bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Identifying Drug Sensitivity Subnetworks with NETPHIX

Yoo-Ah Kim [*†] Rebecca Sarto Basso [*†‡] Damian Wojtowicz [†] Dorit S. Hochbaum [‡]
Fabio Vandin [§¶] Teresa M. Przytycka [†¶]

Abstract

Phenotypic heterogeneity in cancer is often caused by different patterns of genetic alterations. Understanding such phenotype-genotype relationships is fundamental for the advance of personalized medicine. One of the important challenges in the area is to predict drug response on a personalized level. The pathway-centric view of cancer significantly advanced the understanding of genotype-phenotype relationships. However, most of network identification methods in cancer focus on identifying subnetworks that include general cancer drivers or are associated with discrete features such as cancer subtypes, hence cannot be applied directly for the analysis of continuous features like drug response. On the other hand, existing genome wide association approaches do not fully utilize the complex proprieties of cancer mutational landscape. To address these challenges, we propose a computational method, named NETPHIX (NETwork-to-PHenotpe assocIation with eXlusivity), which aims to identify mutated subnetworks that are associated with drug response (or any continuous cancer phenotype). Utilizing properties such as mutual exclusivity and interactions among genes, we formulate the problem as an integer linear program and solve it optimally to obtain a set of genes satisfying the constraints. NETPHIX identified gene modules significantly associated with many drugs, including interesting response modules to MEK1/2 inhibitors in both directions (increased and decreased sensitivity to the drug) that the previous method, which does not utilize network information, failed to identify. The genes in the modules belong to MAPK/ERK signaling pathway, which is the targeted pathway of the drug.

1 Introduction

Genetic alterations in cancer are associated with diverse phenotypic properties such as drug response or patient survival. However, the identification of mutations causing specific phenotypes and the interpretation of the phenotype-genotype relationships remain challenging due to a large number of passenger mutations and cancer heterogeneity. Indeed, the relationships between genotype and phenotype in most tumors are complex and different mutations in functionally related genes can lead to the same phenotype. The pathway-centric view of cancer [1, 2, 3] suggests that cancer phenotypes should be considered from the context of dysregulated pathways rather than from the perspective of mutations in individual genes. Such pathway-centric view significantly advanced the understanding of the mechanisms of tumorigenesis. Many computational methods to identify cancer driving mutations have been developed based on pathway approaches [4, 5, 6, 7, 8, 9]. Network based approaches have been further applied to find subnetworks associated with various disease phenotypes [4, 6, 10, 11, 12].

[*]Equal first author contribution
[†]National Center of Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD, USA
[‡]Department of Industrial Engineering and Operations Research, University of California at Berkeley, CA, USA
[§]Department of Information Engineering, University of Padova, Italy
[¶]Co-corresponding authors: fabio.vandin@unipd.it and przytyck@ncbi.nlm.nih.gov

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Those methods have been developed aiming to find genes whose mutations are associated specifically with given phenotypes rather than finding general cancer drivers.

Recent projects have characterized drug sensitivity for a large number of drugs in hundreds of cancer cell lines [13], [14]. This data, together with information about the genetic alterations in these cells, can be used to understand how genomic alterations impact drug sensitivity. While the success of network based methods in other cancer domains suggests that such approaches should be also useful in the studies of drug response, most of previous approaches focused on discrete phenotypic traits – e.g., cancer vs. healthy, good or bad prognosis, or cancer subtypes – and therefore, cannot be directly applied to the analysis of continuous features such as drug sensitivity.

Several algorithms for the identification of mutations associated with drug response have been previously developed [15, 16] but without considering functional relationships among genes. For example, REVEALER used a re-scaled mutual information metric to iteratively identify a set of genes associated with the phenotype [16]. UNCOVER employs an integer linear programming formulation based on the set cover problem, by designing the objective function to maximize the association with the phenotype and preferentially select mutually exclusive gene sets [15]. However, without interaction information, the genes identified by the algorithms may not belong to the same pathways, making them more likely to include false positives and making it difficult to interpret the uncovered association and the underlying mechanism.

To address these challenges, we introduce a computational tool named NETPHIX (NETwork-to-PHenotype assocIation with eXlusivity). With the goal of identifying mutated subnetworks that are associated with a continuous phenotype, we formulate the problem as an integer linear program and solve it to optimality using CPLEX. For each drug, we attempt to identify both directions of associated subnetworks– a subnetwork whose alterations correlate with increased sensitivity to the drug (decreased cell survival) and a subnetwork that correlates with reduced sensitivity to the drug (increased cell survival). Based on the fact that mutations in cancer drivers tend to be heterogeneous, our algorithm builds on combinatorial optimization techniques involving set cover and network constraints. In addition, NETPHIX preferentially selects mutually exclusive genes as the solution, utilizing an observation that patient groups harboring different cancer driving mutations tend to be mutually exclusive [17, 18, 19, 7, 20, 21]. This approach together with a carefully designed strategy for selecting subnetwork size allows to leave out passenger mutations from the sensitivity networks.

There have been related studies combining GWAS analysis with network constraints [22, 23, 24, 25]. While these methods generally perform well at pointing broadly defined disease related functional pathways, they do not consider complex properties of cancer mutations such as the aforementioned mutual exclusivity of cancer drivers, and are not designed to zoom on subnetworks that are specific enough to help understand drug action. As discussed later in this work, the genomic landscape related to drug response can be complex and mutations in different genes in the same pathway can affect the response differently. Pharmaceutical drugs are often developed to target specific genes, and the response depends on the function and the mutation status of the gene as well as other genes in the same pathway.

We evaluated NETPHIX and other related methods using simulations and showed that NETPHIX outperforms competing methods. Applying NETPHIX to drug response data, we identified sensitivity-associated (increasing or decreasing the sensitivity) subnetworks for a large set of drugs. These subnetworks provided important insights into drug action. Effective computational methods to discover these associations will improve our understanding of the molecular mechanism of drug sensitivity,

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

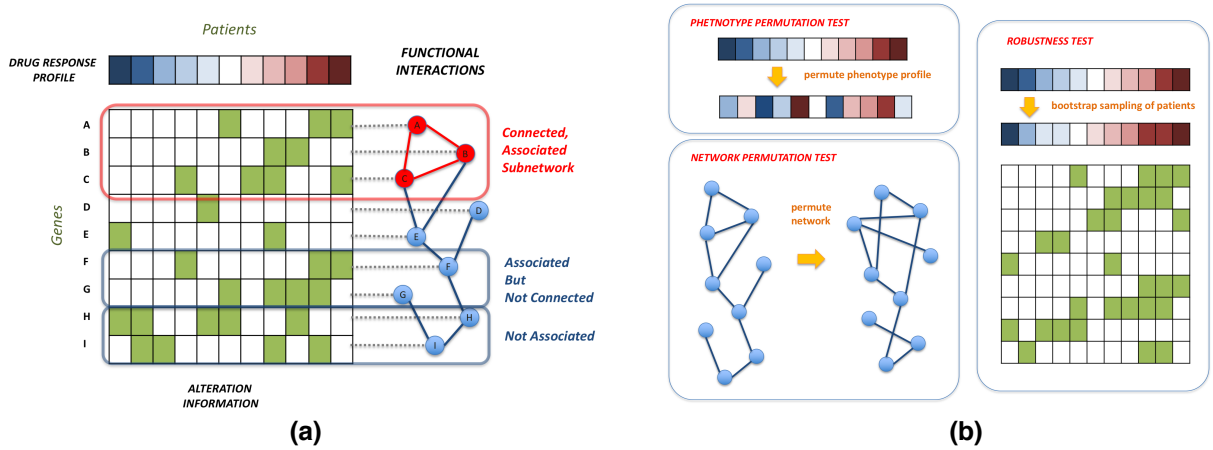


Figure 1: Method Overview. **(a)** NETPHIX finds a connected set of genes for which corresponding mutations are associated with positive phenotype values (red colors in the drug response profile indicate positive values and blue colors are for negative values). **(b)** The significance and the robustness of identified modules are assessed using two different permutation tests and bootstrapping.

help to identify potential dug combinations, and have a profound impact on genome-driven, personalized drug therapy. NETPHIX is available at `https://www.ncbi.nlm.nih.gov/CBBresearch/Przytycka/index.cgi#netphlix`

2 Method

2.1 The NETPHIX method overview

Given gene alteration information of cancer samples and their drug sensitivity profiles (or any cancer-related, continuous phenotypes), NETPHIX aims to identify genetic alterations underlying the phenotype of interest. Starting with the assumption that genes whose mutations lead to the same phenotype must be functionally related, NETPHIX utilizes functional interaction information among genes and enforces the identified genes to be highly *connected* in the network while, at the same time, making sure that the aggregated alterations of these genes are significantly *associated* with the given phenotype (Figure 1a). In addition, to leverage the property of heterogeneity and mutual exclusivity, NETPHIX utilizes a set cover approach and penalizes overlapping mutations. Specifically, it has been observed that patient groups harboring different cancer driving mutations tend to be mutually exclusive. This property may arise when mutations in two different genes lead to dysregulation of the same cancer driving pathway and the role of the two genes for cancer progression is redundant. In such cases, observing mutations in both genes in one patient is unlikely. Building on this observation, NETPHIX identifies a connected set of genes S such that the sum of phenotypic weights of the patients with alterations in S (minus the penalties for overlapping alterations) is maximized. For example, in Figure 1a, the combined alteration of gene set A, B, C would be identified by NETPHIX as the module is functionally connected and has significant positive association with the phenotype (even though individual gene associations may not be as significant). The patients with alterations in genes A and B are completely mutually exclusive while there is only one patient with overlapping mutations in B and C.

We formulated the problem as an integer linear program (ILP) and solved it to obtain the optimal set of genes that satisfies the constraints using CPLEX (`https://www.ibm.com/analytics/cplex-optimizer`). We provide the formal definition of the problem and the detailed ILP formu-

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

lation in Section 2.2 and 2.3, respectively.

Once we obtain the optimal gene modules, we assess both the significance and robustness of the identified modules by performing permutation tests and bootstrapping (Figure 1b and Section 4.2). To assess the significance of the association between the phenotype and the identified subnetwork, we performed permutation tests by permuting the phenotype profile of the patients. Note that our algorithm is designed to identify the modules associated specifically with a given phenotype (e.g., drug sensitivity to each drug) rather than finding general cancer drivers, and the permutation test will estimate the significance of the association of the given phenotype profile compared with randomly generated phenotypes. In addition, we performed another permutation test based on permutations of functional interactions (in a degree preserving way), which assess the importance of the interaction information in the solution. Finally, we also examine the robustness of the gene selections by performing bootstrap sampling of the patients and solved the ILP with the phenotype and alterations profiles for the sampled sets of patients. See Section 4.2 for the details of the permutation and bootstrapping procedures.

2.2 Formal definition of the computational problem

We are given a graph $G = (V, E)$, with vertices $V = \{1, \ldots, n\}$ representing genes and edges E representing interactions among genes. Let P denote the set of m patients (samples). For each sample $j \in P$, we are also given a phenotype profile value $w_j \in \mathbb{R}$ which quantitatively measures a phenotype (e.g., drug response, pathway activation, etc.). Let $P_i \subseteq P$ be the set of samples in which gene $i \in V$ is altered. We say that a patient $j \in P$ is *covered* by gene $i \in V$ if $j \in P_i$ i.e. if gene i is mutated in sample j. We say that a sample $j \in P$ is *covered* by a subset of vertices $S \subseteq V$, if there exists at least one vertex v in S such that $j \in P_v$.

Our goal is to identify a connected subgraph S of G of at most k vertices such that the sum of the weights of the samples covered by S is maximized. Since we are interested in functionally complementary mutations, we also penalize coverage overlap when an element is covered more than once by S by assigning a penalty p_j for each of the additional times sample j is covered by S. As penalty we use the average of the positive phenotype values if the original value of the element was positive. If the original value of the element was negative we assign a penalty equal to its value. Let $c_S(j)$ be the number of times element $j \in P$ is covered by S. For a set S of genes, we define its weight $W(S)$ as:

$$W(S) = \sum_{j \in \cup_{s \in S} P_s} w_j - \sum_{j \in \cup_{s \in S} P_s} (c_S(j) - 1) p_j$$

Thus, we define the optimization problem as follows:

The Phenotype Associated Connected Coverage problem: *G*iven a graph G defined on a set of n vertices V, a set P, a family of subsets $P = \{P_1, \ldots, P_n\}$ where for each i, $P_i \subseteq P$ is associated with $i \in V$, weights w_j and penalties $p_j \geq 0$ for each sample $j \in P$ find the subset $S \subseteq V$ of $\leq k$ vertices maximizing $W(S)$.

The Phenotype Associated Connected Coverage problem is NP-hard since for a complete graph the problem is equivalent to the NP-hard Target Associated k-Set problem studied in [15]. Although the problem is NP-hard, we formulated it as an integer linear programming as described in the next subsection, and solved it to optimality using CPLEX, which can be run in a reasonable amount of time (See Figure S3b for running times for different k's).

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

2.3 ILP formulation

An ILP formulation for Target Associated k-Set problem was considered in [15]. In NETPHIX we include an additional set of constraints that ensures the genes selected in the solution are connected in the network V. Let x_i be a binary variable (denoted with $x_i \in \mathbb{B}$) equal to 1 if gene $i \in V$ is selected and $x_i = 0$ otherwise. Let z_j be a binary variable equal to 1 if sample j is covered and $z_j = 0$ otherwise. Let y_j denote the number of times sample j is covered in the solution. Finally, let w_j be the weight of sample j and p_j be the penalty for sample j. Our ILP formulation is as follows:

$$
\begin{aligned}
z(q) = \max \quad & \sum_j (w_j + p_j) z_j - \sum_j p_j y_j && && (1)\\
\text{s.t.} \quad & \sum_i x_i \leq k, && && (2)\\
& y_j = \sum_{i: j \in P_i} x_i, && \forall j && (3)\\
& y_j \geq z_j, && \forall j && (4)\\
& z_j \geq y_j / k, && \forall j && (5)\\
& \sum_{l: il \in E} x_l \geq D(k-1)(x_i - 1) + D\left(\sum_{l \in V} x_l - 1\right) && \forall i \in V && (6)\\
& x_i, z_j \in \mathbb{B}, y_j \in \mathbb{N} && \forall i, j && (7)
\end{aligned}
$$

Constraint (2) impose that the total number of sets in the solution is at most k. Constraints (3) define how many times each sample has been covered. Constraints (4) ensure that for each element $j \in P$, if j is covered by the current solution then the number of times j is covered in the solution is at least 1. Constraints (5) impose that for each element $j \in P$, if j is covered by at least one element in the current solution then j is covered.

Constraints (6) were used to ensure the high connectivity of selected module. Specifically, the constraints enforce that each selected gene is connected with at least D fraction of genes in the selected module (other than the gene itself). Note that if $D \geq 0.5$, the module is a connected subgraph since for any two non-adjacent vertices, they must have a common neighbor ($D = 0.5$ is used in our analysis). In our study, we used a functional interaction network (from STRING database), which is relatively dense. For sparse networks where highly connected components are rare, we may use an alternative approach based on a branch-and-cut algorithm to ensure the connectivity. See Supplementary Section S1 for the description of an alternative algorithm.

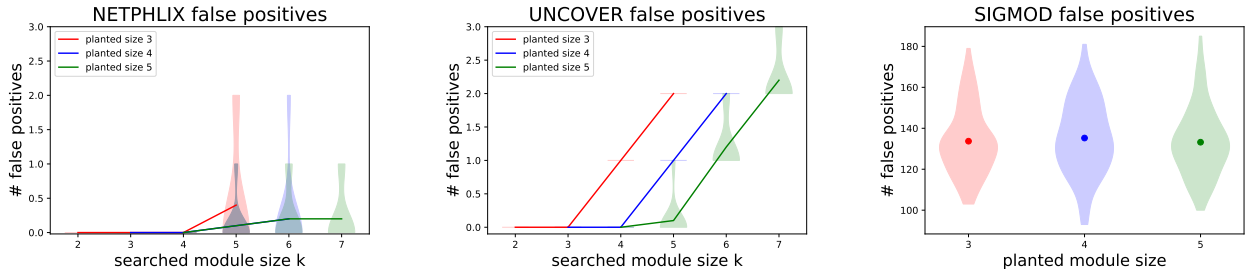
To select an appropriate module size k, we computed modules of increasing sizes, stopping the process if increasing module size does not satisfy the constraints on the objective value of the optimal solution and p-values (See Section 4.2 for details).

3 Results

3.1 Evaluation on simulated data

We generated a set of simulated instances where we planted phenotype associated modules with varying parameters onto the background of real cancer cell mutation data (Section 4.1). We then compared

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a) **(b)** **(c)**

Figure 2: Method comparison on simulated data. The number of false positives in the modules identified by **(a)** NETPHIX, **(b)** UNCOVER, and **(c)** SigMOD. The solid lines and dots show the average number of false positives over all different parameters for each k and the shaded areas indicate the distribution over different instances.

the performance of and two related methods – UNCOVER and SigMOD. UNCOVER [15] was proposed previously as a method to identify a set of phenotype-associated genes by considering a similar objective function but without utilizing interaction information. SigMOD is a recently proposed module identification algorithm combining GWAS and network based approach, and it was found to outperform other related methods [25]. SigMOD requires individual association scores of genes to a phenotype as an input, for which we used the p-value of the association of each gene to a phenotype by performing t-tests on the coefficients of univariate linear regression.

We planted modules of size 3, 4, and 5 and we evaluated the accuracy of the three methods in identifying the planted modules (Figure 2). For NETPHIX and UNCOVER, we ran the algorithm for different k's, while SigMOD automatically adjust all its parameters to find the best module. All the algorithms uncovered the planted modules in almost all instances (Figure S1). However, only NETPHIX shows very low rate of false positives, i.e., falsely identified genes (Figure 2). NETPHIX usually does not extend the best module with spurious genes even if we searched for modules bigger than planted while UNCOVER tends to add more genes when increasing k. SigMOD identified a large number of spurious genes along the planted modules (approx. 100-180 genes) that are not associated with phenotypes.

3.2 Comparison of NETPHIX and UNCOVER on drug response dataset

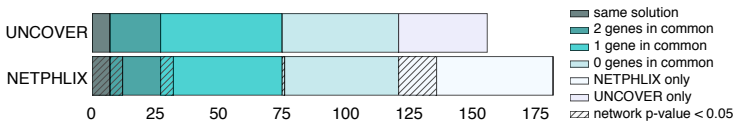
We applied NETPHIX and UNCOVER to analyze a dataset of 736 cancer cell lines for which somatic alterations and drug sensitivity data for 265 drug sensitivity experiments are available (Section 4.1) and we compared the identified modules (Figure 3a). For each drug, we ran both algorithms to identify modules with decreased or increased sensitivity (530 instances in total). For comparison, we considered here modules of size $k = 3$ and the p-value from the phenotype permutation test $p_{ph} \leq 0.05$.

NETPHIX reports 182 modules (out of 530 instances) while UNCOVER finds 156 modules. Although our goal is not to identify cancer drivers but to find the genes associated with sensitivity to each drug, cancer drivers are expected to be most relevant to drug response. The modules reported by included a much higher fraction of cancer genes among the genes in the modules as a whole than the UNCOVER modules, and have a much more significant p-value for the enrichment of cancer driving genes ($p < 10^{-24}$, cancer driver genes reported in [3]). While the UNCOVER modules are also enriched for cancer genes, the enrichment ($p < 10^{-10}$ by Fisher exact test) is lower than for NETPHIX modules. These results show that NETPHIX reports modules that contain many cancer

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

	UNCOVER	NETPHIX	
	pheno sig.	pheno sig.	net sig.
# modules	156	182	33
# genes	273	220	36
# drivers	16	33	14
p-value	10^{-10}	10^{-24}	10^{-16}

(a)



(b)

Figure 3: Comparison of the modules identified by UNCOVER and NETPHIX. **(a)** The table shows the number of genes/modules that are significant with phenotype and network permutations ($p_{ph} \leq 0.05$ and $p_{net} \leq 0.05$) and their cancer driver enrichment. The network permutation test cannot be performed for UNCOVER because it is a network agnostic method. **(b)** The bar chart illustrates similarities and differences between the modules identified by NETPHIX and UNCOVER.

relevant genes with a higher degree of functional coherence with the drug targets than the UNCOVER modules (Figure 3a).

In addition to the phenotype permutation test, we performed the network permutation test for NETPHIX and considered the modules with both p-values $p_{ph} \leq 0.05$ and $p_{net} \leq 0.05$. NETPHIX identifies 15 modules with decreased sensitivity to drug response (increased cell survival) and 18 modules with increased sensitivity to drug response (decreased cell survival). The genes in the NETPHIX modules as a whole are significantly enriched in well-known cancer genes ($p < 10^{-16}$ by Fisher exact test; 27 fold enrichment), showing that NETPHIX identifies modules of genes relevant to the disease (Figure 3a). Of the 33 instances (phenotype and increased/decreased sensitivity association) for which NETPHIX identifies a module, 15 have no module identified by UNCOVER ($p_{ph} \leq 0.05$). Of the remaining 18 instances, in 7 cases the same module is identified by NETPHIX and by UNCOVER, while in 11 cases NETPHIX and UNCOVER report completely or partially different modules (Figure 3b). For the latter, to compare the quality of the modules we checked whether the genes in the module and the drug target (that is unknown to the methods) are part of the same pathway, since one can expect that alterations in different members of the molecular mechanism targeted by the drug have a similar effect on drug response. In 10 cases out of 11, the NETPHIX solution has more members in a pathway (by KEGG or Reactome) that includes the drug target than UNCOVER solutions, while in the remaining case the solutions from the two algorithms have the same number of members in such pathways.

Note that since NETPHIX has additional network constraints compared to UNCOVER, the values of the objective function for NETPHIX's modules cannot be greater than those of UNCOVER for the same instances. Nonetheless, we found that the objective values of NETPHIX's modules are close to the ones of UNCOVER (i.e., at least 75% of UNCOVER's values for most instances, Figure S3) while obtaining more functionally coherent modules.

3.3 Biological implications of drug sensitivity modules identified by NETPHIX

Application of NETPHIX to 530 instances of drug response profiles (increased and decreased sensitivity for 265 drug experiments) with different module sizes k's resulted in 166 modules that are significantly associated with drug sensitivity (Table S2). See Section 4.2 for detailed description on how significant modules are selected.

Many of the modules identified by NETPHIX provide interesting insights related to drug action. In particular, we analyzed the response to drugs targeting the RAS/MAPK pathway (Table S1 and Fig-

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

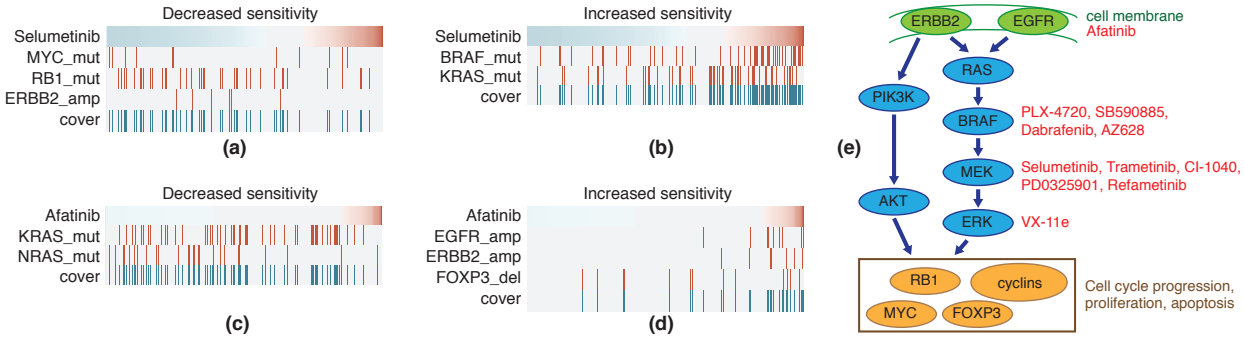


Figure 4: Subnetworks identified by NETPHIX for selected drugs. Each panel shows the values of the phenotype (top row) for all samples (columns), with blue being decreased sensitivity values and red being increased sensitivity values. For each gene in the subnetworks, alterations in each sample are shown in red, while samples not altered are in grey. The last row shows the alteration profile of the entire solution in blue. **(a)** and **(b)** Subnetworks with decreased and increased sensitivity to Selumetinib, respectively. **(c)** and **(d)** Subnetworks with decreased and increased sensitivity to Afatinib, respectively. **(e)** Schematic diagram of MAPK/ERK and AKT signaling pathways with drugs and their drug targets annotated.

ure 4e). This pathway regulates the growth, proliferation and apoptosis and is often dysregulated in various cancers. Among the most common mutations of this pathway are mutations of BRAF. Interestingly, NETPHIX identified the same module (BRAF, KRAS, and NRAS) as associated with increased sensitivity to several of those drugs (CI-1040, PD032590, and Refametinib). All these three drugs act by blocking MEK1 and MEK2 genes that are immediately downstream of BRAF/KRAS/NRAS and thus increased sensitivity of this subnetwork is consistent with the action of these drugs. Moreover, NETPHIX identified the module of genes ERBB2 (amplification), MYC, and RB1 (mutations) as associated with decreased sensitivity to these three drugs.

Selumetinib (another drug targeting MEK 1/2) and VX-11e (which blocks ERK2 gene that is downstream of MEK 1/2) have similar response (Figures 4a,b and S4, and Tables S1 and S2). All the genes in the modules are related to the MAPK/ERK signaling pathway; BRAF, KRAS, NRAS are three core members, ERBB2 is a receptor protein that, in particular, signals through this pathway, while MYC and RB1 are downstream of the MAPK/ERK signaling pathway. These findings indicate that the alterations in different components of the same pathway can contribute to drug sensitivity in different ways.

In contrast to the response to MEK1/2 and ERK2 inhibitors, the drugs directly targeting BRAF are associated with more heterogeneous subnetworks (Table S1), which suggests that patient specific mutational profile can provide important clues in predicting drug response.

The drugs associated with similar modules but with opposite response can be candidates for combination drug therapy. For example, we identified Afatinib as having a subnetwork of EGFR, ERBB2, FOXP3 with increased sensitivity. This suggest that it might be beneficial to use Afatinib in combination with MEK 1/2 and ERK2 targeting drugs. Indeed, clinical trails for the Afatinib and Selumetinib combinations are currently underway (`https://clinicaltrials.gov/ct2/show/NCT02450656`).

There are several MYC-related modules identified by NETPHIX. An interesting example is the module for PHA-793887 (Figure S2a), comprising genes KIT, MYC, and NRAS (phenotype permutation $p_{ph} \leq 10^{-2}$; network permutation $p_{net} \leq 10^{-2}$), all known cancer genes. PHA-793887 targets the cell cycle through the inhibition of members of the cyclin dependent kinase (CDK) family, including CDK2. KIT, MYC, and NRAS are all related to the PI3K-AKT signaling pathway (involved in cell

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

cycle progression) that is upstream of CDK2. Another notable MYC-related module reported by NETPHIX comprises CDKN1B, EGFR, and MYC and is associated (phenotype permutation $p_{ph} \leq 10^{-2}$; network permutation $p_{net} \leq 4 \times 10^{-2}$) with increased sensitivity to Pelitinib (Figure S2b). Pelitinib targets epidermal growth factor receptor (EGFR) and all three genes in the module are related to the ErbB signaling pathway: EGFR is a member of the pathway, while both CDKN1B and MYC are downstream of the pathway (Figure S2b).

In summary,the modules identified by NETPHIX are in good correspondence with the action of the respective drugs, suggesting that NETPHIX can correctly identify relevant modules and the modules can thus be used to predict potential patient-specific drug combinations and to provide guidance to personalized treatment.

4 Materials and method details

4.1 Datasets

Drug sensitivity dataset: The Genomics of Drug Sensitivity in Cancer Project(`https://www.cancerrxgene.org/`) consists of drug sensitivity data generated from high-throughput screening using fluorescence-based cell viability assays following 72 hours of drug treatment. In particular, we considered the area under the curve for each experiment as a phenotype. These scores are provided in the file `portal-GDSC_AUC-201806-21.txt` available through the DepMap data portal (`https://depmap.org`) for 265 compounds and 743 cell lines, with 736 having alteration data available through the DepMap portal. For the DepMap experiments [26, 27], we used the alteration provided at `https://depmap.org/portal/download/all/`. We downloaded the data on July 6th 2018. In particular we used mutation data from the file `portal-mutation-2018-06-21.csv` that includes binary entries for 18,652 gene-level mutations. Additionally we considered 22,746 amplifications and 22,746 deletions computed from the gene copy number data in `portal-copy_number_relative-2018-06-21.csv`, with an amplification defined by a copy number above 2 and a deletion defined by a copy number below -1.

Interaction network For functional interactions among genes, we used the data downloaded from STRING database version 10.0 (`https://string-db.org`). We only included the edges with high confidence scores (≥ 900) as an input to The resulting interaction network includes 9,215 nodes and 160,249 edges.

Preprocessing drug sensitivity data: For every drug response profile, we excluded samples with missing values for that phenotype, which results in a different number of samples for each phenotype. The number of samples varied between 240 and 705. To generate drug sensitivity values for the patients, we took the negatives of cell viability (i.e., increased cell survival indicates decreased sensitivity to the drug and vice versa) and then normalized the phenotype values before running the algorithm, by using standard z-scores (subtracting the average value $\sum_{j \in J} w_j/m$ from each weight w_j and dividing the result by the standard deviation of the (original) w_j's), in order to have both positive and negative phenotype values. Following previously established practice [16], we discarded features with low or high frequency, that correspond to noisy features and to features whose frequency is too high to show a significant association with drug response in combination with other features, respectively. In particular, features present in less than 1% samples or more than 25% samples were excluded from our analyses.

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Generating simulated data: For the background of simulation data, we use the same gene alteration table and interactions from drug sensitivity dataset described previously in this section. The phenotype values for individual samples are randomly drawn from normal distribution $N(0, 1)$. We then planted randomly generated phenotypes and associated modules to the background as follows.

Phenotypes: α fraction of patients $P(\alpha)$ ($\alpha = 0.1, 0.2$, and 0.3) were randomly selected and assigned phenotype values drawn randomly from $N(z, 0.5)$ where z is a z-score corresponding to a cumulative p-value p ($p = 0.005$, 0.1, 0.99, and 0.995).

Associated gene modules: we randomly selected a gene set $S(k)$ of size k ($k = 3, 4$, and 5) and added random alterations in $S(k)$ for patients $P(\alpha)$ so that each patient in $P(\alpha)$ has an alteration in exactly one gene in $S(k)$. Therefore, the added alterations among the patients $P(\alpha)$ are mutually exclusive although there may be overlapping mutations due to the background alterations. We also added random edges among the genes $S(k)$ so that they satisfy the density constraints.

We generated 10 random instances for each combination of parameters (k, α, z) and ran the module identification algorithms.

4.2 Method details

Selecting module size k: To identify significant modules for each of 530 instances of drug response data (increase or decreased sensitivity of each drug experiment), we ran NETPHIX with different k's and choose the best k for each instance as follows: start with $k = 1$ and increase k by one until the improvement is not sufficient (up to $k = 5$). We chose 5% improvement cutoff over the previous k for stop condition ((OPT(k+1)-OPT(k))/OPT(k) < 0.05). Our simulation results show that the improvement of the optimal objective value decreases significantly once the algorithm reaches the size of a correct solution (Fig. S3a). In addition, the algorithm performs phenotype permutation test and stops if p-values starts increasing (i.e., less significant than the previous run). Once the algorithm stops, we define the identified module to be significant if the FDR adjusted p-value (Benjamini/Hochberg) is less than 0.1.

Phenotype permutation test: In the phenotype permutation, the dependencies among alterations in genes are maintained, while the association between alterations and the phenotype is removed. In particular, a permuted dataset under the null distribution is obtained as follows: the graph $G = (V, E)$ and the sets $P_i, i \in V$ are the same as observed in the data; the values of the phenotype are randomly permuted across the samples. To estimate the p-value for the solutions obtained by our methods we used the following standard procedure: 1) we run an algorithm on the real data $\mathcal{D}$, obtaining a solution with objective function $o_{\mathcal{D}}$; 2) we generate N permuted datasets as described above; 3) we run the same algorithm on each permuted dataset; 4) the p-value is then given by $(e + 1)/(N + 1)$, where e is the number of permuted datasets in which our algorithm found a solution with objective function $\geq o_{\mathcal{D}}$.

Network permutation test: In the second permutation test, a permuted dataset under the null distribution is obtained by generating permuted networks (swapping edges to preserve the degree of nodes) while maintaining the same phenotype profile and gene alteration table. To generate each permuted network, we performed edge swapping $100 * |E|$ times. This permutation measures how likely a random network would have a module with the objective value at least the optimal. The test statistics used to compute p-values is again the value of the objective function of the solution and the p-value is calculated with same procedure described above for phenotype permutation test.

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Robustness test: To test the robustness of gene selection in modules, we use the bootstrapping method. More specifically, we sampled patients with replacement to generate random instances of the same number of samples. Let B_i be a random set of patients generated with bootstrapping in the i-th iteration. The phenotype and alteration profiles of the patients B_i were used as inputs to and the optimal solution O_i was computed with the random instances. We repeated bootstrapping 100 times to obtain $\{ B_1, B_2, ... B_{100}\}$, for which optimal solutions $\{ O_1, O_2, ... O_{100}\}$ were computed, respectively. The robustness of a gene (or an edge, resp.) in the optimal solution is obtained by counting the number of time the gene (pair of genes, resp.) appears in $\{ O_1, O_2, ... O_{100}\}$.

5 Conclusions

We developed a new computational method, NETPHIX (NETwork-to-PHenotpe mapping LeveragIng eXlusivity), for the identification of mutated subnetworks that are associated with a continuous phenotype. Using simulations and analyzing experimental data, we showed that NETPHIX can uncover the subnetworks associated with response to cancer drugs with high precision. Using NETPHIX to study drug response in cancer, we found many statistically significant and biologically relevant modules including two distinct MAPK/ERK signaling related modules associated with opposite response to drugs targeting MEK1/2 and ERK2 genes. We also demonstrated that subnetworks identified by NETPHIX can suggest combination drug therapy and guide personalized medicine.

The applicability of NETPHIX can go far beyond the drug response discussed in this paper, to any continuous cancer phenotypes. We expect that NETPHIX will find broad applications in many types of network-to-phenotype association studies.

Acknowledgements

This research was supported in part by the Intramural Research Programs of the National Library of Medicine at National Institutes of Health, USA. FV was supported, in part, by the University of Padova grants “SID2017” and ”STARS: Algorithms for Inferential Data Mining”. We would like to thank Jan Hoinka for helpful discussions.

References

[1] Douglas Hanahan and Robert A Weinberg. Hallmarks of cancer: the next generation. *cell*, 144(5):646–674, 2011.

[2] Levi A Garraway and Eric S Lander. Lessons from the cancer genome. *Cell*, 153(1):17–37, 2013.

[3] Bert Vogelstein, Nickolas Papadopoulos, Victor E Velculescu, et al. Cancer genome landscapes. *science*, 339(6127):1546–1558, 2013.

[4] Y. A. Kim, D. Y. Cho, and T. M. Przytycka. Understanding Genotype-Phenotype Effects in Cancer via Network Approaches. *PLoS Comput. Biol.*, 12(3):e1004747, Mar 2016.

[5] Y. A. Kim, R. Salari, S. Wuchty, and T. M. Przytycka. Module cover - a new approach to genotype-phenotype studies. *Pac Symp Biocomput*, pages 135–146, 2013.

[6] M. Hofree, J. P. Shen, H. Carter, A. Gross, and T. Ideker. Network-based stratification of tumor mutations. *Nat. Methods*, 10(11):1108–1115, Nov 2013.

[7] F. Vandin, P. Clay, E. Upfal, and B. J. Raphael. Discovery of mutated subnetworks associated with clinical data in cancer. *Pac Symp Biocomput*, pages 55–66, 2012.

[8] P. Dao, Y. A. Kim, D. Wojtowicz, et al. BeWith: A Between-Within method to discover relationships between cancer modules via integrated analysis of mutual exclusivity, co-occurrence and functional interactions. *PLoS Comput. Biol.*, 13(10):e1005695, Oct 2017.

[9] H. Y. Chuang, E. Lee, Y. T. Liu, D. Lee, and T. Ideker. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.*, 3:140, 2007.

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

[10] W. Zhang, J. Ma, and T. Ideker. Classifying tumors by supervised network propagation. *Bioinformatics*, 34(13):i484–i493, Jul 2018.

[11] H. Carter, M. Hofree, and T. Ideker. Genotype to phenotype via network analysis. *Curr. Opin. Genet. Dev.*, 23(6):611–621, Dec 2013.

[12] S. R. Gilman, J. Chang, B. Xu, et al. Diverse types of genetic variation converge on functional gene networks involved in schizophrenia. *Nat. Neurosci.*, 15(12):1723–1728, Dec 2012.

[13] W. Yang, J. Soares, P. Greninger, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.*, 41(Database issue):D955–961, Jan 2013.

[14] J. Barretina, G. Caponigro, N. Stransky, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, Mar 2012.

[15] Rebecca Sarto Basso, Dorit S. Hochbaum, and Fabio Vandin. Efficient algorithms to discover alterations with complementary functional association in cancer. In *Research in Computational Molecular Biology (RECOMB) 2018*, pages 278–279, 2018.

[16] J. Kim and et al. Characterizing genomic alterations in cancer by complementary functional associations. *Nature Biotechnology*, 34(5):539–546, May 2016.

[17] G. Ciriello, E. Cerami, B. A. Aksoy, C. Sander, and N. Schultz. Using MEMo to discover mutual exclusivity modules in cancer. *Curr Protoc Bioinformatics*, Chapter 8:Unit 8.17, Mar 2013.

[18] Y. A. Kim, D. Y. Cho, P. Dao, and T. M. Przytycka. MEMCover: integrated analysis of mutual exclusivity and functional network reveals dysregulated pathways across multiple cancer types. *Bioinformatics*, 31(12):i284–292, Jun 2015.

[19] Yoo-Ah Kim, Sanna Madan, and Teresa M Przytycka. Wesme: uncovering mutual exclusivity of cancer drivers and beyond. *Bioinformatics*, page btw242, 2016.

[20] Mark DM Leiserson, Hsin-Ta Wu, Fabio Vandin, and Benjamin J Raphael. Comet: a statistical approach to identify combinations of mutually exclusive alterations in cancer. *Genome biology*, 16(1):160, 2015.

[21] Simona Constantinescu, Ewa Szczurek, Pejman Mohammadi, Jörg Rahnenführer, and Niko Beerenwinkel. Timex: a waiting time model for mutually exclusive cancer alterations. *Bioinformatics*, page btv400, 2015.

[22] C. Li and H. Li. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*, 24(9):1175–1182, May 2008.

[23] P. Jia, S. Zheng, J. Long, W. Zheng, and Z. Zhao. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics*, 27(1):95–102, Jan 2011.

[24] C. A. Azencott, D. Grimm, M. Sugiyama, Y. Kawahara, and K. M. Borgwardt. Efficient network-guided multi-locus association mapping with graph cuts. *Bioinformatics*, 29(13):i171–179, Jul 2013.

[25] Y. Liu, M. Brossard, D. Roqueiro, et al. SigMod: an exact and efficient method to identify a strongly interconnected disease-associated module in a gene network. *Bioinformatics*, 33(10):1536–1544, May 2017.

[26] N. Stransky, M. Ghandi, G. V. Kryukov, et al. Pharmacogenomic agreement between two cancer cell line data sets. *Nature*, 528(7580):84–87, Dec 2015.

[27] J. Barretina, G. Caponigro, N. Stransky, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*, 483(7391):603–607, Mar 2012.

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Materials

S1 Algorithm for Sparse networks

The approach described in the main text works well for dense networks, see Section 2 for results on cancer data. For sparse networks imposing high density for the selected subnetwork might not be the best approach. Therefore we also propose a variation of the formulation presented in section 2.2 to handle sparse networks.

Given a graph $G = (V, E)$ and two distinct nodes h and l from V, a subset of nodes $N \subseteq V \backslash \{h, l\}$ is an (h, l) node separator if and only if after removing N from V there is no path between h and l in G. Let $\mathcal{N}(h, l)$ denote the family of all (h, l) node separators. A separator $N \in \mathcal{N}(h, l)$ is minimal if $N \backslash \{i\}$ is not an (h, l) separator for any $i \in N$.

As an alternative to constraint (6) in the formulation above one could impose the following connectivity constraint:

$$\sum_{i \in N} x_i \geq x_h + x_l - 1, \quad \forall h, l \in V, h \neq l, \forall N \in \mathcal{N}(h, l) \tag{8}$$

Constraint (8) ensure that for any pair of selected nodes h, l there is a path between them in the graph. An analogous constraint was used in [28, 29] and [30] shows that constraints (8) are facet defining for the connected subgraph polytope if N is a minimal (h, l) node separator. The NETPHIX package includes the implementation of this ILP as well. Constraints (8) are treated as lazy constraints and are only introduced when an integer solution that violates these inequalities is found. The branch and cut algorithm used is analogous to the one used in [28] and [29] and it involves finding a minimal node separator for nodes in the disjoint connected components of the found infeasible solution.

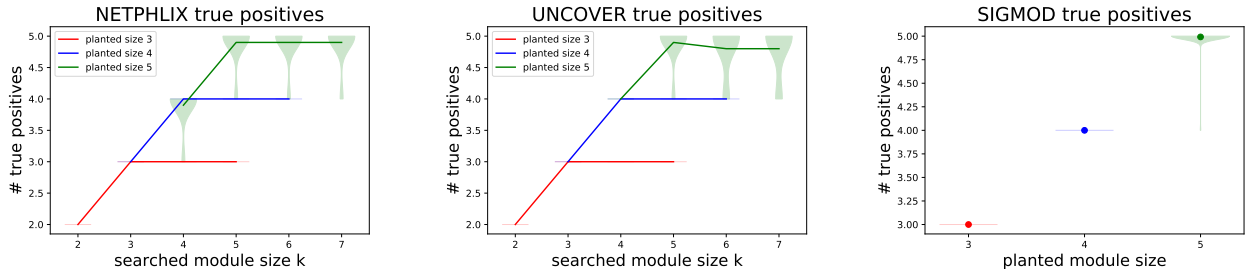
Supplementary References

[28] Matteo Fischetti, Markus Leitner, Ivana Ljubic, Martin Luipersbeck, Michele Monaci, Max Resch, Domenico Salvagnin, and Markus Sinnl. Thinning out steiner trees: a node-based model for uniform edge costs. *Math. Program. Comput.*, 9(2):203–229, 2017.

[29] Anna Bomersbach, Marco Chiarandini, and Fabio Vandin. An efficient branch and cut algorithm to find frequently mutated subnetworks in cancer. In *Algorithms in Bioinformatics - 16th International Workshop, WABI 2016, Aarhus, Denmark, August 22-24, 2016. Proceedings*, pages 27–39, 2016.

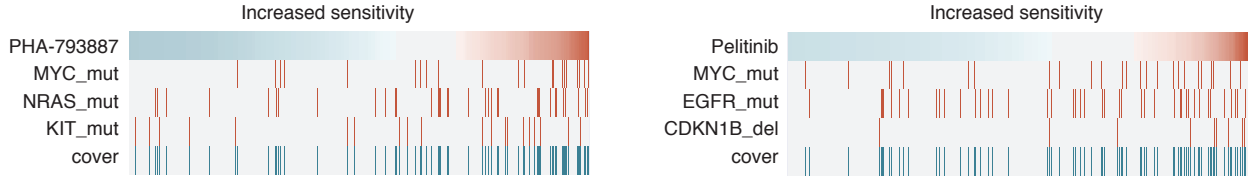
[30] Yiming Wang, Austin Buchanan, and Sergiy Butenko. On imposing connectivity constraints in integer programs. *Math. Program.*, 166(1-2):241–271, 2017.

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a) **(b)** **(c)**

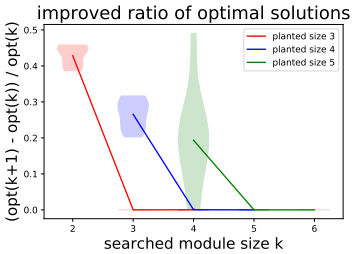
Figure S1: Method comparison on simulated data. The number of true positives in the modules identified by **(a)** NETPHIX, **(b)** UNCOVER, and **(c)** SigMOD. he solid lines and dots show the average number of false positives over all different parameters for each k and the shaded areas indicate the distribution over different instances.



(a) **(b)**

Figure S2: Subnetworks identified by for selected drugs. Each panel shows the values of the phenotype (top row)for all samples (columns), with blue being decreased sensitivity values and red being increased sensitivity values. For each gene in the subnetworks, alterations in each sample are shown in red, while samples not altered are in grey. The last row shows the alteration profile of the entire solution in blue. **(a)** and **(b)** Subnetworks with increased sensitivity to PHA793887 and Pelitinib, respectively.

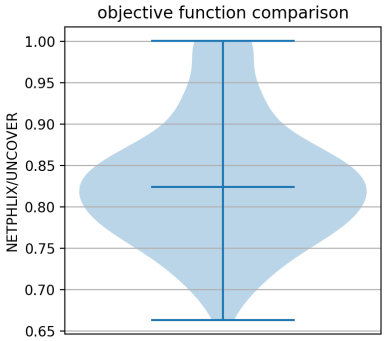
bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(a)

\protect \u

.06em\vbo



_running$_{t}ime.pdf$

(b)

(c)

Figure S3: Performance of NETPHIX. (a) The improvement of objective values over different k's in simulation. **(b)** The average running times of NETPHIX over different k's. **(c)** Comparison of the values of the objective function for NETPHIX's modules and UNCOVER's modules. Since NETPHIX includes additional constraints w.r.t. UNCOVER, the values of the objective function for its optimal solutions cannot be larger than the values of the objective function for UNCOVER's solutions. We display the distribution of the objective values for NETPHIX's modules as respective fractions of the UNCOVER objective values for the same instance.

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

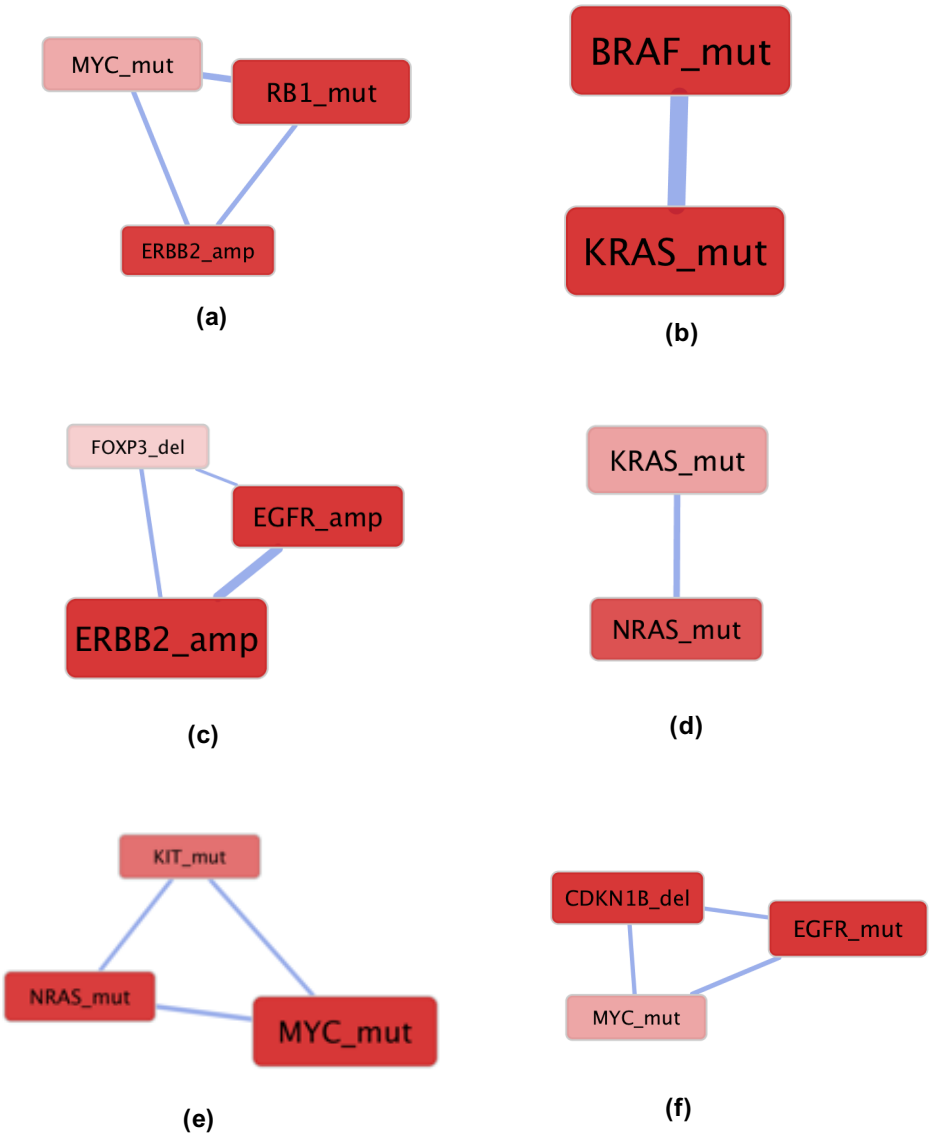


Figure S4: Network view of modules discussed in Section 3.3 and Figures 4 and S2. The size of the modules indicates the robustness of the genes in bootstrapping while the darkness of color represents their individual association scores. The thickness of edges shows the number of times the pair of genes appear in the same run of bootstrapping, i.e., how likely they appear together.

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S1: Sensitivity subnetworks for drugs in MAPK/ERK signaling pathway. Analysis of false positives in simulation studies indicatedthat false positive modules are often composed of an olfactory receptors' subnetwork. Human genome contains hundreds of olfactory receptor genes which are functionally related forming a giant densely connected subnetwork and since they are also subject to frequent mutations, olfactory receptors' subnetworks are expected to be false positives.

drug	id	increased sensitivity module	decreased sensitivity module	drug target	targeted pathway
PLX-4720	175	BRAF_mut	KRAS_mut, NRAS_mut, ERBB2_amp, IL2_del, RHOA_del	BRAF	ERK MAPK signaling
SB590885	199	BRAF_mut	none	BRAF	ERK MAPK signaling
AZ628	20	BRAF_mut, NRAS_mut, PTPN11_mut, KRAS_amp, PIK3CD_del	TCF7L2_mut, SMARCA4_mut, E2F5_mut, RB1_mut, CCND1_amp	BRAF	ERK MAPK signaling
PLX-4720	176	BRAF_mut	KRAS_mut, NRAS_mut, PIK3CD_mut, ERBB2_amp, RHOA_del	BRAF	ERK MAPK signaling
Dabrafenib	64	BRAF_mut,	PPARA_mut, RB1_mut, JUN_mut, DACH1_del, SMAD4_del	BRAF	ERK MAPK signaling
Selumetinib	203	BRAF_mut, KRAS_mut	MYC_mut, RB1_mut, ERBB2_amp	MEK1, MEK2	ERK MAPK signaling
VX-11e	244	BRAF_mut, KRAS_mut, NRAS_mut	RB1_mut, ERBB2_amp, CCND1_amp	ERK2	ERK MAPK signaling
Trametinib	231	BRAF_mut, KRAS_mut, NRAS_mut	LAMA3_mut, COL7A1_del	MEK1, MEK2	ERK MAPK signaling
CI-1040	54	BRAF_mut, KRAS_mut, NRAS_mut	MYC_mut, RB1_mut, ERBB2_amp	MEK1, MEK2	ERK MAPK signaling
PD0325901	160	BRAF_mut, KRAS_mut, NRAS_mut	MYC_mut, RB1_mut, ERBB2_amp	MEK1, MEK2	ERK MAPK signaling
Refametinib	187	BRAF_mut, KRAS_mut, NRAS_mut	MYC_mut, RB1_mut, ERBB2_amp	MEK1, MEK2	ERK MAPK signaling
AS605240	15	BRAF_mut, KRAS_mut, NRAS_mut	Olifactory Receptors	PI3Kgamma	PI3K/MTOR signaling
(5Z)-7-Oxozeaenol	1	BRAF_mut, NRAS_mut	MYC_mut, RB1_mut, ERBB2_amp, CDKN1B_del, RHOA_dell	TAK1	Other kinases
Afatinib	6	EGFR_amp, ERBB2_amp, FOXP3_del	BRAF_mut, KRAS_mut, NRAS_mut	ERBB2, EGFR	EGFR signalling
Pelitinib	162	MYC_mut, EGFR_mut, CDKN1B_del	BRAF_mut, RB1_mut, MAPK1_del	EGFR	EGFR signalling
PHA-793887	168	MYC_mut, NRAS_mut, KIT_mut	none	CDK2, CDK7, CDK5	cell cycle

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S2: All sensitivity subnetworks found by NETPHIX. List of 166 modules selected as significant (out of all 530 instances) and their statistics. Analysis of false positives in simulation studies indicated that false positive modules are often composed of an olfactory receptors' subnetwork. Human genome contains hundreds of olfactory receptor genes which are functionally related forming a giant densely connected subnetwork and since they are also subject to frequent mutations, olfactory receptors' subnetworks are expected to be false positives.

drug	id	sensitivity	k	optimal objective	selected module	p_{ph}	adj p_{ph} (BH)	p_{net}
(5Z)-7-Oxozeaenol	1	decreased	5	57.6	MYC_mut, RB1_mut, ERBB2_amp, CDKN1B_del, RHOA_del	0.0099	0.0557	0.7273
(5Z)-7-Oxozeaenol	1	increased	2	64.2	BRAF_mut, NRAS_mut	0.0099	0.0557	0.0198
Afatinib	6	decreased	2	33.9	KRAS_mut, NRAS_mut	0.0099	0.0557	0.0198
Afatinib	6	increased	3	78.3	EGFR_amp, ERBB2_amp, FOXP3_del	0.0099	0.0557	0.0891
Afatinib	7	decreased	3	48.9	BRAF_mut, KRAS_mut, NRAS_mut	0.0099	0.0557	0.0099
Afatinib	7	increased	3	68.4	MAPK8_mut, ERBB2_amp, NFATC1_del	0.0099	0.0557	0.9091
AICA Ribonucleotide	8	decreased	5	51.6	OR8H2_mut, OR4K13_mut, OR1M1_mut, OR10G8_del, OR4C15_del	0.0099	0.0557	0.4158
AKT inhibitor VIII	9	increased	3	59.4	PIK3CA_mut, ITGA1_mut, ERBB2_amp	0.0099	0.0557	0.9091
AKT inhibitor VIII	10	decreased	1	34.0	MXD1_amp	0.0198	0.0872	1
AR-42	13	decreased	3	44.7	FYN_mut, CTTN_amp, TEK_del	0.0198	0.0872	1
AR-42	13	increased	3	54.5	MYC_mut, NRAS_mut, IL2RB_mut	0.0099	0.0557	0.5455
AS605240	15	increased	3	66.6	BRAF_mut, KRAS_mut, NRAS_mut	0.0099	0.0557	0.0099
AT-7519	16	increased	3	56.1	MYC_mut, NRAS_mut, KIT_mut	0.0099	0.0557	0.1584
Axitinib	19	decreased	3	38.5	SPHKAP_mut, SPHK1_mut, KDSR_del	0.0099	0.0557	0.5455
AZ628	20	decreased	5	35.8	TCF7L2_mut, SMARCA4_mut, E2F5_mut, RB1_mut, CCND1_amp	0.0099	0.0557	0.1089
AZ628	20	increased	5	45.5	BRAF_mut, NRAS_mut, PTPN11_mut, KRAS_amp, PIK3CD_del	0.0099	0.0557	0.0099
AZD8055	24	decreased	5	54.4	OR4C46_mut, OR4K13_mut, OR5K1_mut, OR4N4_del, OR4C15_del	0.0099	0.0557	1
Belinostat	26	increased	3	49.9	MYC_mut, NRAS_mut, KIT_mut	0.0099	0.0557	0.3465
BIX02189	31	decreased	5	49.3	TCF7L2_mut, HDAC3_mut, RB1_mut, CCND1_amp, SMAD4_del	0.0099	0.0557	0.1287
Bleomycin (50 uM)	33	decreased	2	46.1	LAMA2_mut, AGRN_del	0.0099	0.0557	0.6364
BMS-536924	37	decreased	3	47.6	PTEN_mut, GRB7_amp, CAV1_amp	0.0099	0.0557	0.2178
BMS-754807	38	increased	1	38.5	KRAS_mut	0.0099	0.0557	1
Bosutinib	40	decreased	2	34.7	RB1_mut, TFDP1_del	0.0198	0.0872	0.1683
Bryostatin 1	41	decreased	5	45.0	DNMT1_mut, RB1_mut, JUN_mut, HDAC2_del, SMAD2_del	0.0099	0.0557	0.7273
BX-912	42	decreased	5	51.6	PLG_mut, HRG_mut, PROS1_mut, CLU_del, SERPINA1_del	0.0099	0.0557	1
BX-912	42	increased	3	52.7	MYC_mut, NRAS_mut, KIT_mut	0.0099	0.0557	0.0594
BX796	43	decreased	5	53.1	OR1S2_mut, OR6N2_mut, OR5K1_mut, OR5C1_mut, OR4C11_del	0.0099	0.0557	1
Camptothecin	45	decreased	2	39.7	OR4C15_mut, OR4C11_del	0.0099	0.0557	0.8182
Camptothecin	45	increased	3	53.9	CREBBP_mut, HIST1H1E_mut, KAT2B_del	0.0099	0.0557	0.0099
CAY10603	46	increased	3	57.1	MYC_mut, NRAS_mut, KIT_mut	0.0099	0.0557	0.2178
CCT-018159	47	decreased	4	46.6	ARHGEF17_mut, RHOBTB2_del, ARHGAP10_del, ARHGAP18_del	0.0198	0.0872	0.0099
CCT007093	48	increased	5	66.7	PLCG1_mut, VHL_mut, UBC_mut, USP33_mut, JAK2_del	0.0099	0.0557	0.0792
CGP-082996	50	increased	4	40.0	TAF1_mut, MYC_mut, TAF4B_del, TBP_del	0.0099	0.0557	0.0099
CHIR-99021	53	decreased	3	48.4	ASXL1_mut, RB1_mut, RBBP7_del	0.0099	0.0557	0.7273
CI-1040	54	decreased	3	49.9	MYC_mut, RB1_mut, ERBB2_amp	0.0099	0.0557	0.4653
CI-1040	54	increased	3	101.4	BRAF_mut, KRAS_mut, NRAS_mut	0.0099	0.0557	0.0099
Cisplatin	55	decreased	5	50.5	OR56A3_mut, OR8H2_mut, OR6N2_mut, OR4E2_del, OR4C15_del	0.0099	0.0557	0.9091
CP466722	57	increased	3	53.3	MYC_mut, NRAS_mut, KIT_mut	0.0099	0.0557	0.1287
CP724714	58	decreased	5	35.7	EPHA3_mut, TIAM1_mut, NRAS_mut, LYN_mut, RHOA_del	0.0198	0.0872	0.1176
CP724714	58	increased	3	74.3	EIF2AK2_mut, ERBB2_amp, STAT3_del	0.0099	0.0557	0.8182
CUDC-101	60	increased	3	53.0	MYC_mut, NRAS_mut, KIT_mut	0.0099	0.0557	0.4356
CX-5461	61	decreased	2	37.3	KDR_mut, PTK2B_del	0.0198	0.0872	0.0891
Dabrafenib	64	decreased	5	51.7	PPARA_mut, RB1_mut, JUN_mut, DACH1_del, SMAD4_del	0.0099	0.0557	0.8182
Dabrafenib	64	increased	1	111.9	BRAF_mut	0.0099	0.0557	1
Dacinostat	65	decreased	5	62.9	OR4C6_mut, OR5F1_mut, OR2T34_del, OR4C11_del, OR10G2_del	0.0099	0.0557	0.0495
Dactolisib	66	decreased	5	56.9	SIRT1_mut, RB1_mut, CITED2_mut, LPL_del, PPARA_del	0.0099	0.0557	1
Daporinad	67	decreased	2	41.0	SMG1_mut, PPP2R2A_del	0.0099	0.0557	0.5172
Docetaxel	70	decreased	3	47.1	SREBF2_mut, LSS_mut, HSD17B7_amp	0.0099	0.0557	0.9091
Doramapimod	71	decreased	1	31.1	SACS_mut	0.0198	0.0872	1
Doxorubicin	72	decreased	5	50.9	OR52H1_mut, OR5F1_mut, OR2T10_del, OR10G2_del, OR2A4_del	0.0198	0.0872	0.9091
Epothilone B	78	increased	3	45.8	ERCC6_mut, SMAD4_del, KAT2B_del	0.0198	0.0872	0.9091
Fedratinib	81	decreased	3	48.7	TIAM1_mut, SMAD4_del, CDH1_del	0.0099	0.0557	0.9091
Fedratinib	81	increased	3	50.5	MYC_mut, NRAS_mut, KIT_mut	0.0198	0.0872	0.0297
FR-180204	84	decreased	3	39.5	PTEN_mut, CCND1_amp, SPRY2_del	0.0099	0.0557	0.0297
Gefitinib	86	decreased	4	41.8	BRAF_mut, KRAS_mut, SPRY2_del, PTEN_del	0.0099	0.0557	0.0099
Genentech Cpd 10	88	decreased	3	47.9	BRAF_mut, MBP_del, MAPK1_del	0.0099	0.0557	0.5455
Genentech Cpd 10	88	increased	2	44.7	MAP3K5_mut, MINK1_del	0.0099	0.0557	0.0099
GSK269962A	93	decreased	3	43.6	STAT6_mut, SMAD4_del, PIAS4_del	0.0198	0.0872	0.8182
GSK429286A	95	decreased	3	40.5	KRAS_mut, CDKN2A_mut, CCND1_amp	0.0099	0.0557	0.1584
GSK650394	96	increased	3	51.5	CTNNB1_mut, PTPRB_mut, MYC_amp	0.0099	0.0557	0.3762
GSK690693	97	decreased	3	45.5	BRAF_mut, MBP_del, MAPK1_del	0.0099	0.0557	0.7273
GSK690693	97	increased	3	64.4	PTEN_mut, PSMD3_amp, CDKN1B_del	0.0099	0.0557	0.198
HG-5-88-01	102	decreased	5	37.3	PTEN_mut, MAPK8_mut, ERBB2_amp, CTNNB1_del, TP53_del	0.0099	0.0557	0.2079
I-BET-762	104	decreased	5	57.8	MTA3_mut, ERCC6_mut, CCND1_amp, SMAD4_del, KAT2B_del	0.0099	0.0557	1
I-BET-762	104	increased	3	63.7	MYC_mut, NRAS_mut, KIT_mut	0.0099	0.0557	0.1287
Idelalisib	105	decreased	3	37.1	AGAP2_mut, RB1_mut, CCND1_amp	0.0099	0.0557	0.0594
Idelalisib	105	increased	3	57.2	MYC_mut, CDC16_mut, CDKN1B_del	0.0099	0.0557	0.2277
Imatinib	106	decreased	3	19.0	APOB_mut, NRAS_mut, UBC_mut	0.0198	0.0872	0.3465
IPA-3	108	increased	5	66.5	MYC_mut, NRAS_mut, RAC1_mut, CCND1_mut, CDC42_mut	0.0099	0.0557	0.0198
Ispinesib Mesylate	109	increased	3	47.6	MAP3K5_mut, MINK1_del, USP9X_del	0.0198	0.0872	0.2376
JNK Inhibitor VIII	110	decreased	3	43.0	LAMA2_mut, MYC_mut, RAC1_mut	0.0099	0.0557	0.0198
JQ1	113	decreased	2	33.9	MDN1_mut, GTPBP4_del	0.0198	0.0872	0.0891
JW-7-24-1	115	decreased	3	52.7	STK11_mut, PARP1_mut, SMAD4_del	0.0099	0.0557	1
KIN001-244	118	decreased	5	50.7	OR2M2_mut, OR10R2_mut, OR13G1_mut, OR51B5_mut, OR4C15_del	0.0099	0.0557	0.7273
KIN001-260	119	decreased	3	44.7	PIK3CA_mut, CTNNB1_del, SMAD4_del	0.0099	0.0557	0.5455
KIN001-270	121	decreased	2	35.3	KRAS_mut, TEK_del	0.0198	0.0872	0.1386
KIN001-270	121	increased	2	47.9	MYC_mut, SMARCA2_mut	0.0099	0.0557	0.1287
Lapatinib	123	increased	3	37.9	NUP153_mut, CPSF3_mut, SMAD4_del	0.0099	0.0557	0.0594
Lenalidomide	124	decreased	1	30.4	DNAH8_mut	0.0099	0.0557	1
Lestauritinib	125	decreased	5	64.2	SMAD3_mut, LEF1_mut, WWTR1_mut, SMAD4_del, TFDP1_del	0.0099	0.0557	1

bioRxiv preprint doi: https://doi.org/10.1101/543876; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Linsitinib	128	decreased	5	49.6	PTEN_mut, PLCB1_mut, MAPK8_mut, GNAI1_mut, PTEN_del	0.0099	0.0557	0.1188
Linsitinib	128	increased	2	42.0	KRAS_mut, LYN_del	0.0198	0.0872	1
Methotrexate	131	decreased	2	34.1	PPP2R2A_del, RPL17_del	0.0198	0.0872	1
Midostaurin	133	decreased	3	43.3	KRAS_mut, RB1_mut, CCND1_amp	0.0099	0.0557	0.1089
Mitomycin-C	134	decreased	5	55.9	OR51D1_mut, OR10Q1_mut, OR4C11_del, OR10G2_del, OR14I1_del	0.0099	0.0557	0.0693
Navitoclax	138	decreased	3	43.5	BRAF_mut, AKT1_mut, SMAD4_del	0.0198	0.0872	1
NG-25	139	decreased	3	47.2	APC_mut, DACH1_del, SMAD4_del	0.0099	0.0557	0.0099
NSC-207895	142	decreased	5	60.1	KRAS_mut, ERBB2_amp, CCND1_amp, PSMB1_del, PSMD11_del	0.0099	0.0557	0.0297
Nutlin-3a (-)	145	decreased	5	71.0	MET_mut, RB1_mut, SMAD4_del, FOXP3_del, TP53_del	0.0099	0.0557	0.0099
NVP-BHG712	146	decreased	5	55.4	RB1_mut, JUN_mut, CCND1_amp, DACH1_del, SMAD4_del	0.0099	0.0557	0.0792
Olaparib	149	decreased	5	47.1	NCOA1_mut, POLR2A_mut, SMAD4_del, TP53_del, CCNK_del	0.0099	0.0557	0.1188
OSI-027	152	decreased	3	50.9	STK11_mut, TGIF1_mut, SMAD4_del	0.0099	0.0557	1
OSU-03012	154	increased	2	42.6	NOTCH1_mut, SEL1L_mut	0.0099	0.0557	0.1584
PAC-1	155	increased	5	61.7	MYC_mut, EGFR_mut, MAX_mut, ERBB2_amp, ERBB4_del	0.0099	0.0557	0.0198
Palbociclib	157	decreased	3	62.3	RB1_mut, KDM1A_mut, RB1_del	0.0099	0.0557	0.8182
Palbociclib	157	increased	1	63.3	MTAP_del	0.0099	0.0557	1
Pazopanib	159	decreased	3	42.8	PIK3CG_mut, KRAS_mut, RAF1_del	0.0099	0.0557	0.8182
PD0325901	160	decreased	3	50.0	MYC_mut, RB1_mut, ERBB2_amp	0.0099	0.0557	0.0198
PD0325901	160	increased	3	104.3	BRAF_mut, KRAS_mut, NRAS_mut	0.0099	0.0557	0.0099
PD173074	161	decreased	5	45.6	EP300_mut, MED24_amp, LPL_del, MED31_del, MED14_del	0.0099	0.0557	0.5455
Pelitinib	162	decreased	3	39.9	BRAF_mut, RB1_mut, MAPK1_del	0.0198	0.0872	0.1287
Pelitinib	162	increased	3	58.3	MYC_mut, EGFR_mut, CDKN1B_del	0.0099	0.0557	0.0396
PFI-1	166	decreased	1	36.5	RIMS2_mut	0.0198	0.0872	1
PHA-793887	168	increased	3	57.7	MYC_mut, NRAS_mut, KIT_mut	0.0099	0.0557	0.0792
PI-103	170	decreased	5	61.1	OR5W2_mut, MC3R_mut, OR10G2_del, GNB1_del, MC4R_del	0.0099	0.0557	1
Pictilisib	171	decreased	5	49.7	OR52A5_mut, OR4K13_mut, OR5K1_mut, OR11G2_mut, OR4M2_del	0.0099	0.0557	0.7273
Pictilisib	172	decreased	2	37.2	CHD9_mut, ALAS1_del	0.0099	0.0557	0.6364
PIK-93	173	decreased	1	35.1	TG_mut	0.0099	0.0557	1
Piperlongumine	174	decreased	4	56.4	DMD_mut, UBC_mut, ITGB1_del, RHOA_del	0.0099	0.0557	0.0792
PLX-4720	175	decreased	5	43.0	KRAS_mut, NRAS_mut, ERBB2_amp, IL2_del, RHOA_del	0.0099	0.0557	0.0099
PLX-4720	175	increased	1	100.2	BRAF_mut	0.0099	0.0557	1
PLX-4720	176	decreased	5	44.6	KRAS_mut, NRAS_mut, PIK3CD_mut, ERBB2_amp, RHOA_del	0.0198	0.0872	0.2353
PLX-4720	176	increased	1	108.0	BRAF_mut	0.0099	0.0557	1
Ponatinib	177	decreased	5	45.8	KRAS_mut, RB1_mut, ERBB2_amp, CCND1_amp, TFDP1_del	0.0099	0.0557	0.0326
QL-X-138	180	decreased	3	47.2	TG_mut, INS_del, SGK1_del	0.0099	0.0557	1
QL-XI-92	181	decreased	3	47.1	APC_mut, CCND1_amp, SMAD4_del	0.0099	0.0557	0.1683
Quizartinib	185	decreased	3	30.1	BRAF_mut, KRAS_mut, CCND1_amp	0.0198	0.0872	0.0099
Refametinib	187	decreased	3	54.6	MYC_mut, RB1_mut, ERBB2_amp	0.0099	0.0557	0.0198
Refametinib	187	increased	3	104.2	BRAF_mut, KRAS_mut, NRAS_mut	0.0099	0.0557	0.0099
Refametinib	188	decreased	3	51.5	MYC_mut, RB1_mut, ERBB2_amp	0.0099	0.0557	0.1485
Refametinib	188	increased	3	114.0	BRAF_mut, KRAS_mut, NRAS_mut	0.0099	0.0557	0.0099
SB-505124	196	decreased	1	34.7	MACF1_mut	0.0198	0.0872	1
SB216763	197	decreased	3	38.3	GRIN2A_mut, DVL1_mut, RHOA_del	0.0099	0.0557	0.1485
SB52334	198	decreased	3	45.2	EP300_mut, VHL_mut, NFATC1_del	0.0099	0.0557	1
SB590885	199	increased	1	90.7	BRAF_mut	0.0099	0.0557	1
Selumetinib	202	decreased	2	34.1	SMARCA4_mut, RB1_mut	0.0099	0.0557	0.0495
Selumetinib	202	increased	2	64.7	BRAF_mut, KRAS_mut	0.0099	0.0557	0.0198
Selumetinib	203	decreased	3	50.7	MYC_mut, RB1_mut, ERBB2_amp	0.0099	0.0557	0.0198
Selumetinib	203	increased	2	107.8	BRAF_mut, KRAS_mut	0.0099	0.0557	0.0198
Shikonin	207	decreased	4	49.3	ANK2_mut, ITPR2_amp, ITPR1_del, SNAP25_del	0.0198	0.0872	0.396
SN-38	209	decreased	5	55.7	OR2T6_mut, OR2M2_mut, OR2L8_mut, OR52B6_mut, OR4C16_del	0.0099	0.0557	1
SN-38	209	increased	3	53.6	CREBBP_mut, KAT2B_del, TAF1B_del	0.0099	0.0557	0.0198
STF-62247	212	decreased	3	38.1	AGAP2_mut, CCND1_amp, SMAD4_del	0.0099	0.0557	0.1485
Sunitinib	213	decreased	3	28.3	SETD2_mut, HDAC3_mut, RB1_mut	0.0099	0.0557	0.6364
T0901317	214	increased	3	55.2	MYC_mut, GATA1_mut, MMP2_del	0.0198	0.0872	0.1683
TAK-715	215	increased	3	54.5	KIAA1524_mut, MYC_mut, MYB_mut	0.0099	0.0557	0.3663
Talazoparib	216	decreased	5	53.9	OR56A3_mut, OR2T6_mut, OR10V1_mut, OR4C11_del, OR4E2_del	0.0099	0.0557	0.0396
Tanespimycin	218	decreased	3	54.1	MYC_mut, POLA1_mut, RB1_mut	0.0099	0.0557	0.0198
Tanespimycin	218	increased	3	51.2	PIK3CA_mut, NCK1_mut, TEK_del	0.0099	0.0557	0.3762
Temsirolimus	220	decreased	3	47.3	BDP1_mut, RB1_mut, TBP_del	0.0099	0.0557	0.5294
TGX221	221	increased	1	25.8	MRPS27_amp	0.0099	0.0557	1
THZ-2-102-1	223	increased	3	45.8	MYC_mut, PIAS2_mut, MYC_amp	0.0198	0.0872	0.5455
THZ-2-49	224	decreased	4	49.0	RB1_mut, MCM7_mut, CCND1_amp, CCNA1_del	0.0099	0.0557	0.6364
TL-1-85	227	decreased	3	44.9	APC_mut, DACH1_del, SMAD4_del	0.0099	0.0557	0.0693
TPCA-1	230	decreased	3	46.9	PTPRM_mut, CTNNB1_del, TCF4_del	0.0099	0.0557	0.9091
Trametinib	231	decreased	2	41.4	LAMA3_mut, COL7A1_del	0.0099	0.0557	0.1881
Trametinib	231	increased	3	100.6	BRAF_mut, KRAS_mut, NRAS_mut	0.0099	0.0557	0.0099
Tretinoin	232	decreased	3	40.1	PTEN_mut, PLCD3_mut, SMAD4_del	0.0099	0.0557	0.4059
Tubastatin A	233	decreased	5	49.6	TBL1X_mut, HDAC9_mut, WWTR1_mut, CCND1_amp, SMAD4_del	0.0099	0.0557	0.6364
Tubastatin A	233	increased	3	54.6	MYC_mut, NRAS_mut, KIT_mut	0.0198	0.0872	0.7273
TW 37	234	decreased	5	60.2	FPR1_mut, PRKACA_mut, ADCY7_del, EDN1_del, NFATC1_del	0.0099	0.0557	0.8182
UNC0638	235	decreased	3	50.1	PIK3CA_mut, RHOG_del, ARHGAP6_del	0.0198	0.0872	0.9091
Vinblastine	239	decreased	5	55.3	OR10A2_mut, OR1S2_mut, OR52K2_mut, OR51G1_mut, OR4C16_del	0.0198	0.0872	1
Vinblastine	239	increased	3	46.3	CREBBP_mut, PYGO1_mut, MYC_amp	0.0099	0.0557	0.0792
VNLG/124	242	decreased	3	33.6	SPTA1_mut, DCC_del, FYN_del	0.0099	0.0557	1
Vorinostat	243	decreased	5	58.0	OR51A4_mut, OR8B2_mut, OR5AS1_mut, OR4E2_del, OR4C15_del	0.0099	0.0557	1
VX-11e	244	decreased	3	41.1	RB1_mut, ERBB2_amp, CCND1_amp	0.0099	0.0557	0.4653
VX-11e	244	increased	3	99.2	BRAF_mut, KRAS_mut, NRAS_mut	0.0099	0.0557	0.0099
WIKI4	248	decreased	4	47.3	LEO1_mut, LEF1_mut, ELP3_del, CDC73_del	0.0099	0.0557	0.9091
WZ3105	250	increased	3	58.4	MAP3K5_mut, PARK7_del, MINK1_del	0.0099	0.0557	0.0792
XMD13-2	253	decreased	3	42.7	BRAF_mut, NGF_mut, NEDD4L_del	0.0198	0.0872	1
XMD14-99	254	decreased	2	34.3	EP300_mut, SMAD4_del	0.0099	0.0557	0.1683
XMD8-85	256	increased	3	31.7	MYC_mut, PRPF8_mut, HNRNPK_del	0.0198	0.0872	1
Y-39983	258	increased	3	56.4	MYC_mut, CRKL_mut, STAT5B_mut	0.0099	0.0557	0.7273
ZG-10	262	decreased	5	41.0	FUS_mut, SNRPA_mut, POLR2B_mut, POLR2F_del, CTDP1_del	0.0099	0.0557	0.0198
ZM447439	264	decreased	4	40.8	OR4M2_mut, OR6C6_mut, OR5D18_mut, OR4C16_del	0.0198	0.0872	1