ELSEVIER

European Journal of Cancer 40 (2004) 2544–2549

European Journal of Cancer

www.ejconline.com

Transcription profiling of gene expression in drug discovery and development: the NCI experience

Edward A. Sausville, Susan L. Holbeck *

Developmental Therapeutics Program, National Cancer Institute, Information Technology Branch, 6130 Executive Blvd., Room 8014, Bethesda, MD 20892-7444, USA

Received 4 August 2004; received in revised form 9 August 2004; accepted 10 August 2004
Available online 15 September 2004

Abstract

Transcript profiling, using microarray or other analogous technologies, to query on a large-scale the expression of genes in tumours or their derivative cell lines has numerous potential uses in oncology drug discovery and development. Characterisation of genes expressed in tumours may allow tumours to be separated into subsets defining subtypes that have a distinctive pathway utilisation. The molecular entities comprising the pathways which distinguish one disease subset from another then become potential candidate drug targets. Alternatively, gene expression patterns may be correlated with the degree of antiproliferative effect of candidate drug leads. This can reveal aspects of the drug's action that could serve to provide a further basis for benchmarking the generation of analogues or provide important information about pathways potentially modulated by the drug in achieving cytotoxicity. New information is emerging that the expression of drug transport-related molecules is a major variable that can be usefully explored using gene expression data, and the features promoting successful drug handling by the tumour cell may be an additional variable which can be illuminated by gene expression studies.
Published by Elsevier Ltd.

Keywords: Microarray; Technologies; Transcript profiling; Drug candidates; Gene expression; Targets

The advent of large-scale profiling of gene expression using gene chip technology has offered numerous opportunities for drug discovery and development, from which only limited achievements have been realised to date. For example, characterisation of gene expression in distinct subsets of the same disease allows the definition of distinct molecular phenotypes. This is exemplified by the studies of Staudt and colleagues [1] in diffuse large B cell lymphomas, which documented a germinal centre type of gene expression, with a relatively favourable prognosis following therapy with standard chemotherapeutic agents. In addition, gene expression profiling documented the existence of an activated B cell type of lymphoma with evidence of expression of genes responsive to nuclear factor κB (NFκB) signalling that had a very poor prognosis [2]. The activation of this pathway therefore becomes a target of interest in developing therapies directed at this subtype of diffuse large B cell lymphoma. Subsequent examples of this type of approach have been exemplified by the studies of Sorlie and colleagues [3] in breast cancer documenting a basal epithelial-like group, an erbB2-overexpressing group, and a "normal" breast-like group, and of Beer and colleagues [4] in lung cancer. The targets identified by these analyses become the basis for drug discovery strategies.

A complementary strategy is to consider how subsets of a given disease that do or do not respond to "standard" therapies differ in their gene expression profile. For example, Shipp and colleagues [5] retrospectively catalogued gene expression profiles in diffuse aggressive

* Corresponding author. Tel.: +1 301 435 9178; fax: +1 301 480 4808.
E-mail address: holbecks@mail.nih.gov (S.L. Holbeck).

0959-8049/$ - see front matter. Published by Elsevier Ltd.
doi:10.1016/j.ejca.2004.08.006

lymphoma patient specimens from individuals who did or did not respond favourably to standard therapy. A limited set of genes emerged in the unfavourable set, and these might legitimately be considered targets for the evolution of therapies useful in the unfavourable subset of patients. Analogous thinking has been applied to breast cancer by van de Vijver and colleagues [6], again with the evolution of a limited set of genes whose expression portends a poor response to therapy.

To consider how studies of this type suggest concrete discovery strategies, we will consider in somewhat greater detail the studies of Beer and colleagues [4], who analysed gene expression in samples from patients with lung cancer and were able to identify genes that classified the tumours by stage, as well as identifying genes associated with a poor prognosis. One of the genes identified in the Beer and colleagues [4] study is LAMB1, encoding laminin, β1, an extracellular matrix protein involved in adhesion, migration and metastasis. A possible use of the National Cancer Institute (NCI) screening data is to leverage these types of studies into discovery of drugs that might potentially be useful in treatment of the poor prognosis group. We asked whether we could identify Food and Drug Administration (FDA)-approved anticancer compounds that had been tested in the NCI screen whose pattern of sensitivity was correlated with expression of LAMB1 (GC32475), as measured in the Novartis Affymetrix U95A dataset described later in this article. Utilising the entire panel of 59 cell lines, a significant correlation (Pearson correlation coefficient (PCC) 0.57, *P*-value 0.0000024) was found with sensitivity to bleomycin (NSC 125066). Looking at just the subset of nine lung cancer lines within the screening panel, a positive correlation was still observed, although no longer significant with this small number of cell lines (PCC 0.481, *P*-value 0.19). Numerous groups have used bleomycin sensitivity of peripheral blood lymphocytes as a surrogate for identifying individuals with increased sensitivity to environmental mutagens. Those with increased sensitivity to bleomycin have an increased risk of developing lung cancer. The correlation we observed may arise if this group of individuals has an enhanced mutagenic response to DNA damage. This information therefore raises the question of whether LAMB1 production in some way reflects a pro-mutagenic pathway that could be the basis for intervention as a prevention target for lung cancer.

From the standpoint of drug discovery, compounds may be picked *a priori* which alter gene expression in ways that suggest the ability to regulate novel pathways. This type of application was applied very early in the utilisation of the technology to analyses of compounds that altered the expression of the yeast transcriptome [7,8], with the idea of both screening for desired compound effects, as well as in detecting early in a compound's development cycle, the possibility of "off target" effects. Marton and colleagues [8] compared the patterns obtained following compound treatment with those obtained from yeast strains deleted for single genes, with the thesis that compounds inhibiting the activity of a gene product would have a similar pattern of gene expression to cells genetically lacking this gene.

Analysis of changes in signature disease-specific genes has been the 'cornerstone' of the approach developed by various commercial parties after exposure of cultured mammalian cells to novel compounds. While these approaches are certainly of interest in delineating novel approaches to pathway modulation, as yet, it has had limited impact in selecting new leads relevant to oncology drug development.

An alternative approach is to correlate sensitivity to drug effect, usually through some measure of anti-proliferative action, with gene expression profiles prior to drug exposure, with the hypothesis that the pattern of gene expression would correlate with those genes that critically enable or protect against the activity of the drug. Since 1990, the United States (US) has maintained an *in vitro* anti-cancer drug screen, whose operation has been described previously [9,10] and which features the profiling of compound action in approximately 60 different cell lines selected to represent common adult solid tumour types plus a more limited repertoire of haematological neoplasms. Computational techniques have been devised which allow the pattern of cellular response to a new agent to be compared with the patterns of previously screened compounds. The initial approach, called the COMPARE algorithm [11], calculates the similarity of patterns of drug susceptibility utilising the Pearson correlation coefficient (PCC). The complementary Mean Graph facilitates visual inspection of the patterns by plotting the deflection of a particular cell line from the average growth inhibition susceptibility of all cell lines. Very early in the operation of NCI *in vitro* drug screen, it became apparent that the COMPARE algorithm could define compounds with similar mechanisms of action, as they could be identified by high PCCs due to similar patterns of cellular response. For example, the NCI *in vitro* drug screen easily defines agents with anti-microtubule or DNA damage-directed mechanisms of action secondary to topoisomerase or anti-metabolite activity. Subsequent algorithms have been applied to the drug screen data, including neural net and self-organising map informatics approaches to provide alternative filters that recognise distinct relationships between patterns of compound action [12,13].

NCI's major use of this computational database is to recognise compounds with novel mechanisms of action in comparison to previously studied classes, and to discern relationships with molecular target expression that might suggest the mechanism of action of novel compounds. COMPARE results can suggest potentially fruitful avenues of research upon which to focus. For

example, several novel structural classes giving strong correlations with activity of tubulin agents were subsequently confirmed to have this mechanism of action [14,15]. The activity of a novel group of proteasome inhibitors tested in the NCI tumour cell line screen was found not to correlate with known classes of compounds, supporting the view that they were a mechanistically novel class [16]. These related compounds correlated well with one another, with their relative potency in the cell line screen mirroring their potency in inhibition of proteasome activity in a biochemical assay [17]. Screening data for publicly disclosed compounds is available through www.dtp.nci.nih.gov, which also contains information as to how academic and commercial organisations, both in the US and internationally, can enter into confidentiality agreements to permit the evaluation of non-publicly disclosed compound's at no cost to the originating parties.

An extension of the reasoning that led to the COMPARE computational algorithm, which allows recognition of a compound's mechanism of action by similarity of its pattern of cytotoxicity to compounds with known mechanisms, led to the possibility that the pattern of cellular susceptibility to a compound might be correlated not only with the pattern of action of other compounds, but with the patterns of molecular target expression. Lee and colleagues [18] were the first to take this approach, using rhodamine efflux taken as a measure of *mdr* activity to define novel chemotypes acting as modulators of *mdr* drug efflux pump function. This observation led to the idea that the pattern of molecular target expression could define agents that correlated directly with the target's expression (by convention greater drug sensitivity correlating with greater target expression) or inversely with target expression (drug resistance correlating with greater degrees of target expression).

This result led to the effort to characterise the expression of molecular targets in the cell lines comprising the NCI *in vitro* drug screen. Efforts to measure the enzymatic activity of drug metabolising enzymes, the presence of p53 and *ras* mutations, and how these might be used in drug discovery campaigns have been described (summarised in [9]). Information pertaining to the expression of molecular targets in the NCI 60 cell *in vitro* anti-cancer drug screen is also available through www.dtp.nci.nih.gov/mtargets/mt_index.html.

Among the approaches used to analyse target expression was the use of microarray profiling to define the patterns of gene expression in the different cell lines. The initial microarray effort used nearly 10 000 partial-to full-length cDNA probes printed on glass slides [19,20]. These studies utilised cDNA clones generated by the IMAGE consortium, and analysed expression in each cell line relative to that in a reference sample of RNA. This reference was generated by pooling RNA from 12 of the cell lines, to ensure the best chance that this pool would contain most RNAs. There are a couple of caveats in utilisation of this data. First, the cDNAs printed on the arrays were generated by polymerase chain reaction (PCR) from the 96-well plates supplied by the IMAGE consortium. Subsequently, it was found that a significant number of the wells contained multiple clones, confounding the identification of these genes [19]. A second caution is that the data is not absolute gene expression, but expression relative to a reference sample, so large ratios can arise even for poorly expressed genes. Even in this early analysis, the potential of this approach for drug discovery was demonstrated by the observation that 5-fluorouracil action was inversely correlated with the level of dihydropyrimidine dehydrogenase, and the enzymes of asparagine biosynthesis likewise identified cell lines with susceptibility to L-asparaginase, as shown in Fig. 1.

Subsequent microarray characterisations of the cell lines in the NCI *in vitro* drug screen utilised distinct methodologies to array the probe sets and to interpret the information. Millenium Pharmaceuticals employed an early version of the Affymetrix oligonucleotide chips (HUM6000) and software to analyse the expression of roughly 7000 genes. More recently, Novartis analysed expression of approximately 12000 features in the NCI cell line panel, using Affymetrix U95A chips. This latter analysis was performed in triplicate, generating an extremely robust dataset of gene expression. The oligonucleotide arrays have the advantage that absolute levels of gene expression are measured. For the U95A chips, many genes are represented by several features, increasing one's confidence in correlations involving these genes.

Examples of how this data has contributed to the evaluation of novel compounds in the screen include the observation that diethyldithiocarbamate, a prototypic inhibitor of NFκB [21], actually correlates directly with the expression of the type-2 interleukin (IL)-1 receptor and a sphingomyelinase. It is clear now that activation of the IL-1 receptor actually leads to signalling through phospholipase C to modulate sphingomyelinase, as well as the degradation of the IκB protein to activate NFκB signalling. Thus, if one had no idea of how diethyldithiocarbamate functioned, consideration of the expression array data would have pointed to members of the IL-1 signalling pathway, including NFκB.

Additional examples where considering the pattern of expressed genes correlated with the activity of novel compounds include the identification by Wosikowski and colleagues [22] of novel compounds affecting erbB2 signalling, including novel chemotypes directly inhibiting erbB2 kinase activity. Dick and colleagues [23] measured activity of NADPH alkenal/one oxidoreductase in the NCI drug screen panel and demonstrated that this correlated with sensitivity to irofulven. Algeciras-Schim-

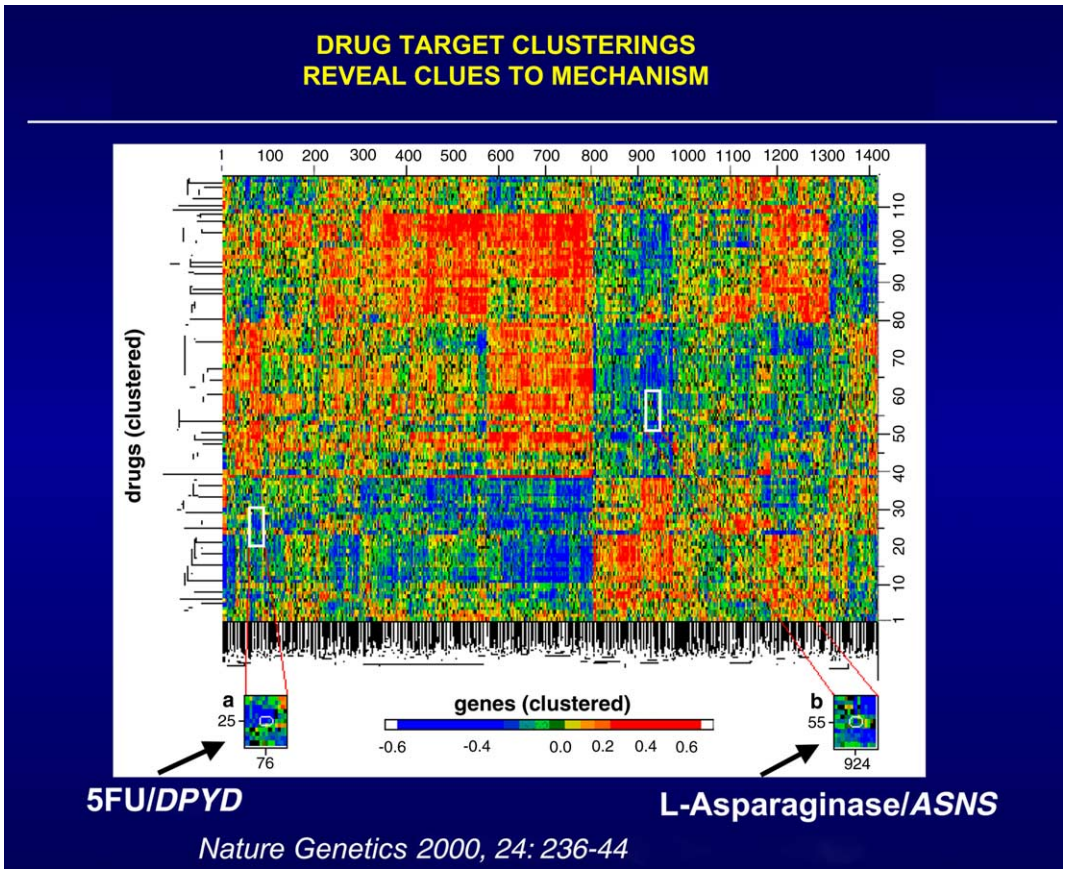

Fig. 1. Clustered image map relating activity patterns of 118 tested compounds to the expression patterns of 1,376 genes in the 60 cell lines. Included, in addition to the gene expression levels, are data for 40 molecular targets assessed one at a time in the cells. A red point (high positive Pearson correlation coefficient (PCC)) indicates that the agent tends to be more active (in the two-day SRB assay) against cell lines that express more of the gene; a blue point (high negative correlation) indicates the opposite tendency. Genes were cluster-ordered on the basis of their correlations with drugs (mean-subtracted, average-linkage clustered with correlation metric); drugs were clustered on the basis of their correlations with genes (mean-subtracted, average-linkage clustered with correlation metric). Inset (a) shows a magnified view of the region around the point (white circle) representing the correlation between *DPYD* and 5-fluorouracil (5-FU). Inset (b) is an analogous magnified view for *ASNS* and the drug L-asparaginase. Originally published in Scherf U, Ross DT, Waltham M *et al.* A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000, 24, 236-44. Used with kind permission of the Nature Publishing Group (www.nature.com/ng/).

nich and colleagues [24] examined the sensitivity of the cell line panel to killing by a soluble form of CD95 ligand, which separates the panel into two classes. These classes were found to largely correspond to the broad clusters of gene expression distinguishing mesenchymal from epithelial cells identified in the cDNA microarrays described by Ross and colleagues [19].

One of the major determinants of drug sensitivity is the ability of a compound to enter the cell, or to be pumped out of the cell. Huang and colleagues [25] utilised a transporter-specific microarray to analyse the expression of approximately 700 transporter and channel genes in the NCI drug screen cell lines. They found that cells with higher expression of the equilibrative nucleoside transporter (ENT, SLC29A1) were more sensitive to several nucleoside analogues. The role of this gene in the transport of these compounds was confirmed by showing decreased sensitivity to these nucleosides following treatment with an inhibitor of the transporter. We analysed the Novartis Affymetrix U95A expression data described earlier for correlations with sensitivity of the cell lines to FDA-approved anticancer compounds. Sensitivity of the cell lines to all-*trans* retinoic acid (ATRA) was higher in cells with higher expression of ABCA4 (with a PCC of 0.69, *P*-value 1.1×10^{-9}), a retinal-specific ATP-binding cassette transporter known to be involved in retinoid transport in the retina [26].

Ideally, drug screening algorithms of the future will use a "blend" of bioinformatic approaches. Where the atomic structure of the intended target is available through physical methodologies such as X-ray crystallography or nuclear magnetic resonance (NMR) spectroscopy, candidate chemotypes might be suggested by the energetics of various potential docking sites. Screens of "focused" libraries of compounds constructed after

consideration of this information would reveal candidate "hits" with evidence of binding to or effect on target function. At the same time, consideration of the gene expression profile of cell types preferentially responsive to the anti-proliferative potential of the compound would provide information about sets of pathways whose presence denotes sensitivity to the drug candidate at an early stage of its evaluation. These data could then provide important clues as to pathways engaged by the molecule, and mechanisms for its transport and metabolic handling, as well as assurance that additional mechanisms are not relevant to the compound's action. Subsequent iterations of the compound structure through early elaboration of the hit molecule toward the generation of a *bona fide* lead would alternate between the structural and functional screens.

Where the antiproliferative activity of a potential anticancer lead is recognised, e.g., from a complex natural product-derived mixture, consideration of effects on gene expression after compound exposure, or on the pattern of expressed genes correlating with compound activity may likewise provide important clues to the mechanism of the compound's action. This could be done in conjunction with efforts to define high affinity binding partners in the intracellular environment by affinity chromatography and proteomic approaches. In this manner, timely integration of gene expression or transcriptional profiling into a drug discovery and development programme contributes greatly to the "smartness" and efficiency of the process, and can provide fascinating insights into compound action that greatly increase the informational package available to decision-makers at key decision points in the compound's life cycle.

In summary, transcriptional profiling of gene expression in tumour cells has provided a wealth of information. Although some may take the point of view that there is actually too much information with no "validated" algorithm to sift through the meaning of the data, we are just at the beginning of refining how the information is to be used. It might be reasonably argued that we actually are coming to a period in which the "growing pains" of the initial use of the methodology, in which everything from how to quality control the data to the development of suitable algorithms to interpret the information, was being road-tested. We now can use this experience to build discovery programmes where the best use of gene expression data is to serve as a hypothesis generator, leading to the likelihood that the most focused set of relevant hypotheses can be quickly queried and useful outcomes incorporated into the molecule's development strategy.

Conflict of interest statement

None declared.

References

1. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, *et al*. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000, **403**(6769), 503–511.
2. Davis RE, Brown KD, Siebenlist U, Staudt LM. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. *J Exp Med* 2001, **194**(12), 1861–1874.
3. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, *et al*. Gene expression patterns of breast carcinomas distinguish tumour subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001, **98**(19), 10869–10874.
4. Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, *et al*. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002, **8**(8), 816–824.
5. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, *et al*. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* 2002, **8**(1), 68–74.
6. van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, *et al*. A gene-expression signature as a predictor of survival in breast cancer. *New Engl J Med* 2002, **347**(25), 1999–2009.
7. Gray NS, Wodicka L, Thunnissen AM, Norman TC, Kwon S, Espinoza FH, *et al*. Exploiting chemical libraries, structure, and genomics in the search for kinase inhibitors. *Science* 1998, **281**(5376), 533–538.
8. Marton MJ, DeRisi JL, Bennett HA, Iyer VR, Meyer MR, Roberts CJ, *et al*. Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat Med* 1998, **4**(11), 1293–1301.
9. Holbeck SL. Update on NCI in vitro drug screen utilities. *Eur J Cancer* 2004, **40**(6), 785–793.
10. Scudiero DA, Shoemaker RH, Paull KD, Monks A, Tierney S, Nofziger TH, *et al*. Evaluation of a soluble tetrazolium/formazan assay for cell growth and drug sensitivity in culture using human and other tumour cell lines. *Cancer Res* 1988, **48**(17), 4827–4833.
11. Paull KD, Shoemaker RH, Hodes L, Monks A, Scudiero DA, Rubinstein L, *et al*. Display and analysis of patterns of differential activity of drugs against human tumour cell lines: development of mean graph and COMPARE algorithm. *J Natl Cancer I* 1989, **81**(14), 1088–1092.
12. van Osdol WW, Myers TG, Paull KD, Kohn KW, Weinstein JN. Use of the Kohonen self-organizing map to study the mechanisms of action of chemotherapeutic agents. *J Natl Cancer I* 1994, **86**(24), 1853–1859.
13. Weinstein JN, Kohn KW, Grever MR, Viswanadhan VN, Rubinstein LV, Monks AP, *et al*. Neural computing in cancer drug development: predicting mechanism of action. *Science* 1992, **258**(5081), 447–451.
14. Li L, Wang HK, Kuo SC, Wu TS, Mauger A, Lin CM, *et al*. Antitumour agents. 155. Synthesis and biological evaluation of 3′,6,7-substituted 2-phenyl-4-quinolones as antimicrotubule agents. *J Med Chem* 1994, **37**(20), 3400–3407.
15. Chen K, Kuo SC, Hsieh MC, Mauger A, Lin CM, Hamel E, *et al*. Antitumour agents.174. 2′,3′,4′,5,6,7-Substituted 2-phenyl-1,8-naphthyridin-4-ones: their synthesis, cytotoxicity, and inhibition of tubulin polymerization. *J Med Chem* 1997, **40**(14), 2266–2275.
16. Adams J, Palombella VJ, Sausville EA, Johnson J, Destree A, Lazarus DD, *et al*. Proteasome inhibitors: a novel class of potent and effective antitumour agents. *Cancer Res* 1999, **59**(11), 2615–2622.
17. Holbeck SL, Sausville EA. The proteasome and the COMPARE algorithm. In Adams J, ed. *Proteasome Inhibitors in Cancer Therapy*. Humana Press, 2004.

18. Lee JS, Paull K, Alvarez M, Hose C, Monks A, Grever M, *et al*. Rhodamine efflux patterns predict P-glycoprotein substrates in the National Cancer Institute drug screen. *Mol Pharmacol* 1994, **46**(4), 627–638.
19. Ross DT, Scherf U, Eisen MB, Perou CM, Rees C, Spellman P, *et al*. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000, **24**(3), 227–235.
20. Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, *et al*. A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 2000, **24**(3), 236–244.
21. Kim CH, Kim JH, Moon SJ, Hsu CY, Seo JT, Ahn YS. Biphasic effects of dithiocarbamates on the activity of nuclear factor-kappaB. *Eur J Pharmacol* 2000, **392**(3), 133–136.
22. Wosikowski K, Schuurhuis D, Johnson K, Paull KD, Myers TG, Weinstein JN, *et al*. Identification of epidermal growth factor receptor and c-erbB2 pathway inhibitors by correlation with gene expression patterns. *J Natl Cancer I* 1997, **89**(20), 1505–1515.
23. Dick RA, Yu X, Kensler TW. NADPH alkenal/one oxidoreductase activity determines sensitivity of cancer cells to the chemotherapeutic alkylating agent irofulven. *Clin Cancer Res* 2004, **10**(4), 1492–1499.
24. Algeciras-Schimnich A, Pietras EM, Barnhart BC, Legembre P, Vijayan S, Holbeck SL, *et al*. Two CD95 tumour classes with different sensitivities to antitumour drugs. *Proc Natl Acad Sci USA* 2003, **100**(20), 11445–11450.
25. Huang Y, Anderle P, Bussey KJ, Barbacioru C, Shankavaram U, Dai Z, *et al*. Membrane transporters and channels: role of the transportome in cancer chemosensitivity and chemoresistance. *Cancer Res* 2004, **64**(12), 4294–4301.
26. Koenekoop RK. The gene for Stargardt disease, ABCA4, is a major retinal gene: a mini-review. *Ophthalmic Genet* 2003, **24**(2), 75–80.