Establishing Connections between Microarray Expression Data and Chemotherapeutic Cancer Pharmacology[1]

**Anders Wallqvist,[2] Alfred A. Rabow,
Robert H. Shoemaker, Edward A. Sausville, and
David G. Covell[2]**

Science Applications International Corporation [A. W., A. A. R.], and
Developmental Therapeutics Program, National Cancer Institute at
Frederick, NIH [R. H. S., E. A. S., D. G. C.], Frederick, Maryland 21702

Abstract

**We have investigated three different microarray
datasets of ~6 K gene expressions across the National
Cancer Institute's panel of 60 tumor cell lines. Initial
assessments of reproducibility for gene expressions
within each dataset, as derived from sequence analysis
of full-length sequences as well as expressed
sequence tags (EST), found statistically significant
results for no more than 36% of those cases where at
least one replicate of a gene appears on the array.
Filtering the data based only on pairwise comparisons
among these three datasets creates a list of ~400
significant concordant expression patterns. The
expression profiles of these smaller sets of genes were
used to locate similar expression profiles of synthetic
agents screened against these same 60 tumor cell
lines. A correspondence was found between mRNA
expression patterns and 50% growth inhibition
response patterns of screened agents for 11 cases that
were subsequently verifiable from ligand-target
crystallographic data. Notable amongst these cases
are genes encoding a variety of kinases, which were
also found to be targets of small drug-like molecules
within the database of protein structures. These 11
cases lend support to the premise that similarities
between expression patterns and chemical responses
for the National Cancer Institute's tumor panel can be
related to known cases of molecular structure and
putative cellular function. The details of the 11
verifiable cases and the concordant gene subsets are
provided. Discussions about the prospects of using
this approach as a data mining tool are included.**

Introduction

Gene expression profiling from mRNA expression microar-
rays has become a powerful tool in assessing the cellular
response in normal and cancer cells (1–3). The prevailing
viewpoint proposes a complex network of genes working
together to regulate homeostasis within the cell. The com-
plexity of this cellular network has often been underesti-
mated, because hundreds of genes may be time-depen-
dently up- or down-regulated in response to a single effector
(4, 5). Despite the complexity of data interpretation, gene
expression profiling has clear clinical applications in its ability
to subgroup tumors that cannot otherwise be differentiated
(6–9). Recently, Alizadeh *et al.* (10) were able to identify
clinically distinct types of diffuse large B-cell lymphomas
using gene expression techniques. Khan *et al.* (11) used an
artificial neural network to successfully classify clinical ex-
amples of small, round blue-cell tumors.

In this paper we examine the connection between gene
expression data across the NCI's[3] 60 tumor cell lines and
chemical screening experiments conducted on the same cell
lines to establish cytotoxicity, as measured by concentration
of agents to GI_{50} (12–14). Correlations between gene expres-
sion patterns and GI_{50} patterns across this tumor cell panel
are postulated, in our analysis, to suggest possible interac-
tions between chemical agents and either gene products or
nucleic acids. One should note that no drugs are introduced
in this set of gene expression measurements. It is only via
their correlated responses that we postulate any connections
between putative targets and chemical response.

Previous attempts to identify relationships between mo-
lecular targets and chemicals, based on expression patterns
observed in the NCI's anticancer drug screen, have been
reported with varied degrees of success (15, 16). Much of
this difficulty results from the lack of abundant gene-drug
relationships, that can also be experimentally verified, thus
also making it difficult here to fully evaluate our general
hypothesis. As an alternative it is possible to question our
general hypothesis about linkages between chemical re-
sponses and putative targets without explicit administration
of these agents (4, 17). However, our strategy does not take
into account additional concerns related to making conclu-
sions from mRNA microarray data. For example, Tamm *et al.*
(18) critique the use of mRNA as a measure of expression by
pointing to the well-known fact that post-transcriptional reg-
ulation of expression is most likely of equal importance for
the expression of some genes. Others have amplified this
viewpoint by suggesting the necessity of measuring actual
protein levels, a prevailing feeling among members of the
growing proteomics community.

[3] The abbreviations used are: NCI, National Cancer Institute; GI_{50}, 50%
growth inhibition; PDB, Protein Data Bank; STO, staurosporine; cAMP,
cyclic AMP; PDGF, platelet-derived growth factor; BDH, 3-hydroxybu-
tyrate dehydrogenase; AND, dehydroepiandrosterone; DHT, dihydrotest-
osterone; QUE, quercetin; DIA4, diaphorase 4 or menadione oxidoreduc-
tase; PTPRC, protein tyrosine phosphatase receptor type C; MMP, matrix
metalloproteinase; APOD, apolipoprotein D; ADH5, alcohol dehydrogen-
ase 5; CTSH, cathepsin H.

In this analysis we examine three gene expression datasets for coherence between related gene expression patterns. At the outset we acknowledge that much controversy has been raised around the fact that gene expression data can be of highly varying quality and that it is necessary to make repeated measurements to get a clear picture of the true expression profile (19–21). As an alternative, our analysis will treat each of these datasets as a single replicate, and based on coherence of their patterns across the 60 tumor cell lines, extract the "best" set of gene expressions amongst these experiments. In fact, the quality of these data supports only a few hundred significant gene expressions; however, based on our requirement of concordance, this smaller data set can be advanced with higher confidence for additional analysis. It is among these few hundred concordant gene expressions that we hunt for evidence of target-drug relationships based on the additional observations of GI_{50} values obtained from these same set of 60 tumor cell lines.

A limitation of this type of analysis is that verifications of postulated target-drug associations are nontrivial. Facing this criticism, we propose our findings as testable hypotheses. Our analysis infers target-drug relations by seeking similar patterns of cellular response as measured in gene space and in chemical space. A substantial amount of information already exists about this chemical space; a detailed clustering analysis has been completed that has organizes the >36 K screened compounds into slightly more than 1 K clusters; the latter of which are grouped according to mechanisms of cellular action that include biomolecular synthesis and cell cycle control (22). This analysis already provides a great deal of information about chemical activity and can be additionally surveyed for matches between gene expression profiles and chemical response profiles across the NCI's tumor cell panel. While this treatment offers only clues about these interactions, validations can be found by seeking chemically similar ligands that have been deposited as ligand-target complexes in the PDB (23). Examinations of these ligand targets for their biochemical function followed by establishing the link between these expressed proteins and mRNA expressions via sequence alignment is used here to make connections between gene expression and chemical activity.

Our analysis finds that in the relatively restricted set of concordant gene expressions, 11 verifiable gene-drug relationships are found. Previous attempts to find similar correlations have yielded, at best, only one or two such relationships (15, 16). While our 11 observations represent only a small number when compared with the total space of possible interactions, these results are encouraging by demonstrating that target-drug interactions can be extracted from diverse measurements of gene expression and GI_{50} data. The tools necessary to connect these measurements are incorporated in the world wide web.[4]

Data Treatment for Finding Significantly Differentiated Gene Expressions

Currently three publicly available datasets using different microarray technologies exist for the 60 human cancer cell lines used by the NCI Developmental Therapeutics Program to screen anticancer agents. Two datasets are referred to as the Millennium and the Weinstein/Botstein/Stanford data sets, and are freely available from the Developmental Therapeutics Program website (15).[5] The Stanford data is derived from cDNA arrays and are referenced to a pool of mRNA from 12 cell lines. The Millennium data set employed chips from Affymetrix where the relative abundance of each gene is calculated from the average difference of intensities between matching and mismatched oligonucleotides. Affymetrixs chips were also used in the third dataset, which originates from the Whitehead Institute (16).[6] The cell lines investigated using these arrays constitute cells derived from cancers of breast, central nervous system, colorectal, leukemia, lung, melanoma, ovarian, prostate, and renal systems of origin. Given that we have three sets of gene expression measurements across the same number of cell lines it is of interest to compare these experimental results. Our primary aim here is not to address quality control for measurements obtained from different technologies by different laboratories. Rather, as we will demonstrate, despite major reported difficulties regarding microarray data reproducibility, these three datasets can provide novel information regarding target-drug associations. Thus, where there may be a tendency to discount highly noisy datasets, our results reveal that meaningful information can be extracted when carefully analyzed. In order to compare the same genes in each dataset we first selected the set of ESTs within each microarray that have a pairwise common Unigene (24) designation, as derived from the sequence accession number. This yields 4791 putative common genes among the Millennium and the Stanford datasets, 4152 between the Whitehead and the Stanford datasets, and 5759 between the Millennium and the Whitehead datasets. There were 2105 genes common to all three of the datasets. The total number of genes contained in the three datasets was 5390, 9564, and 7006 for the Millennium, Stanford, and Whitehead datasets, respectively. For additional data processing we invoke standard Z-score normalization by setting the average measured value for a particular gene across all of the cell lines to zero and normalize each value by the group's standard deviation (22). The Z-score normalization was used to provide dimensionless data designation across the microarray and GI_{50} datasets. This normalization has no effect on the correlation analysis provided herein.

In order to capture similar differential gene expressions among these datasets we look to similarities between the total expression profile across the 60 cell lines. In Fig. 1 we display the normalized distribution of correlation coefficients of the data vectors for each gene that has two or more members within the same Unigene cluster but restricted to
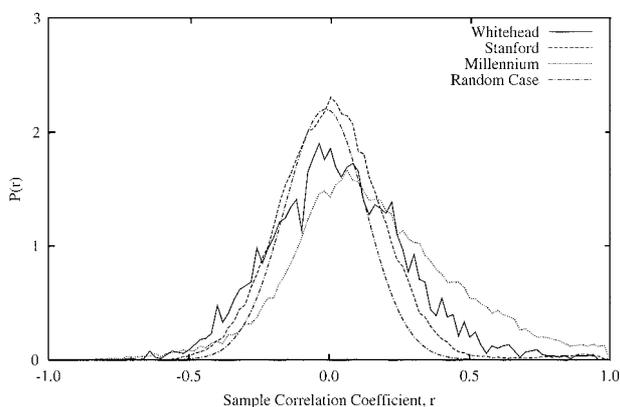

Fig. 1. The distribution of sample correlation coefficients for genes that have the same Unigene designation as compared with the randomized set. The 5% significance level threshold for a correlated measurement across the cell lines is 0.25, indicating that the bulk of all data points are not correlated with each other.
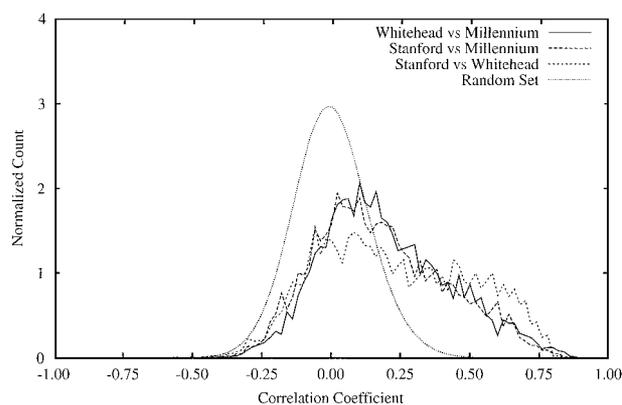

Fig. 2. The distribution of pairwise correlation coefficients between the three microarray datasets. The correlation coefficient is calculated using data vectors across all the cell lines that have the same Unigene designation. These distributions are compared with the random case where the datavectors have been scrambled. The data support, at the 5% significance level, a concordance of about 26–37% of the data vectors examined. This number is slightly higher than the concordance found between genes within a dataset in Fig. 1. The fraction of gene similarity between and within dataset are two unrelated quantities. There is nothing that says that genes that are duplicated within a dataset represent those duplicated across datasets.

the same microarray. There were 987 such genes in the Millennium data set, 869 in the Stanford set, and 520 in the Whitehead dataset. The correlation coefficient of two expression profiles $\vec{A}$ and $\vec{B}$ across the 60 cell lines was calculated from:

$$r_{\vec{A},\vec{B}} = \frac{\vec{A} \cdot \vec{B}}{|\vec{A}|\,|\vec{B}|} \qquad (1)$$

These three distributions of the set of pairwise correlation coefficients between respective microarrays are skewed towards a slightly positive correlation. At the 5% significance level, as derived from the randomized distribution, we find from numerical integration that only 11% of the measurements are correlated in the Stanford datasets compared with 19% and 36% in the Whitehead and Millennium datasets, respectively. Thus, while the bulk of the measurements within each microarray experiment appears to be noise, significant information appears to be retained above background (*i.e.* random) levels. A comparison of the gene expression profiles across the 60 tumor cell lines between the same genes (as defined by Unigene) across these datasets reveals a similar picture. Thus, in Fig. 2 we show the normalized distribution of all of the pairwise correlation coefficients of data vectors for each dataset compared with datavectors derived from a random distribution. At a 5% significance level about 26–37% of the datavectors are correlated with each other.

Our next step is to order each gene expression pattern according to the existing set of response profiles available from chemical screens across the 60 cell-line panel. The basic premise for correlating gene expression and chemical response data assumes an underlying relationship between chemical activity (GI$_{50}$) and gene expression, that when challenged with chemicals evokes a cytotoxic response. The exact nature of this response can be quite complex, most likely involving multiple targets of biochemical pathways. Regardless of mechanism, our premise is that correlations between gene expression patterns and GI$_{50}$ patterns indi-

cate, albeit crudely, linkages between chemical response and gene expression.

Our strategy for making this linkage lies in our previous analysis of the NCI's screening data. As noted earlier, we have clustered GI$_{50}$ values for ~36 K screened compounds (22) to organize this data into ~1 K clusters that represent different types of cellular activity using a self organizing map (SOM). Using this organization, the gene expression profiles are matched to similar GI$_{50}$ profiles from screened data. The matching projection is done by calculating the Euclidian distance between the data vector and all of the node vectors of the GI$_{50}$ map, and selecting the location with the minimal distance. These projections of gene expression data onto chemical response data are a means of relating each measurement according to its activity pattern across the 60 tumor cells. However, our projections are not based on the complete set of measured genes for each microarray, rather the analysis is conducted only on the concordant subset of gene expressions. Thus, the gene expression data is first filtered according to concordance, then projected onto chemical response space.

An appropriate question about gene projections on to chemical clusters is the reliability of placement. The similarity measure for map projections is their Euclidean distance. The data vector from gene expressions measured across the 60 tumor cells is placed in the cluster having the smallest Euclidean distance. We estimate the a priori probability of a chance occurrence of having two data vectors coprojecting to the same location on the GI$_{50}$ map by calculating the ratio of all of the vectors that coproject to those that do not. This yields a P of 3.8×10^{-3} for the coprojection procedure for finding significantly differentiated gene expressions in the datasets.

A total of 106 genes are identified based on pairwise filtering of the Millennium and the Stanford datasets, 175

from the Stanford and the Whitehead datasets, and 154 from the Millennium and Whitehead datasets. This gives a total of 376 unique genes that survive this filtering technique. A listing of these genes is available together with their Unigene identifier.[7] The 376 selected genes represent many types of cellular functions, although this set is dominated by genes involved in signal transduction, with the remainder falling in broadly defined categories: adhesion/extracellular matrix, tumor suppressor, ribosomal/transcriptions, immune response, melanoma, and proliferation clusters. Only 29 of them have been characterized as housekeeping genes (25).

The recent paper by Staunton *et al.* (26) attempts to identify drug-gene relationships in ways similar to ours but using a more extensive prefiltering of the datasets into cells of extreme drug (in)sensitivity. While our analysis cannot be directly compared with theirs, we find a 34% overlap between their sets of reported genes and those found by us to convey the most information in our analysis.

As examples we provide here a brief description of genes that appear to have strongly concordant gene expression profiles across these three datasets. We emphasize strongly that where these genes have similar response patterns across the 60 tumor cell lines, conclusions about these observations in regard to cell function are not addressed.

EDNRB (Hs.82002, endothelin receptor type B, cluster 16.8) or endothelin receptor B is expressed in all of the human melanoma cell lines, though metastatic melanoma expresses this receptor relatively less (27, 28). Inspection of the response pattern for the tumor cell panel reflects the high expression of EDNRB within the melanoma panel (data not shown). A similar strong pattern is observed for a smaller set of breast cancer cell lines.

FN1 (Hs.287820, Fibronectin 1 or *LETS*, cluster 7.22) is a fibronectin, which is an important class of extracellular multi-adhesive matrix proteins. As such, fibronectins are ligands to the integrin family of cell adhesion molecules and partake in the regulation of cytoskeletal organization. The strong signal for fibronectin expression has also been corroborated by previous measurements of cancer expressions profiles using a variety of alternative methods (1, 29, 30). The strong fibronectin signal within the renal panel lines is quite evident and coincides with the observations that fibronectin may be a critical factor in the regulatory role of extracellular matrix proteins in metastatic invasion of renal cancer cells (31).

LCP1 (Hs.76506, lymphocyte cytosolic protein 1 (L-plastin), cluster 23.9) is an actin regulating protein. Structural proteins like actin, may be involved in the development and progression of cancer (32). Regulation of these genes is accomplished by a number of genes, L-plastin among them. L-plastin is an actin binding protein that has tissue-specific expression patterns. L-plastin is specifically expressed in hematopoietic cells but has also been found to be highly expressed in cell lines derived from mammary solid tumors. Dysregulation of actin-binding proteins during carcinogenesis may, thus, be the direct link between the observed upregulation of L-plastin in the cancer cell lines, although the

exact role or L-plastin in the tumor process remains unknown (33). Upregulation of L-plastin has been linked to testosterone in breast and prostate cancer cells (34). This observation might suggest a corresponding subpanel sensitivity to testosterone. The expression profile of L-plastin is strongest within the leukemia and breast cancer panels, near a region on our anticancer map demonstrated to have sensitivity to selected steroid molecules, NSCs 624018 and 633664.

MCAM (Hs.211579, melanoma adhesion molecule, *MUC18*, cluster 15.7) is a transmembrane glycoprotein and is a member of the immunoglobulin superfamily. The protein is closely related to a number of cell adhesion molecules. Tumor progression and metastasis in human malignant melanoma is associated with *MCAM*. Consistent with this expression pattern we observe enhanced expression activities in the melanoma panel.

S100P (Hs.2962, S100 calcium-binding protein P, cluster 8.10) is a low molecular weight calcium-binding protein, which is associated with the regulation of cellular processes such as cell cycle progression and differentiation. Overexpression of *S100P* has been postulated to play an important role in the immortalization of human epithelial cell *in vitro* and in tumor progression *in vivo* (35). Other S100 calcium binding proteins are also found to be correlated among these three expression datasets. These include the *S100A4* gene ($r = 0.70$, $P < 0.01$), whereas the *S100B* gene expression is only weakly correlated ($r = 0.27$, $P < 0.01$). *S100P* is downregulated after androgen deprivation in an androgen-responsive prostate cancer cell line (36). As in the L-plastin case described above, the gene expression profile of *S100P* is most similar to a region on our anticancer map that is sensitive to steroid molecules, NSC 689621 and 652123.

It is important to note that previous analysis of portions of these datasets have also identified L-plastin and *S100P* as important genes. The methods used in these reports were considerably more complicated that the simple filtering method proposed here.

Identifying Molecules That Affect Expression Levels
Relating expression patterns from cell lines to a chemical response represents an important validation step, usually involving considerable biochemical effort. As an alternative we seek verifications based on surveys of the PDB structural library of ligand complexes. The steps taken in this verification are outlined in Fig. 3. The procedure for relating chemical response and gene expression levels is to first identify the set of proteins within the PDB that are homologous to the genes that are coprojected on the GI_{50} map. The entire projected gene datasets that have a homologous PDB sequence spans the entire map. The number of unique genes on the microarray that have at least one ligand-bound PDB homolog is 1231, 1197, and 1523 for the Millennium, Stanford, and Whitehead datasets, respectively. The average coverage of the complete SOM map for these expressed genes is 71%. The imposed concordance criteria narrows the selection to the most likely gene/drug associations. Methods of protein sequence alignments were performed using FASTA, version 3.3, with standard gap-parameters and the BLOSUM50 similarity matrix (37). The next step is to determine whether there

---

[7] Internet address: http://octagon.ncifcrf.gov/~wallqvis/gene.data.html.

Microarray data

⬇

Select common genes
from different datasets
that project to the
same spot on GI50 map

⬇

Find PDB protein
homologs to the
projected genes

⬇

Compare PDB ligand
with the compounds
on the projected spot
on the GI50 map

⬇

If the chemical similarity between a
compound found on the GI50 map
and the homologous PDB protein
ligand is high we have a verification
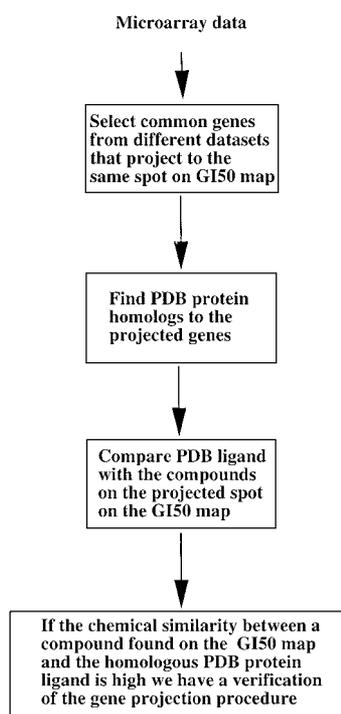of the gene projection procedure

Fig. 3. Flow diagram of the verification steps described in the text. Thus, we begin with the concordant dataset from two microarray measurements that have their datavectors projected to the same location on the GI_{50} map. These genes are searched for homologous proteins in the PDB that are cocrystallized with a ligand. The final step is to verify that those compounds associated with each projected node on the GI_{50} map are chemically similar to the PDB ligand. Thus, if the projected genes have a similar function to the protein in the PDB and the PDB ligand is chemically similar to a compound in the projected spot we have a verification of the gene expression profile association with a GI_{50} measurement.

is a structural match between the ligand bound in the PDB structure and screened NSC compounds. The complete set of ligands within the PDB database were extracted by scanning for heteroatom records, and any fully or partially present ligands were collated as a PDB ligand. This includes ligands that are associated with only DNA records as well as modified residues that are covalently attached to other residues. This collation provides a sample of possible ligand/protein and ligand/DNA interactions. Small ions and unsuitable metal ligands were deselected from this list, leaving a total of 1919 PDB small molecule ligands suitable for structural comparisons to the screened NSC compounds.

In order to describe the chemical similarity of the PDB ligand to the compounds deposited in the NCI database we use a bit-vector assignment to describe each molecule. This is an electronically convenient way of describing a molecule in order to catch the flavor of its possible interactions. In such a description the molecule is dissected for properties that are coded in an on/off bit, *e.g.* presence or absence of aromatic fragments, carboxyl groups, hydrogen-bond donor, and so forth. We have used the properties defined by the regular E-screen bit-vectors (38), which encodes 431 bits. We then use the Tanimoto coefficient as measure to identify compounds containing similar chemical elements or fragments

via a bit-vector similarity, defined as the number of bits in common divided by the total number of bits. This is calculated for bit-vectors $\vec{A}$ and $\vec{B}$ as:

$$T(\vec{A}, \vec{B}) = \frac{\vec{A} \cdot \vec{B}}{|\vec{A}| + |\vec{B}| - \vec{A} \cdot \vec{B}} \qquad (2)$$

The Tanimoto coefficient is a measure of the number of common substructures shared by two molecules as described by this bit-vector mask. In this work the E-screen bit-vector mask was generated and used as a similarity measure between NSC compounds and PDB ligands. Although a high similarity does not guarantee that two compounds will behave the same in a biological screen, high structural similarity can be used to identify structural binding motifs of similar compounds bound to a protein target (39, 40).

Similarities between molecules are thus measured via the Tanimoto coefficient of a discrete bit-vector of length 431 for each compound. This Tanimoto coefficient identifies common molecular fragments between two compared molecules and ranges from 0 to 1. In this case we have used a cutoff of 0.75 as being of significant similarity (41). Thus, if we find a similar ligand in the PDB we query the parent structure for its function, and if its function is similar to that of the original query gene we consider evidence for verification of a target-drug association. Because the number of ligands in the PDB is rather modest, we cannot expect to verify each individually selected significant gene; instead we use this process to verify the basic premise of similarity between gene expression and drug response.

In order to estimate the joint occurrence of a coprojection and the chance occurrence that a PDB ligand has a Tanimoto score >0.75 with a NSC compound we calculate the ratio of all of the PDB ligand: NSC compound pairs that have such a Tanimoto coefficient to those that do not. This yields a P of 8×10^{-4}. The a priori probability of a joint occurrence of these two events is then the product of these to probabilities and yields a final P of 3×10^{-6} for the procedure.

We have used a pairwise comparison strategy to extract information from the three gene expression datasets. Our analysis finds evidence for 11 putative chemical-gene relationships. While this number represents a low percentage of the total number of concordant genes, the remaining not-yet-verifiable genes represent the subject of future investigations into their potential chemical-gene relationships.

Genes and Chemicals Connected via Cellular Profiles

The connection between gene expression and GI_{50} values is investigated for each pairwise concordant set of microarray experiments, Millennium-Whitehead, Millennium-Stanford, and Whitehead-Stanford. Genes identified by our procedure as having a corresponding PDB ligand molecule are listed in Table 1, and their cross-reference to PDB and NSC ligands is listed in Table 2. Fig. 4 provides structural representations of the NSC compounds and their analogous PDB ligand structure. Note that these structural conformations are arbitrary and might not represent the actual bound conformation. The following sections briefly discuss the 11 verifiable cases.

Table 1 Gene designations

Cross-reference of genes with Unigene annotations via sequence accession numbers to the GC numbering system of EST adopted by the Millennium and Stanford dataset as well as the accession numbers given in the Whitehead dataset. The cluster on the cancer map[22] to which these gene expression datavectors project are also listed in this table.

Gene	Unigene	Millennium	Stanford	Cluster
CAMK1	Hs.184402	GC22306	GC18090	10.23
MAP2K4	Hs.75217	GC25430	GC14360	12.15
PDGFRA	Hs.74615	GC25705	GC10098	19.1
BDH	Hs.76893	GC20746	GC15388	27.15
PRKCB1	Hs.77202	GC23444	GC14583	27.6
PRKCB1	Hs.77202	GC23445	GC14583	27.6
		Millennium	Whitehead	
DIA4	Hs.80706	GC21197	J03934	15.10
PTPRC	Hs.17012	GC25403	Y00062	25.6
MMP1	Hs.83169	GC26040	X54925	7.17
APOD	Hs.75736	GC19985	J02611	8.17
		Stanford	Whitehead	
ADH5	Hs.78989	GC18296	M81118	10.14
CTSH	Hs.76476	GC10087	X16832	13.6

The gene expression profile of *CAMK1* (calcium/calmodulin-dependent protein kinase I) is most similar to the GI_{50} profiles observed for a set of chemicals that includes NSC compound 618487, which is identical to the PDB ligand STO shown in Fig. 4*A*. Thus, a direct connection is established between the gene expression of *CAMK1* and the chemical response of STO against its PDB kinase target. In this instance both the tested ligand and the crystallized ligand are identical, to leave little doubt about this connection. The phosphorylation and inhibition of *CAMK1* by other calcium-dependent kinases also makes it a likely candidate to be involved in modulating the balance between cAMP- and Ca^{2+}-dependent signal transduction pathways (42). *CAMK1* is homologous to several other kinase proteins in the PDB database: 1STC, 1BXG, and 1CKP, which contain either the catalytic subunit of the cAMP-dependent protein kinase α or a human cyclin-dependent kinase 2. Our analysis permits only speculations about the potential binding of STO to these other kinase molecules. Examination of the cellular profiles finds the renal panel to be most sensitive to the STO. Surveys for PDB proteins homologous to *CAMK1* find an α-catalytic subunit of a cAMP-dependent protein kinase (1STC). This observation is significant, because 1STC also shares homology with the *MAP2K4* sequence. Mitogen-activated protein kinase pathways are signal transduction cascades with distinct functions in mammals. *MAP2K4* kinase is a potent physiologic activator of the stress-activated protein kinases. 1STC is bound by a ligand having structural similarity to NSC compound 645327 shown in Fig. 4*A*. Although this compound and STO are chemically quite different, both compounds display some structural similarity in their fused ring systems that might suggest a common pharmacophore and cellular activity.

PDGFRA (platelet-derived growth factor receptor, α polypeptide, *PDGFR2*) is a membrane-spanning growth factor receptor with tyrosine kinase activity. Overexpression of the *PDGFRA* subcomponent in the *PDGF* signaling system has been implicated in the development and malignant pro-

Table 2 Gene/PDB linkage

Cross-references of gene data with the PDB and the underlying NSC compounds, and its associated PDB ligand. The expectation value of the alignment of the GenBank EST from the Unigene clustering of ESTs with the PDB sequence is given in the fourth column. The Tanimoto coefficient between the NSC compound and the PDB ligand is given in the last column. The PDB ligands associated with the protein structures are STO, ANP 5′-adenyly-imido-triphosphate, DHT, AND 3-beta-hydroxy-5-androsten-17-one, QUE 3,5,7,3′,4′-pentahydroxyflavone (quercetin), VK3 menadione, OBA 2-(oxalyl-amino)-benzoic acid, PLH methylamino-phenylanalyl-leucyl-hydroxamate, AZE axerophthene, IU5 iso-ursodeoxycholic acid and ALD carbobenzyloxylleucinyl-leucinyl-leucinal.

Gene	EST	PDB	E-value	NSC	Lig	Tan
CAMK1[a]	L41816	1STC	1.1e-27	618487	STO	1.00
MAP2K4[a]	L36870	1STC	4.4e-09	645327	STO	0.75
PDGFRA[a]	M21574	2SRC	1.4e-25	672971	ANP	0.93
BDH[a]	M93107	1DHT	1.3e-13	92227	DHT	0.86
BDH[a]	M93107	3DHE	1.3e-13	92227	AND	0.83
PRKCB1[a]	X07109	2HCK	8.2e-07	169517	QUE	0.82
DIA4[b]	J03934	2QR2	1.0e-44	11897	VK3	0.89
DIA4[b]	J03934	2QR2	1.0e-44	651207	VK3	0.76
PTPRC[b]	Y00062	1C85	8.1e-35	635526	OBA	0.79
MMP1[b]	X05231	1MNC	1.4e-41	672675	PLH	0.76
APOD[b]	J02611	1FEN	1.6e-06	122759	AZE	0.75
ADH5[c]	M30471	1DDA	2.9e-100	49452	IU5	0.75
CTSH[c]	X16832	1BP4	1.5e-24	679678	ALD	0.84

[a] Millennium and Stanford dataset.
[b] Millennium and Whitehead dataset.
[c] Stanford and Whitehead dataset.

gression of diffuse gliomas (43). From their similarities in cellular response profiles, we identify NSC compound 672971 as a candidate ligand based on its structural similarity to the PDB ligand ANP 5′-adenyly-imido-triphosphate shown in Fig. 4*A*, which is bound to the crystal structure 2SRC, a human tyrosine-protein kinase c-src. This gene-drug association links the tyrosine-kinases together with a ligand binding motif similar to ATP, a natural substrate of kinases.

BDH is a lipid-requiring mitochondrial membrane enzyme with an absolute and specific requirement for phosphatidylcholine, which acts as an allosteric activator of BDH enzymatic activity (44). Its gene expression profile links it to two dehydrogenases in the PDB, 3DHE and 1DHT via their ligand similarity to the steroid-like NSC compound 92227. The 3DHE deposition contains estrogenic 17-β hydroxysteroid dehydrogenase complexed with the ligand AND, while 1DHT is the same protein but complexed to DHT. The corresponding similarities of these compounds are shown in Fig. 4*A*. The similarities of their ligands allows us to make a tentative gene-chemical connection for the *BDH* gene and the steroid compound NSC 92227. Bailly *et al.* (45) showed that the gene expression of this mitochondrial enzyme is modulated throughout developmental changes in hormonal and metabolic conditions, especially via corticosterone and estradiol.

PRKCB1 plays an important role in B-cell activation and may be functionally linked to a tyrosine kinase in antigen receptor-mediated signal transduction (46). Berns *et al.* (47) report that *PRKCB1* also functions in angiogenesis and cancer growth. Our analysis finds a structural link between protein kinase C and the PDB structure 2HCK, which contains a
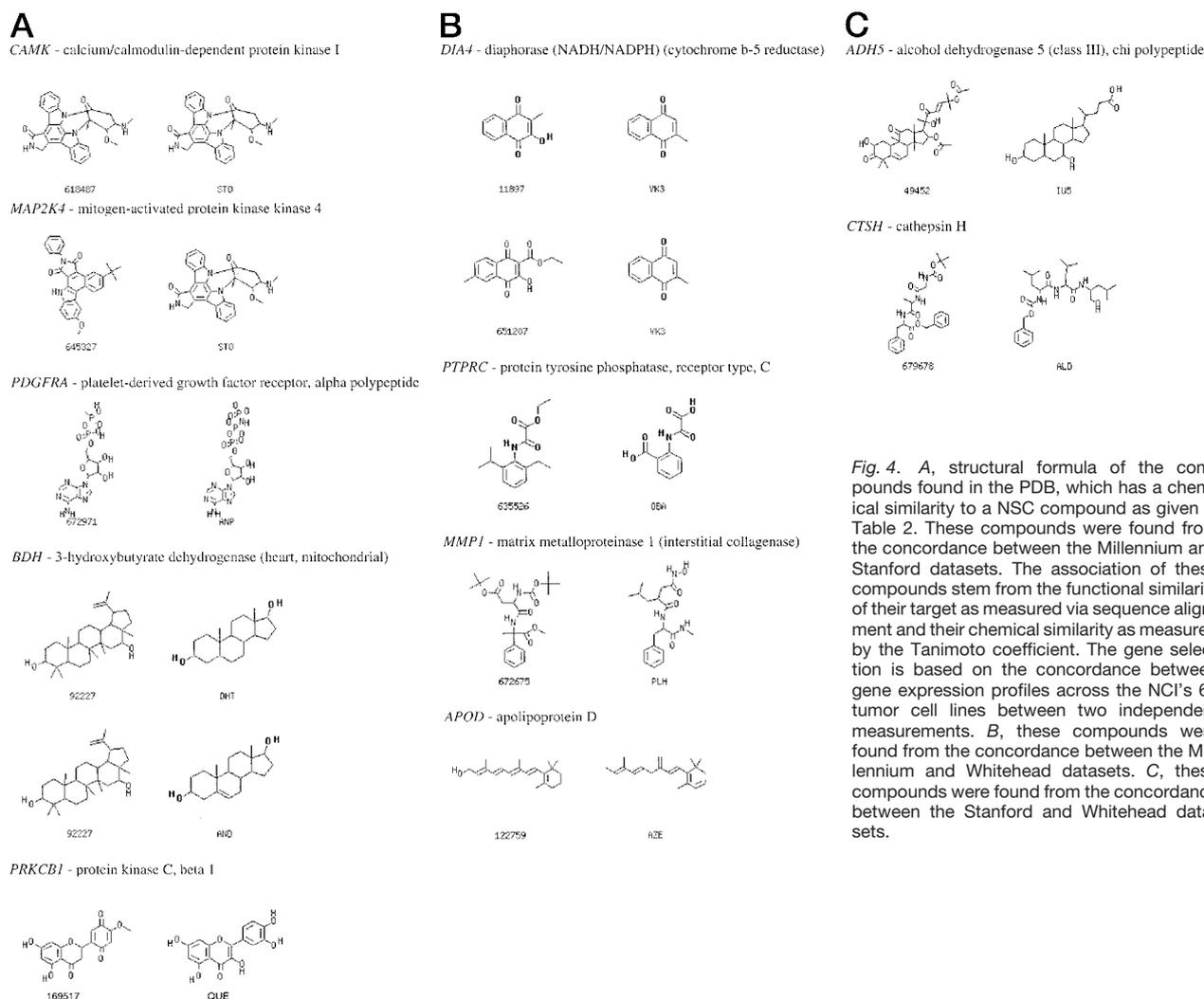
A
CAMK - calcium/calmodulin-dependent protein kinase I

618487   STO

MAP2K4 - mitogen-activated protein kinase kinase 4

645327   STO

PDGFRA - platelet-derived growth factor receptor, alpha polypeptide

672971   HNP

BDH - 3-hydroxybutyrate dehydrogenase (heart, mitochondrial)

92227   DHT

92227   AND

PRKCB1 - protein kinase C, beta 1

169517   QUE

B
DIA4 - diaphorase (NADH/NADPH) (cytochrome b-5 reductase)

11897   VK3

651207   VK3

PTPRC - protein tyrosine phosphatase, receptor type, C

635526   OBA

MMP1 - matrix metalloproteinase 1 (interstitial collagenase)

672675   PLH

APOD - apolipoprotein D

122759   AZE

C
ADH5 - alcohol dehydrogenase 5 (class III), chi polypeptide

49452   IU5

CTSH - cathepsin H

679678   ALD

Fig. 4. *A*, structural formula of the compounds found in the PDB, which has a chemical similarity to a NSC compound as given in Table 2. These compounds were found from the concordance between the Millennium and Stanford datasets. The association of these compounds stem from the functional similarity of their target as measured via sequence alignment and their chemical similarity as measured by the Tanimoto coefficient. The gene selection is based on the concordance between gene expression profiles across the NCI's 60 tumor cell lines between two independent measurements. *B*, these compounds were found from the concordance between the Millennium and Whitehead datasets. *C*, these compounds were found from the concordance between the Stanford and Whitehead datasets.

src family kinase, hck. The similarities in cellular profiles of the gene expression of *PRKCB1* and the GI_{50} response pattern to the compound QUE are consistent with the sequence similarities of its target protein, and structural similarities between NSC compound 169517 and the kinase hck-bound ligand QUE.

DIA4 is part of the detoxification process of quinones derived from the oxidation of benzene metabolites. Diaphorase can also activate bioreductive anticancer drugs. Downregulation of diaphorase has been shown to induce gastric cancer in certain cell lines (48). Menadione is present in the PDB as a ligand to the human quinone reductase type 2, and two analogous NSC compounds 11897 and 651207 are found to have similar gene expression profiles. Their structural similarity is given in Fig. 4*B*.

PTPRC is a major high molecular weight leukocyte cell surface molecule, and it functions as a membrane-bound protein tyrosine phosphatase. It is required for efficient lymphocyte signaling and plays an important role in the human immune system. Its gene expression profile is highly correlated with the cellular profile induced by NSC compound

635526. This molecule is analogous to the PDB ligand OBA in PDB deposition 1C85, which contains the structure of a protein tyrosine phosphatase 1B. The similarity between the ligand and the NSC compound shown in Fig. 4*B*, coupled with the closely matched gene/protein function, clearly establishes their gene/chemical relationship.

MMP1 is a matrix metalloproteinase that helps to break down interstitial collagen. Overexpression of *MMP1* in tumor cells is indicative of the invasive nature of cancer. In the PDB there exists a structure of the catalytic domain of the metalloprotease neutrophil collagenase. The inhibitor bound to this enzyme is PLH, which shares common structural elements with the NSC compound 672675 in Fig. 4*B*. The gene expression profile projects to the nearest neighbor of the cluster containing this compound, providing a tentative link between chemical agent and gene.

APOD is a member of the α (2 μ)-microglobulin superfamily of carrier proteins termed lipocalins. It shares a high degree of homology to retinol-binding protein. This homology allows us to assign the PDB structure 1FEN as possessing similarities with the *APOD* gene. The axerophthene ligand

is closely analogous to the NSC compound 122759 and shares its cellular profile with the gene expression for APOD. The strong similarity between the ligand and the NSC compound in Fig. 4*B* is evidence for a tentative gene/drug relationship between retinol-like molecules and lipocalins.

ADH5 (class III), χ polypeptide is a protein of which the specific function in humans is largely unknown. There exists a highly homologous protein model in the 1DDA PDB deposition, which is an ADH complexed with isoursodeoxycholic acid, a steroid. An analogous NSC steroid compound shown in Fig. 4*C*, 49452, is found to have a strongly similar gene expression profile, indicating a tentative relationship between these two data profiles.

CTSH belongs to a class of cystein-dependent intracellular proteases. The cathepsins have an important function in regulating intracellular protein degradation. The up-regulation of cathepsin gene transcription appears to be characteristic for invasive tumor cells (49). In the structural deposition of 1BP4, papain has been used as a model to test cathepsin inhibitors. The PDB ligand carbobenzyloxylleucinyl-leucinyl-leucinal shares structural similarity to the NSC compound 679678 as shown in Fig. 4*C*, providing a link between the protease functions and the activity of structurally similar ligands that may bind cathepsin.

Conclusion

Computational tools aimed at improving our understanding of chemotherapeutic cancer pharmacology can also aid the drug discovery process. In this paper we present an analysis that links chemical response space via GI_{50} measurements to a set of expression profiles of specific gene targets. Because the cellular environment in which these drugs act is very complex it is advantageous to derive an understanding about what genes are affected by which compounds. Our computational tools are not specific enough to pinpoint all of the possible chemical/gene interactions, but they do serve to provide initial hints about which processes or pathways might be affected, either directly or indirectly. Information of this type provides valuable insight into cause and effect, examinations that serve as the basis for additional biochemical studies.

Using a methodology that seeks similarities in cellular response patterns derived from gene expression measurements and chemical screens, connections between gene and chemical space can be made. Our procedure is grounded in the premise that these similarities in cellular response represent associations between gene products and chemical activity. We additionally verify this association by identifying small structurally similar compounds that imply a putative connection to chemotherapeutic cancer pharmacology. These latter relationships are verified here for 11 test cases. Although not emphasized in this work these measurement also allow us to differentiate gene/chemical responses based on different cell lines and, thus, also on clinically different cancer types. This may aid the identification of drugs that are specific for certain types of cancers and provide a tool for focusing efforts in the drug discovery process.

Different methods for identifying gene-chemical associations have been proposed by Butte *et al.* (16) and by Scherf *et al.* (15), who also describe the paucity of verifiable connections possible from this same dataset; the former case revealing 1 and the latter case another of the 11 associations reported here. The difference between our approach and theirs is the use of multiple datasets as surrogate replicate measurements of the same data, then filtering these data based on concordant response patterns and finally verifying our gene-chemical relationships by seeking actual structural cases. We find, with reasonably high confidence, assignments of gene-drug relationship for 11 verifiable cases, comprising drug binding to a variety of targets. Known kinase effector molecules taken from the PDB were positively correlated with their corresponding genes and NSC compounds based on similarities in their cellular response profiles. Likewise the *BDH* gene was found to be projected to a cluster on our WEB-accessible anticancer map with known steroid activity, which could be verified by the corresponding hydroxysteroid dehydrogenase ligand and structure in the PDB archive. None of these 11 associations appears to be spurious, although this cannot be ruled out without additional biochemical investigations of each specific system.

The wealth of data accompanying the post-genomic era offers high promise for understanding cellular processes and deriving strategies to affect these systems. Harvesting this information will not be simple. As our investigation reveals, this data can be quite noisy, but when confronted with data of poor quality, additional computational efforts can be utilized that lead to the extraction of meaningful information. These results are not unanticipated, given that these analyses involve quite large amounts of data that are collected from extremely complex biological systems. Additional complications related to this system are that these measurements are made on somewhat artificial cell lines and not real tumors (1, 8, 50), the GI_{50} experiments are single valued measurements of a highly complex system, and that only a subset of all the genes in the cell are represented on the microarray chip. Strategies to overcome these criticisms will be devised. Our approach offers one solution by exploring chemical and genetic links that in most cases cannot be easily verified by other means than the route taken here. This strategy does offer hope, by revealing a small set of gene/drug linkages that can be additionally exploited as possible novel data in the search for new chemotherapeutic strategies.

References

1. Zhang, L., Zhou, W., Velculescu, V. E., Kern, S. E., Hruban, R. H., Hamilton, S. R., Vogelstein, B., and Kinzler, K. W. Gene expression profiles in normal and cancer cells. Science (Wash. DC), *276:* 1268–1271, 1997.

2. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science (Wash. DC), *286:* 531–537, 1999.

3. 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. F., Laskhari, D., Shalon, D., Myers, T. G., Weinstein, J. N., Botstein, D., and Brown, P. O. Systematic variation in gene expression patterns in human cancer cell lines. Nat. Genet., *24:* 227–235, 2000.

4. Wilson, M., DeRisi, J., Kristensen, H-H., Imboden, P., Rane, S., Brown, P. O., and Schoolnik, G. K. Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. Proc. Natl. Acad. Sci. USA, *96:* 12833–12838, 1999.

5. Liu, T-X., Zhang, J-W., Tao, J., Zhang, R-B., Zhang, Q-H., Zhoa, C-J., Tong, J-H., Lanotte, M., Waxmani, S., Chen, S-J., Mao, M., Hu, G-X., Zhua, L., and Chen, Z. Gene expression networks underlying retinoic acid-induced differentiation of acute promyelocytic leukemia cells. Blood, *96:* 1496–1504, 2000.

6. Perou, C. M., Jeffrey, S. S., van de Rijn, M., Rees, C. A., Eisen, M. B., Ross, D. T., Pergamenchikov, A., Williams, C. F., Zhu, S. X., Lee, J. C. F., Lashkari, D., Shalon, S., Brown, P. O., and Botstein, D. Distinctive gene expression patterns in human mammary epithelian cells and breast cancers. Proc. Natl. Acad. Sci. USA, *96:* 9212–9217, 1999.

7. Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., Sampas, N., Dougherty, E., Wang, E., Marincola, F., Gooden, C., Lueders, J., Glatfelter, A., Pollock, P., Carpten, J., Gillanders, E., Leja, D., Dietrich, K., Beaudry, C., Berens, M., Albets, D., Sondak, V., Hayward, N., and Trent, J. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature (Lond.), *406:* 536–540, 2000.

8. Hough, C. D., Sherman-Baust, C. A., Pizer, E. S., Montz, F. J., Im, D. D., Rosenshein, N. B., Cho, K. R., Riggins, G. J., and Morin, P. J. Large-scale serial analysis of gene expression reveals genes differentially expressed in ovarian cancer. Cancer Res., *60:* 6281–6287, 2000.

9. Perou, C. M., Sorlie, T., Eisen, M. B., van der Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lonning, P. E., Borresen-Dale, A-L., Brown, P. O., and Botstein, D. Molecular portraits of human breast tumours. Nature (Lond.), *406:* 747–752, 2000.

10. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., Powell, J. I., Yang, L., Marti, G. E., Moore, T., Hudson, J., Jr., Lu, L., Lewis, D. B., Tibshirani, R., Sherlock, G., Chan, W. C., Greiner, T. C., Weisenburger, D. E., Armitage, J. O., Warnke, R., Levy, R., Wilson, W., Grever, M. R., Byrd, J. C., Botstein, D., Brown, P. O., and Staudt, L. M. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature (Lond.), *403:* 503–511, 2000.

11. Khan, J., Wen, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C., Peterson, C., and Meltzer, P. S. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. Nat. Med., *7:* 673–679, 2001.

12. Boyd, M. R. The NCI *in vitro* anticancer drug discovery screen. *In:* B. Teicher (ed.), Anticancer Drug Development Guide: Preclinical Screening, Clinical Trials and Approval, Pp. 23–41. Totowa, NJ: Humana Press, 1995.

13. Boyd, M., and Paull, K. D. Some practical considerations and applications of the National Cancer Institute *in vitro* anticancer drug discovery screen. Drug Dev. Res., *34:* 91–109, 1995.

14. 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. R., Anderson, N. L., Buolamwini, K. 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. An information-intensive approach to the molecular pharmacology of cancer. Science (Wash. DC), *275:* 343–349, 1997.

15. 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. A gene expression database for the molecular pharmacology of cancer. Nat. Genet., *24:* 236–244, 2000.

16. Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R., and Kohane, I. S. Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. Proc. Natl. Acad. Sci. USA, *97:* 12182–12186, 2000.

17. Debouck, C., and Goodfellow, P. N. DNA microarrays in drug discovery and development. Nat. Genet., *21* (Suppl.)*:* 48–50, 1999.

18. Tamm, I., Kornblau, S. M., Segall, H., Krajewski, S., Welsh, K., Kitada, S., Scudiero, D. A., Tudor, G., Qui, Y. H., Monks, A., Andreeff, M., and Reed, J. C. Expression and prognostic significance of IAP-family genes in human cancers and myeloid leukemias. Clin. Cancer Res., *6:* 1796–1803, 2000.

19. Heyer, L. J., Kruglyak, S., and Yooseph, S. Exploring expression data: identification and analysis of coexpressed genes. Genome Res., *9:* 1106–1115, 1999.

20. Coller, H. A., Grandori, C., Tamayo, P., Colbert, T., Lander, E. S., and Eisenman, R. N. Expression analysis with oligonucleotide microarrays reveals that MYC regulated genes involved in growth, cell cycle, signaling, and adhesion. Proc. Natl. Acad. Sci. USA, *97:* 3260–3265, 2000.

21. Lee, M-L. T., Kuo, F. C., Whitmore, G. A., and Sklar, J. Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. Proc. Nat. Acad. Sci. USA, *97:* 9834–9839, 2000.

22. Rabow, A. A., Shoemaker, R. H., Sausville, E. A., and Covell, D. G. Mining the National Cancer Institute's tumor-screening database: identification of compounds with similar cellular activities. J. Med. Chem. *45:* 818–840, 2002.

23. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. The protein data bank. Nucleic Acids Res., *28:* 235–242, 2000.

24. Schuler, G. D., Boguski, M. S., Stewart, E. A., Stein, L. D., Gyapay, G., Rice, K., White, R. E., Rodriguez-Tome, P., Aggarawal, A., Bajorek, E., Bentolila, S., Birren, B. B., Butler, A., Castle, A. B., Chiannilkulchai, N., Chu, A., Clee, C., Cowles, S., Day, P. J. R., Dibling, T., Drouot, D., Dunham, I., Duprat, S., East, C., Edwards, C., Fan, J-B., Fang, N., Fizames, C., Garret, G., Green, L., Hadley, D., Harris, M., Harrison, P., Brady, S., Hicks, A., Holloway, E., Hui, L., Hussain, S., Louis-Dit-Sully, C., Ma, J., MacGilvery, A., Mader, C., Maratukulam, A., Matise, T. C., McKusick, K. B., Morissette, J., Mungall, A., Muselet, D., Nusbaum, H. C., Page, D. C., Peck, A., Perkins, S., Piercy, M., Qin, F., Quackenbush, J., Ranby, S., Reif, T., Rozen, S., Sanders, C., She, Z., Silva, J., Slonim, D. K., Soderlund, C., Sun, W-L., Tabar, P., Thangarajah, T., Vega-Czarny, N., Vollrath, D., Voyticky, S., Wilmer, T., Wu, X., Adams, M. D., Auffray, D., Walter, N. A. R., Brandon, R., Deheija, A., Goodfellow, P. N., Houlgatte, R., Hudson, J. R., Jr., Ide, S. E., Iorio, K. R., Lee, W. Y., Seki, N., Nagase, T., Ishikawa, K., Nomura, N., Phillips, C., Polymeropoulos, M. H., Sandusky, M., Schmitt, K., Berry, R., Swanson, K., Torres, R., Venter, J. C., Sikela, J. M., Beckmann, J. S., Weissenbach, J., Myers, R. M., Cox, D. R., James, M. R., Bentley, D., Deloukas, P., Landers, E. S., and Hudson, T. J. A gene map of the human genome. Science (Wash. DC), *274:* 540–546, 1996.

25. Hsiao, L-L., Dangond, F., Yoshida, T., Hong, R., Jensen, R. V., Misra, J., Dillon, W., Lee, K. F., Clark, K. E., Haverty, P., Weng, Z., Mutter, G. F., Frosch, M. P., MacDonald, M. E., Milford, E. I., Crum, C. P., Bueno, R., Pratt, R. E., Mahadevappa, M., Warrington, J. A., Stephanopoulos, G., Stephanopoulos, G., and Gullans, S. R. A compendium of gene expression in normal human tissue. Physiol. Genomics, *7:* 97–104, 2001.

26. Staunton, J. E., Slonim, D. K., Coller, 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. Chemosensitivity prediction by transcriptional profiling. Proc. Nat. Acad. Sci. USA, *98:* 10787–10792, 2001.

27. Lahav, R., Heffner, C., and Patterson, P. H. An endothelin receptor B antagonist inhibits growth and induces cell death in human melanoma cells *in vitro* and *in vivo*. Proc. Natl. Acad. Sci. USA, *96:* 11496–11500, 1999.

28. Kikuchi, K., Nakagawa, H., Kadono, T., Etoh, T., Byers, H. R., Mihm, M. C., Jr., and Tamaki, K. Decreased ETB receptor expression in human metastatic melanoma cells. Biochem. Biophys. Res. Commun., *219:* 734–739, 1996.

29. Gress, T. M., Wallrapp, C., Frohme, M., Muller-Pillash, F., Lacher, U., Friess, H., Buchler, M., Adler, G., and Hoheisel, J. D. Identification of genes with specific expression in pancreatic cancer by cDNA representational difference analysis. Genes Chromosomes Cancer, *19:* 97–103, 1997.

30. Clark, E. A., Golub, T. R., Lander, E. S., and Hynes, R. O. Genomic analysis of metastasis reveals an essential role for RhoC. Nature (Lond.), *406:* 532–535, 2000.

31. Brenner, W., Gross, S., Steinbach, F., Horn, S., Hohenfellner, R., and Thuroff, J. W. Differential inhibition of renal cancer cell invasion mediated by fibronectin, collagen IV and laminin. Cancer Lett., *155:* 199–205, 2000.

32. Leavitt, J. G., Bushar, T., Kakunaga, T., Hamada, H., Hirakawa, T., Goldman, D., and Merril, C. Variations in expression of mutant β-actin accompanying incremental increases in human fibroblast tumorigenicity. Cell, *28:* 259–268, 1982.

33. Lapillonne, A., Coue, O., Friederich, E., Nicolas, A., del Maestro, L., Louvards, D., Robine, S., and Sastre-Garau, X. Expression patterns of L-plastin isoform in normal and carcinomatous breast tissues. Anticancer Res., *20:* 3177–3182, 2000.

34. Lin, C-S., Lau, A., Yeh, C-C., Chang, C-H., and Lue, T. F. Upregulation of L-plastin gene by testosterone in breast and prostate cancer cells: Identification of three cooperative androgen receptor-binding sequences. DNA Cell Biol., *19:* 1–7, 2000.

35. Guerreiro da Silva, I. D. C., Hu, Y. F., Russo, I. H., Ao, X., Salicioni, A. M., Yang, X., and Russo, J. S100P calcium-binding protein overexpression is associated with immortalization of human breast epithelial cells *in vitro* and early stages of breast cancer development *in vivo*. Int. J. Oncol., *16:* 231–240, 2000.

36. Averboukh, L., Liang, P., Kantoff, P. W., and Pardee, A. B. Regulation of S100P expression by androgen. Prostate, *29:* 350–355, 1996.

37. Pearson, W. R., and Lipman, D. J. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA, *85:* 2444–2448, 1988.

38. Ihlenfeldt, W-D., Takahashi, Y., Abe, H., and Sasaki, S. Computational and management of chemical properties in CACTVS: an extensible networked approach towards modularity and compatibility. J. Chem. Inf. Comput. Sci., *34:* 109–116, 1994.

39. Covell, D. G., Jernigan, R. L., and Wallqvist, A. Structural analysis of inhibitor binding to HIV-1 protease: identification of a common binding motif. J. Molecular Structure (Theochem), *423:* 93–100, 1998.

40. Villoutreix, B. O., Härdig, Y., Wallqvist, A., Covell, D. G., de Frutos, P. G., and Dahlbäck, B. Structural investigation of C4b-binding protein by molecular modeling: localization of putative binding sites. Protein Sci., *31:* 391–405, 1998.

41. Voigt, J. H., Bienfait, B., Wang, S., and Nicklaus, M. C. Comparison of the NCI open database with seven large chemical structural database. J. Chem. Inf. Comput. Sci., *41:* 702–712, 2001.

42. Matsushita, M., and Nairn, A. C. Inhibition of the Ca2+/calmodulin-dependent protein kinase I cascade by cAMP-dependent protein kinase. J. Biol. Chem., *274:* 10086–10093, 1999.

43. Smith, J. S., Wang, X. Y., Qian, J., Hosek, S. M., Scheithauer, B. W., Jenkins, R. B., and James, C. D. Amplification of the platelet-derived growth factor receptor-A (PDGFRA) gene occurs in oligodendrogliomas with grade IV anaplastic features. J. Neuropathol. Exp. Neurol., *59:* 495–503, 2000.

44. Loeb-Hennard, C., and McIntyre, J. O. (R)-3-Hydroxybutyrate dehydrogenase: selective phosphatidylcholine binding by the C-terminal domain. Biochemistry, *39:* 11928–11938, 2000.

45. Bailly, A., Lone, Y. C., and Latruffe, N. Post-transcriptional analysis of rat mitochondrial D-3-hydroxybutyrate dehydrogenase control through development and physiological stages. Biol. Cell, *73:* 121–129, 1991.

46. Leitges, M., Schmedt, C., Guinamard, R., Davoust, J., Schaal, S., Stabel, S., and Tarakhovsky, A. Immunodeficiency in protein kinase C-β-deficient mice. Science (Wash. DC), *273:* 788–791, 1996.

47. Berns, H., Humar, R., Hengerer, B., Kiefer, F. N., and Battegay, E. J. RACK1 is up-regulated in angiogenesis and human carcinomas. FASEB J., *14:* 2549–2558, 2000.

48. Sagara, N., and Katoh, M. Mitomycin C resistance induced by TCF-3 overexpression in gastric cancer cell line MKN28 is associated with DT-diaphorase down-regulation. Cancer Res., *60:* 5959–5962, 2000.

49. Frohlich, E., Schlagenhauff, B., Mohrle, M., Weber, E., Klessen, C., and Rassner, G. Activity, expression, and transcription rate of the cathepsins B, D, H, and L in cutaneous malignant melanoma. Cancer (Phila.), *91:* 972–982, 2001.

50. St. Croix, B., Rago, C., Velculescu, V., Traverso, G., Romans, K. E., Montgomery, E., Lal, A., Riggins, G. J., Lengauer, C., Vogelstein, B., and Kinzler, K. W. Genes expressed in human tumor endothelium. Science (Wash. DC), *289:* 1197–1202, 2000.

Molecular Cancer Therapeutics

Anders Wallqvist, Alfred A. Rabow, Robert H. Shoemaker, et al.

Updated version	Access the most recent version of this article at: http://mct.aacrjournals.org/content/1/5/311

Cited articles	This article cites 46 articles, 20 of which you can access for free at: http://mct.aacrjournals.org/content/1/5/311.full#ref-list-1
Citing articles	This article has been cited by 5 HighWire-hosted articles. Access the articles at: http://mct.aacrjournals.org/content/1/5/311.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.
Permissions	To request permission to re-use all or part of this article, use this link http://mct.aacrjournals.org/content/1/5/311. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.