

Anticancer medicines in development: assessment of bioactivity profiles within the National Cancer Institute anticancer screening data

David G. Covell,¹ Ruili Huang,¹
and Anders Wallqvist²

¹National Cancer Institute-Frederick, Developmental Therapeutics Program, Screening Technologies Branch, Laboratory of Computational Technologies and ²Laboratory of Computational Technologies, Science Applications International Corporation-Frederick, Inc., National Cancer Institute-Frederick, Frederick, Maryland

Abstract

We present an analysis of current anticancer compounds that are in phase I, II, or III clinical trials and their structural analogues that have been screened in the National Cancer Institute (NCI) anticancer screening program. Bioactivity profiles, measured across the NCI 60 cell lines, were examined for a correspondence between the type of cancer proposed for clinical testing and selective sensitivity to appropriately matched tumor subpanels in the NCI screen. These results find strongest support for using the NCI anticancer screen to select analogue compounds with selective sensitivity to the leukemia, colon, central nervous system, melanoma, and ovarian panels, but not for renal, prostate, and breast panels. These results are extended to applications of two-dimensional structural features to further refine compound selections based on tumor panel sensitivity obtained from tumor screening results. [Mol Cancer Ther 2007;6(8):2261–70]

Introduction

The process of discovering new anticancer therapies often begins with screening campaigns to identify potential leads (1). Charting the course of an effective screening campaign involves, minimally, choices about compounds (2, 3), screens (4, 5), and analysis methods (4, 6). Notable

variations of this theme are continuously evolving due to the availability of larger compound libraries (7), technological improvements in assay readouts (8), and new strategies for analyzing large, high-content data sets (9–11). An often overlooked component of modern discovery efforts involves retrospective assessments of screening results for compounds currently tested in clinical trials (12). Here, we examine the preclinical screening performance of clinical trial candidates as well as their structural analogues that have also been screened in the National Cancer Institute (NCI) tumor panel. Our analysis will assess whether there is statistical support for a relationship between *in vitro* screening sensitivity within specific tumor types and the cancer target of each clinical trial proposed for these candidates.

Historically, the NCI Tumor Panel Screening program (referred to hereafter as the NCI-60), which began in 1990, has offered a unique readout for small-molecule and natural product extract screening against a selected set of immortalized human cancer cell lines (13–15). Conventional analyses of the NCI-60 bioactivity data have used the COMPARE program (Pearson correlation coefficients) to identify putative mechanisms of action (16) and, in some cases, lead compounds for further development (17, 18). More recent analyses of the NCI-60 data set have used alternative means to relate similarities in NCI-60 bioactivity profiles (19–22) to reveal the considerable information yet available for mining this unique data set (20, 23–27).

The strategy proposed here revisits the NCI-60 data with specific focus on the performance of clinical anticancer candidates currently in phase I, II, or III trials. The analysis consists of two sequential steps. Briefly, electronic representations are obtained for a published set of clinical candidates (the “parent” set)³ and used to collect two-dimensional structural analogues screened in the NCI-60 (the “analogue” set). The z-score normalized NCI-60 bioactivity profiles, as measured by $-\log(GI_{50})$, are evaluated for these compounds to assess the correspondence between the type of clinical trial proposed for these anticancer agents and appropriately matched selective subpanel sensitivity in their bioactivity profiles. Our findings support selective preferences for only a subset of candidates and their associated clinical trials. Strategies are provided for relating two-dimensional structural features of selected compounds to their NCI-60 subpanel sensitivity. The knowledge gained from this analysis may offer a “reverse engineering strategy” for selecting previously untested compounds as potential anticancer candidates.

Received 3/20/07; revised 5/25/07; accepted 6/18/07.

Grant support: Federal funds from the National Cancer Institute, NIH, under contract no. NO1-CO-12400, and the Developmental Therapeutics Program of the National Cancer Institute Division of Cancer Treatment and Diagnosis.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: David G. Covell, National Cancer Institute-Frederick, Developmental Therapeutics Program, Screening Technologies Branch, Laboratory of Computational Technologies, Frederick, MD 21702. Phone: 301-846-5785; Fax: 301-846-6978. E-mail: covell@ncifcrf.gov

Copyright © 2007 American Association for Cancer Research.

doi:10.1158/1535-7163.MCT-06-0787

³ <http://www.phrma.org>

Materials and Methods

Anticancer Medicines in Development

A recently published³ collection of 399 anticancer medicines in development (AMID) will be our parent set. Each of these compounds has recently reached phase I, II, or III anticancer trials. These AMIDs are therapeutically directed against 19 cancer classes plus designations of "solid," "other," "chemo," and "unspecified." Solid tumor types have been designated as a therapeutic modality for more than 80 of the AMIDs, whereas fewer than 10 AMIDs have stomach cancer, bladder cancer, or neuroblastoma as their cancer trial target. A majority ($n = 47$) of these AMIDs are directed against two cancer types whereas five AMIDs have been proposed for clinical testing against eight or more cancer types.⁴

The two-dimensional structural features of the ~43,000 small molecules in the NCI-60 screen are queried for analogues to our AMID set (28). Summary statistics for these structural analogues are displayed as a histogram (Fig. 1), stratified according to two-dimensional structural similarity as measured by a Tanimoto score. Similarity scores based on two-dimensional structural representations do not take into account stereoisomers. Thus, a perfect Tanimoto score of 1.0 will not necessarily indicate structurally identical compounds. A compound tested in the NCI-60 screen is referred to as an "NSC" compound. The number of structurally unique screened compounds and AMIDs in the 0.5 Tanimoto set are 4,977 and 159, respectively. These numbers progressively decrease with increasing Tanimoto similarity (28) to yield 65 NSCs with a Tanimoto score of 1.0 to 57 AMIDs. Whereas our choice of a lower Tanimoto cutoff of 0.5 is arbitrary, it is motivated in part by findings that biologically similar activities often result from compounds having relatively weak structural similarity (25, 29). Our results will thus be reported across a broad Tanimoto range to provide a relatively comprehensive assessment of weakly as well as strongly similar two-dimensional structures.

Figure 1 also reveals the differential attrition between AMID and analogue NSC counts with increasing Tanimoto scores. The fractional retention of AMIDs with NSCs in their analogue set decreases to ~36% ($57/159 = 0.358$) at the highest Tanimoto score, whereas only 1.3% ($65/4,977 = 0.013$) of the original NSCs are retained at this cut. This observation reflects additional biological components, beyond structural similarity, that dictate selection of lead compounds as candidate drugs for clinical trials. Alternatively, as compounds travel further down the development pipeline, they may become less likely to appear in public domain data.

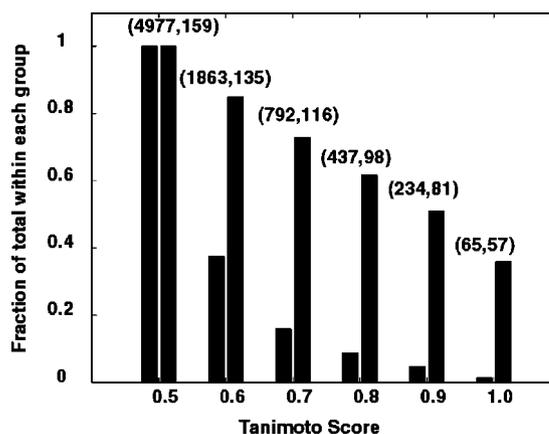


Figure 1. Histogram of NSC and AMID counts at Tanimoto scores ranging from 0.5 to 1.0. The left and rightmost columns and numbers represent counts for NSCs and AMIDs, respectively, at each Tanimoto score.

Cancer Classification

Our analysis will be restricted to the nine tumor types within the NCI-60 panel that have also been designated as a cancer classification for each AMID clinical trial.⁵ Whereas alternative resources exist for selecting compounds entering clinical trials (Clinical Trials Knowledge Area of Prous Science Integrity, the drug discovery and development portal⁶), our analysis will focus on the summaries provided at http://www.phrma.org/medicines_in_development as representative of the current selection of compounds available for cancer therapy. Table 1 stratifies the number counts for compounds in the screen and, in parentheses, AMID counts for each parent-analogue set stratified by cancer classification and Tanimoto similarity scores. Here, we list the total count of compounds in the NCI-60 for each of the clinical trial designations of the parent AMIDs. For example, there are 12 AMIDs currently undergoing clinical trials in brain cancer and there are 435 compounds tested in the NCI-60 that have a Tanimoto score >0.5. These numbers decrease to 6 AMIDs having 7 NSC compounds with a perfect Tanimoto score in this cancer class. The values listed in the "Totals" row of Table 1 reflect the cumulative NSC and AMID counts for each cancer trial; thus, AMIDs tested in more than one clinical trial are counted as separate occurrences in these totals. The last row of Table 1 represents the hit rate for the cumulative totals referenced to the values obtained at a Tanimoto cutoff of 0.5. These hit rates decline gradually with increasing Tanimoto score; however, the effect is more dramatic for NSC counts, consistent with the attrition observed in Fig. 1.

Results

Intersection between NCI-60 Measurements and the Analogue Set

Our analysis tests for a correspondence between the NCI-60 tumor panel sensitivity of our analogue compounds

⁴ Supplementary data for this article are available at Molecular Cancer Therapeutics Online (<http://mct.aacrjournals.org/>). Supplementary Figs. S1 and S2 provide histograms for the number of AMIDs in each cancer class and the number count of AMIDs in each cancer class, respectively.

⁵ http://www.phrma.org/medicines_in_developments

⁶ <http://integrity.prous.com>

and the clinical trial classification of their parent AMIDs. This analysis applies two filters to the raw data set of 4,977 bioactivity profiles. The first filter retains only NCI-60 bioactivity profiles above a threshold coefficient of variation of 8% (19). The second filter retains only bioactivity profiles with an average within-NCI-60 panel bioactivity greater than the group average GI₅₀ activity. Both of these filters eliminate screening records where little or no differential sensitivity is observed across the NCI-60. Table 2 summarizes these parent-analogue-Tanimoto counts. Data collected using these filters result in an average attrition to ~43% of the starting set of NSCs (2,097/4,877 = 0.43), while losing only ~10% of the AMIDs (144/159 = 0.90). Thus, filtering based on NCI-60 screening data eliminates a large fraction of NSC compounds but still retains sufficient numbers to find Tanimoto associations that retain 90% (144) of the starting 159 AMIDs.

NCI-60 Bioactivity Profiles

The next step examines the NCI-60 bioactivity profiles that pass our filters. The intention here is to determine whether NCI-60 panel sensitivity for this surviving set of compounds corresponds to the clinical trial designations for the parent AMIDs from which the analogue set is derived. The results for all nine tumor panels (Fig. 2; z-scores on the left and NCI-60 panel averages on the right) reveals an apparent correspondence between our analogue compound set and panel-specific NCI-60 bioactivity profiles, with the strongest correspondence found for leukemia, lung, colon, and central nervous system (CNS) tumors, and less so for the other cancer classes. These results suggest a possible relationship between the target type of cancer for many compounds in clinical trials and subpanel sensitivity in the NCI-60 screening results.

Randomization Check for Within-Panel Sensitivity

To assess the statistical significance of the apparent correspondence between AMID cancer class and NCI-60 panel sensitivity for analogue sets of NSCs, random bioactivity profiles were generated from the complete database of ~43,000 GI₅₀ measurements. Each random bioactivity profile was divided into two groups, one consisting of the GI₅₀'s within one of the nine NCI-60 panels and the other consisting of pooled measurements from the remaining eight NCI-60 panels. Each randomization step collects this differential statistic for a set of NSCs corresponding in size to those listed in Table 2 (e.g., a random set of 188 bioactivity profiles is collected for the CNS set and divided into CNS and non-CNS data sets). The within-to-between NCI-60 panel differential signal is determined for the actual test set of NSCs and for the randomized set, and a standard *t*-test is used to evaluate the statistical significance of this difference. These results find that only the NCI-60 lung, prostate, and renal tumors are not significantly different from random (*P* values of 0.064, 0.071, and 0.497, respectively). In the remaining six NCI-60 panels, their analogue NSC GI₅₀ measurements exhibited statistically significant differential sensitivity, with the leukemia and CNS tumor panels having the best significance scores (*P* values of 7.58e-6 and 2.65e-8, respectively) and panels colon, melanoma, ovarian, and breast having significance scores of 6.93e-3, 1.12e-5, 5.37e-3, and 4.87e-2, respectively. The marginal significance in the lung panel (*P* = 0.064) and the complete lack of significance in the renal panel (*P* = 0.497) is consistent with the plots (Fig. 2), where the lung test set is also sensitive in the CNS and prostate panels and the renal set shows high sensitivity to the melanoma and prostate panels. The

Table 1. NSC (AMID) counts for the NCI-60 tumor classes, raw counts

	Tanimoto score					
	1.0	0.9	0.8	0.7	0.6	0.5
Leukemia	15 (13)	45 (16)	63 (17)	119 (20)	338 (22)	1,096 (27)
Lung	10 (9)	106 (15)	211 (20)	266 (21)	397 (24)	1,011 (27)
Colon	4 (4)	28 (5)	80 (7)	122 (9)	281 (10)	831 (13)
Brain	7 (6)	22 (6)	67 (6)	80 (8)	151 (12)	435 (12)
Melanoma	6 (6)	34 (7)	81 (8)	150 (9)	285 (10)	824 (10)
Ovarian	8 (6)	22 (6)	79 (8)	118 (9)	272 (10)	734 (10)
Kidney	3 (3)	19 (4)	30 (4)	39 (4)	63 (8)	171 (9)
Prostate	9 (7)	26 (9)	40 (11)	74 (14)	202 (16)	803 (20)
Breast	6 (5)	60 (9)	102 (12)	155 (14)	320 (18)	922 (23)
Totals	68 (59)	362 (77)	753 (93)	1,123 (99)	2,309 (130)	6,827 (151)
Hit rate	0.01 (0.39)	0.05 (0.51)	0.11 (0.62)	0.16 (0.66)	0.34 (0.86)	1.0 (1.0)

NOTE: Counts for NSCs (AMIDs) at different Tanimoto cutoffs across the nine cancer classifications. For each AMID clinical trial type, the two-dimensional structural analogues in the NCI-60 screening set are accumulated at each of the six Tanimoto cutoffs. For example, 27 AMIDs are currently in clinical trials for leukemia. A total of 1,096 NSCs are found with two-dimensional structural greater than 0.5, whereas only 13 of these AMIDs and 15 NSC remain in the set for a Tanimoto score of 1.0. The excess of NSC over AMID counts at a Tanimoto score of 1.0 reflects instances of stereo isomers in the analogue set. The data appearing in the "Totals" row reflect a cumulative count of compounds (AMIDs and NSCs, respectively) across each cancer classification. Thus, compounds tested in more than one clinical trial are treated as separate occurrences in the totals of the last row. Note that 8 of the 159 parent AMIDs (Tanimoto >0.5) do not have clinical trials that appropriately match the NCI-60 tumor panels. The last row represents the hit rate obtained as the ratio of cumulative counts for each Tanimoto category divided by the counts for the Tanimoto greater than 0.5 category.

Table 2. NSC (AMID) counts for the NCI-60 tumor classes, filtered by panel sensitivity

	Tanimoto score						
	1.0	0.9	0.8	0.7	0.6	0.5	
Leukemia	9 (8)	18 (12)	24 (12)	42 (15)	112 (18)	410 (23)	
Lung	6 (6)	68 (11)	133 (14)	154 (19)	189 (22)	353 (26)	
Colon	3 (3)	14 (5)	27 (7)	46 (9)	103 (9)	233 (13)	
Brain	6 (5)	20 (5)	65 (5)	73 (8)	97 (10)	188 (12)	
Melanoma	2 (2)	22 (6)	44 (6)	61 (7)	83 (10)	209 (10)	
Ovarian	7 (6)	14 (6)	37 (8)	48 (9)	89 (9)	212 (10)	
Renal	1 (1)	9 (3)	14 (3)	17 (3)	23 (7)	56 (10)	
Prostate	2 (2)	14 (4)	19 (6)	21 (8)	44 (13)	158 (17)	
Breast	6 (5)	60 (9)	102 (12)	155 (14)	320 (18)	278 (23)	
Totals	42 (38)	239 (61)	465 (73)	617 (92)	1,060 (116)	2,097 (144)	
Hit rate	0.02 (0.26)	0.11 (0.42)	0.22 (0.51)	0.29 (0.64)	0.51 (0.81)	1.0 (1.0)	

NOTE: Counts for NSCs (AMIDs) at different Tanimoto cutoffs across the nine cancer classifications. Numbers are based on those NSCs passing threshold filters for differential GI_{50} profiles (coefficient of variation ≥ 0.08 and within-panel sensitivity above profile average). As in Table 1, each cancer trial (i.e., row) is processed separately, and the row "Totals" reflect cumulative AMID and NSC counts.

universal sensitivity of the prostate cells to all analogue sets is also evident in this image. These results support the existence of differential NCI-60 panel sensitivity within most parent-analogue sets. This result also suggests that NCI-60 panel sensitivity may provide an *a priori* basis for selecting candidate compounds to be tested against specific cancer classes.

A unifying explanation for the failure of the renal, breast, and prostate panels to exhibit sensitivity within their proposed clinical trial target most likely represents a mixture of effects. The few number of prostate cell lines ($n = 2$) limits the statistical power needed to distinguish selective sensitivity within this panel. The breast lines are also known to be quite genetically varied (30, 31), and more

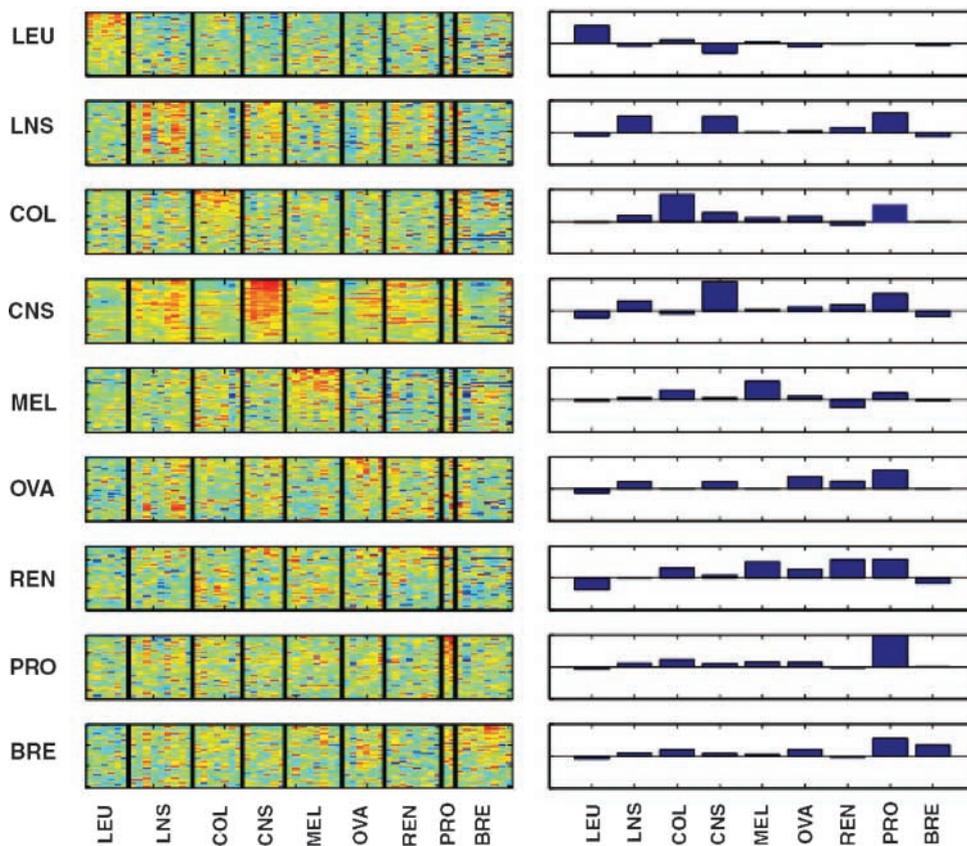


Figure 2. Left, z-score normalized GI_{50} values for analogue compounds with structural similarity (Tanimoto score > 0.5) to the AMID set for the nine NCI-60 tumor panels. Panel-selective sets are displayed separately, from top to bottom. The GI_{50} results for each panel represent the individual measurements for all NCI-60 tumor cell lines, grouped horizontally according to tumor cell panel. Red and blue, most sensitive and least sensitive GI_{50} scores, respectively. Right, column averaged GI_{50} values for each NCI-60 tumor panel. Sensitive and insensitive values appear above and below the horizontal reference of zero differential activity. Ordering of tumor panels is identical to that appearing on the left. LEU, leukemia; LNS, lung; COL, colon; CNS, central nervous system; MEL, melanoma; OVA, ovarian; REN, renal; PRO, prostate; BRE, breast.

recent evidence points to the possibility of misclassification (32). The failure of the renal panel to support within-panel selectivity does not readily suggest an explanation. Renal cancers, in general, are among the most refractory types of cancer, and more effective treatments are desperately being perused. In this instance, this might represent the apparently poor efficacy of renal-specific AMIDs.

Reverse Engineering Strategies for Data Mining

Our results support panel-selective sensitivity for some analogue NSCs associated with our parent AMIDs. This result raises the question of whether NCI-60 subpanel selectivity could be used to mine existing data for additional anticancer candidates. As a preliminary test, we filtered the bioactivity profiles for our analogue AMIDs, as described above, to yield a total of 317 NSCs (Tanimoto score >0.5), corresponding to 64 AMIDs. Progressively higher Tanimoto scores yielded 110 NSCs and 38 AMIDs (Tanimoto score >0.6), 62 NSCs and 22 AMIDs (Tanimoto score >0.7), 48 NSCs and 13 AMIDs (Tanimoto = 0.8), 20 NSCs and 8 AMIDs (Tanimoto score >0.9), and 3 NSCs and 3 AMIDs (Tanimoto = 1.0). Although these results are speculative, they support the use of a simple filter based on NCI-60 panel selectivity and a Tanimoto cutoff of 0.8 for identifying a reasonable number of hits within a set of ~300 screened compounds in our analogue AMIDs.

Figure 3 displays the results of harvesting panel-selective bioactivity profiles across the ~43,000 screened compounds. The number count in each NCI-60 panel is listed in the vertical label at the middle of this figure. A maximum of 4,468 NSCs are found for the leukemia panel and a minimum of 596 for the prostate panel. Approximately 40% (~18,000/43,000) of the NCI-60 bioactivity profiles have significant ($P < 0.05$) within-panel sensitivity, while retaining 114 of the original set of 159 AMIDs at a Tanimoto score >0.5. These compounds are inclusive of the counts listed in Table 2 and thus capture the ability to identify reasonable numbers of NSCs in our original parent-analogue pairs on the basis of subpanel selectivity alone, albeit at the expense of relatively higher NSC counts when compared with our starting analogue AMIDs.

The counts displayed (Fig. 3) can be used as a crude gauge of hit rates for compound selections based on NCI-60 subpanel selectivity versus randomly selected compounds in the screening set. The average value of 2,026 NSCs found across all NCI-60 subpanels (cf. Fig. 3) as well as their median count of 24 NSCs in the AMID analogue set (at a Tanimoto score >0.5) defines a baseline probability of 24/2,026 as a random hit rate. The random probability of finding 24 NSCs from the entire sample of 43,000 screened compounds is determined as $\sum_i p_i = 24/43,000$. This trivial example finds that on the basis of the reduction in sample size alone, an enrichment factor over random of 21.2 [(24/43,000)/(24/2,026)] is achieved by selecting panel-specific NSCs. Alternatively, simulation results can be used to determine the sample size needed to generate an equivalent hit rate across the complete sample space (~43,000) versus the panel-specific sample space (~2,000). These results find that a 26 times smaller sample size is

sufficient to obtain an equivalent hit rate when sampling from the panel specific versus complete data set. Whereas these results might seem to be trivial, their potential effect on improving hit counts can be dramatic. These numbers scale linearly with the fewer counts for higher Tanimoto similarity while maintaining an equivalent enrichment. Thus, subpanel selectivity seems to be useful as an initial filter for compound selection. Summary counts for the intersection between these NSCs and our AMID set (Table 3) find, remarkably, that 16 AMIDs in this set have a perfect Tanimoto score. These results provide additional support for the benefit of selecting screened compounds as candidate AMIDs based on NCI-60 panel selectivity. Also evident (Table 3) is the fact that, on average, less than 10% of panel-specific NSCs are accounted for in our AMID set. It is not possible to determine whether these additional compounds might themselves be equally effective anticancer agents.

Relating NCI-60 Selectivity to Two-Dimensional Structure

The results obtained above can be analyzed further to determine whether the NCI-60 panel sensitive compounds can be selected on the basis of two-dimensional structural descriptors. This step is crucial to addressing issues about strategies for mining novel compounds from an arbitrary library. This analysis begins by expanding the NCI-60 panel-specific "sensitive" NSCs to include the "insensitive" NSCs. The second step separately characterizes the sensitive and insensitive compounds according to their structural fingerprints. The objective here is to determine whether the molecular descriptors (e.g., fingerprints) are sufficiently different in the sensitive versus insensitive compounds to be considered as a complementary filter to compound selections based on panel-specific NCI-60 profiles. Conventional quantitative structure-activity relationship studies exploit these structure-function differences as a means to optimize the biological activity of lead compounds. A simpler objective is sought here: to determine whether structural determinants can be supported as a basis for biological activity and, if so, propose a selection strategy based on combined GI₅₀ activity and two-dimensional structure.

The analysis consists of four steps. First, a Fisher's exact test is used to determine which of the molecular descriptors [in this case defined as Daylight (33) fingerprints based on a fixed size of 2,048 bits assigned using a maximum number of 7 bonds separating atom types] exhibit a statistical difference ($P < 0.05$) between the sensitive and insensitive compounds. Statistically significant molecular descriptors are assigned values of 1 or -1 if the enrichment count is highest in the sensitive or insensitive compounds, respectively.

From a mask, m , defined as a vector with elements -1, 0, or 1, and the fingerprints, f , a modified Tanimoto score, T_m , is defined as;

$$T_m = \frac{m \times f}{\|f\|},$$

where $\|f\|$ is defined as the length of the fingerprint vector f .

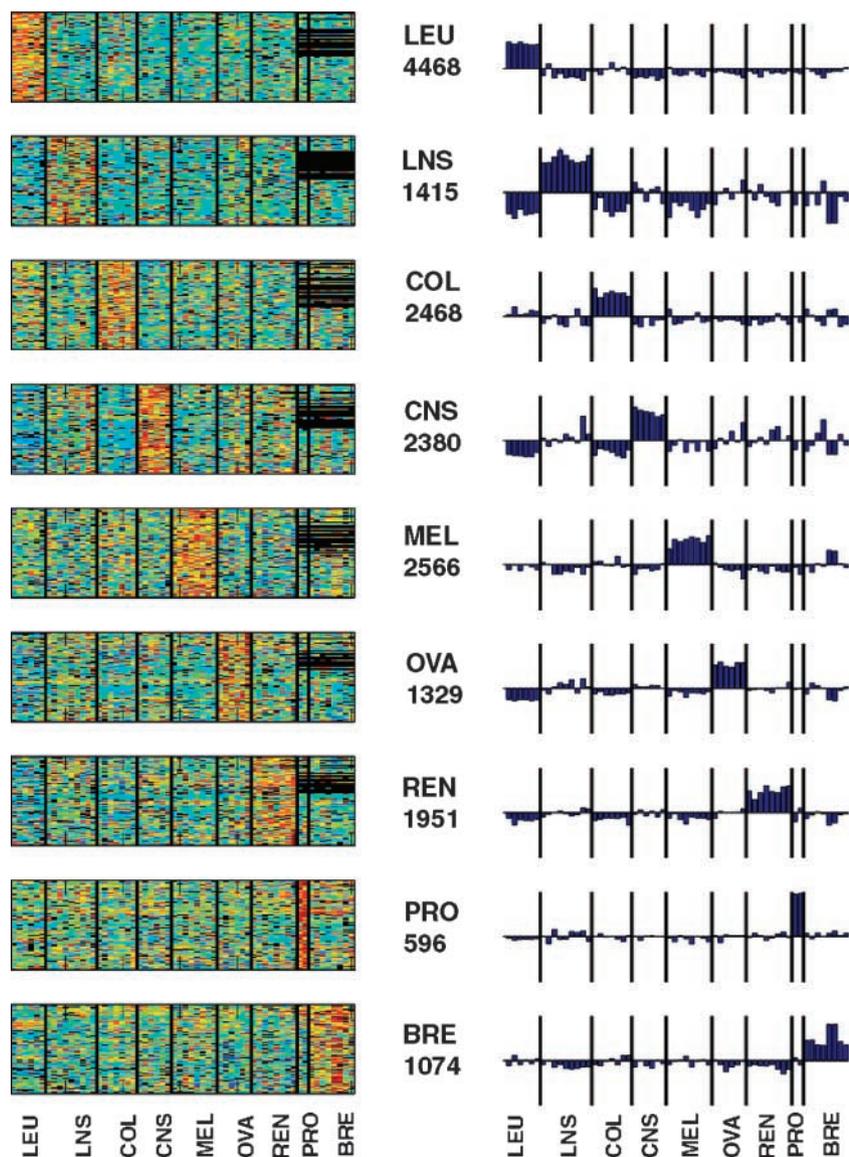


Figure 3. Left, z-score normalized GI_{50} values for intra-panel versus inter-panel sensitivity across the complete NCI-60 data set. Red and blue, most sensitive and least sensitive GI_{50} scores, respectively. Deep blue, missing data points. Right, column averaged GI_{50} values for each NCI-60 tumor panel. Sensitive and insensitive values appear above and below the horizontal reference of zero differential activity. Tumor panel order is identical to that presented in Fig. 2.

Molecular descriptors favoring the positive bits in the mask will receive a positive T_m , whereas molecular descriptors favoring the negative bits in the mask will receive a negative T_m . Third, at discrete intervals, T_{cut} , along the complete range of T_m 's, the sensitive and insensitive NSCs are assigned to one of four categories: true positive, false positive, false negative, and true negative. The measure of interest is the capacity of T_m to select "true positives" from the pooled sensitive and insensitive NSCs. Two metrics commonly used to test this capacity are "recall" and "precision". Recall is defined as the proportion of sensitive NSCs that are identified by their T_m as positive [true positive / (true positive + false negative)]. Similarly, precision is defined as the proportion of insensitive NSCs that are identified by their T_m 's as negative [true negative / (true negative + false positive)]. In general, a good testing procedure is characterized by high

recall and precision, whereas in reality, when the recall is very high, the precision tends to be low. A receiver operating characteristic (ROC; refs. 34, 35) curve, which is a plot of recall versus $1 - \text{precision}$, is commonly used as an efficient way to display the relationship between recall and precision. The preferred test yields the greatest number of true positives with the least number of false positives, resulting in a ROC curve that tends upward while moving from left to right. ROC curves for CNS, colon, breast, and melanoma are shown in Fig. 4. These results exhibit the desired features of a valid testing procedure, with the area under the ROC curve for all nine NCI-60 panels above the threshold of 0.7 for nonrandom associations, indicated by the diagonal line.

These ROC curves can be used to identify two-dimensional structural features that are specifically associated with panel-selective sensitive or insensitive activity. The

Table 3. NSC (AMID) counts for the NCI-60 tumor classes, derived by panel sensitivity

	Tanimoto score					
	1.0	0.9	0.8	0.7	0.6	0.5
Leukemia	3 (3)	13 (7)	25 (15)	50 (23)	103 (35)	289 (63)
Lung	2 (2)	4 (3)	9 (7)	16 (13)	42 (22)	75 (35)
Colon	3 (3)	12 (7)	32 (12)	45 (18)	88 (27)	204 (47)
CNS	5 (5)	26 (8)	57 (15)	69 (20)	104 (34)	191 (50)
Melanoma	1 (1)	5 (4)	7 (6)	19 (11)	55 (23)	181 (45)
Ovarian	1 (1)	6 (4)	15 (12)	26 (18)	54 (26)	113 (45)
Renal	2 (2)	5 (5)	18 (12)	28 (14)	55 (24)	103 (44)
Prostate	1 (1)	9 (7)	21 (11)	32 (17)	53 (29)	93 (42)
Breast	2 (2)	5 (5)	11 (8)	17 (11)	21 (13)	42 (29)
Totals	20 (20)	85 (50)	195 (98)	302 (145)	575 (233)	1,291 (400)
Hit rate	0.01 (0.05)	0.65 (0.125)	0.15 (0.25)	0.23 (0.36)	0.44 (0.58)	1.0 (1.0)

NOTE: Number count for NSCs and, in parentheses, the count for their structurally similar AMIDs. As in Tables 1 and 2, the row "Totals" reflects cumulative counts based only on clinical trials having appropriate matches to the NCI-60 tumor panels. See the legend to Table 1 for more details.

point on each ROC that deviates greatest from the diagonal defines the case of greatest precision and recall. This point often represents the largest fraction of true positives and true negatives and can be used to assess two-dimensional structural features that distinguish sensitive from insensitive compounds. Figure 5 displays the results for the CNS tumor panel, where the top image represents the all-to-all Tanimoto scores for the sensitive (true-positive) and insensitive (true-negative) compounds at the point of maximal deviation from the diagonal of the ROC. These

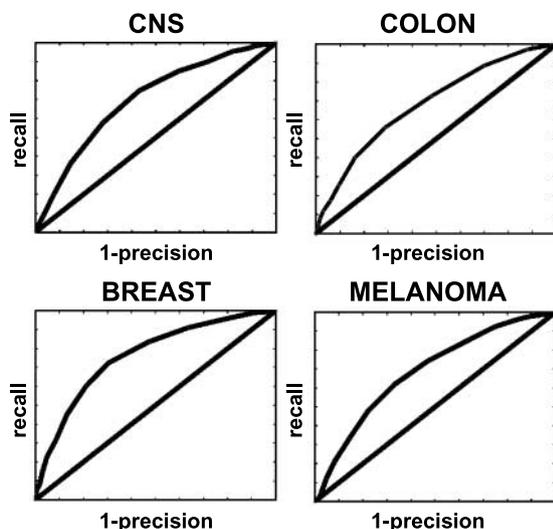


Figure 4. ROCs based on modified Tanimoto scores, T_m , for the intra-NCI-60 panel sensitive and insensitive NSCs. X axis, 1 – precision and y-axis represents recall. Each ROC is generated by counting the fractions of sensitive and insensitive NSCs assigned to the four categories, true positive, false positive, false negative, and true negative, at each Tanimoto cutoff. The area below the ROC and above the diagonal is an indication of the quality of the method used to generate the curve. The observed deviations from diagonal support the application of this procedure for separating sensitive from insensitive NSCs.

Tanimoto scores are obtained in the conventional fashion, and thus represent a completely different metric from the Tanimoto scores described in "Reverse Engineering Strategies for Data Mining." Nonetheless, it is clear that the true-positive and true-negative compounds exhibit stronger intra-Tanimoto versus inter-Tanimoto scores, as indicated visually by their block diagonal appearance, albeit the true-negative set is more structurally diverse than the true-positive set. Interestingly, the true-positive and true-negative compounds represent remarkably different structures, the true-positive set being composed largely of planar nitrogen-containing heterocyclic rings and the true-negative set consisting of nonaromatic oxygen and sulfur-containing structures. Representatives from the true-positive and true-negative sets are displayed in Fig. 5 (bottom). Clearly, the true-positive set is consistent with the known CNS sensitivity of ellipticinium compounds (36), whereas the true-negative set provides a possible rationale for *a priori* exclusion of possible lactone family members as potentially inactive against CNS tumors. Not apparent from this result are clues about the respective mechanisms of activity (or inactivity) for the true-positive and true-negative compounds.

These results support a connection between NCI-60 subpanel sensitivity/insensitivity and the two-dimensional structural features comprising these screened compounds. This procedure offers a potential complement to compound selections resulting from subpanel sensitivity. In support of this claim, reasonable numbers of AMID counts (Table 4) are found for compounds selected at a T_{cut} where the ROC deviates greatest from random (i.e., the diagonal line), which also represents the point on the ROC curve often chosen as a threshold for selecting the combined positive list (e.g., true positive and false positive) for evaluation. In contrast to separating the true-positive set from the true-negative set, as described above, the analysis here selects positive occurrences regardless of whether they are true or false. The numbers of screened compounds

for each NCI-60 tumor panel are shown in parentheses in the first column of Table 4. As a reminder, these numbers result from a compound selection scheme that is based only on NCI-60 panel selectivity and a two-dimensional structural filter based on panel sensitivity/insensitivity. A comparison of these results to Table 3 reveals slightly lower hit counts based on fewer compounds. These results support the use of subpanel selectivity (insensitive and sensitive) and two-dimensional structural descriptors as a selection strategy for mining potential analogue AMIDs. This approach, however, is likely to yield only lead compounds. In fact, inspection (Table 4) finds the nearly complete absence of perfect Tanimoto matches. Thus, in practice, it is rare to find NSCs in the NCI-60 open repository that are sensitive to NCI-60 tumor panels and that correspond to our designated AMID cancer trial type and have identical two-dimensional structural matches. It is less rare, however, to find structural analogues at lower Tanimoto scores (~ 0.8) with panel-

specific sensitivities in their bioactivity profiles that correspond to the clinical trial for AMIDs. Collectively, these results suggest that a selection strategy devoid of any prior knowledge of the parent AMID set can be obtained from NCI-60 bioactivity profiles combined with two-dimensional structural descriptors that yield families of compounds marginally similar in structure to the current set of parent AMIDs. Whether these analogue sets provide a track toward discovery of the additional AMID compounds cannot be easily determined. Virtual modeling combined with secondary testing might offer possibilities for refinements of the initial compound lists.

Discussion

The NCI-60 screening data represent a unique, publicly available, information-rich source that has the potential to be important in the process of discovering candidates for

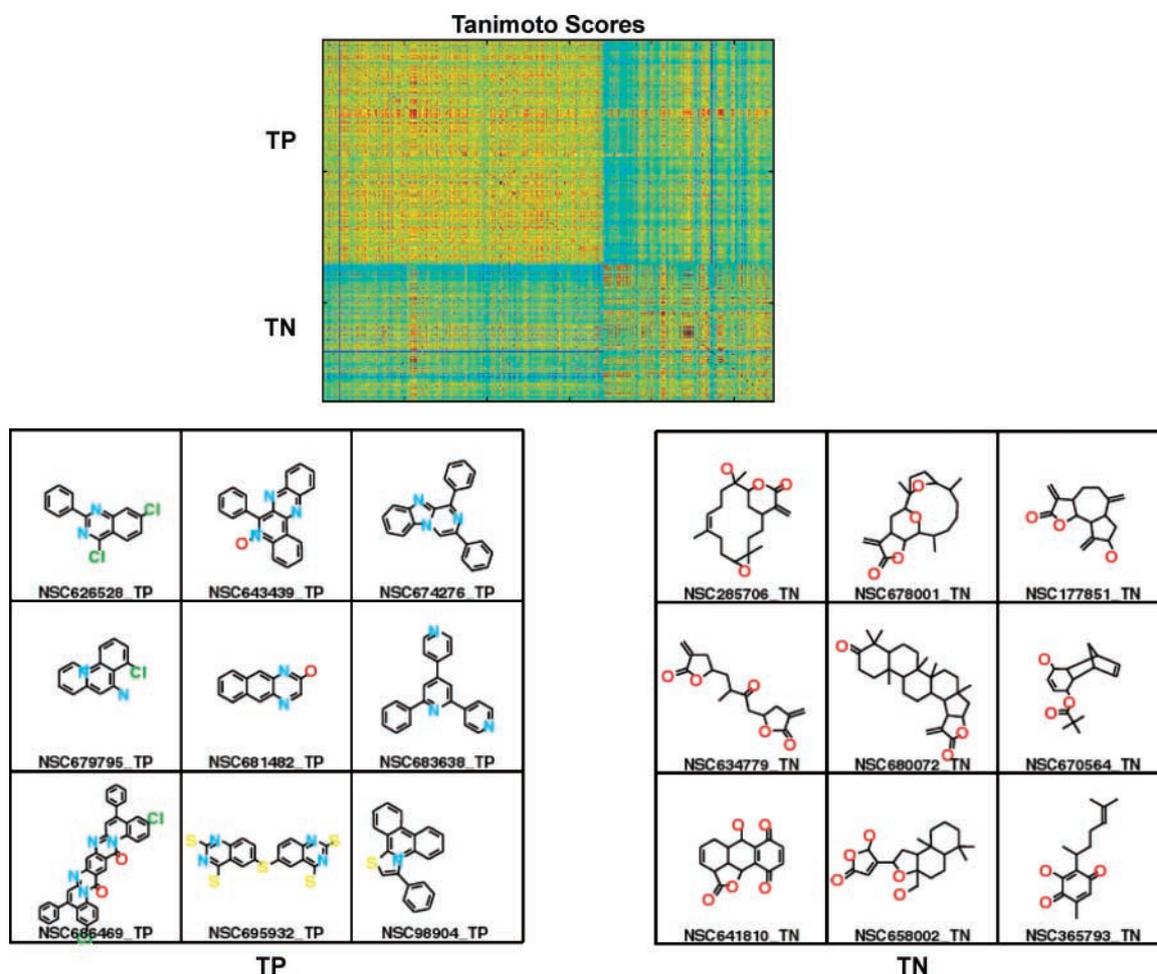


Figure 5. *Top*, all-to-all Tanimoto scores for the true positive and true negative compounds for the CNS tumor panel. Compounds represented here are derived from a ROC analysis based on structural features that distinguish sensitive and insensitive compounds. Tanimoto scores are represented spectrally (*red*, 1.0; *blue*, 0.0). *Bottom*, the two-dimensional structures represent examples taken from the true positive (*TP*; *left*) and true negative (*TN*; *right*) groups. The compound identifier is displayed at the bottom of each structure.

Table 4. NSC (AMID) counts for the NCI-60 tumor classes, derived by panel sensitivity and two-dimensional structural features

	Tanimoto score					
	1.0	0.9	0.8	0.7	0.6	0.5
Leukemia	1 (1)	5 (3)	8 (3)	11 (4)	33 (6)	153 (8)
Lung	0 (0)	4 (1)	5 (2)	6 (3)	7 (3)	16 (7)
Colon	0 (0)	2 (1)	3 (2)	5 (3)	10 (5)	22 (6)
CNS	2 (2)	9 (2)	30 (2)	32 (2)	39 (6)	53 (8)
Melanoma	0 (0)	5 (2)	9 (3)	12 (3)	23 (5)	64 (6)
Ovarian	0 (0)	1 (1)	2 (1)	4 (3)	5 (3)	15 (5)
Renal	0 (0)	0 (0)	0 (0)	0 (0)	4 (3)	4 (3)
Prostate	0 (0)	2 (1)	3 (1)	3 (1)	6 (4)	12 (6)
Breast	0 (0)	2 (2)	4 (2)	4 (2)	3 (5)	12 (9)
Totals	3 (3)	30 (13)	64 (16)	77 (21)	130 (44)	351 (58)
Hit rate	0.01 (0.05)	0.09 (0.22)	0.18 (0.28)	0.22 (0.36)	0.37 (0.76)	1.0 (1.0)

NOTE: NSC (AMID) counts derived from the complete set of NCI-60 screening data using NCI-60 panel selectivity and two-dimensional structural filters as described in the text. See the legend to Table 1 for additional details.

anticancer development (14, 19, 24, 31, 37). As new technologies and their associated data become available, the NCI-60 panel exists as a transitional monitor of past and current discovery efforts. In particular, recent attempts to interface the existing NCI-60 screening profiles with microarray and pathway data (20, 24, 25) seem to be useful for providing clues about the mechanism of action of a compound, and when fortunate, a development candidate for cancer therapy (22, 38, 39).

The analysis presented here examines the correspondence between the type of cancer proposed for clinical testing and the panel-selective bioactivity within the NCI-60 screen. The results find support for the NCI anticancer screen to enrich, over random, selections of analogue compounds that share bioactivity profiles within these clinical candidates. Reverse engineering data mining, based on collecting compounds with subpanel selectivity, finds a good intersection between screened compounds and candidate medicines currently in clinical trials. Mining strategies based on structural considerations derived from compounds with panel-selective sensitive versus insensitive bioactivity profiles provide an additional means for reducing the number of compounds to be tested. Selections based on subpanel selectivity can be contrasted with using potency value [mean $-\log(GI_{50})$] scores across the NCI-60 panel. Based on the parent-analogue sets represented in Fig. 1, an average of 57% of these compounds at each Tanimoto cutoff would have been excluded by considering potencies above -6.0 log units (1 $\mu\text{mol/L}$). The retention jumps to 85% for potencies above -4.0 log units (1.0e -4 mol/L). Thus, it seems that mining strategies based on NCI-60 subpanel selectivity offer a better opportunity to identify this set of AMIDs when compared with potency.

Collectively, our results provide a knowledge-based mining strategy using NCI-60 tumor panel selectivity as a means to harvest novel candidate compounds. The proposed methods can provide an opportunity to reduce large

compound lists into testable subpopulations for the purpose of extracting candidate anticancer medicines from the current NCI database.

Acknowledgments

We thank Drs. John Beutler, Robert Shoemaker, Susan Mertins, and John Cardellina for valuable contributions during the preparation of the manuscript.

References

- Lam LT, Davis RE, Pierce J, et al. Small molecule inhibitors of I κ B kinase are selectively toxic for subgroups of diffuse large B-cell lymphoma defined by gene expression profiling. *Clin Cancer Res* 2005;11:28–40.
- Jacoby E, Schuffenhauer A, Popov M, et al. Key aspects of the Novartis compound collection enhancement project for the compilation of a comprehensive chemogenomics drug discovery screening collection. *Curr Top Med Chem* 2005;5:397–411.
- Orry AJ, Abagyan RA, Cavasotto CN. Structure-based development of target-specific compound libraries. *Drug Discov Today* 2006;11:261–6.
- Fischer HP. Towards quantitative biology: integration of biological information to elucidate disease pathways and to guide drug discovery. *Biotechnol Annu Rev* 2005;11:1–68.
- Caldwell GW, Yan Z. Screening for reactive intermediates and toxicity assessment in drug discovery. *Curr Opin Drug Discov Devel* 2006;9:47–60.
- Brown N, Zehender H, Azzaoui K, Schuffenhauer A, Mayr LM, Jacoby E. A cheminformatics analysis of hit lists obtained from high-throughput affinity-selection screening. *J Biomol Screen* 2006;11:123–30.
- Prasanna MD, Vondrasek J, Wlodawer A, Rodriguez H, Bhat TN. Chemical compound navigator: a web-based chem-BLAST, chemical taxonomy-based search engine for browsing compounds. *Proteins* 2006;63:907–17.
- Haber C, Boillat M, van der Schoot B. Precise nanoliter fluid handling system with integrated high-speed flow sensor. *Assay Drug Dev Technol* 2005;3:203–12.
- Giuliano KA, Cheung WS, Curran DP, et al. Systems cell biology knowledge created from high content screening. *Assay Drug Dev Technol* 2005;3:501–14.
- Perlman ZE, Mitchison TJ, Mayer TU. High-content screening and profiling of drug activity in an automated centrosome-duplication assay. *Chembiochem* 2005;6:145–51.
- Blower PE, Cross KP. Decision tree methods in pharmaceutical research. *Curr Top Med Chem* 2006;6:31–9.
- Hrusovsky K. Getting on the critical path: better evaluation tools for drug discovery and development. *Drug Discov Today* 2006;11:773–4.

13. Boyd MR, Paull KD. Some practical considerations and applications of the National Cancer Institute *in vitro* anticancer drug discovery screen. *Drug Dev Res* 1995;34:91–109.
14. Shoemaker RH, Monks A, Alley MC, et al. Development of human tumor cell line panels for use in disease-oriented drug screening. *Prog Clin Biol Res* 1988;276:265–86.
15. Shoemaker RH, Scudiero DA, Melillo G, et al. Application of high-throughput, molecular-targeted screening to anticancer drug discovery. *Curr Top Med Chem* 2002;2:229–46.
16. Paull KD, Shoemaker RH, Hodes L, et al. 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;81:1088–92.
17. Voeller DM, Grem JL, Pommier Y, Paull K, Allegra CJ. Identification and proposed mechanism of action of thymidine kinase inhibition associated with cellular exposure to camptothecin analogs. *Cancer Chemother Pharmacol* 2000;45:409–16.
18. Kohlhagen G, Paull KD, Cushman M, Nagafuji P, Pommier Y. Protein-linked DNA strand breaks induced by NSC 314622, a novel non-camptothecin topoisomerase I poison. *Mol Pharmacol* 1998;54:50–8.
19. Rabow AA, Shoemaker RH, Sausville EA, Covell DG. Mining the National Cancer Institute's tumor-screening database: identification of compounds with similar cellular activities. *J Med Chem* 2002;45:818–40.
20. Huang R, Wallqvist A, Covell DG. Anticancer metal compounds in NCI's tumor-screening database: putative mode of action. *Biochem Pharmacol* 2005;69:1009–39.
21. Huang Y, Blower PE, Yang C, et al. Correlating gene expression with chemical scaffolds of cytotoxic agents: ellipticines as substrates and inhibitors of MDR1. *Pharmacogenomics J* 2005;5:112–25.
22. Shoemaker RH. The NCI60 human tumour cell line anticancer drug screen. *Nat Rev Cancer* 2006;6:813–23.
23. Bussey KJ, Chin K, Lababidi S, et al. Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. *Mol Cancer Ther* 2006;5:853–67.
24. Covell DG, Wallqvist A, Huang R, Thanki N, Rabow AA, Lu XJ. Linking tumor cell cytotoxicity to mechanism of drug action: an integrated analysis of gene expression, small-molecule screening and structural databases. *Proteins* 2005;59:403–33.
25. Wallqvist A, Huang R, Thanki N, Covell DG. Evaluating chemical structure similarity as an indicator of cellular growth inhibition. *J Chem Inf Model* 2006;46:430–7.
26. Huang R, Wallqvist A, Thanki N, Covell DG. 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;5:381–99.
27. Wallqvist A, Huang R, Covell DG, Roschke AV, Gelhaus KS, Kirsch IR. Drugs aimed at targeting characteristic karyotypic phenotypes of cancer cells. *Mol Cancer Ther* 2005;4:1559–68.
28. Willett P. Similarity-based approaches to virtual screening. *Biochem Soc Trans* 2003;31:603–6.
29. Martin YC, Kofron JL, Traphagen LM. Do structurally similar molecules have similar biological activity? *J Med Chem* 2002;45:4350–8.
30. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000;24:227–35.
31. Scherf U, Ross DT, Waltham M, et al. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000;24:236–44.
32. Ikediobi ON, Davies H, Bignell G, et al. Mutation analysis of 24 known cancer genes in the NCI-60 cell line set. *Mol Cancer Ther* 2006;5:2606–12.
33. Daylight. Daylight Chemical Information Systems, Inc. Aliso Viejo, CA.
34. Lasko TA, Bhagwat JG, Zou KH, Ohno-Machado L. The use of receiver operating characteristic curves in biomedical informatics. *J Biomed Inform* 2005;38:404–15.
35. Stephan C, Wesseling S, Schink T, Jung K. Comparison of eight computer programs for receiver-operating characteristic analysis. *Clin Chem* 2003;49:433–9.
36. Shi LM, Fan Y, Myers TG, et al. Mining the NCI anticancer drug discovery databases: genetic function approximation for the QSAR study of anticancer ellipticine analogues. *J Chem Inf Comput Sci* 1998;38:189–99.
37. Blower PE, Yang C, Fligner MA, et al. Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data. *Pharmacogenomics J* 2002;2:259–71.
38. Solit DB, Garraway LA, Pratilas CA, et al. BRAF mutation predicts sensitivity to MEK inhibition. *Nature* 2006;439:358–62.
39. Garraway LA, Widlund HR, Rubin MA, et al. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* 2005;436:117–22.

Molecular Cancer Therapeutics

Anticancer medicines in development: assessment of bioactivity profiles within the National Cancer Institute anticancer screening data

David G. Covell, Ruili Huang and Anders Wallqvist

Mol Cancer Ther 2007;6:2261-2270.

Updated version Access the most recent version of this article at:
<http://mct.aacrjournals.org/content/6/8/2261>

Cited articles This article cites 38 articles, 6 of which you can access for free at:
<http://mct.aacrjournals.org/content/6/8/2261.full#ref-list-1>

Citing articles This article has been cited by 2 HighWire-hosted articles. Access the articles at:
<http://mct.aacrjournals.org/content/6/8/2261.full#related-urls>

E-mail alerts [Sign up to receive free email-alerts](#) related to this article or journal.

Reprints and Subscriptions To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions To request permission to re-use all or part of this article, use this link
<http://mct.aacrjournals.org/content/6/8/2261>.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.