

# Chemogenomics of Sensitivity and Resistance to Anticancer Drugs

Zunyan Dai\*, Wolfgang Sadee and Paul Blower\*

*Program of Pharmacogenomics, Department of Pharmacology and the Comprehensive Cancer Center, College of Medicine, The Ohio State University, Columbus, Ohio 43210, USA*

**Abstract:** Chemogenomics integrates genomic datasets with biological and chemical characteristics of compounds. It uses genomic or proteomic profiling combined with chemoinformatics and statistics to study the response of a biological system to chemical compounds. This large-scale approach is particularly suitable for addressing chemotherapeutic resistance, a main obstacle to successful treatment of cancer patients. Chemogenomics as a discipline has been molded by a suite of datasets derived from a panel of 60 cancer cell lines that are used for drug discovery by the National Cancer Institute (NCI-60). Offering cytotoxic potencies for >50,000 compounds across the NCI-60, and mRNA profiles, proteomes, mutations, and epigenetic factors, this unparalleled public resource enables rapid discovery of molecular targets and mechanisms of chemosensitivity/resistance. While chemogenomics holds much promise, broader impact in all fields of biology and medicine requires expansion of curated and accessible datasets of diverse biological systems, biological and chemical properties of compound libraries, and novel informatics tools for extracting the valuable information embedded in high-dimensional data. Exploiting the effects of diverse chemical probes, chemogenomics adds an important dimension to systems biology for understanding cellular functions.

## INTRODUCTION

Resistance to chemotherapy prevents the successful treatment of solid tumors and hematological malignancies. Both host factors and genetic or epigenetic changes accumulating during cancer development contribute to the emergence of chemoresistance. Host factors include germ-line polymorphisms in untransformed tissues affecting drug absorption, distribution and metabolism [Gottesman, 2002; Szakacs *et al.*, 2006]. In addition, because of inherent genetic or genomic instability, a tumor's cell population acquires multiple genetic or epigenetic alterations that could lead to innate drug resistance. These changes include alteration of cell cycle and proliferation, decreased uptake and increased efflux of chemotherapeutic drugs, suppression of programmed cell death, and increased DNA repair mechanisms. However more commonly, chemotherapy induces or selects drug-resistant clones leading to acquired drug resistance, the main reason of eventual treatment failure. Frequently, treatment with one drug induces broad-spectrum resistance to multiple other drugs with divergent structures and mechanisms of action. For example, this can occur through upregulation of nonselective enzymes or transporters, such as ABCB1 (MDR1, or P-glycoprotein) [Dai *et al.*, 2004; Huang and Sadee, 2003; 2006; Szakacs *et al.*, 2006].

The human genome project and availability of high-throughput measurements of RNA and protein expression have enabled a broad approach to the search for novel chemo-resistance and chemo-sensitivity factors. With strong investments in genomics and systems biology, large molecular databases on cancers have emerged that can provide

critical information for drug discovery and improved therapies. The combined datasets available for the 60 transformed cancer cell lines used at the NCI for drug discovery (NCI-60) provide a wealth of information on cytotoxic potency of numerous compounds, with possible inference of drug targets and factors involved in chemosensitivity or resistance.

A number of closely related terms are currently in use, such as chemical genetics and chemogenomics, which require proper definition. Chemical genetics studies the function of a protein using chemical compounds to disrupt the protein's function, substituting for mutagenesis studies or gene knockout experiments [Bleicher, 2002; Salemme, 2003; Smukste and Stockwell, 2005; Stockwell, 2000]. The definition of chemogenomics varies somewhat among authors [Bleicher, 2002; Bredel and Jacoby, 2004]. In the context of the current review, chemogenomics is defined as the use of genomics to measure the system-wide effect of a compound on a biological system, either single cells or whole organisms [Bredel and Jacoby, 2004]. It combines high-throughput genomics or proteomics profiling with chemoinformatic and statistical analysis to study the response of a biological system to chemical compounds [Bredel and Jacoby, 2004]. Chemoinformatics in turn deals with a broad informatics approach to understanding diversity of chemical structures, typically using structural descriptors for defining compound classes. Addition of chemoinformatics expands the scope of chemogenomics by enabling one to deal with numerous compounds and compound classes simultaneously. This has proven valuable where the analysis of the effect of single compounds is insufficient to draw inferences on biological mechanisms. Cellular response is measured by phenotypic readouts in a high-throughput assay. Chemogenomics also investigates the consequences of differential gene/protein expression on cellular response to compound treatment. This review focuses on the NCI-60 cell panel as a chemogenomics model system for understanding the chemo-sensitivity and -resistance mechanisms of anticancer compounds. We also discuss briefly other chemogenomics systems for novel

\*Address correspondence to these authors at the Department of Pharmacology and the Comprehensive Cancer Center, College of Medicine, The Ohio State University, Columbus, Ohio 43210, USA; Tel: 614-292-6680; Fax: 614-292-7232; E-mail: dai.15@osu.edu

Department of Pharmacology and the Comprehensive Cancer Center, College of Medicine, The Ohio State University, Columbus, Ohio 43210, USA; Tel: 614-292-5565; Fax: 614-292-7232; E-mail: blower.7@osu.edu

anticancer discovery, drug target validation, and analysis of mechanisms of drug toxicity.

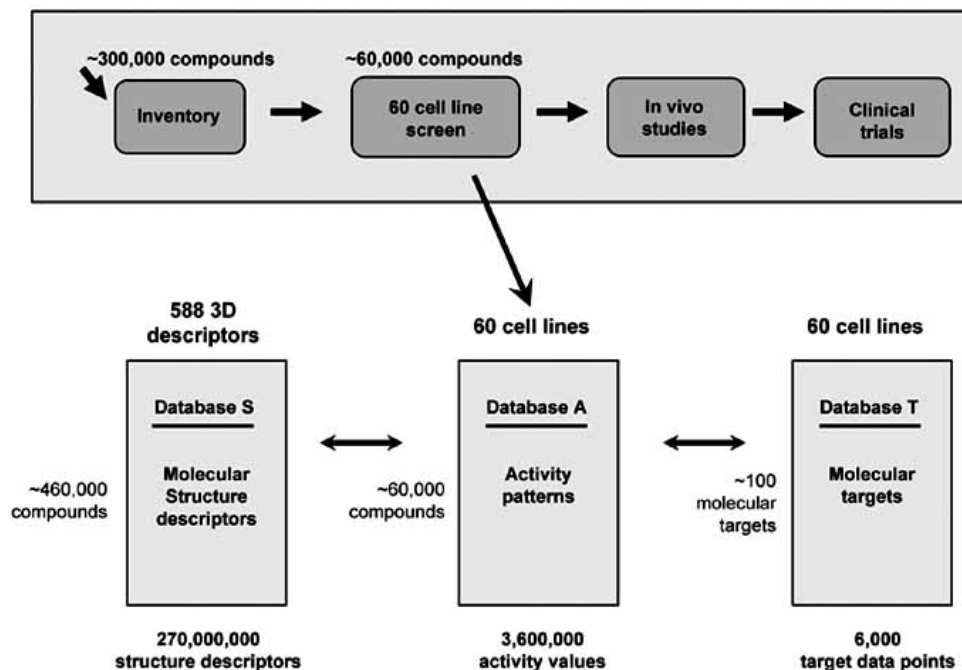
### CHEMOGENOMICS RESEARCH BASED ON NCI DATABASES

Since 1990, the NCI's Developmental Therapeutics Program has been testing compounds for growth inhibition against a panel 60 human tumor cell lines (NCI-60). The data provide  $GI_{50}$  values for each compound-cell line pair, that is the concentration causing 50% growth inhibition [Boyd and Paull, 1995]. The NCI-60 cell line panel are derived from nine broad categories of tumor types: breast (5 cell lines), CNS (6), colon (7), leukemias (6), lung (9), melanomas (10), ovarian (7), prostate (2), and renal (8). Because they have been removed from the *in vivo* tumor environment and selected for growth in culture, the NCI-60 cell lines do not accurately represent primary tumors. However, they offer a number of advantages over primary tumors for chemogenomic studies in cancer [Weinstein and Pommier, 2003]. The cell lines are reasonably stable and reproducible over extended time periods. Thus far more than 100,000 compounds have been profiled over 16 years, many of which have been tested repeatedly. Screening data for approximately 43,000 non-proprietary compounds is publicly available online ([http://dtp.nci.nih.gov/docs/cancer/cancer\\_data.html](http://dtp.nci.nih.gov/docs/cancer/cancer_data.html)).

In addition to pharmacological profiling, the NCI-60 has also been the subject of numerous genomic, proteomic and other systems biology studies (collectively -omic studies, [Weinstein, 2002]). These included transcription profiling [Annereau *et al.*, 2004; Butte *et al.*, 2000; Dai *et al.*, 2006; Huang *et al.*, 2004; Ross *et al.*, 2000; Scherf *et al.*, 2000;

Staunton *et al.*, 2001], proteomics [Nishizuka *et al.*, 2003a; Nishizuka *et al.*, 2003b], single nucleotide polymorphisms [Weinstein and Pommier, 2003], DNA copy-number changes and chromosomal aberrations [Bussey *et al.*, 2006; Roschke and Kirsch, 2005a; Roschke *et al.*, 2005b; Roschke *et al.*, 2003; Wallqvist *et al.*, 2005] and DNA methylation of cancer-related genes [Weinstein and Pommier, 2003]. In the near future, microRNA expression profiles will also be available (Blower *et al.*, unpublished results). The collection of datasets related to the NCI-60 provides an unparalleled public resource for integrated chemogenomic studies, leading to hypotheses that can be tested in the clinic. In this section, we review the development of datasets related to the NCI-60 and chemogenomic studies aimed at molecular drug targets and mechanisms of chemosensitivity/resistance.

In 1997, Weinstein *et al.* laid out a conceptual framework for large-scale integration of databases that reveals connections between potencies of cytotoxic compounds and cellular characteristics [Weinstein *et al.*, 1997]. This approach involves the three databases shown in Fig. (1). Database [A] contains the activity patterns of tested compounds over the NCI-60 cell lines, [S] contains molecular structural features of the compounds, and each row in database [T] contains the pattern (across the NCI-60) of a measured cellular characteristic that may modulate cellular response to compound dosing. At the time, database [T] contained information for approximately 100 cellular characteristics such as mutational status of RAS and p53, protein levels of p53, response to  $\gamma$  radiation, and rhodamine efflux by MDR-1. Databases [A] and [T] were integrated by calculating the correlation between the activity pattern of a compound and the characteris-



**Fig. (1).** Simplified schematic overview of an information-intensive approach to cancer drug discovery and molecular pharmacology at the NCI. Each row of the activity (A) database represents the pattern of activity of a particular compound across the 60 cell lines. As described in the text, the A database can be related to a structure (S) database containing 2D or 3D chemical structure characteristics of the compounds and a target (T) database containing information on possible molecular targets or modulators of activity within the cells. Adapted from Science, Ref. [Weinstein 1997]; statistics are 1997 data; used with permission.

tic pattern of a target. Each row of [A] and [T] was normalized and the inner product of the two matrices was calculated, producing a matrix of Pearson correlation coefficients. Each row in the correlation matrix corresponds to a compound, each column to a target, and entry (i,j) gives the correlation between the activity pattern of compound  $C_i$  and the characteristic pattern of target  $T_j$ .

To facilitate visual pattern recognition of this large database, the authors introduced the Cluster Image Map, a device now widely used in genomic studies and more commonly known as a heatmap. A heatmap is a color image showing the strength of association between row and column elements. Both row and column elements are arranged in cluster order and a dendrogram is added to the left side and to the top. Rows and columns are clustered separately. Rows are clustered based on correlation between compound potency patterns so that compounds with similar patterns are grouped together. The columns are clustered analogously based on target characteristic patterns. The color of the (i,j)-element shows the strength of correlation between the activity pattern of  $C_i$  and the expression pattern  $T_j$  over the NCI-60 cell lines. As such it links molecular pharmacology with cellular targets and provides hypotheses that can be tested experimentally.

Scherf *et al.* were the first study to integrate large databases on gene expression and molecular pharmacology [Scherf *et al.*, 2000]. They used cDNA microarrays to measure gene expression levels in the NCI-60 cell lines and correlated the expression patterns with drug activity patterns. Strong negative correlation between the expression pattern of a gene and the activity pattern of the compound means that cells expressing higher levels of the gene are less sensitive to this compound. Therefore, correlations suggest chemoresistance mechanisms and provide hypotheses that can be tested experimentally. For example, expression of dihydropyrimidine dehydrogenase (DPYD) was negatively correlated with potency of 5-fluorouracil (5-FU) over the NCI-60, so that cell lines with low expression of DPYD which inactivates 5-FU, tend to be more sensitive to 5-FU. Similarly, expression of asparagine synthetase (ASNS) was negatively correlated with L-asparaginase sensitivity, consistent with the fact that cells lacking ASNS are sensitive to L-asparaginase, which depletes extracellular L-asparagine. Although the correlation was only moderately high over the full 60 cell lines, it was much stronger within the leukemia and ovarian subpanels with values of  $r = -0.98$  for leukemias and  $r = -0.88$  for ovarians. These examples illustrate how variations in expression levels of particular genes relate to mechanisms of drug sensitivity and resistance.

Staunton *et al.* investigated the question of whether patterns of gene expression profiles of untreated cells are sufficient for predicting sensitivity or resistance of the NCI-60 cell lines to chemical compounds [Staunton *et al.*, 2001]. The authors developed an algorithm for classification of cell line chemosensitivity based on gene expression profiles. Oligonucleotide microarray profiles for 6,817 genes were measured for the NCI-60 cell lines, and 232 compounds with broad dynamic ranges of drug responses, from resistant to sensitive, over the NCI-60 were chosen for study. A separate classifier, designed to be independent of the cells' tissue of

origin, was built for each compound. Optimal predictor sets, with an average of 68 genes/classifier, were selected by leave-one-out cross validation and then evaluated on independent sets of data. The accuracy of chemosensitivity prediction was considerably better than would be expected by chance, suggesting that for some compounds using gene expression profiles to classify compounds as sensitive or resistant is feasible. However, this approach may have heuristic value applicable only to these cell lines.

Butte *et al.* constructed a large network with genes and drugs as nodes and correlation coefficients as edge weights between all pairs of nodes [Butte *et al.*, 2000]. The authors used Affymetrix oligonucleotide chips to measure expression levels of 7,245 genes over the NCI-60 cell lines, joined this with a database of cytotoxicity patterns for 5,084 drugs measured by the NCI, and calculated all pairwise correlations: gene-gene, gene-drug, and drug-drug. Then edges between nodes with weights below a threshold were removed resulting in a relevance network comprised of approximately 200 connected subnetworks. At the threshold chosen, 88% of the subnetworks contained only drugs, and the majority of these were analog classes. Only one network contained a gene-drug association between lymphocyte cytosolic protein-1 (LCP1) and a thiazolidine carboxylic acid derivative. Therefore, valuable gene-drug correlations are lost in this analysis. Possibly, setting different scales for gene-gene, gene-drug, and drug-drug correlations might enable a chemogenomics analysis of all relevant interactions.

Lee *et al.* compared microarray gene-expression levels for the NCI-60 measured by two different expression-profiling platforms, spotted cDNA arrays and Affymetrix oligonucleotide chips [Lee *et al.*, 2003]. Analysis of the distribution of correlation coefficients between sequences on the two arrays mapping to the same Unigene cluster revealed a bimodal distribution with 63% in a peak around  $r = 0.6$  and 37% in a peak around  $r = 0$ . A correlation cutoff was selected at  $r = 0.3$  because only 3% of the random, non-matching transcript pairs had  $r > 0.3$ , whereas 63% of the UniGene-matched pairs met that criterion. The analysis resulted in a consensus set of 1,493 genes that give similar profiles on both platforms. In a comparison between mRNA profiles of membrane transporters, obtained with spotted oligonucleotide arrays and Affymetrix chips, similar results were produced with both arrays, but only for well expressed genes [Landowski *et al.*, 2004]. This result cautions against over-interpretation of array results where the mRNA is not robustly expressed.

Wallqvist *et al.* compared datasets of microarray gene-expression profiles for the NCI-60 measured by three different laboratories, one using spotted cDNA arrays and two using Affymetrix oligonucleotide chips [Wallqvist *et al.*, 2002]. They found 2105 genes common to all three datasets and approximately 400 significant concordant expression patterns which serve as replicate measurements of the same dataset. The authors then linked expression profiles of the concordant genes with compound classes. The compounds were clustered based on cytotoxicity profiles using a self organizing map (SOM). Each gene was mapped to the compound cluster with minimal Euclidean distance between the expression vector and the SOM node vector. For 11 cases,

they found a correspondence between gene expression patterns and compound cytotoxicity patterns that were verified from ligand-target crystallographic data on the Protein Data Bank.

Wallqvist *et al.* developed data mining tools to study associations between chemosensitivity measured by the activity patterns of 36,000 compounds tested by the NCI and gene expression, molecular target measurements and mutational status over the NCI-60 cell lines [Wallqvist *et al.*, 2003]. The authors clustered the compounds into nodes using a self-organizing map procedure [Rabow *et al.*, 2002]. Each node is assigned a fingerprint which encodes the growth inhibition pattern from dosing by compounds within the node. A similarity score is calculated between each node fingerprint and the fingerprints for measurements of mRNA expression, molecular targets and mutational status. For each gene or molecular target, the similarity scores are mapped onto the SOM using a coloring scheme which gives a visual image of the correlations. The authors found that seventy percent of the response patterns consisted of two anti-correlated maps and concluded that gene expression levels do not, in general, correlate with compound activity patterns but instead they reflect a generic toxic condition. Within the remaining set of non-generic response patterns, numerous examples were found where a correlation suggested a biochemical basis for cellular cytotoxicity.

In an extension of the work of Wallqvist linking cytotoxicity and gene expression data for the NCI-60 [Wallqvist *et al.*, 2002; 2003], Covell *et al.* focused on a previously unexplored region of the SOM comprising some 4,000 compounds which exhibited unique cytotoxicity response profiles [Covell *et al.*, 2005; Rabow *et al.*, 2002]. The authors performed an extensive structural analysis of compounds within the focus region and identified characteristic structural motifs and drug-gene associations supported either from literature reports or binding with analogous compounds in the Protein Data Bank. This revealed connections between compounds clustered in the focus region and genes involved in cellular detoxification, extracellular matrix and cytoskeletal proteins, DNA damage repair processes, oxidative stress responses, and apoptosis.

Although technically more difficult, proteomic profiling should be more directly related to pharmacological response than mRNA expression profiles. Nishizuka *et al.* measured protein levels for the NCI-60 using reverse-phase protein lysate microarrays [Nishizuka *et al.*, 2003a]. This technology permits all of the 60 cell lines to be analyzed for a specific protein side by side in a single array. The authors reported results for 52 proteins and used the data to compare the patterns of protein expression with patterns for mRNA levels for the same genes. They found that clustering of cell lines based on protein levels generally resembled those obtained from clustering by gene expression patterns. However, while cell-structure related proteins showed a high correlation between mRNA and protein levels across the NCI-60 cell lines, other protein classes did not. This result also cautions against use of mRNA level alone as the marker of chemoresistance. Important relationships may be missed. In a follow-on study using the NCI-60 cell lines, Nishizuka *et al.* developed a procedure for identifying molecular markers for the differen-

tial diagnosis of colon and ovarian cancers [Nishizuka *et al.*, 2003b]. They identified villin as a candidate marker for colon cancers and moesin for ovarian cancers, both proteins involved in the structural assembly of cytoskeletal elements. Colon and ovarian cancers are difficult to distinguish clinically, and this determines selection of chemotherapy.

Blower *et al.* pursued a suggestion from Weinstein for projecting genomic information from the NCI-60 through the activity patterns of compounds to molecular structural characteristics of those compounds [Blower *et al.*, 2002; Weinstein *et al.*, 1997]. The relationship between compound activity profiles and gene expression patterns, the matrix  $[A \cdot T^T]$  of Pearson correlation coefficients was computed (Fig. 1). From this, the matrix  $[S \cdot A \cdot T^T]$  (or SAT for Structure-Activity-Target) was generated to associate a structural feature  $F$  with a gene. Each row of the SAT matrix corresponds to a structural feature  $F$ ; each column, a gene  $G$ ; and the  $ij$ -th element reflects the tendency of substructure  $F_i$  to occur in compounds that are more active in cell lines that express large amounts of the  $G_j$  gene product. Thus the SAT matrix identifies structural features that are associated with the observed correlations between gene expression and growth inhibition, and this can be used for structure-based data mining.

Blower *et al.* describe statistical techniques for selecting genes with characteristic expression patterns, and apply structure-based data mining software to identify substructural compound classes correlated with gene expression. This approach can shed light on molecular mechanisms and has the potential to accelerate drug discovery in several ways: (i) It can be used to prioritize genes for follow-up studies as potential chemosensitivity/resistance targets. (ii) Because the analysis projects genomic information to molecular substructure, it allows extraction of a preliminary structure-activity relationship (SAR) directly from the SAT correlations. (iii) The preliminary SAR can, in turn, be used for early pharmacophore development or to select new, untested drug candidates from an actual or virtual library of compounds. (iv) It can be used to prioritize candidate compounds for detailed gene expression analysis or other biological studies.

Roschke *et al.* analyzed chromosomal aberrations in the NCI-60 and studied karyotypic complexity and heterogeneity based on three components of genomic anatomy: ploidy, numerical chromosome changes, and structural chromosome rearrangements [Roschke *et al.*, 2003]. The NCI-60 showed wide variations. In a follow-on study, Roschke *et al.* examined the question of whether karyotypic complexity and instability showed associations with the activity patterns of potential anticancer compounds [Roschke *et al.*, 2005b]. The authors analyzed the correlation between compound activities and five karyotypic parameters and identified several classes of compounds (e.g., ellipticines, benzodithiophenediones) that are more cytotoxic toward cell lines that are more karyotypically complex and unstable. However, most chemotherapeutic agents do not show significant correlations with the karyotypic parameters.

Bussey *et al.* analyzed relationships between DNA copy number, mRNA expression and drug sensitivity in the NCI-60. DNA copy numbers were assessed for 353 cancer related

genes using comparative genomic hybridization (CGH) microarrays [Bussey *et al.*, 2006]. The authors examined correlations between copy number changes and drug activities for 118 drugs with known mechanisms of action. As expected they found strong negative correlations between ABCB1 copy number and known MDR1 substrates. They also found positive correlations for tubulin acting agents and negative correlations for agents that incorporate into DNA. Therefore, gene dosage effects can be tested and have the potential to affect chemoresistance.

### Applications

In this section we review studies that use informatics approaches to associate compounds or compound classes with relevant gene families, followed by experimental validation. These studies exemplify the utility of chemogenomics in combination with rapid *in vitro* testing of drug candidates in selected cell lines expressing the target gene.

Huang *et al.* used long (70-mer) oligonucleotide arrays to analyze gene expression of membrane transporters and channels (632 probes) in the NCI-60 cell lines [Huang *et al.*, 2004]. By analyzing correlations between gene expression and the potencies of 119 standard anticancer drugs, the study identified numerous potential drug-transporter relationships and validated several such relationships by using small molecule inhibitors of transporters or siRNA knockdown. For example, the expression of nucleoside transporter SLC29A1 is positively correlated with the potency of nucleoside analogues, azacytidine and inosine-glycodialdehyde, and inhibition of SLC29A1 significantly reduced the potency of these two drugs. In other examples, the authors identified significant negative correlations between several ABC efflux transporters and multiple drugs, suggesting potential mechanisms of drug resistance. Several negative drug-transporter correlations were validated using specific siRNA which increased potency of several substrates. The authors concluded that measurement of transporter gene expression may prove useful in predicting anticancer drug response.

Huang *et al.* described a systematic study of chemoresistance involving ABCB1 across diverse classes of anticancer drug candidates based on correlation analyses between cytotoxic drug potency and gene expression in NCI-60 cell lines [Huang *et al.*, 2005a]. Ellipticine analogs displayed a range of correlation coefficients with ABCB1, encoding the multi-drug resistance protein MDR1. Cytotoxicity assays (using MDR1 inhibitors and siRNA-mediated MDR1 downregulation in MDR1-overexpressing cells) were used to test MDR1 interactions with five ellipticines with a broad range of MDR1 correlation values. A structure-activity study within the ellipticine class of 61 compounds revealed a strong substituent effect. The subclass of ellipticiniums (with an N-alkyl substituted pyridine) had strong, negative correlations with MDR1 and were shown to be MDR1 substrates. In contrast, the subclass of ellipticines (with an unsubstituted pyridine) had neutral or positive correlations and were not MDR1 substrates. This provides a powerful approach for selecting compounds from a series of congeners that can avoid specific resistance mechanisms.

Huang *et al.* studied the role of SLC7A11 in chemoresistance. SLC7A11 encodes the transporter subunit of the het-

erodimeric amino acid transport system which is thought to mediate resistance against multiple drugs by regulating cellular levels of glutathione [Huang *et al.*, 2005b]. Using a subset of 1,400 compounds tested by the NCI, the authors found numerous instances where the activity pattern of the compound shows a high negative correlation with the expression pattern of SLC7A11 over the NCI-60 cell lines. For example, analogs of geldanamycin yielded widely varying correlations with SLC7A11. The parent compound geldanamycin has a high negative correlation with SLC7A11, and experimental results confirmed that glutathione is the likely substrate mediating chemoresistance. Down-regulation of SLC7A11 by small interfering RNA or inhibition of transport system with small molecule inhibitors increased the potency of geldanamycin in cells expressing large amounts of SLC7A11. However, the geldanamycin analog 17-AAG, for which the potency pattern is uncorrelated with SLC7A11 expression, was not affected.

Dai *et al.* developed a systematic procedure for identifying an optimal subset of genes predictive of cytotoxic potency in the NCI-60 cell lines [Dai *et al.*, 2006]. The study focused on gene families related to growth factor signaling, using a database of 343 genes, including growth factors and receptors, metalloproteinases, and ras-like GTPases. The genes were correlated with the NCI mechanism of action dataset of 119 drugs ([http://dtp.nci.nih.gov/docs/cancer/searches/standard\\_mechanism.html](http://dtp.nci.nih.gov/docs/cancer/searches/standard_mechanism.html)). Negatively correlated genes clustered into two main groups with distinct expression profiles and drug correlations, represented by EGFR and ERBB2 (Her-2/Neu). A subset of only 13 genes was found to be sufficient for near optimal prediction of drug potency against the NCI-60. The approach pares down the number of predictors to a minimum.

### Data Analysis Software

A number of publicly available software systems have been developed for analysis of gene expression data [D'Alimonte *et al.*, 2005; Kapushesky *et al.*, 2004; Reich *et al.*, 2004; Sarkans *et al.*, 2005; Sturn *et al.*, 2003; Yang *et al.*, 2005; Zeeberg *et al.*, 2003; Zeeberg *et al.*, 2005]. GoMiner is a web-based client-server application for analysis and interpretation of results from multiple microarray experiments [Zeeberg *et al.*, 2003; Zeeberg *et al.*, 2005]. It provides several useful visualizations of "interesting" genes. One displays selected genes in the context of the Gene Ontology (GO) hierarchy and identifies GO categories that are enriched in genes that are significantly up- or down-regulated. The system also provides URL links to external data resources.

Expression Profiler is an open-source web-based platform for analysis and visualization of gene expression data developed by a consortium based at the European Bioinformatics Institute [Kapushesky *et al.*, 2004]. The software provides components for filtering, clustering, similarity searching, and visualization as well as statistical techniques for identifying significant genes. GeneCluster is a software package that provides statistical tools for analyzing gene expression data including self-organizing maps for clustering and several supervised learning methods for classifying an unknown gene based on a training set of labeled samples

[Reich *et al.*, 2004]. Genesis is a client-server application for large scale clustering and visualization of gene expression data, although visualization methods are not described [Sturn *et al.*, 2003]. MetNet3D is a 3D virtual reality system, focused primarily on visualization, that allows a user to explore interactions between gene clusters and metabolic networks [Yang *et al.*, 2005].

While we have surveyed a number of programs for linking gene expression or other cellular data with cytotoxic potencies of drugs, none of the programs are publicly available and none provides a full range of software tools for integration, data mining and visualization of disparate datasets. An integrated software platform that enables systematic exploration of integrated chemogenomic data, including compound structure and activity patterns, gene and protein expression levels and other cellular factors, could quickly produce a wealth of hypotheses for follow-up experimental structure-activity studies.

### SUMMARY OF OTHER METHODS

The major goal of chemogenomics is the concurrent target identification and drug discovery using high-throughput genomics technologies and chemical libraries for the treatment of human diseases [Bredel and Jacoby, 2004]. In addition to the NCI-60 human cell lines, advances in high throughput chemogenomic profiling, such as heterozygous yeast deletion libraries for drug target identification [Giaever 2003; Giaever *et al.*, 2002; Giaever *et al.*, 2004; Giaever *et al.*, 1999], gene expression based high-throughput drug screening (GE-HTS) *in vitro* system [Stegmaier *et al.*, 2004], and the development of a large-scale chemogenomics database in rats [Ganter *et al.*, 2005], promise to accelerate the discovery of potential drug targets and/or novel drugs, and increase understanding of complex interactions between components of a biological system.

The heterozygous yeast deletion library for drug target identification is based on the theory that reduced gene dosage of drug targets from two copies in a wild-type strain to one copy after heterozygous deletion will increase drug sensitivity [Giaever 2003; Giaever *et al.*, 2002; Giaever *et al.*, 2004; Giaever *et al.*, 1999]. Giaever *et al.* pooled approximately 6000 heterozygous yeast deletion strains, which represented about 96% of yeast genes in combination [Giaever *et al.*, 2004]. Chemical compounds were added to the culture of the pooled strains. DNA from the culture was hybridized to oligo-microarrays to identify the strains that are sensitive and inhibited by the compounds. The authors characterized gene products interacting with anticancer, antifungal, and anticholesterol agents. The authors applied the method to three anticancer drugs, methotrexate, 5-FU, and cisplatin. In the case of methotrexate, not only the deletion of dihydrofolate reductase (DFR1, DHFR, the known methotrexate target), but also genes affecting methotrexate transportation, increased methotrexate sensitivity. However, no significant targets were found for both 5-FU and cisplatin revealing limited utilities. This technique can accelerate the drug discovery process by revealing unknown drug targets, increase mechanistic understanding, and aid in prioritizing compounds for development [Giaever *et al.*, 2004].

Stegmaier *et al.* described a general approach to identifying compounds that modify cellular states, such as myeloid differentiation of AML (acute myelogenous leukemia) cells [Stegmaier *et al.*, 2004]. Induction of differentiation of AML cells has therapeutic potential. However, since the mechanisms or targets of undifferentiated AML were unknown, the differentiation phenotype had been measured by cellular morphology changes or biochemical tests, such as nitroblue tetrazolium (NBT) reduction, which are both low throughput tests. The authors used microarray technology to identify a gene expression signature, comprising 4 genes, that could distinguish AML cells before and after differentiation. The gene-expression signature, as a surrogate for the differentiation phenotypes, was used to screen a library containing 1,739 chemical compounds. Eight chemical compounds induced the differentiation signature and were confirmed by morphology test and NBT reduction analysis as differentiation inducers. Gene expression based high-throughput screening (GE-HTS), using gene expression signatures as surrogates for biological state changes, could be a useful strategy for identifying novel biological process modulators without knowing the key effectors [Stegmaier *et al.*, 2004].

In a recent study from the Broad Institute, Lamb *et al.* described the creation of a Connectivity Map intended as a public resource of drug-gene signatures for linking drugs, genes and diseases [Lamb *et al.*, 2006]. The authors profiled 164 bioactive small molecules including many FDA approved drugs in four of the NCI-60 cell lines. Typically, cells were treated with 10 $\mu$ M for 6 hours, then mRNA expression levels were measured on Affymetrix GeneChip microarrays giving a database of reference profiles. Each reference profile comprises a rank ordered list of genes ordered by differential expression relative to the batch control. The database can be searched by comparing a query signature of up- and down-regulated genes with reference signatures. The authors provide several examples demonstrating the use of genomics signatures for recognizing structurally diverse drugs with common mechanisms of actions.

In addition to the discovery of drug targets and novel drugs, chemogenomics has also been applied to drug toxicity assessment. For this purpose, Ganter *et al.* from Iconix Pharmaceuticals have developed a large-scale chemogenomics database [Ganter *et al.*, 2005]. The database contains information from a rat *in vivo* model and comprises approximately 600 different compounds, including more than 400 FDA approved drugs. The rats were treated at different dosage levels and time periods, then sacrificed to measure the traditional toxicity endpoints, clinical chemistry, hematology panel, relative organ weights and histopathology. Profiles for over 3200 different drug-dose-time combinations in up to 7 different tissues were collected. In addition, a total of 10,997 gene expression arrays were applied to control and experimental groups based on different drug-dose-time-tissue combinations. The authors used 3 anticancer drugs, carmustine, methotrexate, and thioguanine, to illustrate the use of gene expression profiles to assess drug toxicity compared to traditional toxicity measurements. The three drugs had similar toxicity to decrease blood cell counts, which coincide with the gene express patterns. However, carmustine had different gene expression signature associated with its hepa-

toxicity, which is distinct from the other two. In addition, the gene expression repression of aminolevulinic synthase 2 (Alas2) measured by microarrays in liver samples was a surrogate biomarker for reticulocyte depletion in blood samples. The authors proposed that the database was valuable for evaluating molecular, cellular and toxicological effects important for drug development [Ganter *et al.*, 2005].

## CONCLUSIONS

The scientific community has made large investments in genomic and other systems biology studies, producing large databases of molecular information on cancers that promise to provide critical information for accelerating drug discovery and delivering improved therapies. For example, the data available for the NCI-60 cell lines provide a wealth of information on response to drug dosing and cellular targets and factors that may predict chemosensitivity or resistance to treatment. In the near future, we expect additional data on protein levels, mutational status, and microRNA expression will further enrich this picture and help to explain patterns of chemosensitivity/resistance for classes of compounds. However, to date there have been relatively few cases where strong target-substrate associations have been discovered and experimentally validated.

Most data mining efforts have focused on single genes and used simple correlations between compound potency and gene expression patterns to identify potential substrate associations. Even for strong gene-compound associations such as MDR1-taxol, where the expression pattern of MDR1 (ABCB1) across the NCI-60 is highly negatively correlated with GI<sub>50</sub> values of taxol, the correlation accounts for only a portion of the observed potencies. Indeed, discovering a single gene that can predict a main share of the activity pattern of a discriminating compound is unlikely. Often the relationship between mRNA and protein level is not linear, and hidden regulatory factors such as microRNAs may be at play. For most compounds, combinations of genes and microRNAs will need to be considered, and DNA copy number, methylation, and mutational status may also play a role. Thus, statistical methods for multivariate analysis and feature selection (combinations of genes, microRNAs, and other factors) will need to be employed to predict compound activity patterns. It is also unlikely that accurate global models can be constructed for diverse compounds, and compounds or classes of compounds will need to be treated separately with multiple local models.

Although available data provide a rich picture of the NCI-60 cell lines, this information is largely unexploited, in part, because of the sheer volume of data available and lack of tools to integrate and correlate it. A number of valuable analytical tools are publicly available on NCI websites (<http://discover.nci.nih.gov/> and <http://spheroid.ncicrf.gov/spheroid/>). An important next step is the development of an integrated software platform for exploration, data mining and visualization of integrated data related to the NCI-60 cell lines, including compound structure and activity patterns, gene/protein/microRNA expression levels as well as factors such as DNA copy number, methylation, and mutational status. Such a platform will make systematic exploration of this valuable data a possibility and promises to dramatically

increase the return on the investment the scientific community has already made in data collection.

## REFERENCES

- Annereau, J. P.; Szakacs, G.; Tucker, C. J.; Arciello, A.; Cardarelli, C.; Collins, J.; Grissom, S.; Zeeberg, B. R.; Reinhold, W.; Weinstein, J. N.; Pommier, Y.; Paules, R. S. and Gottesman, M. M. (2004) Analysis of ATP-binding cassette transporter expression in drug-selected cell lines by a microarray dedicated to multidrug resistance. *Mol. Pharmacol.* **66**, 1397-1405.
- Bleicher, K. H. (2002) Chemogenomics: bridging a drug discovery gap. *Curr. Med. Chem.* **9**, 2077-2084.
- Blower, P. E.; Yang, C.; Fligner, M. A.; Verducci, J. S.; Yu, L.; Richman, S. and Weinstein, J. N. (2002) Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data. *Pharmacogenomics J.* **2**, 259-271.
- Boyd, M. R. and Paull, K. D. (1995) Some practical consideration and applications of the National Cancer Institute *in vitro* anticancer drug discovery screen. *Drug Dev. Des.* **34**, 91-109.
- Bredel, M. and Jacoby, E. (2004) Chemogenomics: an emerging strategy for rapid target and drug discovery. *Nat. Rev. Genet.* **5**, 262-275.
- Bussey, K. J.; Chin, K.; Lababidi, S.; Reimers, M.; Reinhold, W. C.; Kuo, W. L.; Gwadry, F.; Ajay, K.; Kouros-Mehr, H.; Fridlyand, J.; Jain, A.; Collins, C.; Nishizuka, S.; Tonon, G.; Roschke, A.; Gehlhaus, K.; Kirsch, I.; Scudiero, D. A.; Gray, J. W. and Weinstein, J. N. (2006) Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. *Mol. Cancer Ther.* **5**, 853-867.
- Butte, A. J.; Tamayo, P.; Slonim, D.; Golub, T. R. and Kohane, I. S. (2000) Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl. Acad. Sci. USA* **97**, 12182-12186.
- Covell, D. G.; Wallqvist, A.; Huang, R.; Thanki, N.; Rabow, A. A. and Lu, X. J. (2005) Linking tumor cell cytotoxicity to mechanism of drug action: an integrated analysis of gene expression, small-molecule screening and structural databases. *Proteins* **59**, 403-433.
- Dai, Z.; Barbacioru, C.; Huang, Y. and Sadee, W. (2006) Prediction of anticancer drug potency from expression of genes involved in growth factor signaling. *Pharm. Res.* **23**, 336-349.
- Dai, Z.; Huang, Y. and Sadee, W. (2004) Growth factor signaling and resistance to cancer chemotherapy. *Curr. Top. Med. Chem.* **4**, 1347-1356.
- D'Alimonte, D.; Lowe, D.; Nabney, I. T.; Mersinias, V. and Smith, C. P. (2005) MILVA: an interactive tool for the exploration of multidimensional microarray data. *Bioinformatics* **21**, 4192-4193.
- Ganter, B.; Tugendreich, S.; Pearson, C. I.; Ayanoglu, E.; Baumhueter, S.; Bostian, K. A.; Brady, L.; Browne, L. J.; Calvin, J. T.; Day, G. J.; Breckenridge, N.; Dunlea, S.; Eynon, B. P.; Furness, L. M.; Ferng, J.; Fielden, M. R.; Fujimoto, S. Y.; Gong, L.; Hu, C.; Idury, R.; Judo, M. S.; Kolaja, K. L.; Lee, M. D.; McSorley, C.; Minor, J. M.; Nair, R. V.; Natsoulis, G.; Nguyen, P.; Nicholson, S. M.; Pham, H.; Roter, A. H.; Sun, D.; Tan, S.; Thode, S.; Tolley, A. M.; Vladimirova, A.; Yang, J.; Zhou, Z. and Jarnagin, K. (2005) Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.* **119**, 219-244.
- Giaever, G. (2003) A chemical genomics approach to understanding drug action. *Trends Pharmacol. Sci.* **24**, 444-446.
- Giaever, G.; Chu, A. M.; Ni, L.; Connelly, C.; Riles, L.; Veronneau, S.; Dow, S.; Lucau-Danila, A.; Anderson, K.; Andre, B.; Arkin, A. P.; Astromoff, A.; El-Bakkoury, M.; Bangham, R.; Benito, R.; Brachet, S.; Campanaro, S.; Curtiss, M.; Davis, K.; Deuschbauer, A.; Entian, K. D.; Flaherty, P.; Foury, F.; Garfinkel, D. J.; Gerstein, M.; Gotte, D.; Guldener, U.; Hegemann, J. H.; Hempel, S.; Herman, Z.; Jaramillo, D. F.; Kelly, D. E.; Kelly, S. L.; Kotter, P.; LaBonte, D.; Lamb, D. C.; Lan, N.; Liang, H.; Liao, H.; Liu, L.; Luo, C.; Lussier, M.; Mao, R.; Menard, P.; Ooi, S. L.; Revuelta, J. L.; Roberts, C. J.; Rose, M.; Ross-Macdonald, P.; Scherens, B.; Schimmack, G.; Shafer, B.; Shoemaker, D. D.; Sookhai-Mahadeo, S.; Storms, R. K.; Strathern, J. N.; Valle, G.; Voet, M.; Volckaert, G.; Wang, C. Y.; Ward, T. R.; Wilhelm, J.; Winzler, E. A.; Yang, Y.; Yen, G.; Youngman, E.; Yu, K.; Bussey, H.; Boeke, J. D.; Snyder, M.; Philippsen, P.; Davis, R. W. and Johnston, M.

- (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387-391.
- Giaever, G.; Flaherty, P.; Kumm, J.; Proctor, M.; Nislow, C.; Jaramillo, D. F.; Chu, A. M.; Jordan, M. I.; Arkin, A. P. and Davis, R. W. (2004) Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc. Natl. Acad. Sci. USA* **101**, 793-798.
- Giaever, G.; Shoemaker, D. D.; Jones, T. W.; Liang, H.; Winzeler, E. A.; Astromoff, A. and Davis, R. W. (1999) Genomic profiling of drug sensitivities via induced haploinsufficiency. *Nat. Genet.* **21**, 278-283.
- Gottesman, M. M. (2002) Mechanisms of cancer drug resistance. *Annu. Rev. Med.* **53**, 615-627.
- Huang, Y.; Anderle, P.; Bussey, K. J.; Barbacioru, C.; Shankavaram, U.; Dai, Z.; Reinhold, W. C.; Papp, A.; Weinstein, J. N. and Sadee, W. (2004) Membrane transporters and channels: role of the transporter in cancer chemosensitivity and chemoresistance. *Cancer Res.* **64**, 4294-4301.
- Huang, Y.; Blower, P. E.; Yang, C.; Barbacioru, C.; Dai, Z.; Zhang, Y.; Xiao, J. J.; Chan, K. K. and Sadee, W. (2005a) Correlating gene expression with chemical scaffolds of cytotoxic agents: ellipticines as substrates and inhibitors of MDR1. *Pharmacogenomics J.* **5**, 112-125.
- Huang, Y.; Dai, Z.; Barbacioru, C. and Sadee, W. (2005b) Cystine-glutamate transporter SLC7A11 in cancer chemosensitivity and chemoresistance. *Cancer Res.* **65**, 7446-7454.
- Huang, Y. and Sadee, W. (2003) Drug sensitivity and resistance genes in cancer chemotherapy: a chemogenomics approach. *Drug Discov. Today* **8**, 356-363.
- Huang, Y. and Sadee, W. (2006) Membrane transporters and channels in chemoresistance and -sensitivity of tumor cells. *Cancer Lett.* **239**, 168-182.
- Kapuskesky, M.; Kemmeren, P.; Culhane, A. C.; Durinck, S.; Ihmels, J.; Komer, C.; Kull, M.; Torrente, A.; Sarkans, U.; Vilo, J. and Brazma, A. (2004) Expression Profiler: next generation--an online platform for analysis of microarray data. *Nucleic Acids Res.* **32**, W465-W470.
- Lamb, J.; Crawford, E. D.; Peck, D.; Modell, J. W.; Blat, I. C.; Wrobel, M. J.; Lerner, J.; Brunet, J. P.; Subramanian, A.; Ross, K. N.; Reich, M.; Hieronymus, H.; Wei, G.; Armstrong, S. A.; Haggarty, S. J.; Clemons, P. A.; Wei, R.; Carr, S. A.; Lander, E. S. and Golub, T. R. (2006) The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **313**, 1929-1935.
- Landowski, C. P.; Anderle, P.; Sun, D.; Sadee, W. and Amidon, G. L. (2004) Transporter and ion channel gene expression after Caco-2 cell differentiation using 2 different microarray technologies. *AAPS J.* **6**, e21.
- Lee, J. K.; Bussey, K. J.; Gwadry, F. G.; Reinhold, W.; Riddick, G.; Pelletier, S. L.; Nishizuka, S.; Szakacs, G.; Annereau, J. P.; Shankavaram, U.; Lababidi, S.; Smith, L. H.; Gottesman, M. M. and Weinstein, J. N. (2003) Comparing cDNA and oligonucleotide array data: concordance of gene expression across platforms for the NCI-60 cancer cells. *Genome Biol.* **4**, R82.
- Nishizuka, S.; Charboneau, L.; Young, L.; Major, S.; Reinhold, W. C.; Waltham, M.; Kouros-Mehr, H.; Bussey, K. J.; Lee, J. K.; Espina, V.; Munson, P. J.; Petricoin, E., 3rd; Liotta, L. A. and Weinstein, J. N. (2003a) Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc. Natl. Acad. Sci. USA* **100**, 14229-14234.
- Nishizuka, S.; Chen, S. T.; Gwadry, F. G.; Alexander, J.; Major, S. M.; Scherf, U.; Reinhold, W. C.; Waltham, M.; Charboneau, L.; Young, L.; Bussey, K. J.; Kim, S.; Lababidi, S.; Lee, J. K.; Pittaluga, S.; Scudiero, D. A.; Sausville, E. A.; Munson, P. J.; Petricoin, E. F., 3rd; Liotta, L. A.; Hewitt, S. M.; Raffeld, M. and Weinstein, J. N. (2003b) Diagnostic markers that distinguish colon and ovarian adenocarcinomas: identification by genomic, proteomic, and tissue array profiling. *Cancer Res.* **63**, 5243-5250.
- Rabow, A. A.; Shoemaker, R. H.; Sausville, E. A. and Covell, D. G. (2002) Mining the National Cancer Institute's tumor-screening database: identification of compounds with similar cellular activities. *J. Med. Chem.* **45**, 818-840.
- Reich, M.; Ohm, K.; Angelo, M.; Tamayo, P. and Mesirov, J. P. (2004) GeneCluster 2.0: an advanced toolset for bioarray analysis. *Bioinformatics* **20**, 1797-1798.
- Roschke, A. V. and Kirsch, I. R. (2005a) Targeting cancer cells by exploiting karyotypic complexity and chromosomal instability. *Cell Cycle* **4**, 679-682.
- Roschke, A. V.; Lababidi, S.; Tonon, G.; Gehlhaus, K. S.; Bussey, K.; Weinstein, J. N. and Kirsch, I. R. (2005b) Karyotypic "state" as a potential determinant for anticancer drug discovery. *Proc. Natl. Acad. Sci. USA* **102**, 2964-2969.
- Roschke, A. V.; Tonon, G.; Gehlhaus, K. S.; McTyre, N.; Bussey, K. J.; Lababidi, S.; Scudiero, D. A.; Weinstein, J. N. and Kirsch, I. R. (2003) Karyotypic complexity of the NCI-60 drug-screening panel. *Cancer Res.* **63**, 8634-8647.
- Ross, D. T.; Scherf, U.; Eisen, M. B.; Perou, C. M.; Rees, C.; Spellman, P.; Iyer, V.; Jeffrey, S. S.; Van de Rijn, M.; Waltham, M.; Pergamenschikov, A.; Lee, J. C.; Lashkari, D.; Shalon, D.; Myers, T. G.; Weinstein, J. N.; Botstein, D. and Brown, P. O. (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* **24**, 227-235.
- Salemme, F. R. (2003) Chemical genomics as an emerging paradigm for postgenomic drug discovery. *Pharmacogenomics* **4**, 257-267.
- Sarkans, U.; Parkinson, H.; Lara, G. G.; Oezcimen, A.; Sharma, A.; Abeygunawardena, N.; Contrino, S.; Holloway, E.; Rocca-Serra, P.; Mukherjee, G.; Shojatalab, M.; Kapuskesky, M.; Sansone, S. A.; Farne, A.; Rayner, T. and Brazma, A. (2005) The ArrayExpress gene expression database: a software engineering and implementation perspective. *Bioinformatics* **21**, 1495-1501.
- Scherf, U.; Ross, D. T.; Waltham, M.; Smith, L. H.; Lee, J. K.; Tanabe, L.; Kohn, K. W.; Reinhold, W. C.; Myers, T. G.; Andrews, D. T.; Scudiero, D. A.; Eisen, M. B.; Sausville, E. A.; Pommier, Y.; Botstein, D.; Brown, P. O. and Weinstein, J. N. (2000) A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.* **24**, 236-244.
- Smukste, I. and Stockwell, B. R. (2005) Advances in chemical genetics. *Annu. Rev. Genomics Hum. Genet.* **6**, 261-286.
- Stanton, J. E.; Slonim, D. K.; Collier, H. A.; Tamayo, P.; Angelo, M. J.; Park, J.; Scherf, U.; Lee, J. K.; Reinhold, W. O.; Weinstein, J. N.; Mesirov, J. P.; Lander, E. S. and Golub, T. R. (2001) Chemosensitivity prediction by transcriptional profiling. *Proc. Natl. Acad. Sci. USA* **98**, 10787-10792.
- Stegmaier, K.; Ross, K. N.; Colavito, S. A.; O'Malley, S.; Stockwell, B. R. and Golub, T. R. (2004) Gene expression-based high-throughput screening (GE-HTS) and application to leukemia differentiation. *Nat. Genet.* **36**, 257-263.
- Stockwell, B. R. (2000) Chemical genetics: ligand-based discovery of gene function. *Nat. Rev. Genet.* **1**, 116-125.
- Sturn, A.; Mlecnik, B.; Pieler, R.; Rainer, J.; Truskaller, T. and Trajanoski, Z. (2003) Client-server environment for high-performance gene expression data analysis. *Bioinformatics* **19**, 772-773.
- Szakacs, G.; Paterson, J. K.; Ludwig, J. A.; Booth-Genthe, C. and Gottesman, M. M. (2006) Targeting multidrug resistance in cancer. *Nat. Rev. Drug Discov.* **5**, 219-234.
- Wallqvist, A.; Huang, R.; Covell, D. G.; Roschke, A. V.; Gehlhaus, K. S. and Kirsch, I. R. (2005) Drugs aimed at targeting characteristic karyotypic phenotypes of cancer cells. *Mol. Cancer Ther.* **4**, 1559-1568.
- Wallqvist, A.; Rabow, A. A.; Shoemaker, R. H.; Sausville, E. A. and Covell, D. G. (2002) Establishing connections between microarray expression data and chemotherapeutic cancer pharmacology. *Mol. Cancer Ther.* **1**, 311-320.
- Wallqvist, A.; Rabow, A. A.; Shoemaker, R. H.; Sausville, E. A. and Covell, D. G. (2003) Linking the growth inhibition response from the National Cancer Institute's anticancer screen to gene expression levels and other molecular target data. *Bioinformatics* **19**, 2212-2224.
- Weinstein, J. N. (2002) 'Omics' and hypothesis-driven research in the molecular pharmacology of cancer. *Curr. Opin. Pharmacol.* **2**, 361-365.
- Weinstein, J. N.; Myers, T. G.; O'Connor, P. M.; Friend, S. H.; Fornace, A. J., Jr.; Kohn, K. W.; Fojo, T.; Bates, S. E.; Rubinstein, L. V.; Anderson, N. L.; Buolamwini, J. K.; van Osdol, W. W.; Monks, A. P.; Scudiero, D. A.; Sausville, E. A.; Zaharevitz, D. W.; Bunow, B.; Viswanadhan, V. N.; Johnson, G. S.; Wittes, R. E. and Paull, K. D. (1997) An information-intensive approach to the molecular pharmacology of cancer. *Science* **275**, 343-349.
- Weinstein, J. N. and Pommier, Y. (2003) Transcriptomic analysis of the NCI-60 cancer cell lines. *C. R. Biol.* **326**, 909-920.

- Yang, Y.; Engin, L.; Wurtele, E. S.; Cruz-Neira, C. and Dickerson, J. A. (2005) Integration of metabolic networks and gene expression in virtual reality. *Bioinformatics* **21**, 3645-3650.
- Zeeberg, B. R.; Feng, W.; Wang, G.; Wang, M. D.; Fojo, A. T.; Sunshine, M.; Narasimhan, S.; Kane, D. W.; Reinhold, W. C.; Lababidi, S.; Bussey, K. J.; Riss, J.; Barrett, J. C. and Weinstein, J. N. (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* **4**, R28.
- Zeeberg, B. R.; Qin, H.; Narasimhan, S.; Sunshine, M.; Cao, H.; Kane, D. W.; Reimers, M.; Stephens, R. M.; Bryant, D.; Burt, S. K.; Elnekave, E.; Hari, D. M.; Wynn, T. A.; Cunningham-Rundles, C.; Stewart, D. M.; Nelson, D. and Weinstein, J. N. (2005) High-throughput GoMiner, an 'industrial-strength' integrative gene ontology tool for interpretation of multiple-microarray experiments, with application to studies of common variable immune deficiency (CVID). *BMC Bioinformatics* **6**, 168.