Gene expression

SPLINDID: a semi-parametric, model-based method for obtaining transcription rates and gene regulation parameters from genomic and proteomic expression profiles

Kavitha Bhasi, Alan Forrest and Murali Ramanathan*

Department of Pharmaceutical Sciences, State University of New York at Buffalo, Buffalo, NY 14260-1200, USA

Received on May 18, 2005; revised on July 25, 2005; accepted on August 10, 2005

Advance Access publication August 11, 2005

ABSTRACT
Purpose: To evaluate a semi-parametric, model-based approach for obtaining transcription rates from mRNA and protein expression.

Methods: The transcription profile input was modeled using an exponential function of a cubic spline and the dynamics of translation; mRNA and protein degradation were modeled using the Hargrove–Schmidt model. The transcription rate profile and the translation, and mRNA and protein degradation rate constants were estimated by the maximum likelihood method.

Results: Simulated datasets generated from the stochastic, transit compartment and dispersion signaling models were used to test the approach. The approach satisfactorily fit the mRNA and protein data, and accurately recapitulated the parameter and the normalized transcription rate profile values. The approach was successfully used to model published data on tyrosine aminotransferase pharmacodynamics.

Conclusions: The semi-parametric approach is effective and could be useful for delineating the genomic effects of drugs.

Availability: Code suitable for use with the ADAPT software program is available from the corresponding author.

Contact: murali@acsu.buffalo.edu

INTRODUCTION
DNA arrays are capable of simultaneously measuring the expression of thousands of genes from single samples and increasingly, large-scale protein measurements are also becoming possible through developments in proteomics methodologies. In the near future, therefore, many large datasets containing the time courses of simultaneous measurements of mRNA and protein will become available.

The primary motivation for measuring a large number of molecular endpoints is to delineate the mechanisms of gene regulation. Expression profiling and proteomics experiments provide a rich snapshot of mRNA and protein levels at each time point. Unfortunately, these techniques do not directly estimate the underlying changes in the transcription rate and gene regulatory processes caused by the actions of drugs or experimental stimuli because transcriptional regulation effects occur in the context of mRNA and protein elimination processes, which differ considerably between gene products. Direct experiments on transcription rates and key gene regulatory parameters are generally not possible in the majority of experimental settings. For example, direct transcription rate measurements are difficult to make and are often experimentally infeasible because the nuclear run-on assay, the usual method of choice for this purpose, can only be done in vitro after extensive cellular fractionation; it is therefore used only when absolutely necessary (Ausubel et al., 2005; Strand et al., 1994). The ability to infer the transcript production profile from array data could therefore prove valuable in a variety of contexts. Likewise, the mRNA and protein half-lives, which determine the dynamics of drug effects are difficult to measure directly; experimental estimation of mRNA and protein half-lives require the use of transcription/translation blockers such as actinomycin and cycloheximide, respectively; these agents are non-specific and can alter the system under investigation (Glynn et al., 1992; Sisler et al., 1967; Tang et al., 1999). Novel modeling techniques based on pharmacokinetics and pharmacodynamics (PK/PD) can therefore play a rich role in the analysis of these pharmacogenomic time series datasets but have not been extensively investigated.

Here, we propose a novel modeling method that yields multiple gene regulation parameters from mRNA and protein profiles. The strategy combines the results from modified cubic spline approximations of the transcription profile with the deterministic Hargrove-Schmidt model (Hargrove and Schmidt, 1989). The overall strategy is thus semi-parametric and allows additional useful information, e.g. the experimentally inaccessible normalized transcript production rate profile to be harvested from expression profiles.

DERIVATIONS AND RESULTS
The pharmacodynamic model for mRNA and protein
The Hargrove–Schmidt model was used to represent the dynamics of messenger RNA and protein. The Hargrove–Schmidt model (Fig. 1) is a two-compartment model that assumes information flow from mRNA (M) to protein (P) via a first-order translation rate constant, $k_T$ and independent, first-order degradation of mRNA and protein with rate constants $k_M$ and $k_P$, respectively (Hargrove and Schmidt, 1989). The M and P are concentrations of mRNA and protein, respectively; the rate of mRNA transcription per unit volume is $R(t)$. With these definitions, the rates of change of

$$
\begin{align*}
\frac{dM}{dt} &= R(t) - \frac{M}{k_M}, \\
\frac{dP}{dt} &= k_T M - \frac{P}{k_P}.
\end{align*}
$$

*To whom correspondence should be addressed at Department of Pharmaceutical Sciences, 543 Cooke Hall, State University of New York at Buffalo, Buffalo, NY 14260-1200, USA
mRNA and protein, $dM/dt$ and $dP/dt$, respectively are described by the differential equations in Equation (1):

$$\frac{dM}{dt} = R(t) - k_M M = k_M \left( \frac{R(t)}{k_M} - M \right), \quad \text{(1a)}$$

$$\frac{dP}{dt} = k_T M - k_P P. \quad \text{(1b)}$$

The ratio $R(t)/k_M$ is referred to as the normalized transcription rate profile.

The non-parametric cubic spline approximation

The non-parametric component of the modeling process uses flexible, relatively ‘non-parametric’ functions such as those based on splines to describe the transcription profiles in combination with the deterministic Hargrove–Schmidt model for describing gene dynamics.

We refer to the overall strategy and its implementation as SPLINDID because it employs functions of splines to describe the transcription rate profile. Specifically, the transcription rate profile, $R(t)$ is modeled as an exponential function of splines of the form,

$$R(t) = e^{\text{Spline}(t)} - 1. \quad \text{(2)}$$

Cubic splines were used for modeling the Spline(t) term in the exponential of Equation (2). A cubic spline is a piecewise polynomial of order 4 (degree 3) that can be represented by (Bar-Joseph et al., 2003; deBoor, 1978; Schumaker, 1981):

$$\text{Spline}(t) = \sum_{i=1}^{n-1} C_{ij} \frac{(t-x_j)^{i-1}}{(i-1)!} \quad \text{for} \quad x_j \leq t \leq x_{j+1},$$

where the $C_{ij}$ are coefficients and the piecewise polynomial terms are defined to be non-zero only between $n$ breakpoints $x_j$ that are strictly increasing.

The exponential function term of $R(t)$ in Equation (2) is always positive; however, this term approaches zero only when the Spline(t) function in the exponent approaches negative infinity. However, in modeling it is necessary to provide initial conditions that frequently are zero, and this is numerically problematic with the exponential term alone. Therefore, the term containing $-1$ was employed in Equation (2) so that an initial condition of $R(0) = 0$ could be imposed.

The implementation of the cubic spline was accomplished by setting the order of spline in a B-spline basis function formulation to 4. This implementation via B-spline basis function minimizes the likelihood of numerical problems that can sometimes arise with the normal equations for splines because the B-spline basis functions produce well-conditioned banded matrices. For spline calculations, three subroutines, DBSINT, DBSOPK and DBSVAL, from the Fortran programming language version of the International Mathematical Statistical Library (IMSL, Visual Numerics Inc., San Ramon, CA) for the Unix platform were used. The positions of the knots of the B-splines were optimized with the DBSOPK subroutine; the spline coefficients were computed using the DBSINT subroutine. The outputs from the DBSINT subroutine were provided as input to DBSVAL for interpolation.

Implementation of modeling strategy

We explicitly integrated Equation (1a) using the integration factor method, which results in the following equation for $M$.

$$M = M_0 e^{-k_M t} + e^{-k_M t} \int_0^t e^{k_M \tau} R(\tau) d\tau. \quad \text{(3)}$$

The $M_0$ term is the constant of integration and represents the initial value of $M$.

The integrated expression for $M$ (Equation 3) was incorporated into the right-hand side of Equations (1a) and (1b). The system consisting of integral term in Equation (3) and the resultant forms of Equations (1a) and (1b) was integrated within the ADAPT PK/PD systems analysis software (D’Argenio and Schlumitzky, 1997) in the first stage of the heuristic estimation procedure.

The input data consisted of values of the mRNA time profile, $M(t)$, and the protein time profile, $P(t)$. The principal model outputs were estimates for the normalized transcription rate profile, i.e. the ratio $R(t)/k_M$, and parameter estimates for $k_T$ and $k_P$.

Over the infinite time horizon $t = 0$ to $t = \infty$, the values of $R(t)$ and $k_M$ are not completely independent of each other because mass conservation requires that the total production of mRNA equal the total degradation of mRNA; i.e. $\langle R \rangle$, the total area under the transcript production $R(t)$ curve from $t = 0$ to $t = \infty$ must equal the total area under the mRNA degradation curve from $t = 0$ to $t = \infty$. Likewise, the total translation of mRNA must match the total protein degraded. Mathematically, these mass conservation requirements can be expressed as:

$$\langle R \rangle = k_M \langle M \rangle, \quad \text{(4a)}$$

$$k_T \langle M \rangle = k_P \langle P \rangle. \quad \text{(4b)}$$

In Equations (4a) and (4b), $\langle M \rangle$ and $\langle P \rangle$ are the areas under the mRNA and protein curves (from $t = 0$ to $t = \infty$), respectively. The constraint due to Equation (4b) was imposed for avoiding covariation between $k_T$ and $k_T$: the model was constrained to appropriately keep the magnitude of $k_T$ greater than $k_P$ when $P$ was greater than $M$ and vice versa.

Several nested modeling runs that varied the number of estimated ordinates for interpolation by the spline function were conducted; the time points corresponding to the estimated ordinates were equally spaced over the interval of the data. Model selection was user driven and based on a combination of graphical visualization by the user and the Akaike Information Criterion (Akaike, 1974). The models identified by the visual inspection and the Akaike
Information Criteria were generally the same or only differed incrementally in complexity.

From the selected model, the parameter estimates for the ordinates for interpolating the spline [which describes the transcription rate profile \( R(t) \)] and the parameters \( k_{3M}, k_T \) and \( k_p \) of the Hargrove–Schmidt model were obtained from ADAPT. The estimated values of \( R(t) \) and \( k_{3M} \) were then used to compute the ratio \( R(t)/k_{3M} \).

**Parameter estimation**

The parameter estimation procedures employed the maximum likelihood method in the ADAPT software package for the UNIX platform (D’Argenio and Schlumitzky, 1997). The variance model employed assumed that residual error standard deviation, \( \sigma_i \), was related to the true value of each output \( Y_i \), as approximated by its fitted value \( \hat{Y}_i \), via the relationship: \( \sigma_i^2 = \text{SD}_{\text{intercept}}^2 + (\text{SD}_{\text{slope}} \cdot \hat{Y}_i)^2 \). The SD_{slope} is a measure of precision whereas SD_{intercept} is a measure of sensitivity.

**Generation of simulated datasets from signaling models**

The ground truth in gene expression and proteomics experiments is rarely known with certainty and we employed simulated data to assess the performance of the method. The parameter values used for the simulation and the results from simulations without noise were used as reference or ‘true’ values to which the performance of the proposed method was compared.

Three signal transduction models, the stochastic, the transit compartment (Sun and Jusko, 1998) and the dispersion models (Ramanathan, 2002) were used to generate simulated data for transcriptional signaling, i.e. for the \( R(t) \) profiles. These models are prototypical examples of different model classes: the stochastic model is a linear compartmental model, the transit compartment model is compartmental model containing non-linearities whereas the dispersion model is a linear, lumped parameter model.

The overall model used to generate simulated mRNA and protein dynamics (Fig. 2D) incorporated first-order drug elimination and each of the three signaling models (stochastic, transit compartment or dispersion models) was linked to a Hargrove–Schmidt module (Ramanathan, 2002). All three signal transduction models were provided with a drug input as a bolus that resulted in an initial concentration of 100 concentration units and was eliminated with a first-order elimination rate constant, \( k_{3M} \), of 0.5 h\(^{-1} \). The parameters for the Hargrove–Schmidt modules for the three signal transduction models were identical: the mRNA degradation rate constant, \( k_{3M} \), was set to 1.0 h\(^{-1} \); the protein degradation rate constant, \( k_p \), was set to 0.5 h\(^{-1} \); and the translation rate constant, \( k_T \), was set to 1.5 h\(^{-1} \); the initial mRNA and protein concentrations were set equal to zero. Noisy data with a coefficient of variation (CV, defined as the ratio of the standard deviation to the mean) of 0.2 were obtained in triplicate at the following time points: 0, 1.5, 3, 4.5, 6, 7.5, 9, 10.5, 12, 13.5, 15, 16.5 and 18 h. The mRNA and protein levels obtained from noisy simulations with these signal transduction models were used as input for the SPLINID method.

The stochastic or tanks-in-series model (Fig. 2A) is a series of well-stirred compartments with identical residence times (Sun and Jusko, 1998) and its bolus or impulse response is an Erlang distribution, a special case of the gamma distribution, a special case of the gamma distribution with shape parameter \( \alpha \) and scale parameter \( \beta \). The Erlang distribution represents the time for completing a sequence of \( N \) tasks whose durations are identical exponential probability distributions. The transit compartment model (Fig. 2B) is an extension of the tanks-in-series model in which one (or more) of the signaling compartments incorporates non-linearity via a Hill equation. The input for both models was modulated by non-equilibrium receptor binding with on and off rates denoted by \( k_1 \) and \( k_{-1} \), respectively (Sun and Jusko, 1998). The stochastic and transit compartment models used the following parameters, defined in the legend to Figure 2: \( k_1 = 0.5 \) (concentration units \( \cdot \) h\(^{-1} \)), \( k_{-1} = 2 \) h\(^{-1} \); \( \tau = 0.75 \) h; the initial receptor concentration was set to 10 concentration units; all remaining initial conditions were set equal to zero. The Hill coefficient, \( h \), was set to 1.2 for the transit compartment model.

We have previously developed a lumped parameter, non-compartmental model called the dispersion model (Fig. 2C) and demonstrated its ability to describe signaling dynamics (Ramanathan, 2002). The dispersion model element is described by the following non-dimensional partial differential equation:

\[
D_N \frac{\partial C}{\partial Z^2} = \frac{\partial C}{\partial Z} - R_N C = \frac{\partial C}{\partial t}.
\]

In Equation 5, \( C \) is the concentration of the signal, \( Z \) is the non-dimensional distance (defined as \( x/L \), where \( x \) is the distance and \( L \) is...
the length of the system) and $T$ is the non-dimensional time (defined as $\tau T_\text{p}$, where $t$ is the time and $T_\text{p}$ is the residence time of the system). The non-dimensional axial dispersion number $D_\text{N} = D/uL$, where $D$ is the diffusion coefficient, $u$ is the signal velocity and $L$ is the length, measures the rate of axial dispersion relative to the rate of convective signal transfer and correspondingly, the non-dimensional number $R_\text{N} = kL/u$, where $k$ is the elimination rate constant, measures the rate of elimination relative to the rate of convective signal transfer. The dispersion equation was solved by numerical inversion of the output Laplace transform using a publicly available Fortran implementation of Weeks’ method (Garbow et al., 1988; Weeks, 1966) as previously described (Ramanathan, 2002). The following parameter values were used: $D_\text{N} = 0.2$, $R_\text{N} = 0$, $\tau_\text{D} = 3$, $\alpha = 1$.

We assumed closed boundary conditions that do not allow signaling molecules to leave the system. The mathematical description of the closed boundary conditions for $T = 0$ is:

$$\frac{\partial C}{\partial Z} = 0 \quad \text{at } Z = 0, \quad C - D_N \frac{\partial C}{\partial Z} = C_{\text{in}} \quad \text{and at } Z = 1, \quad \frac{\partial C}{\partial Z} = 0. \quad (6)$$

The $C_{\text{in}}$ term is the drug input concentration. These boundary conditions are reasonable approximations for the many signaling cascades initiated by ligand binding to cell surface receptors and whose effects are mediated by the binding transcription factors to DNA.

**Evaluation of the semi-parametric approach**

**Assessing the ability of the model to fit simulated data**

As a first step in evaluating the performance of the approach, we assessed its ability to fit the simulated data. The inputs to the model were simulated mRNA and protein time profiles.

The results from simultaneously fitting the model to the mRNA and protein data are summarized in Figure 3A and 3B for the stochastic model, Figure 3C and D for the transit compartment model and Figure 3E and F for the dispersion model datasets. In each case, the selected models contained seven estimated ordinates for the spline interpolation. For all three simulated datasets, the SPLINDID approach provided excellent fits to the mRNA and protein time profiles as assessed by the approach of the mRNA and protein curves to the true values obtained from simulations without noise. The standardized residuals were normally distributed ($p$-values all $>0.05$ in a one-way Kolmogorov Smirnov test; data not shown).

**Obtaining translational and protein degradation rate constants**

As the next step in evaluating the performance of the approach, we assessed its ability to predict the translational ($k_T$) and protein degradation ($k_P$) rate constants employing the noisy mRNA and protein datasets simulated from the stochastic, transit compartment and dispersion models as inputs. The values of $k_T$ and $k_P$ used in the simulations were treated as ‘true’ values for comparison purposes.

Table 1 summarizes the results; the parameters estimates $k_T$ and $k_P$ and the corresponding CV values are shown. The estimated values of both parameters were similar to their true values and the CV values were all <15%. This indicates that the SPLINDID approach provides accurate estimates of these parameters, which as indicated earlier, are experimentally difficult to access.

**Obtaining estimates for the normalized transcriptional rate profile**

In the next step of evaluating the performance of the approach, we assessed its ability to predict the normalized transcriptional rate profiles, $R(t)/k_M$. The $R(t)/k_M$ can be viewed as the time profile of mRNA levels that would be obtained if each instantaneous value of $R(t)$ were to hypothetically reach steady state. It was necessary to focus on the normalized transcriptional profile rather than the individual estimates for the $R(t)$ profile and $k_M$ because of the covariation between the $R(t)$ and $k_M$. Mathematically, the cause of the covariation between the $R(t)$ and $k_M$ can be understood from the mRNA mass balance in Equation (1a): a given right-hand-side value can be obtained in a multitude of ways because increases (decreases) in mRNA input via the $R(t)$ term can be offset by appropriate increases (decreases) in mRNA degradation rate via the $k_M$ term.

The accuracy of the normalized transcriptional rate estimates was confirmed by comparison to the values from noise-free simulations in several numerical experiments. Figure 4A–C compare the $R(t)/k_M$ profiles obtained by fitting the simulated mRNA from the stochastic, transit compartment and dispersion models, respectively. The profiles obtained using the SPLINDID approach closely resemble those predicted by the noise-free simulation.

The values predicted by SPLINDID were slightly time shifted compared to the simulated values; however, the magnitude of the shift was relatively modest. In additional numerical experiments (not shown), the extent of the shift was found to be dependent on time and the value of $k_M$. Additional investigations are currently underway to identify the causes of the time shift.

**Application of method to pharmacodynamics**

We further assessed the performance of the semi-parametric method against the experimental results from Jusko and co-workers on the dynamics of tyrosine aminotransferase (TAT, an enzyme that catalyzes the transamination of tyrosine to hydroxyphenylpyruvate oxoacid in the liver and its transcription is known to be specifically increased by glucocorticoid treatment) mRNA and protein in livers of methylprednisolone treated rats (Xu et al., 1995). This dataset was suitable for testing the proposed approach because it reports parallel measurements of the dynamics of an mRNA and its cognate protein.

Figure 5A and 5B show the TAT mRNA and protein data, respectively, and the corresponding profiles obtained using the SPLINDID approach for the selected model containing seven estimated ordinates for spline interpolation. Visually, the semi-parametric modeling approach provided an excellent fit to the mRNA and protein data because the fitted curve satisfactorily captures the key characteristics of the dataset. Figure 5C shows the predicted normalized transcription rate profile based on the SPLINDID approach.

**DISCUSSION**

The objective of this report was to develop a novel semi-parametric, model-based analysis method for time series data obtained from gene expression profiling experiments and to evaluate the method by applying it to a test bed of data from genomic studies.
In general, the semi-parametric model is robust, requires relatively few assumptions regarding the transcription rate profile, and by providing a smooth function through the middle of the mRNA and protein data, allows the data ‘to speak for itself’. The functions are a visual guide that can potentially facilitate subsequent selection of mechanistic models. The approach provides a broad range of quantitative information including the normalized transcription rate profile and the rate constants for protein degradation and translation, all of which are difficult to access experimentally.

Although we have specifically focused on time series containing both mRNA and protein dynamics, the methods can be used for datasets containing only mRNA dynamics with relatively minor changes. With the necessary changes, the approach could be used for analyzing experiments in which only mRNA profiles were acquired and for describing the regulation of genes, e.g. rRNA genes, which encode RNA but are not translated to protein. We are in the process of evaluating the applicability of the SPLINDID approach to more complex gene regulation mechanisms.

![Fig. 3](image)

**Table 1.** Accuracy of the parameter estimates from SPLINDID approach

<table>
<thead>
<tr>
<th>Parameter</th>
<th>True value</th>
<th>Simulated dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Stochastic model</td>
</tr>
<tr>
<td>$k_p$</td>
<td>0.5 h$^{-1}$</td>
<td>0.489 (8.18%)</td>
</tr>
<tr>
<td>$k_T$</td>
<td>1.5 h$^{-1}$</td>
<td>1.49 (10.5%)</td>
</tr>
</tbody>
</table>

The parameter estimates (CV in percent in parentheses) are shown.
Our approach was motivated by Wagner–Nelson deconvolution (Wagner, 1974) and by the non-parametric, spline-based modeling techniques proposed for input deconvolution for oral dosage forms in PK (Fattinger and Verotta, 1995; Gillespie and Veng-Pedersen, 1985; Gries et al., 1996; Veng-Pedersen and Gillespie, 1986). While these deconvolution techniques have not been systematically investigated for pharmacogenomic modeling, the pharmacogenomic problem presents some specific challenges. For example, it is possible to obtain the unit impulse response in PK by intravenous bolus dosing but it is not possible to obtain the unit impulse response for the pharmacogenomic problem.

Input deconvolution is numerically a challenging ill-posed problem; e.g. in pharmacogenomic datasets, many different combinations $R(t)$ and $k_{34}$ could produce the same outcomes. It is usually necessary to provide constraints to allow the stable implementation of the procedures so that unrealistic options are weeded out prior to deconvolution. For this reason, many investigators in the PK literature have used polyexponentials as the unit impulse response functions for deconvolutions (Gillespie and Veng-Pedersen, 1985; Hovorka et al., 1998). Because the unit impulse responses of biological signaling pathways leading to transcription are not well characterized, it could be argued that such polyexponential-type constraints are potentially excessively restrictive for the pharmacogenomics problem; we considered and were successful with more ‘flexible’ constraints that have not been previously used in this context. First, we proposed the

---

**Fig. 4.** The solid line is the SPLINDID approach estimate for the normalized transcription rate profile for the stochastic model (A), the transit compartment model (B) and the dispersion model (C). The dashed line represents the normalized transcription rate profile in the absence of added noise. The $x$-axes for all the plots are identical and are shown only in the bottom panel; the units on the $y$-axes are in arbitrary concentration units/h.

**Fig. 5.** The fit of the semi-parametric model to the TAT mRNA (open circles in A, in pmol/g) and activity level data (filled circles in B, in absorbance at 331 nm/min g) from Xu et al. (1995). The lines in A and B show the fit of the SPLINDID approach to the TAT mRNA and TAT activity data, respectively. (C) The predicted transcription rate $R(t)$. 

---

K.Bhasi et al.
use of exponential functions of splines to constrain the input \( R(t) \) profiles to positive values and to overcome the undesirable inflections that are often associated with unconstrained cubic splines. We evaluated several approaches including the use of unconstrained splines and the direct fitting of the mRNA profiles [rather than \( R(t) \)] with exponentially modified splines. These alternatives provided inferior fits but also had the serious problem of sometimes yielding negative values of \( R(t) \) values over certain time periods, which are inadmissible on physical grounds. Although algorithms to constrain the splines to preserve convexity/concavity or minimize oscillation between breakpoints are available, it is generally difficult to constrain these algorithms to ensure positive values (Asim and Brodlie, 2003). The approach proposed on the other hand, was not associated with these problems over the datasets examined. The additional constraint that we found useful was the integration factor method, which is used to solve first-order, non-homogeneous differential equations.

The cubic splines used for fitting the data are piecewise polynomials and, therefore, require specification of the number and location of breakpoints. In general, splines are a more ‘flexible’ class of approximating function and can provide the best uniform approximation to a continuous function in a manner that the error has the alternation properties. However, if a spline uses too few breakpoints, the resulting spline is inflexible and provides biased predictions and if it uses too many breakpoints, the spline is too flexible and the predictions are unreliable because of high variance. In SPLINDID, the number of breakpoints is directly determined by the number of estimated ordinates used for exponential spline interpolation. This allows the user to interact with the system and by selecting the number of breakpoints appropriately, the effects of noise can be minimized. We positioned the breakpoints using the DBSOPK program. However, alternative schemes such as simple equispacing and log-equispacing are also feasible and have been investigated in the context of population PK (Fattinger and Verotta, 2003). A cubic spline (Akaike, 1974). In this initial report, we have focused on obtaining proof-of-principle using splines but the methods can be potentially adapted for use with any suitable approximating or interpolating function.

In conclusion, the semi-parametric deconvolution method proposed is both effective, and versatile. It could facilitate an expanded data mining of clinical gene expression data obtained after treatment with drugs, e.g. glucocorticoids and interferon-\( \beta \), that mediate effects predominantly via transcription and could further our understanding of the molecular mechanisms underlying disease and treatment effect variability.

ACKNOWLEDGEMENTS

We thank Dr Donald Mager for his assistance with the ADAPT program. This work was supported in part by grants from the Kapoor Foundation, National Science Foundation (Research Grant 0234895) and the National Institutes of Health (P20-GM 067650).

Conflict of Interest: none declared.

REFERENCES


Glynn,J.M. et al. (1992) Apoptosis induced by actamycin D, camptothecin or aphidicolin can occur in all phases of the cell cycle. Biochem. Soc. Trans., 20, 848.


