

Cue-Signal-Response Analysis of TNF-Induced Apoptosis by Partial Least Squares Regression of Dynamic Multivariate Data

KEVIN A. JANES,¹ JASON R. KELLY,² SUZANNE GAUDET,² JOHN G. ALBECK,²
PETER K. SORGER,^{1,2} and DOUGLAS A. LAUFFENBURGER^{1,2}

ABSTRACT

Biological signaling networks process extracellular cues to control important cell decisions such as death–survival, growth–quiescence, and proliferation–differentiation. After receptor activation, intracellular signaling proteins change in abundance, modification state, and enzymatic activity. Many of the proteins in signaling networks have been identified, but it is not known how signaling molecules work together to control cell decisions. To begin to address this issue, we report the use of partial least squares regression as an analytical method to glean signal–response relationships from heterogeneous multivariate signaling data collected from HT-29 human colon carcinoma cells stimulated to undergo programmed cell death. By partial least squares modeling, we relate dynamic and quantitative measurements of 20–30 intracellular signals to cell survival after treatment with tumor necrosis factor alpha (a death factor) and insulin (a survival factor). We find that partial least squares models can distinguish highly informative signals from redundant uninformative signals to generate a reduced model that retains key signaling features and signal–response relationships. In these models, measurements of biochemical characteristics, based on very different techniques (Western blots, kinase assays, etc.), are grouped together as covariates, showing that heterogeneous data have been effectively fused. Importantly, informative protein predictors of cell responses are always multivariate, demonstrating the multicomponent nature of the decision process.

Key words: tumor necrosis factor-alpha (TNF- α), insulin, apoptosis, partial least squares (PLS), principal component analysis (PCA).

INTRODUCTION

THE RESPONSE OF CELLS TO CYTOKINES, growth factors, and hormones is mediated by receptors that transduce extracellular cues into changes in intracellular physiology. Downstream of receptors, signal

¹Biological Engineering Division, Massachusetts Institute of Technology, Cambridge, MA 02139.

²Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139.

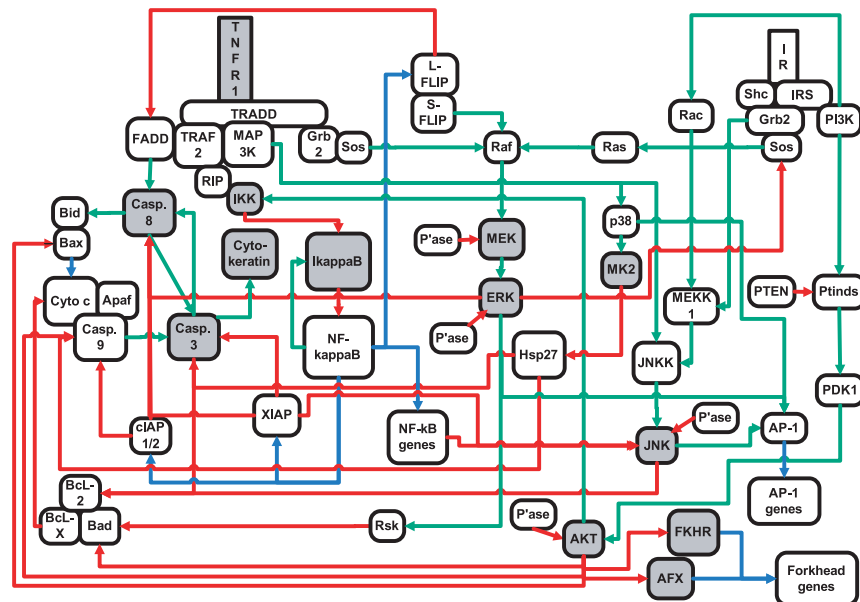
transduction networks are controlled by large sets of proteins acting in concert (Lodish *et al.*, 2000). In the case of programmed cell death, cytokines such as tumor necrosis factor- α (TNF- α) (Wajant *et al.*, 2003) function as prodeath cues, whereas growth factors such as insulin (Avruch, 1998) exert prosurvival effects. The magnitudes of the responses vary with cell type, but the pathways downstream of cytokine receptors are conserved and highly interconnected (Fig. 1(a)). It appears that the determination of whether a cell will live or die involves a balance between prodeath and prosurvival signals (Xia *et al.*, 1995). Thus, the intracellular signal transduction network stimulated by TNF, insulin, and other cytokines acts as a signal processor that converts opposing cues into a functional response that controls cell fate. Although many of the components of cue–signal–response signaling systems have now been identified, it is not yet known how binary decisions, such as death–survival, are made. Moreover, the complexity of cell signaling networks precludes a simple protein-by-protein assignment of function.

Increasingly, systematic methods are being applied to the interpretation and computational analysis of cell signaling (Brent, 2000; Kitano, 2002). These computational methods lie on a spectrum of approaches that vary in their level of abstraction and specificity (Fig. 1(b)) (Ideker *et al.*, 2003). The most abstract methods, such as multivariate analysis and clustering (Rives *et al.*, 2003), are powerful because they can handle empirical data and prior knowledge about the underlying phenomena in a flexible fashion. At the other extreme, differential equation-based models (Schoeberl *et al.*, 2002) are useful for codifying existing prior knowledge in pursuit of *in silico* predictions. A number of groups have used basic multivariate analysis techniques to mine DNA microarray datasets (Alter *et al.*, 2000; Holter *et al.*, 2000; Kuruvilla *et al.*, 2002; Misra *et al.*, 2002; Yeung *et al.*, 2002). In addition, several reports have been published using partial least squares (PLS) methods to generate microarray-based classification systems (Datta, 2001; Nguyen *et al.*, 2002). Unlike abstracted, data clustering algorithms, PLS can relate entire datasets to prior knowledge or to other data by tuning the relative emphasis of a particular measurement based upon its contribution to a datawide model. Unlike specified, mechanistic models, PLS does not require fine-grained knowledge about the interrelationship of measured components. Rather, PLS enables a data set to “speak for itself” in terms of global variations and covariations.

Cue–signal–response relationships in cell signaling are ripe for applying multivariate analysis techniques, such as PLS, to tease out the key biochemical changes associated with critical cell decisions. However, assaying cell signaling involves the measurement of a wide range of parameters including protein abundance, localization, enzymatic activity, and posttranslational modification. In addition, protein measurements involve a variety of techniques, including Western blots, kinase assays, protein microarrays, and imaging. Thus, protein-based cell signaling studies present a challenging problem in the fusion of heterogeneous data into a single coherent model (von Mering *et al.*, 2002). From previous work on transcriptional profiling (which involves a single type of data, usually from a single type of measurement), it was not known if multivariate methods would be useful with heterogeneous, sparse sets of protein measurements. Moreover, it was not clear whether PLS models could be used to generate hypotheses that are testable by molecular-level or cell-level experimentation.

In this paper, our purpose is to determine whether PLS regression can be used to uncover important aspects of biological cue–signal–response systems (Fig. 1(a)). Specifically, we examine the TNF- and insulin-mediated death-versus-survival response of HT-29 human colon carcinoma cells by PLS analysis of multivariate, heterogeneous experimental measurements of nearly 30 signaling protein levels, states, and activities. We wish to attract the interest of other experimental molecular cell biologists, so we emphasize the conceptual details of PLS as a practical modeling technique and refer readers elsewhere for more rigorous discussion of the mathematics (Geladi *et al.*, 1986; Martens *et al.*, 2001). A noteworthy feature of PLS applied to biological data, in contrast to PLS modeling in other areas (e.g., consumer models of perceived food quality), is that biological measurements reflect tangible molecular entities with known, mechanistic roles in intracellular processes. Thus, these data-driven models of signaling are empirical, but not phenomenological, and suggest mechanistic dependencies. We show here that PLS and related methods can uncover key contributors to death–survival decisions. These contributions always involve multiple proteins working in concert, but the informative proteins consist of only a fraction of the original protein dataset. Thus, our results suggest that within a cue–signal–response system lies a reduced set of information-rich protein measurements that together constitute an efficient model of the signaling network state and the relevant signal–response relationships.

(a) Specified, mechanistic



(b)

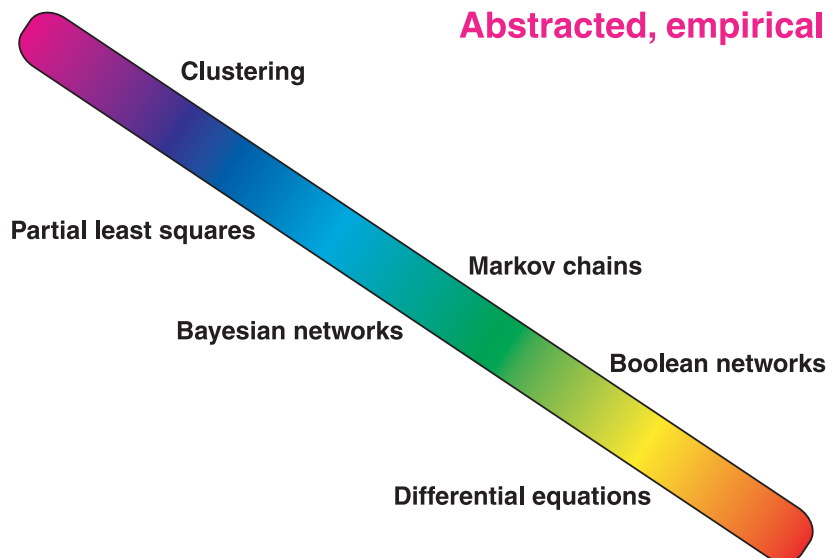


FIG. 1. Partial least squares regression is an analytical technique of intermediate specificity and abstraction that is useful for modeling complex protein networks, such as the signaling pathways downstream of $\text{TNF-}\alpha$ and insulin. (a) Schematic illustration of the biomolecular signaling network governing cell death-versus-survival decisions in response to prodeath cues ($\text{TNF-}\alpha$) and prosurvival cues (insulin) (Janes *et al.*, 2003). Network state is determined by many component characteristics, including protein expression levels, protein–protein interactions, protein enzymatic activities, and protein locations. Approximately 30 of these characteristics are measured in quantitative and dynamic terms for multivariate systems analysis. Green arrows indicate activating interactions, red arrows indicate inhibitory interactions, and blue arrows indicate transcriptional interactions. Shaded nodes highlight proteins whose characteristics were measured experimentally. Diagram is not implied to be comprehensive (e.g., location-dependent interactions have been abstracted). (b) Spectrum of computational approaches for modeling protein networks: clustering (Rives *et al.*, 2003), partial least squares, Markov models (Gomez *et al.*, 2002), Bayesian networks (Sachs *et al.*, 2002), Boolean networks (Huang *et al.*, 2000), and differential equations (Schoeberl *et al.*, 2002). The spectrum is not implied to be comprehensive.

MATERIALS AND METHODS

Cell culture and treatment

HT-29 cells (ATCC) were grown in McCoy's 5A medium, supplemented with 10% fetal bovine serum, 2 mM glutamine, 100 Units/ml penicillin, and 100 $\mu\text{g/ml}$ streptomycin (Invitrogen). For signaling experiments, cells were seeded at 50,000 cells/cm², grown for 24 h and then treated in complete medium supplemented with 200 U/ml interferon gamma (IFN γ , Roche) for an additional 24 h. After sensitization in medium with IFN γ , cells were rinsed once with medium and treated with complete medium supplemented with 50 ng/ml TNF- α (Peprotech) with or without 100 nM insulin (Roche) for the indicated time.

For cell death experiments, HT-29 cells were seeded at 1.6×10^4 per well in 96-well plates. After 24 hours of growth, the cells were treated with IFN γ for 25 hours. Pretreatment with the inhibitors (50, 16.7, 5.6, or 1.8 μM PD98059; 1000, 333, 111, or 37 nM wortmannin; 10, 3.3, 1.1, or 0.37 μM SB203580; 10, 3.3, 1.1, or 0.37 μM SP600125; and combinations thereof) was performed during the final 90 minutes of IFN γ stimulation. After removal of the pretreatment medium, the cells were treated with the same inhibitors as during the pretreatment, in medium alone or in the presence of 50 ng/ml TNF- α or 50 ng/ml TNF- α + 50 nM insulin.

Quantitative Western blot assay

Whole cell lysates of both adherent and nonadherent cells were prepared in 125 mM Tris-HCl pH 6.8, 4% SDS, 20% glycerol, 10 mM NaF, 10 mM Na₄P₂O₇, 10 mM β -glycerophosphate, 1 mM Na₃VO₄, 10 $\mu\text{g/ml}$ leupeptin, 10 $\mu\text{g/ml}$ pepstatin, and 10 $\mu\text{g/ml}$ chymostatin. Protein concentration was determined by the bicinchoninic acid assay (Pierce). Proteins (50 μg of lysate) were separated by SDS-PAGE and then transferred to nitrocellulose. After blocking (30–60 min at room temperature), blots were probed overnight at 4°C in primary antibody, washed 3 \times 5 min in TBS-T (20 mM Tris-HCl pH 7.5, 137 mM NaCl, 0.1% Tween 20), incubated 1 h at room temperature in secondary antibody (1:2500 Alexa-488 conjugated goat anti-rabbit or 1:2500 phycoerythrin-conjugated goat anti-mouse, Pierce), and finally washed 3 \times 5 min in TBS-T. Blots were scanned on a FluorImager 495 (Molecular Dynamics) and quantitated using ImageQuant 5.0 software (Molecular Dynamics). Primary antibodies used were anti-phospho-Akt substrates (1:1000), anti-Akt, anti-phospho-Akt (S473, 1:1000), anti-caspase-8 (1C12, 1:1000), anti-phospho-FKHR and AFX (S256, 1:1000), anti-phospho-I κ B- α (S32, 1:500), anti-MEK1/2 (1:1000), anti-phospho-MEK1/2 (S217/221, 1:1000), anti-phospho-SAPK/JNK (T183/Y185, 1:500), anti-caspase-3 (1:500), anti-cleaved caspase-3 (1:1000) (Cell Signaling Technology), and anti-TNFR1 (1:1000, Biovision). Data points are averages of two independent time courses.

High-throughput kinase activity assay

The microtiter-based kinase activity assays were performed as described (Janes *et al.*, 2003) from lysates prepared under similar conditions, but in different experiments, from those used for quantitative Western blotting. Data points are the averages of three independent time courses.

Cell death assay

After 14.5 hours of cytokine treatment, cells were rinsed with PBS and trypsinized. The supernatant and rinse from treated cells were saved and combined with the trypsinized cells to ensure capture of both floating and adherent cells. The cells were washed with PBS, then fixed in 100% MeOH. Fixed cells were stained with a fluorescein-labeled monoclonal antibody against caspase-cleaved cytokeratin (M30, Roche) according to the manufacturer's instructions. Cells were washed once after staining, then analyzed on a FACScalibur (Becton–Dickinson) to determine the percentage of M30-positive cells in each sample.

Computational analysis

Prior to all analyses, the dependent and independent blocks (i.e., matrices) were mean centered and variance scaled. Principal component analysis of the preprocessed data matrix (X) was done in

SIMCA-P 10.0 (Umetrics) software suite according to the following iterative formula:

$$E_1 = X - t_1 p_1^T; E_2 = E_1 - t_2 p_2^T; E_i = E_{i-1} - t_i p_i^T \quad (1)$$

where E_i represents the residual of the i^{th} principal component, with score vector t_i and loading vector p_i , and T represents transpose. The information captured by the i^{th} component was determined by the following formula:

$$\% \text{ information captured} = \|t_i p_i\|^2 / \|X\|^2. \quad (2)$$

For the discriminant PLS regression (DPLSR) model, a binary approach was used to generate a dependent block classifying TNF- α (\equiv 0)- and TNF- α + insulin (\equiv 1)-treated samples. DPLSR was done in SIMCA-P according to (1) for the independent block and the following iterative formula for the dependent (Y) block:

$$F_1 = Y - b_1 t_1 q_1^T; F_2 = F_1 - b_2 t_2 q_2^T; F_i = F_{i-1} - b_i t_i q_i^T \quad (3)$$

where F_i represents the residuals of the i^{th} dependent principal component, with score vector t_i and loading vector q_i , and b_i represents the weight characterizing the inner relation between the independent and dependent principal components. The class structure captured by the i^{th} component was determined by the following formula:

$$\% \text{ class structure captured} = \|b_i t_i q_i\|^2 / \|Y\|^2. \quad (4)$$

A block shuffling algorithm was programmed in MATLAB (Mathworks Inc.) to determine the regression coefficients that were significant in predicting the dependent variables in these models: the dependent block was randomized, and then PLS analysis was conducted via the NIPALS algorithm (Geladi *et al.*, 1986) using a PLS Toolbox (Eigenvector Research). The results of 500 shuffling iterations were recorded, and the mean \pm S.D. was reported. Regression coefficients greater than one S.D. outside of the shuffling iterations were considered significant.

For the inhibitor datasets, the effects of the pharmacological inhibitors on kinase activity could not be experimentally measured. Instead, approximate activities using the known *in vivo* IC₅₀ values of the various inhibitors (Alessi *et al.*, 1995; Arcaro *et al.*, 1993; Bennett *et al.*, 2001; Cuenda *et al.*, 1995) were derived and scaled to values from zero (complete inhibition) to one (no inhibitor) for each kinase. When the estimated ERK, JNK, p38, and PI3K activities were cast as a four-dimensional signaling space and a one-dimensional response space of measured levels of apoptosis, the problem was not well posed for PLS analysis, because little insight was gained by reducing four dimensions to two or three. (PLS models are most useful for data reduction when the original signaling space exceeds 10 measurements [K.A.J., unpublished observations].) To address this issue (and to include that certain kinase activities might have nonadditive effects on cell survival), two-factor (Signal 1 \times Signal 2), square (Signal 1 \times Signal 1), and cubic terms were appended to the independent block of calculated activities, and the expanded measurement set was analyzed with the same iterative formula as the DPLSR approach.

RESULTS AND DISCUSSION

Directed data acquisition and biological significance

To generate models, PLS relies on a data matrix of informative, quantitative measurements. Our work focuses on the regulation of programmed cell death in response to the prodeath factor TNF- α and the survival factor insulin. We directed our experimental measurements to a subset of the downstream kinases, caspases, and other regulators that are known to influence cell death or survival (Fig. 1(a)). In particular, we focused on those proteins whose signaling varies in response to TNF- α , insulin, or both, and whose perturbation, by genetic or other means, affects cell survival. By imposing both these criteria, we increase the likelihood that the data we collect will make a useful contribution to models of apoptotic signaling.

In our experimental analysis of apoptosis, we have examined four types of protein-based signals: phosphorylation, enzymatic activity, abundance, and cleavage state. Protein phosphorylation is an important mechanism of signal transduction in cells (Hunter, 2000), and measurements of the phosphorylation states of signaling proteins and the enzymatic activity of the protein kinases that catalyze these phosphorylation events represent one source of information on death and survival signals. Similarly, measurement of protein abundance captures important information on both intracellular signaling proteins and cytokine receptors. Finally, measurement of the cleavage states of caspases, the biochemical executors of apoptosis (Nicholson *et al.*, 1997), makes it possible to gauge how far cells have progressed towards death.

The specific data points collected in this paper come from three types of measurements, which each characterize one or more of these four types of signals. In one set of data, we explored signaling dynamics in depth by quantifying nearly 30 protein states, levels of abundance, and activity from HT-29 cells treated with 50 ng/ml TNF- α alone or in combination with 100 nM insulin at various time points subsequent to cytokine addition (Fig. 2). The kinase activities of five protein kinases (ERK, Akt, JNK, IKK, and MK2) were measured with a kinase activity assay (Fig. 2(c-d)) (Janes *et al.*, 2003), in which kinases are immunoprecipitated from cells and mixed with protein or peptide substrates. Three of these kinases (ERK, MK2, and JNK) are components of mitogen-activated protein kinase cascades, Akt is a key mediator of survival signals, and IKK regulates the nuclear factor- κ B transcription factor (Fig. 1(a)). Regulatory phosphorylation events were monitored for MEK, Akt, and JNK by Western blotting (Fig. 2(a-b)). Western blotting was also used to examine the phosphorylation of several Akt and IKK substrates, as well as the cleavage state of caspase 8 and the executioner caspase, caspase 3. One important feature of the measurements is that they yield heterogeneous data with regard to experimental technique and biological significance. In spite of this heterogeneity, the data all capture signals that meet the criteria for multivariate analysis: signals that vary in response to cues and signals that, when perturbed, affect cell response.

In another set of experiments, the apoptotic response of HT-29 cells was measured in the presence of different small molecule kinase inhibitors that act as drugs. By examining the combinatorial effects of these molecular-level perturbations on cell survival, we indirectly explored signal-response relationships and broadly surveyed the quantitative range of the apoptotic response in HT-29 cells. For these experiments, changes in apoptosis were measured by flow cytometry with an apoptotic marker: caspase-cleaved cytokeratin.

Compact representations of signaling by principal component analysis

As a first approach to reducing the complexity of our data, we used principal component analysis (PCA), a nondirected multivariate analysis technique. We cast our data as a matrix of M rows, representing m_i experimental samples, and N columns, representing n_i measurements on the experimental samples. For multivariate analysis to work most effectively, this matrix should have as few empty elements as possible, implying that the same measurements be performed on all samples.

Here, all 29 measurements were performed on all 26 samples (13 time points per treatment \times 2 treatments), leaving no empty elements in the data matrix. The different time points were collected from separate tissue culture plates, so the individual time points from both treatments were blindly considered as independent samples for the purposes of the multivariate analysis. Since all of the measurements were made on samples prepared under identical conditions, we can safely include these in the same sample row. However, not all measurements used the same cell extract material (since the different assays require different extract procedures; see Materials and Methods), so it is possible that some of the variation in the measurements might be due to variability in the lysate preparations.

An $M \times N$ matrix is frequently viewed as a flat, two-dimensional spreadsheet of numbers; alternatively, the data can be thought to represent a multidimensional "signaling" space, where the different protein measurements form a set of N axes (Fig. 3). The M samples are then vectors describing coordinates in the N -dimensional space, such that the projection of a sample vector m_i on to axis n_i represents one measurement. In this N -dimensional space, the axes are not orthogonal, meaning that the measurements defining these axes are partially redundant. Moreover, the m_i sample vectors project on some axes more than others, because some measurements have a wider range of variation than others.

PCA reduces the dimensionality of the data space and the size of the $M \times N$ matrix, by factoring it into the product of an $M \times P$ scores matrix, containing the same number of scores as samples in the data, and a $P \times N$ loadings matrix, containing the same number of loadings as original measurements.

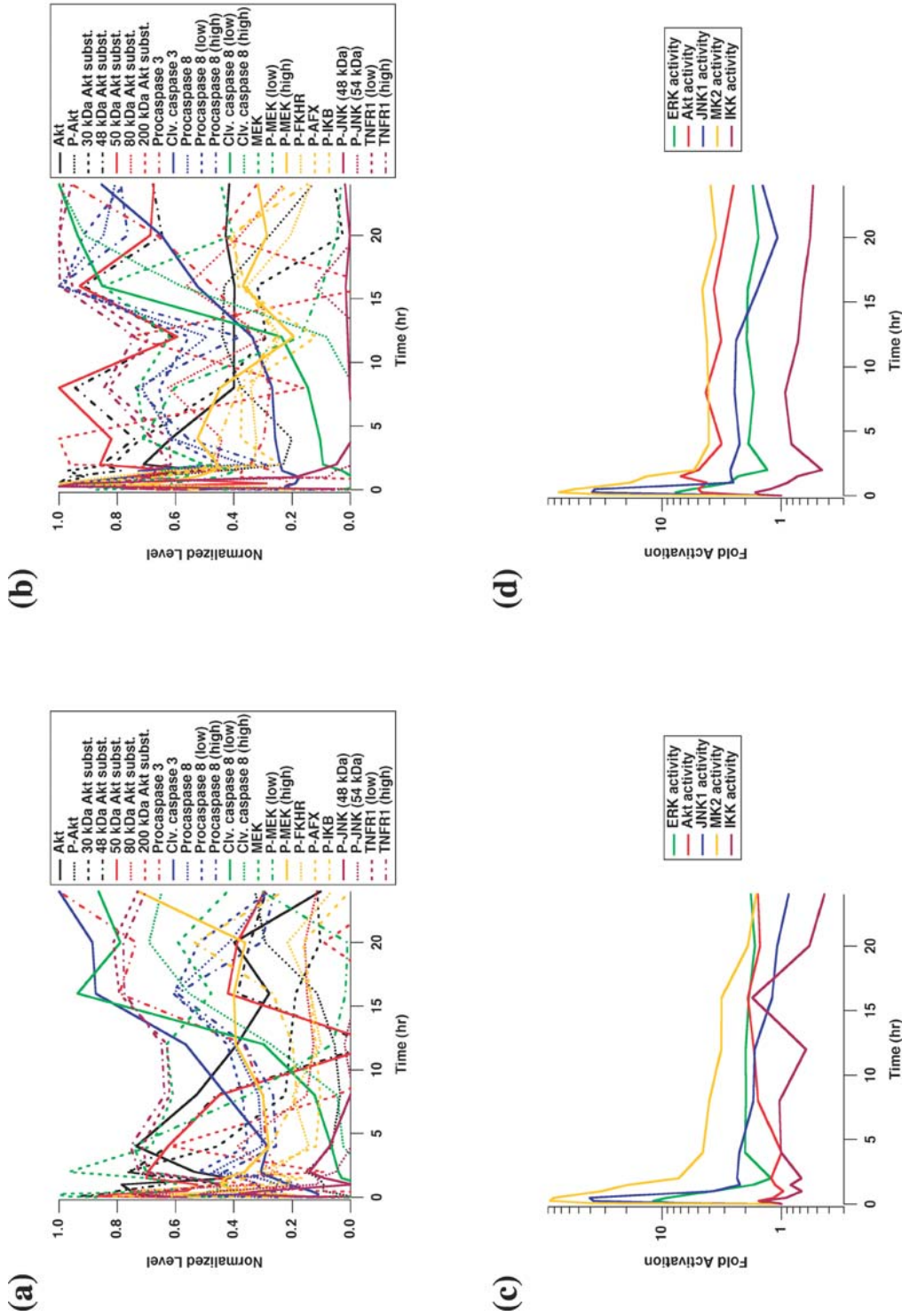


FIG. 2. TNF- α alone and TNF- α + insulin cotreatment elicit distinct signaling patterns in HT-29 cells. Protein state, level ((a) and (b)), and activity ((c) and (d)) measurements in HT-29 cells over 24 hours treated with 50 ng/ml TNF- α ((a) and (c)) or both 50 ng/ml TNF- α and 100 nM insulin ((b) and (d)). For (a) and (b), “P-” signifies phosphorylation state. Note the difference in scale between (a–b) and (c–d). Activity measurements are reprinted from Janes *et al.*, (2003) with permission.



FIG. 3. A data matrix of N signaling measurements can be represented as an N -dimensional signaling space and reduced via principal components analysis to a P -dimensional principal component space. Column shade represents the magnitude of the measurement in the different samples. Note the colinearity of axes in signaling space and the orthogonality of axes in principal component space.

This decomposition represents the data in P dimensions, with p_i principal components, or latent variables, of the system. These principal components form a new, orthogonal coordinate system. The scores matrix defines how far the samples project along the coordinate system of the p_i principal components, and the loadings matrix defines how the principal component axes “point” relative to the original measurement space. These two matrices enable one to flip from principal component space to signaling space, and vice versa, by matrix multiplication and division, respectively. The multiplication of a row in the scores matrix and a column in the loadings matrix is called an outer product.

Most computational algorithms determine the principal components iteratively, with the first component capturing as much of the measurement information (i.e., the changes in measurement values over all samples) as possible. Then, the outer product of the first score-loading vector pair is subtracted from the original data matrix, and the second principal component is calculated to capture as much of the residual information as possible. To avoid biasing the decomposition toward measurements with disproportionately large absolute magnitudes or dynamic ranges, we subtract the mean of each measurement, taken across all samples, and divide by the measurement variance. These preprocessing techniques (called mean centering and variance scaling, respectively) eliminate difficulties associated with measurement heterogeneity by putting all data evenly on a unitless measurement axis with a mean of zero and variance of one. This enables scale-free, intermeasurement comparisons in the decomposition procedure.

If the data matrix is completely decomposed by iteration, then P will equal the lesser of M and N , and the product of the scores and loadings matrices will exactly equal the original dataset. However, because the principal components were iteratively calculated to maximize information capture, good approximations of the original dataset can be obtained with only the first few principal components. In this way, PCA generates a data-driven model, comprised of the most significant principal components, and extracts the most salient features of a dataset while removing spurious fluctuations (usually from measurement error and noise). The critical question in PCA is whether a significant fraction of the measurement information ($\geq 60\%$) can be captured effectively when P is much less than N .

To begin to determine the most critical factors in our dataset for influencing the death-versus-survival decision, we performed PCA on a complete set of preprocessed protein data and examined the amount of information captured as a function of increasing principal component number. In Fig. 4(a), we see that the first two principal components together capture over 62% of the information in the original data matrix, with progressively smaller contributions from higher dimensions. Often with PCA, the residual information from higher components (in this model, 38% of the total) is postulated to contain mostly random experimental fluctuations. However, with only 29 measurements, our sample set was too small to permit a rigorous treatment of signal to noise (see Fig. 4(a) legend).

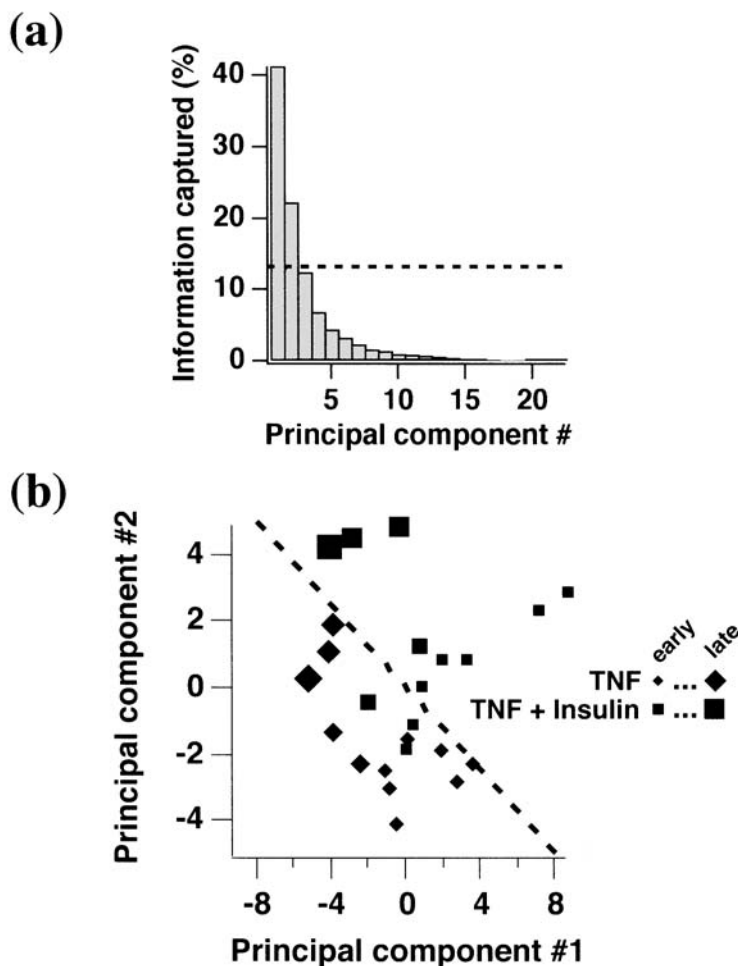


FIG. 4. Principal component analysis reduces a complex dataset of heterogeneous signaling measurements to a data-driven model that retains the differences in the original measurement set. (a) Information captured as a function of increasing principal component number. The hatched line represents the arbitrary information cutoff for the two-component PCA model. Higher components were very fluctuant in their scores and loadings and did not indicate consistent data-wide variations in the measurement set (data not shown), suggesting that these components were fitting measurement error and noise. (b) Scores plot for the two-component PCA model. The hatched line highlights the data-driven separation of the TNF- α (\blacklozenge) and TNF- α + insulin (\blacksquare) samples. The size of the markers indicates increasing time.

Therefore, we further examined the quality of the two-component PCA model by plotting how the different samples mapped along the two principal components via their scores (Fig. 4(b)). The size of the data markers represents increasing time; we see that that earlier time points fall largely south and southeast on the two principal components in Fig. 4(b), whereas later time points are located more north and northwest. Importantly, when we applied an objective test to the TNF- α (diamonds) and TNF- α + insulin (squares) time courses, we found that the two sets of measurements were separated in a highly significant fashion, with $p < 0.005$ that the data represents two distinct populations by a two-dimensional Kolmogorov–Smirnov test (Fasano *et al.*, 1987), illustrated schematically by the hatched line in Fig. 4(b). This result contrasts the separation achieved by a PCA model with only one principal component (when the null hypothesis was not rejected by this same test, with $p = 0.19$), and it therefore supports the two-component model as a compact, but not oversimplified, representation of the sample set.

The information contained in the reduced model (see first two columns in Fig. 4(a)) has general implications for the study of protein networks. Starting with heterogeneous measurements from a complex signaling network (Fig. 1(a)), we created a model that efficiently reduced the dimensional complexity from

29 to 2 dimensions (Fig. 4(a)). This reduced model was informative, in that it clearly determined from the differences in protein state (but without any prior knowledge) that the data derived from two sets of cells that had been exposed to different death and survival mediators.

As mentioned in the Introduction, the original measurements themselves contain real biological meaning, it is therefore valuable to explore the contribution of individual signals to the datawide model, as well as the interplay between signals; the relationship between the multivariate models and the original signaling measurements is contained in the loadings matrix. Upon inspection of the loadings, we first found that none of the reduced axes (the principal components) corresponded to an individual measurement (data not shown). Rather, the axes were composed of combinations of measurements: the first principal component points largely toward early, prosurvival signals, such as phosphorylated Akt substrates and phospho-I κ B, whereas the second component points more toward late, prodeath signals, such as cleaved and procaspase levels (data not shown). Second, we found that not all measurements were equally informative. As mentioned earlier, high-dimensional data usually contains measurements that are collinear, and we identified several groups of measurements with equally weighted loadings (data not shown), indicating that these signals were redundant in their information content. Importantly, this grouping was consistent with the known biology. For example, the regulatory phosphorylation of Akt and the enzymatic activity of Akt were highly collinear, and the phosphorylation states of many Akt substrates were clustered. Therefore, we can conclude that PCA has achieved an effective data fusion of heterogeneous sources of data.

Overall, our results with PCA suggest that a multivariate, but by no means comprehensive, set of signaling measurements is sufficient to classify phenotypically distinct samples of cells, whereas no single measurement has this power of discrimination. PCA also appears to be effective at fusing diverse data sources into a single consistent model. More generally, the analysis suggests that there exists a set of highly informative measurements that, in combination, compactly describes the relevant variations in signaling network state and might in the future obviate the need to characterize all of the signals in the network.

Identification of survival signals by discriminant partial least squares

Our next step in data analysis focused on attempts to incorporate prior knowledge about cell treatments rather than to derive them *post facto*. In general, data analysis is much more powerful if prior knowledge can be captured effectively; the specific prior knowledge in our data is that some cells were treated only with TNF- α whereas others were cotreated with insulin. HT-29 cells die in response to TNF- α (Abreu-Martin *et al.*, 1995; Wilson *et al.*, 2002) and are rescued from TNF-induced death when cotreated with insulin (Remacle-Bonnet *et al.*, 2000). To focus on this death-versus-survival decision, we used PLS, since it is designed to generate data-driven models that relate a matrix, or block, of independent measurements to a block of dependent measurements or classifications.

Cue, signal, and response measurements can act as independent or dependent blocks, but it is conceptually helpful to organize the data so that cause–effect relationships are appropriately directed in relation to the original biological setting (i.e., so that an independent block of signals affects a dependent block of responses, rather than the reverse). In our case, the independent block contains the levels and activities of proteins, and the dependent block contains the binary information on cytokine treatments that constitutes our prior knowledge about the outcome:

- Samples that are associated with cell death (i.e., with TNF- α treatment) \equiv 0.
- Samples that are associated with cell survival (i.e., with TNF- α + insulin treatment) \equiv 1.

For the classification, we assigned time points from each treatment to one of the two classes, ignoring the time component. This class architecture highlights proteins whose state, level, or activity is consistently different between the two responses at all time points. (In future work, we expect to have sufficient data to look at each time point individually.) Since our model is meant to bring out the differences between the two outcome classes, it belongs to a subcategory of PLS called discriminant partial least squares regression (DPLSR).

To perform DPLSR, the original M samples \times N measurements data matrix is supplemented with a separate, dependent block of M samples \times one classification. Principal components are extracted iteratively

from the independent block, analogously to PCA, and the dependent block might also be decomposed, if there were multiple classifications. The most important modification is that, whereas PCA extracts latent variables to maximize the information captured from the independent block, DPLSR decomposes the data matrix to maximize the correlation between the principal components and the (possibly decomposed) classification of the dependent block. The principal components are related to the dependent block by additional coefficients, called weights, that quantify the contribution of each component to the class model through an “inner product.”

Whereas PCA generates a reduced model of the variations in the independent block alone, the decomposition criterion for DPLSR focuses on the relationship between the independent and dependent blocks, highlighting the measurements that strongly covary with the known outcome and deemphasizing those that do not. In other words, it is insufficient for a measurement to simply change from sample to sample—to be regarded as significant in the DPLSR model, these changes must covary with the changes in apoptotic response. This is a powerful quality of PLS, since it is frequently difficult to identify the determinative changes in a particular response, when the characteristics of many other proteins are simultaneously varying.

We decomposed our signaling dataset with DPLSR and the aforementioned binary classification system, after preprocessing the independent and dependent block as described earlier. The decomposition criterion makes the data in the original dependent block more valuable than the independent block for assessing information capture, so we concentrated on model capture of the class structure rather than capture of the signaling variation, which was the focus of PCA. As with the PCA model, however, we observed that the first two components were highly informative (78% capture of the class structure) and that information content fell rapidly with increasing component number (data not shown).

To cross-validate the two-component DPLSR model, we first used a Cooman’s plot (Fig. 5(a)) to determine the ability of the model to discriminate between two samples classes. The abscissa and ordinate mark the deviations of the samples from the two classes. The four quadrants demarcate the samples that, according to the model, would fall into (I) neither class, (II) the death class, (III) both classes, or (IV) the survival class. We see in Fig. 5(a) that the TNF- α samples (diamonds) fall largely in (II) and TNF- α + insulin samples (squares) fall in (IV). Thus, the DPLSR model allows good discrimination of the two outcomes, with only a few samples falling into (III) and no samples in (I). Interestingly, the sample with the largest deviation from the death class (TNF- α + insulin at 30 min, marked by an arrowhead in Fig. 5(a)), corresponds to the time point at which the largest number of measurements differ in a pairwise comparison of the TNF- α and TNF- α + insulin samples (data not shown).

Although both PLS and PCA reduce data complexity, PLS is generally more powerful because of the predictive ability of the resulting models. After a PLS model has been generated linking independent and dependent blocks, one can attempt to predict which dependent variable is linked to a particular set of measurements. PLS model predictions are achieved in three steps: 1) an experiment is projected onto the reduced principal component space of the PLS model; 2) in this principal component space, the model then uses the inner product, consisting of scores, loadings, and weights, to calculate a predicted dependent component; and 3) if the dependent block has been reduced in the model, the predicted dependent component is recomposed to form a predicted dependent block.

We used this predictive power as an additional form of cross-validation to evaluate how well the PLS model recapitulated the original outcome classes from the independent block of signaling measurements. In Fig. 5(b), for each sample, we plotted the observed (assigned) class against the class predicted by the DPLSR model using only the independent block of the sample. These predictions were not *de novo* predictions, since all the data are part of the training set. Nevertheless, these predicted outcome classes do provide an indication of model quality: if the model were to perfectly fit the training set, then all of the TNF- α treated samples would collapse upon (0, 0) and the TNF- α + insulin treated samples upon (1, 1) in the observed-versus-predicted plot. Instead, the data spread laterally from these coordinates, because we retained only two principal components in the PLS model, but we considered the two-component model valid, since it achieved the desired separation of the samples, with no overlap between the two sample classes.

As mentioned before, our original motivation for developing the DPLSR model was to quantify the contributions of various signaling measurements to the outcome classification, and we therefore examined the measurement space of the PLS model. In the PLS model, the contributions of a measurement take the

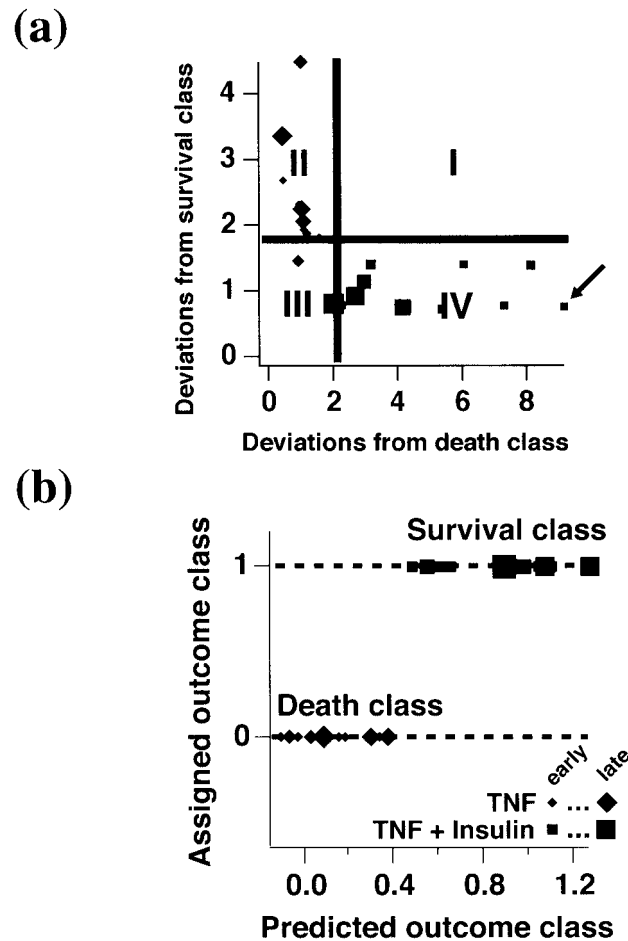


FIG. 5. Discriminant partial least squares regression explicitly distinguishes distinct phenotypic outcomes from a dataset of heterogeneous signaling measurements. (a) Cooman's plot for the DPLSR model. The four quadrants represent (I) neither class, (II) the death class, (III) both classes, and (IV) the survival class for the TNF- α (\blacklozenge) and TNF- α + insulin (\blacksquare) samples. The solid lines represent 95% confidence intervals for excluding a sample from the model. The size of the markers indicates increasing time. The arrow highlights the 30 min TNF- α + insulin time point as the most clearly classified survival sample. (b) Observed-versus-predicted plot for the DPLSR model. The hatched lines represent the original class assignments for the TNF- α (\blacklozenge) and TNF- α + insulin (\blacksquare) samples. The size of the markers indicates increasing time. Perfect recapitulation of the dataset would collapse the TNF- α samples on (0, 0) and the TNF- α + insulin samples on (1, 1). Note that the two sample sets do not overlap horizontally.

form of a regression coefficient that includes the multiplicative and additive influences of the loadings and scores of the measurement in the two principal components that are retained, along with the weights relating these principal components to the class outcome. The columns in Fig. 6 shows the regression coefficients for all state, level, and activity measurements included in the model. Positive regression coefficients suggest correlations with survival, by contributing to a "1" classification by addition, and negative regression coefficients suggest correlations with death, by contributing to a "0" classification by subtraction.

Next, we assessed the significance of these regression coefficients in the context of the model, the class structure, and the original dataset. Techniques for estimating significance in PLS models vary, but we had success with a Monte Carlo-type strategy, where the rows of the dependent block (here, the classifications) are shuffled randomly, and a DPLSR model is created with a new set of regression coefficients (Rohlf *et al.*, 2000). These regression coefficients are naïve, since the prior knowledge about the outcome of these samples is lost by the randomization procedure. After several hundred shuffling iterations, one can gain an indication of each mean naïve regression coefficient (usually near zero) in a particular data structure

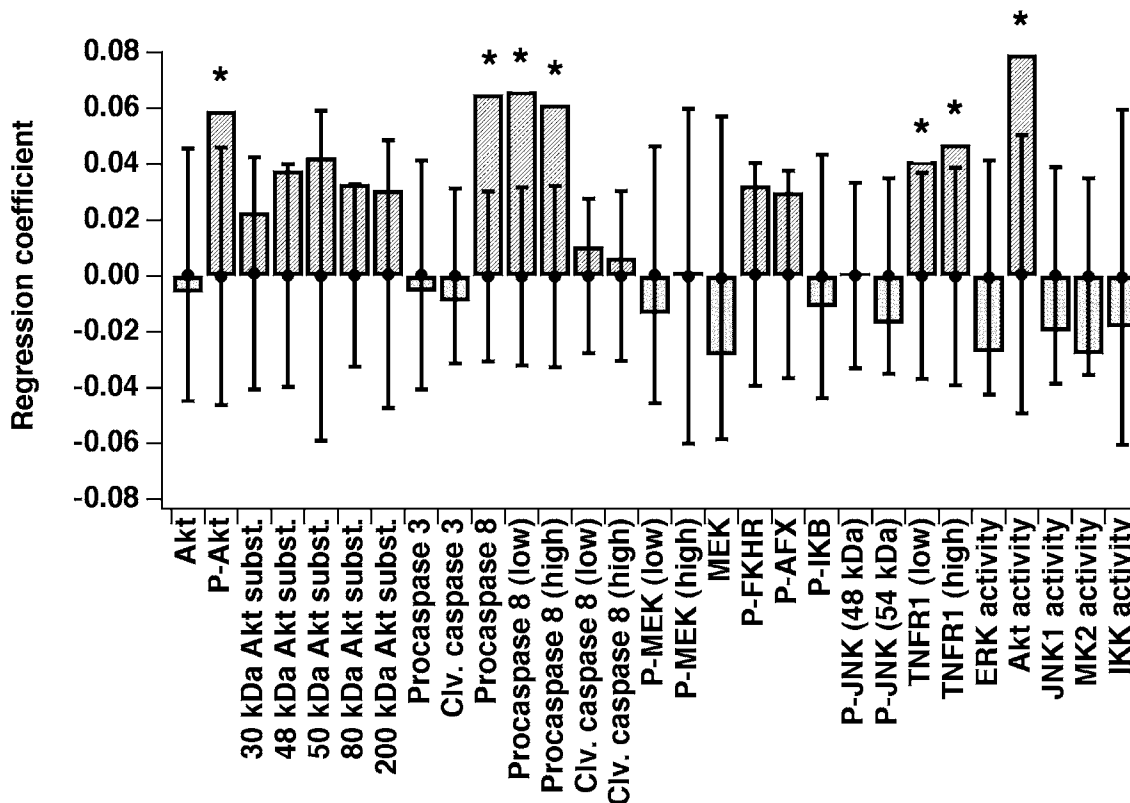


FIG. 6. Discriminant partial least squares regression identifies Akt phosphorylation state, procaspase 8 levels, TNFR1 levels, and Akt activity as important signals for cell survival. The columns represent the model regression coefficients, whereas the points represent the mean naïve coefficient (\pm S.D.) from 500 block shuffling iterations. The asterisks indicate measurements with regression coefficients outside of those expected by a naïve model.

and, importantly, the variation of the regression coefficient about its mean. The results of these shuffling iterations are included as markers in Fig. 6, with asterisks indicating the seven “outcome classifiers.” These measurements have regression coefficients substantially different from the naïve coefficients that would be expected by chance classification.

Most significantly, we found that, despite the simple classifications in the DPLSR model, the multicomponent properties of the apoptosis network were clear (Fig. 7). When we generated DPLSR models with partial datasets consisting of certain isolated classifiers, such as the levels of TNF receptor 1 (TNFR1, leftmost column in Fig. 7), we recapitulated the class structure poorly (38% capture). In contrast, with the complete set of outcome classifiers, we generated a model of nearly the same quality as when the entire dataset was used (74% versus 78% capture, compare third and fourth column in Fig. 7). By this analysis, we identified those signals that the data support as a “minimal model” for predicting the dependent block. Therefore, both the PCA and DPLSR models support the hypothesis that measurement of individual molecules is not sufficient to capture network properties, but an appropriate set of information-rich measurements can, in combination, create useful models of the signaling network.

A number of the survival classifiers found by DPLSR were intuitively reasonable from a biological perspective: Akt phosphorylation state, procaspase 8 level (both forms), and Akt activity. Akt is phosphorylated in response to insulin treatment (Lawlor *et al.*, 2001), and phosphorylation leads to an increase in catalytic activity (Alessi *et al.*, 1996; Kohn *et al.*, 1996) to constitute a potent prosurvival signal (Dudek *et al.*, 1997). Caspase 8 is an initiator caspase along one of the death pathways (Nicholson *et al.*, 1997). Early in the TNF- α time course, the inactive pro-form was processed to the active, cleaved form (Fig. 2(a)). In the TNF- α + insulin time course, however, this cleavage was blocked, retaining high levels of the procaspase 8 zymogen (Fig. 2(b)).

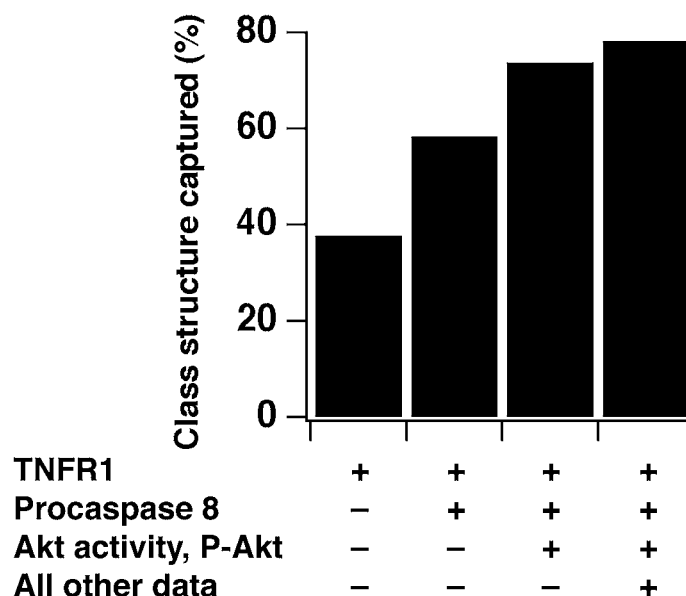


FIG. 7. A combined subset of critically important survival signals quantitatively captures outcome classes to the same extent as an entire dataset. Smaller subsets of the data were modeled individually by DPLSR and characterized by their ability to capture the death-versus-survival class structure. Note that the dataset consisting of measurements of TNFR1, procaspase 8, and Akt, in combination, captures nearly as much of the assigned class structure as all 29 measurements.

In contrast, it was intriguing and counterintuitive that the model identified the level of expression of TNFR1 as a survival classifier. Comparison of the original time courses showed that the DPLSR result was supported by the original sample set (Fig. 2(a–b)), perhaps suggesting that insulin competes with, or inhibits, the internalization and downregulation of these receptors (Kull, 1988). Alternatively, there may be other potentiating effects of insulin on TNFR1 expression (Dunger *et al.*, 1995). The identification of TNFR1 levels as a predictor for outcome emphasizes that DPLSR models can incorporate prior knowledge without biasing against novel or counterintuitive signal–response relationships.

In addition to extracting outcome classifiers, the DPLSR model also identified measurements with zero, or near zero, contribution to any class discrimination (e.g., 48 kDa JNK phosphorylation state, total Akt level). Thus, in the same way that PCA filters noise and identifies redundant measurements within a dataset, PLS models deemphasize signals that do not productively contribute to a correlation with the dependent block. This is useful from a practical standpoint, since future experiments can be made more efficient and less costly by eliminating these measurements.

Interestingly, the DPLSR model exclusively highlighted survival classifiers from this set of signaling measurements, with no death classifiers significantly outside of random fluctuations in the data. We believe that failure to identify death classifiers is simply a reflection of the two treatment conditions that generated the samples: TNF- α treatment alone, placed in the death class, and TNF- α + insulin cotreatment, placed in the survival class. Insulin was expected to activate a number of pathways that DPLSR would identify as survival classifiers, but since TNF- α was present as a cue in both the death and survival outcomes, insulin would need to rapidly attenuate a TNF-induced pathway at all time points for the pathway to be highlighted as a death classifier. We expect that our models will extract additional classifiers, as we work toward supplementing the dataset with more diverse cue combinations (e.g., insulin alone, no cytokines).

Exploration of signal-response space by small molecule inhibitors and partial least squares

We hypothesized that data-driven models could be made to efficiently capture signal–response space when cells were treated with small molecules that mimic the action of drugs. We sought to fuse prior knowledge of key pharmacological parameters (namely, the IC₅₀ of a small molecule) with measurements

of the phenotypic effects of these drugs. Small molecule inhibitors of four protein kinases (MEK, JNK, p38, and PI3K) were used in various combinations and concentrations to quantitatively change important kinase activities (see Fig. 1(a) for the position of these kinases in the signaling network), and the effect of these inhibitors on basal, TNF-induced, and TNF- α + insulin-induced cell death in HT-29 cells was measured by flow cytometry using an apoptotic marker (see Materials and Methods). We calculated kinase activities as functions of inhibitor concentration but not cytokine treatment (see equation in top left of Fig. 8(a) and Materials and Methods), thus we treated the three stimulation conditions separately (inhibitors alone, TNF- α with inhibitors, and TNF- α + insulin with inhibitors) to model sample groups that had their kinases activated by the same cytokines. In addition, the models included nonlinear representations of kinase activity as separate measurements to account for synergistic influences of the target kinases on apoptosis (see Materials and Methods).

These nonlinear terms, together with the original calculated activities, were preprocessed and decomposed to generate PLS models for the different cytokine treatments. Figs. 8(b–d) show cross-validated predictions for the models of the three inhibitor datasets compared with measured death. We see that the PLS models capture the dynamic range of the measured dependent block with good correlation ($R^2 \sim 0.6$ – 0.8) between the predicted and observed cell death in the training set, and the models consistently identified a number of signaling activities that contributed positively or negatively to apoptosis. For instance, both p38 activity and the square of p38 activity were assigned strong positive regression coefficients (data not shown), suggesting that this pathway is a hypersensitive prodeath predictor. In contrast, the multiplicative activity of MEK \times JNK was a significant prosurvival signal for cells treated with TNF- α or both TNF- α and insulin, implying synergistic regulation of apoptosis by these pathways. These preliminary models indicate that PLS can easily incorporate pharmacological knowledge with experimental data into a coherent model that captures the combinatorial effects of drug treatments on cell response to suggest modes of drug action. It remains to be determined whether this can now be used to design combinations of drugs that will have a synergistic effect on apoptosis.

CONCLUSIONS

Signal transduction in mammalian cells involves the coordinated activities of many proteins. Modeling techniques, such as PLS, developed to explicitly treat multivariate data are helpful tools for gaining insight from large-scale biological experiments. To gain network-level understanding of the cytokine-induced death-versus-survival decision in HT-29 cells, we have applied PLS to a large set of heterogeneous protein measurements from cells treated with the death factor, TNF- α , alone and in combination with the survival factor, insulin. Our goal was to determine the properties of the signaling network that were most determinative for cell survival. We found that heterogeneous data on apoptotic signaling can be reduced, without any prior knowledge, to a PCA model that captures the most salient variations and discards less informative contributions from higher dimensions. This analysis highlights the multicomponent properties of cell signaling and identifies points of redundancy in the dataset.

Additionally, we have found that PLS can filter protein data in a more directed fashion to correlate measurements with cell phenotypes. We used outcome classes and DPLSR-based modeling to identify Akt, caspase 8, and TNFR1 as critical nodes in the signaling network that together act as a reduced set of highly informative network measurements for classifying cell survival. However, no single measurement of a single protein was as informative as the combined set of classifiers. Initial experiments also suggest that it will be possible to incorporate pharmacological data into our models for the purpose of exploring drug action. The predictive and data-fusing capabilities of PLS models suggest that a number of other biological cause-effect relationships might be productively explored with PLS and systematic measurements of underlying signal transduction pathways.

ACKNOWLEDGMENTS

This work was supported by the DARPA Bio-Info-Micro Program, the NIGMS Cell Migration Consortium, and the NIGMS Cell Decision Processes Center at MIT. K.A.J. is grateful for support from the Whitaker Foundation.

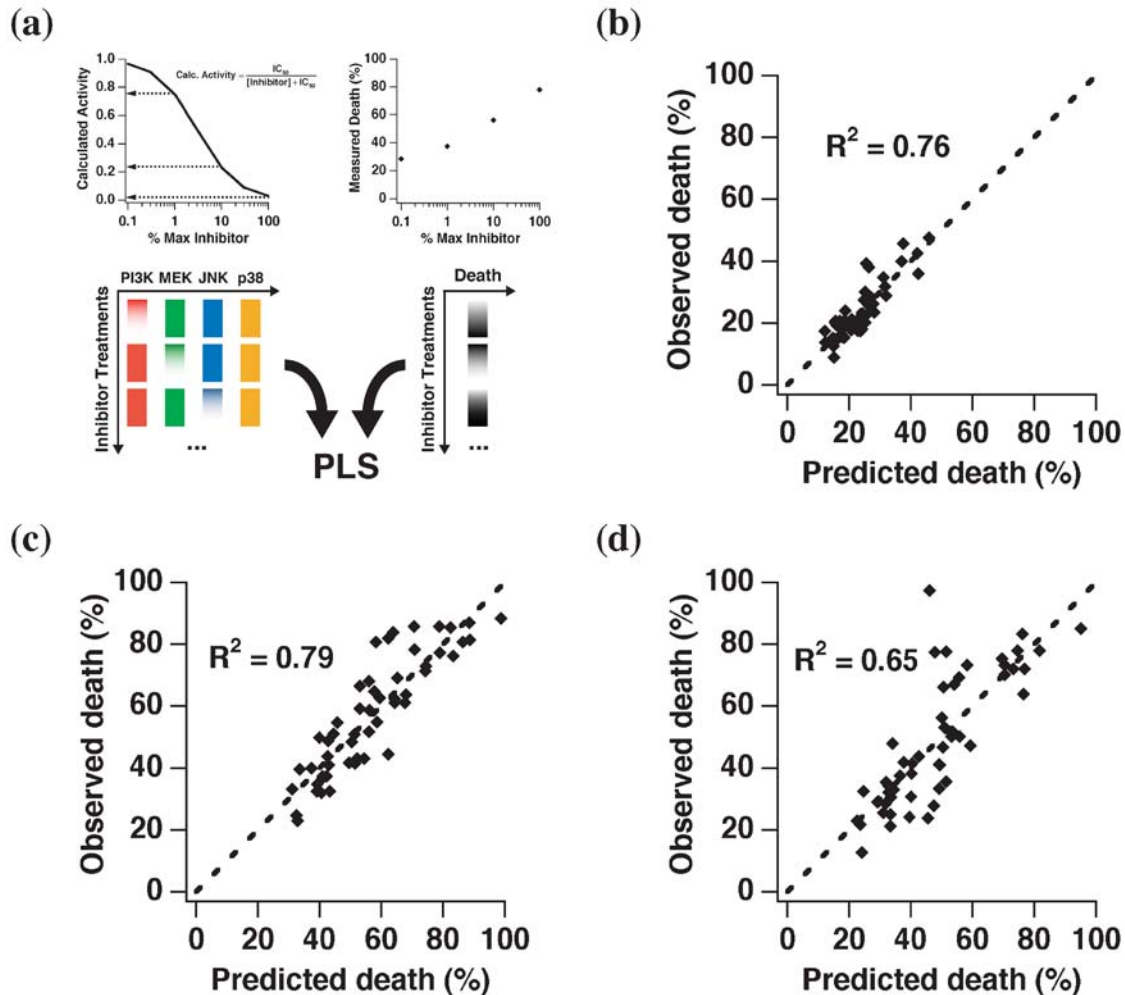


FIG. 8. Partial least squares regression quantitatively relates inhibitor-induced changes in cell signaling to continuous measurements of cell response. (a) PLS modeling framework for relating dose-response curves of pathway inhibitors to cell function. The inhibitor concentrations were used to calculate an estimated effect on pathway activity (**top left**) that was then used as an independent block with the measured effects of the inhibitor on cell death (**top right**) as the dependent block. Each inhibitor was assumed to act specifically on the intended target, with no secondary effects on the other pathways. Column shade represents the magnitude of the different samples in the independent block (**bottom left**) and dependent block (**bottom right**). Observed-versus-predicted plots for PLS models of death in response to (b) inhibitors alone, (c) TNF- α in the presence of inhibitors, (d) TNF- α + insulin in the presence of inhibitors. The models included linear and nonlinear signaling activities in the independent block, as described in Materials and Methods. The hatched lines represent what would be expected if there were perfect agreement between the model predictions and the original dataset.

REFERENCES

- Abreu-Martin, M.T., Vidrich, A., Lynch, D.H., and Targan, S.R. 1995. Divergent induction of apoptosis and IL-8 secretion in HT-29 cells in response to TNF- α and ligation of Fas antigen. *J. Immunol.* 155, 4147–4154.
- Alessi, D.R., Andjelkovic, M., Caudwell, B., Cron, P., Morrice, N., Cohen, P., and Hemmings, B.A. 1996. Mechanism of activation of protein kinase B by insulin and IGF-1. *Embo J.* 15, 6541–6551.
- Alessi, D.R., Cuenda, A., Cohen, P., Dudley, D.T., and Saltiel, A.R. 1995. PD 098059 is a specific inhibitor of the activation of mitogen-activated protein kinase in vitro and in vivo. *J. Biol. Chem.* 270, 27489–27494.
- Alter, O., Brown, P.O., and Botstein, D. 2000. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. USA* 97, 10101–10106.

- Arcaro, A., and Wymann, M.P. 1993. Wortmannin is a potent phosphatidylinositol 3-kinase inhibitor: The role of phosphatidylinositol 3,4,5-trisphosphate in neutrophil responses. *Biochem. J.* 296 (Pt 2), 297–301.
- Avruch, J. 1998. Insulin signal transduction through protein kinase cascades. *Mol. Cell. Biochem.* 182, 31–48.
- Bennett, B.L., Sasaki, D.T., Murray, B.W., O'Leary, E.C., Sakata, S.T., Xu, W., Leisten, J.C., Motiwala, A., Pierce, S., Satoh, Y., Bhagwat, S.S., Manning, A.M., and Anderson, D.W. 2001. SP600125, an anthrapyrazolone inhibitor of Jun N-terminal kinase. *Proc. Natl. Acad. Sci. USA* 98, 13681–13686.
- Brent, R. 2000. Genomic biology. *Cell* 100, 169–183.
- Cuenda, A., Rouse, J., Doza, Y.N., Meier, R., Cohen, P., Gallagher, T.F., Young, P.R., and Lee, J.C. 1995. SB 203580 is a specific inhibitor of a MAP kinase homologue which is stimulated by cellular stresses and interleukin-1. *FEBS Lett.* 364, 229–233.
- Datta, S. 2001. Exploring relationships in gene expressions: A partial least squares approach. *Gene Expr.* 9, 249–255.
- Dudek, H., Datta, S.R., Franke, T.F., Birnbaum, M.J., Yao, R., Cooper, G.M., Segal, R.A., Kaplan, D.R., Greenberg, M.E. 1997. Regulation of neuronal survival by the serine-threonine protein kinase Akt. *Science* 275, 661–665.
- Dunger, A., Schroder, D., Augstein, P., Witstruck, T., Wachlin, G., Vogt, L., Ziegler, B., and Schmidt, S. 1995. Impact of metabolic activity of beta cells on cytokine-induced damage and recovery of rat pancreatic islets. *Acta Diabetol.* 32, 217–224.
- Fasano, G., and Franceschini, A. 1987. A multidimensional version of the Kolmogorov–Smirnov test. *Monthly Notices of the Royal Astronomical Society* 225, 155–170.
- Geladi, P., and Kowalski, B.R. 1986. Partial least-squares regression—a tutorial. *Analytica Chimica Acta* 185, 1–17.
- Gomez, S.M., and Rzhetsky, A. 2002. Towards the prediction of complete protein–protein interaction networks. *Pac. Symp. Biocomput.*, 413–424.
- Holter, N.S., Mitra, M., Maritan, A., Cieplak, M., Banavar, J.R., and Fedoroff, N.V. 2000. Fundamental patterns underlying gene expression profiles: Simplicity from complexity. *Proc. Natl. Acad. Sci. USA* 97, 8409–8414.
- Huang, S., and Ingber, D.E. 2000. Shape-dependent control of cell growth, differentiation, and apoptosis: Switching between attractors in cell regulatory networks. *Exp. Cell Res.* 261, 91–103.
- Hunter, T. 2000. Signaling—2000 and beyond. *Cell* 100, 113–127.
- Ideker, T., and Lauffenburger, D. 2003. Building with a scaffold: Emerging strategies for high-to-low-level cellular modeling. *Trends Biotechnol.* 21, 255–262.
- Janes, K.A., Albeck, J.G., Peng, L.X., Sorger, P.K., Lauffenburger, D.A., and Yaffe, M.B. 2003. A High-throughput quantitative multiplex kinase assay for monitoring information flow in signaling networks: Application to sepsis-apoptosis. *Mol. Cell. Proteomics* 2, 463–473.
- Kitano, H. 2002. Systems biology: A brief overview. *Science* 295, 1662–1664.
- Kohn, A.D., Takeuchi, F., and Roth, R.A. 1996. Akt, a pleckstrin homology domain containing kinase, is activated primarily by phosphorylation. *J. Biol. Chem.* 271, 21920–21926.
- Kull, F.C., Jr. 1988. The TNF receptor in TNF-mediated cytotoxicity. *Nat. Immun. Cell Growth Regul.* 7, 254–265.
- Kuruwilla, F.G., Park, P.J., and Schreiber, S.L. 2002. Vector algebra in the analysis of genomewide expression data. *Genome Biol.* 3, RESEARCH0011.1–0011.11.
- Lawlor, M.A., and Alessi, D.R. 2001. PKB/Akt: A key mediator of cell proliferation, survival and insulin responses? *J. Cell. Sci.* 114, 2903–2910.
- Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., and Darnell, J.E. 2000. *Molecular Cell Biology*, W.H. Freeman, New York.
- Martens, H., and Martens, M. 2001. *Multivariate Analysis of Quality: An Introduction*, John Wiley, Chichester, UK.
- Misra, J., Schmitt, W., Hwang, D., Hsiao, L.L., Gullans, S., and Stephanopoulos, G. 2002. Interactive exploration of microarray gene expression patterns in a reduced dimensional space. *Genome Res.* 12, 1112–1120.
- Nguyen, D.V., and Rocke, D.M. 2002. Tumor classification by partial least squares using microarray gene expression data. *Bioinformatics* 18, 39–50.
- Nicholson, D.W., and Thornberry, N.A. 1997. Caspases: Killer proteases. *Trends Biochem. Sci.* 22, 299–306.
- Remacle-Bonnet, M.M., Garrouste, F.L., Heller, S., Andre, F., Marvaldi, J.L., and Pommier, G.J. 2000. Insulin-like growth factor-I protects colon cancer cells from death factor-induced apoptosis by potentiating tumor necrosis factor alpha-induced mitogen-activated protein kinase and nuclear factor kappaB signaling pathways. *Cancer Res.* 60, 2007–2017.
- Rives, A.W., and Galitski, T. 2003. Modular organization of cellular networks. *Proc. Natl. Acad. Sci. USA* 100, 1128–1133.
- Rohlf, F.J., and Corti, M. 2000. Use of two-block partial least-squares to study covariation in shape. *Syst. Biol.* 49, 740–753.
- Sachs, K., Gifford, D., Jaakkola, T., Sorger, P., and Lauffenburger, D.A. 2002. Bayesian network approach to cell signaling pathway modeling. *Sci. STKE* 2002, PE38.
- Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D., and Muller, G. 2002. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.* 20, 370–375.

- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. 2002. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* 417, 399–403.
- Wajant, H., Pfizenmaier, K., and Scheurich, P. 2003. Tumor necrosis factor signaling. *Cell Death Differ.* 10, 45–65.
- Wilson, C.A., and Browning, J.L. 2002. Death of HT29 adenocarcinoma cells induced by TNF family receptor activation is caspase-independent and displays features of both apoptosis and necrosis. *Cell Death Differ.* 9, 1321–1333.
- Xia, Z., Dickens, M., Raingeaud, J., Davis, R.J., and Greenberg, M.E. 1995. Opposing effects of ERK and JNK-p38 MAP kinases on apoptosis. *Science* 270, 1326–1331.
- Yeung, M.K., Tegner, J., and Collins, J.J. 2002. Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl. Acad. Sci. USA* 99, 6163–6168.

Address correspondence to:

D.A. Lauffenburger
Biological Engineering Division
56-341
Massachusetts Institute of Technology
Cambridge, MA 02139

E-mail: lauffen@mit.edu