Chemoinformatic Analysis of NCI Preclinical Tumor Data: Evaluating Compound Efficacy from Mouse Xenograft Data, NCI-60 Screening Data, and Compound Descriptors

Anders Wallqvist,*,‡ Ruili Huang,† and David G. Covell†

Developmental Therapeutics Program, National Cancer Institute, NCI-Frederick, Frederick, Maryland 21702, and Laboratory of Computational Technologies, SAIC-Frederick, Inc., NCI-Frederick, Frederick, Maryland 21702

Received April 11, 2007

We provide a chemoinformatic examination of the NCI public human tumor xenograft data to explore relationships between small molecules, treatment modality, efficacy, and toxicity. Efficacy endpoints of tumor weight reduction (TW) and survival time increase (ST) compared to tumor bearing control mice were augmented by a toxicity measure, defined as the survival advantage of treated versus control animals (TX). These endpoints were used to define two independent therapeutic indices (TIs) as the ratio of efficacy (TW or ST) to toxicity (TX). Linear models predictive of xenograft endpoints were successfully constructed $(0.67 < r^2 \leq 0.74)_{\text{observed\_versus\_predicted}}$ using a model comprised of variables in treatment modality, chemoinformatic descriptors, and in vitro cell growth inhibition in the NCI 60-cell assay. Cross-validation analysis based on randomly chosen training subsets found these predictive correlations to be robust. Model-based sensitivity analysis found chemistry and growth inhibition to provide the best, and treatment modality the worst, indicators of xenograft endpoint. The poor predictive power derived from treatment alone appears to be of less importance to xenograft outcome for compounds having strongly similar chemical and biological features. ROC-based model validation found a 70% positive predictive value for distinguishing FDA approved oncology agents from available xenograft tested compounds. Additional chemoinformatic applications are provided that relate xenograft outcome to biological pathways and putative mechanism of compound action. These results find a strong relationship between xenograft efficacy and pathways comprised of genes having highly correlated mRNA expressions. Our analysis demonstrates that chemoinformatic studies utilizing a combination of xenograft data and in vitro preclinical testing offer an effective means to identify compound classes with superior efficacy and reduced toxicity.

INTRODUCTION

A fundamental guiding principle in anticancer drug discovery is that of chemical and biological similarity, i.e., closely related molecules will elicit a comparable response in a biological system,[1−4] albeit with noteworthy exceptions.[5] Chemoinformatics is now recognized among the many 'omics' disciplines for its focus on chemistry, biology, and discovery.[6−8] Current chemoinformatic studies encompass a wide range of chemical considerations[9,10] applied across an equally diverse set of biologies.[6,11] Within the area of anticancer drug discovery, a cascade of test systems exist for identifying compounds sufficiently effective for tumor cell killing, while still retaining acceptable safety margins for human use.[12] Widely used test systems include human tumor cells grown in culture or grown as xenografts in murine hosts. As a testament to the importance of currently used xenograft modes, no currently accepted anticancer compound has failed in xenograft testing.[13] Chemoinformatic efforts that attempt to relate chemical features to screening or xenograft endpoints are, to our knowledge, rare. Although not documented, chemoinformatics studies utilizing mouse xenograft experiments for preclinical assessment of efficacy appear to be complicated by many difficulties inclusive of devising appropriate methods for standardizing xenograft data across diverse dosing protocols, proper consideration of effects due to physiochemical property descriptors of early development candidates, and unanticipated off-target effects on efficacy and toxicity.[14,15] Untoward results in any of these areas can complicate chemoinformatic interpretations and prematurely terminate an otherwise promising investigational new drug.[16] Consequently, the value and relevance of preclinical cancer xenograft models as well as other in vitro and in vivo tests have recently been debated.[13,17−19] Currently there are no preclinical models approved for directing clinical decisions.[20] Thus detailed chemoinformatic studies of these systems, such as that described herein, remain highly desirable.[12,21−24]

The bulk of available xenograft data from the Developmental Therapeutics Program (DTP) represents an early drug screening effort[25−27] predating the use of growth inhibition data from the NCI 60 tumor cell line screen. Xenograft outcome measures were based on established thresholds for increased survival time and/or tumor weight reduction. Variations in dosage, injection site, vehicle, and mouse strain

* Corresponding author phone: (301)846-5665; fax: (301)846-6798; e-mail: wallqvist@ncifcrf.gov.
† National Cancer Institute, NCI-Frederick.
‡ SAIC-Frederick, Inc., NCI-Frederick.

10.1021/ci700132u
Published on Web 06/08/2007

were explored for evidence of compound activity. If no activity was observed in at least three xenograft tests, then the compound was not considered further. The predictive power of xenograft activity for human efficacy within a given histology was recently assessed with mixed results.[12,14,28] Using the DTP xenograft data, efficacy in phase II clinical trials for breast or colon cancers was not well-predicted with these models, whereas predictability was found for non-small-cell lung and ovarian cancer. Hollow fiber[29] activity was also found to be predictive in certain cases. Noteworthy in these studies was the finding that xenograft activity independent of tumor histology was found to be well-correlated with outcome in phase II trials. None of these studies conducted a detailed analysis of the chemotypes related to xenograft activity.

In this paper we revisit the DTP xenograft data with two distinct goals. The first is to provide a model-based evaluation of xenograft efficacy and toxicity across variations in treatment, biology, and chemistry. The sheer magnitude and diversity of xenograft data alone has traditionally proved difficult to comprehensively analyze[15] and may have fostered controversy over its utility as a predictive model for anticancer drug development. These results of this analysis will be presented in three subsections: *Data Overview and Analysis*, strategies for describing the raw data; *Modeling of Outcome*, model development; *Dependency of Outcome on Treatment, Chemistry, and Biology*, chemosensitivity analysis of parameter variations; and *Modeling of Therapeutic Index for FDA Approved Cancer Compound*, model application. These sections will explore the relationships between treatment modality, in vitro biological response in the NCI-60 screen,[30,31] and chemical descriptors, on xenograft efficacy and toxicity. A practical extension of our model will be provided to address the *Implications for Varying Treatment Parameters*. Taken collectively, these results will demonstrate that a self-consistent model, encompassing many of the experimental variations, can be obtained and applied toward assessing the impact of experimental design on xenograft outcome and identifying optimal experimental conditions.

The second goal of our analysis is to integrate the results of this model with prior analyses of a compound's mechanism of action as defined by a molecular target or pathway target. Two subsections will be provided. The first, titled *Therapeutic Index and Drug Mechanism of Action*, will assess whether combining xenograft data with complementary chemoinformatic data obtained from the NCI-60 screening efforts can add value to the existing, separately derived, results. The second subsection will integrate our xenograft model with previously published pathways results[32,33] to identify novel relationships between *Pathway Fitness and Xenograft Toxicity Data*. Although highly speculative, our results reveal that casting the existing DTP xenograft data into a self-consistent model provides a systematic method to better assess the results of xenograft data and a powerful means to integrate prior information, derived separately from the NCI-60 tumor panels, into the process of data interpretation. The opportunity to cohesively interpret these parallel efforts provides useful suggestions regarding strategies for future data analyses.

RESULTS

Data Overview and Data Treatment. The experimental procedures used for the xenograft data generation and collection are given in a technical summary accessible on the DTP Web site; the available data can be obtained from PubChem.[34] Each xenograft experiment generates an outcome measured as tumor weight reduction (TW) or survival time advantage (ST) in mice relative to a tumor bearing control cohort, typically with each treated and control set containing an average of 10−20 mice. Each experiment is characterized by a number of variables pertaining to the xenograft tumor and type of mouse, drug concentration, vehicle, site of administration, dose, and schedule. In summary, there are seven different sites for drug administration, of which intraperitoneal is the most common. A total of 31 different drug formulations have been used, of which saline solutions with or without Tween-80, hydroxypropylcellulose, and carboxymethylcellulose are typical formulations. Numerous drug schedules were used, totaling 187 unique protocols, reflecting the empirical effort to find the best treatment modality for each drug in combination with all other variables. The most common schedule is a daily drug administration repeated nine times. A total of 50 different tumor models are recorded, of which the P388 and L1210 leukemia models are the most frequently investigated. Six different tumor implantation sites were used with the intraperitoneal site being the most common. Fifteen different mouse strains are used with the B6D2F1 and CD2F1 strains being typical. Not all combinations of test parameters are investigated here as there are a total of slightly over 5000 variable combinations in this experimental data set; the most frequent combination uses a drug formulation of saline with Tween-80 injected intraperitoneally daily with a P388 leukemia tumor model implanted intraperitoneally in CD2F1 mice.

Instead of examining each data point and judging whether a test compound was meritorious for that particular set of experimental conditions, we attempted to establish a global perspective of this data by constructing dose−response curves for each set of xenograft outcomes. For a specific set of assay conditions, the xenograft outcome was recorded for each test condition, and a particular endpoint was chosen. All concentration units were converted to mg/kg/injection divided by the molecular weight of the compound, i.e., a concentration unit. A specific endpoint would be the drug concentration at which the tumor weight in treated animals is reduced by 50% when compared to the tumor weight in control animals (TW50). Likewise, we define ST150 as the drug concentration in a specific assay where the lifetime of the treated animal is 50% longer than the control animals (ST150). In Figure 1 we show this concept for adriamycin and colon carcinoma 38 tumor cells subcutaneously implanted in B6D2F1 mice. In this figure the solid curves are drawn to guide the eye, but no direct functional relationship is implied. However, these lines provide support for model-based characterizations of what clearly involves quite noisy data. Endpoints other than TW50 and ST150 could be used to gauge a compound's effect; however, all assay endpoints cannot necessarily be evaluated with this model due to paucity of data and, in some cases, a complete lack of compound efficacy in these xenograft tests.

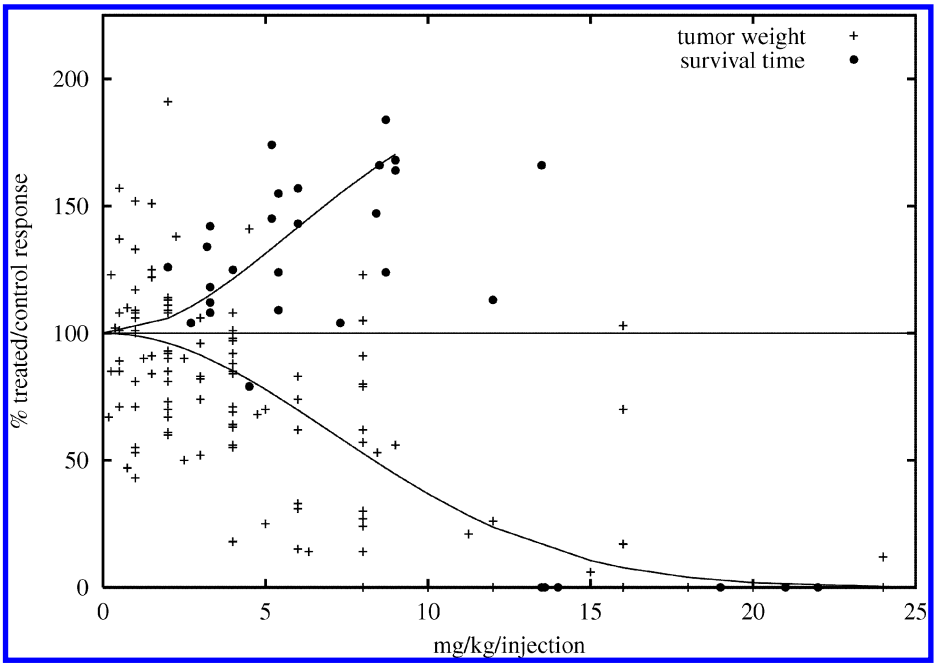

Figure 1. Dose−response curves for xenograft endpoints. Outcome of xenograft experiments for adriamycin in the colon carcinoma 38 model of cancer implanted in mice. The tumor weight reduction is calculated at the end of treatment as the percentage of tumor weight in treated mice versus an untreated tumor bearing control group for each tested drug concentration. Likewise the survival time increase of treated mice is given relative to control. Although the data show considerable variability, the trend of a larger effect at higher dosages are evident. From the dose response curve, drawn here to guide the eye, we can estimate the concentration of each drug in the data set that achieves a predetermined endpoint, e.g., a 50% reduction in tumor weight (TW50) or a 50% increase in survival time (ST150).

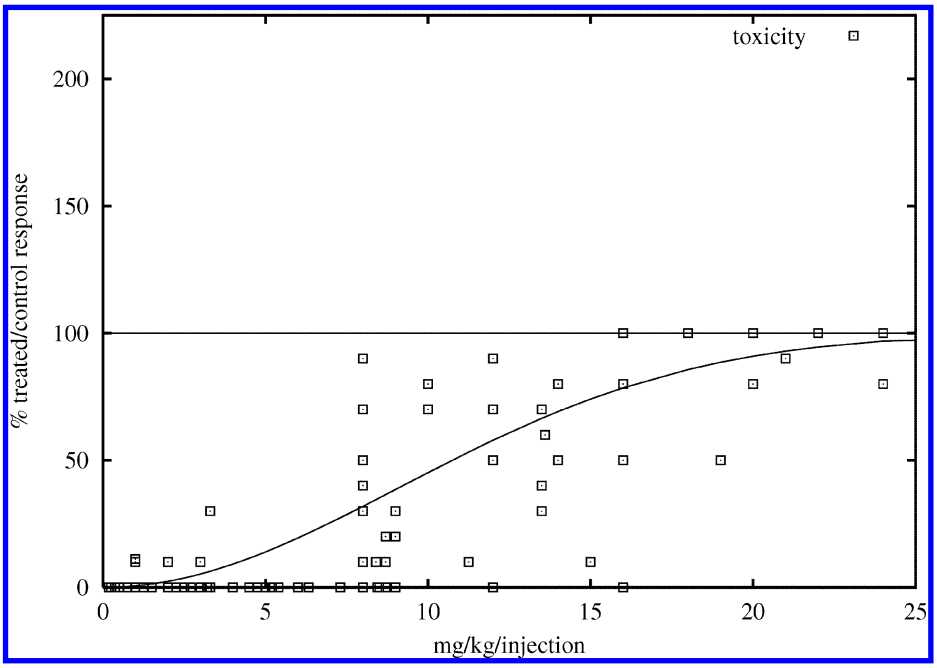

Figure 2. Dose−response curves for toxicity. In this figure we illustrate the evaluation of toxicity for the same system as in Figure 1. The quantity measures the excess deaths in the treated over the control population due to drug effects. The qualitative behavior reveals few deaths occurring at low dosages with an increased death rate at higher doses. This effect defines a dose response curve that can be evaluated for a fixed response, e.g., the concentration at which there is a 50% excess death in the treated population (TX50).

A further characterization of the assay data is captured in a toxicity measure, defined similarly to the tumor weight reduction (TW) and survival time (ST) values. By counting the number of treated mice versus the number of control mice that survived the assay we can define a dose−response curve as the (1-treated/control)*100 and test when we achieve a particular endpoint, e.g., the concentration at which half the treated mice do not survive as compared to the control group (TX50). This is illustrated in Figure 2 for the same assay system as in Figure 1. Quantifying data in this manner provide an effective measure of toxicity; i.e. at low concentrations there is no substantial death increase in this example for the adriamycin-treated group versus the tumor-bearing control group, whereas at the highest concentrations most of the treated mice have died compared to the control group bearing the same tumor.

The three endpoints, TW50, ST150, and TX50, are arbitrarily chosen to characterize this data; different values

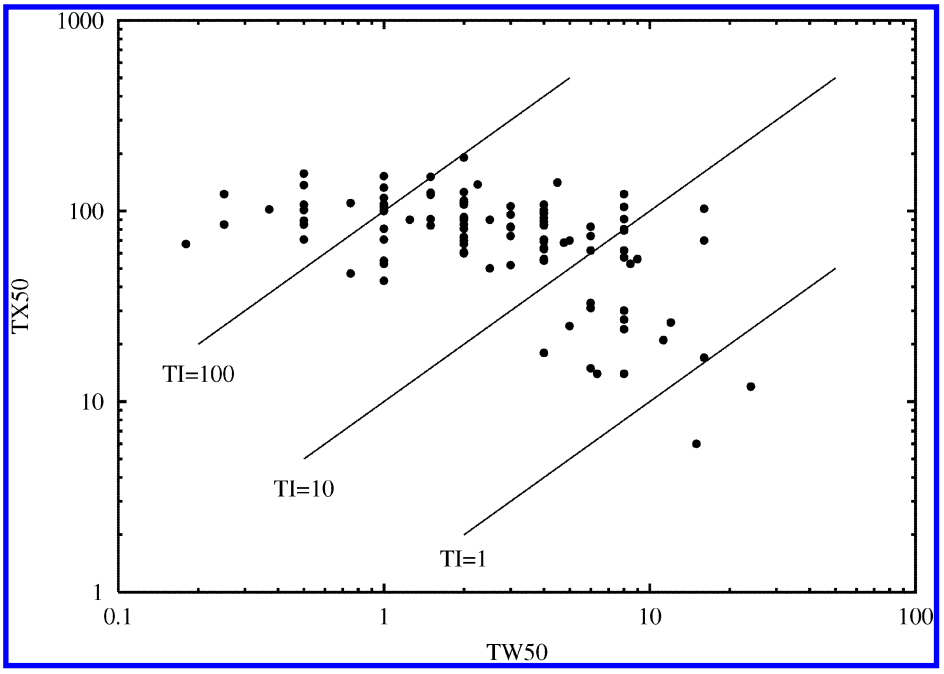

Figure 3. TI dependence on assay. As different cancer models and treatment options are evaluated a range of efficacy and toxicity values are generated. Here, TW50 and TX50 values for all adriamycin assays is given. The lines refer to constant ratios of TX50/TW50, i.e., a measure of therapeutic index. The efficacy and toxicity of the drug can vary substantially with treatment and cancer model, similar to the differences in patient response observed in clinical trials of the same drug.

or more stringent antitumor thresholds could also be chosen, e.g., TW90, ST200, and TX10, with a concomitant reduction in the number of evaluated endpoints. Explorations of relationships between these parameters across this diverse set of experimental designs yield systematically quantifiable global characterizations of the xenograft data, rather than simply selecting the "best" subset of responses. This global perspective can then be used to relate dosing protocols and molecular properties to xenograft outcome. A further quantity of interest is the therapeutic index (TI), defined as the ratio of the concentration achieving a particular efficacy to toxicity, i.e., TI = (TW or ST)/TX. A large TI indicates that the effect can be achieved at a minimally toxic concentration. An example of the range of TIs using TW50 and TX50 for assay endpoints in adriamycin testing is shown in Figure 3 (see legend for details). The performance of a single agent, as widely accepted as adriamycin, can thus vary from a TI of 1 to more than 100 depending assay conditions, i.e., scheduling, tumor-type, drug-formulation, vehicle, and route. The scatter in this data accurately reflects the underlying challenges associated with using xenograft outcomes within a decision network (DN). Evident in this example is that outcomes from selected experimental designs can be used to reject agents as clinical candidates.

Modeling Outcome. Our experimental design assumes that xenograft outcome depends on treatment parameters, chemistry parameters, and biology parameters. Based on this premise it is possible to construct models of TW50, ST150, and TX50 as a function of these parameters. Here the treatment parameters consist of all the possible variations in the xenograft assay results due to scheduling, drug-formulation, tumor model, etc. Our *treatment variable set* was collected from the most frequent treatment combinations. The chemical properties of these compounds are derived from calculated property descriptors to form the *chemical descriptor set*. For the bulk of the compounds, corresponding GI_{50} measurement exists across the NCI-60 tumor cell lines. Thus the individual growth inhibition values for each cell line can be used as input variables for the *biological descriptor set*. These descriptor sets are not necessarily independent, but this is not a requirement for modeling outcome. These descriptor sets, applied to the TW50 data set and shown in Figure 4, are comprised of 81 different xenograft treatment schedules, 216 chemical descriptors derived from the compound structure, and 70 GI_{50} measurements from the NCI-60 cell line screen, ordered from left to right. These data are further ordered from bottom to top according to the concentration required to achieve an endpoint of TW50, sorted to place the smallest and most potent compounds at the bottom. A general trend is evident as compounds that have more sensitive GI_{50} values (i.e., potent) correspond with increased tumor weight reduction (TW50). It has been shown previously that GI_{50} values are the most predictive of xenograft activity.[28] A further observation is that the most potent compounds also exhibit the most extreme chemical properties, as reflected in the chemical descriptor set.[32,35]

A stepwise linear regression procedure automatically selects the most important variables from these sets to make the best fit of the xenograft data. Xenograft outcome can be described very well using a small number of parameters. Shown in Figure 5 are the results of modeling the TW50 endpoint; spanning 6 orders of magnitude in dosing concentration. The final modeling results of our three xenograft outcomes are summarized in Table 1, with the r^2 values for selecting variables from the full variable set (treatment, chemistry, and biology) ranging from 0.67 to 0.74. Assessments of the strengths of these correlations were performed using randomly chosen data sets sized to represent one-third of the overall data size and used to predict these same correlations for the remaining data fraction. The distribution for r^2 from 100 simulations yielded a mean of 0.68 and a coefficient of variation of 18% for correlations based on the

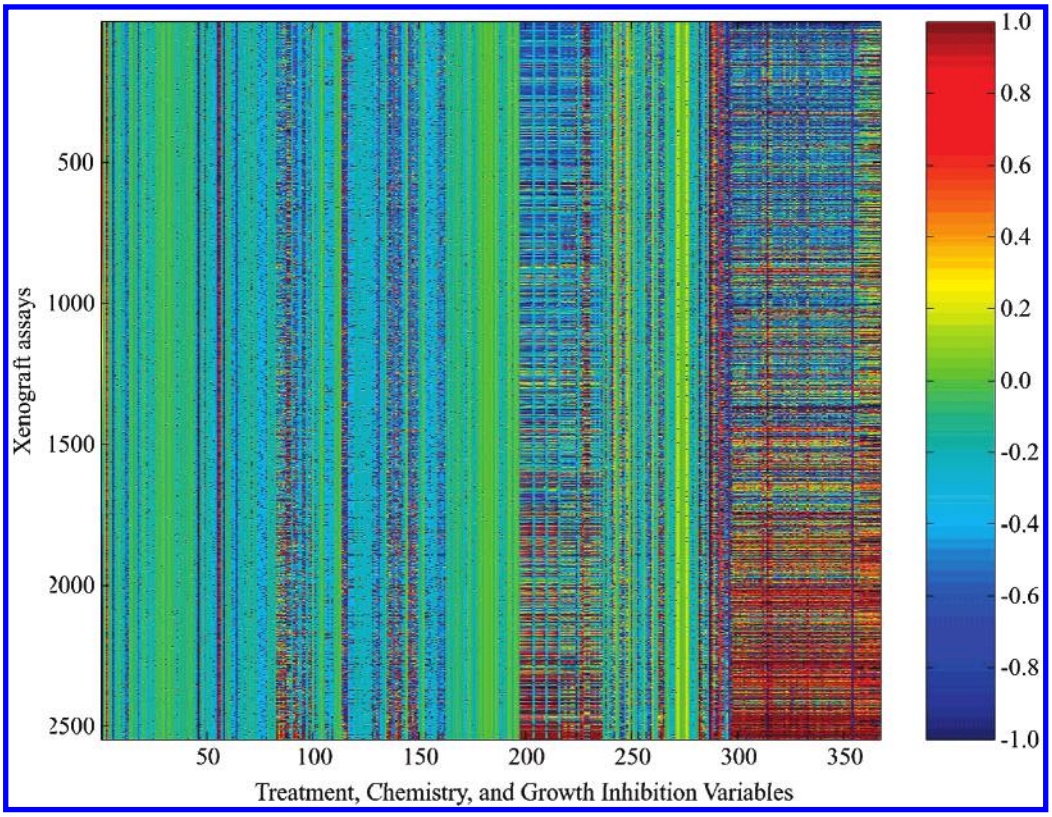

Figure 4. Overview of data vectors. Graphic illustration of all variables associated with a particular xenograft outcome; this includes treatment options (variables 1−212), chemical descriptors of molecular properties and structure (variables 213−296), and growth inhibition data (variables 297−366). The outcome is sorted so that the least sensitive results are at the top and the most sensitive at the bottom. The data has been z-score normalized to emphasize the variation in each variable set.

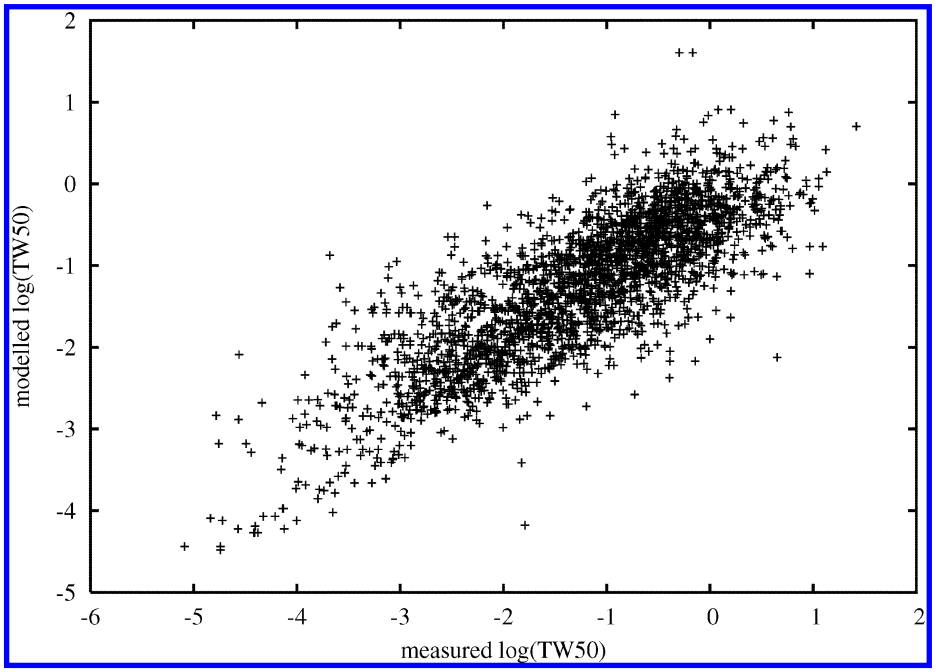

Figure 5. Modeling of TW50. Comparison between measured and calculated values of TW50 using a linear regression model with 50 parameters optimally chosen from the data set in Figure 4. The correlation coefficient was 0.82; details of variations in the model are given in Table 1.

three model variables of treatment, chemistry, and biology. Thus it is possible to model the xenograft data with a modest number of parameters using variables selected from all three fundamental processes that define our xenograft study: treatment, chemistry, and biological activity. No single set of variables can capture the full extent of the xenograft assays, as outlined in Table 1. As expected, treatment alone was the worst predictor of outcome, yet the treatment variables did carry some information relevant to the outcome, regardless of the administered compound. The biological descriptor set, derived from GI_{50} data vectors, are better predictors of outcome when compared to the treatment descriptor set, with the chemical descriptor set alone carrying the most information relevant to assay outcome. Here the

Table 1. Evaluation of Xenograft Modeling[a]

			correlations r/r^2			
endpoint	N	M	all	treatment	chemistry	biology
ST150	11381	6549	0.86/0.74	0.45/0.20	0.79/0.62	0.75/0.56
TW50	6642	2546	0.82/0.67	0.36/0.13	0.72/0.52	0.65/0.54
TX50	25511	6549	0.84/0.71	0.39/0.15	0.76/0.58	0.72/0.52

[a] Our model-based evaluation of the xenograft parameters uses a correlation coefficient to measure goodness of fit. The three endpoints modeled are survival time increase of 50% compared to control (ST150), tumor weight reduction by 50% (TW50), and a measure of toxicity (TX50) indicating a 50% death rate. The original number of xenograft data points (N) could not all be matched to GI_{50} data vectors, and the total number of data used in the modeling is M. The maximum number of variables included in the modeling was 50; further variable inclusion only marginally improved the fit. Both the correlation coefficient r and r^2 are given for all variable sets tested. Selecting data from all three categories of variables yielded optimal solutions; each individual variable set was inferior to the combination.

combination of all three variable sets produced the best modeling results for xenograft outcome. These results were further borne out by simulations based on using one-third of the data to predict the results for the remaining fraction of data. Distributions of r^2 from these simulations support the observation that treatment has the poorest predictive value, while chemistry and biology yielded mean r^2 values comparable to those listed in Table 1.

Dependency of Outcome on Treatment, Chemistry, and Biology. Models of xenograft data are not necessarily predictive. For example, models predictive of the numerous different treatment schemes are less likely to be successful. And since treatment can have profound effects on outcome, one is forced to address the issue of when and if a model, albeit theoretical or animal based, can be predictive of xenograft outcome. The mouse xenograft system allows us to begin to answer this question self-consistently across a reasonably broad spectrum of treatment choices. This perspective does not directly address the fidelity of using mouse xenograft models to predict human outcome but rather questions how sensitive a model is to variations in treatment protocol.

A specific set of variables from treatment, chemistry, and biological parameters are chosen for our linear regression model. This model's parameters can be ordered according to groups such that the outcome is predicted from a set linear in treatment $\mathbf{B}(\text{treatment})\cdot\mathbf{X}(\text{treatment})$, a set linear in chemistry $\mathbf{B}(\text{chemistry})\cdot\mathbf{X}(\text{chemistry})$, and a set linear in biological variables $\mathbf{B}(\text{biology})\cdot\mathbf{X}(\text{biology})$. The outcome ($\mathbf{Y}$) is then the sum of these terms, schematically written as

$$\mathbf{Y} = \mathbf{B}(\text{treatment})\cdot\mathbf{X}(\text{treatment}) + \mathbf{B}(\text{chemistry})\cdot\mathbf{X}(\text{chemistry}) + \mathbf{B}(\text{biology})\cdot\mathbf{X}(\text{biology}) \quad (1)$$

We can now introduce a similarity measure between the same classes of variables to gauge their difference. For two compounds (i and j) we can measure their chemical similarity by calculating the Pearson correlation coefficient $r_{ij}^{\text{chemistry}}$ between their chemical variables, $\mathbf{X}(\text{chemistry})_i$ and $\mathbf{X}(\text{chemistry})_j$. Likewise we can calculate a corresponding correlation coefficient between biology variables r_{ij}^{biology} as well as between treatment variables $r_{ij}^{\text{treatment}}$. This procedure provides a pairwise measure of similarity across the complete range of variation for all selected variables. These correlations can be used to select regions of chemistry, biology, and treatment conditions that yield the most consistent xenograft response. For example we can construct groups of xenograft outcomes based on experiments that fulfill a particular set of conditions, e.g.

$$0.2 < r_{ij}^{\text{chemistry}} \leq 0.4 \quad (2)$$

$$0.8 < r_{ij}^{\text{biology}} \leq 1.0 \quad (3)$$

$$0.4 < r_{ij}^{\text{treatment}} \leq 0.6 \quad (4)$$

This example selects experimental outcomes that have low chemical similarity among the tested compounds, high biological similarity, as measured by the correlation strength on their GI_{50} profiles, and medium treatment similarity, arbitrarily defined as experiments having moderately similar dosing protocols. For all such combinations we gather the individual outcomes $\mathbf{Y}_i$ and $\mathbf{Y}_j$ and form a pair of data vectors containing all first members of the pair $\mathbf{O}_1$ and all second members of the pair $\mathbf{O}_2$. The similarity between the data vectors $\mathbf{O}_1$ and $\mathbf{O}_2$ is then calculated using the Pearson correlation coefficient. We can then examine similarity of xenograft response based on similarity of each variable, e.g., the parameter range given in the above example would give a single number, e.g., 0.2, descriptive of xenograft outcome for compounds with low chemical similarity, high biological similarity, and medium dosing similarity. Since this a four-dimensional space it is hard to directly visualize, but in Figure 6 we present three cuts. The data in this figure are taken from the tumor weight reduction (TW50) endpoint of the xenograft data. These cuts are constructed so that one variable has a high similarity, $r > 0.8$, and the xenograft response similarity (TW50 in this case) is drawn as a contour surface depending on variation in the two remaining experimental variables.

This abstraction of the data allows a number of qualitative conclusions about modeling outcome of anticancer agents in xenograft studies as well as suggesting similar results in other settings. In Figure 6(A) the treatment similarity is held high ($r_{\text{treatment}} > 0.8$), and similarity in chemical and biological activity is varied. These contour lines are color-coded in cyan, magenta, blue, and green, to indicate xenograft outcome at correlation cuts of 0.2, 0.4, 0.6, and 0.8, respectively. At low similarity in chemical properties, defined at the left edge of this figure, and biological similarity nearly spanning its entire range ($0.0 <= r_{\text{biology}} <= 0.9$), the concordance of xenograft outcome is quite low (the correlation cutoff of 0.2 is indicated by the left-most cyan contour line in this figure). If on the other hand there is high similarity in both chemical and biological properties (upper right portion of this figure), xenograft outcome is highly correlated (correlation cutoff of 0.8 as indicated by the green contour line), as should be expected.

The graph in Figure 6(B) shows the effect of variations in treatment and biology, fixed at high chemical similarity, again contoured according to strength of xenograft outcome similarity. Here concordant xenograft outcomes ($r > 0.8$, green contour) correspond with high treatment similarity, while biological similarity as low as $r_{ij}^{\text{biology}} = 0.5$ can be tolerated with a high coherence in xenograft outcome; as

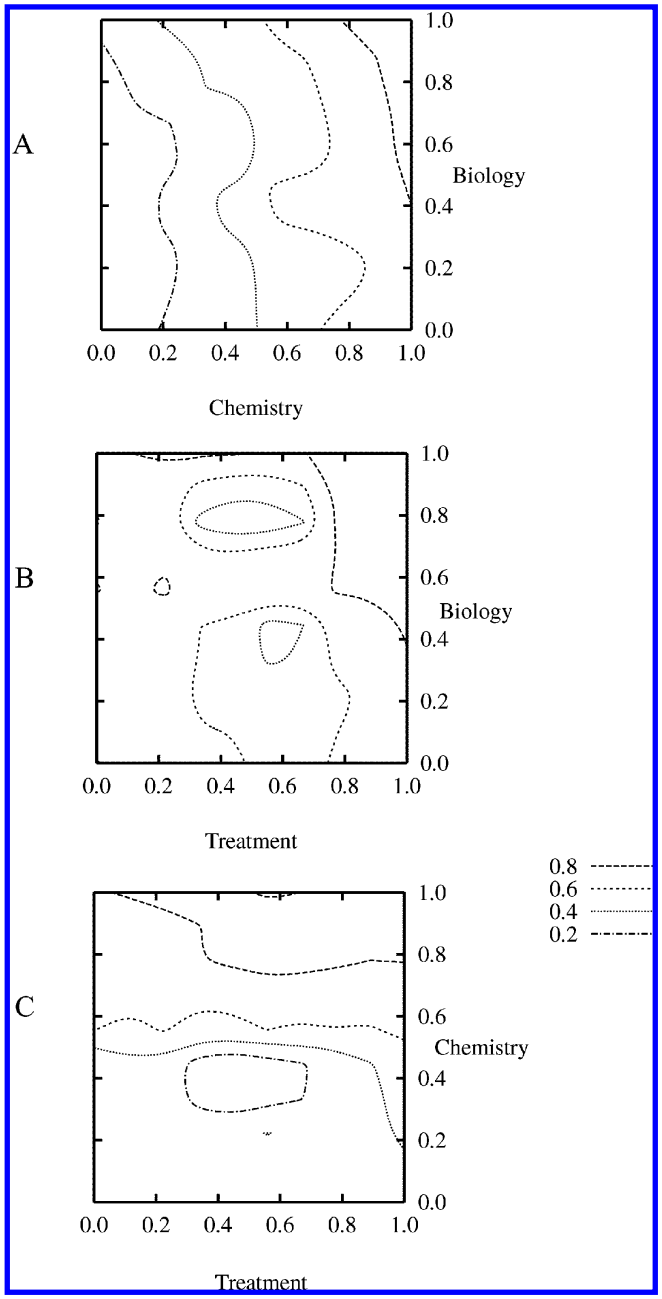

Figure 6. Influence of treatment, chemistry, and biology. Variations in xenograft outcome as a function of similarity measured across our chemical, biological, and treatment descriptor sets. Here we test the change in tumor weight reduction outcome as a function of similarity in the selected variables from treatment, physical/chemistry properties, and the GI_{50} data. Each graph shows the dependency of xenograft outcome similarity (shown as different colored contour lines) as a function of variation in two of our three descriptor sets, as the third set is held fixed. Panel (A) shows the similarity in outcome for fixed treatment, (B) fixed physical/chemistry properties, and (C) fixed GI_{50} parameters.

derived from the xenograft outcome similarity cut of 0.8 (i.e., green contour line) passing through the upper right portion of this Figure. The most interesting observation for modeling and prediction of outcome is shown in Figure 6(C) where the biological similarity is held high with varying chemistry and treatment similarity. If the chemical similarity is high, at or above $r_{ij}^{\text{chemistry}}$ of 0.8, there is considerable tolerance for treatment variation. The green contour boundary for high xenograft outcome similarity has a lower limit for treatment similarity of $r_{ij}^{\text{treatment}} \geq 0.4$. Thus, although treatment is important, predictions and models where compounds have high chemical and biological similarity will show considerable outcome similarity over a moderate range of treatment variation. It is of course always true that individual examples can vary from the global results, and care must always be taken when interpreting each compound assay. Nonetheless, this analysis provides support for chemically similar compounds that also produce comparable in vitro screening profiles (i.e., biological similarity) and yield similar xenograft responses across a range of treatment schedules. Although arguably challenged as an oversimplification, these results support a case for in vitro potent, structurally similar compounds appearing to yield robust xenograft responses across diverse treatment schedules.

Modeling of TI for FDA Approved Cancer Compounds. For all xenograft experiments on a particular compound, i.e., different xenograft models and treatment protocols, there is an optimal set of conditions that yields a maximum TI, TI_{opt}. The following analysis uses tumor weight reduction (TW50) as an endpoint for selecting TI_{opt}. Selecting only optimal outcomes from the complete set of experiments has been considered by others as a useful index of xenograft efficacy (i.e., Johnson et al.[28] selected xenograft data sets where acceptable activity was observed in at least three experiments). The set of small molecule compounds that is FDA-approved for clinical use in chemotherapy constitutes an objective set that reflects on the suitability of a compound as active against cancer. Selecting FDA approved agents for our evaluation assumes their clinical efficacy, thus eliminating the step of acquiring treatment response data for compounds entering clinical trial testing.[14] TI_{opt} distributions of xenograft outcome for FDA-approved drugs versus compounds randomly picked from our xenograft data find a significant difference between these two populations ($p \leq 0.05$, Kolmogorov test, data not shown), with a median TI_{opt} for FDA-approved drugs of 10.1 versus 1.5 for the random set.

This difference can be additionally exploited in a predictive manner by constructing a sensitivity versus specificity analysis within the FDA-approved compound set, which as noted above represents bona fide drugs. The parameter used to discriminate between compound classes is a TI_{opt}-cut, the value of which is continuously varied across its 0−1 range. Compounds above each TI_{opt}-cut are classified in the sensitivity/specificity calculation as drugs, and otherwise they are classified as nondrugs. Using this scheme the number count is obtained for true positives, TP (i.e., compounds above the TI_{opt}-cut that are in the FDA set), false negatives, FN (compounds below the TI_{opt}-cut that are in the FDA set), and the converse false positive, FP, and true negative, TN, counts for the non-FDA compounds. The results of these counts can be used to construct a receiver operator curve (ROC) that provides a measure of how well an outcome, TW50 in this case, distinguishes the FDA approved oncology agents from all the xenograft tested compounds. This ROC is shown in Figure 7, with the finding that at the "best" selection of the TI_{opt}-cut parameter, defined as the point on the ROC of maximal deviation from the random (i.e., diagonal) case, ~70% coverage is observed, i.e., 0.7 of the FDA drugs are recovered, with only a 20% false detection rate (FDR), i.e. more that 80% of the nondrugs are correctly identified as such. These results provide quantifiable support

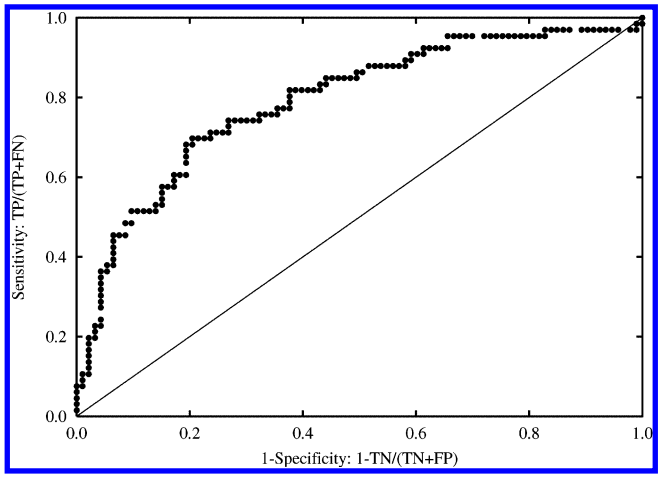

Figure 7. ROC analysis of FDA-approved oncology agents. The receiver-operator characteristic (ROC) for testing the optimal TI as a selective parameter to differentiate FDA-approved anticancer drugs from a random selection of xenograft tested compounds. Sensitivity (*x*-axis) measures how well we detect FDA-approved agents, whereas specificity (*y*-axis) measures how well we can identify agents not in the FDA set. Using this graph it is possible to select an optimal TI cut sufficient to select both positive and negative identification to 70% accuracy.

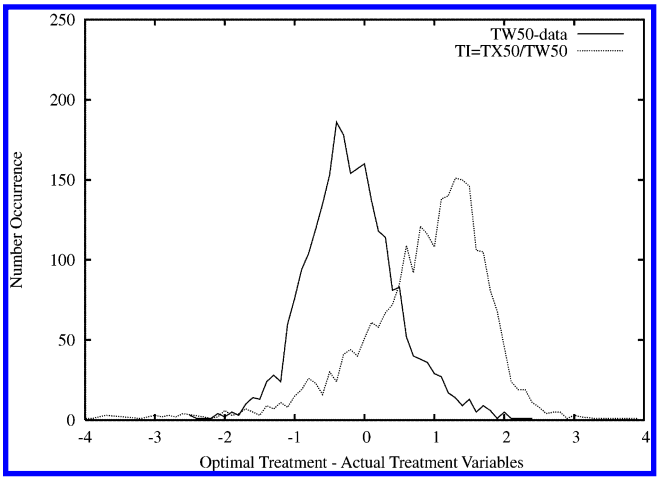

Figure 8. Distribution of the effect of optimization of treatment options. Distribution function of concentration efficacies (logarithmic scale) for the TW50 data set when taking into account possible variations in treatment to achieve an optimal result. Variation in treatment can typically change the dosage by an order of magnitude. A similar variation in treatment also indicates that the therapeutic index TX50/TW50 can change even more than efficacy.

for the use of a data-derived parameter, TW50 in this case, as a useful predictor of clinical efficacy for this set of FDA approved compounds. This result is in contrast to the previously published findings of limited xenograft utility for predicting clinical efficacy, albeit completely different calculations have been used in this comparison.

Implication for Varying Treatment Parameters. In the linear model of xenograft tumor weight reduction (TW50) and survival time increase (ST150) endpoints, *Y*, for compound *i*, administered under the specified treatment conditions can be given generally by

$$Y_i = \sum_j B_j X_j \tag{5}$$

which we can write as a function of our previous three variable sets as

$$Y_i = \sum_j B_j X_j^{\text{treatment}} + \sum_k B_k X_k^{\text{chemistry}} + \sum_l B_l X_l^{\text{biology}} \tag{6}$$

or equivalently by denoting each summation as a separate function as

$$Y_i = F^{\text{treatment}}(X) + F^{\text{chemistry}}(X) + F^{\text{biology}}(X) \tag{7}$$

We can now define an estimator of the relationship between *treatment* and xenograft outcome against the other two variables, i.e.

$$\hat{F}^{\text{treatment}}(X) = Y_i - F^{\text{chemistry}}(X) - F^{\text{biology}}(X) \tag{8}$$

Since the treatment variables are discrete, i.e., there can only be one injection site, there can only be one treatment schedule, and for each xenograft experiment we can find all treatment conditions that have a particular outcome. In principle we could then estimate the optimal treatment schedule for any particular drug and outcome. If we define the optimal treatment as requiring the least amount of drug we can vary the allowable treatment parameter to find the best $F^{\text{treatment}}(X)$, i.e., the most negative value.

$$F_{\text{opt}}^{\text{treatment}}(X) = \max\{\sum_j B_j X_j^{\text{treatment}}\} \tag{9}$$

In Figure 8 we show the distribution of the differences between optimal treatment schedules and actual treatment schedules in the xenograft data.

$$D = F_{\text{opt}}^{\text{treatment}}(X) - F^{\text{treatment}}(X) \tag{10}$$

The width of this distribution, at half its height, gives a representative change of TW50 concentration between the optimal versus overall treatment schedule of 1.3 log-units, i.e., scheduling effects can typically change the required dosage to achieve the same effect by more than an order of magnitude. Similarly we can also optimize the therapeutic index for the endpoints available in this data set. If we use the definition of TI_{TW} as TX50/TW50 the analogous distribution of *D*-values defined above are also shown in Figure 8. The half-width value of 1.5 log units for this distribution is comparable to that found above for TW50; however, the complete distribution is right-shifted to higher values. This result is an indication that treatment schedules can have a profound effect on the balance between toxicity and efficacy, even under highly controlled xenograft conditions. Alternatively, interpretations of xenograft data themselves are less than straightforward.

Therapeutic Index and Drug Mechanism of Action. The growth inhibition ($-\log(GI_{50})$) pattern across the NCI-60 tumor cell screen resulting from drug exposure has been used to identify and classify drug mechanism of action. We have previously used GI_{50} measurements to map regions in response space in order to classify broader groups of drug mechanism of action, such as antimitotic agents, membrane active agents, agents targeting nucleic acid metabolism, metabolic stress, and cell survival as well as agents involved in targeting kinases/phosphatases signaling and oxidative stress.[30,32,33] We have used this framework to explore global

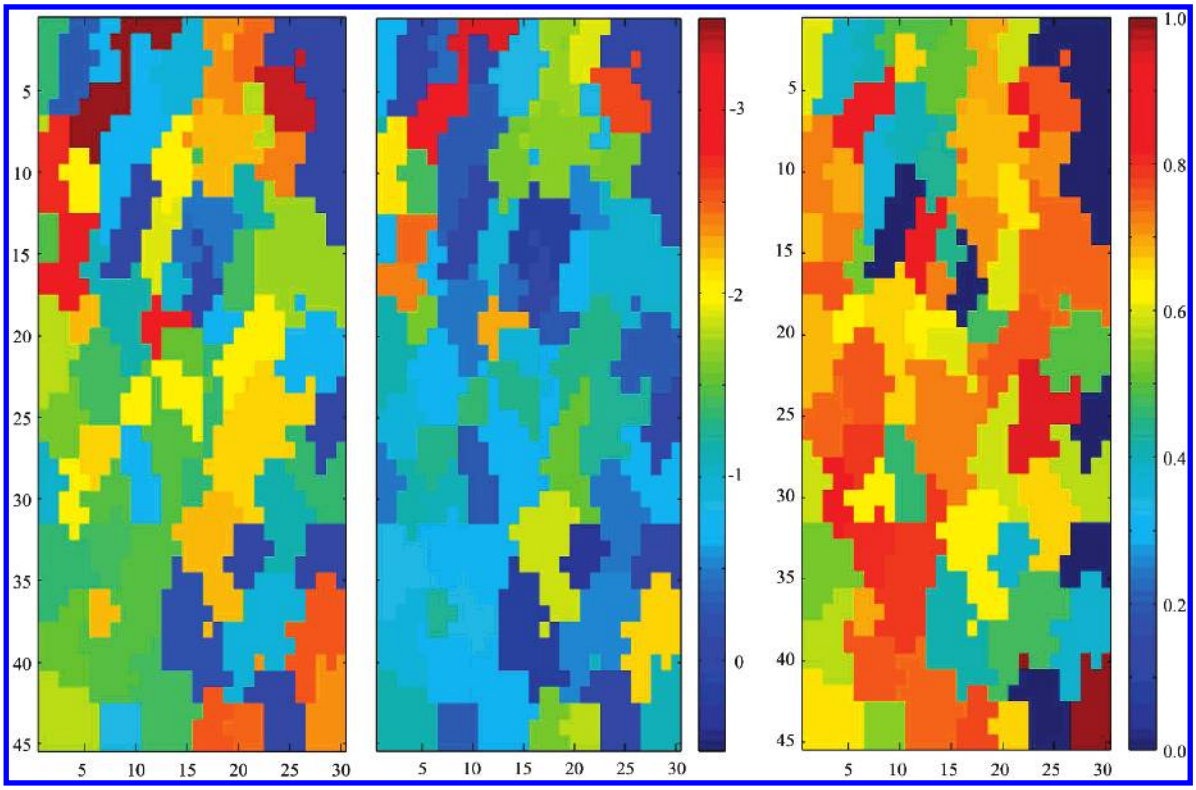

Figure 9. Distribution of efficacy and toxicity values on the GI_{50} SOM. Using the previously developed self-organizing map of the GI_{50} data vectors we can map xenograft outcome to regions of the SOM via drugs that have been tested in both assay systems. The division of the SOM into regions is reflective of different mechanisms of action; here these regions are used as a delimiter for the average xenograft properties derived from the compounds in this region. Specific patterns of efficacy, toxicity, and therapeutic index are connected to different mechanisms of drug action via the SOM. Potent compounds in the xenograft assay are shown as red in the left graph; however, their toxicity also tends to be large which restricts their therapeutic index and hence potential clinical usefulness. The TW50 and TX50 concentration variations span 4 orders of magnitude. The TIs in the right most graph range from 1 to the minimum of 10 or the lowest TI in the subregion.

properties of the xenograft data. In particular we examine the therapeutic index (TI) for compounds that have been tested both in the xenograft system and the NCI60 tumor cell growth inhibition assay. These compounds and their associated data can be placed into distinct subregions on the mechanism of action SOM via each compound's GI_{50} screening profile. Subregions of the map have proven to be a good compromise between individual GI_{50} data vectors and mechanism of action classes identified via literature.[30] In Figure 9 we show the subregion averaged efficacy, toxicity, and therapeutic index for the survivial time (ST150) endpoint in the xenograft data. In Figure 9A,B, dark blue colors identify subregions requiring the highest concentration to achieve their effect, i.e., lower potency compounds, while dark red indicates the most potent compounds. In general the most potent compounds are found in map regions known to be associated with antimitotic activity. The mapping of TI on the mechanism of action map in Figure 9C shows the highest TI as dark red. Variations in TI are visible across the map, with no specific regionalization, thus the most potent compound groups are not necessarily associated with a good TI. It is noteworthy that the subregion known to contain agents that affect oxidative metabolism, via targeting the mitochondrial system I electron transport system, have a consistently high TI. Although the literature is sparse for associating TIs to mechanism of action, the known mitochondrial system I targeting acetogenins appears to achieve a high TI.[36]

Particular compound groups can also be investigated with respect to their average therapeutic index. Here we have identified the corresponding xenograft data that belong to known sets of compound classes.[4] In Table 2 we have collected the average TI, TW50, and GI_{50} for 15 drug classes. Associated with these compound classifications is a measure of structural similarity and biological response similarity for members within each group. The similarities in structure and biological response provide a measure of how well these compound classes are defined, e.g., topoisomerase I inhibitors have both a high structural similarity and biological response similarity, while the group of alkylators has an essentially random structural and biological response similarity. The therapeutic index for these compound classes is not correlated with any of the other quantities in Table 2. While the GI_{50} and the efficacy data are correlated, as also noted in Figure 4, the toxicity of these compounds is not related to their efficacy.

Pathway Fitness and Toxicity. The identification of pathways as a more reliable indicator of drug efficacy in the growth inhibition data,[32,33] and as a possible classifier for detecting cancer genes,[37] also raises the possibility of relating xenograft data to specific biological processes that are important in cancer. In our terminology, a pathway consists of a group of genes that have been independently identified as contributing to an aspect of biological function that has been annotated in either the KEGG, BIOCARTA, or GO schemes. While the concept of a pathway lacks a standard definition, it is gaining support as a means to better understand cellular biology.[38,39] The identification of co-regulated genes in pathways via mRNA gene expression profiles measured across the NCI-60 cell lines has established

Table 2. TI Properties for Selected Compound Groups[a]

drug class	<logTI>	<−log(TW50)>	N_{TI}	<−log(GI_{50})>	$<t>_{intra}$	$<r>_{intra}$
alkylators	0.37	1.60	234	4.91	0.20	0.29
antibiotics	0.32	2.77	111	6.91	0.52	0.45
antifolates	0.26	1.67	46	5.33	0.53	0.56
channel agents	0.43	0.97	30	4.85	0.47	0.26
DNA polymerase inhibitors	0.31	0.53	51	4.64	0.80	0.60
direct membrane	0.23	1.20	60	5.15	0.45	0.36
Golgi disruptive agents	0.25	2.52	11	6.33	0.49	0.20
intercalating agents	0.29	1.13	23	5.66	0.45	0.31
mitotic	0.34	2.57	18	6.69	0.50	0.48
phosphatase/kinase	0.30	2.47	33	5.37	0.42	0.41
purine antimetabolites	0.61	1.42	84	4.80	0.40	0.52
pyrimidine antimetabolites	0.35	0.52	112	4.85	0.51	0.32
steroids	0.59	2.01	36	6.07	0.35	0.23
topoisomerase I	0.58	2.16	45	6.22	0.75	0.78
topoisomerase II	0.38	2.10	106	6.15	0.32	0.44

[a] The average xenograft data and GI_{50} growth inhibition data for a set of selected compound groups potentially relevant for anticancer therapy. For this group definition there is no overall correlation between the toxicity and the growth inhibition potency, though the three most active groups, antibiotics, antimitotic, and Golgi disruptive agents are the most potent groups in both assay systems.

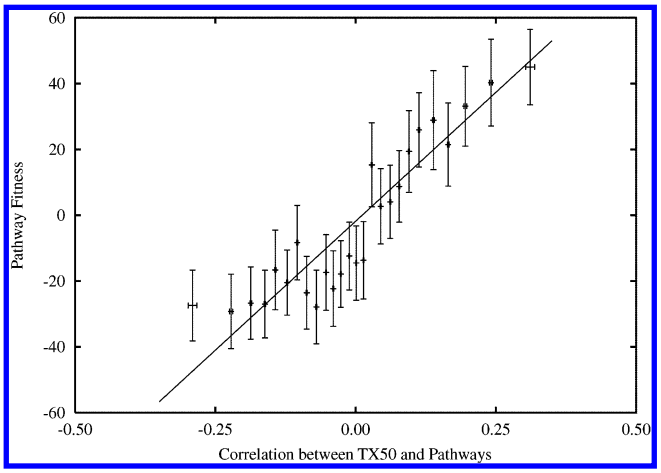

Figure 10. Pathway fitness. Relationships between pathway fitness and correlation between pathway coherence and toxicity (TX50). Coherent pathways usually belong to biological processes that need to be tightly regulated both in time and space. The graph suggests that coherent pathways correspond to compounds with the highest potency. Conversely, compounds associated with more cohesive pathways are more potent than compounds associated with less coherent pathways.

two general categories of pathways: cohesive and noncohesive pathways.[32,33] The cohesive pathways contain genes that are significantly coregulated, i.e., the gene expression for pairs of genes in a pathway match expression values across the NCI-60 GI_{50} profile, compared to a random set of genes. Groups of genes that are required to be close both in space and time, e.g., proteins of the ribosome and proteins in mitosis, are found to be cohesive.[33] The concept of linking the growth inhibition data to specific pathways via a self-organizing map was introduced by us and provides insight into drugs or sets of drugs that are hypothesized as targeting a particular pathway, and its implied function, in the cell.[32]

Previously defined pathway coherence levels[32,33] can be projected on to the GI_{50} SOM. These coherence values can be compared to their subregion-matched toxicity scores, as projected in Figure 9, to obtain a measure of correlation strength. These correlations, plotted in Figure 10, reflect a strong positive relationship between pathway coherence and toxicity/efficacy. This observation is an indication that the pathways identified as coherent appear to be intrinsically more susceptible to chemical intervention, i.e., a greater degree of compound efficacy. The lack of coherence in a pathway does not necessarily preclude its ability to target individual pathway members; it does imply that the function of the pathway itself may be more challenging to affect via drug intervention when compared to more coherent pathways. Tightly regulated processes such as mitochondrial electron transport, protein biosynthesis, and centromere and kinetochore complex maturation would thus be potentially good drug targets, if these pathways can be differentiated between cancer and normal tissue. Likewise, less appropriate pathway targets include induction of apoptosis, chromatin remodeling and methylation, and signal transduction via kinase/phosphatases as well as lipid, protein, and ion transport pathways. However, if these pathways show a change in coherence between normal and cancer tissue, they too may become good targets for specific cancers.

DISCUSSION/CONCLUSION

The DTP xenograft repository provides a potentially important source for chemoinformatic studies of relevant data in the quest for better cancer therapies. While these data are mature, in the sense that the bulk of them were collected decades ago, their ability to match some phase II clinical trial efficacy results have been observed in prior studies.[14,28] Positive xenograft activity remains an important property in propelling a compound through the myriad of stages, usually ending in clinical recommendation for a limited number of compounds. Here, we have recast these data in a chemoinformatic perspective that allows a consistent evaluation of compound efficacy useful for retrospective explorations into aspects of these data that might otherwise have been overlooked. A linear model of xenograft efficacy was constructed based on treatment, chemistry, and biology. Using this model, explorations of the sensitivity of xenograft efficacy to variations in these model variables find a strong dependence on treatment scheduling, with a greater than 10-fold difference in efficacy between optimal scheduling versus the least effective schedule. Receiver operator characteristic analysis finds a 70% accuracy between model predicted xenograft efficacy and FDA approved oncology agents, with a false detection rate of only 20%. Taken together, these results support a xenograft model based on preclinical in vitro

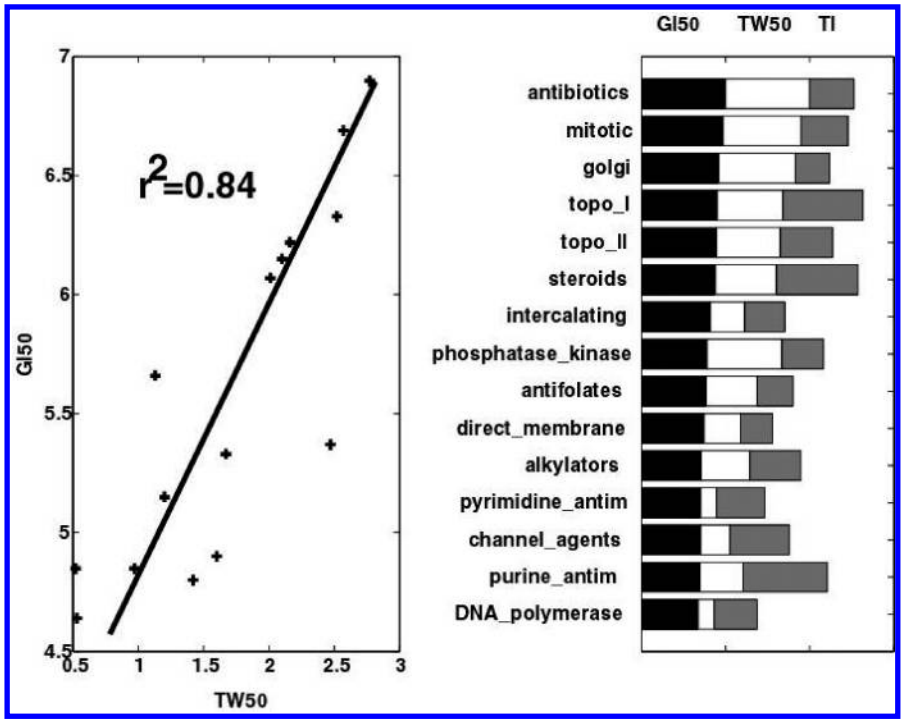

Figure 11. Averaged GI50, TX. and TI for selected compound classes. Plot at the left represents columns 2 and 4 of Table 3. The histograms appearing at the left of this figure represent averaged GI50, TW50, and TI from Table 2.

screening results and chemoinformatic characterization as potentially useful for prioritizing compounds for development.

Our model yields additional chemical insights with respect to compound classes and considerations of efficacy, toxicity, and preclinical in vitro screening results. Figure 11 provides a graphical representation of the data in Table 2. The left panel plots the average GI_{50} values for each of the 15 compound classes versus the average tumor weight reduction (TW50) in xenograft tests. The observed correlation coefficient (r^2=0.84) supports the relevance of in vitro screening data in predicting xenograft outcome. The histogram at the right of this figure displays the average GI_{50}, TX, and TI for each compound class, ordered by in vitro potency; i.e. the left-hand plot in this figure represents the GI_{50} and TX components for each histogram bar shown at the right. A noteworthy outlier in this figure is the phosphatase/kinase drug class, where high TW50 values occur at a log lower GI_{50} concentration when compared to the most potent compound classes that yield the greatest tumor weight reduction. This finding suggests the potential importance of considering lower potency compounds in the in vitro NCI-60 tumor cell screen, especially if they act on phosphatase/kinase targets. This comes with a price since a lower TI is associated with this compound class. An additional observation involving a therapeutic index finds that although TI does not significantly correlate with either GI_{50} or TX, there are important trends evident in this figure. The longest bars this histogram represent the best case scenario for the three measures of GI_{50}, TX, and TI; examples being steroids, topoisomerase I, mitotic, and the antibiotic drug classes. TI for the phosphatase/kinase drug class is among the lowest in this set and provides an indication that agents in this class may pose challenges for designing treatment modalities that reduce toxicity. Although outside the scope of our analysis, microdosing strategies and their effect on xenograft outcome could be tested for selected drug classes.

A few caveats accompany our results. First the ability of our model to retrospectively predict experimental outcome suggests that it may have prospective predictive power. However, the considerable variability in treatment modalities across these xenograft experiments can only partly be taken into account, and hence predicting treatment based on desired outcome is limited by the small number of scheduling parameters available for modeling. Second, our current model of toxic outcome was not associated with any particular adverse reaction or organ failure but focused only on animal survival. The extension of any xenograft model, such as the one proposed here, into phase I clinical trial data/observations is certainly possible, but it requires access to a large amount of clinical data. The ability to model xenograft outcome, and use these results to optimize experimental designs, is an additionally important aspect when considering strategies for reduced animal testing. The successful modeling of compound toxicology at the FDA[40,41] emphasizes the importance of in silico models as a practical tool for regulatory evaluation. And finally, the analysis presented here represents one perspective of this extensive data set. Additional and alternative efforts to maximize the information contained in this data set will be needed to enhance future anticancer drug discovery efforts.

In summary, our model-based analysis relating chemistry, as derived from chemical descriptors, to biology, as defined by xenograft outcome, provides the opportunity to systematically connect these data to agents investigated in the NCI-60 cell line growth inhibition assay. Further associations with cellular pathways are established that link biological processes to in vitro screening potency and xenograft efficacy. These results extend the previous reports of a statistical association between in vitro NCI-60 screening potency and xenograft efficacy into a chemoinformatics arena than can be used prospectively to assess compounds evaluated in the NCI-60 screen but not yet considered for xenograft testing.

METHODS

DTP Resources of GI_{50} Data. The NCI-60 tumor cell drug discovery panel was developed as a tool to assess anticancer activity of compounds against a range of cell lines derived from different tumors, including lung, renal, colorectal, ovarian, breast, prostate, central nervous system, melanoma, and hematological malignancies.[42] The data consist of concentration values (GI_{50}) for each cell line at which the drug results in a 50% reduction in the net protein increase relative to untreated control cells during a 48 h drug incubation. The GI_{50} data vectors used in our analysis were log-transformed and selected to have a maximum of 20 missing data elements and a signal covariance of at least 0.02. The concentration of $-\log(GI_{50})$ values typically ranges from 4.0 to 8.0. The pattern of GI_{50} measurements across the tumor cell lines has proven useful for identifying mechanisms of action for some drug classes and aids in the classification of novel drugs submitted to the NCI's tumor screen.[30,43,44]

DTP Resources for Xenograft Data. The protocols used for the xenograft data accumulations can be found at http://www.dtp.nci.nih.gov/branches/btb/available_documents.html. The DTP xenograft assay data can be downloaded from the ftp-site on PubChem at ftp://ftp.ncbi.nlm.nih.gov/pubchem/Bioassay/.

Self-Organizing Map of GI_{50} Data. Cytotoxicity measurements for nearly 30 000 small molecules screened against the NCI's tumor cell panel were clustered into 1350 groups or nodes using a SOM.[30,32,33] Each node on this map represents a set of similar cytotoxicity responses which can be separated into nine major response categories (regions) based on literature references: mitosis (M), membrane function (N), nucleic acid metabolism (S), metabolic stress and cell survival (Q), kinases/phosphatases, oxidative stress (P), and four unexplored regions R, F, J, and V. Each of these regions is further divided into a total of 80 clades (a clade is a group of clusters that share similar cytotoxic responses) or subregions: M_1-M_8, N_1-N_{13}, P_1-P_8, Q_1-Q_7, R_1-R_7, S_1-S_{13}, F_1-F_8, J_1-J_8, and V_1-V_8.

Gene Expression Data. Constitutive gene expression data from Novartis, measured in triplicate across the 60 tumor cell lines using the Affymetrix DNA oligonucleotide microarray technology, were downloaded from the Developmental Therapeutics Program (DTP) Web server at http://www.dtp.nci.nih.gov. This data set contains 12 626 mRNA expression profiles and is publicly available. The data set is first filtered to include only measurements that exhibited the strongest intensity in signal ($p \leq 0.05$). The logarithm of each signal is taken to suppress extreme data values. Replicate measurements for each gene are then averaged by taking the median. Finally, only gene expression profiles having data available for at least 40 cell lines are included yielding a total of 4923 genes for the pathway analysis.

FDA Approved Anticancer Drugs. The compilation of current anticancer drugs in the clinic and their specification can be found at the FDA Oncology Tools site: http://www.accessdata.fda.gov/scripts/cder/onctools/druglistframe.htm.

Molecular Descriptors. The molecular descriptor set is obtained by calculating the families of fast descriptors, which include 214 descriptors, in the Cerius^2 v. 4.10 molecular modeling software from Accelrys Inc. (San Diego, CA). We also included the Leadscope unique feature count (LSUFC), calculated by counting the number of unique structural features defined by Leadscope present in a molecule, as an additional parameter to cover the structural complexity aspect of the molecules, resulting in a total of 215 independent variables (descriptors) to describe the chemical nature of each compound.

Stepwise Linear Regression. A linear regression model of xenograft data was employed to explore optimal solutions of xenograft outcome (**Y**) defined by a set of variables (**X**) linearly dependent on xenograft treatment parameters, chemical descriptors, and biological activity captured in the GI_{50} measurements. The formal solution to the problem of finding the optimal coefficients (**B**) is given by the matrix equations:

$$\mathbf{Y} = \mathbf{X}\,\mathbf{B} \tag{11}$$

$$\mathbf{X'Y} = (\mathbf{X'X})\,\mathbf{B} \tag{12}$$

$$\mathbf{B} = (\mathbf{X'X})^{-1}\,(\mathbf{X'Y}) \tag{13}$$

The number of dependent variables included in the solution for **B** is iteratively constructed by selecting the variable that contributes the most to the solution. Variables are added until their contribution exceeds a *p*-value of 0.05. This results in a typical parameter set of 50 variables to model the outcome. The range for the ratio of variables to observables is less than 1%.

Correlation Coefficient. The Pearson or sample correlation coefficient of two data vectors $\vec{u}$ and $\vec{v}$ is defined as

$$r(\vec{u},\vec{v}) = \frac{\sum_i (u_i - \bar{u})\cdot(v_i - \bar{v})}{\sqrt{\sum_i (u_i - \bar{u})\cdot\sum_i (v_i - \bar{v})}} \tag{14}$$

where $\bar{u}$ denotes the average of all elements in$\vec{u}$. The correlation coefficient measures the fidelity of a linear fit of v(u) and takes on values between -1 and $+1$. A correlation coefficient of 1 indicates that each vector is linearly dependent on the other. It does not mean that the vectors are the same.

Pathways. Three databases are used for pathway gene analysis: Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.ad.jp/kegg/), GO (http://www.geneontology.org/), and BioCarta (http://www.biocarta.com/). Annotations for 134 human pathways containing 2804 genes are downloaded from the KEGG ftp site (ftp://ftp.genome.ad.jp/pub/kegg/pathways/hsa/). BioCarta annotations for 314 pathways containing 1406 human genes are downloaded from NCI's Cancer Genome Anatomy Project (CGAP, http://cgap.nci.nih.gov/) ftp site (ftp://ftp1.nci.nih.gov/pub/CGAP). Annotations for 3564 GO terms containing 10 921 human genes are downloaded from the GO ftp site (ftp://ftp.geneontology.org/pub/go/). Within our gene expression data set, 1047 genes are present in KEGG, 604 are present in BioCarta, and 3210 are present in GO.

Pathway$-GI_{50}$ SOM Clade Correlation Scores. For each pathway and each SOM clade, the clade correlation coefficients of all genes in the pathway and the genes that are not in the pathway are compared as two sample

populations using the Kruskal−Wallis rank sum procedure and a *H* statistic (*H*-score).[33] A large *H* ($H > 3.84$) indicates a statistically significant difference ($p \leq 0.05$) between the two sample populations. A negative sign is added to the *H*-score if the sum of ranks of the pathway clade−gene correlation coefficients is smaller than the nonpathway clade−gene correlation coefficients, that is, the GI_{50} clade correlates stronger with nonpathway genes than pathway genes.

ACKNOWLEDGMENT

This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under Contract NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services nor does the mention of trade names, commercial products, or organization imply endorsement by the U.S. Government. This research was supported in part by the Developmental Therapeutics Program in the Division of Cancer Treatment and Diagnosis of the National Cancer Institute. This project was funded by contract no. NO1-CO-12400.

REFERENCES AND NOTES

(1) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002**, *45* (19), 4350−8.
(2) Willett, P. Similarity-based approaches to virtual screening. *Biochem. Soc. Trans.* **2003**, *31* (Pt 3), 603−6.
(3) Voigt, J. H.; Bienfait, B.; Wang, S.; Nicklaus, M. C. Comparison of the NCI Open Database with Seven Large Chemical Structural Databases. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 702−712.
(4) Wallqvist, A.; Huang, R.; Thanki, N.; Covell, D. G. Evaluating chemical structure similarity as an indicator of cellular growth inhibition. *J. Chem. Inf. Model*. **2006**, *46* (1), 430−7.
(5) Maggiora, G. M. On outliers and activity cliffs−why QSAR often disappoints. *J. Chem. Inf. Model.* **2006**, *46* (4), 1535.
(6) Kelly, D. E.; Clark, A. Modern approaches to drug discovery and design: setting the scene. *Biochem. Soc. Trans.* **2003**, *31* (2), 428.
(7) Jonsdottir, S. O.; Jorgensen, F. S.; Brunak, S. Prediction methods and databases within chemoinformatics: emphasis on drugs and drug candidates. *Bioinformatics* **2005**, *21* (10), 2145−60.
(8) Brown, F. Editorial opinion: chemoinformatics-a ten year update. *Curr. Opin. Drug Discovery Dev.* **2005**, *8* (3), 298−302.
(9) Oprea, T. I. *Chemoinformatics in Drug Discovery*; Wiley-VCH: Weinheim, 2005; Vol. 23.
(10) Oprea, T. I.; Allu, T. K.; Fara, D. C.; Rad, R. F.; Ostopovici, L.; Bologa, C. G. Lead-like, drug-like or "Pub-like": how different are they? *J. Comput.-Aided Mol. Des.* **2007**, *21* (1−3), 113−9.
(11) Gagna, C. E.; Winokur, D.; Lambert, C. W. Cell biology, chemogenomics and chemoproteomics. *Cell Biol. Int.* **2004**, *28* (11), 755−64.
(12) Kerbel, R. S. Human tumor xenografts as predictive preclinical models for anticancer drug activity in humans: better than commonly perceived-but they can be improved. *Cancer Biol. Ther.* **2003**, *2* (4 Suppl 1), S134−9.
(13) Sausville, E. A.; Burger, A. M. Contributions of human tumor xenografts to anticancer drug development. *Cancer Res.* **2006**, *66* (7), 3351-4, discussion 3354.
(14) Voskoglou-Nomikos, T.; Pater, J. L.; Seymour, L. Clinical predictive value of the in vitro cell line, human xenograft, and mouse allograft preclinical cancer models. *Clin. Cancer Res.* **2003**, *9* (11), 4227−39.
(15) Teicher, B. A. Tumor models for efficacy determination. *Mol. Cancer Ther.* **2006**, *5* (10), 2435−43.
(16) Rogatko, A.; Babb, J. S.; Tighiouart, M.; Khuri, F. R.; Hudes, G. New paradigm in dose-finding trials: patient-specific dosing and beyond phase I. *Clin. Cancer Res.* **2005**, *11* (15), 5342−6.
(17) Sausville, E. A. Optimizing target selection and development strategy in cancer treatment: the next wave. *Curr. Med. Chem. Anticancer Agents* **2004**, *4* (5), 445−7.
(18) Suggitt, M.; Bibby, M. C. 50 years of preclinical anticancer drug screening: empirical to target-driven approaches. *Clin. Cancer Res.* **2005**, *11* (3), 971−81.
(19) Hariparsad, N.; Sane, R. S.; Strom, S. C.; Desai, P. B. In vitro methods in human drug biotransformation research: implications for cancer chemotherapy. *Toxicol. in Vitro* **2006**, *20* (2), 135−53.
(20) Schrag, D.; Garewal, H. S.; Burstein, H. J.; Samson, D. J.; Von Hoff, D. D.; Somerfield, M. R. American Society of Clinical Oncology Technology Assessment: chemotherapy sensitivity and resistance assays. *J. Clin. Oncol*. **2004**, *22* (17), 3631−8.
(21) Berger, D. P.; Henss, H.; Winterhalter, B. R.; Fiebig, H. H. The clonogenic assay with human tumor xenografts: evaluation, predictive value and application for drug screening. *Ann. Oncol.* **1990**, *1* (5), 333−41.
(22) Fiebig, H. H.; Maier, A.; Burger, A. M. Clonogenic assay with established human tumour xenografts: correlation of in vitro to in vivo activity as a basis for anticancer drug discovery. *Eur. J. Cancer* **2004**, *40* (6), 802−20.
(23) Langdon, S. P.; Hendriks, H. R.; Braakhuis, B. J.; Pratesi, G.; Berger, D. P.; Fodstad, O.; Fiebig, H. H.; Boven, E. Preclinical phase II studies in human tumor xenografts: a European multicenter follow-up study. *Ann. Oncol.* **1994**, *5* (5), 415−22.
(24) Gutmann, D. H.; Hunter-Schaedle, K.; Shannon, K. M. Harnessing preclinical mouse models to inform human clinical cancer trials. *J. Clin. Invest.* **2006**, *116* (4), 847−52.
(25) Shoemaker, R. H.; Wolpert-DeFilippes, M. K.; Kern, D. H.; Lieber, M. M.; Makuch, R. W.; Melnick, N. R.; Miller, W. T.; Salmon, S. E.; Simon, R. M.; Venditti, J. M.; et al. Application of a human tumor colony-forming assay to new drug screening. *Cancer Res.* **1985**, *45* (5), 2145−53.
(26) Marsh, J. C.; Shoemaker, R. H.; Salmon, S. E.; Kern, D. H.; Venditti, J. M. Relationship between in vitro tumor stem cell assay and in vivo antitumor activity using the P388 leukemia. *Int. J. Cell Cloning* **1988**, *6* (1), 60−8.
(27) Goldin, A.; Johnson, R. K.; Venditti, J. M. Usefulness and limitations of murine tumor models for the identification of new antitumor agents. *Antibiot. Chemother.* **1980**, *28*, 1−7.
(28) Johnson, J. I.; Decker, S.; Zaharevitz, D.; Rubinstein, L. V.; Venditti, J. M.; Schepartz, S.; Kalyandrug, S.; Christian, M.; Arbuck, S.; Hollingshead, M.; Sausville, E. A. Relationships between drug activity in NCI preclinical in vitro and in vivo models and early clinical trials. *Br. J. Cancer* **2001**, *84* (10), 1424−31.
(29) Decker, S.; Hollingshead, M.; Bonomi, C. A.; Carter, J. P.; Sausville, E. A. The hollow fibre model in cancer drug screening: the NCI experience. *Eur. J. Cancer* **2004**, *40* (6), 821−6.
(30) Rabow, A. A.; Shoemaker, R. H.; Sausville, E. A.; Covell, D. G. Mining the National Cancer Institute's tumor-screening database: Identification of compounds with similar cellular activities. *J. Med. Chem.* **2002**, *45*, 818−840.
(31) Covell, D. G.; Wallqvist, A.; Huang, R. L.; Thanki, N.; Rabow, A. A.; Lu, X. J. Linking tumor cell cytotoxicity to mechanism of drug action: An integrated analysis of gene expression, small-molecule screening and structural databases. *Proteins: Struct. Funct. Bioinformatics* **2005**, *59* (3), 403−433.
(32) Huang, R.; Wallqvist, A.; Thanki, N.; Covell, D. G. Linking pathway gene expressions to the growth inhibition response from the National Cancer Institute's anticancer screen and drug mechanism of action. *Pharmacogenomics J.* **2005**, *5* (6), 381−99.
(33) Huang, R.; Wallqvist, A.; Covell, D. G. Comprehensive analysis of pathway or functionally related gene expression in the National Cancer Institute's anticancer screen. *Genomics* **2005**.
(34) Wheeler, D. L.; Barrett, T.; Benson, D. A.; Bryant, S. H.; Canese, K.; Chetvernin, V.; Church, D. M.; DiCuccio, M.; Edgar, R.; Federhen, S.; Geer, L. Y.; Helmberg, W.; Kapustin, Y.; Kenton, D. L.; Khovayko, O.; Lipman, D. J.; Madden, T. L.; Maglott, D. R.; Ostell, J.; Pruitt, K. D.; Schuler, G. D.; Schriml, L. M.; Sequeira, E.; Sherry, S. T.; Sirotkin, K.; Souvorov, A.; Starchenko, G.; Suzek, T. O.; Tatusov, R.; Tatusova, T. A.; Wagner, L.; Yaschenko, E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2006**, *34* (database issue), D173−80.
(35) Huang, R. L.; Wallqvist, A.; Covell, D. G. Anticancer metal compounds in NCI's tumor-screening database: putative mode of action. *Biochem. Pharmacol.* **2005**, *69* (7), 1009−1039.
(36) Raynaud-Le Grandic, S.; Fourneau, C.; Laurens, A.; Bories, C.; Hocquemiller, R.; Loiseau, P. M. In vitro antileishmanial activity of acetogenins from Annonaceae. *Biomed. Pharmacother.* **2004**, *58* (6−7), 388−92.
(37) Bild, A. H.; Yao, G.; Chang, J. T.; Wang, Q.; Potti, A.; Chasse, D.; Joshi, M. B.; Harpole, D.; Lancaster, J. M.; Berchuck, A.; Olson, J. A., Jr.; Marks, J. R.; Dressman, H. K.; West, M.; Nevins, J. R. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **2006**, *439* (7074), 353−7.
(38) Dopazo, J. Functional interpretation of microarray experiments. *OMICS* **2006**, *10* (3), 398−410.
(39) Curtis, R. K.; Oresic, M.; Vidal-Puig, A. Pathways to the analysis of microarray data. *Trends Biotechnol.* **2005**, *23* (8), 429−35.

(40) Matthews, E. J.; Kruhlak, N. L.; Weaver, J. L.; Benz, R. D.; Contrera, J. F. Assessment of the health effects of chemicals in humans: II. Construction of an adverse effects database for QSAR modeling. *Curr. Drug Discovery Technol.* **2004**, *1* (4), 243−54.

(41) Matthews, E. J.; Kruhlak, N. L.; Benz, R. D.; Contrera, J. F. Assessment of the health effects of chemicals in humans: I. QSAR estimation of the maximum recommended therapeutic dose (MRTD) and no effect level (NOEL) of organic chemicals based on clinical trial data. *Curr. Drug Discovery Technol.* **2004**, *1* (1), 61−76.

(42) Monks, A.; Scudiero, D.; Skehan, P.; Shoemaker, R.; Paull, K.; Vistica, D.; Hose, C.; Langley, J.; Cronise, P.; Vaigro-Wolff, A.; et al. Feasibility of a high-flux anticancer drug screen using a diverse panel of cultured human tumor cell lines. *J. Natl. Cancer Inst.* **1991**, *83* (11), 757−66.

(43) Keskin, O.; Bahar, I.; Jernigan, R. L.; Beutler, J. A.; Shoemaker, R. H.; Sausville, E. A.; Covell, D. G. Characterization of anticancer agents by their growth inhibitory activity and relationships to mechanism of action and structure. *Anticancer Drug Des.* **2000**, *15* (2), 79−98.

(44) Paull, K. D.; Shoemaker, R. H.; Hodes, L.; Monks, A.; Scudiero, D. A.; Rubinstein, L.; Plowman, J.; Boyd, M. R. Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. *J. Natl. Cancer Inst.* **1989**, *81* (14), 1088−92.

CI700132U