test) and the CTRP dataset ($p = 0.04$, t-test; Figure 15).

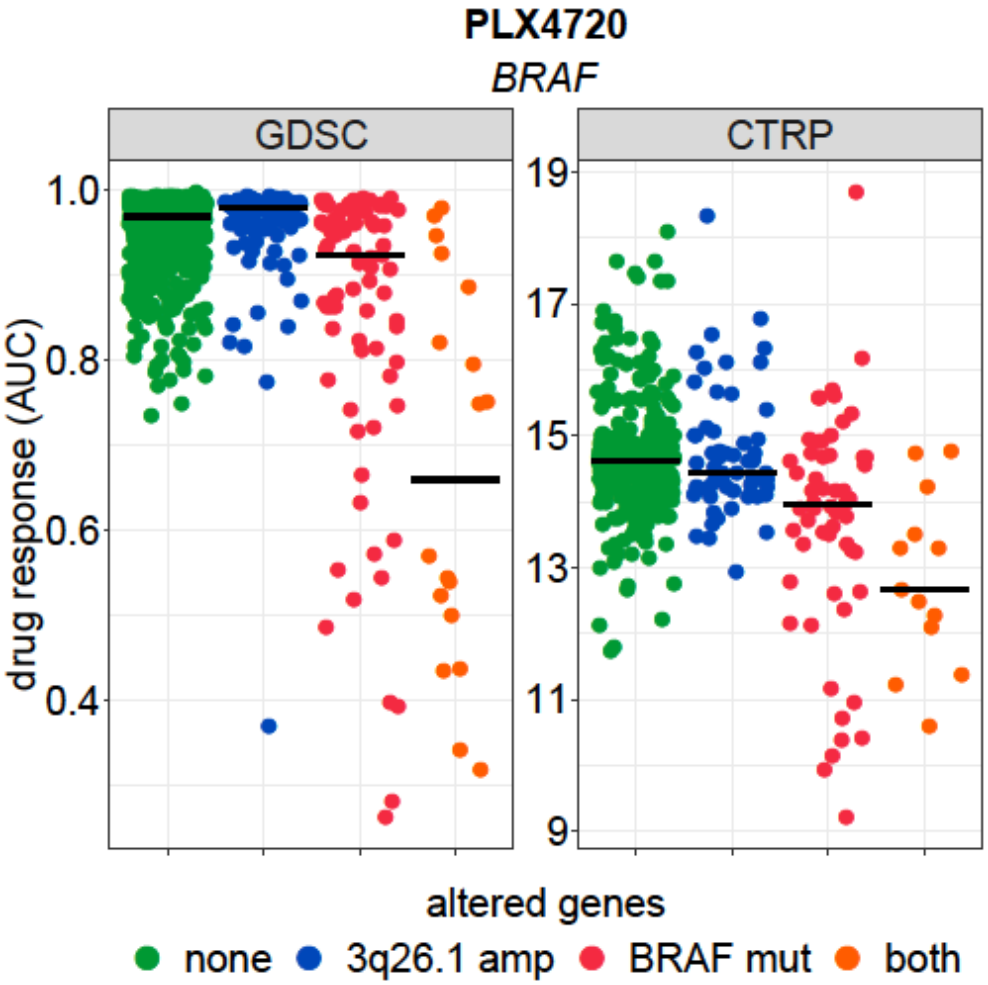

Figure 15: Simultaneous 3q26.1 amplification and BRAF mutation mediates PLX4720 sensitivity in the GDSC and the CTRP dataset. The drug target BRAF is indicated in italics. Each point represents an individual cell line. Cell lines are grouped by dataset (GDSC or CTRP) and alteration status (amp: amplification, mut: mutation). The median drug response for each group is specified by horizontal lines.

When restricting the BRAF-mutated group to cancer cell lines with $BRAF^{V600E}$ mutations, the sensitizing interaction effect was significant in the GDSC dataset ($p < 10^{-17}$, t-test) and marginally significant in the CTRP dataset ($p = 0.1$, t-test).

1.3.7. Duplicate model

In addition to the lower performance boundary represented by the default background model, we used the agreement of duplicate screens as an upper performance boundary. In the GDSC dataset, 14 unique drugs were screened twice such that 28 drugs can be compared. The drug response values of the duplicate screens showed Pearson correlations between 0.13 ($p = 0.0002$) for the Bromodomain-containing protein inhibitor JQ1 and 0.84 ($p < 10^{-15}$) for the EGFR/ ERBB2 inhibitor Afatinib. For each of the 28 drugs, we fitted a univariate model with the drug response values of the duplicate screen as the predictive variable. We termed these models duplicate models.

We found that the duplicate model performed better than models with up to two alterations (Figure 12) for 16 drugs and worse for 12 drugs ($p < 0.05$, Mann-Whitney-Wilcoxon test). We further assessed the 12 drugs for which the duplicate model was outperformed by models with alterations.

For 8 of the 12 drugs, the duplicate model did not only perform worse than models with alterations but also worse than a null model predicting the mean drug response across cell lines. Since the mean drug sensitivity is a better predictor than the duplicate drug response, this suggests low variability of the drug response data.

For another 3 of the 12 drugs, the duplicate model was outperformed by the default background model. This suggests that the covariates tissue, growth medium, growth properties, and CNA count are more robust predictors than the duplicate drug response.

For 1 of the 12 drugs, the MEK inhibitor RDEA119 (drug ID 1014), a mutation pair model with interaction was the only complexity for which the duplicate model was outperformed. The model that was most frequently selected in the cross-validation instances was ATM*NRAS (frequency: 16%). This model was also the best model

based on adj. R^2 and BIC. The model predicts that cell lines with simultaneous ATM and NRAS mutations respond worse to RDEA119 than cell lines with only NRAS mutations (Figure 16). We could not validate this model in the CTRP dataset since RDEA119 was not screened.

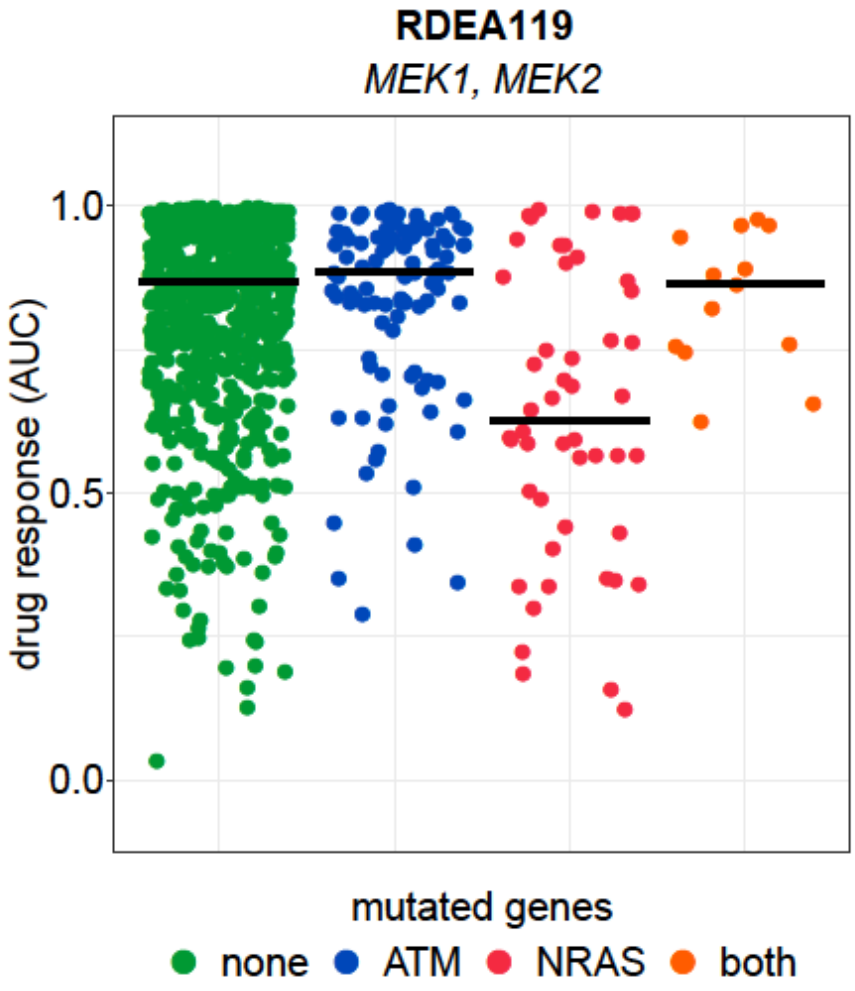

Figure 16: Simultaneous mutation of ATM and NRAS mediates resistance to the MEK inhibitor RDEA119 in the GDSC dataset. Each point represents a single cell line. Cell lines are grouped by mutation status of ATM and NRAS and the median drug response per group is shown (horizontal lines). The drug target is depicted in italics.

2. Tissue specificity

Having evaluated the role of genetic interactions in drug response (Chapter 1), we tested whether the association between genomic alterations and drug response depends on factors other than alterations. Since the tissue of origin strongly affects drug response (Figure 2), we assessed whether the tissue covariate can modulate the association between a genomic alteration and drug response. Due to the high number of tissue categories (see Methods), the alteration frequency across cell lines limits the detection of statistical associations. Since mutations are more frequent than CNAs (Figure 5), we analyzed the tissue specificity of mutation associations.

To distinguish mutation associations that are tissue-specific from associations that can be generalized across tissues, we introduced interactions between individual mutations and the tissue covariate in single-mutation models (see Methods). We defined associations between mutations and drug response as tissue-specific if the mutation-tissue interaction term improved the model ($p < 0.05$, F-test) and as generalizable across tissues otherwise. We found that 17% (374 out of 2232) of the mutation associations were tissue-specific.

To assess whether the tissue specificity we observed is driven by random differences between data subsets, we generated 1000 randomized datasets. In each randomization instance, we shuffled the tissue annotation while maintaining the uneven mutation distribution across tissues (see Methods). The percentage of tissue-specific associations in the original dataset was larger than in the randomized datasets ($p < 0.001$, randomization test; Figure 17).

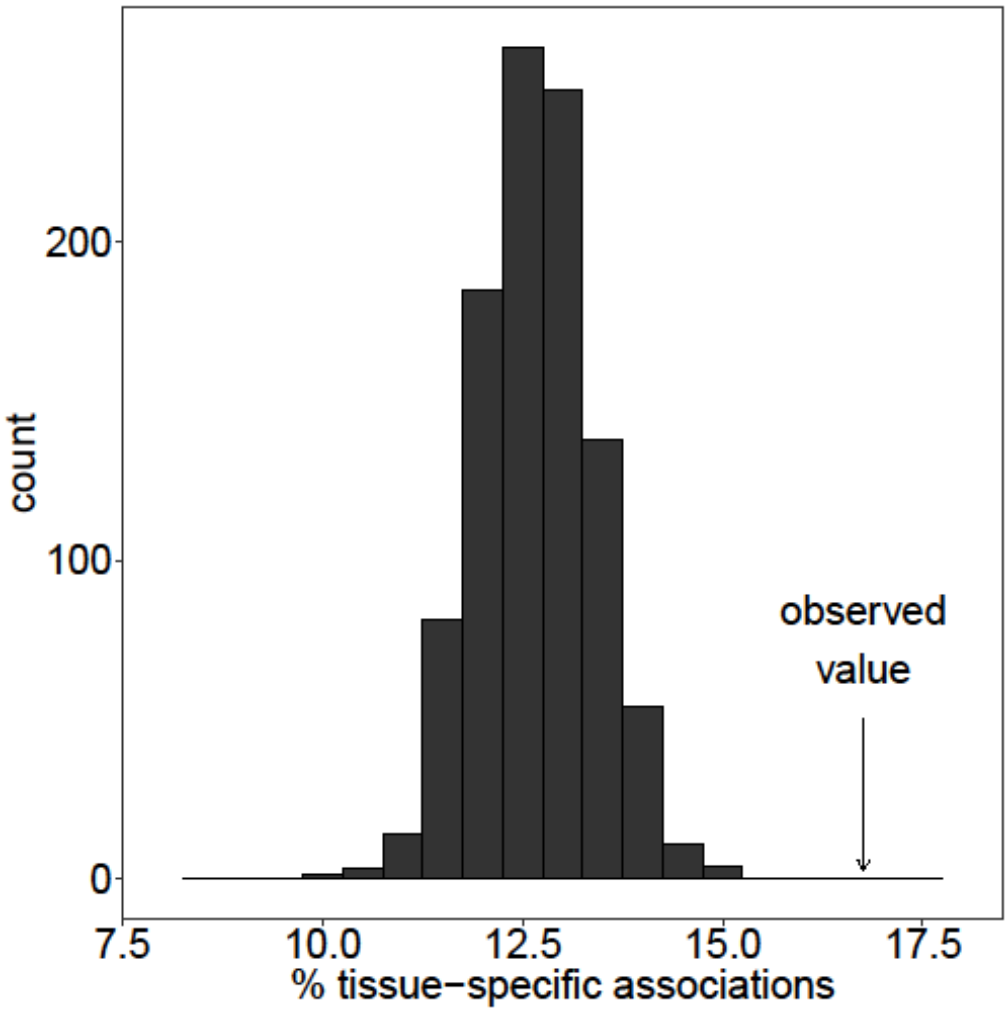

Figure 17: Proportion of tissue-specific mutation associations. Single-mutation models with mutation-tissue interaction term were classified as tissue-specific ($p < 0.05$, F-test) or general. The observed percentage of tissue-specific associations is higher than in datasets with randomly sampled tissue annotations ($p < 0.001$, randomization test).

2.1. Models with tissue-specific mutation associations outperform models with general mutation associations

To test whether tissue-specific and general mutation associations show differences in model performance, we assessed the adj. R^2 of single-mutation models with mutation-tissue interaction term. Since the mutation-tissue interaction term is significant ($p < 0.05$, F-test) for tissue-specific associations only, we expected a

positive contribution to the adj. R^2 for tissue-specific rather than associations. Accordingly, we found that tissue-specific models tend to explain higher proportions of the variation in drug response than general models ($p < 10^{-18}$, Mann-Whitney-Wilcoxon test).

To test whether the improvement in performance is greater than expected by chance, we generated 100 randomized datasets by permuting the tissue annotation and maintaining the unequal mutation frequencies across tissues (see Methods). For the original dataset and the randomized datasets, we computed the estimated performance difference between tissue-specific and general models. The observed performance difference between tissue-specific and general models exceeded random expectation ($p = 0.01$, randomization test, Figure 18).

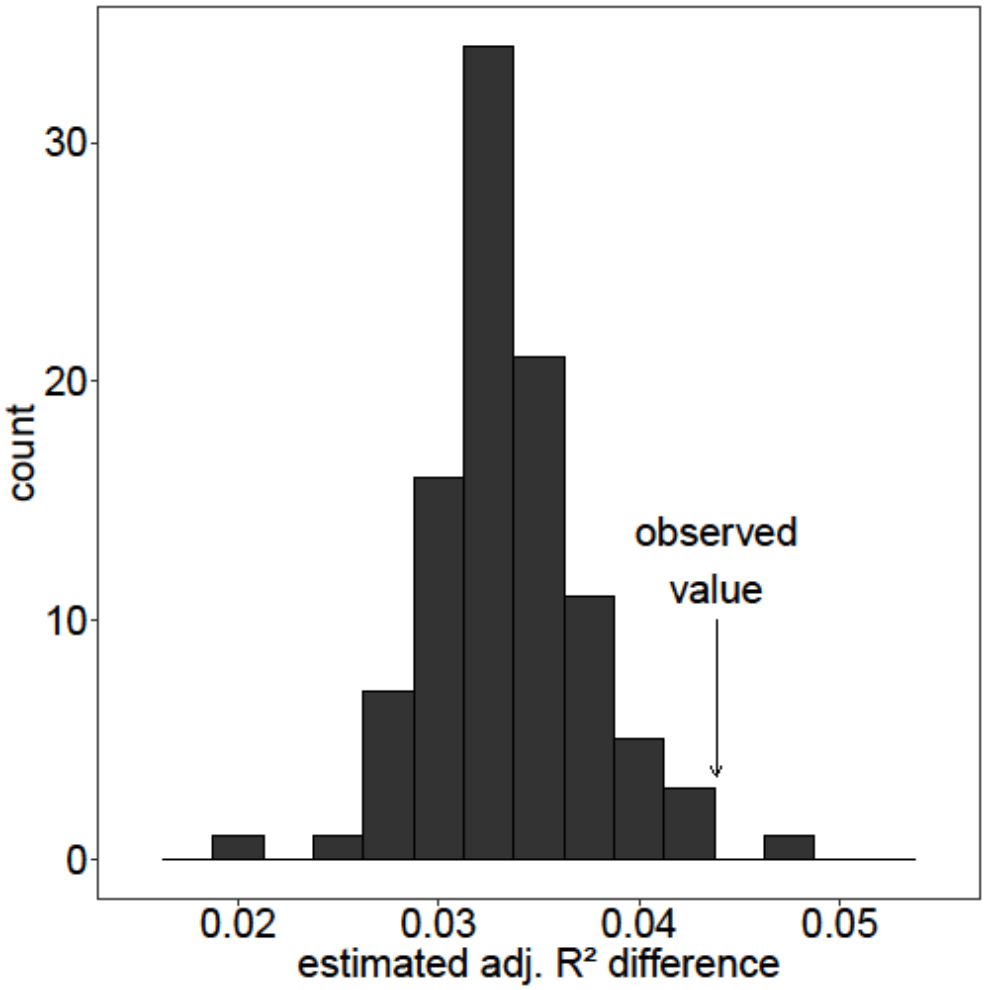

Figure 18: Performance difference between models with tissue-specific and general mutation associations. Models were compared based on the adj. R^2. The performance difference in the original data (observed value) is greater than in data with randomly sampled tissue annotations ($p = 0.01$, randomization test).

2.2. Mutations can mediate drug sensitivity and resistance in different tissues

To further characterize tissue-specific associations between a mutation and drug response, we filtered for associations where the mutation coefficient was significant in more than one single-tissue model ($p < 0.05$, t-test). The sign of the mutation coefficient estimate defines the direction of the effect. Negative mutation coefficients represent sensitivity associations while positive coefficients represent resistance associations. For the same gene and the same drug, the mutation coefficient switched

the sign across tissues in 15% (17 out of 115) of the associations. Since most mutations consistently mediate either drug sensitivity or drug resistance in all cancer types, this indicates that differences between tissues of origin tend to be quantitative rather than qualitative.

To compare the percentage of sign-switching mutation associations to random expectation, we generated 100 randomized datasets by shuffling the tissue annotation (see Methods). For the original dataset, the percentage of mutation associations showing different signs across tissues was larger than expected at random ($p < 0.01$, randomization test; Figure 19). This suggests that mutations can be associated with drug sensitivity and resistance in different tissues.

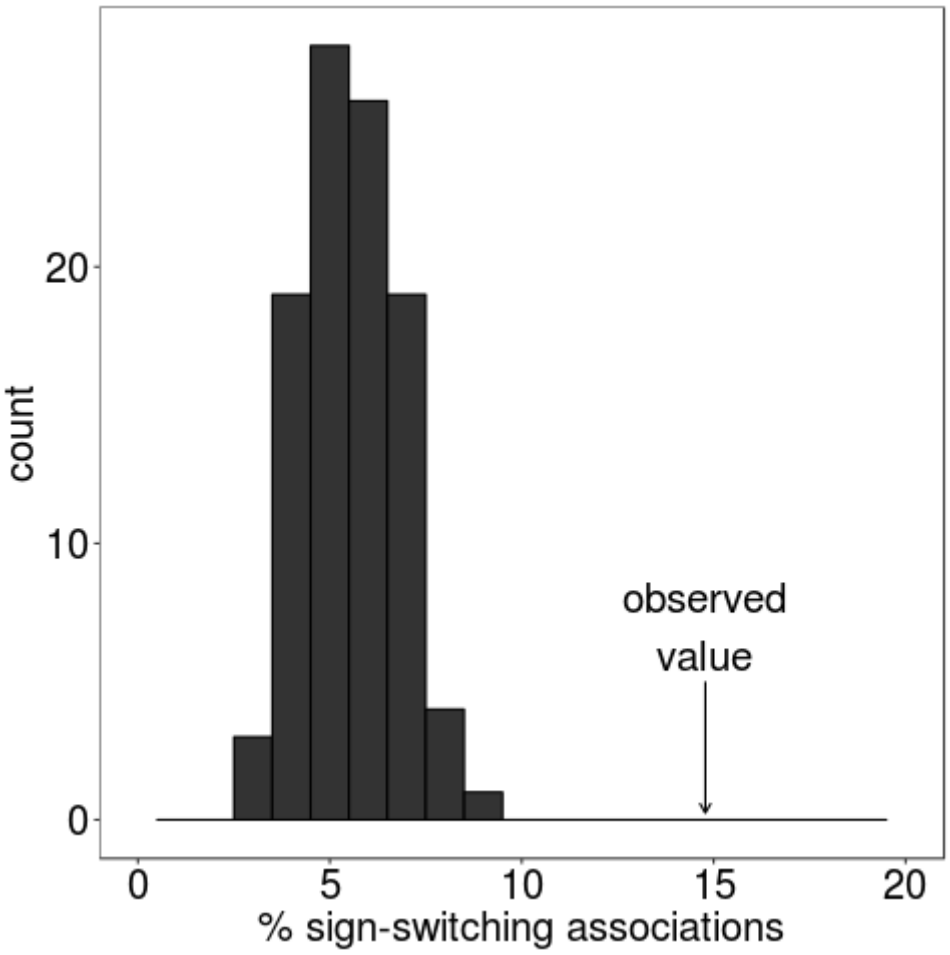

Figure 19: Proportion of mutations mediating drug resistance and sensitivity in different tissues. The percentage of sign-switching mutation associations in the original data (observed value) is higher than in data with randomly sampled tissue annotations ($p < 0.01$, randomization test).

The largest difference in mutation coefficient estimates between tissues was observed for the association of BRCA1 mutations with response to the AURKB inhibitor ZM-447439 (Figure 20). Mutations in BRCA1 tend to sensitize to the drug in large intestine derived cancer cell lines ($p = 0.03$, t-test) and confer resistance in leukemia cell lines ($p = 0.003$, t-test). In cancer cells originating from the aerodigestive tract, the digestive system (other than large intestine), the urogenital system and the lung (non-small cell lung cancer), BRCA1 mutations were associated with ZM-447439 resistance ($p \geq 0.05$, t-test). We could not validate this finding in the CTRP dataset since the drug was not screened.

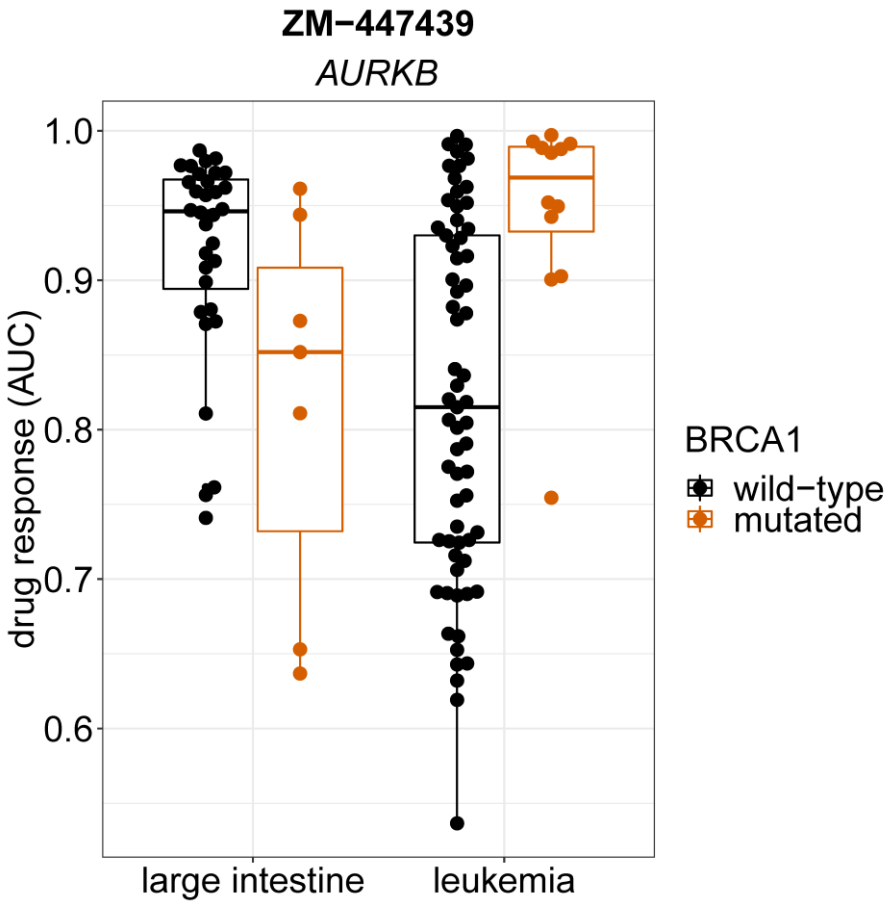

Figure 20: BRCA1 mutations mediate sensitivity or resistance to the AURKB inhibitor ZM-447439 in different tissues. BRCA1 mutations are associated with drug sensitivity in large intestine cancer cell lines and with resistance in leukemia cell lines. Drug response values are grouped by tissue of origin and by mutation status of BRCA1. Each point represents one cancer cell line.

2.3. Examples of tissue-specific and general mutation associations

To increase confidence in our list of tissue-specific and general mutation associations, we filtered for consistency between the GDSC and the CTRP dataset. Compared to the unfiltered list, we observed an increase in the percentage of tissue-specific mutation associations. In the validated list, 32% of the mutation associations were tissue-specific (cf. Figure 17).

To illustrate the differences between tissue-specific and general associations, we fitted single-mutation models to data from individual tissues. The association between a given mutation and a given drug was compared across tissues. Figure 21 displays all tissue-specific and general mutation associations in the GDSC dataset that show consistent association patterns in the CTRP dataset.

We identified the association of EGFR mutations with response to the EGFR/ ERBB2 inhibitor Afatinib (drug screened twice; $p < 10^{-4}$ (drug ID 1032) and $p = 0.007$ (drug ID 1377), F-test) and the EGFR inhibitor Gefitinib ($p = 10^{-4}$, F-test) as examples of tissue-specific mutation associations. The sensitivity association between EGFR mutations and Afatinib and Gefitinib was especially pronounced in non-small cell lung cancer cell lines. In accordance with our results, Afatinib and Gefitinib approval is confined to the treatment of patients with EGFR-mutated non-small cell lung cancer (Kazandjian et al., 2016; Wecker and Waller, 2018).

Likewise, the association of TP53 mutations with resistance to the MDM2 inhibitor Nutlin-3a shows tissue specificity ($p < 10^{-8}$, F-test). Although several tissues of origin showed strong effects, the strongest association in the single-tissue models was observed in cancer cell lines originating from the nervous system. In line with this, mutations in TP53 were previously reported as negative predictors of Nutlin-3a response in medulloblastoma (Kunkele et al., 2012) and glioblastoma multiforme (Villalonga-Planells et al., 2011).

In addition, the association of BRAF mutations with response to the BRAF inhibitor PLX4720 is tissue-specific (drug screened twice; $p < 10^{-7}$ (drug ID 1036) and $p < 10^{-8}$ (drug ID 1371), F-test). The sensitizing effect of BRAF mutations is especially pronounced in skin and thyroid cancer cell lines. In accordance with our results, BRAF

inhibitors showed clinical efficacy in melanoma (Hauschild et al., 2012) and thyroid cancer (Falchook et al., 2015).

Similarly, we found that the association between NRAS mutations and response to the BRAF inhibitor PLX4720 shows tissue specificity (drug screened twice; $p < 10^{-4}$ (drug ID 1036 and 1371), F-test). We observed a particularly strong resistance association in skin cancer cell lines. In line with this, NRAS mutations were previously reported to confer resistance to BRAF inhibitors in melanoma by reactivating mitogen-activated protein kinase signaling (Nazarian et al., 2010).

We also observed a tissue-specific association between TP53 mutations and response to PLX4720 ($p = 0.007$ (drug ID 1371), F-test). Mutations of TP53 tend to confer resistance to PLX4720, an effect which is especially pronounced in skin cancer cell lines. In accordance with our results, a joint role for BRAF and TP53 mutations was reported in skin cancer development (Yu et al., 2009).

Besides tissue-specific mutation associations, Figure 21 shows the estimated mutation coefficients across tissues for general mutation associations ($p \geq 0.05$, F-test). For a given mutation and a given drug, the estimated mutation coefficients were almost identical which visually confirms our classification of these associations as generalizable across tissues. Overall, the strongest mutation coefficient estimates for individual tissues were weaker than for tissue-specific associations ($p < 10^{-15}$, Mann-Whitney-Wilcoxon test). This confirms our finding that models with tissue-specific mutation associations tend to achieve better performance (Figure 18).

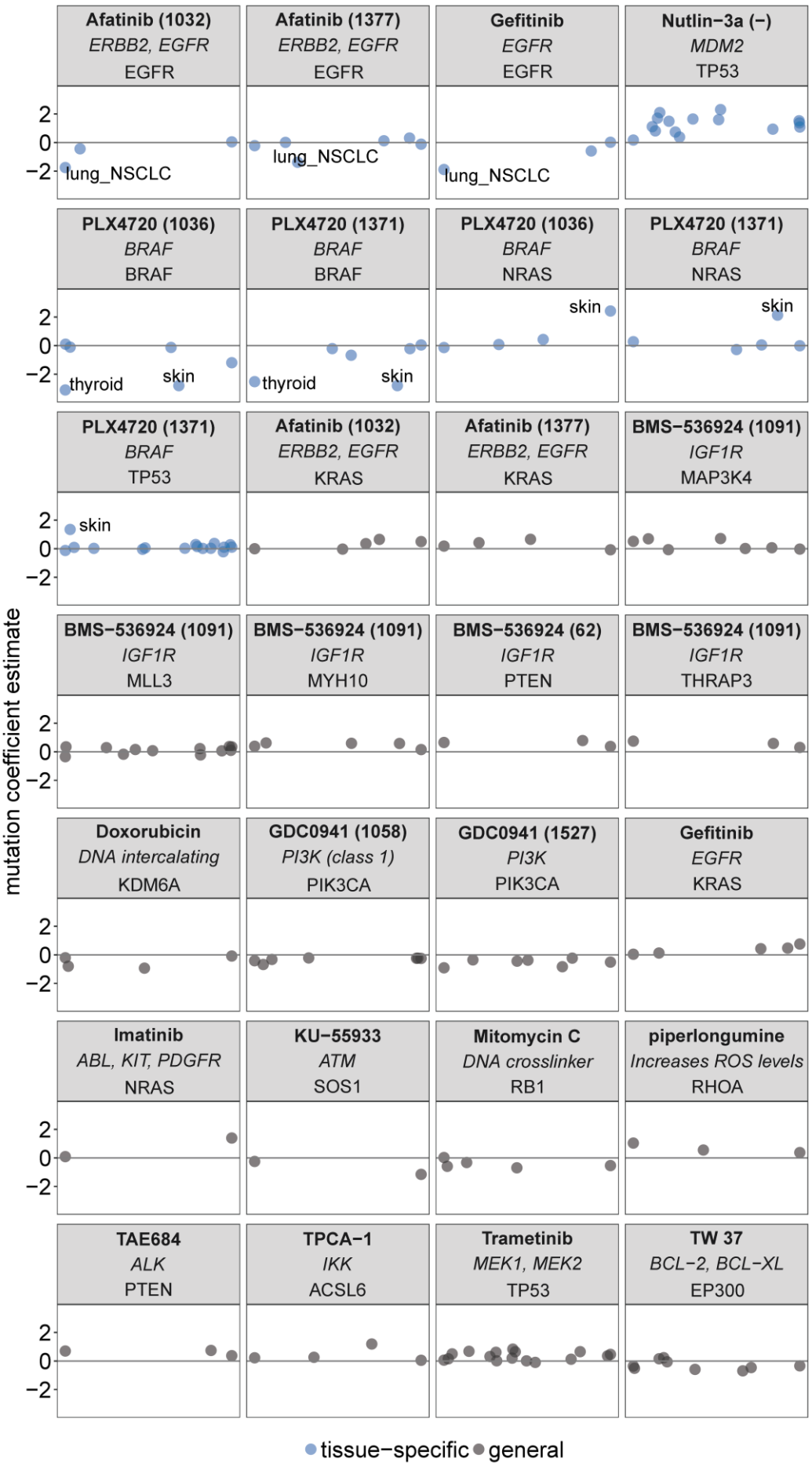
Afatinib (1032)
ERBB2, EGFR
EGFR
Afatinib (1377)
ERBB2, EGFR
EGFR
Gefitinib
EGFR
EGFR
Nutlin−3a (−)
MDM2
TP53
PLX4720 (1036)
BRAF
BRAF
PLX4720 (1371)
BRAF
BRAF
PLX4720 (1036)
BRAF
NRAS
PLX4720 (1371)
BRAF
NRAS
PLX4720 (1371)
BRAF
TP53
Afatinib (1032)
ERBB2, EGFR
KRAS
Afatinib (1377)
ERBB2, EGFR
KRAS
BMS−536924 (1091)
IGF1R
MAP3K4
BMS−536924 (1091)
IGF1R
MLL3
BMS−536924 (1091)
IGF1R
MYH10
BMS−536924 (62)
IGF1R
PTEN
BMS−536924 (1091)
IGF1R
THRAP3
Doxorubicin
DNA intercalating
KDM6A
GDC0941 (1058)
PI3K (class 1)
PIK3CA
GDC0941 (1527)
PI3K
PIK3CA
Gefitinib
EGFR
KRAS
Imatinib
ABL, KIT, PDGFR
NRAS
KU−55933
ATM
SOS1
Mitomycin C
DNA crosslinker
RB1
piperlongumine
Increases ROS levels
RHOA
TAE684
ALK
PTEN
TPCA−1
IKK
ACSL6
Trametinib
MEK1, MEK2
TP53
TW 37
BCL−2, BCL−XL
EP300
mutation coefficient estimate
2
0
−2
lung_NSCLC
thyroid
skin
tissue−specific
general

Figure 21: Comparison of tissue-specific and general mutation associations across tissues in the GDSC dataset. For each tissue of origin, a separate single-mutation model was fitted. Mutation coefficients across tissues are shown for tissue-specific (blue) and general (grey) mutation associations with drug response. Tissue-specific mutation associations show higher variability than general associations. Drug names (bold, with drug IDs in brackets), drug targets (italic) and mutated genes (plain) are indicated. NSCLC: non-small cell lung cancer.

2.3.1. Correlation between tissue-specific mutation associations and frequencies

To test whether the tissue specificity of mutation associations is linked to tissue-specific mutation frequencies, we correlated mutation coefficient estimates and mutation frequencies in single tissues. Since mutation coefficients can have positive or negative signs, absolute values represent the strength of the effect.

In the GDSC dataset, TP53 mutations showed a strong association with resistance to the MDM2 inhibitor Nutlin-3a in tissues with low TP53 mutation frequencies (Figure 22; Pearson's $r = -0.55$, $p = 0.03$). Likewise in the GDSC dataset, BRAF mutations were strongly associated with sensitivity to the BRAF inhibitor PLX4720 in tissues with high BRAF mutation frequencies (drug screened twice, Pearson's r = -0.89, p 0.01 (drug ID 1036) and r = -0.92, p = 0.008 (drug ID 1371)). In both datasets, higher TP53 mutation frequencies were related to weaker associations between TP53 mutations and PLX4720 resistance (Pearson's $r = -0.52$, $p = 0.04$ (GDSC, drug ID 1371) and $r = -0.93$, $p = 10^{-4}$ (CTRP)). This was the only tissue-specific association showing significant correlations with mutation frequencies in both datasets. In summary, we found that some tissue-specific differences in mutation associations are linked to tissue-specific mutation frequencies.

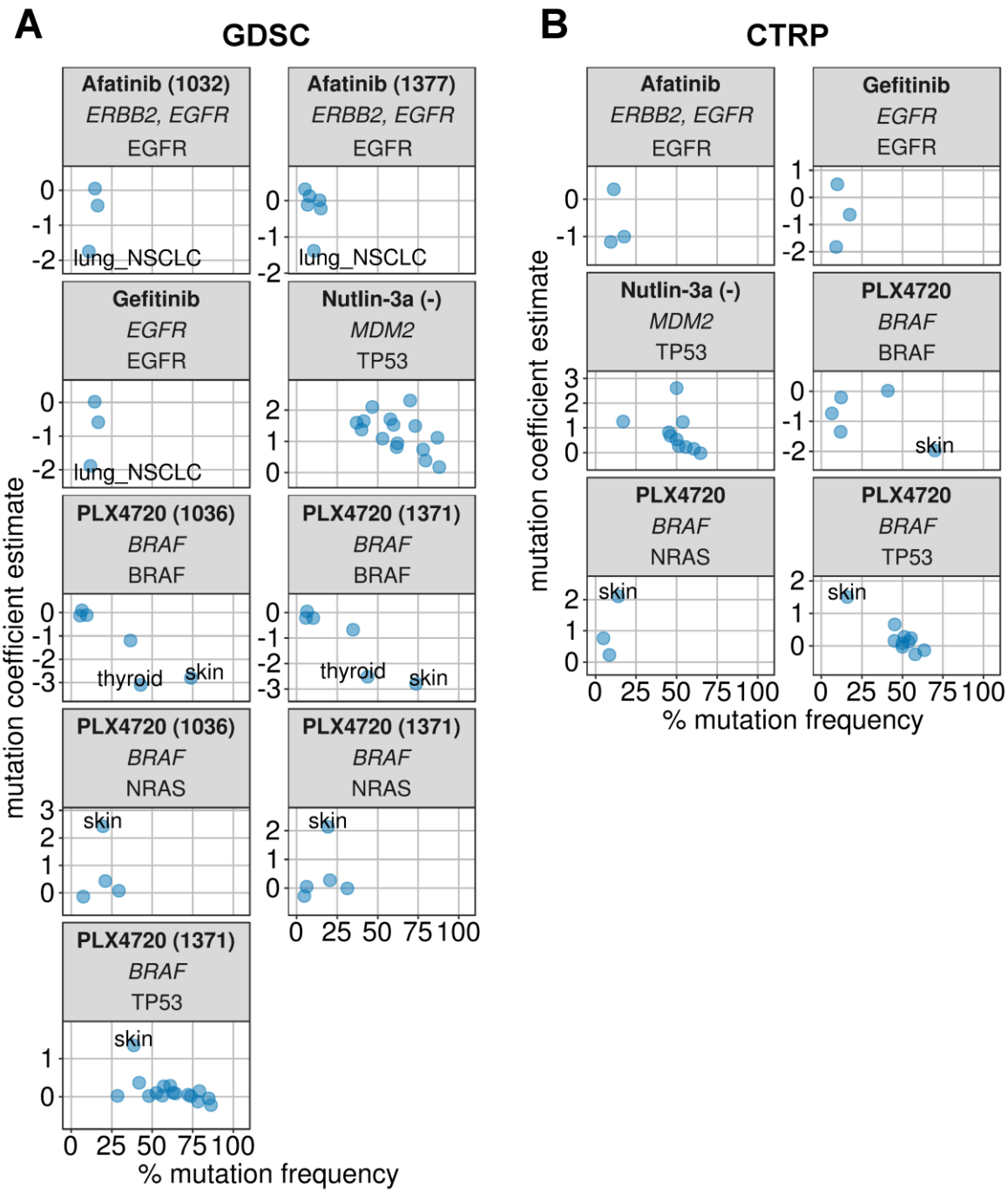

Figure 22: Correlation between tissue-specific mutation coefficient estimates and mutation frequencies in the A) GDSC and B) CTRP dataset. Each point represents one tissue of origin. Tissues with particularly strong mutation associations are depicted. Drug names (bold, with drug IDs in brackets), targets (italic) and mutated genes (plain) are indicated.

3. Combining genetic interactions and tissue specificity

3.1. Drug-specific background models

To summarize our findings that (i) the association between covariates and drug response is variable across drugs (Figure 2 and Figure 3), (ii) single mutations can be strongly associated with drug response (Figure 4A) and (iii) mutation associations can depend on the tissue of origin (Figure 17 and Figure 21), we created new, drug-specific background models. Compared to the default background model, drug-specific background models can exclude covariates and include single mutations as well as mutation-tissue interactions (see Methods).

Depending on the drug, the resulting drug-specific background models contained between 0 and 13 variables. The covariates of the default background model tissue (93%), growth medium (42%), CNA count (27%) and growth properties (26%) were the most frequently included variables, followed by TP53 (5%), NRAS (5%), BRAF (4%) and KRAS mutations (3%). For five drugs, the drug-specific background model did not have any variables, meaning that it corresponds to a null model predicting the mean drug response across cell lines. For another nine drugs, the variables of the drug-specific and the default background model were identical.

We found that drug-specific background models explained up to 60% of the variation in drug response (Figure 23). Compared to the default background model, drug-specific background models performed better for 62% of the drugs (161 drugs) based on the adj. R^2 and for 90% of the drugs (238 drugs) based on the BIC (Figure 23). This implies that drug-specific background models improve predictive performance for most drugs.

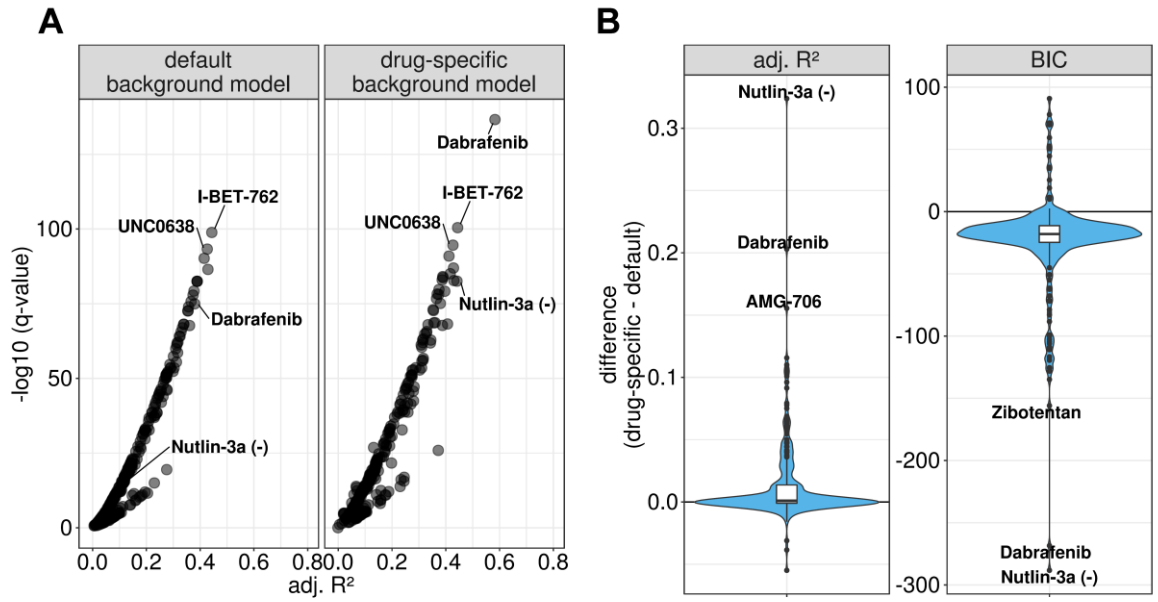

Figure 23: Performance of default and drug-specific background models. A) For each drug, the coefficient of determination (adj. R^2) and the significance (adjusted p-value (Benjamini-Hochberg correction); F-test) of the default and the drug-specific background models are shown. Selected drugs are labeled, the labeled model for UNC0638 (drug screened twice) corresponds to drug ID 245. B) Performance differences between drug-specific and default background models based on adj. R^2 and Bayesian information criterion (BIC). Drugs showing the largest differences are labeled.

To complement the comparisons based on adj. R^2 and BIC, we performed a cross-validation analysis and compared the performance of both background models based on test errors. For 11 drugs, the drug-specific background model outperformed the default background model ($p < 0.05$, Mann-Whitney-Wilcoxon test). The corresponding drug-specific background models were enriched for models containing mutations ($p = 10^{-4}$, chi-square test), suggesting that the inclusion of mutations drives the improvement in model performance. For one drug, the receptor tyrosine kinase inhibitor AMG-706, the drug-specific background model performed worse than the default background model ($p < 0.001$, Mann-Whitney-Wilcoxon test). Since the model for AMG-706 had the highest number of variables (13) across all drug-specific

background models, this could imply model overfitting. No significant differences were found for the remaining drugs ($p \geq 0.05$, Mann-Whitney-Wilcoxon test).

Compared to the default background model, the drug-specific background model for the MDM2 inhibitor Nutlin-3a showed the most significant decrease in test errors ($p < 10^{-15}$, Mann-Whitney-Wilcoxon test). The background models for Nutlin-3a also differed most in terms of adj. R^2 and BIC (Δ adj. $R^2 = 0.32$, Δ BIC = -288; Figure 23). The drug-specific background model for Nutlin-3a contained the covariates tissue and CNA count, the mutation status of TP53 and RB1 and the interaction term between TP53 and tissue (cf. Figure 21). This illustrates the predictive potential of individual mutations and mutation-tissue interactions.

3.2. Identification of mutation-mutation interactions and synthetic lethal triplets

To find examples of mutation-mutation interactions in drug response, we used mutation pair models containing the covariates of the drug-specific background models. To select mutation-mutation interaction examples (Table S1), we filtered for mutation pair models with interaction that outperformed simpler models and showed consistent association patterns in both datasets.

We found that the interaction between BRAF and TP53 mutations is associated with resistance to the BRAF inhibitor PLX4720 (drug ID 1371, $p < 10^{-3}$ (GDSC) and $p = 0.047$ (CTRP), t-test). This confirms our previous finding based on a model including the covariates of the default background model (Figure 13). The drug-specific background model for PLX4720 included the mutation status of BRAF and NRAS as well as the mutation-tissue interaction terms for both genes. This is in line with the tissue specificity that we observed for the association between BRAF or NRAS and PLX4720 response (Figure 21).

In addition, we observed that the interaction between CREBBP and FGFR2 mutations mediates resistance to the cytotoxic drugs Gemcitabine ($p = 0.01$ (GDSC) and 0.02 (CTRP), t-test), Bleomycin (drug ID 1378, $p = 10^{-4}$ (GDSC) and 0.008 (CTRP), t-test) and SN-38 ($p < 10^{-5}$ (GDSC) and $p = 0.002$ (CTRP), t-test; Figure 24). Since both

CREBBP (Dutto et al., 2018) and FGFR2 (Huang et al., 2015) are involved in DNA repair, drug resistance may arise due to increased DNA repair capacity or increased DNA damage tolerance (Cheung-Ong et al., 2013).

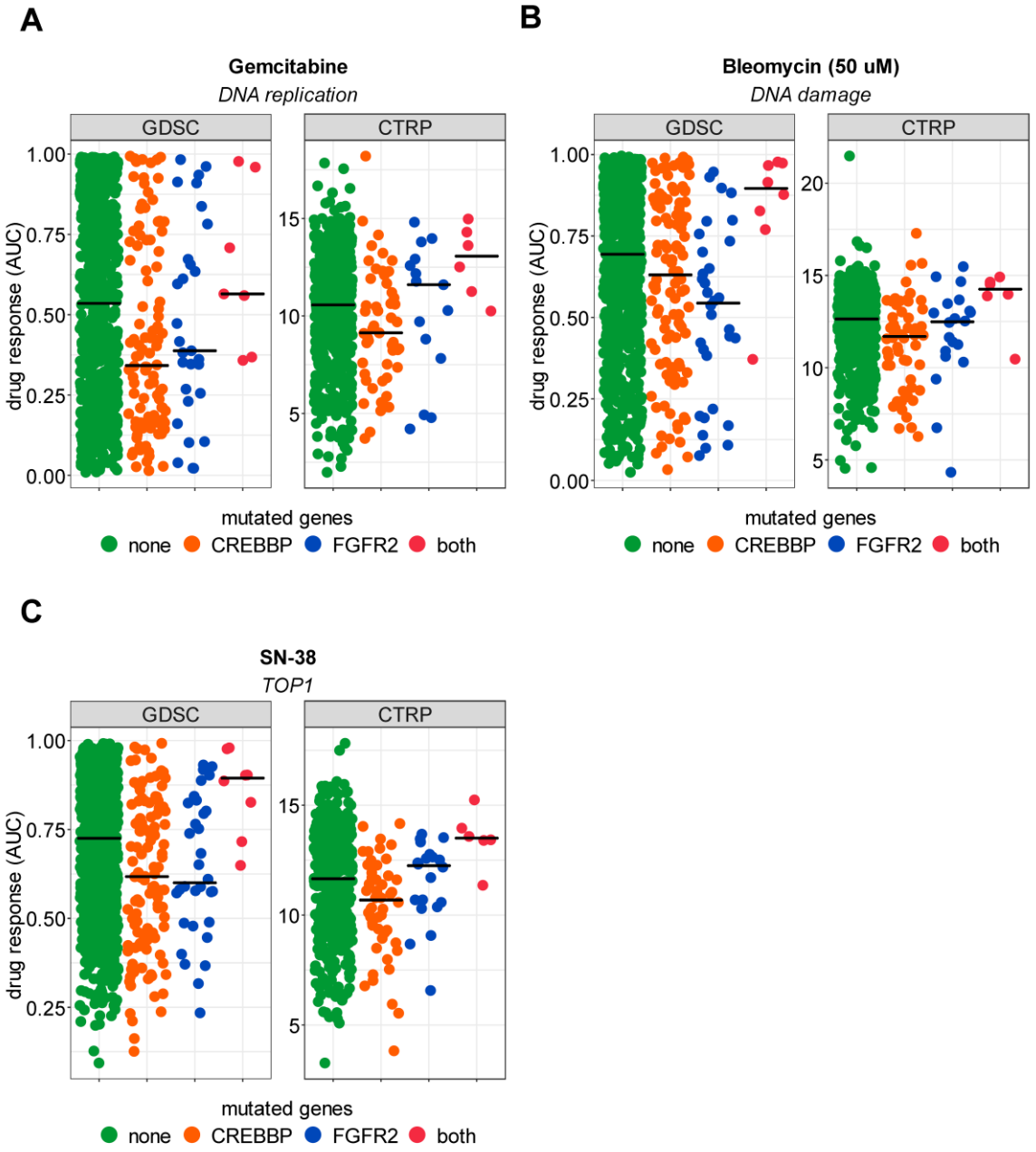

Figure 24: Simultaneous mutation of CREBBP and FGFR2 mediates resistance to several cytotoxic drugs in the GDSC and the CTRP dataset. The response to (A) Gemcitabine, (B) Bleomycin and (C) SN-38 is shown. Drugs (bold) and drug targets (italic) are specified. Cell lines are represented as points and grouped by mutation status of CREBBP and FGFR2. Horizontal lines depict the median drug response for each group.

To identify mutation pairs that jointly mediate drug sensitivity, we used the concept of synthetic lethality. Two genes are termed synthetically lethal if their simultaneous perturbation is lethal although the perturbation of each individual gene is viable (O'Neil et al., 2017). Both genes can be perturbed by molecular alterations or targeted drugs. We extended the concept of synthetic lethality to three perturbations consisting of two mutations and one drug. These sets of three perturbations, which we termed synthetic lethal triplets, can be considered as special cases of mutation-mutation interactions. Accordingly, we applied additional conditions to identify them (see Methods and Table S1).

We found that simultaneous mutation of CTNNB1 and PIK3CA sensitizes cell lines to the MDM2 inhibitor Nutlin-3a (Figure 25A). According to our definition, CTNNB1 and PIK3CA mutations and Nutlin-3a treatment form a synthetic lethal triplet. The sensitizing contribution of the CTNNB1-PIK3CA interaction was consistent between the GDSC and the CTRP dataset ($p = 0.02$ in both datasets, t-test). The drug-specific background model for Nutlin-3a contained the covariates tissue and CNA count, the mutation status of TP53 and RB1 and the interaction term between TP53 and tissue. The TP53-tissue interaction confirms our previous finding regarding the tissue-specific association of TP53 mutations with Nutlin-3a response (cf. Figure 21 and Chapter 3.1). The pro-apoptotic transcription factor FOXO3 may integrate signals from CTNNB1, PIK3CA and Nutlin-3a (Figure 25B). On the one hand, FOXO3 is inhibited by CTNNB1 and PI3K/ AKT signaling (Tenbaum et al., 2012). On the other hand, FOXO3 is degraded through ubiquitination by MDM2, which is the Nutlin-3a target (Fu et al., 2009).

In addition, we found that cell lines with both KRAS and MAP3K4 mutations show increased sensitivity to the DNA synthesis inhibitor Cytarabine ($p = 0.003$ (GDSC) and 0.001 (CTRP), t-test; Figure 25C). The drug-specific background model for Cytarabine contained the covariates growth properties and tissue. On one hand, cell proliferation including DNA synthesis (Zhang and Liu, 2002) is regulated by mitogen-activated protein kinase signaling in which both KRAS and MAP3K4 are involved (MetaCore, Ras family GTPases in kinase cascades; http://pathwaymaps.com/maps/379;Figure 25D). On the other hand, DNA synthesis is inhibited by Cytarabine. These relationships may explain why the three perturbations form a synthetic lethal triplet.

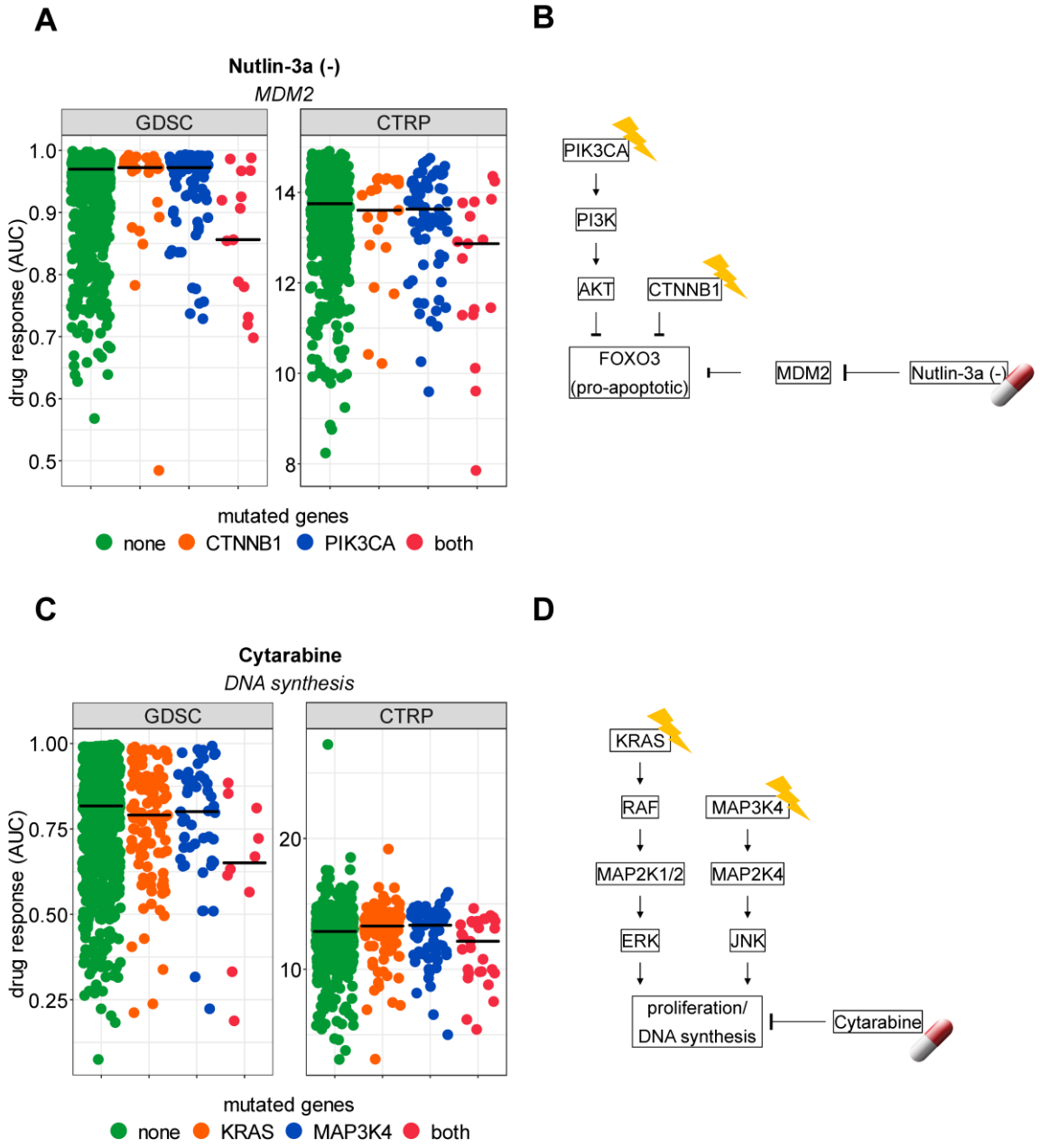

Figure 25: Synthetic lethal triplets between a mutation pair and a drug in the GDSC and the CTRP dataset. A) Simultaneous mutation of CTNNB1 and PIK3CA mediates sensitivity to the MDM2 inhibitor Nutlin-3a. B) The transcription factor FOXO3 may link the synthetic lethal triplet between CTNNB1 and PIK3CA mutations, and Nutlin-3a. C) Simultaneous mutation of KRAS and MAP3K4 mediates sensitivity to the DNA synthesis inhibitor Cytarabine. D) KRAS and MAP3K4 act in parallel pathways to promote DNA synthesis which may explain their synthetic lethal interaction with Cytarabine. Cell lines in A) and C) are grouped by mutation status, horizontal lines indicate the median drug response. Drugs (bold) and drug targets (italic) are depicted.

4. Mutation interaction networks

In Chapter 1, we established that the association of a single alteration with drug response can depend on other alterations in the cancer cell. In principle, the analysis of interactions can be extended beyond alteration pairs and triplets. However, the number of possible combinations greatly increases for larger sets of alterations. Combinatorial complexity and the associated computational cost limit the model size for which an exhaustive search according to our analytical framework is feasible. For example, we already tested several hundreds of thousands of models for combinations of three alterations (Chapter 1.3). Since larger sets of alterations do not allow to fit models for all possible combinations of alterations, a subset of models must be preselected for testing.

We developed a network approach to select alteration subsets for drug response prediction models in a three-step procedure. Since the co-occurrence of different CNAs can be strongly correlated (Chapter 1), we restricted the following analyses to mutations. First, we constructed networks that represent all interactions among cancer driver mutations at once. In these networks, nodes represent mutations and edges represent their interaction with respect to drug response. We termed these networks mutation interaction networks. Second, we used module search algorithms to partition these networks (Blondel et al., 2008; Clauset et al., 2004; Rosvall and Bergstrom, 2008). The resulting modules are defined as network regions with dense connections within a given module and sparse connections between two different modules (Girvan and Newman, 2002). Since network edges represent interactions between mutations, modules represent sets of mutations that strongly interact with respect to drug response. Third, we used linear regression models with interaction to predict drug response based on these sets of mutations. This approach entails several advantages. By partitioning mutation interaction networks into modules of arbitrary size, we increased the maximal number of mutations that can be included in a model beyond mutation pairs and triplets. By using module search algorithms to preselect sets of mutations, we decreased the number of fitted models for a given number of mutations compared to all possible combinations of mutations.

The main idea of our approach is that every set of mutations can be represented by both a network and a linear regression model. Independent effects are represented by

nodes or main effects while pairwise cooperative effects are represented by edges or interaction effects. We distinguish three different network levels, namely edges, networks, and modules (Figure 26). In this context, the smallest possible network is an edge which corresponds to a mutation pair model with interaction.

An edge model predicts drug response based on a mutation pair whose joint effect corresponds to the sum of both node effects and the edge effect. The node or main effects are additive while the edge or interaction effect is non-additive. Two mutations are synergistic if the joint effect of both mutations is larger than the sum of their individual effects, and antagonistic otherwise.

To quantify deviations from additivity in our mutation interaction network, we used weighted edges. Large edge weights represent strong antagonism or synergy between two mutations, which could imply that the corresponding proteins interact physically or participate in the same biological process. By emphasizing edge weights, we aim to facilitate the identification of mutation sets that show strong cooperativity with respect to drug response. Accordingly, modules can be interpreted as functional units that jointly influence drug response.

4.1. Algorithm description

For the initial evaluation of our algorithm, we selected the ten compounds with the highest mean drug sensitivity across cell lines in the GDSC dataset. Our entire algorithm (Figure 26) was embedded in a cross-validation loop (see Methods). In each cross-validation instance, we split the full dataset for a given drug into training and test set. We used the training set for model fitting on the level of input edges and result modules and the test set to compute the prediction error on the level of result modules.

To assemble edge models (Figure 26), which represent the basic building block of our mutation interaction networks, we fitted mutation pair models with interaction for all mutation pairs that co-occur in cancer cell lines with available drug response data (see Methods). For each edge model, the edge weight is based on the mutation-mutation interaction term. To find a statistical measure for the edge weights that is robust to changes in the training data, we computed the coefficient of variation across training sets for different statistics. We compared the coefficient estimate, the t-statistic and the

significance (log-transformed FDR-corrected p-value) of the interaction term. The coefficient estimate and the t-statistic can have positive or negative signs depending on whether the interaction term mediates drug sensitivity or resistance. Since module search algorithms (Blondel et al., 2008; Clauset et al., 2004; Rosvall and Bergstrom, 2008) cannot deal with negative edge weights, absolute values were used. We found that the absolute value of the t-statistic, termed variable importance (Kuhn, 2008), showed less variability than the absolute coefficient estimate and the significance of the interaction term ($p < 10^{-15}$; Mann-Whitney-Wilcoxon test). Therefore, we used the variable importance of the interaction term as the edge weight. When assessing the relation between edge weight and variability, we observed that stronger edges were less variable (Spearman's rho = -0.91; $p < 10^{-15}$), indicating higher robustness.

In the mutation interaction networks, an edge has no meaning by itself–it solely represents a mutation pair for which a model was fitted. Only the edge weight is informative since it defines the strength of the mutation-mutation interaction. However, module search algorithms (Blondel et al., 2008; Clauset et al., 2004; Rosvall and Bergstrom, 2008) consider both edge number and edge weight. For the decision whether a network region is a module or not, many weak edges within this region can have the same positive contribution as few strong edges. To reduce noise originating from high numbers of weak edges, we used different edge thresholds that include only a defined percentage of strongest edges (10 to 100% in 10% steps; Figure 26). The thresholding step removes weak edges from the input network, thereby facilitating module detection.

To assemble mutation interaction networks (Figure 26), we compiled a list of weighted edges based on all mutation pair models. This edge list defines the mutations to be included in the network as nodes and the connections among them.

To select mutation subsets for drug response prediction models (Figure 26), we partitioned our mutation interaction network using a module search algorithm that maximizes the sum of edge weights within modules and minimizes the sum of edge weights between modules (Blondel et al., 2008). To keep mutation numbers as small as possible, the module search algorithm was applied recursively on every identified module until no smaller module could be found. Both the primary modules and the submodules were kept for further analysis. Each module was translated into a regression model by interpreting nodes as main effects and edges as interaction

effects. We fitted the resulting model using the training data and computed the prediction error using the test data.

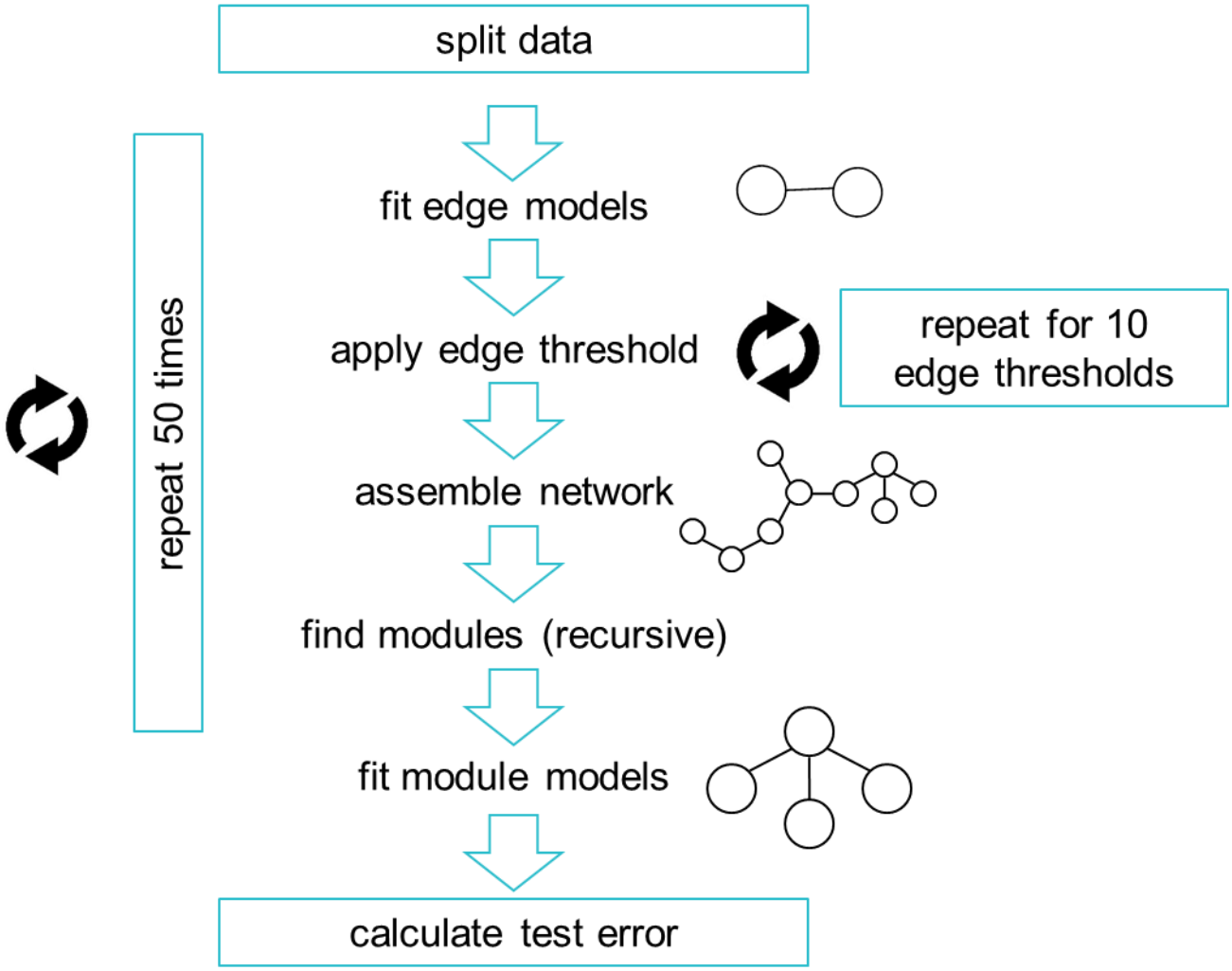

Figure 26: Schematic overview of our network-based drug response prediction algorithm.

4.2. Algorithm improvement

To improve the performance of module-based models, we assessed the effect of different changes to the algorithm. As described above, the covariates tissue, growth medium, growth properties, and CNA count explain large proportions of the variation in drug response (Figure 2 and Figure 23A). To test whether the covariates of the

default background model affect the performance of our method, we included the covariates in edge and module models (Figure 26). Including the covariates significantly improved the test error of the module models ($p < 10^{-15}$, Mann-Whitney-Wilcoxon test).

To test whether including information about the functional impact of mutations improves model performance, we replaced the binary mutation events in the input data by Combined Annotation Dependent Depletion (CADD) scores (Kircher et al., 2014). In contrast to binary mutation data that encodes the presence or absence of a mutation in a given gene, CADD scores encode the deleteriousness of mutations such that different mutations within the same gene can have different scores. Based on the test error of the module models, using binary mutation data resulted in better performance than using CADD scores ($p < 10^{-15}$, Mann-Whitney-Wilcoxon test).

To assess the influence of the module search algorithm on model performance, we compared the algorithms Infomap (Rosvall and Bergstrom, 2008), Fast Greedy (Clauset et al., 2004) and Louvain (Blondel et al., 2008). In contrast to the other two algorithms, the Infomap algorithm can include node weights in addition to edge weights. We defined node weights as the variable importance of the corresponding mutation in a univariate model. The Infomap algorithm runs a random walk on the network, with edge weights determining the node visit frequency and node weights enabling random jumps between nodes. Modules are defined as network regions with a long persistence time of the random walker. However, the Infomap algorithm often did not accomplish to partition the input network and returned the full network instead. The module search algorithms Fast Greedy and Louvain split the network into modules by maximizing the modularity of the partition (see Methods). This implies that the sum of edge weights is maximized within modules and minimized across modules. When running the analysis with Fast Greedy and Louvain, Louvain's method generated module models with lower test errors than Fast Greedy ($p = 0.002$, Mann-Whitney-Wilcoxon test). Thus, we included the covariates of the default background model, used binary mutation data and chose Louvain's method as our module search algorithm.

4.3. Module evaluation

To assess model robustness and performance, we ran our network-based algorithm (Figure 26) for all 265 drugs in the GDSC dataset and retrieved more than 14 million result modules for evaluation. The modularity of the network partition (see Methods) correlated with better performance (Spearman's rho = 0.04; $p < 10^{-15}$). This supports using the modularity as the optimizing function for the module search.

When analyzing the relationship between model performance and module size, we observed a positive correlation between module size and test error (Spearman's rho = 0.27; $p < 10^{-15}$). Based on the median test error across all models with a given number of mutations, seven mutations were determined as the optimal model size (Figure 27).

When analyzing the frequency of result modules across cross-validation instances, we found that smaller modules occurred more frequently than bigger ones (Spearman's rho = -0.31; $p < 10^{-15}$). Together, these results imply that smaller modules tend to show increased model performance and robustness.

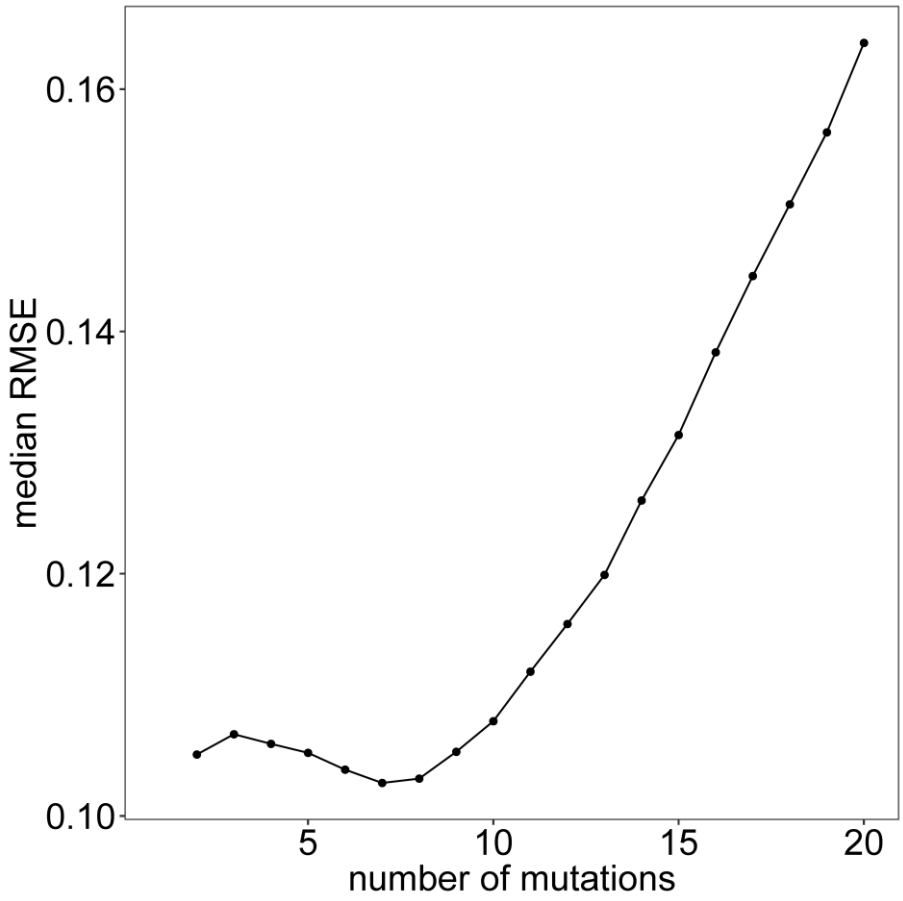

Figure 27: Median test error of all modules with a given number of mutations. For visualization purposes, the horizontal axis is truncated at 20 mutations. For larger modules (up to 192 mutations), the median test error continues to increase. RMSE: root mean squared test error.

To evaluate the performance of module models compared to other model complexities, we retrieved the module with the lowest test error in each cross-validation instance. We first compared the distribution of test errors for the best module and the default background model. For 47% of the drugs, the best module performed significantly better than the default background model ($p < 0.05$, Mann-Whitney-Wilcoxon test). Compared to mutation pair models with interaction, the number of drugs for which the default background model was outperformed decreased by 1% (47% - 48%; cf. Figure 12A).

When comparing mutation module models to mutation pair models with interaction, there was a significant difference in test errors ($p < 0.05$, Mann-Whitney-Wilcoxon test) for only one drug. For the MDM2 inhibitor Nutlin-3a, mutation pair models with

interaction outperformed mutation module models ($p = 0.006$, Mann-Whitney-Wilcoxon test). With the exception of Nutlin-3a, we thus observed comparable performance for mutation module models and mutation pair models.

4.4. Method comparison

To compare the performance of our network method to other methods, we selected the regularized regression approach elastic net (Zou and Hastie, 2005), which is a standard method for drug response prediction (Barretina et al., 2012; Basu et al., 2013; Garnett et al., 2012; Iorio et al., 2016). We compared model performance based on test error distributions by selecting the module with the lowest test error in each cross-validation instance. Our method performed significantly better than elastic net without interaction for 38% of the drugs and better than elastic net with interaction for 42% of the drugs ($p < 0.05$, Mann-Whitney-Wilcoxon test). Elastic net with or without interaction never outperformed our method ($p \geq 0.05$, Mann-Whitney-Wilcoxon test). Our module models were significantly smaller than the models generated by elastic net with or without interaction ($p < 10^{-15}$, Mann-Whitney-Wilcoxon test) which could explain the differences in performance.

4.5. Ensemble models

Since the module models with the lowest test error differ across cross-validation instances, we aimed to select a single ensemble model for each drug. We tried two different approaches to create ensemble models (see Methods). First, we selected the most frequent nodes and edges across cross-validation instances. We iterated over different frequency thresholds and included all variables above a certain threshold. Second, we clustered the result modules that were generated in the cross-validation instances. For each cluster, we retrieved a medoid module that represents the modules in this cluster. For both approaches, we selected the model with the best BIC as the ensemble model.

When comparing the BIC values of the ensemble models generated by both approaches, the frequency-based approach performed better than the cluster-based approach for 91% of the drugs and worse for 8% of the drugs. For the remaining two drugs, the ensemble models generated by both approaches were identical. In both cases, the selected ensemble models were mutation pair models with interaction.

When comparing the number of mutations in the ensemble models, we observed that the frequency-based approach yielded smaller models than the clustering-based approach ($p < 10^{-15}$, Mann-Whitney-Wilcoxon test). Since we observed higher test errors with increasing model size (cf. Chapter 4.3), the differences in model size could explain the better performance of the frequency-based approach.

For eight drugs, we validated the frequency-based ensemble model in the CTRP dataset (see Methods and Table 2). For seven of these drugs, the ensemble model was a single-mutation model or a mutation pair model with or without interaction. For the BRAF inhibitor Dabrafenib, we identified a model consisting of AXIN2, BRAF, CDH1 and RB1 mutations as the ensemble model. Besides the mutations, the model included interactions between BRAF and all other mutations (Figure 28).

Table 2: Validated frequency-based ensemble models.

Drug name (ID)	Drug target	Ensemble model
Gefitinib	EGFR	EGFR
Afatinib (1032)	ERBB2, EGFR	KRAS+MAP3K4
PLX4720 (1036)	BRAF	BRAF*BRCA1
Nutlin-3a (-)	MDM2	RB1*TP53
MK-2206	AKT1, AKT2	HRAS+PIK3CA
PLX4720 (1371)	BRAF	BRAF
Dabrafenib	BRAF	cf. Figure 28
GDC0941 (1527)	PI3K	PIK3CA

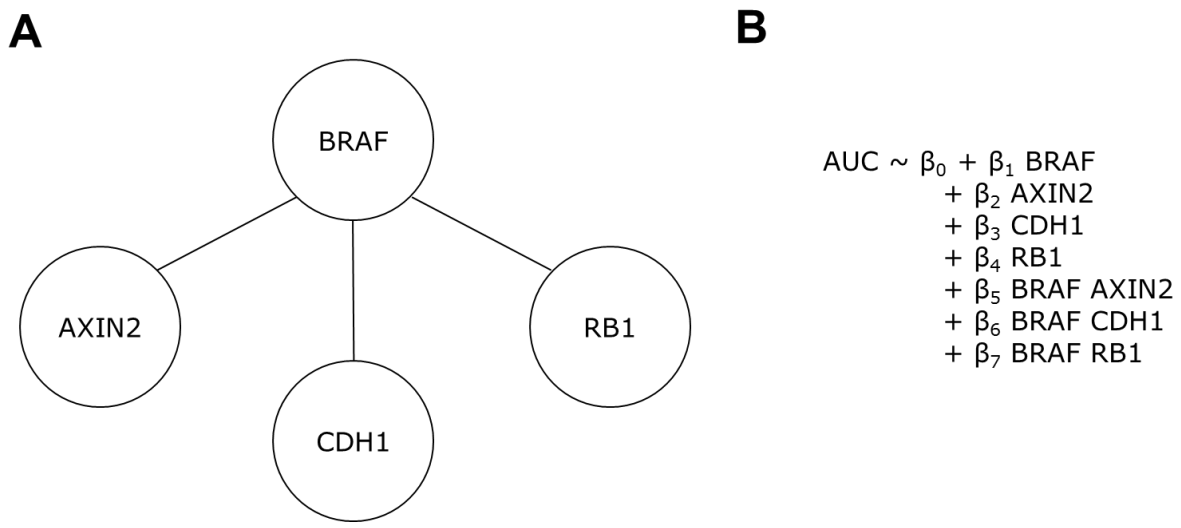

Figure 28: Ensemble model for the BRAF inhibitor Dabrafenib. A) Network representation. B) Model representation. AUC: area under the dose-response curve.

The ensemble model (Figure 28) predicts that BRAF mutations mediate sensitivity to Dabrafenib ($p < 10^{-39}$ (GDSC) and $< 10^{-4}$ (CTRP), t-test) while the interaction between BRAF and AXIN2 ($p = 0.003$ (GDSC) and 0.008 (CTRP), t-test), and BRAF and RB1 ($p = 0.0003$ (GDSC) and $< 10^{-4}$ (CTRP), t-test) mediate Dabrafenib resistance (Figure 29 and Figure 30). This means that cell lines with only BRAF mutations tend to respond while cell lines with BRAF and either AXIN2 or RB1 mutations are resistant to Dabrafenib. Of note, the main effects for AXIN2, CDH1 and RB1 and the BRAF-CDH1 interaction term were not significant ($p \geq 0.05$ in both datasets). Moreover, no cell line in the CTRP dataset contained both CDH1 and BRAF mutations.

Compared to the BRAF*TP53 model that we identified in Chapter 1.3.6.1, the ensemble model for Dabrafenib performed worse based on adj. R^2 and BIC (Δ adj. $R^2 = -0.02$ (GDSC and CTRP), Δ BIC = 50 (GDSC) and 2 (CTRP)). Together with the observation that 7 out of 8 validated ensemble models contained one or two mutations, these results imply that single-mutation or mutation pair models predict drug response best for most drugs.

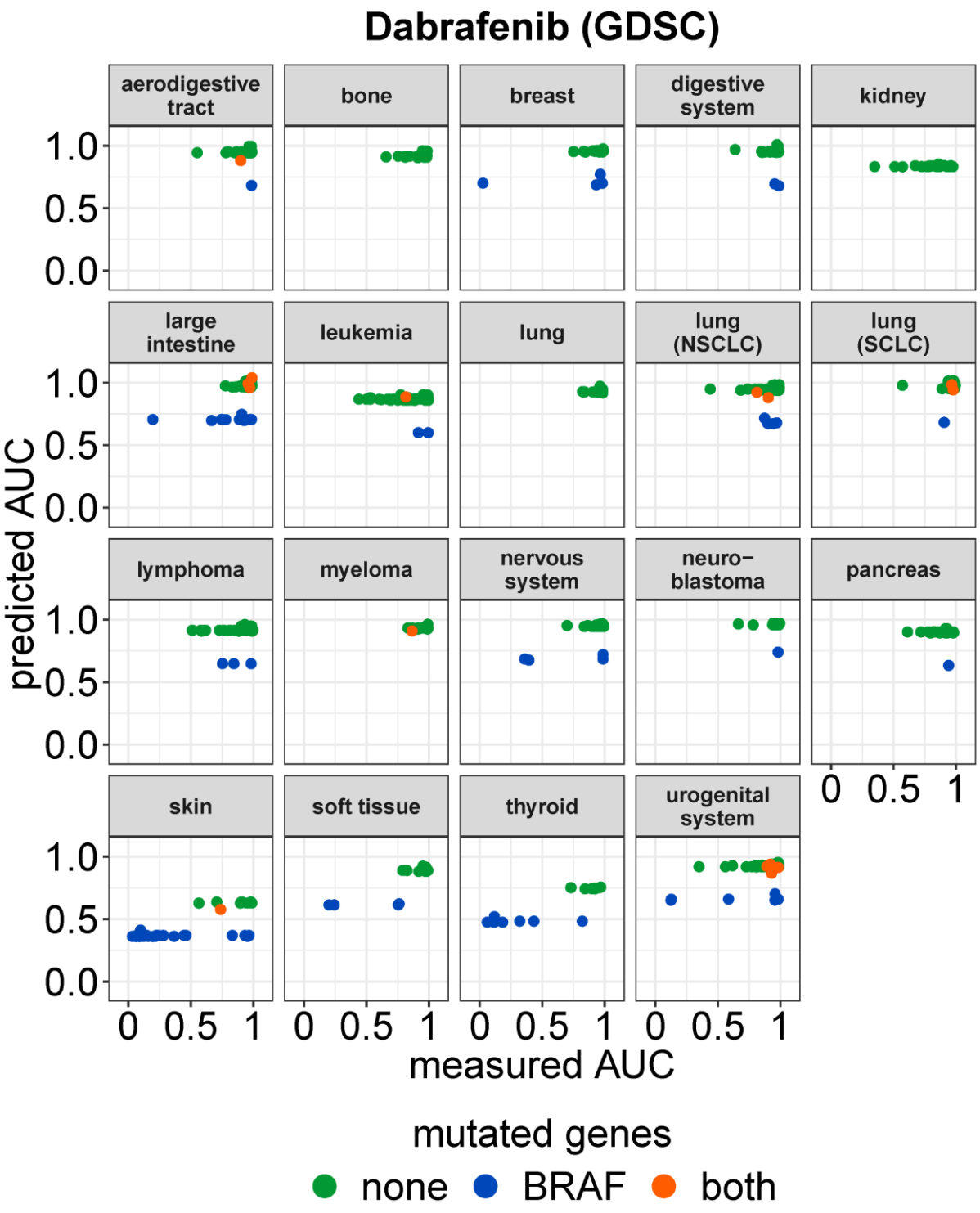

Figure 29: A module consisting of AXIN2, BRAF, CDH1, and RB1 explains response to the BRAF inhibitor Dabrafenib in the GDSC dataset. Cell lines are grouped by tissue of origin and by mutation status of BRAF, AXIN2 and RB1 (none: BRAF, AXIN2, and RB1 wild-type; both: BRAF and AXIN2 or BRAF and RB1 mutated).

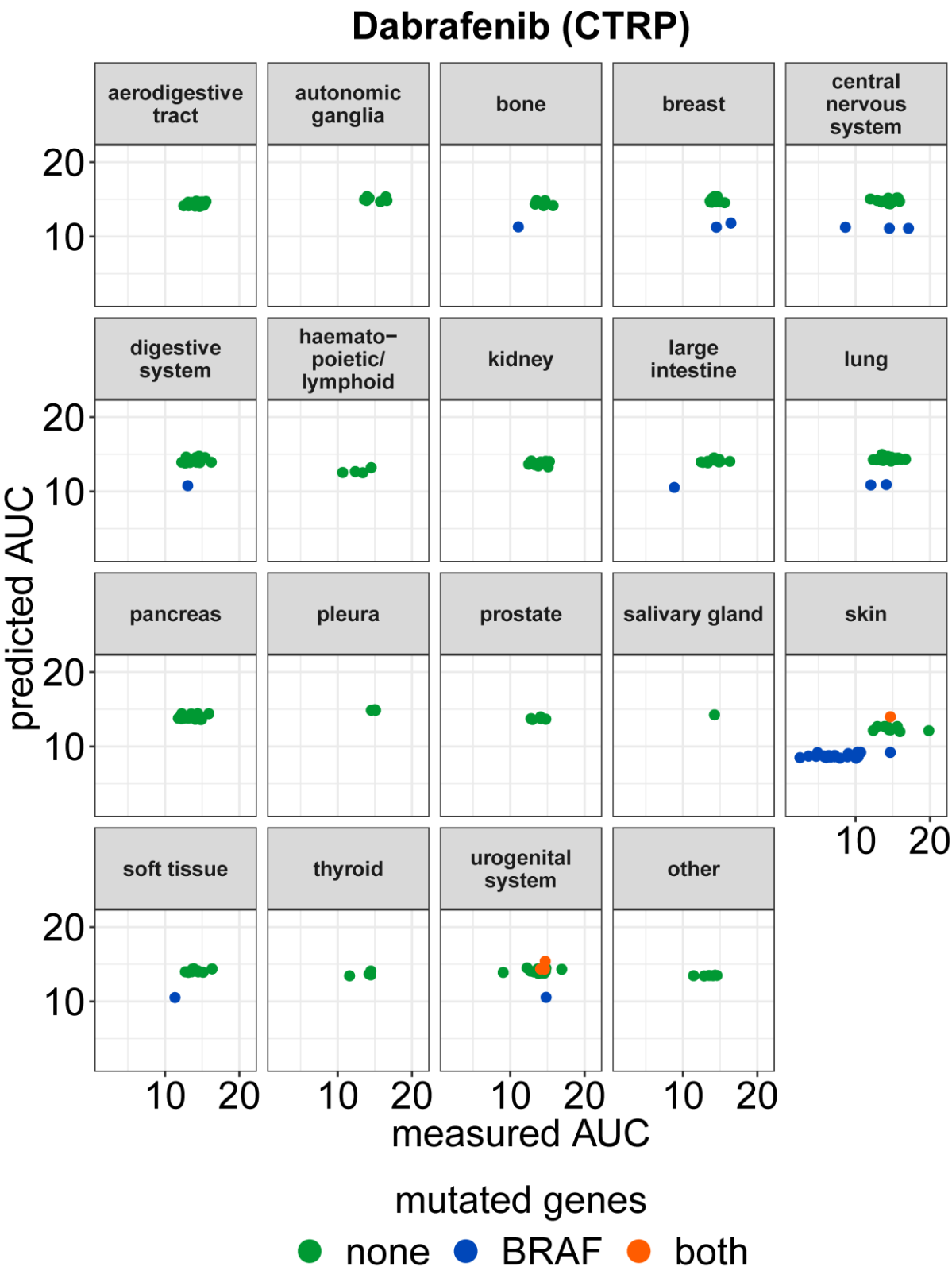

Figure 30: A module consisting of AXIN2, BRAF, CDH1, and RB1 explains response to the BRAF inhibitor Dabrafenib in the CTRP dataset. Cell lines are grouped by tissue of origin and by mutation status of BRAF, AXIN2 and RB1 (none: BRAF, AXIN2, and RB1 wild-type; both: BRAF and AXIN2 or BRAF and RB1 mutated).

Discussion

In this thesis, we studied the role of interactions in drug response (Figure 1). In Chapter 1, we presented an analytical framework for the exhaustive testing of models combining up to three genomic alterations. We found that genetic interactions unlock a high predictive potential. In Chapter 2, we assessed whether mutation associations depend on the tissue of origin. Compared to mutation associations that can be generalized across cancer types, tissue-specific mutation associations showed higher predictive performance. In Chapter 3, we combined the results of the first two Chapters by fitting models that consider both genetic interactions and tissue specificity. We identified candidate mutation pairs that interact with respect to drug response. In Chapter 4, we generated mutation interaction networks that represent interdependencies between all mutations at once. We partitioned these networks into modules and used the corresponding sets of mutations to predict drug response (Figure 26). Overall, models with seven mutations showed optimal predictive performance, but mutation pair models with interaction performed best for many drugs.

We found that considering interactions between genomic alterations improves the performance and robustness of drug response prediction models. Models without interaction assume that individual alteration effects are additive, but biological systems rely on the joint action of molecular components, which can involve synergism or antagonism. We found that synergistic mutation-mutation interactions were more frequent and stronger than antagonistic interactions (Chapter 1.3.1), indicating that neglecting interactions results in an underestimation of the association between two mutations and drug response.

1. Analytical framework

To analyze the association of genomic alterations with drug response, models including up to three alterations with or without interaction were set up. To assess the predictive performance for different model complexities, we fitted models on the entire GDSC dataset and evaluated model performance based on the adj. R^2 and the BIC.

When choosing the best model per complexity for each drug, the models that were selected based on both metrics contained identical variables for 98% of the drugs (Chapter 1.3.2). In contrast to the high concordance of best models within complexities, the evaluation of best models across complexities was rather discordant (Figure 10). Although the BIC penalizes the number of parameters in the model more heavily, both metrics agree that models with interaction account for larger proportions of best models per drug than models without interaction.

To test whether additive or non-additive effects dominate model performance, we assessed whether the best models per complexity were nested, meaning that more complex models include the variables of less complex models. For most drugs, the best model with interaction did not include the variables of the best model without interaction (Figure 7), which implies that interactions strongly contribute to model performance. Since the best model with interaction can rarely be foreseen knowing the variables of the best model without interaction, our interaction-based approach identifies predictive combinations of variables that additive models miss.

To control for confounding factors, we summarized important covariates as background models. Although previous models also included covariates (Iorio et al., 2016), our study goes beyond other approaches and systematically assesses the contribution of a large panel of potential confounders (Chapter 1.1). We defined models with genomic alterations as useful models if they outperformed the default background model that includes the covariates tissue, growth properties, growth medium, and CNA count. Both the adj. R^2 and the BIC agree that a useful model with up to three alterations can be identified for most drugs (85% or 100%; Chapter 1.3.2.2). When comparing models with alterations to the default background model, a model comparison test ($q < 0.1$, Benjamini-Hochberg correction, F-test) is more stringent than comparing the BIC values of both models without requiring a specific BIC difference. This explains why the percentage of drugs outperforming the default background model is lower (cf. Figure 8 and Figure 9).

To evaluate the relevance of interactions, we compared the number of useful models with and without interaction. Based on model comparison tests (Figure 8) and test errors (Figure 12), including interactions lead to a consistent increase in the number of drugs for which a useful model could be found. This implies that interactions unlock a predictive potential that additive models cannot capture.

The number of drugs that benefit from adding an interaction term to an alteration pair model (between 19 and 26%; Figure 12) resembled the number of drugs that benefit from adding a second alteration to a single-alteration model (between 17% and 22%; Figure 12). Similarly, we observed that the best model with two alterations and interaction can predict drug response better than the best model with three alterations, but without interaction (Figure 6). Together, these findings imply that including an interaction has a similar or even better potential to improve drug response prediction models as adding an alteration.

Adding a second alteration to a single-alteration model results in a large decrease in model robustness (Figure 11). In contrast, adding an interaction term to an alteration pair model increases model robustness. In summary, our results show that interactions positively contribute to model performance and robustness.

By defining interactions based on linear regression models, the approach we adopted is similar to a previous study (Jiang et al., 2018). However, Jiang and colleagues focused on alteration pairs where one alteration affects the drug target, which restricts the analysis to targeted drugs. In contrast, we identified predictive markers for both targeted drugs and cytotoxic chemotherapeutics. For instance, our findings suggest that the interaction between CREBBP and FGFR2 mutations consistently mediates resistance to several cytotoxic drugs (Figure 24).

Moreover, including the alteration status of the target gene as a default predictive variable (Jiang et al., 2018) presumes a prominent predictive influence. However, we found that alterations of the drug target do not always show the strongest association with drug response in univariate models (Figure 4 and Table 1). For instance, the association between TP53 mutations and resistance to the MDM2 inhibitor Nutlin-3a is the strongest mutation association across drugs. Nevertheless, TP53 is not directly targeted, but only indirectly affected since MDM2 mediates its ubiquitin-dependent degradation (Kojima et al., 2006).

Since the method proposed by Jiang and colleagues predicts responders and non-responders by computing the correlation with a genome-wide signature (Jiang et al., 2018), it is difficult to trace back which marker genes drove the prediction. In contrast, our models contain far fewer alterations, which facilitates their interpretability.

2. Network method

To integrate the information about pairwise mutation-mutation interactions, we constructed mutation interaction networks. In these networks, nodes represent mutations while edges represent mutation pairs. Edges are weighted by the non-additive effect of the mutation pair on drug response. Our approach is conceptually similar to a network method that uses expression data to define interactions with respect to disease (Watkinson et al., 2008). While our method weights edges by the variable importance in a linear regression model (Kuhn, 2008), Watkinson and colleagues employ the information-theoretic measure of synergy.

When applying our network method, we identified more than 14 million modules and used the corresponding sets of mutations to predict drug response. By restricting model fitting to module-based mutation sets instead of all possible combinations of mutations, the model space is decreased. Thus, module search algorithms replace a model selection technique.

When partitioning mutation interaction networks into modules, each partition has an assigned modularity value that is computed based on the sum of weighted edges within and across modules. The modularity is the optimizing function of the module search algorithm we used (Blondel et al., 2008). We found a weak correlation between the modularity of the partition and the performance of the resulting module models (Chapter 4.3). Since the modularity is computed based on edge weights, this observation highlights the importance of the edge weight definition. Choosing an alternative edge weight measure that results in a stronger correlation between modularity and performance could improve the overall performance of our method. Since the module search algorithm used in the present study cannot consider node weights, future studies could evaluate edge weight measures that combine the information about the variable importance of nodes and edges. Alternatively, algorithms that consider node and edge weights, but optimize a function other than modularity, could be assessed (Dittrich et al., 2008; Lecca and Re, 2015). However, the Infomap algorithm, which belongs to this class, did not yield satisfactory results (Chapter 4.2).

When linking the number of mutations in a module model to predictive performance, we found that fewer mutations were associated with better performance (Chapter 4.3). We identified seven mutations as the optimal number of mutations in a model (Figure 27). However, the direct comparison between module models and mutation pair models with interaction showed significant differences for only one drug, the MDM2 inhibitor Nutlin-3a, for which mutation pair models with interaction outperformed module models. Similarly, the assessment of ensemble models that are based on the most frequently included alterations across cross-validation instances suggests that models with up to two mutations often achieve the best performance (Chapter 4.5 and Table 2). In summary, the model sizes we identified as optimal are by far smaller than models generated by conventional computational methods. In accordance with our results, a previous study modeling drug response based on gene expression data determined 6 to 12 genes as the optimal model size (Dong et al., 2015).

A previous study highlighted elastic net, a standard method for drug response prediction (Barretina et al., 2012; Basu et al., 2013; Garnett et al., 2012; Iorio et al., 2016), as one of the best-performing algorithms (Jang et al., 2014). In comparison, our network method showed better performance for 42% of the drugs and comparable performance for the remaining drugs (Chapter 4.4). Notably, the elastic net implementation we used allows to include interaction effects without the corresponding main effects, which violates variable hierarchies. In this model setting, an interaction term represents the association between two co-occurring mutations and drug response, but can no longer be interpreted as a non-additive contribution. Future studies could implement elastic net with a hierarchy constraint as it was proposed for lasso regression (Bien et al., 2013).

3. Candidate genetic interactions

To select candidate interactions for a given drug, we compared the performance of alteration pair models with and without interaction based on test errors. In each cross-validation instance, 80% of the data were used for model training while the remaining 20% of the data were held out for model testing. We extracted the best test error in each cross-validation instance and used a statistical test ($p < 0.05$, Mann-

Whitney-Wilcoxon test) to compare the test errors of alteration pair models with and without interaction.

Our procedure to select candidate interactions is very stringent for the following reasons. First, training models based on 80% of the available cell lines decreases the number of detectable associations, which strongly depends on the number of training cell lines (Iorio et al., 2016). Second, since we compared distributions of only 50 test errors that correspond to 50 cross-validation instances, the statistical test we applied sets a high hurdle for an improvement by interactions. Increasing the number of cross-validation instances would increase the statistical power, but goes beyond the scope of this thesis. Third, the covariates of the default background model account for large proportions of the variation in drug response. Therefore, alteration pair models with and without interaction can have very similar test errors, which impedes the detection of statistically significant performance differences. In summary, the stringent procedure we chose filters for high-confidence interactions.

For example, we found that the interaction between BRAF and TP53 mutations mediates resistance to several BRAF inhibitors. Gain-of-function BRAF mutations and loss-of-function TP53 mutations were reported to interact in promoting tumorigenesis (Yu et al., 2009). Here, we found evidence for the relevance of the BRAF-TP53 interaction in drug response. We show that the simultaneous mutation of BRAF and TP53 mediates resistance to BRAF inhibitors, whereas cell lines with only BRAF mutations tend to respond (Figure 13). We observed consistent association patterns for the BRAF inhibitors Dabrafenib and PLX4720 in the GDSC and the CTRP dataset. In accordance with our results based on *in vitro* data, we found that melanoma patients with only BRAF mutations tend to respond better to the BRAF inhibitor Vemurafenib than patients with BRAF and TP53 mutations (Chapter 1.3.6.1).

In line with our results, two experimental studies showed that TP53 activation can potentiate the effect of BRAF inhibitors. One study showed that the microRNA-3151 links BRAF and TP53 signaling (Lankenau et al., 2015). As a downstream effector of BRAF, the microRNA-3151 downregulates TP53 expression and nuclear localization. The overexpression of microRNA-3151 in Vemurafenib-resistant skin and thyroid cancer cell lines points towards combining BRAF and microRNA-3151 inhibition. According to another study, resistance to the BRAF inhibitor Vemurafenib can be overcome by a combination therapy with the TP53 reactivator PRIMA-1^{Met} in

melanoma cell lines and patient-derived xenograft models (Krayem et al., 2016). PRIMA-1^{Met} completed a phase I clinical trial (Lehmann et al., 2012) and a combination trial with the BRAF inhibitor Dabrafenib is ongoing (https://www.clinicaltrials.gov/ct2/show/NCT03391050). Accordingly, the BRAF-TP53 interaction we identified represents a promising target for combination therapies, which underlines the utility of our modeling approach.

As an example of a mutation-CNA interaction, we found that the interaction between mutation of BRAF and amplification of the genomic region 3p14.1 is associated with sensitivity to the BRAF inhibitor SB590885 (Figure 14). However, we could not validate our finding in the CTRP dataset because SB590885 was not screened. Since amplification of 3p14.1 in the GDSC cancer cell line panel is rare, very few cell lines have only the amplification or both the amplification and a BRAF mutation. Future studies could increase confidence in the association we identified by downsampling the group of wild-type and BRAF-mutated cells to match these small sample numbers and repeatedly fitting the same model to balanced data subsets.

Previous research identified the master lineage transcription factor MITF as the target of the 3p14.1 amplification (Garraway et al., 2005). The authors showed that MITF overexpression and BRAF mutation cooperate in tumorigenesis, which supports the interaction between both alterations that we identified in drug response. However, MITF amplification was discussed as a clinical resistance mechanism in BRAF-mutated melanoma (Van Allen et al., 2014) which contradicts our results. Intriguingly, both high and low expression of MITF have been associated with BRAF inhibitor resistance (Müller et al., 2014), suggesting that MITF plays a complex role in determining BRAF inhibitor responsiveness.

For the BRAF inhibitor PLX4720, we identified a sensitivity interaction between BRAF mutations and amplification of 3q26.1 in the GDSC and the CTRP dataset (Figure 15). This interaction was not identified in a previous study (Iorio et al., 2016). Although amplification of 3q26.1 is frequent in melanoma cell lines (Tanami et al., 2004), the co-occurrence of BRAF mutations and 3q26.1 amplifications was not restricted to skin cancer cell lines in the datasets we used. In contrast to Tanami and colleagues, we did not observe increased frequencies of the 3q26.1 amplification in BRAF-mutated cell lines ($p = 0.57$ (GDSC) and 0.76 (CTRP), chi-square test)

Using our network method, we found that the interaction between BRAF and either AXIN2 or RB1 mutations confers resistance to Dabrafenib in the GDSC and the CTRP dataset (Figure 29 and Figure 30). Follow-up studies could combine the sensitivity associations (BRAF mutations alone or in combination with either 3q26.1 or 3p14.1 amplifications) and the resistance associations (BRAF mutations with either TP53, AXIN2 or RB1 mutations) that we identified for different BRAF inhibitors into a single drug response prediction model.

We used drug-specific background models as a performance reference to identify additional mutation-mutation interaction candidates. Since drug-specific background models can include single mutations and mutation-tissue interactions, their performance is usually higher compared to the default background model (Figure 23). We required that the final mutation pair model outperformed the drug-specific background model with either of the mutations ($p < 0.05$, F-test), which results in a high hurdle for detecting an improvement by interactions.

Based on this selection criterion, we identified several candidate interactions between two mutations that jointly mediate drug sensitivity or resistance (Chapter 3.2 and Table S1). We observed that the corresponding proteins often act in parallel signaling pathways upstream of the drug target. Future investigations could build on our observations and analyze pathway topologies systematically as it was done in a study on drug combinations (Menden et al., 2017).

We identified synthetic lethal triplets between two mutations and a drug, for instance CTNNB1 and PIK3CA mutations that sensitize cancer cell lines to the MDM2 inhibitor Nutlin-3a (Figure 25A and B). The interaction between CTNNB1 and PIK3CA has already been discussed in the context of tumorigenesis (Riemer et al., 2017), but not drug response. Since synthetic lethality databases usually list synthetic lethal gene pairs (Jerby-Arnon et al., 2014; Li et al., 2014), they do not allow to validate the synthetic lethal triplets we found. Instead, we identified candidate synthetic lethal triplets in the GDSC dataset and validated them in the CTRP dataset.

A previous study (Liu et al., 2016) aimed to identify alteration pairs consisting of one alteration that sensitizes to a drug and a second alteration that reverts this effect. If perturbation of the first gene is lethal whereas the simultaneous perturbation of both genes is viable, such a relationship is termed synthetic viability (Motter et al., 2008).

Since targeting the second, resistance-mediating alteration by a second drug could have a synergistic effect, synthetic viable relationships represent potential opportunities for combination therapies. Although identifying synthetic viable alteration pairs was not the primary objective of this work, the BRAF-TP53 interaction we found to be associated with BRAF inhibitor resistance is in fact a synthetic viable relationship. Likewise, the BRAF-AXIN2 and the BRAF-RB1 interaction are synthetic viable. In all three examples, the BRAF mutation mediates BRAF inhibitor sensitivity, while the second mutation, which suppresses this effect, could potentially be targeted by a second drug. This shows that our analytical framework can be used to generate hypotheses for two-drug combinations.

4. Tissue specificity

In previous studies, tissue specificity was analyzed in models based on cell lines from individual tissues of origin (Iorio et al., 2016). However, the tissue specificity of drug response in cancer cell lines is controversial (Garnett et al., 2012; Iorio et al., 2016; Jaeger et al., 2015). Therefore, we modeled drug response in a pan-cancer setting using the tissue of origin as a covariate. By considering mutation-tissue interactions in pan-cancer models, our analytical framework allows to distinguish tissue-specific mutation associations from associations that are generalizable across tissues. As an additional advantage, regression coefficients can be estimated based on all cell lines.

We observed that models with tissue-specific mutation associations explain drug response better than models with general mutation associations (Figure 18). This is in line with previous findings (Iorio et al., 2016) and underlines the relevance of mutation-tissue interactions.

The candidates we identified as tissue-specific mutation associations show consistency across identical and similar drugs. For example, the tissue-specific association between EGFR mutations and EGFR inhibitors was consistent for two Afatinib screens and one Gefitinib screen (Figure 21). Both drugs are approved for the treatment of EGFR-mutated non-small cell lung cancer (Kazandjian et al., 2016; Wecker and Waller, 2018), the cancer type for which we observed the strongest association between EGFR mutations and drug sensitivity.

We found that some mutations mediate drug sensitivity and resistance in different tissues (Figure 19 and Figure 20). Since the sign of the mutation coefficient determines the direction of the effect, we termed these associations sign-switching associations. Due to the limited number of cell lines with a specific mutation in different tissues of origin, we could not validate any of the sign-switching mutation associations in the CTRP dataset. Nevertheless, sign-switching phenomena were previously observed in cancer patients with BRAF-mutated tumors receiving BRAF inhibitor treatment. While melanoma patients showed high response rates (60-80%; Flaherty et al., 2010; Kefford et al., 2010), only about 5% of colorectal cancer patients responded (Kopetz et al., 2010). Mechanistic studies revealed that BRAF inhibition activates EGFR via a feedback loop, thereby promoting continued proliferation in presence of a BRAF inhibitor (Corcoran et al., 2012; Prahallad et al., 2012). In line with the good response rates of melanoma patients, this feedback activation does not affect melanoma cells since they express lower levels of EGFR.

5. Strengths and limitations

A review of computational methods for drug response prediction highlighted the need for a thorough evaluation and validation of model performance (Azuaje, 2017). In line with these recommendations, our models were evaluated by different performance metrics and examined by both cross-validation and an independent validation dataset. Additionally, we used a background model as a lower performance boundary. The background model contains important covariates, of which the tissue of origin explains the highest variation in drug response. Although gene expression data is most predictive of drug response (Costello et al., 2014; Iorio et al., 2016; Jang et al., 2014), we did not consider gene expression in our models because its predictive power correlates with the tissue of origin. By including the tissue of origin as a covariate, we keep our models small enough to be readily comprehensible.

In addition to the lower performance boundary, which is represented by the background models, we defined the duplicate drug response as an upper performance boundary. To estimate the agreement of duplicate drug screens, we created duplicate models that predict drug response based on the drug response values of the

corresponding duplicate drug screen. Surprisingly, duplicate models were outperformed by a null model or the default background model (Chapter 1.3.7) for 11 out of 28 drugs, which raises concerns about data quality. In accordance with our results, the inconsistency of duplicate screens has been criticized (Safikhani et al., 2016). A possible explanation is that most cell lines are resistant to most drugs which results in a low variability of drug response values such that biological signals can be confounded by measurement noise (Geeleher et al., 2016). To cope with the data quality issues, we believe that a stringent selection of candidate associations and their validation in a second cancer cell line panel is needed. Although many cell lines in the GDSC and the CTRP dataset are identical, the experimental procedure and the readout for cell viability are different (Iorio et al., 2016; Seashore-Ludlow et al., 2015). Therefore, we consider the datasets as independent.

The majority of candidate models we identified explained response to BRAF inhibitors. This shows that the predictive performance strongly depends on the drug that is screened. We hypothesize that the high number of sensitive cell lines and the specificity of the BRAF inhibition facilitate the prediction of BRAF inhibitor response.

Similar to the influence of the screened drug, our results may also depend on the choice of genomic alterations. By using a preselected set of mutations and CNAs (Iorio et al., 2016), we aim to enrich for alterations that are recurrent and functional in cancer. This implies the assumption that the genes involved in cancer development and progression correspond to the set of genes that determine drug response.

We focused our analyses in Chapter 2, 3 and 4 on mutations since CNAs are less frequent (Figure 5) and since the occurrence of two CNAs can be strongly correlated. When evaluating the performance and robustness of models including different alteration types, mutation and mutation-CNA pair models with interaction were more robust than CNA pair models with interaction (Figure 11) and were never outperformed by any other model complexity (Figure 12). Therefore, mutation-CNA pair models could be a promising alternative to mutation pair models. However, interpreting the functional consequences of CNAs is more challenging since the amplified or deleted regions often span several genes, only some of which may undergo positive selection during tumorigenesis (Cramer et al., 2016).

6. Conclusions and outlook

In this thesis, we developed an interaction-centered approach to predict drug response based on genomic features. The strength of our approach lies in the exhaustive model search we conducted, resulting in millions of tested models (Chapter 1.3 and 4.3) and exceeding the number of models that were assessed in previous systematic studies (Jang et al., 2014). We show that interactions can unlock a considerable predictive potential. Since drug response prediction models with interaction show improved performance and robustness, our study strongly encourages the use of interactions. Our findings suggest that existing models can be refined by testing whether alteration effects depend on the alteration status of other genes or the tissue of origin. By considering genetic interactions and tissue specificity, this work contributes to a holistic view on the determining factors of drug response.

Novel drug response prediction models should not only aim at improving predictive performance but also to ensure model interpretability (Azuaje, 2017; Knijnenburg et al., 2016). Our study complements previous attempts to create small, easily interpretable drug response prediction models. Unlike conventional computational methods that output ranked lists of important features, our method generates amenable hypotheses that can guide experimental studies. Integrating prior knowledge about signaling pathways, protein-protein interactions and chemical drug properties may further enhance the amenability of drug response prediction models in the future.

Both the analytical framework and the network-based approach presented here can be used for the analysis of other drug screens, but also viability screens based on shRNA and CRISPR/Cas9 (Aguirre et al., 2016; McDonald et al., 2017; Tsherniak et al., 2017). To validate our candidate interactions and to find mechanistic explanations for strong non-additive effects, follow-up studies should focus on complementing the theoretical work presented here by experiments. Since statistical interactions can reflect biological interactions, associations with drug response that involve interactions can help to understand a drug’s mode of action.

Methods

Statistical analyses

We used R version 3.4.2 to conduct statistical analyses. All statistical tests were two-tailed.

GDSC data

Drug response data was retrieved from The Genomics of Drug Sensitivity in Cancer Project (GDSC; Yang et al., 2013; release 6, version 17; v17_fitted_dose_response.xlsx). The dataset comprises 1001 cancer cell lines and 265 drugs. Most of the drugs (n = 242) target specific biological processes or pathways, some of them (n = 9) are cytotoxic chemotherapeutics and the remaining drugs (n = 14) lack a defined mode of action. Since 14 out of the 265 drugs were screened twice, we indicate the unique drug ID where required.

In contrast to previous studies (Garnett et al., 2012; Iorio et al., 2016; Knijnenburg et al., 2016), we summarized drug response by the area under the dose-response curve (AUC) instead of the half-maximal inhibitory concentration (IC_{50}). The reason for this is that the AUC is more reliable although the IC_{50} is most commonly used (Huang and Pang, 2012). Additionally, estimating the IC_{50} in large drug screens can be imprecise since many cell lines do not achieve an inhibition of 50% (Bouhaddou et al., 2016; Haverty et al., 2016). In the GDSC dataset, the AUC takes values between 0 and 1 where 0 represents drug sensitivity and 1 represents drug resistance. When fitting drug response prediction models, we transformed the AUC values into z-scores.

To test for potential confounders of drug response, we retrieved cell line annotation data (Yang et al., 2013; release 6; Cell_Lines_Details.xlsx). We extracted tissue (“GDSC Tissue descriptor 1”), growth medium (“Screen Medium”), growth properties (“Growth properties”) and microsatellite instability status (“Microsatellite instability Status (MSI)”). We used mutation data from the Catalogue Of Somatic Mutations In

Cancer (COSMIC; Forbes et al., 2017; version 80; GRCh37/hg19; CosmicCLP_MutantExport.tsv) to count the total number of mutations (mutation IDs), including silent mutations, per cell line. CNA counts (CNV IDs) per cell line were computed using CNA data from COSMIC (Forbes et al., 2017; version 80; GRCh37/hg19; CosmicCLP_CompleteCNA.tsv), considering only CNAs with known minor allele and total copy number. In addition, we retrieved the proportion of 30 mutational signatures from the literature (Jarvis et al., 2018; Supplementary Table 2-COSMIC signatures.xls). Signature 27 did not have any non-zero values.

To investigate the association between mutations and drug response, we retrieved a list of 267 potential cancer driver genes (Iorio et al., 2016). For each cell line, we extracted the mutation status for these genes from COSMIC (Forbes et al., 2017; version 80; GRCh37/hg19; CosmicCLP_MutantExport.tsv) while excluding silent mutations. Compared to Iorio and colleagues (Iorio et al., 2016), we used a more recent data version that includes a genome reduction step. Therefore, only 248 out of 267 genes were mutated in any of the cell lines.

To investigate the association between CNAs and drug response, we used a list of 425 recurrently altered chromosomal segments (RACS) as defined by Iorio and colleagues (PANCAN_CNA_BEM.rdata.txt, Iorio et al., 2016). The location of genomic regions is indicated as chromosome, short (“p”) or long (“q”) arm, and cytogenic band, sub-band and sub-sub-band (Strachan and Read, 1999).

CTRP data

We used the data from the Cancer Therapeutic Response Portal (CTRP) as a validation dataset. Drug response data was retrieved from the publication (Seashore-Ludlow et al., 2015). In total, 76 drugs were screened in both the GDSC (Iorio et al., 2016) and the CTRP. Of those, 10 drugs were screened in duplicates in the GDSC project.

In the CTRP dataset, AUC values can take values between 0 and 20. Like in the GDSC dataset, low AUC values correspond to sensitive cell lines while high values

correspond to resistant cell lines. As described above, AUC values were transformed into z-scores to fit drug response prediction models.

The covariates tissue (“ccle_primary_site”) and growth medium (“culture_media”) were retrieved from the CTRP publication (Seashore-Ludlow et al., 2015). We did not include the growth properties covariate since all cell lines in the CTRP dataset are adherent. To increase the consistency of covariates between the GDSC and the CTRP dataset, we summarized “upper_aerodigestive_tract” and “oesophagus” as “aero_dig_tract”, “endometrium”, “ovary” and “urinary_tract” as “urogenital_system”, and “biliary_tract”, “liver” and “stomach” as “digestive_system”. Likewise, we reduced the number of medium categories by summarizing media with the prefix “DMEM0” as “DMEM” and media with the prefix “RPMI0” as “RPMI”. We grouped all remaining media into the category “other”. We computed CNA counts for each cancer cell line based on binary copy number calls for amplifications and deletions (CCLE_MUT_CNA_AMP_DEL_binary_Revealer.gct) that we retrieved from the CCLE website (portals.broadinstitute.org/ccle/data; Kim et al., 2016).

To investigate the association between genomic alterations and drug response, we used the same set of 248 mutations and 425 RACS as for the GDSC dataset (see above). We retrieved mutation data from the CTRP publication (Seashore-Ludlow et al., 2015). RACS data was retrieved from the GDSC dataset such that data availability is limited to overlapping cell lines. However, since there can be cell lines with missing drug response values, but available RACS data in the GDSC dataset, there can still be cell lines that are unique to the validation dataset (CTRP).

Linear regression models

Single-mutation model

$$AUC \sim \beta_0 + \beta_1 mut + covariates + \varepsilon$$

Single-CNA model

$$AUC \sim \beta_0 + \beta_1 cna + covariates + \varepsilon$$

Mutation pair model

$$AUC \sim \beta_0 + \beta_1 mut_1 + \beta_2 mut_2 + \beta_3 mut_1 mut_2 + covariates + \varepsilon \quad (I)$$

CNA pair model

$$AUC \sim \beta_0 + \beta_1 cna_1 + \beta_2 cna_2 + \beta_3 cna_1 cna_2 + covariates + \varepsilon$$

Mutation-CNA pair model

$$AUC \sim \beta_0 + \beta_1 mut + \beta_2 cna + \beta_3 mut\ cna + covariates + \varepsilon$$

Mutation triplet model

$$AUC \sim \beta_0 + \beta_1 mut_1 + \beta_2 mut_2 + \beta_3 mut_3 + \beta_4 mut_1 mut_2 + \beta_5 mut_1 mut_3 + \beta_6 mut_2 mut_3 + \beta_7 mut_1 mut_2 mut_3 + covariates + \varepsilon$$

CNA triplet model

$$AUC \sim \beta_0 + \beta_1 cna_1 + \beta_2 cna_2 + \beta_3 cna_3 + \beta_4 cna_1 cna_2 + \beta_5 cna_1 cna_3 + \beta_6 cna_2 cna_3 + \beta_7 cna_1 cna_2 cna_3 + covariates + \varepsilon$$

Mutation-tissue interaction model

$$AUC \sim \beta_0 + \beta_1 mut + \beta_2 mut\ tissue + covariates + \varepsilon$$

In the alteration models, the AUC is the response variable, β_0 is the intercept, $\beta_{1,2,3,4,5,6,7}$ are the coefficient estimates, $mut_{1,2,3}$ and $cna_{1,2,3}$ are binary variables that encode the alteration status of mutations and RACS, and ε is an unobserved error term. Models indicated as with or without interaction differ in that they include or exclude interaction terms. Candidate alteration pair models without interaction are denoted as alt_1+alt_2, alteration pair models with interaction as alt_1*alt_2, where alt (alteration) can be either a mutation or a CNA. Of note, alteration triplet models include three-way interactions between all alterations in addition to two-way interactions between pairs of alterations.

The included covariates are defined by the default or the drug-specific background model. All models were fitted using the R function lm. Models for combinations of alterations were fitted if the respective alterations co-occurred in at least five cell lines with drug response data.

Alteration distribution

For every single alteration, the number of cell lines with the respective alteration and drug response data was computed. Analogously, the number of cell lines with all alterations and drug response data was computed for alteration pairs and triplets. For the co-occurrence plots, alteration pairs or triplets that co-occurred in less than five cell lines with available drug response data were set to NA and not considered for calculating the mean co-occurrence across drugs.

We used the function phicoef in the R package “GenomicRanges” to compute pairwise correlations for mutation pairs, CNA pairs, and mutation-CNA pairs.

Role of interactions

Synergy is defined as situations where the whole is bigger than the sum of its parts (Anastassiou, 2007). In mutation pair models with interaction (equation I), the sum of both mutation coefficients ($\beta_1 + \beta_2$) represents the sum of parts while the sum of both mutation coefficients and the interaction coefficient ($\beta_1 + \beta_2 + \beta_3$) represents the whole. To assess the magnitude, but not the direction of the effect, we used absolute values for the whole and the sum of parts. Interaction types were classified as synergistic if

$|\beta_1 + \beta_2 + \beta_3| > |\beta_1 + \beta_2|$

or antagonistic if

$|\beta_1 + \beta_2 + \beta_3| < |\beta_1 + \beta_2|$.

We used all fitted mutation pair models with interaction and did not require a minimal difference for the comparison. We refer to the absolute value of β_3 as the strength of the association of an interaction with drug response.

Background models

To select covariates, we assessed the predictive performance of potential confounders in three model settings. In all model equations, AUC is the response variable, β_0 is the intercept, $\beta_{1\text{-}36}$ are the coefficient estimates, $\text{confounder}_{1\text{-}36}$ are the confounder variables, and ε is the error term.

Univariate model

$$\text{AUC} \sim \beta_0 + \beta_1 \text{ confounder} + \varepsilon$$

Univariate models were fitted for all pairwise combinations of drugs and confounders. We extracted the adj. R^2 and counted the number of drugs for which a given confounder was significantly associated with drug response ($p < 0.05$, F-test).

Full model

$$\text{AUC} \sim \beta_0 + \beta_1 \text{ confounder}_1 + \beta_2 \text{ confounder}_2 + \ldots + \beta_{36} \text{ confounder}_{36} + \varepsilon$$

Using model comparison tests, the full model was compared against the full model without one confounder. We counted the number of drugs for which the full model significantly outperformed the full model without a given confounder ($p < 0.05$, F-test). We extracted the adj. R^2 for the full model and the full model without one confounder and computed the difference.

Default background model

$$\text{AUC} \sim \beta_0 + \beta_1 \text{ tissue} + \beta_2 \text{ growth properties} + \beta_3 \text{ growth medium} + \beta_4 \text{ CNA count} + \varepsilon$$

We defined the default background model as a model with the covariates tissue of origin, growth properties, growth medium, and total number of CNAs (CNA count). To estimate the effect of adding a confounder to the default background model, we counted the number of drugs for which the default background model including an additional confounder outperformed the default background model ($p < 0.05$, F-test).

To create drug-specific background models, we used the default background model as a starting point. First, we tested whether the default background model significantly outperformed a model without one of the four covariates ($p < 0.05$, F-test). Covariates not passing this test were excluded from the model. Second, we assessed the effect on performance if one of the 248 cancer driver mutations was included in the model. We included single mutations if the resulting model significantly outperformed a model with the preselected covariates (Holm-corrected $q < 0.1$, t-test). For models including mutations and the tissue covariate, we assessed whether mutation-tissue interaction terms improved the model. For this test, we excluded data for all tissue with less than five cell lines with or without the mutation. We added mutation-tissue interaction terms if the resulting model performed significantly better than a model with the preselected covariates and mutations ($p < 0.05$, F-test).

Model performance

To assess model performance, we fitted models using the entire GDSC dataset and computed the BIC and the adj. R^2. Both metrics penalize the number of variables in a given model. To select the best model per complexity, we assessed all fitted models of a given size and chose the model with the highest adj. R^2 or the lowest BIC. To compute the overlap between best models based on adj. R^2 and BIC, we extracted the variables of the best models and searched for identical models. To compare models to the default background model, we used model comparison tests (F-test) for the best models based on adj. R^2. For the best models based on BIC, we compared the BIC values without requiring a minimal difference.

In addition, we used cross-validation to compare the performance of different models. We used the createMultiFolds function in the R package “caret” (version 6.0-47; https://cran.r-project.org/web/packages/caret) to create 50 independent cross-validation instances. In each cross-validation instance, we generated a training set containing 80% of the data and a test set containing 20% of the data. We used the training set for model fitting and the test set to compute the root mean squared error (RMSE). In each cross-validation instance, the model with the lowest test error was selected as the best model. The test errors of the 50 best models, corresponding to the 50 cross-validation instances, were used to compare the performance of different model complexities. To assess model robustness, we extracted the variables of the 50 best models and determined the frequency of the most frequently selected model.

We used the agreement of duplicates as an upper performance boundary for drugs in the GDSC dataset that were screened twice. For each drug for which a duplicate screen was available, we fitted a univariate model,

$$AUC \sim \beta_0 + \beta_1 AUC_{dup} + \varepsilon$$

where AUC is the drug response to predict, β_0 is the intercept, AUC_{dup} is the drug response of the corresponding duplicate screen, and ε is the error variable. We used 50 cross-validation instances (see above) to compute the test errors of the duplicate model.

Nested models

To identify nested models, we extracted the variables of the best models per complexity based on the adj. R^2. For each drug, we compared alteration pair models with interaction to alteration pair models without interaction and alteration triplet models with interaction to alteration triplet models without interaction. We defined nested models as two models with identical main effects that differ only by the interaction terms.

Clinical drug response

We retrieved clinical response data and mutation data and for 31 melanoma patients that were treated with the BRAF inhibitor Vemurafenib (Van Allen et al., 2014). Out of the 31 patients, all except patient 53 had BRAF-mutated tumors (https://cancer.sanger.ac.uk/cosmic/study/overview?paper_id=34281). We defined two patient subgroups. One group included patients with only BRAF-mutated tumors (n = 26) and the other group included patients with BRAF- and TP53-mutated tumors (n = 4). Clinical response was available as RECIST (Response Evaluation Criteria in Solid Tumors) criteria. Disease control rates for both patient subgroups were computed by dividing the number of patients with the RECIST criteria “complete response”, “partial response” and “stable disease” by the total number of patients. We computed the relative risk using the riskratio function in the R package “fmsb” (https://cran.r-project.org/web/packages/fmsb; version 0.6.3). The study (Van Allen et al., 2014) also included patients receiving the BRAF inhibitor Dabrafenib, but none of the Dabrafenib-treated patients had tumors with both BRAF and TP53 mutations.

Mutation-tissue interactions

To distinguish tissue-specific and general mutation associations in the GDSC dataset, we first selected mutations that were significantly associated with drug response in single-mutation models ($p < 0.05$, t-test). Second, we added mutation-tissue interaction terms and compared models with and without interaction term. We restricted this comparison to genes where at least two tissues with at least five cell lines with and without mutations were available. Mutation associations were defined as tissue-specific if the model with interaction was significantly better than the model without interaction ($p < 0.05$, F-test) and as general otherwise ($p \geq 0.05$, F-test).

To validate our list of tissue-specific and general associations in the CTRP dataset, we applied the same tests and filtered for mutations that mediate drug sensitivity or drug resistance in both datasets.

To test whether the percentage of tissue-specific associations in the GDSC dataset exceeds random expectation, we generated 1000 randomized datasets by permuting the tissue labels of the cancer cell lines. We maintained the mutation frequency of the original data within each tissue of origin. To compute an empirical p-value, we compared the percentage of tissue-specific mutation associations in the original data to the percentages in the randomized data.

To compare the predictive performance of tissue-specific and general mutation associations, we used the adj. R^2 of models with mutation-tissue interaction term. We compared the adj. R^2 distribution of tissue-specific associations to the adj. R^2 distribution of general associations. We used the estimated difference in location of the R function wilcox.test as the effect size. We generated 100 randomized data instances by permuting the tissue annotation while maintaining the unequal distribution of mutations across tissues. We compared the adj. R^2 distributions of tissue-specific and general associations for each randomization instance. We compared the effect size of the original data to the distribution of effect sizes of the randomized data.

To identify mutations that mediate resistance or sensitivity in different tissues, we retrieved a list of tissue-specific mutation associations and fitted single-mutation models to data from individual tissues. We required a significant mutation association in at least two tissue of origins ($p < 0.05$, t-test) and at least one resistance and one sensitivity association. We generated 100 randomized datasets as described above and applied the same condition to the randomized datasets. We computed empirical p-values by comparing the number of sign-switching mutation associations in the original dataset to the number of sign-switching mutation associations in the randomized datasets.

For tissue-specific associations, we assessed whether the strength of mutation associations in single-tissue models is linked to tissue-specific mutation frequencies. We used the R function cor.test to compute the Pearson correlation between the absolute coefficient estimates in single-tissue models and the mutation frequencies in the corresponding tissues.

To identify influential observations, we computed Cook's distance based on the pan-cancer models. We used models with mutation-tissue interaction for tissue-specific mutation associations and models without interaction for general

associations. We excluded models with influential observations (Cook's distance > 0.5) from Figure 20, Figure 21 and Figure 22.

For pan-cancer models, the covariates of the default background model were included. For single-tissue models, we excluded the tissue of origin. We additionally excluded the growth medium and growth properties if all cancer cell lines within a given tissue of origin belonged to the same category.

Identification of mutation-mutation interactions and synthetic lethal triplets

To select examples of mutation-mutation interactions, we used a mutation pair model including the covariates of the drug-specific background model (equation I). We used the GDSC dataset for identification of mutation-mutation interaction candidates and the CTRP dataset for their validation. For each mutation-mutation interaction, three conditions had to be fulfilled. First, the mutation pair model with interaction outperforms the drug-specific background model ($p < 0.05$, F-test) and the drug-specific background model including one of the two mutations ($p < 0.05$, F-test). Second, the mutation-mutation interaction term is significant ($p < 0.05$, t-test). Third, the overall effect in cancer cells with both mutations ($\beta_1+\beta_2+\beta_3$; equation I) has a consistent sign in the GDSC and the CTRP dataset. For synthetic lethal triplets, we additionally filtered for negative interaction effects (β_3; equation I) and negative overall effects ($\beta_1+\beta_2+\beta_3$; equation I). We excluded examples with influential observations (Cook's distance > 0.5) from Figure 24, Figure 25 and Table S1.

Network method

We developed a method to construct mutation interaction networks from which mutation interaction modules can be identified. In a first step, mutation pair models were fitted for all mutation pairs that co-occur in at least five cell lines with drug response data (see above). We used a cross-validation scheme with 50 independent training sets containing 80% of the data. In each cross-validation instance, the respective training set was used for model fitting. Second, we constructed mutation

interaction networks where nodes represent single mutations and edges represent the interaction of two mutations with respect to drug response. The edges in the mutation interaction network are weighted. We defined edge weights as the absolute value of the t-statistic for the interaction term. Third, we applied an edge threshold. We sorted all edges by their strength and kept only a defined percentage of strongest edges. We removed 0-90% of the edges in 10% steps. Of note, nodes can only exist together with an edge. We deleted nodes that were no longer connected after edge thresholding. The resulting network can have a lower number of edges and nodes. Fourth, we ran a module search algorithm. This algorithm partitioned the full network into modules. We translated these modules to linear models by interpreting nodes as main effects and edges as interaction effects. These multivariate linear models were fitted using the training data. The prediction error was computed on the test data. For each module, we applied the module search algorithm recursively until no smaller modules could be identified. Each submodule was translated into a linear model and fitted using the training data.

Algorithm improvement

We evaluated performance based on the ten drugs with the highest mean drug response in the GDSC dataset. We used all modules across drugs, cross-validation instances and edge thresholds to evaluate whether changes in our algorithm improve performance.

To compute CADD scores, we retrieved coding mutations from COSMIC (Forbes et al., 2017; version 80; GRCh37/hg19; CellLinesCodingMuts.vcf.gz) and applied CADD v1.3 (Kircher et al., 2014). We replaced binary mutation data by CADD scores and applied our network method.

Module search algorithms

We used the R package “igraph” for module search algorithms. The functions multilevel.community, fastgreedy.community and infomap.community were used to search for modules using the algorithms Louvain, Fast Greedy, and Infomap.

The module search algorithms Louvain (Blondel et al., 2008) and Fast Greedy (Clauset et al., 2004) aim to maximize the modularity function,

$$Q = \frac{1}{2m}\sum_{ij}\left[A_{ij}\frac{k_i k_j}{2m}\right]\delta(c_i c_j),$$

where A_{ij} represents the edge weight between nodes i and j, k_i and k_j represent the sum of all weighted edges that are attached to node i and j, respectively, 2m is the sum of all weighted edges in the network, and c_i and c_j are the modules to which the nodes i and j belong. The δ-function is 1 if the nodes i and j belong to the same module and 0 otherwise.

The Infomap algorithm (Rosvall and Bergstrom, 2008) is based on a random walk over the network. Edge weights determine the node visit frequency and node weights enable random teleportation between nodes. During the random walk, nodes are encoded by code words whose lengths are inversely proportional to their visit frequency. The aim is to minimize the description length of the network’s topology.

Of the three tested module search algorithms, Infomap was the only one that allows to define node weights. Node weights for a given mutation were defined as the absolute value of the t-statistic for the mutation coefficient. All node weights were normalized such that the sum of all node weights in a network adds up to 1.

Module evaluation

We used all 265 drugs in the GDSC dataset for module evaluation. Across all drugs, all cross-validation instances and all edge thresholds, models for more than 14 million modules were fitted.

Ensemble models

To summarize the 50 result modules that correspond to the 50 cross-validation instances for a given drug, we aimed to generate ensemble models. We tested two approaches to create ensemble models, a cluster-based approach and a frequency-based approach. All ensemble models included the covariates of the default background model. We used the BIC to compare the performance of both approaches.

For the frequency-based approach, we retrieved the variable inclusion frequencies (VIFs) of all nodes and edges across the 50 result modules. We ranked the nodes and edges by VIF. We generated a list of VIF thresholds that includes all unique VIFs. For a given VIF threshold, we fitted a linear model including all nodes (main effects) and edges (interaction effects) that have at least the VIF of the current threshold. Model sizes can theoretically vary between 1 and the sum of all nodes and edges. When the VIF threshold is decreased, one or more variables are added. For each drug, we computed the BIC of all possible ensemble models and selected the model with the best BIC.

For the cluster-based approach, we generated all pairwise combination of modules. For each module pair, we extracted the number of edges in both modules and the number of overlapping edges. We computed the cosine similarity as

$$\frac{\#\ \text{overlapping edges}}{\sqrt{\#\ \text{edges (module 1)}} * \sqrt{\#\ \text{edges (module 2)}}}.$$

Using the matrix of cosine similarities, we created a dissimilarity object by computing the Euclidian distance between all pairs of modules. We performed k-medoid clustering on this dissimilarity object using the function pamk in the R package “fpc”. The function pamk estimates the number of clusters which corresponds to the number of result medoids. All result medoids were translated to linear models. For each drug, we computed the BIC for all medoid models and selected the model with the best BIC.

For validation, we required that an ensemble model outperformed the default background model based on the BIC and based on a model comparison test ($p < 0.05$, F-test) in both the GDSC and the CTRP dataset. All mutations in the ensemble model

had to occur in both datasets. We excluded models with influential observations (Cook's distance > 0.5) in at least one dataset.

Method comparison

We compared our network method against elastic net. We used cross-validation with 50 independent splits (training: 80%, test: 20%) as described above. We used the function cv.glmnet in the R package "glmnet" (version 2.0-16). We enforced the inclusion of the default background model covariates by setting the penalty.factor argument for the respective coefficients to 0. We tested alpha values between 0 and 1 in 0.1 steps. For each value of alpha, we searched the lambda value yielding the lowest mean squared error by 10-fold cross-validation. To calculate the test set RMSE, we used the combination of alpha and lambda values that yielded the lowest RMSE. Since the elastic net implementation we used does not respect variable hierarchies, we implemented elastic net with and without interaction.

Acknowledgments

I would like to thank Eike Staub and Roland Eils for giving me the opportunity to pursue my PhD project in their groups. I thank Merck KGaA for the financial support.

I thank Eike Staub for giving me broad scientific freedom, for insisting on his convictions “One slide – one message” and “Explain it to your grandmother”, for pushing me to my borders and beyond, and for believing in my abilities to develop this project independently.

I would like to thank Matthias Schlesner for formally taking over my supervision and for the feedback he gave. I thank Johanna Mazur for her advice and support, Daniel Hübschmann for his feedback and endorsement, and Octavio Espinosa for accompanying my project from the beginning.

I am very grateful for Wolfgang Huber’s valuable comments in the context of my Thesis Advisory Committee. I thank Benedikt Brors, Michael Boutros and Ursula Kummer for agreeing to be part of my thesis defense committee.

I particularly thank Verena and Nasrin, my sisters Miriam and Sophie, and my brother David for their vigilant proofreading, Çirts for his thousand questions and my mother and my siblings Noemi and Tobias for their emotional support. I would like to thank my colleagues in Heidelberg and Darmstadt for the scientific inspiration, the company, and the exchange. I thank Corinna Sprengart and Manuela Schäfer for their administrative support and the pleasant interaction that regularly saved my day.

Appendix

1. Abbreviations[1]

ACSL6	Acyl-CoA Synthetase Long Chain Family Member 6
adj. R^2	adjusted coefficient of determination
AKT	cf. AKT1
AKT1	AKT Serine/Threonine Kinase 1
AKT2	AKT Serine/Threonine Kinase 2
ALK	ALK Receptor Tyrosine Kinase
ATM	ATM Serine/Threonine Kinase
AURKB	Aurora Kinase B
AXIN2	Axin 2
BCL2	BCL2, Apoptosis Regulator
BCL2L1	BCL2 Like 1
BCL-XL	cf. BCL2L1
BCL-2	cf. BCL2
BIC	Bayesian information criterion
BRAF	B-Raf Proto-Oncogene, Serine/Threonine Kinase
BRCA1	BRCA1, DNA Repair Associated
CADD	Combined Annotation Dependent Depletion
CCLE	Cancer Cell Line Encyclopedia
CDH1	Cadherin 1
CNA	copy number alteration
COSMIC	Catalogue Of Somatic Mutations In Cancer
CREBBP	CREB Binding Protein
CTNNB1	Catenin Beta 1
CTRP	Cancer Therapeutic Response Portal
EGFR	Epidermal Growth Factor Receptor
EGLN1	Egl-9 Family Hypoxia Inducible Factor 1
EP300	E1A Binding Protein P300

[1] Gene names according to (Stelzer et al., 2016).

ERBB2	Erb-B2 Receptor Tyrosine Kinase 2
ERK	cf. MAPK1
FGFR2	Fibroblast Growth Factor Receptor 2
FOXO3	Forkhead Box O3
FOXP1	Forkhead Box P1
GDSC	Genomics of Drug Sensitivity in Cancer
HRAS	HRas Proto-Oncogene, GTPase
IGF1R	Insulin Like Growth Factor 1 Receptor
IKK	Inhibitor Of Nuclear Factor Kappa B Kinase (protein complex)
JNK	cf. MAPK8
KDM6A	Lysine Demethylase 6A
KMT2C	Lysine Methyltransferase 2C
KRAS	KRAS Proto-Oncogene, GTPase
MAPK1	Mitogen-Activated Protein Kinase 1
MAPK8	Mitogen-Activated Protein Kinase 8
MAP2K1	Mitogen-Activated Protein Kinase Kinase 1
MAP2K2	Mitogen-Activated Protein Kinase Kinase 1
MAP3K4	Mitogen-Activated Protein Kinase Kinase Kinase 4
MDM2	MDM2 Proto-Oncogene
MEK1	cf. MAP2K1
MEK2	cf. MAP2K2
MITF	Melanocyte Inducing Transcription Factor
MLL3	cf. KMT2C
MYH10	Myosin Heavy Chain 10
NRAS	NRAS Proto-Oncogene, GTPase
PIK3CA	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha
PI3K	Phosphatidylinositol 3-kinase (composed of a regulatory and a catalytic subunit)
PTEN	Phosphatase And Tensin Homolog
RAF	Raf kinases, a family of three serine/threonine kinases (ARAF, BRAF and RAF1)
RB1	RB Transcriptional Corepressor 1
RHOA	Ras Homolog Family Member A

SOS1	SOS Ras/Rac Guanine Nucleotide Exchange Factor 1
THRAP3	Thyroid Hormone Receptor Associated Protein 3
TOP1	DNA Topoisomerase I
TP53	Tumor Protein P53

2. Supplementary Tables

Table S1: Mutation-mutation interactions (cf. Chapter 3.2). Synth. lethal: synthetic lethal.

Drug name (ID)	Target	Model	Synth. lethal
Crizotinib	MET, ALK	KALRN*NSD1	
Doxorubicin	DNA intercalating	KRAS*PIK3CB	yes
Doxorubicin	DNA intercalating	BRCA2*RAD21	
Gemcitabine	DNA replication	CREBBP*FGFR2	
Obatoclax Mesylate	BCL-2, BCL-XL, MCL-1	CHEK2*EGFR	
CAL-101	PI3K delta	APC*SUZ12	yes
YM155	BIRC5 (Survivin)	BMPR2*FLT3	yes
YM155	BIRC5 (Survivin)	AKAP9*MECOM	yes
PHA-793887	CDK-pan	ACVR1B*THRAP3	
PI-103	PI3K alpha,DNAPK	EP300*NOTCH2	yes
PIK-93	PI4K,PI3K	BRAF*MYH10	yes
SNX-2112	HSP90	NF1*SPTAN1	
SNX-2112	HSP90	SUZ12*TP53	
Cytarabine	DNA synthesis	CREBBP*TP53	
Cytarabine	DNA synthesis	BMPR2*RAD21	
Cytarabine	DNA synthesis	KRAS*MAP3K4	yes
Gefitinib	EGFR	EGFR*FLT3	
Gefitinib	EGFR	ARID1A*EGFR	
Gefitinib	EGFR	DICER1*EGFR	
Nilotinib	ABL	CTNNB1*MAP4K3	yes
Nilotinib	ABL	HSP90AB1*MLL	yes
Nilotinib	ABL	FN1*MAP4K3	yes
Temsirolimus	mTOR	ATRX*PTEN	
Bosutinib	SRC, ABL, TEC	ARID1A*EGFR	
Bosutinib	SRC, ABL, TEC	EGFR*MYH9	
Afatinib (1032)	ERBB2, EGFR	ARID1A*EGFR	
Nutlin-3a (-)	MDM2	CSNK1G3*TP53	
Nutlin-3a (-)	MDM2	CTNNB1*PIK3CA	yes
Nutlin-3a (-)	MDM2	MYH11*TSC1	
MK-2206	AKT1, AKT2	MLL*SMAD4	yes
MK-2206	AKT1, AKT2	BMPR2*KALRN	
BMS-536924 (1091)	IGF1R	AKAP9*MLL3	
BMS-536924 (1091)	IGF1R	CREBBP*TAOK2	
TW 37	BCL-2, BCL-XL	ATR*FAM123B	yes
TW 37	BCL-2, BCL-XL	CDC73*TRIO	
Tamoxifen	ER	APC*KALRN	

UNC0638 (1236)	G9a(EHMT2), GLP(EHMT1)	CHD9*MYD88	
UNC0638 (1236)	G9a(EHMT2), GLP(EHMT1)	ATM*PTPRU	
UNC0638 (1236)	G9a(EHMT2), GLP(EHMT1)	CREBBP*MYD88	
PLX4720 (1371)	BRAF	BRAF*TP53	
Trametinib	MEK1, MEK2	ATR*BRCA2	
Bleomycin (50 uM)	DNA damage	CREBBP*FGFR2	
Bleomycin (50 uM)	DNA damage	PIK3R1*SETDB1	
SN-38	TOP1	RASA1*TRIO	
SN-38	TOP1	CREBBP*FGFR2	
SN-38	TOP1	ATM*BMPR2	

References

Aguirre, A.J., Meyers, R.M., Weir, B.A., Vazquez, F., Zhang, C.Z., Ben-David, U., Cook, A., Ha, G., Harrington, W.F., Doshi, M.B., et al. (2016). Genomic copy number dictates a gene-independent cell response to CRISPR/Cas9 targeting. Cancer Discov.

Alexandrov, L.B., Nik-Zainal, S., Wedge, D.C., Aparicio, S.A.J.R., Behjati, S., Biankin, A. V., Bignell, G.R., Bolli, N., Borg, A., Børresen-Dale, A.L., et al. (2013). Signatures of mutational processes in human cancer. Nature.

Van Allen, E.M., Wagle, N., Sucker, A., Treacy, D.J., Johannessen, C.M., Goetz, E.M., Place, C.S., Taylor-Weiner, A., Whittaker, S., Kryukov, G. V., et al. (2014). The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. Cancer Discov. *4*, 94–109.

Anastassiou, D. (2007). Computational analysis of the synergy among multiple interacting genes. Mol. Syst. Biol.

Azuaje, F. (2017). Computational models for predicting drug responses in cancer research. Brief. Bioinform.

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

Basu, A., Bodycombe, N.E., Cheah, J.H., Price, E. V., Liu, K., Schaefer, G.I., Ebright, R.Y., Stewart, M.L., Ito, D., Wang, S., et al. (2013). An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules. Cell *154*, 1151–1161.

Bateson, W., Punnett, R., and Hurst, C. (1905). Reports to the Evolution Committee of the Royal Society, report II. Harrison and Sons, London.

Bayer, I., Groth, P., and Schneckener, S. (2013). Prediction Errors in Learning Drug Response from Gene Expression Data - Influence of Labeling, Sample Size, and Machine Learning Algorithm. PLoS One.

Bien, J., Taylor, J., and Tibshirani, R. (2013). A lasso for hierarchical interactions. Ann.

Stat.

Blondel, V.D., Guillaume, J.-L., Lambiotte, R., and Lefebvre, E. (2008). Fast unfolding of communities in large networks.

Bouhaddou, M., DiStefano, M.S., Riesel, E.A., Carrasco, E., Holzapfel, H.Y., Jones, D.C., Smith, G.R., Stern, A.D., Somani, S.S., Thompson, T.V., et al. (2016). Drug response consistency in CCLE and CGP. Nature *540*, E9–E10.

Cheung-Ong, K., Giaever, G., and Nislow, C. (2013). DNA-damaging agents in cancer chemotherapy: Serendipity and chemical biology. Chem. Biol.

Clauset, A., Newman, M.E.J., and Moore, C. (2004). Finding community structure in very large networks.

Corcoran, R.B., Ebi, H., Turke, A.B., Coffee, E.M., Nishino, M., Cogdill, A.P., Brown, R.D., Pelle, P. Della, Dias-Santagata, D., Hung, K.E., et al. (2012). EGFR-mediated reactivation of MAPK signaling contributes to insensitivity of BRAF-mutant colorectal cancers to RAF inhibition with vemurafenib. Cancer Discov.

Costello, J.C., Heiser, L.M., Georgii, E., Gönen, M., Menden, M.P., Wang, N.J., Bansal, M., Ammad-Ud-Din, M., Hintsanen, P., Khan, S.A., et al. (2014). A community effort to assess and improve drug sensitivity prediction algorithms. Nat. Biotechnol. *32*, 1202–1212.

Cramer, D., Serrano, L., and Schaefer, M.H. (2016). A network of epigenetic modifiers and DNA repair genes controls tissue-specific copy number alteration preference. Elife *5*.

Crowley, E., Di Nicolantonio, F., Loupakis, F., and Bardelli, A. (2013). Liquid biopsy: Monitoring cancer-genetics in the blood. Nat. Rev. Clin. Oncol.

Dittrich, M.T., Klau, G.W., Rosenwald, A., Dandekar, T., and Müller, T. (2008). Identifying functional modules in protein-protein interaction networks: An integrated exact approach. Bioinformatics.

Dong, Z., Zhang, N., Li, C., Wang, H., Fang, Y., Wang, J., and Zheng, X. (2015). Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection. BMC Cancer.

Dutto, I., Scalera, C., and Prosperi, E. (2018). CREBBP and p300 lysine acetyl

transferases in the DNA damage response. Cell. Mol. Life Sci.

Falchook, G.S., Millward, M., Hong, D., Naing, A., Piha-Paul, S., Waguespack, S.G., Cabanillas, M.E., Sherman, S.I., Ma, B., Curtis, M., et al. (2015). BRAF inhibitor dabrafenib in patients with metastatic BRAF-mutant thyroid cancer. Thyroid *25*, 71–77.

Flaherty, K.T., Puzanov, I., Kim, K.B., Ribas, A., McArthur, G.A., Sosman, J.A., O'Dwyer, P.J., Lee, R.J., Grippo, J.F., Nolop, K., et al. (2010). Inhibition of Mutated, Activated BRAF in Metastatic Melanoma. N. Engl. J. Med.

Forbes, S.A., Beare, D., Boutselakis, H., Bamford, S., Bindal, N., Tate, J., Cole, C.G., Ward, S., Dawson, E., Ponting, L., et al. (2017). COSMIC: somatic cancer genetics at high-resolution. Nucleic Acids Res. *45*, D777–D783.

Fu, W., Ma, Q., Chen, L., Li, P., Zhang, M., Ramamoorthy, S., Nawaz, Z., Shimojima, T., Wang, H., Yang, Y., et al. (2009). MDM2 acts downstream of p53 as an E3 ligase to promote FOXO ubiquitination and degradation. J. Biol. Chem.

Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. Nature *483*, 570–575.

Garraway, L.A. (2013). Genomics-driven oncology: Framework for an emerging paradigm. J. Clin. Oncol. *31*, 1806–1814.

Garraway, L.A., Widlund, H.R., Rubin, M.A., Getz, G., Berger, A.J., Ramaswamy, S., Beroukhim, R., Milner, D.A., Granter, S.R., Du, J., et al. (2005). Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature.

Geeleher, P., Cox, N.J., and Huang, R.S. (2014). Clinical drug response can be predicted using baseline gene expression levels and in vitro drug sensitivity in cell lines. 1–12.

Geeleher, P., Gamazon, E.R., Seoighe, C., Cox, N.J., and Huang, R.S. (2016). Consistency in large pharmacogenomic studies. Nature *540*, E1.

Goossens, N., Nakagawa, S., Sun, X., and Hoshida, Y. (2015). Cancer biomarker discovery and validation. Transl. Cancer Res. *4*, 256–269.

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

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: The next generation. Cell.

Hauschild, A., Grob, J.J., Demidov, L. V., Jouary, T., Gutzmer, R., Millward, M., Rutkowski, P., Blank, C.U., Miller, W.H., Kaempgen, E., et al. (2012). Dabrafenib in BRAF-mutated metastatic melanoma: A multicentre, open-label, phase 3 randomised controlled trial. Lancet *380*, 358–365.

Haverty, P.M., Lin, E., Tan, J., Yu, Y., Lam, B., Lianoglou, S., Neve, R.M., Martin, S., Settleman, J., Yauch, R.L., et al. (2016). Reproducible pharmacogenomic profiling of cancer cell line panels. Nature *533*, 333–337.

Huang, S., and Pang, L. (2012). Comparing statistical methods for quantifying drug sensitivity based on in vitro dose-response assays. Assay Drug Dev. Technol. *10*, 88–96.

Huang, Y.L., Chou, W.C., Hsiung, C.N., Hu, L.Y., Chu, H.W., and Shen, C.Y. (2015). FGFR2 regulates Mre11 expression and double-strand break repair via the MEK-ERK-POU1F1 pathway in breast tumorigenesis. Hum. Mol. Genet.

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

Jaeger, S., Duran-Frigola, M., and Aloy, P. (2015). Drug sensitivity in cancer cell lines is not tissue-specific. Mol. Cancer.

Jang, I.S., Neto, E.C., Guinney, J., Friend, S.H., and Margolin, A.A. (2014). Systematic assessment of analytical methods for drug sensitivity prediction from cancer cell line data. Pac. Symp. Biocomput.

Jarvis, M.C., Ebrahimi, D., Temiz, N.A., and Harris, R.S. (2018). Mutation Signatures Including APOBEC in Cancer Cell Lines. JNCI Cancer Spectr.

Jasnos, L., and Korona, R. (2007). Epistatic buffering of fitness loss in yeast double deletion strains. Nat. Genet. *39*, 550–554.

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

Jiang, P., Lee, W., Li, X., Johnson, C., Liu, J.S., Brown, M., Aster, J.C., and Liu, X.S. (2018). Genome-Scale Signatures of Gene Interaction from Compound Screens Predict Clinical Efficacy of Targeted Cancer Therapies. Cell Syst.

Kazandjian, D., Blumenthal, G.M., Yuan, W., He, K., Keegan, P., and Pazdur, R. (2016). FDA Approval of Gefitinib for the Treatment of Patients with Metastatic EGFR Mutation-Positive Non-Small Cell Lung Cancer. Clin Cancer Res.

Kefford, R., Arkenau, H., Brown, M.P., Millward, M., Infante, J.R., Long, G. V, Ouellet, D., Curtis, M., Lebowitz, P.F., and Falchook, G.S. (2010). Phase I/II study of GSK2118436, a selective inhibitor of oncogenic mutant BRAF kinase, in patients with metastatic melanoma and other solid tumors. J. Clin. Oncol. *28*, 8503.

Kim, J.W., Botvinnik, O.B., Abudayyeh, O., Birger, C., Rosenbluh, J., Shrestha, Y., Abazeed, M.E., Hammerman, P.S., Dicara, D., Konieczkowski, D.J., et al. (2016). Characterizing genomic alterations in cancer by complementary functional associations. Nat. Biotechnol. *34*, 539–546.

King, A.J., Patrick, D.R., Batorsky, R.S., Ho, M.L., Do, H.T., Zhang, S.Y., Kumar, R., Rusnak, D.W., Takle, A.K., Wilson, D.M., et al. (2006). Demonstration of a genetic therapeutic index for tumors expressing oncogenic BRAF by the kinase inhibitor SB-590885. Cancer Res.

Kircher, M., Witten, D.M., Jain, P., O'Roak, B.J., Cooper, G.M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. Nat. Genet. *46*, 310–315.

Knijnenburg, T.A., Klau, G.W., Iorio, F., Garnett, M.J., McDermott, U., Shmulevich, I., and Wessels, L.F.A. (2016). Logic models to predict continuous outputs based on binary inputs with an application to personalized cancer therapy. Sci. Rep. *6*, 36812.

Kojima, K., Konopleva, M., McQueen, T., O'Brien, S., Plunkett, W., and Andreeff, M. (2006). Mdm2 inhibitor Nutlin-3a induces p53-mediated apoptosis by transcription-dependent and transcription-independent mechanisms and may overcome Atm-

mediated resistance to fludarabine in chronic lymphocytic leukemia. Blood.

Kopetz, S., Desai, J., Chan, E., Hecht, J.R., O'Dwyer, P.J., Lee, R.J., Nolop, K.B., and Saltz, L. (2010). PLX4032 in metastatic colorectal cancer patients with mutant BRAF tumors. J. Clin. Oncol.

Krayem, M., Journe, F., Wiedig, M., Morandini, R., Najem, A., Salès, F., van Kempen, L.C., Sibille, C., Awada, A., Marine, J.-C., et al. (2016). p53 Reactivation by PRIMA-1Met (APR-246) sensitises V600E/KBRAF melanoma to vemurafenib. Eur. J. Cancer *55*, 98–110.

Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. J. Stat. Softw. *28*, 1–26.

Kunkele, A., De Preter, K., Heukamp, L., Thor, T., Pajtler, K.W., Hartmann, W., Mittelbronn, M., Grotzer, M.A., Deubzer, H.E., Speleman, F., et al. (2012). Pharmacological activation of the p53 pathway by nutlin-3 exerts anti-tumoral effects in medulloblastomas. Neuro. Oncol. *14*, 859–869.

Lane, D.P. (1992). p53, guardian of the genome. Nature.

Lankenau, M.A., Patel, R., Liyanarachchi, S., Maharry, S.E., Hoag, K.W., Duggan, M., Walker, C.J., Markowitz, J., Carson, W.E., Eisfeld, A.-K., et al. (2015). MicroRNA-3151 inactivates TP53 in *BRAF* -mutated human malignancies. Proc. Natl. Acad. Sci.

Lecca, P., and Re, A. (2015). Detecting modules in biological networks by edge weight clustering and entropy significance. Front. Genet.

Lehmann, S., Bykov, V.J.N., Ali, D., Andreń, O., Cherif, H., Tidefelt, U., Uggla, B., Yachnin, J., Juliusson, G., Moshfegh, A., et al. (2012). Targeting p53 in vivo: A first-in-human study with p53-targeting compound APR-246 in refractory hematologic malignancies and prostate cancer. J. Clin. Oncol.

Li, X.J., Mishra, S.K., Wu, M., Zhang, F., and Zheng, J. (2014). Syn-lethality: An integrative knowledge base of synthetic lethality towards discovery of selective anticancer therapies. Biomed Res. Int. *2014*.

Liu, Y., Fei, T., Zheng, X., Brown, M., Zhang, P., Liu, X.S., and Wang, H. (2016). An Integrative Pharmacogenomic Approach Identifies Two-drug Combination Therapies for Personalized Cancer Medicine. Sci. Rep. *6*.

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

Menden, M.P., Iorio, F., Garnett, M., McDermott, U., Benes, C.H., Ballester, P.J., and Saez-Rodriguez, J. (2013). Machine Learning Prediction of Cancer Cell Sensitivity to Drugs Based on Genomic and Chemical Properties. PLoS One.

Menden, M.P., Wang, D., Guan, Y., Mason, M., Szalai, B., Bulusu, K.C., Yu, T., Kang, J., Jeon, M., Wolfinger, R., et al. (2017). Community assessment of cancer drug combination screens identifies strategies for synergy prediction. BioRxiv.

Michaelis, M., Rothweiler, F., Nerreter, T., Van Rikxoort, M., Sharifi, M., Wiese, M., Ghafourian, T., and Cinatl, J. (2014). Differential effects of the oncogenic BRAF inhibitor PLX4032 (vemurafenib) and its progenitor PLX4720 on ABCB1 function. J. Pharm. Pharm. Sci.

Motter, A.E., Gulbahce, N., Almaas, E., and Barabási, A.L. (2008). Predicting synthetic rescues in metabolic networks. Mol. Syst. Biol. *4*.

Müller, J., Krijgsman, O., Tsoi, J., Robert, L., Hugo, W., Song, C., Kong, X., Possik, P.A., Cornelissen-Steijger, P.D.M., Foppen, M.H.G., et al. (2014). Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. Nat. Commun.

Nass, S.J., and Moses, H.L. (2007). Cancer Biomarkers: The Promises and Challenges of Improving Detection and Treatment.

Nazarian, R., Shi, H., Wang, Q., Kong, X., Koya, R.C., Lee, H., Chen, Z., Lee, M.K., Attar, N., Sazegar, H., et al. (2010). Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. Nature *468*, 973–977.

Nguyen, L., Dang, C.C., and Ballester, P.J. (2016). Systematic assessment of multi-gene predictors of pan-cancer cell line sensitivity to drugs exploiting gene expression data. F1000Research *5*.

O'Neil, N.J., Bailey, M.L., and Hieter, P. (2017). Synthetic lethality and cancer. Nat. Rev. Genet. *18*, 613–623.

Paziewska, A., Dabrowska, M., Goryca, K., Antoniewicz, A., Dobruch, J., Mikula, M.,

Jarosz, D., Zapala, L., Borowka, A., and Ostrowski, J. (2014). DNA methylation status is more reliable than gene expression at detecting cancer in prostate biopsy. Br. J. Cancer.

Prahallad, A., Sun, C., Huang, S., Di Nicolantonio, F., Salazar, R., Zecchin, D., Beijersbergen, R.L., Bardelli, A., and Bernards, R. (2012). Unresponsiveness of colon cancer to BRAF(V600E) inhibition through feedback activation of EGFR. Nature.

Prasad, V., and Mailankody, S. (2017). Research and development spending to bring a single cancer drug to market and revenues after approval. JAMA Intern. Med.

Rheault, T.R., Stellwagen, J.C., Adjabeng, G.M., Hornberger, K.R., Petrov, K.G., Waterson, A.G., Dickerson, S.H., Mook, R.A., Laquerre, S.G., King, A.J., et al. (2013). Discovery of dabrafenib: A selective inhibitor of Raf Kinases with antitumor activity against B-Raf-driven tumors. ACS Med. Chem. Lett. *4*, 358–362.

Riemer, P., Rydenfelt, M., Marks, M., van Eunen, K., Thedieck, K., Herrmann, B.G., Blüthgen, N., Sers, C., and Morkel, M. (2017). Oncogenic β-catenin and PIK3CA instruct network states and cancer phenotypes in intestinal organoids. J. Cell Biol.

Rosvall, M., and Bergstrom, C.T. (2008). Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. U. S. A. *105*, 1118–1123.

Safikhani, Z., El-Hachem, N., Smirnov, P., Freeman, M., Goldenberg, A., Birkbak, N.J., Beck, A.H., Aerts, H.J.W.L., Quackenbush, J., and Haibe-Kains, B. (2016). Safikhani et al. reply. Nature *540*, E2.

Seashore-Ludlow, B., Rees, M.G., Cheah, J.H., Coko, M., Price, E. V., Coletti, M.E., Jones, V., Bodycombe, N.E., Soule, C.K., Gould, J., et al. (2015). Harnessing connectivity in a large-scale small-molecule sensitivity dataset. Cancer Discov. *5*, 1210–1223.

Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Iny Stein, T., Nudel, R., Lieder, I., Mazor, Y., et al. (2016). The GeneCards suite: From gene data mining to disease genome sequence analyses. Curr. Protoc. Bioinforma.

Strachan, T., and Read, A.P. (1999). Human molecular genetics two (Wiley-Liss).

Tanami, H., Imoto, I., Hirasawa, A., Yuki, Y., Sonoda, I., Inoue, J., Yasui, K., Misawa-Furihata, A., Kawakami, Y., and Inazawa, J. (2004). Involvement of overexpressed

wild-type BRAF in the growth of malignant melanoma cell lines. Oncogene.

Tang, Y.C., and Amon, A. (2013). Gene copy-number alterations: A cost-benefit analysis. Cell.

Tenbaum, S.P., Ordóñez-Morán, P., Puig, I., Chicote, I., Arqués, O., Landolfi, S., Fernández, Y., Herance, J.R., Gispert, J.D., Mendizabal, L., et al. (2012). β-Catenin confers resistance to PI3K and AKT inhibitors and subverts FOXO3a to promote metastasis in colon cancer. Nat. Med.

Tong, A.H., Evangelista, M., Parsons, A.B., Xu, H., Bader, G.D., Pagé, N., Robinson, M., Raghibizadeh, S., Hogue, C.W., Bussey, H., et al. (2001). Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science *294*, 2364–2368.

Tong, A.H.Y., Lesage, G., Bader, G.D., Ding, H., Xu, H., Xin, X., Young, J., Berriz, G.F., Brost, R.L., Chang, M., et al. (2004). Global mapping of the yeast genetic interaction network. Science *303*, 808–813.

Tsai, J., Lee, J.T., Wang, W., Zhang, J., Cho, H., Mamo, S., Bremer, R., Gillette, S., Kong, J., Haass, N.K., et al. (2008). Discovery of a selective inhibitor of oncogenic B-Raf kinase with potent antimelanoma activity. Proc. Natl. Acad. Sci. *105*, 3041–3046.

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

Villalonga-Planells, R., Coll-Mulet, L., Martínez-Soler, F., Castaño, E., Acebes, J.-J., Giménez-Bonafé, P., Gil, J., and Tortosa, A. (2011). Activation of p53 by Nutlin-3a Induces Apoptosis and Cellular Senescence in Human Glioblastoma Multiforme. PLoS One *6*, e18588.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., and Kinzler, K.W. (2013). Cancer Genome Landscapes. Science (80-.). *339*, 1546–1558.

Wan, P.T.C., Garnett, M.J., Roe, S.M., Lee, S., Niculescu-Duvaz, D., Good, V.M., Project, C.G., Jones, C.M., Marshall, C.J., Springer, C.J., et al. (2004). Mechanism of activation of the RAF-ERK signaling pathway by oncogenic mutations of B-RAF. Cell *116*, 855–867.

Watkinson, J., Wang, X., Zheng, T., and Anastassiou, D. (2008). Identification of gene

interactions associated with disease from gene expression data using synergy networks. BMC Syst. Biol. *2*, 10.

Wecker, H., and Waller, C.F. (2018). Afatinib. In Recent Results in Cancer Research. Fortschritte Der Krebsforschung. Progres Dans Les Recherches Sur Le Cancer, pp. 199–215.

Yang, W., Soares, J., Greninger, P., Edelman, E.J., Lightfoot, H., Forbes, S., Bindal, N., Beare, D., Smith, J.A., Thompson, I.R., et al. (2013). Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. *41*, D955.

Yu, H., McDaid, R., Lee, J., Possik, P., Li, L., Kumar, S.M., Elder, D.E., Van Belle, P., Gimotty, P., Guerra, M., et al. (2009). The role of BRAF mutation and p53 inactivation during transformation of a subpopulation of primary human melanocytes. Am. J. Pathol.

Zhang, W., and Liu, H.T. (2002). MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Res. *12*, 9–18.

Zou, H., and Hastie, T. (2005). Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol.