Mining functional relationships in feature subspaces from gene expression profiles and drug activity profiles

Lei Bao, Tao Guo, Zhirong Sun*

Institute of Bioinformatics, Department of Biological Sciences and Biotechnology, Tsinghua University, Beijing 100084, PR China

Received 30 November 2001; revised 29 January 2002; accepted 29 January 2002

First published online 6 March 2002

Edited by Giulio Superti-Furga

Abstract In an effort to determine putative functional relationships between gene expression patterns and drug activity patterns of 60 human cancer cell lines, a novel method was developed to discover local associations within cell line subsets. The association of drug–gene pairs is an explorative way of discovering gene markers that predict clinical tumor sensitivity to therapy. Nine drug–gene networks were discovered, as well as dozens of gene–gene and drug–drug networks. Three drug–gene networks with well studied members were discussed and the literature shows that hypothetical functional relationships exist. Therefore, this method enables the gathering of new information beyond global associations. © 2002 Federation of European Biochemical Societies. Published by Elsevier Science B.V. All rights reserved.

Key words: Drug activity; Gene expression; Feature subspace; Local association

1. Introduction

Recently, Ross et al. [1] used cDNA microarrays to assess gene expression profiles of 60 human cancer cell lines. This gene expression dataset, together with the independently developed drug activity dataset recording compounds' anticancer profiles against the same 60 cell lines [2], provided the opportunity to take a global, systematic look at tumor molecular biology and pharmacology. These two datasets were used to discover gene–gene, drug–drug and drug–gene correlations in tumor cells. Since the gene expression profiles are those for untreated cells, the drug–gene correlations reflect sensitivity to therapy rather than the molecular consequences of therapy. Scherf et al. [3] indicated that some causally related drug–gene pairs had highly correlated profiles. They used the average-linkage clustering method to assess the relationships. The major drawback was the difficulty in identifying positives from the noise. Butte et al. [4] tried to solve this problem by data filtering and permutation trials. However, they found only one hypothetical drug–gene pair.

Here a new strategy was developed to discover additional biological associations that could not easily be discovered by already existing methods. The motivation of this work was to find drug–gene, gene–gene, and drug–drug local associations. A local association is defined here as an association that may exist within a cell line subset, but that may not necessarily exist in the whole cell line set. Anticancer drugs may have significantly different efficacies for different types of cells and two drugs may have similar activity patterns only in certain cell line subsets. Analogously, some genes may be co-regulated by tissue-specific transcription factors and may display similar patterns only in cell lines from this tissue. Under such conditions, global correlations may not exist and all the biological effects may be embodied in the local associations. In our approach, the measurements (across 58 cell lines) for each drug or gene were first partitioned into several subsets according their probability density distribution, and then the pair-wise correlations within these subsets were examined to identify statistically significant pairs. This method for determining local associations enabled the discovery of new biological facts and information relating these two profiles.

2. Materials and methods

2.1. Gene expression data and drug activity data

Gene expression was measured in 60 cell lines by microarrays, and the ratios of the relative mRNA level of each cell line to a common reference sample pool were log transformed [1]. For each cell line, drug activity was expressed as the negative logarithm of GI_{50}, where GI_{50} was the concentration of the compound needed to cause 50% cell growth inhibition [5]. A 1376-gene subset with greater expression variations and a well validated 1400-compound subset were selected as according to Scherf et al. [3] from public databases (http://discover.nci.nih.gov). Because there were only two prostate cell lines, these two samples were eliminated. The cell lines used are listed in Table 1. Thus, there was a 1376×58 gene matrix and a 1400×58 drug matrix, each row of which was a profile vector. The data were normalized so that each profile vector had a zero mean and a standard deviation of one.

2.2. Partitioning of each profile vector into feature subspaces

For every gene or drug profile vector, the 58 measurements (one measurement per cell line) were partitioned into several subsets according to the probability density distribution of the measurements. The resulting cell line subsets are each called a 'feature subspace' (FS). This partitioning procedure was repeated for all the genes and drugs. A multi-scale clustering algorithm based on multiple scale theory was used here [6]. The key idea of multi-scale clustering is that data can be represented in the transformed space (scale-space) at different resolution levels called scales, and any prominent data structure should survive over many scales. For one-dimensional data, a scale-space representation of the data points, x_k, can be realized by convolving them with a Gaussian kernel Φ_σ of scale size σ,

*Corresponding author. Fax: (86)-10-62772237.
E-mail address: sunzhr@mail.tsinghua.edu.cn (Z. Sun).

Abbreviations: FS, feature subspace; CFS, common feature subspace; EST, expressed sequence tag; MDR, multidrug resistance; AR, androgen receptor

PII: S0014-5793(02)02515-2

Table 1
Cancer cell lines and their numerical labels used in this work

Tissue origin	Cell line name
Breast (1–8)	1: BT-549, 2: HS578T, 3: MCF7, 4: MCF7/ADF-RES, 5: MDA-MB-231/ATCC, 6: MDA-MB-435, 7: MDA-N, 8: T-47D
CNS (9–14)	9: SNB-19, 10: SNB-75,11: SF-268, 12: SF-295, 13: SF-539, 14: U251
Colon (15–21)	15: COLO205, 16: HCC-2998, 17: HCT-116, 18: HCT-15, 19: HT29, 20: KM12, 21: SW-620
Lung (22–30)	22: A549/ATCC, 23: EKVX, 24: HOP-62, 25: HOP-92, 26: NCI-H226, 27: NCI-H23, 28: NCI-H322M, 29: NCI-H460, 30: NCI-H522
Leukemia (31–36)	31: CCRF-CEM, 32: HL-60, 33: K-562, 34: MOLT-4, 35: RPMI-8226, 36: SR
Melanoma (37–44)	37: LOXIMVI, 38: M14, 39: MALME-3M, 40: SK-MEL-2, 41: SK-MEL-5, 42: SK-MEL-28, 43: UACC-62, 44: UACC-257
Ovarian (45–50)	45: IGROV1, 46: OVCAR-3, 47: OVCAR-4, 48: OVCAR-5, 49: OVCAR-8, 50: SK-OV-3
Renal (51–58)	51: 786-0, 52: A498, 53: ACHN, 54: CAKI-1, 55: RXF-393, 56: SN12C, 57: TK-10, 58: UO-31

CNS: central nervous system.

$$\Phi_\sigma(t) = -\frac{1}{\sigma\sqrt{(2\pi)}}\exp\left[-\frac{1}{2}\left(\frac{t}{\sigma}\right)^2\right] \qquad (1)$$

to generate the following potential field function:

$$\Psi_\sigma(\cdot) = \sum_{k=1}^{n}\Phi_\sigma(\cdot - x_\mathrm{k}) \qquad (2)$$

From the statistical pattern recognition viewpoint, the potential field function can be viewed as a special case of the Pazen window density estimate of the probability density function for a scale size σ. The local minima of this function correspond to the locally highest concentration of data points in each cluster, and locating all of them by optimization algorithms gives a certain classification of data for that scale size. The data were then classified in this way for various scale sizes. The classification that survived over the widest range of scales was chosen as the optimal classification.

To give prominence to a single gene's effect, a FS whose gene expression level varied far from the overall average expression level was preferable (see Section 4). Therefore, only FSs whose absolute mean values exceeded 1.0 were used in the following calculations.

2.3. Finding the common feature subspace

For every possible drug–gene pair (gene–gene pair and drug–drug pair were essentially similar), the intersection of the gene's FS and the drug's corresponding FS containing at least six cell lines was recorded. FSs with too few cell lines were not interesting and a minimum of six was required because every tissue category contains at least six members (Table 1). Ideally, the gene's FS, the drug's FS and their intersection should be just the same. Yet, since there were many missing values in the two profiles, and since, according to Golub et al. [7], a single gene would rarely completely coincide with any phenotype pattern, few such cases were expected. Therefore, common feature subspace (CFS) was defined as the intersection that contained the majority of the cell lines appearing in the two FSs. Within the union of the two FSs, the number of cell lines not belonging to the intersection should be less than or equal to half the number of cell lines belonging to the intersection and should also be less than six.

2.4. Forming relevance networks

The Pearson correlation coefficient was calculated within the CFS for every drug–gene pair. Statistical hypotheses that the correlations were significantly different from zero were tested at a significance level of 0.02. For the gene–gene and drug–drug pairs, correlations were further tested that they not only significantly differed from zero, but could also explain at least 50% of the variance [8]. Finally, relevance networks were formed by grouping pairs sharing common members.

3. Results

3.1. Permutation trials

The method for finding a candidate functional pair is depicted in Fig. 1. All pairs were determined in this way and represented as relevance networks. All the networks obtained can be found in the supplementary material. Permutation trials were done to test how well the method determines positives from noise. For each drug and gene, measurements were randomly permutated 100 independent times. The results are shown in Table 2. The positives obtained from the true dataset are significantly more than those obtained from randomized datasets. Only one false positive was reproduced by 100 permutations, so candidate pairs identified by the method could rarely be generated through random chance.

3.2. Gene–gene networks

There were, in total, 24 gene–gene networks, 13 of which were single pairs. Many of the networks contained synonymy genes, which demonstrated the validity of this approach. One network listed here (Fig. 2, network 1) contained AKR1C1, AKR1C3 and AKR1C4, all of which belong to the aldo-keto reductase family 1 [9].

The largest network contained nine unique genes (Fig. 2, network 2). They all up-regulated in the CFS having cells with stromal origin, but down-regulated in blood cells and cells with epithelial origin. The literature showed that these genes are related to the metabolism of steroids and fatty acids. For

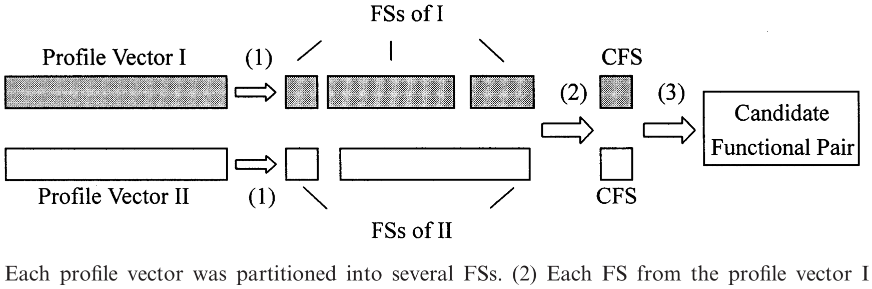


Fig. 1. Methodology. (1) Each profile vector was partitioned into several FSs. (2) Each FS from the profile vector I was intersected with the corresponding FS from the profile vector II. If the resulting intersection covered the majority of cell lines from the two FSs, it was called a CFS. Only the CFS was retained for the next procedure. (3) The correlation coefficient was calculated within the CFS. If the correlation was statistically different from zero, the pair was considered as a functional candidate.

Table 2
Permutation results for drug–gene pairs

Dataset	Number of CFSs found	Number of candidate pairs found
Original dataset	131	12
Permutated datasets	58[a] (0.58[b])	1[a] (0.01[b])

[a] The sum of 100 independent permutations.
[b] Averages.

example, PRKAG1 is a subunit of AMPK that is known to play a key role in regulating both fatty acid and cholesterol synthesis [10]. Fibronectin was found to regulate membrane lipid biosynthesis through AMPK [11]. LXRA mediates cross-talk between fatty acid and cholesterol metabolism [12]. CCND2 is a cyclin gene regulated by the steroid hormone level [13]. These lines of evidence demonstrate that the network caught the common biological properties of the genes and imply the genes may represent key positions in the cross-links of different signaling pathways.

3.3. Drug–drug networks

There were, in total, 18 drug–drug networks, 16 of which were single pairs. Many of the networks had chemically related or biologically related drugs. One network that contained two drugs with known mechanisms is shown in Fig. 2 (network 3). These two drugs are both DNA synthesis inhibitors and the biological mechanisms are similar [14]. Their CFS lacked most of the colon cells and some other epithelial cells. Both drugs displayed high activity in the CFS, which indicates that the two drugs more effectively inhibit growth of blood cells and cells of stromal origin, as compared to epithelial origin. Associating drugs with cells having similar chemotherapeutic susceptibilities may provide useful clues in designing effective drug treatment.

3.4. Drug–gene networks

There were, in total, nine drug–gene networks, seven of which were single pairs. Four of these networks with well studied members are listed in Fig. 3.

The first network (Fig. 3, network 1) contained the drug thaliblastine and the gene caveolin-2 in the CFS of mainly leukemia cell lines. Thaliblastine inhibits a tumor cell's multi-drug resistance (MDR) by direct interaction with the P-glycoprotein, a drug transporter localized in the plasma membrane microdomain called caveola [15,16]. Caveolin-2 is one of the main structural proteins of caveolae and hetero-oligo-

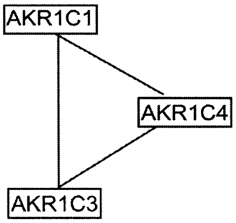
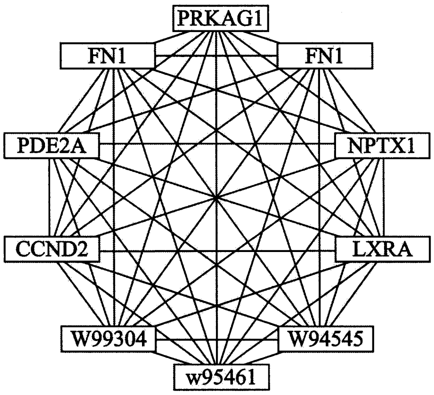
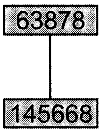
N	Diagram	
1	AKR1C1: Dihydrodiol dehydrogenase AKR1C3: Trans-1,2-dihydrobenzene -1,2-diol dehydrogenase AKR1C4: Chlordecone reductase CFS: 12 15 16 22 23 28 29 50 52 High gene expression	
2	PRKAG1: AMP-activated protein kinase, gamma-1 subunit LXRA: Liver X receptor alpha NPTX1: Neuronal pentraxin 1 FN1: Fibronectin 1 CCND2: Cyclin D2 PDE2A: cGMP-stimulated Phosphodiesterase 2A W99304,W95461,W94545: ESTs CFS: 1 2 4 10-13 22 24 25 29 35 37-40 42-44 49 51 53-55 57 58 High gene expression	 PRKAG1 FN1 FN1 PDE2A NPTX1 CCND2 LXRA W99304 W94545 w95461
3	63878: Cytosine arabinoside, NSC 63878 145668: Cyclocytidine, NSC 145668 CFS: 3 4 7 9 11 12 14 17 22 24 25 27 29 31 32 34 36-39 41 43 48 49 51-54 56 58 High drug activity	

Fig. 2. Examples of gene–gene and drug–drug networks. The first two networks are examples of gene–gene associations. The third network is an example of drug–drug associations. Nodes representing the expression levels of a single gene are in white and labels drawn within the nodes correspond to the GeneCard or GenBank accession codes. Nodes representing measures of susceptibility to a single compound are shaded gray and labels drawn within the nodes correspond to the compound's National Cancer Institute NSC number. The genes can be found at http://bioinfo.weizmann.ac.il or http://www.ncbi.nlm.nih.gov/, and the compounds can be found at http://dtp.nci.nih.gov/docs/dtp_search.html. The cell line labels in the CFS correspond to those in Table 1. The levels of gene expression or drug activity in each CFS are also indicated. Further details regarding all the networks found are located as Supplementary Material in the online version of this article.

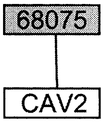
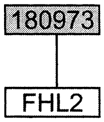
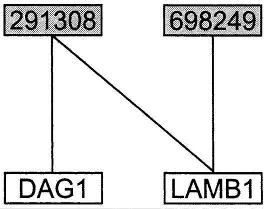
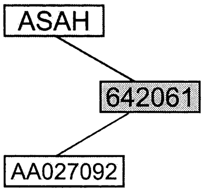
N	Diagram	
1	68075: Thaliblastine, NSC 68075 CAV2: Caveolin-2 CFS: 3 30 31 32 34 35 36 Low drug activity and low gene expression	
2	180973: Tamoxifen, NSC 180973 FHL2: Four and a half LIM domain protein CFS: 29 30 31 32 34 36 Low drug activity and low gene expression	
3	291308: Terrein, NSC 291308 698249: NSC 698249 LAMB1: Laminin, beta 1 DAG1: Dystroglycan 1 CFS: 17 21 31 32 33 34 35 36 Low drug activity and low gene expression.	
4	642061: NSC 642061 ASAH: Acid ceramidase AA027092: EST CFS: 1 6 7 8 38 39 40 41 42 43 44 46 High drug activity and high gene expression	

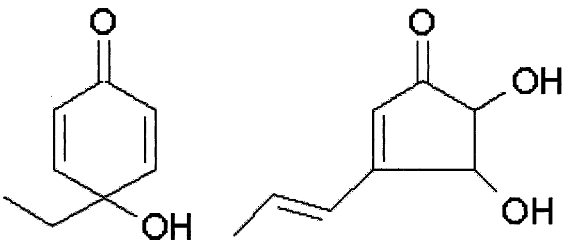
Fig. 3. Examples of networks between drugs and genes. Symbols are the same as in Fig. 2.

mers formed by caveolin-1 and caveolin-2 are the functional assembly units that drive caveola formation in vivo. Lavie et al. found that P-glycoprotein-related MDR was evidently associated with increased caveola number and a massive up-regulation of caveolin expression [17]. Following their hypothesis that precaveolae and caveolins might facilitate the delivery of drugs from intracellular compartments to plasma membrane-resident drug transporters [18], we suggested that caveolin-2 might play a role in bringing thaliblastine to the vicinity of its molecular target, P-glycoprotein.

The second network (Fig. 3, network 2) contained the drug tamoxifen and the gene FHL2 in the CFS of mainly leukemia cell lines. Tamoxifen is an anti-estrogen drug that is widely used in the therapy of breast cancer. Graham indicated that tumor sensitivity to tamoxifen might be governed by a complex set of transcriptional coregulators for steroid hormone receptors [19]. Recently, Zhou et al. found that in breast epithelium, tamoxifen treatment down-regulated androgen receptor (AR) expression and affected AR-dependent gene transcriptions [20]. The authors proposed that tamoxifen might interact with AR to regulate AR target gene transcriptions. FHL2 encodes a LIM-only protein with four and a half LIM domains and its over-expression can induce apoptosis [21]. Very interestingly, FHL2 was recently found to be a specific transcriptional coactivator of the AR [22]. Therefore, it is possible that the drug tamoxifen and the gene FHL2 might have a certain kind of interaction in AR-dependent transactivations.

The third network (Fig. 3, network 3) contained two drugs (Terrein, a 2-cyclopenten-1-one derivative and the drug 698249) as well as two genes (laminin β1 and dystroglycan 1) in the CFS of all leukemia cell lines as well as two colon cell lines. 2-Cyclopenten-1-one exerts its anticancer activity through inducing Waf1 gene expression and the α,β-unsaturated carbonyl structure is essential for the drug's activity [23]. Drug 698249 showed a high positive correlation (0.90) with Terrein in the CFS. Consistently, drug 698249 also has the same structural determinants (Fig. 4), which may be an important clue for understanding its mechanism. Therefore, this method seems useful (at least in this case) for identifying molecular structures needed for each drug's activity. Laminin β1 encodes an extracellular matrix (ECM) protein subunit and dystroglycan 1 encodes a laminin receptor. In a recent paper of Staunton et al. for identifying marker genes that predict chemosensitivity, they noted that the marker gene sets for a number of drugs are enriched in cytoskeleton/ECM genes [24]. Therefore, this association is not surprising. It is possible that cytoskeletal signatures may reflect cellular components that influence sensitivity to a variety of compounds rather than functioning as direct targets of compound activity.

The fourth network (Fig. 3, network 4) is an example of networks with high drug activity and high gene expression.



NSC 698249

Terrein, NSC 291308
(a cyclopentenone derivative)

Fig. 4. Structures of two drugs from Fig. 3, network 3. They both have α,β-unsaturated carbonyl structures that are important for cyclopentenone to exert its anticancer activity. Therefore, the similarity of their anticancer mechanisms was suggested.

The CFS contained most of the melanoma cell lines as well as two breast cell lines that supposedly originated from melanoma [1]. ASAH encodes an acid ceramidase. ASAH mRNA expression was found to be different between cancers and their normal counterparts [25], which suggested that ASAH might be involved in the neoplastic process. The drugs mechanisms are unknown at present.

4. Discussion

Scherf et al. [3] suggested the use of a global correlation coefficient to associate genes and drugs. However, even at a significance level of 0.001, there were still about 10 000 drug–gene pairs whose correlation coefficients were statistically different from zero. The majority of these were false positives, partly due to the outliers' dominant contribution to the correlation coefficient. True functional positives of interest cannot be easily located in such a noisy pair set. Butte [4] used the permutation method to deduce a very stringent correlation coefficient threshold. Although they did find one drug–gene pair that passed the threshold, the sensitivity seemed too low (one functional pair out of 33 million drug–gene pairs). The alternative approach used here had both relatively high sensitivity and high specificity. Nine candidate functional drug–gene networks were found out of 2 million drug–gene pairs. Some of the networks were verified in the literature.

The methodology was designed to find local functional associations, which enabled the gathering of new information beyond global associations. The drug–gene networks found here had only moderate global correlations, which demonstrated the ineffectiveness of using the global approach for discovery. This methodology had an additional advantage in a biological sense. Many drugs have complex effects on several cross-linked gene networks, making it difficult to observe the effect of a single gene. The global association approach is inefficient in such a context, because in the majority of cell lines, a single gene's expression level is usually constitutive. However, in our approach, only FSs with gene expression obviously up-regulated or down-regulated were used for further calculations. Therefore, this approach in some sense simulated the effect of gene knockout or over-expression, and gave prominence to a single gene's effect. It is interesting to note that the majority of CFSs found here contained cell lines from the same origin (e.g. tissue origin; stromal or epithelial type).

This method was unsupervised and could easily be scaled up. One only needs to provide the specificity control parameters of the CFSs. Such parameters could be adjusted on the trade-off between sensitivity and specificity. Furthermore, although we applied this approach to a small subset of the original database, it could be easily scaled up with no modification. The methodology was relatively insensitive to the outliers which can bias the global correlation coefficient. Here, multi-scale clustering was used to obtain FSs, because it has two advantages over other unsupervised methods. First, it automatically determines the cluster number objectively. Second, the resulting classification reflects the natural underlying structure of the original data.

This approach has several limitations. First, the gene expression profiles were collected in untreated cell lines. Therefore, the relationships established between drugs and genes were correlative and not causal. Second, this method did not take into account the synergetic effects of different genes. In fact, drug efficacy is affected by a set of genes, and they should be considered simultaneously. Third, local associations in the FSs beyond our definition could not be found by this approach. An example is the known functional pair of the drug ASNS and the gene L-asparaginase [3]. The drug and gene displayed extremely high negative correlation in six leukemia cell lines but only moderately high negative correlation in all cell lines. This is a good example illustrating the value of local associations. However, their measurements in these cell lines were very different and belonged to different FSs in our results, so we failed to find this pair. This pair suggested subsets other than our FS definition might also be biologically significant. These limitations indicate the direction of future research.

Acknowledgements: The authors would like to thank Professor Patrick O. Brown and John Weinstein for providing the GI_{50} data and the gene expression data and thank Professor Atul J. Butte, Richard Beliveau and Dr. H.B. Ji for their helpful suggestions. This work was supported by a National Natural Science Grant (China) (No. 39980007).

References

[1] Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Rees, C., Spellman, P., Iyer, V., Jeffrey, S.S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J.C., Lashkari, D., Shalon, D., Myers, T.G., Weinstein, J.N., Botstein, D. and Brown, P.O. (2000) Nature Genet. 24, 227–235.

[2] Weinstein, J.N., Myers, T.G., O'Connor, P.M., Friend, S.H., Fornace Jr., A.J., Kohn, K.W., Fojo, T., Bates, S.E., Rubinstein, L.V., Anderson, N.L., Buolamwini, J.K., van Osdol, W.W., Monks, A.P., Scudiero, D.A., Sausville, E.A., Zaharevitz, D.W., Bunow, B., Viswanadhan, V.N., Johnson, G.S., Wittes, R.E. and Paull, K.D. (1997) Science 275, 343–349.

[3] Scherf, U., Ross, D.T., Waltham, M., Smith, L.H., Lee, J.K., Tanabe, L., Kohn, K.W., Reinhold, W.C., Myers, T.G., Andrews, D.T., Scudiero, D.A., Eisen, M.B., Sausville, E.A., Pommier, Y., Botstein, D., Brown, P.O. and Weinstein, J.N. (2000) Nature Genet. 24, 236–244.

[4] Butte, A.J., Tamayo, P., Slonim, D., Golub, T.R. and Kohane, I.S. (2000) Proc. Natl. Acad. Sci. USA 97, 12182–12186.

[5] Boyd, M.R. and Paull, K.D. (1995) Drug Dev. Res. 34, 91–109.

[6] Nakamura, E. and Kehtarnavaz, N. (1998) Pattern Recognit. Lett. 19, 1265–1283.

[7] 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. (1999) Science 286, 531–537.

[8] Yuan, Z.F. and Zhou, J.Y. (2000) Design and Analysis of Experiment, pp. 171–178. Higher Education Press, Beijing.

[9] Khanna, M., Qin, K.N., Klisak, I., Belkin, S., Sparkes, R.S. and Cheng, K.C. (1995) Genomics 25, 588–590.

[10] Hardie, D.G. and Carling, D. (1997) Eur. J. Biochem. 246, 259–273.

[11] Page, K. and Lange, Y. (1997) J. Biol. Chem. 272, 19339–19342.

[12] Tobin, K.A., Steineger, H.H., Alberti, S., Spydevold, O., Auwerx, J., Gustafsson, J.A. and Nebb, H.I. (2000) Mol. Endocrinol. 14, 741–752.

[13] Robker, R.L. and Richards, J.S. (1998) Mol. Endocrinol. 12, 924–940.

[14] Novotny, L., Reichelova, V., Balazova, E. and Ujhazy, V. (1990) Neoplasma 37, 13–22.

[15] Chen, G., Ramachandran, C. and Krishan, A. (1993) Cancer Res. 53, 2544–2547.

[16] Demeule, M., Jodoin, J., Gingras, D. and Beliveau, R. (2000) FEBS Lett. 466, 219–224.

[17] Lavie, Y., Fiucci, G. and Liscovitch, M. (1998) J. Biol. Chem. 273, 32380–32383.

[18] Liscovitch, M. and Lavie, Y. (2000) Trends Biochem. Sci. 25, 530–534.

[19] Graham, J.D., Bain, D.L., Richer, J.K., Jackson, T.A., Tung, L. and Horwitz, K.B. (2000) J. Steroid Biochem. Mol. Biol. 74, 255–259.

[20] Zhou, J., Ng, S., Adesanya-Famuiya, O., Anderson, K. and Bondy, C.A. (2000) FASEB J. 14, 1725–1730.

[21] Scholl, F.A., McLoughlin, P., Ehler, E., de Giovanni, C. and Schafer, B.W. (2000) J. Cell Biol. 151, 495–506.

[22] Muller, J.M., Isele, U., Metzger, E., Rempel, A., Moser, M., Pscherer, A., Breyer, T., Holubarsch, C., Buettner, R. and Schule, R. (2000) EMBO J. 19, 359–369.

[23] Bui, T. and Straus, D.S. (1998) Biochim. Biophys. Acta 1397, 31–42.

[24] 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. (2001) Proc. Natl. Acad. Sci. USA 98, 10787–10792.

[25] Maeda, I., Takano, T., Matsuzuka, F., Maruyama, T., Higashiyama, T., Liu, G., Kuma, K. and Amino, N. (1999) Int. J. Cancer 81, 700–704.