Drugs aimed at targeting characteristic karyotypic phenotypes of cancer cells

Anders Wallqvist,1 Ruili Huang,2 David G. Covell,2 Anna V. Roschke,3 Kristen S. Gelhaus,3 and Ilan R. Kirsch3

1Science Applications International Corp., National Cancer Institute; 2Developmental Therapeutics Program, Screening Technologies Branch, National Cancer Institute at Frederick, Frederick, Maryland; and 3Genetics Branch, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland

Abstract
The karyotypic features of cancer cells have not been a particular focus of anticancer drug targeting either as guidance for treatment or as specific drug targets themselves. Cancer cell lines typically have considerable, characteristic, and variable chromosomal aberrations. Here, we consider small-molecule screening data across the National Cancer Institute’s 60 tumor cell line drug screening panel (NCI-60) analyzed for specific association with karyotypic variables (numerical and structural complexity and heterogeneity) determined for these same cell lines. This analysis is carried out with the aid of a self-organizing map allowing for a simultaneous assessment of all screened compounds, revealing an association between karyotypic variables and a unique part of the cytotoxic response space. Thirteen groups of compounds based on related specific chemical structural motifs are identified as possible leads for anticancer drug discovery. These compounds form distinct groups of molecules associated with relatively unexplored regions of the NCI-60 self-organizing map where anticancer agents currently standard in the clinic are not present. We suggest that compounds identified in this study may represent new classes of potential anti-cancer agents. [Mol Cancer Ther 2005;4(10):1559–68]

Introduction
Most cancers have an abnormal chromosomal content, called aneuploidy, characterized by changes in chromosomal structure and number. Chromosomal aberrations are more often found in malignant tumors than in benign ones and are associated with poorer prognoses, aggressive clinical characteristics, and distinctive histopathology (1–4). It is possible that these quantitative or qualitative changes in the karyotypic state of malignancies may represent potential determinants for anticancer therapies and, ultimately, might allow targeting of the most aggressive and incurable cancers.

The individual immortalized cancer cell lines in the National Cancer Institute (NCI) drug discovery panel (NCI-60) can be characterized and distinguished by a variety of abnormal karyotypic features in both the number and the structure of their component chromosomes. We have quantified the chromosomal aberrations of this cell line panel using spectral karyotyping to delineate ploidy as well as structural and numerical karyotypic complexity and heterogeneity (5). Numerical complexity describes the change in chromosome number compared with the ploidy level of the cell line; structural complexity reflects translocations, deletions, duplications, amplifications, insertions, and inversions of the chromosomes. Ongoing instability is revealed by the heterogeneity variable, which measures metaphase-to-metaphase variation in numerical or structural complexity within a given cell line. These quantifiable spectral karyotyping variables can be used as a descriptor of the chromosomal “state” for each cell line.

The connection between chromosomal instabilities and cancer has focused attention on defects in chromosomal segregation, telomere stability, cell cycle checkpoint regulation, and repair of DNA damage (6–10). The specific components of such processes are not yet fully delineated, but if a protein could be identified as a gatekeeper to maintaining chromosomal stability, manipulation of the function of this protein with small-molecule drugs, either directly or indirectly, might provide a powerful weapon for cancer therapy.

There is no requirement, however, that a single protein be the sole determinant of a cellular karyotype. Although contrary to the current development of precise molecularly targeted drugs, it is also possible that a particular karyotypic phenotype itself could be drugable. In an initial effort to identify lead compounds whose activity could be related to particular karyotypic observables, we studied a subset of drugs (n = 1,429, including many standards of the chemotherapeutic armamentarium) from the Developmental Therapeutics Program data repository that had been repeatedly screened against the NCI-60 (11). This prior study explored correlations between karyotypic variables and growth inhibition for this relatively restricted set of agents. Although infrequent, positive correlations were found, but, in general, for current commonly used anticancer drugs this analysis did not find evidence for a direct positive (i.e., relatively increased sensitivity with increased karyotypic

Received 7/5/05; accepted 8/10/05.
Grant support: NCI, NIH contract N01-C0-12400.

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.

Note: The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. government.

Requests for reprints: Ilan R. Kirsch, Amgen, 1201 Amgen Court West, AW1-J 4144, Seattle, WA 98119-3105. Phone: 206-265-7318; Fax: 206-216-5930. E-mail: ikirsch@amgen.com

Copyright © 2005 American Association for Cancer Research.
doi:10.1158/1535-7163.MCT-05-0224

Mol Cancer Ther 2005;4(10). October 2005
complexity) association between cytotoxic profiles and the experimentally determined variables of karyotypic state. These results suggested that, among other possibilities, the mechanisms of action for many well-known anticancer agents were most likely not associated with chromosomal abnormalities, consistent with their somewhat limited utility, for the most part, in epithelial cancers that reside at the more karyotypically complex end of the cancer spectrum.

In this current study, we have explored the utility of the full set of publicly available screening data, consisting of cell-based growth inhibition data for ~30,000 potential anticancer compounds, to delineate aspects of karyotypic variability that are distinct and uniquely identifiable within this data set. Association between the screening data organized via a self-organizing map (SOM) and karyotypes can be used to distinguish unique mechanisms of action associated with chromosomal aberrations. We delineate groups of compounds heretofore less well-characterized relative to agents already commonly used in oncology clinical practice. These agents or derivatives of them would be candidates for drugs developed for the treatment of karyotypically complex cancers. In addition, exploration of the structures and mechanisms of action of these compounds may provide insight into the nature of karyotypic instability itself.

Materials and Methods

**Developmental Therapeutics Program Resources of GI50 Data**

The NCI-60 cell line 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 hematologic malignancies (12). The data consist of concentration values (GI50) 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-hour drug incubation. The GI50 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(GI50) values typically ranges from 4.0 to 8.0. The pattern of GI50 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 (13–15).

**Karyotypic Variables**

Details of the karyotypic analysis of the NCI-60 have been described previously (5).

**Self-Organizing Map**

To simultaneously describe similarities between all GI50 data vectors, we have used a SOM (16) to organize cellular growth inhibition data derived from the NCI-60 tumor cell panels (15). The SOM algorithm identifies cluster vectors in the 60-dimensional data space by minimizing the deviation between the GI50 data vectors and the cluster vectors. Regions in GI50 space that are dense with data vectors attract many cluster vectors, and regions with few data vectors attract fewer cluster vectors, resulting in a division of response space that mimics information content. An advantage of SOM reordered data is the ability to visualize the global clustering results in an interpretable manner. Our preferred method of display is the uniform projection of SOM clustering in high-dimensional space to a two-dimensional map. This mapping both is simple and retains a great deal of the original high-dimensional information. Additional details regarding the creation and access to the GI50 SOM are given (e.g., ref. 15).

Each GI50 data vector is thus uniquely assigned to a cluster vector on the SOM. Different compounds result in different profiles that are associated with different locations on the SOM. Occasionally, two compounds will generate similar profiles. The more similar, the closer they are then grouped, which can be visualized on the two-dimensional SOM where separate cluster vectors designate groupings of compounds demonstrating related profiles and cluster neighborhoods represent areas of associations between possible mechanisms of action.

**Statistical Analysis and Correlations**

The growth inhibition data vectors used for the SOM construction encode drug concentrations and were Z-score normalized before clustering (i.e., the mean GI50 was subtracted off each measurement and divided by the SD). Normalization to unitless data facilitates comparisons between other independently derived measures from these same tumor cells. The same normalization procedure was applied to construct karyotypic data vectors. Because our SOM is a representation of the normalized GI50 vector space, other data vectors may or may not be appropriately described by this same space. Of course, any data vector will have a minimum distance to one particular cluster vector (e.g., trivially, the null vector will find the cluster vector with the least variance). Quality-control checks of these results using standard measures of similarity revealed no anomalies (i.e., their profiles are similar). In addition, the capacity to find significant matches between cytotoxic profiles and karyotypic profiles is an indication that the karyotypic data can be described by the cytotoxicity-derived space encompassed by our SOM.

The next stage of our analysis addressed appropriate and robust means for assessing significance of similarity and possible means of data reduction to clarify the observables associated with the karyotypic state. The Pearson or sample correlation coefficient (PCC) between two vectors \( \bar{u} \) and \( \bar{v} \) is defined as

\[
\text{PCC}(\bar{u}, \bar{v}) = \frac{\sum_i (u_i - \bar{u}) \cdot (v_i - \bar{v})}{\sqrt{\sum_i (u_i - \bar{u}) \cdot \sum_i (v_i - \bar{v})}}
\]

where \( \bar{u} \) denotes the average of all elements in \( \bar{u} \). 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 exactly the same.

Mol Cancer Ther 2005;4(10). October 2005

Downloaded from mct.aacrjournals.org on April 6, 2017. © 2005 American Association for Cancer Research.
A measure of how similar two data vectors are to each other can be gauged using linear regression. Associated with the correlation (PCC) is a $P$ derived from PCC and $N$ (number of data points) that gives the probability that a correlative relationship exists (i.e., if $P = 0.05$, there is only a 1 in 20 chance that the observed correlation is due to random chance). Moderate values of PCC are sometimes thought of as indicating strong relationships, but this may produce misleading results. Instead, the real strength of the relationship is best indicated by PCC$^2$. Technically, this is the proportion of variance in one vector “explained” by linear regression on the other vector. Thus, even if there exists a nonrandom correlation (e.g., moderate PCC), the strength of the correlation need not be great (e.g., low PCC$^2$). The observation that a correlative relationship exists can then be used to construct testable hypotheses to verify the existence of a statistically supportable connection between the observables. Thus, although the magnitude of the $r$ may be small, a correlative analysis can establish important connections between variables.

**Singular Value Decomposition**

Methods of singular value decomposition are used to investigate the properties of the karyotypic variable vectors themselves. Briefly, we can form a matrix $K$ of $N$ karyotypic variables as columns and $M$ cell lines as rows. This matrix can always be decomposed into three matrices, $U$, $S$, and $V^T$, to form the matrix equation:

\[ K = [U]_N \times [S]_N \times [V^T]_M, \]

where $U$ is a matrix of the same size as $K$, whose columns are orthonormal to each other, $S$ is a $N \times N$ diagonal matrix of singular values, and $V^T$ is another $N \times N$ matrix whose rows are orthonormal to each other (17). The orthonormal columns of $U$ form a basis for the karyotypic profiles, and the orthonormal rows of $V^T$ form the basis for a GI50 response. Here, a basis refers to a simplified representation of the data whose component values can be systematically analyzed. From the orthogonal conditions of the matrices, we can write the $U$ matrix and its elements as linear combination of the other matrices, $U = KVS^{-1}$. The $U$ matrix is thought to better represent the underlying biological process responsible for the observables in $K$. The diagonal elements of $S$ give a measure of importance to each of the basis vectors in $U$, $U_i$, which represent the decomposition of the data. If all the observables in $K$ are independent of each other, the elements of $S$ will be comparable, and the data cannot be reduced to a fewer number of data vectors. The square of the diagonal elements in $S$ is proportional to the variance and thus to the amount of variance explained by each base vector in $U$. By examining the relative variance of each elements, $w$, defined as 

\[ w_i = S^2_i / \sum_j S^2_j, \]

the data can be filtered to determine the dimensionality of the problem. We use the accepted heuristic that if a value of $w_i$ exceeds the threshold of $0.7/N$, the data vector is retained; below this, it can be discarded.

**Results**

**Karyotypic Properties of Cancer Cell Lines**

We have determined previously the spectral karyotype for each of the cell lines included in the NCI-60. We had delineated the following variables for each cell line: number of reconfigured chromosomes (structural complexity), number of chromosomal gains or losses compared with ploidy level (numerical complexity), and evidence of ongoing chromosomal reconfiguration (structural heterogeneity) or numerical gain or loss (numerical heterogeneity; ref. 5).

**Karyotypic Observable Projection on GI50 SOM**

A projection of data derived from all the NCI-60 cell lines on the SOM generated from GI50 cytotoxicity data proceeds by finding the smallest distance between this data vector and all the cluster vectors describing the map. The growth inhibition pattern of sensitivity and insensitivity in GI50 reflect the characteristic cellular differences in growth inhibition after drug exposure in the assay. Correspondingly, the normalized karyotypic observables reflect the characteristic karyotypic differences between cells. Linking karyotypic data to GI50 data via the SOM attempts to delineate hypotheses about relationships between karyotype and drug sensitivity. This linkage presumes that a statistically significant similarity between the karyotypic differential pattern and the GI50 pattern of a drug provides a basis for hypothesizing that cells displaying relatively higher karyotypic measures are more sensitive to that drug than other cells with relatively lower karyotypic measures.

Some typical correlation values and associated $P$s between specific karyotypic observables and GI50 data vectors are given in Table 1. For these compounds, the mean GI50 values across the cell lines given in Table 1 varied between the least sensitive $10^{-4.3}$ mol/L to the most sensitive $10^{-7.3}$ mol/L with an average growth inhibition concentration of $10^{-4.9}$ mol/L, with individual cell GI50 values ranging from the highest test concentration of $10^{-4}$ mol/L down to $10^{-8}$ mol/L. As a reference for the GI50 values, it is worth noting that this same general range of mean sensitivities (based on GI50) is observed for compounds currently used as standard of care in clinical oncology practice (e.g., leucovorin $10^{-4.3}$ mol/L, 5-fluorouracil $10^{-4.6}$ mol/L, cisplatin $10^{-5.5}$ mol/L, gemcitabine $10^{-6.7}$ mol/L, and docetaxel $10^{-7.6}$ mol/L).

Although the projection of karyotypic data vectors on the SOM yields a “best” match to a cytotoxic profile, it is instructive to identify regions of response that correlate well with a karyotypic data vector. A coherent region of responses on the SOM suggests a similar mechanism of action of the underlying drugs in that region. Investigating the association of karyotypic measures with all GI50 data vectors at a given correlation threshold (unadjusted $P$ of 0.05) implicates a predominant association with the P region.
and, in particular, with the P3 subregion. In Fig. 1, this is illustrated for the GI_{50} data vectors that are positively correlated with each of the karyotypic variables (numerical heterogeneity, numerical complexity, structural heterogeneity, and structural complexity), where each variable identified 1,198, 1,006, 1,271, and 382 compounds, respectively, at the \( P \leq 0.05 \) cut. In practical terms, this result indicates that the karyotypic profiles best match the cytotoxic profiles within the subregion P3. These compounds represent an unexplored set of chemical motifs whose activities correlate with the variability of the cellular karyotypes.

The regions on the SOM mapped out by the karyotypic data vectors, shown in Fig. 1, share many similarities but also have unique distinctions that invite the possibility for separate mechanisms responsible for specific activities of certain compounds.

Independent assessments between each of our karyotypic measures find strong correlations between structural heterogeneity and numerical complexity as shown in columns 2 to 4 in Table 2, with a correlation coefficient of 0.66. Both of these karyotypic measures are then correlated to the numerical heterogeneity data vector with correlation coefficients of 0.50 and 0.56, respectively. The structural complexity variable contains the least information, as the other karyotypic measures, with correlation values ranging from 0.03 to 0.40.

It is notable to compare the projection onto the SOM of the karyotypic phenotypes related to the other karyotypic measures, with correlation coefficients of 0.50 and 0.56, respectively. The structural complexity variable contains the least information, as with other correlative methods is a blunt tool for drug discovery, requiring secondary experimental confirmation as with other correlative methods.

<table>
<thead>
<tr>
<th>Karyotypic variables</th>
<th>Formula</th>
<th>Drug name</th>
<th>PCC</th>
<th>( P )</th>
<th>GI_{50}</th>
<th>NSC no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structural heterogeneity</td>
<td>C_{2}H_{14}BrNSe</td>
<td>2-Pyridineselenenyl bromide</td>
<td>0.64</td>
<td>0.1E-06</td>
<td>10^{-4.8}</td>
<td>610578</td>
</tr>
<tr>
<td>Numerical heterogeneity</td>
<td>C_{18}H_{17}N_{3}O_{3}</td>
<td>Ethyl 3-(4-methoxyanilino)-2-quinoxalinecarboxylate</td>
<td>0.63</td>
<td>0.7E-06</td>
<td>10^{-4.3}</td>
<td>680551</td>
</tr>
<tr>
<td>Numerical complexity</td>
<td>C_{32}H_{18}F_{4}N_{4}O_{2}</td>
<td>N-(3,4-dimethoxyphenyl)-3-phenyl-7-(trifluoromethyl)-2-quinoxalineamine</td>
<td>0.52</td>
<td>0.2E-03</td>
<td>10^{-4.7}</td>
<td>631581</td>
</tr>
<tr>
<td>Numerical heterogeneity</td>
<td>C_{32}H_{11}FN_{4}O_{2}S</td>
<td>3-(4-Fluorophenyl)-2-(4-oxo-2H-chromen-3-yl)-1,3-thiazol-2-yl)acrylonitrile</td>
<td>0.50</td>
<td>0.2E-03</td>
<td>10^{-4.5}</td>
<td>684985</td>
</tr>
<tr>
<td>Numerical heterogeneity</td>
<td>C_{36}H_{32}N_{2}O_{2}</td>
<td>N-(8-(1-Naphthylamino)octyl)-1-naphthamide</td>
<td>0.47</td>
<td>0.6E-02</td>
<td>10^{-4.5}</td>
<td>629738</td>
</tr>
<tr>
<td>Numerical heterogeneity</td>
<td>C_{32}H_{44}O_{8}</td>
<td>5-(2,16-Dihydroxy-4,4,9,14-tetramethyl-3,11-dioxaestra-1,5-dien-17-yl)-5-hydroxy-1,1-dimethyl-4-oxo-2-hexenyl acetate</td>
<td>0.45</td>
<td>0.7E-02</td>
<td>10^{-7.3}</td>
<td>106399</td>
</tr>
<tr>
<td>Structural heterogeneity</td>
<td>C_{32}H_{22}N_{4}O_{3}</td>
<td>11-(4-(4-Methoxyphenyl)-3-(4-(4-Methoxyphenyl)diazeyl)-3,5-diphenyl-1H-pyrazol-1-yl)-3-oxopropanamide</td>
<td>0.40</td>
<td>0.3E-02</td>
<td>10^{-4.5}</td>
<td>637921</td>
</tr>
<tr>
<td>Structural complexity</td>
<td>C_{38}H_{30}Cl_{3}N_{6}O_{2}</td>
<td>N-(3-8-(4-anilino-5-(2,4-dichlorophenoxy)methyl)-4H-1,2,4-triazol-3-yl)octyl-5-(2,4-dichlorophenoxy)methyl)-4H-1,2,4-triazol-4-yl-N-phenylamine</td>
<td>0.39</td>
<td>0.5E-02</td>
<td>10^{-5.0}</td>
<td>697169</td>
</tr>
<tr>
<td>Numerical complexity</td>
<td>C_{37}H_{35}Cl_{3}N_{2}O_{7}</td>
<td>3-(4-Pentachlorophenyl)-2-(4-(benzylxoy)phenyl)-2-(3-(tert-butoxycarbonyl)amino)propanoyl)amino)propanoate</td>
<td>0.35</td>
<td>0.9E-02</td>
<td>10^{-4.6}</td>
<td>668884</td>
</tr>
</tbody>
</table>

NOTE: The GI_{50} value associated with each compound represents the log-averaged value across all 60 cell lines employed in the screen. The NSC number identifies the compound in the screening database. A more extensive tabulation of the correlation analysis is given in the supplementary documentation [available at Molecular Cancer Therapeutics Online (http://mct.aacrjournals.org)].
Figure 1. Projection of positively correlated GI50 vectors with karyotypic variables. At the selected $P$ of 0.05, the karyotypic variables are correlated with 1,198, 1,006, 1,271, and 381 GI50 data vectors for numerical heterogeneity ($NH$), numerical complexity ($NC$), structural heterogeneity ($SH$), and structural complexity ($SC$), respectively, of which the top 500 hits (yellow dots) are shown. Earlier versions of this and the subsequent figure (both of which also used a different threshold) were included in a forward-looking, solicited perspective (34).

(18). Toward this endeavor, we have used the associated GI50 data vectors to collect diverse structures that are present in the data and represent distinguishable chemical motifs. This technique has been used before to identify typical chemical structures that could be related to the underlying mechanism of action of these drugs (19). In our analysis, we selected structures with GI50 data vectors that were positively associated with the karyotypic variables and have an unadjusted $P$ of ≤0.01. These results are summarized in Table 3, and a more detailed list is available as

<table>
<thead>
<tr>
<th>Karyotypic variable</th>
<th>Numerical complexity</th>
<th>Numerical heterogeneity</th>
<th>Structural complexity</th>
<th>Structural heterogeneity</th>
<th>$\hat{U}_1$</th>
<th>$\hat{U}_2$</th>
<th>$\hat{U}_3$</th>
<th>$\hat{U}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numerical complexity</td>
<td>1.00</td>
<td>0.56</td>
<td>0.40</td>
<td>0.66</td>
<td>0.90</td>
<td>0.03</td>
<td>0.08</td>
<td>0.46</td>
</tr>
<tr>
<td>Numerical heterogeneity</td>
<td>1.00</td>
<td>0.03</td>
<td>0.50</td>
<td>0.72</td>
<td>$-0.56$</td>
<td>$-0.48$</td>
<td>$-0.13$</td>
<td>$-0.13$</td>
</tr>
<tr>
<td>Structural complexity</td>
<td>1.00</td>
<td>0.28</td>
<td>0.51</td>
<td>0.82</td>
<td>$-0.24$</td>
<td>$-0.13$</td>
<td>$-0.13$</td>
<td>$-0.13$</td>
</tr>
<tr>
<td>Structural heterogeneity</td>
<td>1.00</td>
<td>0.86</td>
<td>$-0.11$</td>
<td>0.43</td>
<td>$0.43$</td>
<td>$-0.29$</td>
<td>$-0.29$</td>
<td>$-0.29$</td>
</tr>
</tbody>
</table>
supplementary documentation. A total of 13 classes of potential agents are delineated as motifs A to L. The majority of these compounds has not been identified in the previous pilot study (11) on a much smaller group of compounds ($n = 1,429$).

Motif A in Table 3 consists of cucurbitacins, a class of natural products that are thought to be anti-inflammatory, possibly via actin/vimentin disruption (20, 21) and/or signal transduction modulation (22). Cucurbitacin-like molecules are found in the N11 region of the SOM. Another intensively studied group of molecules included in motif B are the cytochalasins. Cells treated with cytochalasins are arrested in anaphase due to actin depolymerization and a block of cytokinesis (23, 24). Combined with a shorter distance between the spindle poles, this can cause increased frequency of polyploid cells (25, 26).

No mechanisms of action are known for the bis-naphthylcarboxamides, bis-naphthylureas, and anilinomalonyl phenylazopyrazoles, shown in motifs C and D. These drugs are most closely associated with the P region and correlate most strongly with the numerical complexity, numerical heterogeneity, and structural heterogeneity patterns. Compounds of motif D have been found inactive in NCI’s anti-HIV screen. One of the pyridinethione carbonitrile nucleosides listed as motif E in Table 3 is a P-glycoprotein antagonist (27).

The pentachlorophenyl polypeptide esters defined as motif F correlate specifically with the numerical complexity. One compound in this set has been identified as a potential modifier of the c-erbB2 pathway (28), which is intimately connected to cell cycle control (29) and is thus indirectly or directly related to the chromosomal state of the cell.

Motifs G to J encompassing thiazolyl coumarins, anilino/phenoxy-carboxy/phenyl-6(7)-substituted quinoxalines,
1,8-bis(5-aryloxymethyl-4-anilino-1,2,4-triazol-2-yl)octanes, and 3-alkylidene-5,5-disubstituted tetrahydro-2-furanones, listed in Table 3, are not associated with any known mechanism of action or target. They are associated with all karyotypic variables, except for structural complexity, and cluster mainly in the P region of the SOM. Motif I appears in the S6 region of the map, which is colocalized with an abundance of topoisomerase inhibitor GI50 data vectors. This motif carries specificity for the structural heterogeneity karyotypic variable.

The 2-substituted mercapto-3H-quinazolines listed as motif K and mainly found in the P region of the SOM were originally tested for antibacterial, antifungal, and antiacetylcholinesterase activities (30). Subsequent studies involving these compounds have identified them as kinesin inhibitors (31). Because kinesin is directly involved in the mitotic spindle function, these compounds have received attention as antimitotic agents.

The N-(p-(substituted azole)phenyl) benzenesulfonylamides defined as motif L are largely uncharacterized; however, it is interesting to note that this motif and motifs E and J are the only ones indicating specificity toward structural complexity. The 1,1-dimethyl-3-phenyl-3-pyrrolidinyl/4-morpholinyl naphthalans (motif M) are again a relatively unexplored group of structures, although loosely related substructures have been shown to be inhibitors of thymidylate synthase (32), which is critical for DNA repair and replication.

The structurally very different compound classes identified above can be agents of similar target groups as well as different pathways that are common to the particular chromosomal state. Using the GI50 responses for these structures, we have organized these around the karyotypic features.

Exploration of Karyotypic States

The correlations between the data vector of the four karyotypic variables (Table 2) are not totally independent of each other; thus, the information content of each variable is not unique to itself. We therefore did a single value decomposition of the matrix formed by the four karyotypic data vectors and found $w_1 = 0.55$, $w_2 = 0.25$, $w_3 = 0.11$, and $w_4 = 0.08$. Applying the threshold of $0.7/4 = 0.175$, $w_1$ and $w_2$, which accounts for 80% of the variance, indicates that there are truly only two independent biological processes represented in these data. The decomposition does not identify what these processes are. The bulk of the karyotypic data vectors can thus be reconstructed with only the data of the first two karyotypic base vectors $U_1$ and $U_2$. The correlation of all the karyotypic base vectors with the original karyotypic data vectors is given in Table 2. It is evident that $U_1$ carries the largest correlation with all four variables, each contributing differently to the base vector. On the other hand, $U_2$ is strongly positively correlated with the structural complexity variable, which was least related to the other three karyotypic variables.

The representation of the data in terms of the karyotypic base vectors encompasses a model of the data that can be used separately to explore to GI50 correlations. The strong similarity of $U_1$ to the numerical complexity, numerical heterogeneity, and structural heterogeneity karyotypic variables ensures that the same structural motif groups in Table 3 are retrieved using the karyotypic base vectors, albeit with different correlation strengths. Further analysis of the correlations among karyotypic base vectors with compound-induced growth inhibition patterns represented by the structural motifs in Table 3 confirms that these agents are mostly targeted toward the first and most important karyotypic base vector. Motifs E, J, and L are the only ones that seem to be targeting both the first and the second karyotypic base vectors. The other classes are to a greater or lesser degree correlated with $U_2$.

Discussion

The aim of this study was to identify chemical motifs and lead anticancer drug candidates based on their association with the karyotypic state of cancer cells. This analysis correlated karyotypic variables with drug-induced cytotoxicity and gene expression. For generating hypotheses about the mechanism of action for screened compounds (19, 33). Each set of growth inhibition values measured across the NCI-60 tumor cell panel defines the differential response pattern established for each compound tested in the screen. This differential
Table 3. Identified structural motifs and their correlations with the karyotypic variables

<table>
<thead>
<tr>
<th>Structural motif</th>
<th>SOM region</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Numerical</td>
</tr>
<tr>
<td></td>
<td></td>
<td>complexity</td>
</tr>
<tr>
<td>N11</td>
<td>—</td>
<td>0.42</td>
</tr>
<tr>
<td>P3, P7, N11</td>
<td>0.36</td>
<td>0.42</td>
</tr>
<tr>
<td>P3, P4</td>
<td>0.41</td>
<td>0.43</td>
</tr>
<tr>
<td>P3</td>
<td>0.36</td>
<td>0.46</td>
</tr>
<tr>
<td>M2, P1, P4, P10, P14, N11</td>
<td>0.41</td>
<td>0.40</td>
</tr>
<tr>
<td>P10</td>
<td>0.41</td>
<td>—</td>
</tr>
<tr>
<td>P3</td>
<td>—</td>
<td>0.42</td>
</tr>
<tr>
<td>P3, P4, P6, P10, R3</td>
<td>0.47</td>
<td>0.45</td>
</tr>
</tbody>
</table>

(Continued on the following page)
pattern can be used to establish similarities to patterns generated by other compounds and to construct hypotheses regarding the putative mechanism of action for these compounds.

To this aim, we have used SOMs (16) to investigate global trends in the NCI-60 growth inhibition data (15). A SOM attempts to describe the multidimensional space of growth inhibition patterns from all the screened compounds by assigning representative cluster vectors to describe this space. In essence, the algorithm is done via an iterative process designed to organize individual data vectors into clusters, where a single vector represents each cluster’s members. The cluster vectors are then placed on a two-dimensional grid, organized to locate the most similar cluster vectors as nearest neighbors. The practical effect of this algorithm is first to cluster similar data vectors and second to display these results such that the most similar cluster vectors are close in this space (i.e., to provide a global perspective of the complete data set). The visualization of these results is conveniently done via a two-dimensional map that represents a significant reduction in dimensionality from the initial 60-dimensional space. An earlier analysis of the anticancer agents in this data set found that certain regions on the SOM could be associated with putative biological mechanisms of growth inhibition.

In particular, regions on the SOM were delineated that account for agents described previously in the literature as active against DNA synthesis, mitosis, membranes, xenobiotic metabolism, etc. (19, 33). In addition to cataloging compounds according to a mechanism of action, the results revealed an inherent interconnectedness between various cellular processes and specific growth inhibition patterns.

We investigated the raw karyotypic variable data as well as linear combinations as needed to extract the most representative data set that could be associated with chromosomal aberrations. This strategy revealed many compound classes associated with cellular growth inhibition that have not been identified previously as potential effectors of the karyotype. The observation that two of the structural motifs, cytochalasins and 2-substituted mercapto-3H-quinazolines, are already known to act on mitotic function lends credence to this data mining effort. Some of the unexplored chemical motifs identified have been associated previously with a variety of cellular processes, including signal transduction, drug efflux, and DNA maintenance that only circumstantially can be linked to the karyotype of a cell.

It is of some interest that the karyotypic variable whose relationships are the most unrelated to any of the others is structural complexity. 

### Table 3. Identified structural motifs and their correlations with the karyotypic variables (Cont’d)

<table>
<thead>
<tr>
<th>Structural motif</th>
<th>SOM region</th>
<th>Numerical complexity</th>
<th>Numerical heterogeneity</th>
<th>Structural complexity</th>
<th>Structural heterogeneity</th>
<th>( \hat{U}_1 )</th>
<th>( \hat{U}_2 )</th>
<th>( \hat{U}_3 )</th>
<th>( \hat{U}_4 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. (2H-thiophene-2-carboxamide-3H-quinazolines)</td>
<td>S6</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>0.37</td>
<td>—</td>
<td>0.35</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>1. (2H-thiophene-2-carboxamide-3H-quinazolines)</td>
<td>P11, P13</td>
<td>0.43</td>
<td>—</td>
<td>0.40</td>
<td>0.30</td>
<td>0.27</td>
<td>0.29</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>2. Substituted mercapto-3H-quinazolines</td>
<td>P2-P4, P12</td>
<td>0.41</td>
<td>0.41</td>
<td>—</td>
<td>0.43</td>
<td>0.41</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>3. Substituted mercapto-3H-quinazolines</td>
<td>P4, P13, R1, R3</td>
<td>0.37</td>
<td>—</td>
<td>0.39</td>
<td>0.37</td>
<td>0.28</td>
<td>0.49</td>
<td>0.33</td>
<td>0.32</td>
</tr>
<tr>
<td>4. Substituted mercapto-3H-quinazolines</td>
<td>P10</td>
<td>0.39</td>
<td>0.38</td>
<td>—</td>
<td>—</td>
<td>0.37</td>
<td>—</td>
<td>-0.30</td>
<td>—</td>
</tr>
</tbody>
</table>
might be expected given that, of all the variables, it is most difficult to imagine how a cell might detect the fact of an established chromosomal reconfiguration, which is what structural complexity measures. Excess or decreased numbers of chromosomes might be appreciated by a spindle or kinetochore sensor. Ongoing chromosomal gain or loss (numerical heterogeneity) or ongoing chromosomal breakage and rejoining (structural heterogeneity) might similarly be recognized by checkpoint or DNA repair mechanisms, but how a cell would recognize a reconfigured chromosome that contains a single centromere of one or another of the chromosomes involved in the reconfiguration is less easy to hypothesize given current knowledge about cellular function.

In summary, the set of drugs that have been identified via our karyotype/drug correlation analysis provides a set of lead compounds for further study and draws attention to several regions of the SOM. A striking correlation pattern indicates that the karyotypic observables are often correlated with a relatively unexplored region on the SOM. The SOM now provides the identity of compounds that share these growth inhibition patterns but until now have not been recognized in the literature as having an association with the karyotypic state of a cell. This provides several compounds that can be hypothesized to act, directly or indirectly, in a manner relevant to that state. Elucidation of the effect of these drugs is proposed for future assays using, for example, an interface with gene expression array analysis or, for a smaller set of representative compounds, investigation in the yeast haploid deletion system. If such screens identify genes or pathways implicated via the karyotype/drug correlations, we are in a position to use these discoveries to provide a novel set of cancer-relevant targets.

References
Molecular Cancer Therapeutics

Drugs aimed at targeting characteristic karyotypic phenotypes of cancer cells

Anders Wallqvist, Ruili Huang, David G. Covell, et al.


Updated version
Access the most recent version of this article at:
http://mct.aacrjournals.org/content/4/10/1559

Supplementary Material
Access the most recent supplemental material at:
http://mct.aacrjournals.org/content/suppl/2006/01/30/4.10.1559.DC1

Cited articles
This article cites 32 articles, 9 of which you can access for free at:
http://mct.aacrjournals.org/content/4/10/1559.full.html#ref-list-1

Citing articles
This article has been cited by 2 HighWire-hosted articles. Access the articles at:
/content/4/10/1559.full.html#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, contact the AACR Publications Department at permissions@aacr.org.