Simultaneous Modeling of Concentration-effect and Time-course Patterns in Gene Expression Data from Microarrays

YSEULT F. BRUN[1,2], RAM VARMA[3], SUZANNE M. HECTOR[3], LAKSHMI PENDYALA[3], RAMAKUMAR TUMMALA[3] and WILLIAM R. GRECO[1,2]*

[1]Cancer Prevention and Population Sciences and
[3]Department of Medicine, Roswell Park Cancer Institute, Buffalo, NY 14263;
[2]University at Buffalo, School of Pharmacy, Buffalo, NY 14260, U.S.A.

Abstract. *Background: Time-course and concentration-effect experiments with multiple time-points and drug concentrations provide far more valuable information than experiments with just two design-points (treated vs. control), as commonly performed in most microarray studies. Analysis of the data from such complex experiments, however, remains a challenge. Materials and Methods: Here we present a semi-automated method for fitting time profiles and concentration-effect patterns, simultaneously, to gene expression data. The submodels for time-course included exponential increase and decrease models with parameters, such as initial expression level, maximum effect, and rate-constant (or half-time). The submodel for concentration-effect was a 4-parameter Hill model. Results: The method was applied to an Affymetrix HG-U95Av2 dataset consisting of 51 arrays. The specific study focused on the effects of two platinum drugs, cisplatin and oxaliplatin, on A2780 human ovarian carcinoma cells. Replicates were available at most time points and concentrations. Eighteen genes were selected, and after selection, time-course and concentration-effect were modeled simultaneously. Conclusion: Comparisons of model parameters helped to distinguish genes with different expression patterns between the two drug treatments. This overall paradigm can help in understanding the molecular mechanisms of the agents, and the timing of their actions.*

*Supplementary material should be requested from the corresponding author *via* e-mail.

Correspondence to: William R. Greco, Roswell Park Cancer Institute, Cancer Prevention and Population Sciences, Elm and Carlton St, Buffalo, NY 14263, U.S.A. e-mail: William.Greco@roswellpark.org

Microarray-based experiments have been used to study changes in gene expression due to drug treatment as a function of either time or concentration (1-3). Most published data on such experiments tend to have a limited number of design points – often just two, such as before-treatment *versus* after-treatment or treated *versus* untreated. However, with improved accuracy and better affordability of the technology, microarray experiments of increasingly complex designs are being conducted. Among these are time-course and concentration-effect experiments that seek to understand response to drugs as functions of time or concentration or both, including possible interactions between these two variables (4-6). Time-course experiments with many design points ranging from prior to treatment to some significant time after completion of treatment, and concentration-effect experiments spanning the entire logarithmic range of concentrations in small increments will clearly provide a deeper understanding of the cellular pharmacokinetics and pharmacodynamics of the agents.

Methods that have become standard in microarray data analysis such as class comparisons (7, 8), clustering (9, 10), and supervised or unsupervised learning techniques (11-13) have limited use for analyzing such data. Class comparisons with methods such as SAM/Significance Analysis of Microarrays (7) are useful in obtaining lists of genes that are significantly differentially expressed between two or more classes. Supervised machine learning techniques such as k-Nearest Neighbor and Support Vector Machines are useful in distinguishing sets of genes with predictive power in classification. These methods have specific purposes and are not directly applicable to the time-course and concentration-effect data described here. Clustering methods such as hierarchical clustering and self-organizing maps provide good tools for visualization and clues to underlying patterns that often need to be investigated further. However, they provide little quantitative information regarding these patterns.

It is reasonable to assume that gene expression due to drug treatment varies continuously with time and

concentration, making it feasible to describe the data with a relatively simple mathematical model. Such a model should capture important features in the expression pattern. A typical time-course profile or concentration-effect pattern for a single gene consisting of many design points with replicates can be reduced to a simple model with a small set of parameters. The model effectively summarizes the data, reduces the dimensionality of the problem and facilitates further analysis.

The choice of the model is therefore important. Simple nonparametric models such as splines can be fit to the data (14). However, there are advantages to using models with parameters that can be interpreted meaningfully. Obvious choices for models for gene expression microarray data are those that have been previously published and used extensively in studying response to drugs using traditional slower nonmultiplexed throughput methods. These include models such as the Hill model for concentration-effect patterns and exponential models for time-course profiles.

We searched for details of previously-published methods used for the analysis of data from both time-course and concentration-effect (dose-response) gene expression microarray experiments. We reviewed full papers and abstract summaries of publications listed in the website, http://www.sinica.edu.tw/~hmwu/SMDA/TimeCourse/index.htm, and in the reference lists of three time-course microarray review articles (15-17). These sources referenced over 100 articles dealing with time-course data modeling, but only three (18-20) dealing with concentration-effect modeling. Most of the time-course data analysis approaches come from the machine learning subfield of computer science. A representative list of 20 phrases (21-40) that characterize approaches found among the time-course data analysis articles is presented in Table I.

The three articles that mention microarray concentration-effect data analysis contain no actual analyses. Zhou *et al.* (18) used only two different doses, and did not model the pattern of concentration-effect. Pedadda *et al.* (19) applied an approach only to time series data, and only implied that the same approach could be also applied to concentration-effect data. Eckel *et al.* (20) suggested an empirical Bayes gene screening tool applicable to time series as well as to concentration-effect data, but doesn't apply it to concentration-effect data.

The field of pharmacometrics has a long history of successful modeling of both time-course and concentration-effect phenomena, commonly by fitting mechanistic, empirical or quasi-mechanistic nonlinear models to the response data with nonlinear regression or maximum likelihood approaches. Our modeling approach presented in this paper is inspired by this field, but is unique. We fit innovative variants of standard pharmacometric models to

gene-expression microarray data, both time course and concentration-effect, with nonlinear regression. It is especially rare to model concentration-effect gene expression data; the only other paper to which we are aware was recently published by our group (41). We improved our data analysis approach in the present paper by modeling simultaneously the data from time-course and concentration effect for each gene, whereas they were not fitted simultaneously in the previous paper (41).

In this study, we modeled microarray gene expression data from a set of 51 Affymetrix HG-U95Av2 GeneChips. The time-course data consisted of gene expression profiles obtained prior to treatment, and at five time-points up to 24h after treatment with IC_{90} growth-inhibitory concentrations for both drugs. Gene expression profiles obtained 16h after treatment with IC_{10}, IC_{25}, IC_{50}, IC_{75} and IC_{90} growth-inhibitory concentrations for cisplatin and oxaliplatin comprised the concentration-effect data. Data from control samples, treated with drug-free medium, was available at four time points. Mathematical models, with expression levels as functions of time or drug concentration, were fitted to the time-course and concentration-effect data, and estimated model parameters were further analyzed to extract significant patterns.

Preliminary results were presented as a poster at the American Association of Cancer Research (AACR) meeting in April 2007 in Los Angeles.

The data discussed in this publication have been deposited in NCBIs Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE8057 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE8057).

Materials and Methods

Experiment. All experiments were performed on A2780 human ovarian carcinoma cells. For the time-course experiments, samples were exposed to IC_{90} concentrations of oxaliplatin (32 μM) or cisplatin (25 μM) for 2 hours, and then allowed to grow in drug free medium. Cells were harvested and processed at the following times (Table II): before treatment, immediately after treatment (0 h), and at 2 h, 6 h, 16 h and 24 h after treatment. Control experiments were done following a similar design, but with exposure to drug-free medium. For the concentration-effect experiments, samples were treated with specific growth-inhibitory concentrations of cisplatin or oxaliplatin for 2 hours. The previously-estimated IC_{10}, IC_{25}, IC_{50}, IC_{75} and IC_{90} concentrations (for 2 h drug exposure; 72 h total growth) for cisplatin and oxaliplatin growth inhibition of A2780 were 2.6, 4.0, 6.4, 12 and 25 μM and 2.8, 4.8, 8.5, 17 and 32 μM, respectively. Cells were harvested and processed at 16 h after completion of treatment. There were 3 successive washings with PBS after removing the drug; and floating cells were not collected, only those that were adherent were analyzed. The time-course and concentration-effect experiments were repeated and data collected in duplicate for most conditions. Some design points

Table I. *Listing of time series data analysis methods.*

Phrase that characterizes Data Analysis Method	Ref.
Symbolic dynamical clustering method	21
Clustering based on a simulated annealing procedure	22
Bayesian method for model-based clustering of gene expression dynamics, using autoregressive equations	23
Dynamic Bayesian network	24
Cubic spline - piecewise polynomial	25
Mixed effect model with splines for clustering ; time-lagged correlation coefficient	26
Mixed-effect model using B-splines via an EM algorithm	27
Hidden Markov model	28
Hidden Markov model within a model-based clustering framework	29
Markov chain models using an EM algorithm	30
Mixture functional discriminant analysis	31
Least squares and estimating equations techniques	32
Time warping algorithm	33
Differential expression analysis	34
Permutation test on residuals from ANOVA	35
Support Vector Machine (SVM) algorithm	36
Periodogram as an exploratory device	37
Correspondence analysis	38
Combination of time series and static (average) expression data analysis	39
Functional hierarchical model using a Monte Carlo EM algorithm	40

were identical between the time-course and concentration-effect experiments. At all steps, the quality of each RNA sample was evaluated using gel electrophoresis and spectrophotometry. The biotinylated RNA products were then purified, fragmented, and hybridized to Affymetrix Gene Chip arrays and scanned following streptavidin-phycoerythrin staining. The quality of each fragmented sample was assessed by hybridization to test chips. cRNAs passing the quality checks were then hybridized to Affymetrix HGU95Av2 chips as per established Affymetrix protocols. The quality of the hybridization and image processing was checked using the array outlier computations in dChip (42). After a 2 h drug exposure at the IC_{90} concentrations of these platinum drugs, and a further 24 h incubation in drug-free growth medium, no more than 10-20% of the cells lift off (dead) from these adherent (live) cell cultures. In fact as evident from these data from the time course experiments, this 24 h incubation is an active period for many of the genes controlling apoptosis and cell cycle arrest; these genes are actively being turned on in a progressive, coordinated fashion. Cell death, under our experimental conditions, occurs generally between 72-96 h after drug exposure, so our data obtained after 24 h should be reflective of gene activation or inactivation only, without material interference from cell death.

Probe level analysis. Probe level expression summaries were obtained using the Robust Multichip Average (RMA) method (43) for all arrays that satisfied quality checks. In this method a linear model is fitted to the background corrected, quantile normalized, log transformed probe intensities. Expression summaries were generated using median polish. Only the Perfect Match signal intensities are used in this model. The RMA calculations were performed using the justRMA function in affy1.4.32 of the Bioconductor Project Release 1.4 Package (44) (http://www.bioconductor.org).

Table II. *Study design and sample replicates.*

	Pretreatment	0 h	2 h	6 h	16 h	24 h
	5					
Cisplatin		2	2	2	5	1
Oxaliplatin		2	2	2	4	1
Control		1	1	1	4	

	Control	IC_{10}	IC_{25}	IC_{50}	IC_{75}	IC_{90}
	4					
Cisplatin		2	2	2	2	5
Oxaliplatin		2	2	2	2	4

Description of data. In the time-course study, for each probeset there were 5 data points prior to treatment, 12 data points after cisplatin treatment, 11 after oxaliplatin treatment and 7 control data points. In the concentration-effect study, for each probeset, there were 4 drug free medium treated control data points, 13 data points with cisplatin treatment, and 12 with oxaliplatin treatment. There were four groups with four or more replicates: (i) pre-treatment, (ii) 16 h after treatment with IC_{90} concentration of cisplatin, (iii) 16 h after treatment with IC_{90} concentration of oxaliplatin, and (iv) control samples at 16 h after processing with drug-free medium. Table II summarizes the number of samples for the various conditions.

Time-course model. The expression level of a gene (probe-set) was modeled as a function of time with two parts: an initial steady state response (or effect), followed by an exponential increase or decrease to an asymptotic final response.

For $t \leq 0$, $y = y_0$

For $t \geq 0$,

$$y = y_0 + (y_{max} - y_0) * \left(1 - \exp\left(-\frac{\ln 2}{t_{1/2}} * t\right)\right) \qquad (1)$$

Here, y is expression level of a gene at time t. Treatment started at $t = -2$ h and stopped at $t = 0$ h. Data was available up to $t = 24$ h. Parameter y_0 is the initial constant steady state expression level. Parameter y_{max} is the final asymptotic expression level. The expression, $(y_{max} - y_0)$ is positive for increases and negative for decreases. Parameter $t_{1/2}$ is the half-time, *i.e.* time at which the expression level y of a gene has reached one-half of $(y_{max} - y_0)$, the maximum change in effect from the baseline steady state effect level. The expression shown above can be derived based on the assumption that instantaneous rate of change of expression is proportional to the difference between the expression level at that time and the final expression level.

Concentration-effect model. A four parameter Hill model (*e.g.*, Faessel *et al*, 2003, 46) was used to fit the concentration-effect data for a gene (probeset). The model equation for excitatory and inhibitory response is of the form

$$y = B + (E - B)\frac{x^m}{c_{50}{}^m + x^m} \qquad (2)$$

Here y is the expression level of a gene at the concentration x. Parameter B is the background, baseline or minimum level of expression. For an inhibitory response, B is y_{max}, the level at very high concentration-effect, and for an excitatory response, B is y_0, the level without any drug treatment. Parameter E is the maximum level of expression. For an inhibitory response, E is y_0, the control level when no drug is present, and for an excitatory response, E is y_{max}, the asymptotic level at a very high concentration. Parameter c_{50} is the drug concentration that induces 50% of the maximum effect $(E - B)$. For an inhibitory response, c_{50} is called IC_{50}, and for excitatory response c_{50} is referred to as EC_{50}. Parameter m is the Hill parameter, which describes the steepness (sigmoidicity) of the curve. For an inhibitory response, m is negative and for an excitatory response, m is positive.

Results

Selection of genes. Initial datasets included gene expression values for more than 12600 gene probe sets (all the gene probes and controls of the Affymetrix HGU95Av2 chip) (41, 45). All final nonlinear regression model fitting estimations were performed in SAS® NLIN version 9.1 (SAS Institute), and all final graphs were created using SigmaPlot® version 9.0 (Systat Software, Inc.), with Intel-based microcomputers.

From the 12,600 gene probesets, 34 genes were finally selected for modeling based on a 3-part decision process (I-III) described below:

I. To select genes where the expression was increased by both oxaliplatin and cisplatin, or decreased by both, we first performed two-class comparisons using Significance Analysis of Microarrays (SAM) with control of the False Discovery Rate (FDR) (7). The minimum FDR was the q-value; and the value of delta was chosen to give the minimum False Discovery Rate. Lists of significantly expressed genes were generated at the minimum FDR for the following comparisons: platinum drugs at 16 hours *versus* pre-treatment (q-value for cisplatin and oxaliplatin: <0.1% and 0.15% respectively), and platinum drugs at 16 hours *versus* control samples at 16 hours (q-value for cisplatin and oxaliplatin: 0.12% and 0.14%, respectively). Genes were then eliminated from the list if the estimated increase or decrease were less than 2-fold for either drug treatment, or if mean maximum expression levels with either treatment were less than 1000 (41, 45). After this selection, we visually inspected the time-course profiles and concentration-effect patterns of increase or decrease, and finally chose 12 genes with increases, and 5 genes with decreases, respectively, based on data quality.

II. To select genes differentially affected by the two platinum drugs, we first calculated for each gene the average expression for the control at 16 hours post drug-free medium exposure, and for oxaliplatin and cisplatin treatments at 16 hours post drug exposure, and then we calculated the ratios of average cisplatin/average control and average oxaliplatin/average control. These calculations were automated with a program written in the SAS® macro language.

Then we selected: a) Genes with no material changes in oxaliplatin *versus* control but increases in cisplatin *versus* control at 16 hours: the oxaliplatin/control ratio had to be between 0.9 and 1.1, the cisplatin/control ratio had to be more than 2. Three genes were selected. b) Genes with no material changes in oxaliplatin *versus* control but decreases in cisplatin *versus* control at 16 hours: the oxaliplatin/control ratio had to be between 0.9 and 1.1, the cisplatin/control ratio had to be less than 0.5, and the median expression after cisplatin had to be above 100 expression units. One gene was selected. c) Genes with no material changes in cisplatin *versus* control but increases in oxaliplatin *versus* control at 16 hours: the cisplatin/control ratio had to be between 0.9 and 1.1, the oxaliplatin/control ratio had to be more than 2, and the median expression after oxaliplatin had to be above 100. Six genes were selected. d) Genes with no material changes in cisplatin *versus* control but decreases in oxaliplatin *versus* control at 16 hours: the cisplatin/control ratio had to be between 0.9 and 1.1, the oxaliplatin/control ratio had to be less than 0.5, and the median expression after oxaliplatin had to be above 100 expression units. Ten genes were selected.

The thresholds for selection in decision process II were chosen by us, to select profiles that seemed to show trends.

III. After this decision processes I and II, we visually inspected the time-course profiles and concentration-effect

Table III. *all genes selected.*

Probeset	Category	Symbol	Gene title	Included in modeling	Included in graphs and tables
37842_at	cisplatin increases, oxaliplatin constant	MDFIC	MyoD family inhibitor domain containing	Yes	No
33334_at		ACYP1	acylphosphatase 1, erythrocyte (common) type	Yes	Yes
39742_at		TANK	TRAF family member-associated NFκB activator	Yes	Yes
2018_at	cisplatin decreases, oxaliplatin constant	GJA1	gap junction protein, alpha 1, 43 kDa (connexin 43)	No	No
39633_at	oxaliplatin increases, cisplatin constant	S100A3	S100 calcium binding protein A3	Yes	No
38287_at		PSMB9	proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional peptidase 2)	Yes	No
38576_at		HIST1H2	histone 1, H2bd	Yes	Yes
32805_at		AKR1C1	aldo-keto reductase family 1, member C1 (dihydrodiol dehydrogenase 1; 20-alpha (3-alpha)-hydroxysteroid dehydrogenase)	Yes	No
31792_at		ANXA3	annexin A3	No	No
34678_at		FER1L3	fer-1-like 3, myoferlin (C. elegans)	Yes	Yes
1536_at	oxaliplatin decreases, cisplatin constant	CDC6	CDC6 cell division cycle 6 homolog (S. cerevisiae)	Yes	No
39269_at		RFC3	replication factor C (activator 1) 3, 38 kDa	Yes	No
572_at		TTK	TTK protein kinase	Yes	No
40726_at		KIF11	kinesin family member 11	Yes	Yes
1801_at		BARD1	BRCA1 associated RING domain 1	Yes	No
1515_at		FEN1	flap structure-specific endonuclease 1	Yes	No
41569_at		DNAJC9	DnaJ (Hsp40) homolog, subfamily C, member 9	Yes	Yes
33222_at		FZD7	frizzled homolog 7 (Drosophila)	No	No
2086_s_at		TYRO3	TYRO3 protein tyrosine kinase	Yes	No
39631_at		EMP2	epithelial membrane protein 2	Yes	No
37374_at	both increase	ANXA4	annexin A4	Yes	Yes
36634_at		BTG2	BTG family, member 2	Yes	Yes
2031_s_at		CDKN1A	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	Yes	Yes
37579_at		CYFIP2	cytoplasmic FMR1 interacting protein 2	Yes	Yes
1243_at		DDB2	damage-specific DNA binding protein 2, 48 kDa	Yes	Yes
40336_at		FDXR	ferredoxin reductase	Yes	Yes
41191_at		KIAA0992	palladin	Yes	No
1890_at		PLAB	prostate differentiation factor	Yes	Yes
1173_g_at		SAT	Spermidine/Spermine N1-Acetyltransferase, Alt. Splice 2	Yes	No
33322_i_at		SFN	stratifin	Yes	No
39708_at		STAT3	signal transducer and activator of transcription 3 (acute-phase response factor)	Yes	Yes
36079_at		TP53I3	tumor protein p53 inducible protein 3	Yes	Yes
1945_at	both decrease	CCNB1	cyclin B1	Yes	Yes
527_at		CENPA	centromere protein A, 17kDa	Yes	No
40117_at		MCM6	MCM6 minichromosome maintenance deficient 6 (MIS5 homolog, S. pombe) (S. cerevisiae)	Yes	Yes
37228_at		PLK	polo-like kinase (Drosophila)	Yes	Yes
40619_at		UBE2S	ubiquitin carrier protein	Yes	No

patterns to verify the consistency in increase or decrease of gene expression. We finally selected the following genes: a) 3 genes with no changes due to oxaliplatin but increases due to cisplatin, afterwards referred to as category A. b) 0 genes with no changes caused by oxaliplatin but decreases from cisplatin. c) 5 genes with no changes in caused by cisplatin but increases from oxaliplatin, afterwards referred to as category B. d) 9 genes with no changes from cisplatin but decreases from oxaliplatin, afterwards referred to as category C. e) 12 genes with increases for both drugs, afterwards referred to as category D. f) 5 genes with decreases for both drugs, afterwards referred to as category E.

The data from these 34 genes were modeled. All 34 genes are listed in Table III.

Modeling process. For the design points with 4 or more replicates, for the 34 genes to be modeled, a mean gene expression value and a variance were first calculated. It was assumed and verified that the simple power model for data variance (the variance is proportional to the mean raised to a power) was appropriate for our data. The estimated slope (unweighted linear regression) of a linearized plot of log10 (variance) *versus* log10 (mean) was 1.67; for the subsequent nonlinear regression runs, data were iteratively reweighted by the reciprocal of the estimated variance, *i.e.*, the estimated gene expression raised to the 1.67th power. Time-course and concentration-effect data for all 51 arrays were fitted simultaneously with a common y_0 parameter . The common ymax parameter and the sigmoidicity m parameters for both cisplatin and oxaliplatin were fixed to values found in penultimate nonlinear regression runs. The $t_{1/2}$ parameter for each drug was constrained as a function of the c_{50} and m parameters, as shown in equation (3) below. It was assumed that $y_0=B$ and $y_{max}=E$ for a stimulatory drug; and $y_0=E$ and $y_{max}=B$ for an inhibitory drug. The derivation of Equation (3) is provided as supplementary material*.

$$t_{1/2} = \frac{t * \ln 2}{\ln\left(1 + \left(\frac{c}{c_{50}}\right)^{|m|}\right)} \qquad (3)$$

In Equation (3), t is the time, 16 hours; c is the concentration of drug at that point (32 μM for oxaliplatin and 25 μM for cisplatin). This modeling "trick", where $t_{1/2}$ was set as a function of m and c_{50}, forced the predicted time-course profile and the predicted concentration-effect pattern for each drug to cross at time=16 h and concentration = previously-estimated IC_{90}. Thus a typical nonlinear regression fit of data from 51 arrays for one gene when both drugs are active resulted in 5 final parameter estimates (y_0, $t_{1/2}$ for cisplatin, $t_{1/2}$ for oxaliplatin, c_{50} for cisplatin, c_{50} for oxaliplatin) with accompanying standard errors; and 3 parameter estimates from the penultimate run (ymax, m for cisplatin, m for oxaliplatin). The SAS code with extensive comments, for a typical run for the data from 51 arrays for one gene, MDFIC, is included in supplementary material*.

Also, Figure 1 shows an example of the simulated effect of cisplatin on the PLK gene, to give an illustration of how the time and concentration data cross and share a common design point.

Discussion

From the 34 genes fitted, we selected 18 genes for presentation in this report, based on quality of fit (visual graphical fit) and known importance of the gene involved.
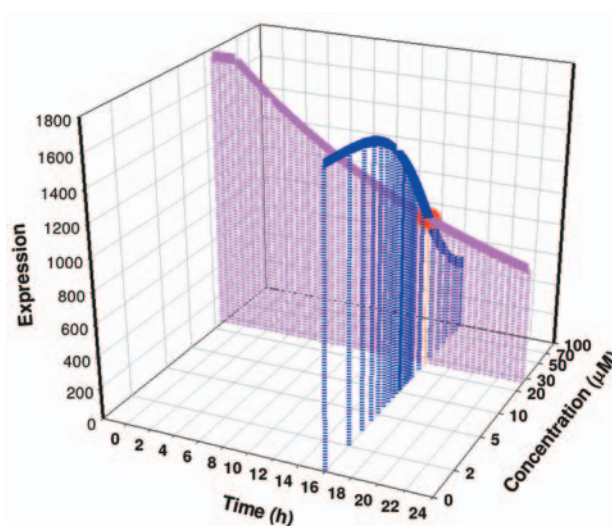

Figure 1. *Simulation of gene expression for gene PLK after cisplatin exposure. Pink line: the time-course profile. Blue line: the concentration-effect profile. Large red sphere: intersection point.*

The parameter estimates obtained after the fits for the 18 genes are presented in Table IV. The fits for the 18 genes are also represented in two figures, Figures 2 (time-course) and Figure 3 (concentration-effect). However, as described above, all of the data (time-course and concentration-effect) for each gene (51 arrays) was fit simultaneously.

Ideally, when both cisplatin and oxaliplatin are active against a particular gene, we should be able to estimate all 8 parameters listed in Table IV for the specific gene, simultaneously, and with an accompanying precision estimate (standard error or 95% confidence interval). However, because of the limited data per gene in this study (one drug time-course profile, one concentration-effect pattern, with one intersection point), we needed to use the constraint of Equation 3, which made two of the eight model parameters nonidentifiable. This means that for each gene, only two of the three parameters ($t_{1/2}$, c_{50}, m) for each drug can be independently estimated; the third parameter is constrained to be calculated from the other two. We chose the least important of the three parameters, the m sigmoidicity parameter, to be not estimated for each drug in the final nonlinear regression run. Although the y_{max} parameter is important and biologically meaningful, from our past experience, it is often estimated with poor precision. Therefore the estimation of this parameter was also excluded from the final run. Even though more design points and more replicates would have been useful for this study, the cost of Affymetrix microarrays precluded expanding the experimental design. This present study is one of the larger studies of its kind to date.

It is possible that more complex nonmonotonic models, with both rising and falling phases with time and/or concentration, might have been more appropriate for some

Table IV. *Parameter estimates for selected genes.*

Symbol	Probeset	Gene title	Cat.	y_0	y_{max}	Cisplatin-specific parameter estimates			Oxaliplatin-specific parameter estimates		
						m	$t_{1/2}$ h	c_{50} μM	m	$t_{1/2}$ h	c_{50} μM
ACYP1	33334_at	acylphosphatase 1, erythrocyte (common) type	A	314 291 to 336	720	1.50	8.38 5.19 to 11.6	12.7 8.05 to 20.1			
TANK	39742_at	TRAF family member-associated NFκB activator	A	52.4 45.6 to 59.1	257	0.197	20.2 12.8 to 27.6	122 10.9 to 1382			
FER1L3	34678_at	fer-1-like 3, myoferlin (*C. elegans*)	B	78.9 65.9 to 91.9	326				2.15	2.49 1.78 to 3.20	4.05 2.24 to 7.34
HIST1H2	38576_at	histone 1, H2bd	B	60.8 55.2 to 66.3	206				1.02	5.79 4.22 to 7.36	5.71 3.14 to 10.4
DNAJC9	41569_at	DnaJ (Hsp40) homolog, subfamily C, member 9	C	1201 1119 to 1282	240				–0.599	8.70 6.63 to 10.8	6.58 3.26 to 13.3
KIF11	40726_at	kinesin family member 11	C	1033 944 to 1121	207				–0.653	5.73 4.29 to 7.16	2.09 0.878 to 4.98
PLAB	1890_at	prostate differentiation factor	D	338 269 to 407	25446	0.762	41.4 34.8 to 48.1	118 92.7 to 150	0.644	29.0 24.6 to 33.5	105 78.8 to 140
CDKN1A	2031_s_at	cyclin-dependent kinase inhibitor 1A (p21, Cip1)	D	653 562 to 744	14271	0.701	15.5 13.4 to 17.5	23.4 18 to 30.4	0.549	12.8 11.1 to 14.5	17.7 12.3 to 25.6
FDXR	40336_at	ferredoxin reductase	D	812 707 to 917	12901	0.610	23.8 20.2 to 27.4	58.8 43.3 to 80.0	0.322	23.9 20.3 to 27.5	164 91.2 to 294
TP53I3	36079_at	tumor protein p53 inducible protein 3	D	382 312 to 451	12446	0.602	33.3 27.6 to 39.1	117 83.5 to 164	0.418	24.7 20.6 to 28.8	124 75.6 to 204
CYFIP2	37579_at	cytoplasmic FMR1 interacting protein 2	D	321 282 to 360	3322	0.498	28.6 23.5 to 33.8	112 72.5 to 174	0.420	18.6 15.6 to 21.7	52.3 31.0 to 88.1
DDB2	1243_at	damage-specific DNA binding protein 2, 48 kDa	D	429 373 to 485	3354	0.493	15.6 12.6 to 18.5	23.1 13.4 to 39.6	0.135	18.9 15.0 to 22.7	169 23.0 to 1247
STAT3	39708_at	signal transducer and activator of transcription 3 (acute-phase response factor)	D	267 230 to 304	2869	0.754	16.6 13.5 to 19.7	26.8 19.1 to 37.6	0.621	27.8 21.9 to 33.8	101 66.6 to 153
ANXA4	37374_at	annexin A4	D	208 175 to 240	2754	0.582	30.6 24.0 to 37.1	104 66.8 to 161	0.734	26.4 20.3 to 32.4	77.5 52.9 to 113
BTG2	36634_at	BTG family, member 2	D	391 349 to 433	2785	1.41	4.71 4.13 to 5.29	5.06 4.03 to 6.34	0.899	5.27 4.50 to 6.04	3.56 2.41 to 5.26
MCM6	40117_at	MCM6 minichromosome maintenance deficient 6 (MIS5 homolog, *S. pombe*) (*S. cerevisiae*)	E	2861 2640 to 3082	572	–0.460	15.2 10.7 to 19.7	21.4 8.59 to 53.3	-0.960	4.17 3.48 to 4.85	2.15 1.32 to 3.51
PLK	37228_at	polo-like kinase (Drosophila)	E	1755 1645 to 1866	351	–1.72	13.2 9.48 to 16.8	21.2 16.7 to 27.0	-1.44	9.82 7.46 to 12.2	19.2 14.5 to 25.3
CCNB1	1945_at	cyclin B1	E	1473 1375 to 1572	295	–1.19	23.6 13.4 to 33.7	38.3 24.3 to 60.5	-0.713	7.47 6.01 to 8.94	5.73 3.38 to 9.71

Cat.: category of gene. A: cisplatin increases gene expression, oxaliplatin doesn't change. B : cisplatin doesn't change, oxaliplatin increases. C: cisplatin doesn't change, oxaliplatin decreases. D: both increase. E: both decrease. y_0: expression at control level (no drug). y_{max}: expression at maximum level of drug. m: hill slope coefficients for cisplatin and oxaliplatin. $t_{1/2}$: half lives of effect for cisplatin and oxaliplatin. c_{50}: concentrations at 50% effect for cisplatin and oxaliplatin respectively, inhibitory or excitatory depending on the gene category. For ymax, m: the values presented are the values fixed in the final nonlinear regression run. For y_0, $t_{1/2}$, c_{50}: the values presented are the estimate and 95% confidence interval.

genes. However, such patterns were never obvious from visual inspections of the recorded gene expression profiles. In addition, our frugal experimental design precluded the routine fitting of more complex models to the gene expression data.

Changes in gene expression were always in the same direction for the time-course profile and concentration-effect pattern for each specific gene. We found five different patterns in gene expression: a) no change with oxaliplatin but increase with cisplatin (*e.g.*, ACYP1, TANK). b) no change with cisplatin but increase with oxaliplatin (*e.g.*, FER1L3, HIST1H2). c) no change with cisplatin but decrease with oxaliplatin (*e.g.*, DNAJC9, KIF11). d) increase with both
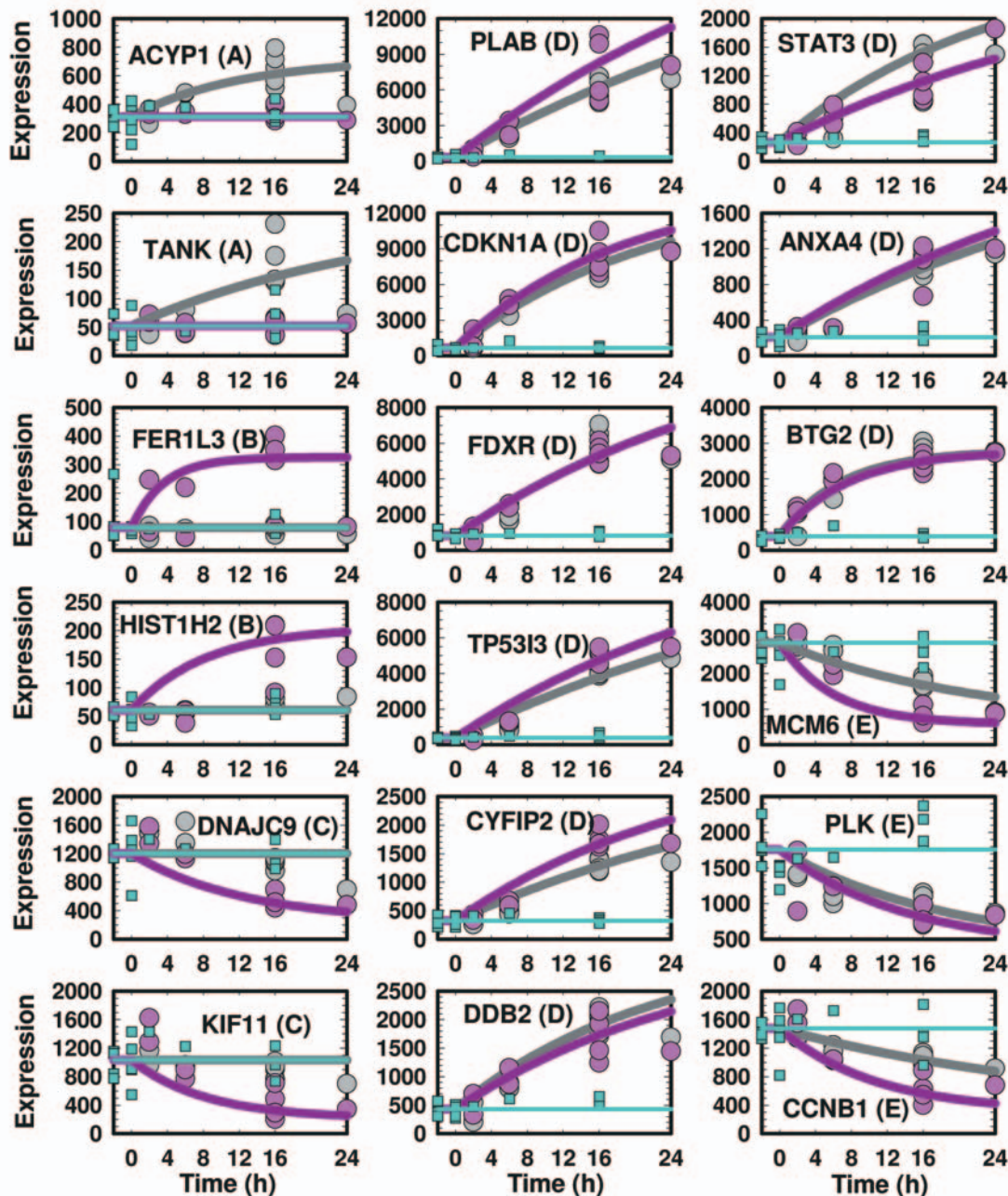
Figure 2. *Time-course of gene expression. Circles, squares: observed data points; curves: predicted fits from the nonlinear regression. Blue squares and curves: control; grey circles and curves: cisplatin; pink circles and dark pink curves: oxaliplatin. Drug concentration is IC_{90} (or drug-free controls); model fitted is an exponential increase or decrease. The letter between parentheses indicates the category of the gene.*

drugs (*e.g.*, PLAB, CDKN1A, FDXR, TP53I3, CYFIP2, DDB2, STAT3, ANXA4, BTG2). e) decrease with both drugs (*e.g.*, MCM6, PLK, CCNB1).

Figure 2 shows the time-course of expression of the 18 selected genes. The expression of genes after exposure only to drug-free medium does not change in time, and the replicates were reasonably consistent for all genes. The exponential model, Equation 1, fit well the change of gene expression over time for all genes. However, for three genes (ACYP1, TANK, and FER1L3) the last time point (24 h) seemed to return to baseline. Overall, when cisplatin and oxaliplatin both had an effect on gene expression, the time course profiles for the two drugs were very similar. Possible exceptions (based on 95% confidence intervals) include STAT3 (the $t_{1/2}$ for cisplatin is 1.7 fold less than that for oxaliplatin), MCM6 (the $t_{1/2}$ for oxaliplatin is 3.6-fold less
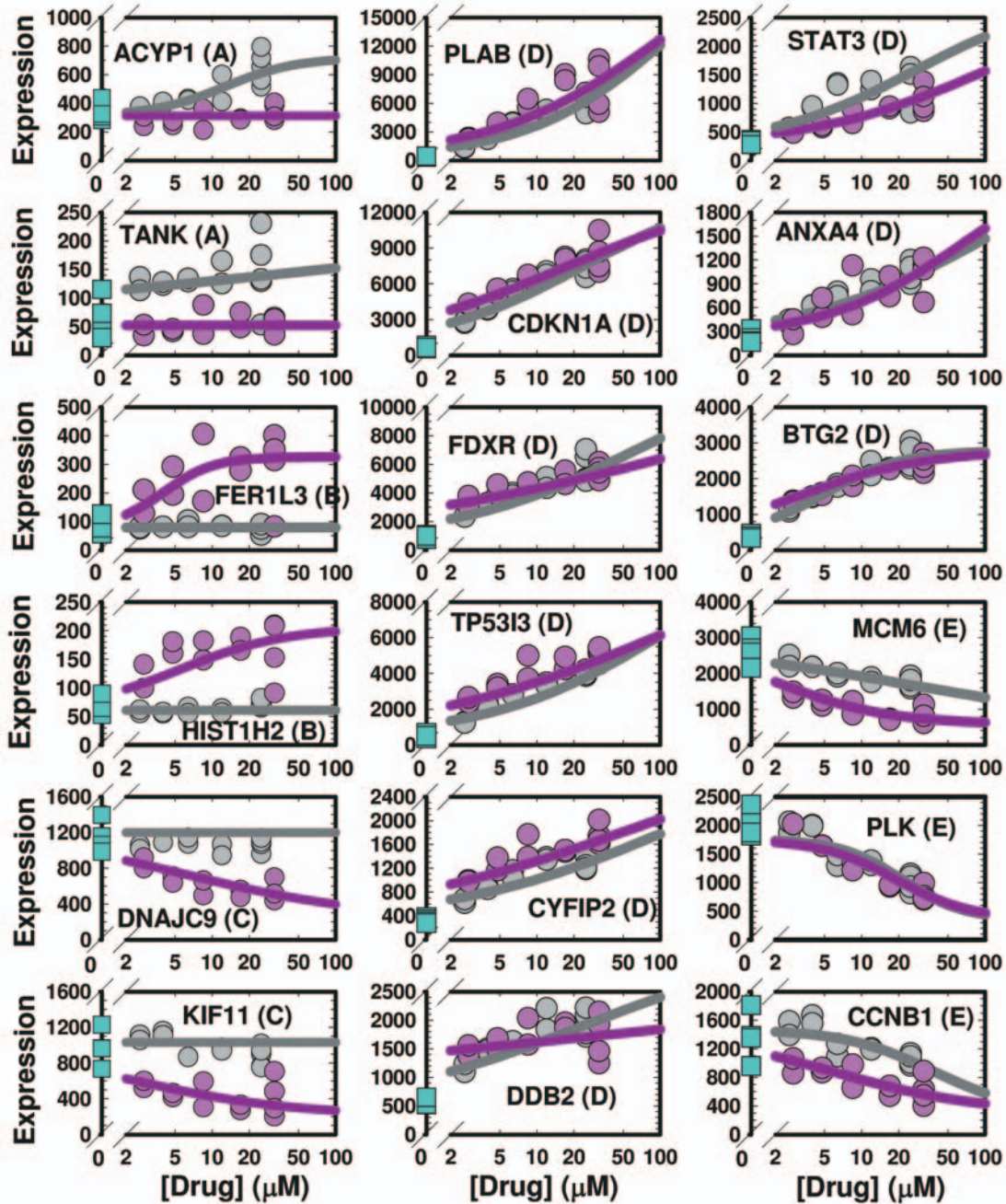
Figure 3. *Concentration-effect of gene expression. Circles, squares: observed data points; curves: predicted fits from the nonlinear regression. Blue squares: control; grey circles and curves: cisplatin; pink circles and dark pink curves: oxaliplatin. Effect is measured at 16 hours. Model fitted is a Hill model. The letter between parentheses indicates the category of the gene.*

than that for cisplatin), and CCNB1 (aka cyclin B1; the $t_{1/2}$ for oxaliplatin is 3.2-fold less than that for cisplatin).

Figure 3 shows the concentration-effect of expression of the 18 selected genes. The Hill model fitted the change of expression *versus* drug concentration well for most genes. Overall, when cisplatin and oxaliplatin both had an effect on gene expression, the concentration-effect patterns for the two drugs were very similar. Possible exceptions (based on 95% confidence intervals) include STAT3 (the c_{50} for cisplatin is 3.8-fold less than that for oxaliplatin), MCM6 (the c_{50} for oxaliplatin is 10-fold less than that for cisplatin), and CCNB1 (aka cyclin B1; the c_{50} for oxaliplatin is 6.7-fold less than that for cisplatin). Although the 95% confidence intervals for the c_{50} for cisplatin and for oxaliplatin for the

FDXR gene (aka ferredoxin reductase) do not overlap, almost all of the data does overlap (Figure 3); the "statistically significant" difference in extrapolated c_{50}s is unlikely to be reproducible or meaningful. Note that STAT3, MCM6 and CCNB1 all showed significant differences in the same direction in key parameters between cisplatin and oxaliplatin for both the time-course profile and the concentration-effect pattern.

For cisplatin, the c_{50}s ranged from 5.06 μM (BTG2) to 122 μM (TANK) and the half-lives ranged from 4.71 h (BTG2) to 41.4 h (PLAB). The steepest hill curve was for PLK ($m = -1.72$), and the shallowest for TANK ($m = 0.197$). BTG2 was the gene affected the most potently as well as the most rapidly by cisplatin.

For oxaliplatin, the c_{50}s ranged from 2.09 μM (KIF11) to 169 μM (DDB2) and the half-lives ranged from 2.49 h (FER1L3) to 29 h (PLAB). The steepest hill curve was for FER1L3 ($m = 2.15$), and the shallowest for DDB2 ($m = 0.135$). KIF11 was the gene affected the most potently by oxaliplatin, and FER1L3 was the gene affected the most rapidly by oxaliplatin. The long half lives found for some of the genes suggest it would be interesting to conduct experiments for a longer period than the 24 hours studied here.

The gene whose expression was the most increased was PLAB (category D). It increased by roughly 25,000 expression units ($y_0 = 338$, $y_{max} = 25446$), or 74-fold. The gene whose expression was the most decreased was MCM6 (category E). It decreased by roughly 2,300 expression units ($y_0 = 2861$, $y_{max} = 572$), or 4-fold. It is to be noted that all five genes with expression decreases (categories C and E) have similar proportional decreases, around 4-fold; this is probably due to our selecting the genes so that the median expression at 16 hours was at least 100 expression units.

Gene expression profiling of human ovarian carcinoma A2780 cells after exposure to either cisplatin or oxaliplatin resulted in the up or down-regulation of several genes belonging to apoptosis, DNA damage, cell cycle progression, signal transduction and metabolic pathways. In this present study 18 genes were selected for modeling and were presented in 5 (A-E) categories as described above. The known functions of these genes are briefly described in supplementary material*.

Conclusion

In summary, gene expression profiles obtained from time-course and concentration-effect microarray experiments were quantitively characterized by fitting models to the data with non-linear regression. Though it doesn't take into account eventual gene co-regulation, for such data, this method offers an attractive alternative to more standard methods, such as clustering. Fitted curves improve visualization of time-course and concentration-effect patterns, and the associated model parameters provide quantitative summaries of the data. The modeling approach presented here provides a unique and different perspective and clearly complements traditional methods to visualize and analyze microarray data. This approach can also help in understanding the molecular mechanisms of the agents, and the timing of their actions.

Acknowledgements

References

1 Rouzier R, Perou CM, Symmans WF *et al*: Breast cancer molecular subtypes respond differently to preoperative chemotherapy. Clin Cancer Res *11(16)*: 5678-5685, 2005.

2: Febbo PG, Thorner A, Rubin MA *et al*: Application of oligonucleotide microarrays to assess the biological effects of neoadjuvant imatinib mesylate treatment for localized prostate cancer. Clin Cancer Res *12(1)*: 152-158, 2006.

3 Cheok MH, Yang W, Pui Ch *et al*: Treatment-specific changes in gene expression discriminate *in vivo* drug response in human leukemia cells. Nat Genet *34(1)*: 85-90, 2003. Erratum in: Nat Genet *34(2)*: 231.

4 Almon RR, Dubois DC, Jin JY and Jusko WJ: Pharmacogenomic responses of rat liver to methylprednisolone: an approach to mining a rich microarray time series. AAPS J *7(1)*: E156-194, 2005.

5 Hu J, Kapoor M, Zhang W, Hamilton SR and Coombes KR: Analysis of dose-response effects on gene expression data with comparison of two microarray platforms. Bioinformatics *17*: 3524-3529, 2005.

6 Zhou Y, Gwadry FG, Reinhold WC *et al*: Transcriptional regulation of mitotic genes by camptothecin-induced DNA damage: microarray analysis of dose- and time-dependent effects. Cancer Res *62(6)*: 1688-1695, 2002.

7 Tusher VG, Tibshirani R and Chu G: Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA *98(9)*: 5116-5121, 2001.

8 Datta S: Empirical Bayes screening of many *p*-values with applications to microarray studies. Bioinformatics *21(9)*: 1987-1994, 2005.

9 Tamayo P, Slonim D, Mesirov J *et al*: Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci USA *96(6)*: 2907-2912, 1999.

10 Eisen MB, Spellman PT, Brown PO and Botstein D: Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA *95(25)*: 14863-14868, 1998.

11 Dudoit S, Fridlyand J and Speed TP: Comparison of discrimination methods for the classification of tumors using gene expression data. J Am Stat Assoc *97(457)*: 77-87, 2002.

12 Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M and Haussler D: Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics *16(10)*: 906-914, 2000.

13 Wei JS, Greer BT, Westermann F *et al*: Prediction of clinical outcome using gene expression profiling and artificial neural networks for patients with neuroblastoma. Cancer Res *64(19)*: 6883-6891, 2004.

14 De Hoon MJ, Imoto S and Miyano S: Statistical analysis of a small set of time-ordered gene expression data using linear splines. Bioinformatics *18(11)*: 1477-1485, 2002.

15 Bar-Joseph Z: Analyzing time series gene expression data. Bioinformatics *20(16)*: 2493-2503, 2004.

16 Kim SY, Imoto S and Miyano S: Inferring gene networks from time series microarray data using dynamic Bayesian networks. Briefings in Bioinformatics *4(3)*: 228-235, 2003.

17 Moller-Levet C, Cho KH, Yin H and Wolkenhauer O: Clustering of gene expression time-series. second year report, UMIST, 2003.

18 Zhou Y, Gwadri FG, Reinhold WC *et al*: Transcriptional regulation of mitotic genes by camptothecin-induced DNA damage: microarray analysis of dose- and time-dependant effects. Cancer Research *62*: 1688-1695, 2002.

19 Peddada SD, Lobenhofer EK, Li L, Afshari CA, Weinberg CR and Umbach DM: Gene selection and clustering for time-course and dose-response microarray experiments using order-restricted inference. Bioinformatics *19(7)*: 834-841, 2003.

20 Eckel JE, Gennings C, Chinchilli VM, Burgoon LD and Zacharewski TR: Empirical Bayes gene screening tool for time-course or dose-response microarray daya. J Biopharm Stat. *14(3)*: 647-670, 2004.

21 Costa IG, de Carvalho F and de Souto M: A Symbolic Approach to Gene Expression Time Series Analysis. VII Brazilian Symposium on Neural Networks (SBRN'02) p. 25, 2002.

22 Lukashin AV and Fuchs R: Analysis of temporal gene expression profiles: clustering by simulated annealing and determining the optimal number of clusters. Bioinformatics *17(5)*: 405-414, 2001.

23 Ramoni MF, Sebastiani P and Kohane IS: Cluster analysis of gene expression dynamics. Proceedings of the National Academy of Science USA *99(14)*: 9121-9126, 2002.

24 Ong IM, Glasner JD and Page D: Modelling regulatory pathways in *E. coli* from time series expression profiles. Bioinformatics *18*: 241S-248S, 2002.

25 Bar-Joseph Z, Gerber GK, Gifford DK, Jaakkola DS and Itamar S: A new approach to analyzing gene expression time series data. (2002). J Computational Biology *10(3)*: 341-356, 2003.

26 Li H, Luan Y, Hong F and Li Y: Statistical methods for analysis of time course gene expression data. Frontiers in Bioscience a90-98, 2002.

27 Luan Y and Li H: Clustering of time-course gene expression data using a mixed-effects model with B-splines. Bioinformatics *9(4)*: 474-482, 2003.

28 Schliep A, Steinhoff C and Schönhuth A: Robust inference of groups in gene expression time-courses using mixtures of HMM. Proceedings of the ISMB 2004 Bioinformatics *20(Suppl 1)*: I283-I289, 2003.

29 Schliep A, Schonhuth A and Steinhoff C: Using hidden Markov models to analyze gene expression time course data. Bioinformatics *19(Suppl 1)*: i255-263, 2003.

30 Wu FX, Zhang WJ and Kusalik AJ: Model-Based Clustering with Genes Expression Dynamics for Time-Course Gene Expression Data. Fourth IEEE Symposium on Bioinformatics and Bioengineering (BIBE'04) p. 267, 2004.

31 Gui J and Li H: Mixture Functional Discriminant Analysis for Gene Function Classification Based on Time Course Gene Expression Data. Proceedings of the Joint Statistical Meeting (Biometrics Section), 2003.

32 Xu XL, Olson JM and Zhao LP: A regression-based method to identify differentially expressed genes in microarray time course studies and its application in an inducible Huntington's disease transgenic model. Hum Mol Genet *11(17)*: 1977-1985, 2002.

33 Aach J and Church G: Aligning gene expression time series with time warping algorithms. Bioinformatics *17*: 495-508, 2001.

34 Whiteside MA, Chen DT, Desmond RA, Abdulkadir SA and Johanning GL: A novel time-course cDNA microarray analysis method identifies genes associated with the development of cisplatin resistance. Oncogene *23(3)*: 744-752, 2004.

35 Park T, Yi SG, Lee S *et al*: Statistical tests for identifying differentially expressed genes in time-course microarray experiments. Bioinformatics *19(6)*: 694-703, 2003.

36 Azzini I, Dellnna R, Ciocchetta F *et al*: Simple methods for peak detection in time series microarray data. Workshop notes of CAMDA 2004 critical Assessment of Microarray Data Durham, North Carolina, 2004.

37 Wichert S, Fokianos K and Strimmer K: Identifying periodically expressed transcripts in microarray time series data. Bioinformatics *20(1)*: 5-20, 2004.

38 Tan Q, Brusgaard K, Kruse T *et al*: Correspondence analysis of microarray time-course data in case-control design. J Biomed Inform *37(5)*: 358-65, 2004.

39 Simon I, Siegfried Z, Ernst J and Bar-Joseph Z: Combined static and dynamic analysis for determining the quality of time-series expression profiles. Nature Biotechnology *23(12)*: 1503-1508, 2005.

40 Hong F and Li H: Functional hierarchical models for identifying genes with different time-course expression profiles. Biometrics *62(2)*: 534-544, 2006.

41 Varma R, Hector S, Greco WR, Clark K, Hawthorn L, Porter C and Pendyala L: Platinum drug effects on the expression of genes in the polyamine pathway – time course and dose response analysis based on Affymetrix gene expression profiling of A2780 ovarian carcinoma cells. Cancer Chemother. Pharmacol *59(6)*: 711-721, 2007.

42 Li C and Wong WH: Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci USA *98(1)*: 31-36, 2001.

43 Irizarry RA, Hobbs B, Collin F *et al*: Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4(2)*: 249-264, 2003.

44 Gentleman RC, Carey VJ, Bates DM *et al*: Bioconductor: open software development for computational biology and bioinformatics. Genome Biol *5(10)*: R80, 2004.

45 Varma R, Hector SM, Clark K, Greco WR, Hawthorn L and Pendyala L: Gene expression profiling of a clonal isolate of oxaliplatin resistant ovarian carcinoma cell line A2780/C10. Oncol Rep *14*: 925-932, 2005.

46 Faessel HM, Slocum HK, Rustum YM and Greco WR: Thymidine and hypoxanthine protection patterns of the folic acid-enhanced synergies for combinations of trimetrexate plus a poly glutamylatable inhibitor of purine or thymidylate synthesis against human ileocecal HCT-8 cells. Int J Oncol *23*: 401-409, 2003.