MIT Open Access Articles

A mammalian functional-genetic approach to characterizing cancer therapeutics

Massachusetts Institute of Technology

A mammalian functional-genetic approach to characterizing cancer therapeutics

Hai Jiang[1], **Justin R Pritchard**[1], **Richard T Williams**[2], **Douglas A Lauffenburger**[1,3], and **Michael T Hemann**[1,*]

[1] The Koch Institute for Integrative Cancer Research at MIT, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

[2] Department of oncology, St. Jude Children's research Hospital, Memphis, Tennessee, USA

[3] Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA

Abstract

Identifying mechanisms of drug action remains a fundamental impediment to the development and effective use of chemotherapeutics. Here we describe an RNA interference (RNAi)-based strategy to characterize small-molecule function in mammalian cells. By examining the response of cells expressing short hairpin RNAs (shRNAs) to a diverse selection of chemotherapeutics, we could generate a functional shRNA signature that was able to accurately group drugs into established biochemical modes of action. This, in turn, provided a diversely sampled reference set for high-resolution prediction of mechanisms of action for poorly characterized small molecules. We could further reduce the predictive shRNA target set to as few as eight genes and, by using a newly derived probability-based nearest-neighbors approach, could extend the predictive power of this shRNA set to characterize additional drug categories. Thus, a focused shRNA phenotypic signature can provide a highly sensitive and tractable approach for characterizing new anticancer drugs.

Chemotherapy remains the frontline therapy for systemic malignancies. However, drug development has been severely hampered by an inability to efficiently elucidate mechanisms of drug action. This limits both the development of modified compounds with improved efficacy and the capability to predict mechanisms of drug resistance and select optimal patient populations for a given agent. Although drug-target interactions have traditionally been examined using biochemical approaches[1], a number of genetic strategies have been developed to identify pathways targeted by uncharacterized small molecules. A well-established genetic approach to drug classification is chemogenomic profiling in yeast[2–6]. In this approach, bar-coded yeast deletion strains are exposed to select agents, and genotype-dependent drug sensitivity is used to identify genes and pathways affected by a given drug,

as well as to develop a response signature that can be compared with other chemical or genetic perturbations[5,7,8]. This approach has proven quite powerful and has been broadly disseminated; however, its efficacy in interrogating cancer chemotherapeutics is limited by the lack of conservation of certain drug targets from yeast to mammals. This is a particular problem in the context of targeted therapeutics, which are frequently directed toward alterations that are specific to mammalian tumors.

More recently, genetic approaches have been developed to examine drug action in mammalian settings. One such approach is to examine drug response in a diverse panel of tumor cell lines[9]. In this case, the pattern of cell line sensitivity and resistance can serve as a signature that defines drug mechanism. Additionally, drug response can be correlated with the presence of specific cancer-related alterations, although this analysis can be confounded by the large diversity of alterations present in a given tumor. An alternative approach is to compare the global transcriptional changes induced by test compounds to those induced by known drugs or defined genetic alterations[10–13]. Here gene expression changes are used as signatures that are characteristic of exposure to a given agent or the presence of a specific cellular state, and common expression changes can be used to cluster similar small molecules. Although each of these approaches have yielded important new insights into drug action, these strategies retain a level of technical variability and resource requirement that limits both disseminated use and overall efficacy. Here we report a tractable RNAi-based approach that represents a simple yet powerful platform for drug screening and characterization.

RESULTS

Clustering drugs via shRNA-mediated phenotypes

We hypothesized that RNAi-mediated suppression of cell death regulators in mammalian cells would uniquely affect the cellular response to certain types of drugs and that drugs with similar mechanisms of action would elicit similar shRNA-dependent responses. To test this strategy, we started with a cell line derived from tumors from a well-established mouse model of Burkitt's lymphoma[14,15]. This cell line was chosen as an experimental system for two reasons. First, these cells are highly sensitive to a diverse set of chemotherapeutics, allowing small molecules to be used at pharmacologically relevant doses. Second, like many high-grade lymphomas, these cells undergo rapid apoptosis, as opposed to prolonged cell cycle arrest, following treatment. This common biological outcome after treatment allows for a systematic comparison of drugs.

In determining which genes to knock down for our studies, we chose two classes of genes known to be critical for cell fate decisions after drug treatment. The Bcl2 family of genes includes both central mediators and inhibitors of cell death, and different members of this gene family are involved in the response to distinct cell death stimuli[16]. The transcription factor p53 functions upstream of components of the Bcl2 family and is another important cell death regulator[17]. Mutation or deletion of p53 has been shown to affect the cellular response to many types of chemotherapeutic drugs[18,19]. As the stabilization and activity of p53 is strongly regulated by phosphorylation, we also targeted a panel of p53-activating kinases, including ATM, ATR, Chk1, Chk2, DNA-PKcs, Smg-1, JNK1 and p38 (refs. [20,21]). Importantly, aside from their roles as regulators of p53, these kinases are also involved in additional cellular responses to chemotherapy, such as DNA replication and repair, the activation of cell cycle checkpoints, regulation of RNA stability and stress signaling[22–26]. Thus, we generated shRNA vectors targeting the Bcl2 family, p53 and its activating kinases (Supplementary Results, Supplementary Fig. 1 and Supplementary Table 1).

To enable a quick and accurate analysis of how the suppression of a given gene affects drug-induced cell death, we used a single-cell flow cytometry-based GFP competition assay. Lymphoma cells were infected with retroviruses coexpressing a given shRNA and green fluorescent protein (GFP) and subjected to 72 h of drug treatment (Fig. 1a). In this assay, GFP-negative cells in the same population serve as an internal control. Using this approach, we systematically investigated how suppression of individual genes affected drug-induced cell death. As an initial proof of principle, we chose 15 chemotherapeutics representing major categories of anticancer drugs in clinical use. To compare different drugs using an objective criterion, all drugs were used at their LD80–90—a concentration at which 80–90% of uninfected lymphoma cells were killed (Supplementary Table 2). A control retrovirus lacking an shRNA or retroviruses expressing shRNAs targeting 29 genes were individually used to infect lymphoma cells. Each infected population was separately treated with 15 chemotherapeutic drugs, and the effect of a particular gene knockdown on therapeutic response was compiled as values of the GFP-determined 'resistance index' (RI) (Fig. 1b). Drugs with similar mechanism of action were expected to have similar patterns of genetic dependence on these 29 genes, which would manifest as similar patterns of RI values. To test this hypothesis in an unbiased manner, we used an unsupervised agglomerative hierarchical clustering approach to compare the RI values of different drugs (Fig. 1b). The significance of this hypothesis was then evaluated using a Monte Carlo principal components analysis-based method[27]. Notably, all 15 drugs tested in this initial experiment formed six distinct clusters that were consistent with their molecular mechanisms of action (Supplementary Fig. 2). Specifically, clear groupings were seen between topoisomerase II (TopoII) poisons doxorubicin (Dox) and etoposide (VP-16), DNA cross-linking agents cisplatin (CDDP), mitomycin C (MMC) and chlorambucil (CBL), single-strand break (SSB)-inducing agents camptothecin (CPT), 6-thioguanine (6-TG) and temozolomide (TMZ)[28,29], nucleic acid synthesis inhibitors methotrexate (MTX), 5-flurouracil (5-FU) and hydroxyurea (HU), and spindle poisons vincristine (VCR) and paclitaxel (Taxol). Taken together, these data showed that a simple comparison of drug response in cells expressing a small set of shRNAs could effectively categorize established chemotherapeutic drugs into subgroups that demarcate common target proteins and pathways.

To investigate whether this platform could be used to characterize mechanisms of drug action, we examined several recently developed chemotherapeutics: suberoylanilide,hydroxamic acid (SAHA), decitabine and roscovitine. Although the immediate biochemical targets of these new chemotherapeutics are known, the mechanisms of cell death induced by these drugs are less well defined. Using our RNAi-based approach, we compiled RI values for each of these three drugs and compared them with the 15 reference drugs mentioned earlier. We observed that the CDK inhibitor roscovitine (Rosco) was most similar to the RNA polymerase inhibitor actinomycin D (ActD) (Fig. 1c and Supplementary Fig. 3a). This is consistent with the findings of several studies showing that roscovitine inhibits CDK7, a component of the general transcription factor TFIIH, to inhibit RNA transcription[30–32]. Notably, the HDAC inhibitor SAHA and the DNA methyltransferase inhibitor decitabine (DAC) formed a distinct cluster outside of the 15 reference drugs (Fig. 1c), suggesting that these two drugs may share a similar mechanism of cell death. To extract the most relevant genes for distinguishing the SAHA-DAC cluster, shRNAs were ranked upon their ability to classify this cluster relative to the rest of the dataset. The most unique aspects of the new SAHA-DAC cell death signature were the (i) p53-independence ($\log_2 RI \approx 0$) and (ii) Bim-dependence ($\log_2 RI \approx 2$) of cell death, consistent with previous studies of SAHA treatment in mouse lymphoma models[33]. Indeed, both SAHA and DAC treatment resulted in an increase in the levels of the proapoptotic BH3-only protein Bim (Supplementary Fig. 3b). Furthermore, suppression of the Bim transcription regulator Chop, but not Foxo3a, resulted in resistance to both SAHA and DAC

(Fig. 1d). Thus, the RI patterns of these newly established drugs could effectively identify their mechanism of action.

Functional characterization of derivatized compounds

A significant challenge in drug development is determining whether lead compound derivatives with enhanced efficacy share the same mechanism of action as the original small molecule. Theoretically, derivatized compounds could show enhanced efficacy, owing to either the activation of additional cell death pathways or, alternatively, through altered pharmacodynamic properties. To examine whether our approach could be used to differentiate between these possibilities, we performed an shRNA-based functional analysis of CY190602, a chemical derivative of the nitrogen mustard bendamustine (Fig. 2a). Compared to the parental drug, CY190602 shows approximately 20–100-fold enhanced toxicity toward cell lines from patients with multiple myeloma (Fig. 2b), an indication for which bendamustine is currently in clinical use. However, the mechanism underlying this increase in cytotoxicity remains unclear. Notably, CY190602's modification on bendamustine occurs on a side chain well away from the nitrogen mustard functional group. To address whether CY190602's toxicity could still be attributed to the nitrogen mustard or whether it was a result of altered target specificity caused by the side chain modification moieties, we compiled the RI values of bendamustine and CY190602 and compared them to those of our 18 reference drugs. Notably, bendamustine and CY190602 showed highly similar patterns of RI values (Fig. 2c), despite a 100-fold-lower dose of CY190602. Additionally, both drugs clustered together with chlorambucil, another nitrogen mustard (Fig. 2d), and a supervised K-nearest-neighbors approach (see Supplementary Methods for a detailed rationale) predicted a chlorambucil-like mechanism for both drugs. This suggests that the primary mode of action of CY190602 is nitrogen mustard-mediated DNA damage rather than an off-target effect conferred during drug optimization.

Screening for compounds on the basis of shRNA signatures

Next, we asked whether this approach could be adapted to phenotype-based screens for new drug candidates without well-established mechanisms of action. Suppression of ATM, Chk2 and p53 all led to significant resistance to genotoxic drugs such as Dox, VP-16, CPT, TMZ, 6TG, CDDP, MMC and CBL (Fig. 1b). This suggested that the *shATM-Chk2-p53* 'resistance signature' might be used to identify genotoxic drugs. To test this hypothesis quantitatively, we examined whether a supervised K-nearest-neighbors approach could accurately characterize all of the drugs in our dataset as either genotoxic or nongenotoxic. Indeed, when a broad panel of chemotherapeutic drugs was tested, all 16 genotoxic chemotherapeutics, but none of 15 nongenotoxic chemotherapeutics, showed a distinct *shATM-Chk2-p53* resistance signature (Fig. 3a). This three-gene resistance signature was subsequently used to screen a chemical library for genotoxic compounds. Two compounds, apigenin and NSC3852, were identified on the basis of their strong *shATM-Chk2-p53* resistance signature (Fig. 3b). We then compiled the full 29-gene RI values for these two compounds and compared them with reference drugs (Fig. 3c). Notably, the K-nearest-neighbors approach predicted apigenin to be most similar to the TopoII poisons doxorubicin and etoposide and NSC3852 to be most like the SSB-inducing agents. Subsequent clustering showed NSC3852 to be most similar to the topoisomerase I (TopoI) poison camptothecin. Our previous studies demonstrated that TopoII poisons are ineffective in killing TopoII-deficient cells, while showing enhanced toxicity for cells lacking TopoI[34]. Consistent with the clustering-based functional predictions, apigenin showed a pattern of shTopoII resistance and shTopoI sensitivity similar to the established TopoII poisons doxorubicin, etoposide and mitoxantrone (Fig. 3d). Conversely, NSC3852 showed a characteristic pattern of resistance, similar to established TopoI poisons camptothecin and irinotecan (CPT11). Notably, none of the other genotoxic drugs showed these resistance and sensitivity patterns with shTopoI and

shTopoII (Fig. 3d). We also found that apigenin and NSC3852 failed to induce DNA damage in TopoII-and TopoI-deficient cells, respectively (Supplementary Fig. 4). Moreover, in a long-term survival assay, TopoII deficiency resulted in significant protection from apigenin, whereas TopoI deficiency significantly protected cells from NSC3852 (Fig. 3e). Taken together, these assays confirmed our classification of apigenin and NSC3852 as TopoII and TopoI poisons, respectively. Thus, small shRNA signatures can be used to screen chemical libraries to identify and characterize new compounds with particular target specificities.

An eight-shRNA set for accurate drug mechanism predication

Given that a three-gene signature could effectively predict and classify genotoxic drugs, we hypothesized that the combined resistance and sensitivity pattern of a small number of genes may be sufficient to accurately characterize most of our chemotherapeutic drugs in this cell line. To test this hypothesis, we examined the seven drug clusters demarcated in our secondary analysis (Fig. 1c) and asked which smaller sets of shRNAs could similarly define these groupings. Here we used a K-nearest-neighbors cross-validation-based approach and a randomized search through 50,000 potential gene subsets. Although most smaller shRNA sets showed a significant loss in resolution relative to the reference set, we found that a set of eight shRNAs, targeting p53, ATR, Chk1, Chk2, Smg-1, DNA-PKcs, Bok and Bim, was able to classify the reference dataset with 100% accuracy and was highly correlated ($r^2 = 0.81$) with the original 29 shRNA signature (Fig. 4a and Supplementary Fig. 5a,b). Although several other sets of eight shRNAs could also classify chemotherapeutics with 100% accuracy, this eight-shRNA signature had the highest range of measurement across all drugs. Notably, this eight-shRNA signature could also correctly classify bendamustine, CY190602, apigenin and NSC3852—drugs that were not included in the feature reduction and cross-validation of the eight-shRNA signature (Supplementary Fig. 5c).

Given the known off-target potential of RNAi, we next sought to determine whether the functional signature derived from these eight shRNAs was attributable to the specific effect of shRNA target gene suppression on therapeutic response. To do this, we used a second set of shRNAs targeting the same eight genes to generate an independent drug response signature. Comparison of shRNA pairs revealed a high correlation between drug response signatures ($r^2 = 0.86$) in cells transduced with distinct shRNAs targeting the same gene, suggesting that the major effects of these shRNAs are 'on target' (Fig. 4b). Additionally, unsupervised hierarchical clustering of the first eight-shRNA response signature or the combined response signatures generated using the first and second eight-shRNA sets revealed the same seven drug classes identified with the original 29-shRNA signature (Fig. 4c). Notably, however, the second set of eight shRNAs could independently predict only five out of seven drug classes. This loss of resolution in the second shRNA set may represent trace 'off-target' shRNA activity in either eight-shRNA set. Alternatively, these differences may be attributable to small differences in the degree of target gene knockdown conferred by distinct hairpins. Consistent with the latter argument, shRNAs in the second set frequently showed reduced target gene suppression (Supplementary Fig. 1 and Supplementary Table 1) and yielded more subtle biological effects, as evidenced by the relative RI values seen in shRNA pairwise comparisons (Fig. 4b).

To extend our eight-shRNA signature approach in a scalable and stringent manner, we revisited a common problem in machine learning. A nonparametric classification method like K-nearest-neighbors will classify any test compound according to its closest neighbor(s), even if the two compounds are quite distinct. Thus, it becomes difficult to determine how distantly a given compound can reside from a reference category of drugs and still be considered to share a similar mechanism of action (Fig. 5a). To overcome this problem, we took advantage of the carefully selected mechanistic diversity of our training

set to create specific empirical cumulative distribution functions for each drug category (Fig. 5b and c). This allowed us to determine whether a test compound was likely to belong to either an existing or a new drug category—a process critical to the broader applicability of this approach.

To determine whether this methodology could correctly categorize chemotherapeutics absent from our initial reference set, we examined a set of 16 additional anticancer drugs (Table 1 and Supplementary Fig. 6). In each case, the eight-shRNA approach successfully grouped drugs according to their mechanism of action. Importantly, when compounds that represent new drug categories were examined, they were not misclassified into the 'nearest' drug category. Rather, they were identifiable as distinct agents that were significantly different from all other drug categories. Consequently, although this eight-shRNA panel was assembled on the basis of responses to seven drug classes, it was also successful in predicting other classes of chemotherapeutics when the training set was updated with new reference compounds. For example, the eight-shRNA signature accurately predicted that the proteasome inhibitor gliotoxin belonged to a drug category not represented by any of the existing reference drugs. However, when the proteasome inhibitors bortezomib (PS341) and MG132 were used to update the training set, the eight-shRNA signature was able to successfully classify gliotoxin and epoxomycin as proteasome inhibitors (Table 1). The eight-shRNA set could be similarly trained to identify two entirely distinct drug categories —Hsp90 inhibitors and EGFR inhibitors— neither of which was used to create the eight-shRNA reference set. Notably, the eight-shRNA signature could also distinguish functional drug subclasses within larger targeted classes of therapeutics. For example, the HER2 inhibitors lapatinib and AEE788 and the multikinase inhibitor sunitinib clustered in distinct categories relative to EGFR inhibitors (Supplementary Fig. 7), despite all of these drugs belonging to the broader category of tyrosine kinase inhibitors. Although the use of more optimized sets of shRNAs may be necessary to probe fine details of certain drug categories, these data suggest that this eight-shRNA set has resolution over a broad range of cytotoxic activities.

Although the cells used in this study are responsive to a number of targeted chemotherapeutics, such as EGFR inhibitors, a potential limitation of this approach is that it lacks resolution for certain compounds requiring cellular targets not present in lymphoma cells. To determine whether this approach could be adapted to cell lines expressing targetable genetic lesions, we examined the performance of the eight-shRNA signature in cells derived from a BCR-Abl-driven model of acute B cell leukemia (B-ALL)[35]. Strikingly, a robust functional signature for alkylating agents could be generated in these cells using the same eight-shRNA set (Fig. 6). Notably, however, the response signature in B-ALL cells differed from that in lymphoma cells. For example, leukemia cells showed distinct genetic dependencies on ATR, DNA-PKcs and Bok. Thus, informative signatures can be derived in distinct cell lines, even if the signatures differ between cell types. Notably, this eight-shRNA signature may not be optimal for B-ALL cells, as feature reduction from the 29-shRNA signature was not performed in this context. Additionally, this signature may not have the same resolution as in lymphoma cells. However, these data suggest that even suboptimal signatures may provide resolution sufficient to cluster classes of chemotherapeutics.

Discussion

The functional genetic approach described here has similarities to well-characterized chemogenomic profiling strategies in lower organisms. However, this approach also has notable advantages over existing genetic approaches for examining drug mechanisms of action and identifying drug targets. First, this approach is sufficiently sensitive to differentiate drugs with distinct targets but common downstream signaling pathways. For

example, TopoI and II poisons produce distinct shRNA sensitivity profiles, yet both ultimately engage common transcriptional networks. Microarray approaches that focus on downstream changes in gene expression are, consequently, less able to distinguish between conventional anticancer agents. In fact, previous microarray studies have shown limited resolution over a number of frontline chemotherapeutics (Supplementary Table 3). Second, this approach is unaffected by pharmacodynamic variability, such as distinctions in drug efflux or detoxification, that obscures comparisons between different cancer cell lines. Finally, and most importantly, this approach is both simple and tractable. Although microarray studies suffer from significant variability between experiments and laboratories, RNAi-based functional arrays are highly reproducible and can be widely disseminated.

Perhaps the most unanticipated aspect of this work lies in the quantity of information that can be derived from a small set of mammalian loss-of-function phenotypes. This focused shRNA signature can characterize a diverse range of drug categories at high resolution and is extendable to completely new drug categories and distinct cell types, suggesting that such signatures might serve as a tractable approach to screen chemical libraries for diverse functional classes of small molecules in a high-throughput manner. Although this specific set of shRNAs may not provide optimal resolution for all cell types or small molecules, these data also suggest that alternative small sets of shRNAs may yield similar information content. For example, although this work focuses on cell viability, it is likely that—given appropriate phenotypic resolution—bioactive compounds affecting diverse aspects of biology can similarly be interrogated with distinct targeted sets of shRNAs.

METHODS

Cell lines and drugs

Eμ-Myc p 19$^{Arf-/-}$ mouse lymphoma cells were cultured in B cell medium as described[15]. MM1S and RPMI8226 cells were cultured in RPMI medium supplemented with glutamate and 10% (v/v) FBS. Drugs were obtained from Sigma, Tocris, Calbiochem, VWR, LC Laboratories and other suppliers. shRNA vectors were generated as described[36,37]. *p185+ p19$^{Arf-/-}$* acute lymphoblastic leukemia cells were derived and cultured according to the procedures outlined in ref. [35].

Drug treatment and flow cytometry

Eμ-Myc pl9$^{Arf-/-}$ cells were counted and seeded at 1 million cells per ml in 48-well plates and treated with various concentrations of drugs. To approximate therapeutic situations in which drug dose decreases over time, half of the volume from each experiment was removed and replenished with fresh medium every 24 h. Cells were analyzed by fluorescence-activated cell sorting (FACS), with propidium iodide as a viability marker. LD80–90 of drugs are defined as concentrations at which the lowest viability reading out of three FACS time points (24, 48 and 72 h) is between 10% and 20%. After we determined drug dose, *Eμ-Myc p19$^{Arf-/-}$* cells were infected with retroviruses encoding shRNAs targeting particular genes. Individual infected cell populations were counted and seeded at 1 million cells per ml in 48-well plates and treated with drugs using the aforementioned protocol. At 72 h, treated and untreated cells were analyzed by flow cytometry. GFP percentages of live (PI-negative) cells were recorded and used to calculate relative resistance index. To avoid outgrowth of untreated control cells, we typically seeded them at 0.25 million per ml, and 75% of medium was replaced at 24 and 48 h.

Calculation of relative resistance index (*X*)

To compare the relative level of chemoresistance and sensitization conferred by each gene knockdown, we introduced the concept of RI (see definition above), to more accurately

analyze the GFP competition results. We define the value of RI as *X*. The biological meaning of this factor *X* is that in a mixture of uninfected and infected (knockdown) cells, the infected (knockdown) cells will be *X*-fold as likely to survive drug treatment when compared to uninfected cells. By our definition of *X*, if one out of *n* uninfected cells survives a drug treatment, then *X* our of *n* infected cells should survive. If we define the total number of uninfected and infected cells as *T* and the GFP percentage of untreated population as G1, then the number of surviving, uninfected cells (un) can be defined as $n - \text{un} = T \times (1 - \text{G1}) \times 1/n$, and the number of surviving, infected cells (in) can be defined as $n - \text{in} = T \times \text{G1} \times X/n$. Hence, the GFP percentage of the treated, surviving population (G2) can be calculated as $\text{G2} = (n - \text{in})/((n - \text{un}) + (n - \text{in}))$. From this equation, we can derive that $X = (\text{G2} - \text{G1 x G2})/(\text{G1} - \text{G1 x G2})$. This equation was used in our studies to compute RI values.

Enhanced K-nearest-neighbors methods

K-nearest-neighbors modeling is a weighted-voting methodology in which the proximity to the training set is used to predict drug class membership. We include this analysis for four reasons. (i) It provides independent validation of the clustering result. (ii) It allows us to quantify the predictive power of the reference set through leave-one-out cross-validation. (iii) Leave-one-out cross-validation allows us to perform a feature reduction to discover smaller gene sets. (iv) It provides an objective prediction of classes for new compounds.

K-nearest-neighbors predictions were performed using a correlation-based metric and a consensus voting scheme. The MATLAB knnclassify.m function was used as a basis for the feature reduction search, as well as cross-validation and predictions. The cross-validation for the K-nearest-neighbors approach was done by systematically leaving out one of the 18 drugs at a time in the final dataset (Fig. 1c) and using the remaining 17 to predict the left-out drugs' identities.

To reduce the size of the feature set to a smaller group of key shRNAs, we randomly searched a subset of 2,000 unique shRNA sets of increasing size. Sampled subsets were scored on the basis of their ability to cross-validate. We then performed a much more extensive search (> 50,000 subsets) of eight shRNA signatures that would be able to correctly classify all of the drugs in our reference set. The shRNA subsets that cross-validated at 100% were then ranked by their least-squares correlation with the distances between drugs in the 29-shRNA signature, and the eight-shRNA set with the highest correlation score was chosen for later experiments.

A K-nearest-neighbors-based approach will always yield a prediction of drug class on the basis of proximity. Therefore, to evaluate the similarity of a new drug to its predicted class we developed a linkage ratio *p*-value test. Briefly, we calculated the initial cluster size of each of the seven drug groups (Fig. 1c) by evaluating the average of all pairwise linkage distances amongst all members of a drug group. When a test compound was predicted to belong to a drug group on the basis of proximity, then the cluster size of that particular drug group was calculated again with the new test drug included. A linkage ratio was then calculated by comparing the cluster size with and without the tested compound. A linkage ratio of less than one indicated that the addition of the drug to a cluster made the average distance between drugs in that category smaller, whereas a linkage ratio greater than one indicated that the cluster expanded. An obvious tradeoff exists between cluster expansion to accommodate modestly distinct compounds with highly homologous mechanisms and expanding the definition to a point where one masks the existence of a completely new compound. This tradeoff varies among drug classes as a function of the inter-class distances. To estimate the significance of a K-nearest-neighbors prediction, as well as to determine whether a compound had a mechanism of action different from those of our original seven drug groups, we sampled the negative control distributions of drug classifications. This was

done on a class-by-class basis by taking the previously studied compounds and forcing them to erroneously classify. We then calculated a linkage ratio for all of these erroneous classifications. On a class-by-class basis we fit a normal distribution to the range of misclassified linkage ratios. The value of the cumulative distribution function was used to calculate the p-value of the new classifications (Fig. 5c), using the null hypothesis that the linkage ratio for a prediction is identical to the linkage ratios of the negative control distribution. The complete MATLAB algorithm used to perform this analysis is provided as the "Drug Prediction Score.M" file found at http://web.mit.edu/icbp/data/index.html.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Sato S, Murata A, Shirakawa T, Uesugi M. Biochemical target isolation for novices: affinity-based strategies. Chem Biol. 2010; 17:616–623. [PubMed: 20609411]

2. Giaever G, et al. Genomic profiling of drug sensitivities via induced haploinsufficiency. Nat Genet. 1999; 21:278–283. [PubMed: 10080179]

3. Giaever G, et al. Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. Proc Natl Acad Sci USA. 2004; 101:79–98.

4. Lum PY, et al. Discovering modes of action for therapeutic compounds using a genome-wide screen of yeast heterozygotes. Cell. 2004; 116:121–137. [PubMed: 14718172]

5. Parsons AB, et al. Integration of chemical-genetic and genetic interaction data links bioactive compounds to cellular target pathways. Nat Biotechnol. 2004; 22:62–69. [PubMed: 14661025]

6. Hillenmeyer ME, et al. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. Science. 2008; 320:362–365. [PubMed: 18420932]

7. Parsons AB, et al. Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. Cell. 2006; 126:611–625. [PubMed: 16901791]

8. Hillenmeyer ME, et al. Systematic analysis of genome-wide fitness data in yeast reveals novel gene function and drug action. Genome Biol. 2010; 11:R30. [PubMed: 20226027]

9. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. Nat Rev Cancer. 2006; 6:813–823. [PubMed: 16990858]

10. Hughes TR, et al. Functional discovery via a compendium of expression profiles. Cell. 2000; 102:109–126. [PubMed: 10929718]

11. Gardner TS, di Bernardo D, Lorenz D, Collins JJ. Inferring genetic networks and identifying compound mode of action via expression profiling. Science. 2003; 301:102–105. [PubMed: 12843395]

12. Lamb J, et al. The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. Science. 2006; 313:1929–1935. [PubMed: 17008526]

13. Hieronymus H, et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. Cancer Cell. 2006; 10:321–330. [PubMed: 17010675]

14. Adams JM, et al. The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. Nature. 1985; 318:533–538. [PubMed: 3906410]

15. Schmitt CA, McCurrach ME, de Stanchina E, Wallace-Brodeur RR, Lowe SW. INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. Genes Dev. 1999; 13:2670–2677. [PubMed: 10541553]

16. Youle RJ, Strasser A. The BCL-2 protein family: opposing activities that mediate cell death. Nat Rev Mot Cell Biol. 2008; 9:47–59.

17. Lu C, El-Deiry WS. Targeting p53 for enhanced radio-and chemosensitivity. Apoptosis. 2009; 14:597–606. [PubMed: 19259822]

18. Lowe SW, Ruley HE, Jacks T, Housman DE. p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. Cell. 1993; 74:957–967. [PubMed: 8402885]

19. Lowe SW, et al. p53 status and the efficacy of cancer therapy in vivo. Science. 1994; 266:807–810. [PubMed: 7973635]

20. Bode AM, Dong Z. Post-translational modification of p53 in tumorigenesis. Nat Rev Cancer. 2004; 4:793–805. [PubMed: 15510160]

21. Brumbaugh KM, et al. The mRNA surveillance protein hSMG-1 functions in genotoxic stress response pathways in mammalian cells. Mol Cell. 2004; 14:585–598. [PubMed: 15175154]

22. Lavin MF. Ataxia-telangiectasia: from a rare disorder to a paradigm for cell signalling and cancer. Nat Rev Mol Cell Biol. 2008; 9:759–769. [PubMed: 18813293]

23. Cimprich KA, Cortez D. ATR: an essential regulator of genome integrity. Nat Rev Mol Cell Biol. 2008; 9:616–627. [PubMed: 18594563]

24. Bartek J, Lukas J. Chk1 and Chk2 kinases in checkpoint control and cancer. Cancer Cell. 2003; 3:421–429. [PubMed: 12781359]

25. Reinhardt HC, Aslanian AS, Lees JA, Yaffe MB. p53-deficient cells rely on ATM-and ATR-mediated checkpoint signaling through the p38MAPK/MK2 pathway for survival after DNA damage. Cancer Cell. 2007; 11:175–189. [PubMed: 17292828]

26. Pearce AK, Humphrey TC. Integrating stress-response and cell-cycle checkpoint pathways. Trends Cell Biol. 2001; 11:426–433. [PubMed: 11567876]

27. Pritchard JR, et al. Three-kinase inhibitor combination recreates multipathway effects of a geldanamycin analogue on hepatocellular carcinoma cell death. Mol Cancer Ther. 2009; 8:2183–2192. [PubMed: 19671754]

28. Swann PF, et al. Role of postreplicative DNA mismatch repair in the cytotoxic action of thioguanine. Science. 1996; 273:1109–1111. [PubMed: 8688098]

29. Mojas N, Lopes M, Jiricny J. Mismatch repair-dependent processing of methylation damage gives rise to persistent single-stranded gaps in newly replicated DNA. Genes Dev. 2007; 21:3342–3355. [PubMed: 18079180]

30. Akhtar MS, et al. TFIIH kinase places bivalent marks on the carboxyterminal domain of RNA polymerase II. Mol Cell. 2009; 34:387–393. [PubMed: 19450536]

31. Ljungman M, Paulsen MT. The cyclin-dependent kinase inhibitor roscovitine inhibits RNA synthesis and triggers nuclear accumulation of p53 that is unmodified at Ser15 and Lys382. Mol Pharmacol. 2001; 60:785–789. [PubMed: 11562441]

32. MacCallum DE, et al. Seliciclib (CYC202, R-Roscovitine) induces cell death in multiple myeloma cells by inhibition of RNA polymerase II-dependent transcription and down-regulation of Mcl-1. Cancer Res. 2005; 65:5399–5407. [PubMed: 15958589]

33. Lindemann RK, et al. Analysis of the apoptotic and therapeutic activities of histone deacetylase inhibitors by using a mouse model of B cell lymphoma. Proc Natl Acad Sci USA. 2007; 104:8071–8076. [PubMed: 17470784]

34. Burgess DJ, et al. Topoisomerase levels determine chemotherapy response in vitro and in vivo. Proc Natl Acad Sci USA. 2008; 105:9053–9058. [PubMed: 18574145]

35. Williams RT, Roussel MF, Sherr CJ. Arf gene loss enhances oncogenicity and limits imatinib response in mouse models of Bcr-Abl-induced acute lymphoblastic leukemia. Proc Natl Acad Sci USA. 2006; 103:6688–6693. [PubMed: 16618932]

36. Dickins RA, et al. Probing tumor phenotypes using stable and regulated synthetic microRNA precursors. Nat Genet. 2005; 37:1289–1295. [PubMed: 16200064]

37. Jiang H, et al. The combined status of ATM and p53 link tumor development with therapeutic response. Genes Dev. 2009; 23:1895–1909. [PubMed: 19608766]
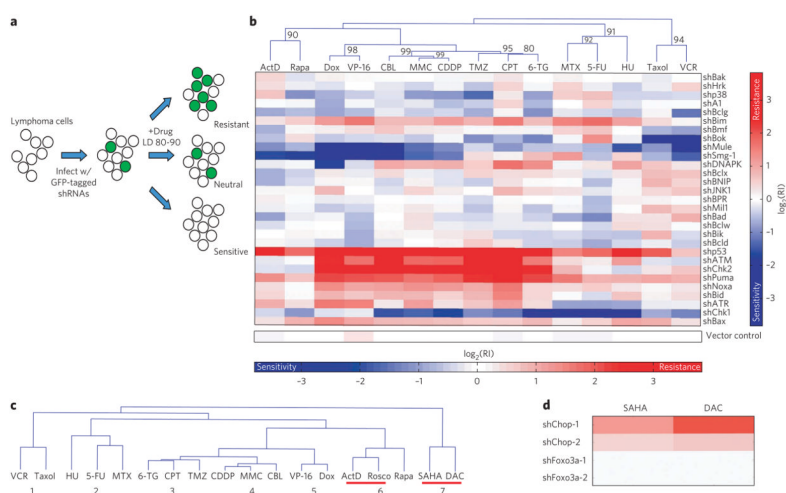
Figure 1. Functional characterization of chemotherapeutic drugs according to patterns of shRNA-conferred drug resistance or sensitivity

(a) A diagram showing the principle of GFP-based competition assays. Suppression of genes that alter drug sensitivity leads to changes in the percentage of GFP-positive cells after treatment, which can be used to calculate the RI (see Methods). **(b)** Unsupervised clustering of RI values of 15 reference compounds. Agglomerative hierarchical clustering was performed on log-transformed RI values for the initial 15 reference drugs, using a correlation metric and centroid linkage. Bootstrapping data is shown to indicate clustering robustness. 'Approximately unbiased' (AU) values from the PVclust function are indicated next to the relevant branches in the clustergram. **(c)** The branching pattern for SAHA, DAC and Rosco and the 15 reference chemodrugs. Numbers below the dendogram demarcate drug categories. **(d)** A heat map showing the response of cells expressing shRNAs targeting the Bim transcriptional regulator Chop and Foxo3a to SAHA and DAC. Log-transformed RI values are shown.
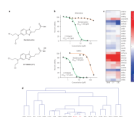
Figure 2. RNAi-based characterization of a compound derivative of bendamustine
(a) The chemical structures of bendamustine and a chemical derivative, CY190602. **(b)**
Dose response curves comparing the viability of the multiple myeloma cell lines
RPMI-8226 (top) and MM1S (bottom) following treatment with bendamustine or
CY190602. **(c)** RI patterns for bendamustine, CY190602 and a related compound,
chlorambucil (CBL). Bendamustine and CY190602 were used at LD80–90 of 110 μ|M and
1.4 μM, respectively. **(d)** The branching pattern for the 18 reference drugs plus
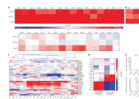bendamustine and CY190602.

Figure 3. Identification and functional characterization of ill-defined genotoxic drugs
(a) A heat map showing the response of cells expressing shATM, shChk2 or shp53 to 16 genotoxic (upper panel) and 15 nongenotoxic (lower panel) chemotherapeutics (see Supplementary Table 2 for drug abbreviations). **(b)** The *shATM-Chk2-p53* response signature for apigenin (APG) and NSC3852 (NSC). **(c)** The branching pattern for the 18 reference compounds plus APG and NSC. APG clusters with the TopoII poisons Dox and VP-16, whereas NSC clusters with the TopoI poison CPT. **(d)** a comparison of the shTopoI and shTopoII response signatures for APG and NSC3852 with response signatures derived from established TopoI (CPT and CPT11) and TopoII poisons (Dox, Mito and VP-16). Although NSC3852 and APG show response patterns characteristic of TopoI and TopoII poisons, respectively, none of the other genotoxic drugs showed either of these resistance and sensitivity patterns. **(e)** A graph showing the number of surviving shTopoII, shTopoI or vector control-expressing cells 12 days after drug treatment with APG or NSC3852. In each case, one million cells were plated before treatment. Data shown are mean ± s.e.m. from three independent experiments.
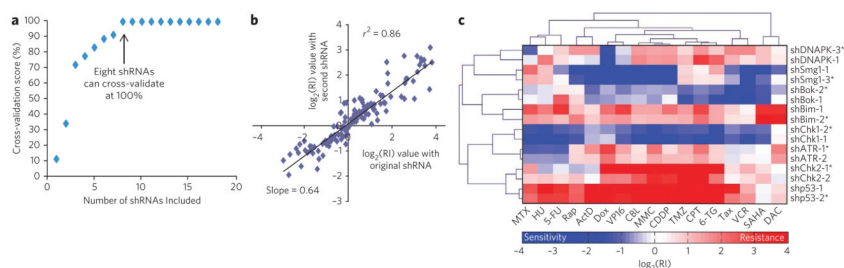
Figure 4. A feature reduction identifies a reduced eight-shRNA set
(a) Analysis of the dataset used for Figure 1c, using a randomized search strategy. The graph shows the relative efficacy of drug prediction as a function of increasing shRNA subset size. The maximum predictability for 2,000 iterations at each shRNA subset size is shown. (b) a graph showing the correlation between enrichment or depletion of cells expressing shRNAs from the original eight-shRNA set and cells expressing shRNAs from the additional eight-shRNA set after drug treatment. Each square represents the $\log_2$RI values following single-drug treatment of cells expressing an original shRNA (*x* axis) or the second shRNA targeting the same gene (*y* axis). The slope of the best-fit line is 0.64, indicating that the absolute RI values are consistently lower in cells expressing hairpins from the second eight-shRNA set. (c) A heat map showing the relative enrichment (red) or depletion (blue) of a second set of shRNAs (labeled with asterisks) targeting each of the genes in the eight-shRNA signature. The associated dendograms show clustering between shRNA pairs, as well as clustering of small molecules into the same seven categories predicted from the 29-shRNA signature.
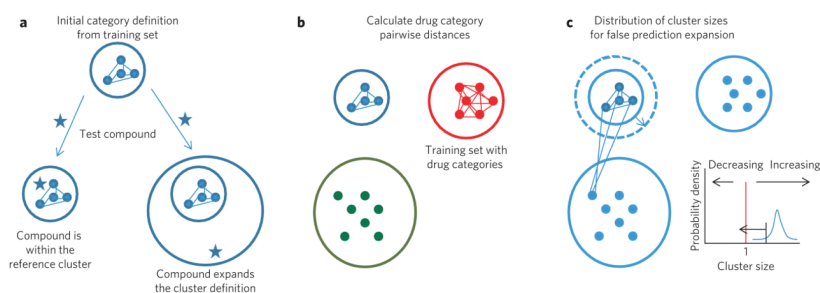
Figure 5. A reduced shRNA signature can accurately predict drug mechanism of action
(a) A diagram of the possible outcomes for a test compound when it is compared to the training set. A test compound could be interpolated within the definition of a drug category that is provided by the training set (left). Alternatively, a test compound could be outside of the drug category (right). Our probabilistic nearest-neighbors algorithm attempts to define an 'acceptable' category extension. **(b)** A schematic depicting the methodology behind probabilistic nearest-neighbors predictions. An initial training set with empirically validated drug categories is used to calculate the drug category-specific cluster sizes. This same methodology is used for compounds whose known mechanism of action is distinct from a particular drug category. **(c)** The increase in the drug category definition that is observed by forcing these empirically derived negative controls to cluster in an erroneous category is used to build a null distribution and an empirical cumulative distribution function.
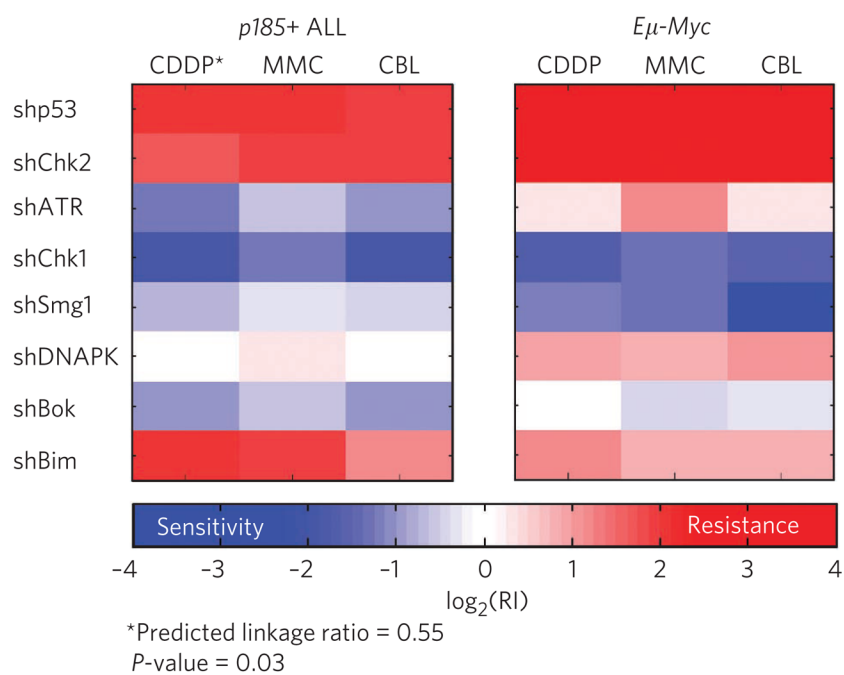
Figure 6. Adaptation of the eight-shRNA signature to a distinct cell line
A heat map comparing the eight-shRNA response signatures of *Myc p19^Arf−/−* lymphoma cells and *p185+ BCR-Abl* leukemia cells following treatment with alkylating agents in a model of acute lymphoblastic leukemia. The eight-shRNA signature from *p185+ BCR-Abl p19^Arf−/−* leukemia cells can identify CDDP as an alkylating agent when CBL and MMC are used as a reference set.

NIH-PA Author Manuscript

Table 1

Using the eight-shRNA signature to predict drug mechanism

(a)

Compounds	Prediction	Linkage ratio	*P*-value	Correct prediction
Daunorubicin	TopoII poison	0.83	0.0004	Yes
Zebularine	HDAC/DNMT inhibitor	0.82	0.002	Yes
Busulfan	DNA cross-linking	0.97	0.007	Yes
2-Methoxyestradiol	Antimicrotubule	1.08	0.009	Yes
Vinblastine	Antimicrotubule	0.95	0.001	Yes
Scriptaid	HDAC/DNMT inhibitor	0.85	0.004	Yes
Carmustine	DNA cross-linking	0.95	0.005	Yes
Mitoxantrone	TopoII poison	0.94	0.002	Yes
Thiotepa	DNA cross-linking	0.92	0.003	Yes
Gemcitabine	Nucleotide synthesis inhibitor	0.93	0.002	Yes
Melphalan	DNA cross-linking	0.91	0.002	Yes
Carboplatin	DNA cross-linking	0.86	0.0008	Yes
Streptozocin	SSB inducer	1.03	0.02	Yes
Maphosphamide	DNA cross-linking	0.93	0.003	Yes
Irinotecan	SSB inducer	0.88	0.002	Yes
Noscapine	Antimicrotubule	0.83	0.0002	Yes
Cantharidin	Novel mechanism	1.27	0.10	Negative control
AA2	Novel mechanism	1.39	0.3	Negative control
Gliotoxin	Novel mechanism	1.11	0.09	Negative control
AG1478	Novel mechanism	1.08	0.06	Negative control

(b)

Updated training set with new categories:

New category	Reference drugs added	New compound	Class prediction	Linkage ratio	*P*-value	Correct prediction
Proteasome inhibitor	PS341 and MG132	Epoxomycin	Proteasome inhibitor	1.02	0.03	Yes
Proteasome inhibitor	PS341 and MG132	Gliotoxin	Proteasome inhibitor	1.22	0.05	Yes
Hsp90 inhibitor	17AAG and PU-H71-Br	VER–50589	Hsp90 inhibitor	1.09	0.0002	Yes
Hsp90 inhibitor	17AAG and PU-H71-Br	BIIB021	Hsp90 inhibitor	0.80	<0.0001	Yes

(b)

Updated training set with new categories:

New category	Reference drugs added	New compound	Class prediction	Linkage ratio	P-value	Correct prediction
Hsp90 inhibitor	17AAG and PU-H71-Br	Neopentylamine-42	Hsp90 inhibitor	0.75	<0.0001	Yes
EGFR inhibitor	Erlotinib and Gefitinib	AG1478	EGFR inhibitor	0.88	0.002	Yes
Negative control	All	PD173074 (FGFR inhibitor)	Novel mechanism	1.30	0.2	Negative control
Negative control	All	GDC0941 (PI3K inhibitor)	Novel mechanism	1.13	0.2	Negative control

(a) A table showing the predictive power of the eight-shRNA signature on a set of drugs that were not used to derive the signature. The Prediction column indicates the mechanism of action of the compound as predicted by a nearest-neighbors approach. The linkage ratio describes the proximity of a test compound to a particular class of compounds and defines the observed increase (or decrease) in drug category size. For example, a linkage ratio of 1.1 indicates that the addition of a new drug expands the drug category by 10%. The p-value describes whether the proximity of a compound to a given drug category is significant when compared to a negative control distribution for that drug category. Cantharidin (a protein phosphatase inhibitor), apoptosis activator 2 (AA2, a direct activator of the apoptosome), gliotoxin (a proteasome inhibitor) and AG1478 (an EGFR inhibitor) were used as negative controls and were predicted to be distinct from any of the existing reference drugs. (b) Category predictions and significance levels upon adding three new drug categories (proteasome, Hsp90 and EGFR inhibitors) that were not used to develop the initial eight-shRNA signature. PU-H71-Br is a chemical derivative of the benzyladenine–based Hsp90 inhibitor PU-H71. VER–50589 and neopentylamine-42 are Hsp90 inhibitors. PD 173074 (a FGFR inhibitor) and GDC 0941 (a PI3K inhibitor) were used as negative controls to test the stringency of predictions after the incorporation of these new drug categories.