DDT Vol. 8, No. 8 April 2003

Drug sensitivity and resistance genes in cancer chemotherapy: a chemogenomics approach

Ying Huang and Wolfgang Sadée

Resistance to antineoplastic drugs represents a serious obstacle to successful cancer treatment. Genome-wide studies correlating drug response phenotypes with large DNA/tissue microarray and proteomic datasets have been performed to identify the genes and proteins involved in chemosensitivity or drug resistance. The goal is to identify a set of chemosensitivity and/or resistance genes for each drug that are predictive of treatment response. Therefore, validated pharmacogenomic biomarkers offer the potential for the selection of optimal treatment regimens for individual patients and for identifying novel therapeutic targets to overcome drug resistance.

Ying Huang
Wolfgang Sadée*
Program of Pharmacogenomics
Dept of Pharmacology
College of Medicine and Public Health
5078 Graves Hall
333 W. 10th Avenue
The Ohio State University
Columbus
OH 43210, USA
tel: +1 614 292 5593
fax: +1 614 292 7232
*e-mail: sadee-1@medctr.osu.edu

▼ Chemotherapy is an important line of defense against hematological malignancies and aggressive forms of solid tumor. However, only a limited proportion of cancer patients responds favorably to commonly used chemotherapeutic drugs, and drug efficacy varies widely among these patients. Although inappropriate treatment can result in the selection of more resistant and aggressive cells, cancer patients are often treated with a standard regimen protocol without considering individual differences in chemosensitivity; this invariably leads to accruing resistance during treatment. Previous studies have revealed several genetic factors that influence the chemosensitivity of cancer cells, including genes involved in drug uptake and secretion, drug metabolism, DNA repair and apoptosis (for review, see Ref. [1]). However, considering the effect of single factors in isolation is insufficient because chemosensitivity involves multiple interacting factors that contribute to the overall response. We will use the term chemosensitivity here to cover both sensitivity and resistance to anticancer drugs.

The availability of the human genome sequence permits a global approach to understanding complex genetic contributions to drug sensitivity. Novel and powerful technologies, including analysis of single-nucleotide polymorphisms (SNPs), DNA, protein and tissue microarrays and proteomics, provide the means for identifying the genes involved in chemosensitivity. The term chemogenomics encompasses the application of genome-wide scanning technology (measuring gene expression and resultant protein profiles) to target identification and drug discovery [2]. A schematic of the overall chemogenomic strategy is shown in Figure 1. Using this chemogenomics strategy, large integrated mRNA expression and drug potency datasets are measured in a panel of cancer tissues, thus providing correlations between cell–gene, cell–drug and drug–gene pairs. Proteomic analyses introduce additional and complementary matrices, yielding gene–protein, protein–drug and protein–cell correlations. A significant correlation for individual drug–gene and drug–protein pairs suggests the presence of chemosensitivity genes, whereas gene–cell and protein–cell matrices reveal cellular characteristics. Polymorphisms and gene dosage effects (gene and/or chromosome deletion or amplification) are also relevant to chemosensitivity and contribute to target identification, but more commonly are regarded in relation to individualized therapy – a hallmark of pharmacogenomics.

The combined use of these rich information sources could lead to improved chemotherapy by: (1) using pharmacogenomic biomarkers for predictive individualization of cancer treatment; and (2) revealing molecular targets

PII: S1359-6446(03)02654-0

for anticancer drug discovery, with the goal of overcoming or bypassing drug resistance. In this article, genomic and proteomic studies that are being used to identify chemosensitivity genes are reviewed and some of the advances and pitfalls in chemogenomics are discussed. Representative studies and their findings are summarized in Table 1. Although current studies provide a wealth of information on chemosensitivity genes, these have been poorly integrated, and clinical applications remain uncertain at present.

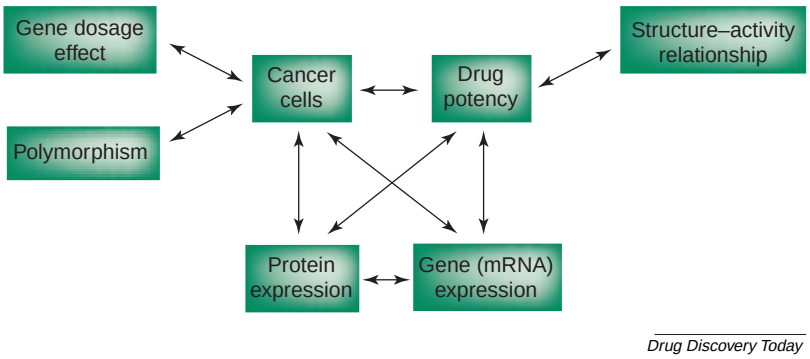

Figure 1. Principal components of chemogenomics – in search of chemosensitivity genes. Integration of multiple large mRNA expression and drug potency datasets for cancer cells provides putative relationships between cell–gene, cell–drug and drug–gene pairs. By ranking the significance of the correlation coefficients between individual gene and drug activity profiles, candidate chemosensitivity genes are identified. For an individual chemosensitivity gene, the putative anticancer drugs whose efficacy is influenced by abnormal expression of this gene in cancer cells can be identified. Proteomics introduces additional and complementary info rmation. Polymorphisms and gene dosage effects (gene/chromosome deletion or amplification) contribute additional layers of relationships relevant to chemosensitivity. Lastly, structure–activity relationships are determined for drugs clustered into functional groups by correlation analysis of the gene–drug or protein–drug matrix.

Correlation of mRNA expression data from DNA microarrays with drug response data

Microarray technology

Variations in gene expression determine the phenotype of a cancer cell and its response to anticancer drugs. High-throughput analysis of gene expression at the mRNA level using DNA microarrays has enabled the remarkable recent progress in understanding the genetic factors involved in cancer biology. The two most commonly used DNA microarrays are based on oligonucleotide and cDNA probes. Short oligonucleotide probes (~25mers), first introduced by Affymetrix (http://www.affymetrix.com), can be synthesized on the array *in situ* using photolithography [3]. The newest Affymetrix array (HG-U133) contains ~45,000 probe sets representing >39,000 transcripts derived from ~33,000 well-substantiated human genes. Alternatively, longer oligonucleotide probes (50–80 bases) can be deposited onto solid supports such as glass slides using robotic printers. cDNA microarrays are produced by robotically printing PCR products or plasmids onto glass slides [4].

A commonly used approach to identify genes involved in chemosensitivity is to correlate gene expression profiles in multiple cancer tissues with drug response. The key issues to be considered include the selectivity, precision and accuracy of the mRNA expression data obtained from microarray experiments. The results from microarrays can vary considerably in quality, especially when these rely only on single measurements for each gene. Most spotted oligonucleotide and cDNA array experiments are performed using two RNA samples (a reference and an experimental sample) that are labeled with different fluorescent dyes. Quantitation of absolute mRNA levels is problematic and thus only ratios between the two samples can be assessed. By contrast, Affymetrix arrays permit the measurement of absolute mRNA levels with the use of analysis metrics for multiple probes and controls for each transcript (perfect match and mismatch probes). However, accurate measurement of expression might vary among genes, and furthermore, comparison between Affymetrix arrays and ratio arrays is difficult. Proper normalization procedures are crucial to obtaining reliable data. Therefore, links between anticancer drugs and important chemosensitivity genes might be difficult to assess. In addition, microarrays are inherently less sensitive than other quantitative methods, such as quantitative RT–PCR, so less abundant transcripts could be missed.

Chemosensitivity studies using transformed cell lines

Scherf and coworkers have pioneered the use of cDNA microarrays for profiling chemosensitivity [5]. These researchers analyzed the gene expression patterns of a panel of 60 untreated human tumor cell lines (termed the NCI60) used at the National Cancer Institute (NCI; http://www.nci.nih.gov). Gene expression was then correlated with the growth inhibitory activity of a subset of ~70,000 compounds that had been previously tested against the NCI60 [these data can be found at the NCI development of therapeutics (DTP) website (http://dtp.nci.nih.gov)]. Increased expression of a given gene in cell lines sensitive to a given drug yields a positive correlation between gene expression and drug activity, whereas increased expression in resistant cells results in a negative correlation. The NCI study focused

Table 1. Approaches for identifying chemosensitivity genes[a]

Approach	Sample	Analytic method	Representative gene–drug pairs	Refs
DNA microarray				
cDNA array	NCI60	Correlation	*DPYD*/5-FU, *ASNS*/L-asparaginase	[5][b]
Affymetrix array	NCI60	Prediction/correlation	Keratin 8/cytochalasin D	[6][c]
Affymetrix array	NCI60	Relevance networks	*LCP1*/4-thiazolidinecarboxylic acid	[7][c]
cDNA array	NCI60	Multi-scale clustering	Caveolin-2/thaliblastine	[8]
cDNA and Affymetrix	NCI60	Correlation	*CAMK1*/STO	[9][d]
cDNA array	NCI60	Correlation	Melanoma-specific genes/ benzodithiophenedione-containing compounds	[41][e]
cDNA array	Cell lines	Correlation	*AKR1B1*/24 drugs, *BIRC5*/5-FU	[10]
cDNA array	Cell lines	Relevance networks	*TAP1*/mitoxantrone	[11]
cDNA array	Cell lines	Differential expression	*XRCC1*/doxorubicin	[13]
cDNA array	Cell lines	Differential expression	Apoptosis genes/doxorubicin	[14]
cDNA array	Cell lines	Differential expression	*MPP1*/cisplatin	[15]
cDNA array	Xenografts	Correlation	*GPX2*/CPM	[16]
cDNA array	Tissues	Prediction	*GSTA3*	[17]
cDNA array	Tissues	Correlation/prediction	*HMG1*	[18]
Tissue microarray				
Tissue array	Tissues	Differential expression	*IGFBP2*	[28]
Tissue array	Tissues	Differential expression	*ABCC2*	[32]
Proteomics				
2D-PAGE	Cell lines	Correlation	Annexin I/daunorubicin	[35]
2D-PAGE	Cell lines	Correlation	Cofilin/daunorubicin	[36]
2D-PAGE	Cell lines	Correlation	Breast cancer specific gene 1/mitoxantrone	[37]
2D-PAGE	Cell lines	Correlation	Elongation factor 1-d/etoposide	[38]
2D-PAGE	Cell lines	Correlation	HSP27/etoposide	[39]

[a]Abbreviations: *ABCC2*, the gene encoding multidrug resistance-associated protein 2; *AKR1B1*, the gene encoding aldose reductase; *ASNS*, the gene encoding asparagine synthetase; *BIRC5*, the gene encoding survivin; *CAMK1*, the gene encoding calcium/calmodulin-dependent protein kinase 1; CPM, cyclophosphamide; *DPYD*, the gene encoding dihydropyrimidine dehydrogenase; 5-FU, 5-flurouracil; *GPX2*, the gene encoding glutathione peroxidase 2; *GSTA3*, the gene encoding glutathione S-transferase A3; *HMG1*, the gene encoding high-motility group protein 1; HSP27, 27-kD heat shock protein; *IGFBP2*, the gene encoding insulin-like growth factor-binding protein 2; *LCP1*, the gene encoding lymphocyte cytosolic protein-1; *MPP1*, the gene encoding palmitoylated membrane protein 1; NCI60, 60 cell lines of the National Cancer Institute's *in vitro* anticancer drug screen; *TAP1*, the gene encoding transporter associated with antigen processing 1; 2D-PAGE, two-dimensional polyacrylamide gel electrophoresis; *XRCC1*, the gene encoding X-ray repair cross complementing protein 1.
[b]http://discover.nci.nih.gov; [c]http://www.genome.wi.mit.edu/MPR; [d]http://spheroid.ncifcrf.gov; [e]http://www.leadscope.com.

on 1376 genes (out of a total of 8000) that showed the strongest patterns of variation across the cell lines, and 118 drugs with established mechanisms of action. Two previously known drug–gene relationships revealed by the correlation analysis served to validate this approach. First, *DPYD* (dihydropyrimidine dehydrogenase) expression was negatively correlated with 5-flurouracil (5-FU) activity, which is consistent with the finding that high levels of DPYD, which is involved in the catabolism of pyrimidines and 5-FU, decrease exposure of cells to the active phosphorylated forms of 5-FU. Second, the expression of *ASNS* (asparagine synthetase) was negatively correlated with sensitivity to L-asparaginase in the NCI60. Because L-asparaginase is used in the treatment of leukemia, this correlation was tested in a sub-panel of leukemic cells, resulting in an even stronger negative correlation. This is consistent with the notion that leukemia cells lacking *ASNS* are more sensitive to L-asparaginase because it depletes extracellular L-asparagine. The data described in this study are available in a searchable public database (http://discover.nci.nih.gov). We searched the database for a relationship between *ABCB1* [encoding multidrug resistance 1 (MDR1)] expression and drug activity, and identified several known substrates of MDR1 through their significant negative correlations. Although the drug–gene matrix in the study by Scherf [5] reveals numerous putative chemosensitivity genes for

selected drugs, one must use such datasets with caution because there is considerable 'noise' associated with mRNA measurements across the NCI60. This necessitates the application of filters and stringent criteria that prevent the detection of numerous potentially significant drug–gene pairs. Nevertheless, the data provide a valuable first screen for chemoresistance genes.

Affymetrix oligonucleotide microarrays have also been used to measure expression levels of ~6800 genes in the NCI60 at the Whitehead Institute (http://www.genome.wi.mit.edu) [6,7]. The gene expression data (http://www.genome.wi.mit.edu/MPR) were similarly applied to the NCI database of cancer susceptibility to anticancer drugs. Staunton and coworkers generated a panel of genes, the expression of which was predictive of sensitivity or resistance to 232 drugs [6]. The authors suggested that for some drugs prediction of chemosensitivity in clinical samples might become feasible. However, none of the genes in the predictive panel was directly associated with known mechanisms of drug resistance, but might represent characteristics of the tumor cell, such as proliferative ability. Moreover, correlation between gene expression in the NCI60 measured by Affymetrix arrays and that measured by cDNA arrays was moderate [5]. Therefore, array results must be validated before conclusions can be drawn concerning the mechanism of the interaction and the utility for clinical applications.

Owing to the complex nature of the array data, several studies have applied advanced computational models for mining the NCI60 databases. One study used relevance networks to deduce a stringent correlation coefficient threshold for extracting significant gene–drug relationships from the noise associated with mRNA expression data in the NCI60 [7]. Another study developed a refined strategy to identify the gene–drug, gene–gene and drug–drug associations that might exist within a cell line subset, but not in the complete NCI60 set [8]. In addition, the DTP website also contains the Affymetrix microarray dataset from Millennium Pharmaceuticals (http://www.mlnm.com). Therefore, three datasets derived from different array technologies for the NCI60 panel are currently publicly available. Wallqvist and coworkers assessed the reproducibility of the gene expression data within each of the three array datasets and found statistically significant results for no more than 36% of those cases where at least one replicate of a gene appears in each dataset [9]. By focusing on only ~400 genes that showed significant concordant expression patterns, these researchers identified 11 verifiable gene–drug relationships. For example, *CAMK1* (calcium/calmodulin-dependent protein kinase 1) expression was found to be correlated with compound NSC 618487, which has a structural scaffold identical to that of staurosporine, a known CAMK1 ligand. This provides indirect support for the notion that NSC 618487 is an effector of CAMK1 and raises the confidence level that a new gene–drug relationship has been identified in this study. These procedures are useful for extracting the true information content from the databases. However, it is important to note that gene–protein networks, rather than individual genes, are responsible for chemosensitivity. Analysis of the relevant networks identified additional drug–gene correlations that were missed by hierarchical clustering of gene expression profiles versus drug potencies [7]. Novel methods are required to deal with this wealth of data and to prevent erroneous conclusions being made.

An independent study similar to that of Scherf and coworkers involved the development of an integrated chemosensitivity database of 55 anticancer drugs and the gene expression profiles of 39 human cancer cell lines [10]. This study identified several genes with expression patterns that showed significant correlation to patterns of drug responsiveness. Some of these genes were correlated with entire classes of drugs. For example, *AKR1B1* (aldose reductase) and *DDB2* (damage-specific DNA binding protein 2) showed positive correlation with multiple drugs, indicating that they are common predictive markers of chemosensitivity. They also identified genes that were correlated only with specific drugs that had similar mechanisms of action. For example, *BIRC5* (survivin) and *BIRC2* (apoptosis inhibitor 1), genes involved in apoptosis, were negatively correlated with 5-FU derivatives. Another group evaluated the relationship between chemosenstivity to eight anticancer drugs and gene expression profiles in eight human hepatoma cell lines using cDNA microarray analysis, and analyzed the data by constructing relevance networks [11]. They identified 42 genes that showed significant correlation. Nearly 20% of these represented various types of membrane transporters, most of which were negatively correlated with chemosensitivity. For example, *TAP1* (transporter associated with antigen processing 1) was associated with resistance to mitoxantrone, consistent with previous reports [12].

These studies demonstrate the utility of correlating gene expression and drug activity in transformed human cell lines. Because gene expression was measured in untreated cells, these studies focused on preexisting sensitivity to drug treatment rather than on the consequences of treatment. Experiments performed in cultures of transformed cell lines suffer from several drawbacks. First, these cell lines have further evolved *in vitro* and might not reflect the tumor they were originally isolated from *in vivo*. Moreover, interactions between the stroma and tumor *in vivo* are

probably crucial parameters in chemosensitivity. Consequently, all *in vitro* correlations need to be verified mechanistically and validated *in vivo*.

Gene expression changes as a result of drug exposure

Each of the studies discussed above focused on the sensitivity of cancer cells to anticancer drugs by analyzing gene expression patterns from untreated cells. By contrast, several studies have focused on the molecular consequences of exposure to anticancer drugs or have addressed differential expression patterns between drug-sensitive cell lines and those with acquired resistance. Kudoh and colleagues [13] used cDNA microarrays to monitor the expression profiles of MCF-7 cells that were either transiently treated with doxorubicin or selected for resistance to doxorubicin. These researchers identified a set of genes with altered expression that overlap between doxorubicin-induced and -selected cells. The authors suggested that these genes represent a profile indicative of putative doxorubicin resistance. Furthermore, gene expression profiles of doxorubicin- and cisplatin-resistant cells were found to differ – as one would expect for drugs with different mechanisms of action. This approach is promising because the candidate genes are cross-validated by two distinct experiments. A similar study used cDNA microarrays to search for differentially expressed genes between a human multiple myeloma cell line and doxorubicin-selected sub-clones that express *ABCB1* and are multidrug resistant [14]. This study identified many differentially expressed genes, including *ABCB1* and genes involved in apoptotic signaling. Wittig and coworkers determined the difference in gene expression between a drug-sensitive melanoma cell line and three derived sub-lines with acquired resistance to the DNA-damaging agents cisplatin, etoposide and fotemustine [15]. They found that 110 genes were transiently or permanently deregulated in at least two resistant sub-lines, and 14 genes were differentially expressed in all three drug-resistant sub-lines. In addition, chromosomal aberrations (such as deletions and duplications) were investigated by comparative genomic hybridization and compared with gene expression, which was used as a criterion for the selection of promising candidate genes. However, the majority of genes identified in these studies have no known role in chemosensitivity. Therefore, the underlying mechanisms must be clarified, and additional testing in clinical trials is needed to assess the utility of this approach in predictive therapy.

In vivo *chemosensitivity studies*

Zembutsu and coworkers extended the method of correlation analysis to *in vivo* animal models [16]. They used a cDNA microarray representing 23,040 genes to analyze expression profiles in a panel of 85 cancer xenografts derived from nine human organs. Furthermore, the xenografts were examined for sensitivity to nine anticancer drugs and 1578 genes displaying expression levels that were correlated with chemosensitivity were identified. Owing to the large number of genes, it is difficult to assess which genes are causatively linked to chemosensitivity, and more filters need to be applied to the dataset before it can be used as a predictive tool in chemotherapy.

Drug efficacy in mouse models might differ from that in humans. To avoid this confounding factor, gene expression profiles of clinical samples were compared with drug response. The expression profiles of 20 esophageal cancer tissues from patients who were treated with the same chemotherapy regimen after tumor removal by surgery were examined using cDNA microarray analysis of 9216 genes [17]. Comparison of these expression profiles with the duration of survival identified 52 genes having expression profiles correlated with prognosis, and probably with sensitivity and/or resistance to drugs. However, these candidate genes could affect tumor aggressiveness rather than chemosensitvity, which cannot be separated in this study. Surprisingly, the correlation between *ABCB1* expression and response was moderate at best, even though the chemotherapy regimen includes MDR1 substrates. The controversy over whether MDR1 represents a clinically important chemoresistance factor remains unresolved. Moreover, it is unclear how sensitive this approach is in detecting chemosensitivity genes, which is dependent on the heterogeneity of the tumor samples, accuracy of the analytical array technology, computational models and the statistical analysis employed in filtering the primary data. No standards have been established by which the validity of such results can be readily assessed.

Sotiriou and coworkers used cDNA microarrays to study gene expression profiles obtained from fine-needle aspirations of primary breast tumors before and after systemic chemotherapy [18]. This study identified candidate genes that might distinguish tumors with complete response to chemotherapy from tumors that do not respond. However, it remains unclear whether gene expression-based predictors reflect sensitivity to treatment or more fundamental aspects of tumor cell biology that are important to disease progression [19]. Although several studies claim that clinical outcome of individual cancer patients can be predicted using gene expression profiles of primary tumors at diagnosis [20–23], further work needs to be done to confirm the role of these genes in chemosensitivity. Similar to the studies performed *in vitro*, the gene–drug correlations identified in these studies represent correlative, but not causal,

relationships. Therefore, validation of these studies at a mechanistic level will be challenging.

Drawbacks of using DNA microarrays to study chemosensitivity

Although studies applying DNA microarray technologies identify different sets of genes that might act as predictive markers for chemosensitivity, and provide a rationale for optimizing chemotherapies for individual patients, they have common drawbacks. First, gene–drug relationships described in these studies, such as DPYD–5-FU and ASNS–L-asparaginase correlations, represent only a small fraction of relationships thought to be relevant to chemotherapy. Many interesting and novel relationships are probably hidden in the data noise. Second, although DNA arrays contain thousands of genes, limited sensitivity and precision restrict the analysis to genes that show sufficiently large variations in expression across the samples analyzed. To obtain usable data from DNA array measurements, rigorous filtering based on quality control and variance is necessary to whittle the set of analyzable genes down to a small subset. We suspect that numerous genes that are important in chemosensitivity are filtered out from the data analysis in these cases. For example, only three of the 48 ABC transporters present in the human genome were examined in the Scherf study [5]. Furthermore, only one of these (*ABCB1*) yielded significant negative correlations, even though at least seven ABC transporters have been previously implicated in drug resistance [1]. Third, when different array types were used to study the same samples, for example when cDNA arrays [5] and Affymetrix Hu6800 arrays [6] were used to study the NCI60 panel, the gene–drug relationships identified in the two studies differed substantially. Fourth, significant correlations between drug activity and gene expression do not always indicate causal relationships. For example, if expression of gene A is highly correlated with that of gene B, both genes might show significant correlations with the same drug, but only one has a causal effect. Alternatively, coexpressed genes might act synergistically in affecting drug response. For example, the gene encoding vaculolar H^+-ATPase plays a role in resistance to daunorubicin, but only in cells overexpressing *ABCC1* [encoding multidrug resistance-associated protein (MRP)]. This is because daunorubicin is sequestered into intracellular vesicles by vacuolar H^+-ATPase and then extruded from the cell by MRP [24]. Therefore, novel computational methods are needed to extract this information from high-dimensional datasets. In each case, the quality of the primary array data are key to the success of large-scale correlation studies. It will be essential to standardize the methodology to permit the comparison of datasets and the integration of the knowledge gained thus far.

Application of tissue microarrays

Although DNA microarray methods have been successful in identifying candidate genes involved in chemosensitivity, gene expression does not necessarily correlate with protein level and function [25,26]. It is therefore necessary to validate the findings derived from DNA array experiments by studying the relevant protein products. To test whether candidate genes are involved in drug response and can serve as clinical markers, each gene and protein candidate must be tested against a large pool of clinical tumor specimens with known clinical outcomes. The recently developed tissue microarray approach permits the analysis of candidate genes in multiple tumor specimens in a timely and cost-efficient manner [27]. The construction of tissue microarrays involves the removal of tissue cylinders (0.6–1.0 mm in diameter) from paraffin-embedded tumors. Sliced sections from up to 1000 tumors are arrayed and analyzed simultaneously for mRNA and proteins [27]. Several studies have combined tissue microarray technology with DNA microarrays to analyze candidate genes in clinical cancer samples. For example, Bubendorf and coworkers identified differentially expressed genes between hormone-refractory and hormone-sensitive human prostate cancer xenografts by using cDNA microarray technology [28]. They validated two of the most consistently overexpressed genes, *IGFBP2* (insulin-like growth factor-binding protein 2) and *HSP27* (27-kD heat shock protein) by immunohistochemistry of tissue microarrays containing specimens of benign hyperplasias, primary prostate cancers, and hormone-refractory local recurrences. IGFBP2 was highly expressed in all of the hormone-refractory tumors, in 36% of the primary tumors and in none of the benign specimens. Overexpression of HSP27 protein was detected in 31% of the hormone-refractory tumors, in 5% of the primary tumors and in none of the benign specimens. A similar approach was applied to renal cell carcinoma [29], human gliomas [30] and thyroid carcinomas [31]. Tissue microarrays have been used to study chemoresistance conferred by *ABCC2* (MRP2), which is involved in multidrug resistance [32]. Significant expression of *MRP2* was observed in various common human tumors. MRP2 expression was highest in moderate to poorly differentiated tumors, suggesting that these tumors might have developed drug resistance to a variety of anticancer drugs. This demonstrated the value of using tissue microarray technology in the validation of novel candidates. High-throughput validation assays at the level of proteins complement the conclusions drawn from mRNA array studies.

Correlation of proteomic data with drug response phenotype

Information on the genetic basis of cancer, obtained at the level of genomic DNA and mRNA, does not accurately

predict what occurs at the protein level within the cells [33]. Proteomics has emerged as a valuable tool for identifying proteins that affect chemotherapeutic efficacy or could serve as drug targets, although it poses daunting technical challenges. Proteomics can be divided into two categories: 'expression proteomics' and 'functional proteomics'. In expression proteomics, a given tissue is analysed for differentially expressed proteins. The techniques used include two-dimensional gel electrophoresis, multidimensional chromatography and protein microarrays [34]. Functional proteomics includes the high-throughput analysis of protein–protein or protein–DNA and/or RNA interactions and post-translational modifications. The technologies used include immunoprecipitation, protein array-based approaches and the yeast two-hybrid system [34].

Sinha and coworkers used proteomics tools to study chemoresistance mechanisms in cell culture systems derived from stomach and pancreatic carcinoma, fibrosarcoma, breast cancer, colorectal cancer and melanomas [35–39]. The parental cell lines and chemoresistant subclones were cultured and analyzed for differential protein expression using high-resolution two-dimensional electrophoresis that fractionates proteins according to charge and mass. Proteins that were overexpressed in the chemoresistant cell lines were identified by mass spectrometry. Annexin I and thioredoxin were found to be overexpressed in drug resistant stomach cancer cells [35]. Proteomic approaches have also been used to understand the effects of doxorubicin on protein expression by surveying global changes in protein patterns after doxorubicin treatment in MCF-7 cells [40]. Doxorubicin caused a marked decrease in the levels of three isoforms of HSP27. However, few of the proteins identified in these studies have previously been linked to chemoresistance, and it is unknown how these proteins might cause chemoresistance. Therefore, it is necessary to validate these findings to determine whether the candidate protein is a causative factor or is one that is not functionally involved with chemosensitivity.

Structure–activity relationships identified from correlation results

Identification of a validated gene–drug or protein–drug correlation can facilitate identification of compounds that bind to the encoded protein of interest. For example, one can use the NCI60 databases to calculate correlation coefficients between *MDR1* expression and all compounds (>70,000) with measured potencies. Therefore, this provides an opportunity to study extensive structure–activity relationships (SARs). In one study, the NCI pharmacogenomic databases were exploited for developing SARs by relating gene expression patterns to ~27,000 substructures and partial chemical features (pharmacophores) [41]. This study identified two subclasses of quinones with a pattern of activity across the NCI60, correlating strongly with the expression patterns of several genes. This concept and computational approach [structure activity target (SAT)] are particularly helpful in target and lead identification. Careful analysis of the chemical features of compounds identified as correlating with the same gene provides additional confidence that the deduced relationships are valid.

Wallquist and colleagues studied mRNA expression patterns and the growth inhibitory activity of compounds in the NCI60 by focusing on genes with concordant expression and drug response patterns in multiple DNA array datasets [9]. These researchers identified 11 novel drug–gene relationships that were subsequently verifiable by searching for chemically similar ligands that have been deposited as ligand–target complexes in a protein databank (http://spheroid.ncifcrf.gov). For example, they found that some kinases were positively correlated with known kinase ligand molecules and compounds with similar structures and chemical properties. This provided a validation of the correlation approach and enabled SAR studies to address the structural features necessary for kinase–ligand interactions. This approach can be particularly valuable where one needs to identify functional and structural substrate selectivity among closely related kinases. Another study showed that a significant proportion of chemoresistance genes is represented by transporters [11]. Owing to the diverse nature of a multitude of drug transporters, information on functional drug–transporter interactions is particularly useful in pharmacological studies.

Conclusions

High-throughput chemogenomic methods have proven to be powerful tools for resolving complex genetic factors involved in chemosensitivity. This is expected to accelerate the drug discovery process because each factor could provide a novel drug target. DNA, protein and tissue microarrays all need to be used in combination with drug response phenotype, such as *in vitro* toxicity or clinical outcome in cancer patients. Furthermore, genotyping of polymorphisms in candidate genes has promise for refining drug target validation. This should aid the identification of novel genetic factors that are predictive of chemotherapy response and molecular targets for drug discovery. However, despite the promise of these technologies, there are numerous obstacles that need to be overcome. For example, the sensitivity and accuracy of the analytical methods must improve dramatically when searching for candidate genes. These methods must be standardized to permit comparisons across studies. Moreover, relationships established through

correlation analysis between drug response phenotype and genetic variations are correlative, not causal. Methods of candidate validation must be improved, with particular attention paid to the microenvironment of the cancer tissue *in vivo*. Finally, genes and proteins do not work in isolation, but function within integrated systems that are responsible for the biological properties of the cancer cell and determine chemosensitivity. Resolving the intricate interactions among proteins in these systems will require the development of novel mathematical models that describe interactions among genes and encoded proteins in quantitative terms. This will enable multi-dimensional gene regulatory networks and complex chemosensitivity phenotypes to be resolved, thus providing the necessary framework for bringing chemogenomics in drug discovery and clinical applications to fruition.

Acknowledgements

Work described in this review was supported by NIH Research Grant GM61390 and the Plasma Membrane Transporter group at the University of California at San Francisco (UCSF).

References

1 Gottesman, M.M. *et al.* (2002) Multidrug resistance in cancer: role of ATP-dependent transporters. *Nat. Rev. Cancer* 2, 48–58
2 Bleicher, K.H. (2002) Chemogenomics: bridging a drug discovery gap. *Curr. Med. Chem*. 9, 2077–2084
3 Lipshutz, R.J. *et al.* (1999) High density synthetic oligonucleotide arrays. *Nat. Genet.* 21 (Suppl. 1), 20–24
4 DeRisi, J. *et al.* (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* 14, 457–460
5 Scherf, U. *et al.* (2000) A gene expression database for the molecular pharmacology of cancer. *Nat. Genet.* 24, 236–244
6 Staunton, J.E. *et al.* (2001) Chemosensitivity prediction by transcriptional profiling. *Proc. Natl. Acad. Sci. U. S. A.* 98, 10787–10792
7 Butte, A.J. *et al.* (2000) Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc. Natl. Acad. Sci. U. S. A.* 97, 12182–12186
8 Bao, L. *et al.* (2002) Mining functional relationships in feature subspaces from gene expression profiles and drug activity profiles. *FEBS Lett*. 516, 113–118
9 Wallqvist, A. *et al.* (2002) Establishing connections between microarray expression data and chemotherapeutic cancer pharmacology. *Mol. Cancer Ther.* 1, 311–320
10 Dan, S. *et al.* (2002) An integrated database of chemosensitivity to 55 anticancer drugs and gene expression profiles of 39 human cancer cell lines. *Cancer Res.* 62, 1139–1147
11 Moriyama, M. *et al.* (2003) Relevance network between chemosensitivity and transcriptome in human hepatoma cells. *Mol. Cancer Ther.* 2, 199–205
12 Lage, H. *et al* (2001) Enhanced expression of human ABC-transporter tap is associated with cellular resistance to mitoxantrone. *FEBS lett.* 503, 179–184
13 Kudoh, K. *et al.* (2000) Monitoring the expression profiles of doxorubicin-induced and doxorubicin-resistant cancer cells by cDNA microarray. *Cancer Res.* 60, 4161–4166
14 Watts, G.S. *et al.* (2001) cDNA microarray analysis of multidrug resistance: doxorubicin selection produces multiple defects in apoptosis signaling pathways. *J. Pharmacol. Exp. Ther.* 299, 434–441
15 Wittig, R. *et al.* (2002) Candidate genes for cross-resistance against DNA-damaging drugs. *Cancer Res.* 62, 6698–6705
16 Zembutsu, H. *et al.* (2002) Genome-wide cDNA microarray screening to correlate gene expression profiles with sensitivity of 85 human cancer xenografts to anticancer drugs. *Cancer Res.* 62, 518–527
17 Kihara, C. *et al.* (2001) Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. *Cancer Res.* 61, 6474–6479
18 Sotiriou, C. *et al.* (2002) Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. *Breast Cancer Res.* 4, 1–8
19 Ramaswamy, S. *et al.* (2003) A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* 33, 49–54
20 van' t Veer, L.J. *et al.* (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530–536
21 Alizadeh, A.A. *et al.* (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511
22 Beer, D.G. *et al.* (2002) Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* 8, 816–824
23 Rosenwald, A. *et al.* (2002) The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *New Engl. J. Med.* 346, 1937–1947
24 Benderra, Z. *et al.* (1998) Role of the vacuolar H+-ATPase in daunorubicin distribution in etoposide-resistant MCF7 cells overexpressing the multidrug-resistance associated protein. *Int. J. Oncol.* 12, 711–715
25 Anderson, L. and Seilhamer, J. (1997) A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18, 533–537
26 Gygi, S.P. *et al.* (1999) Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* 19, 1720–1730
27 Kononen, J. *et al.* (1998) Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat. Med.* 4, 844–847
28 Bubendorf, L. *et al.* (1999) Hormone therapy failure in human prostate cancer: analysis by complementary DNA and tissue microarrays. *J. Natl. Cancer Inst.* 91, 1758–1764
29 Moch, H. *et al.* (1999) High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. *Am. J. Pathol.* 154, 981–986
30 Sallinen, S.L. *et al.* (2000) Identification of differentially expressed genes in human gliomas by DNA microarray and tissue chip techniques. *Cancer Res.* 60, 6617–6622
31 Huang, Y. *et al.* (2001) Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. *Proc. Natl. Acad. Sci. U. S. A.* 98, 15044–15049
32 Sandusky, G.E. *et al.* (2002) Expression of multidrug resistance-associated protein 2 (MRP2) in normal human tissues and carcinomas using tissue microarrays. *Histopathology* 41, 65–74
33 Verma, M. *et al.* (2001) Proteomic approaches within the NCI early detection research network for the discovery and identification of cancer biomarkers. *Ann. New York Acad. Sci.* 945, 103–115
34 Wu, W. *et al.* (2002) Proteomics in cancer research. *Int. J. Gynecol. Cancer* 12, 409–423
35 Sinha, P. *et al.* (1998) Increased expression of annexin I and thioredoxin detected by two-dimensional gel electrophoresis of drug resistant human stomach cancer cells. *J. Biochem. Biophys. Methods* 37, 105–116
36 Sinha, P. *et al.* (1999) Increased expression of epidermal fatty acid binding protein, cofilin, and 14-3-3-sigma (stratifin) detected by two-dimensional gel electrophoresis, mass spectrometry and microsequencing of drug-resistant human adenocarcinoma of the pancreas. *Electrophoresis* 20, 2952–2960
37 Sinha, P. *et al.* (1999) Search for novel proteins involved in the development of chemoresistance in colorectal cancer and fibrosarcoma cells *in vitro* using two-dimensional electrophoresis, mass spectrometry and microsequencing. *Electrophoresis* 20, 2961–2969
38 Sinha, P. *et al.* (2000) Identification of novel proteins associated with the development of chemoresistance in malignant melanoma using two-dimensional electrophoresis. *Electrophoresis* 21, 3048–3057
39 Poland, J. *et al.* (2002) Study of therapy resistance in cancer cells with functional proteome analysis. *Clin. Chem. Lab. Med.* 40, 221–234
40 Chen, S.T. *et al.* (2002) Proteomics reveals protein profile changes in doxorubicin-treated MCF-7 human breast cancer cells. *Cancer Lett.* 181, 95–107
41 Blower, P.E. *et al.* (2002) Pharmacogenomic analysis: correlating molecular substructure classes with microarray gene expression data. *Pharmacogenomics J* 2, 259–271