A Compendium of Signals and Responses Triggered by Prodeath and Prosurvival Cytokines*  
Suzanne Gaudet‡, Kevin A. Janes§, John G. Albeck‡, Emily A. Pace¶, Douglas A. Lauffenburger‡§, and Peter K. Sorger‡$  

Cell-signaling networks consist of proteins with a variety of functions (receptors, adaptor proteins, GTPases, kinases, proteases, and transcription factors) working together to control cell fate. Although much is known about the identities and biochemical activities of these signaling proteins, the ways in which they are combined into networks to process and transduce signals are poorly understood. Network-level understanding of signaling requires data on a wide variety of biochemical processes such as posttranslational modification, assembly of macromolecular complexes, enzymatic activity, and localization. No single method can gather such heterogeneous data in high throughput, and most studies of signal transduction therefore rely on series of small, discrete experiments. Inspired by the power of systematic datasets in genomics, we set out to build a systematic signaling dataset that would enable the construction of predictive models of cell-signaling networks. Here we describe the compilation and fusion of ~10,000 signal and response measurements acquired from HT-29 cells treated with tumor necrosis factor-α, a proapoptotic cytokine, in combination with epidermal growth factor or insulin, two prosurvival growth factors. Nineteen protein signals were measured over a 24-h period using kinase activity assays, quantitative immunoblotting, and antibody microarrays. Four different measurements of apoptotic response were also collected by flow cytometry for each time course. Partial least squares regression models that relate signaling data to apoptotic response data reveal which aspects of compendium construction and analysis were important for the reproducibility, internal consistency, and accuracy of the fused set of signaling measurements. We conclude that it is possible to build self-consistent compendia of cell-signaling data that can be mined computationally to yield important insights into the control of mammalian cell responses.  

Molecular & Cellular Proteomics 4:1569–1590, 2005.

The development of numerical models of complex biological processes such as signal transduction depends on the availability of accurate, self-consistent, and quantitative experimental data. Large scale collection and systematization of such data is likely to have as great an impact on cell biology as complete genome sequencing has had on genetics (1, 2). A few high throughput signaling datasets describing lymphokine-induced protein networks (3) and phosphotyrosine networks (4) have recently become available, and the Alliance for Cellular Signaling continues its efforts to collect data on large numbers of signaling proteins (5). However, the assembly of cell-signaling datasets is in its infancy and faces challenges not encountered in genomics. First, unlike gene sequences, signaling data are necessarily heterogeneous (6). Activation of intracellular signal transduction networks involves biochemical processes as distinct as receptor-ligand binding, protein modification, assembly of multicomponent complexes, translocation among cellular compartments, and protein synthesis and destruction (hereafter referred to as “signals”). Whereas the sequence of an entire genome can be determined using a single mature technology and genome-wide transcriptional profiling can be accomplished with microarray methods alone, monitoring cell signaling requires a wide variety of experimental techniques including enzymatic assays (7), immunoblotting (8), flow cytometry (3), immunofluorescence (9), and mass spectrometry (10). Only a subset of these approaches are currently used in large scale data collection (phosphoprotein mass spectrometry for example) in part because techniques that are mainstays of standard cell biology, such as immunoblotting and immunofluorescence, are not well adapted to high throughput analysis.  

A second complication in systematizing cell-signaling measurements is the lack of inherent structure in the data. Structured data, such as DNA sequence, can be referenced to a single data model and systematized via relational databases. Manipulating unstructured data, such as microscopy and immunoblot images, is a significant informatic challenge to which current databases are not well suited (11). Third, unlike gene sequences, the meaning of data on kinase activation, receptor-ligand interaction, signalosome assembly, and phospholipid modification depends heavily on context (6, 11). Critical aspects of biological context include cell type,
nature and concentration of a cytokine stimulus, and cell cycle or growth phase. Experimental methods also contribute to context because quantitative measurements vary with reagent source and protocol. Finally signaling data are necessarily incomplete: whereas a genomic sequence has a well defined start and finish, the characterization of a complex signaling network is an open-ended process.

This study examined signaling networks that control the apoptosis-survival decision in HT-29 human colon adenocarcinoma cells treated with combinations of the prodeath cytokine tumor necrosis factor-α (TNF)\(^1\) and the prosurvival growth factors epidermal growth factor (EGF) and insulin. TNF activates intracellular signals by binding to trimeric death receptors and promoting assembly of intracellular death-inducing signaling complexes (12, 13). Death-inducing signaling complexes then activate several parallel signaling pathways, including apoptotic caspases (14), stress-activated c-jun N-terminal kinase (JNK) and p38 pathways (15), and the proinflammatory IκB kinase (IKK)-nuclear factor-κB pathway (16).

Activation of the EGF receptor (EGFR) tyrosine kinase occurs through receptor dimerization, conformational change, and autophosphorylation (17, 18). Phosphorylated receptors recruit adaptor proteins, and these then activate multiple signaling proteins including extracellular signal-regulated kinase (ERK) via Ras and the Akt kinase via phosphatidylinositol 3-kinase (19). The binding of insulin to the insulin receptor also activates ERK and Akt, but in contrast to EGFR, the insulin receptor is constitutively dimerized, and most insulin-induced signaling involves modification of insulin receptor substrate 1 (IRS1), a multidomain adaptor protein (20). Antagonism between TNF and EGF or TNF and insulin is well documented and relevant to many diseases, including cancer, inflammatory bowel disease, and diabetes.\(^2\) However, there exists only limited molecular level understanding of the strategies cells use to process conflicting cytokine stimuli and regulate responses such as apoptosis. The limitation arises in part because current practice in signal transduction research, as in other areas of cell biology, is to interpret data only with respect to contemporaneous controls. Categorical conclusions are drawn from direct comparison of experimental and control samples (e.g. ERK is more, less, or equally active between samples), and the next experiment is designed. In studies of complex networks, however, many variables must be measured, and the piecemeal accumulation and analysis of data become both cumbersome and limiting.

In this study we asked whether a validated and self-consistent dataset could be constructed from diverse biochemical measurements of cell-signaling proteins in human cells exposed to TNF, EGF, and insulin. Using techniques available in the average research laboratory, we normalized and fused into a single dataset ~10,000 data points collected over an 18-month period. We have not yet tackled the informatic challenge of building a database to hold the information, relying instead on a series of customized spreadsheets. We therefore refer to the assembled dataset as a compendium. The heterogeneity, lack of intrinsic structure, context dependence, and open-ended nature of the signaling data make the construction of this cytokine-signal-response (CSR) compendium nontrivial (22). However, we show that the compendium makes it possible to compare, with high reliability, data that were not collected contemporaneously. We also show that data can be mined for insights that would not be obvious from more conventional approaches. Finally because the value of a data compendium can only be unlocked computationally, we analyzed the full dataset using a predictive statistical model based on partial least squares regression (PLSR).\(^3\) We will describe elsewhere the use of modeling to elucidate the mechanisms by which cells process conflicting cytokine stimuli; here we apply modeling to the more general problem of assessing the quality and information content of a CSR compendium. In contrast to many other proteomic efforts, we incorporated modeling early in the data collection process by linking protein signaling to relevant phenotypic responses. The inclusion of responses is necessary to constrain statistical models sufficiently to provide useful quantitative assessment of compendium design. Our key findings are that heterogeneous, time-resolved information is indeed critical for predictive power as is sampling broadly across the network. Nonetheless it appears sufficient to monitor a subset of network reactions, significantly reducing the data collection burden on future proteomic compendia.

**EXPERIMENTAL PROCEDURES**

*Cell Culture and Treatments—* HT-29 human colon adenocarcinoma cells (ATCC) were maintained in McCoy’s 5A supplemented with 10% fetal bovine serum, 2 mm glutamine, 10 units/ml penicillin, and 10 units/ml streptomycin (Invitrogen). Cells were used within passage 10 after thaw for all experiments and passaged 2 days prior to plating. For each experiment, HT-29 cells were plated at 50,000 cells/cm\(^2\). After 24 h, the cells were sensitized by changing to medium

---

1. The abbreviations used are: TNF, tumor necrosis factor α; CSR, cytokine-signal-response; EGF, epidermal growth factor; EGFR, epidermal growth factor receptor; ERK, extracellular signal-regulated kinase; FKHR, Forkhead transcription factor; pFKHR, phospho-FKHR; FOXO3, Forkhead box O3a; IKK, IκB kinase; IL-1α, interleukin-1α; IL-1R, IL-1 receptor; IL-1ra, IL-1 receptor antagonist; IR, insulin receptor; IRS1, insulin receptor substrate 1; pIRS1, phospho-IRS1; JNK, c-jun N-terminal kinase; MEK, mitogen-activated protein kinase and extracellular signal-regulated kinase kinase; MK2, mitogen-activated protein kinase-activated protein kinase 2; PLSR, partial least squares regression; ptAkt, phospho-to-total Akt; ptEGFR, phospho-to-total EGFR; SGK, serum- and glucocorticoid-inducible kinase; TGF-α, transforming growth factor α; VIP, variable importance in the projection.

2. K. A. Janes, S. Gaudet, J. G. Albeck, U. B. Nielsen, D. A. Lauffenburger, and P. K. Sorger, submitted for publication.

3. K. A. Janes, J. G. Albeck, S. Gaudet, P. K. Sorger, D. A. Lauffenburger, and M. B. Yaffe, submitted for publication.
supplemented with 200 units/ml interferon-γ at a ratio of −0.13 ml of medium/cm². The cultures were stimulated 24 h later by adding the stimulus diluted in 1/20 of the culture volume of serum-free medium for final concentrations of 0, 0.2, 5, or 100 ng/ml TNF (Peprotech); 0, 1, or 100 ng/ml EGF (Peprotech); and 0, 1, 5, or 500 ng/ml insulin (Sigma). The C225 anti-EGFR monoclonal antibody (H. S. Wiley laboratory) and IL-1α (Amgen) inhibitors were used at 10 μg/ml. Triplicate plates were lysed at 0, 5, 15, 30, 60, and 90 min and 2, 4, 8, 12, 16, 20, and 24 h or prepared for flow cytometry at 12, 24, and 48 h.

**Cell Lysis**—Cell lysis had to be compatible with multiple assays and combine dead (floating) and viable cells. Floating cells were pelleted from culture supernatant and rinsed with ice-cold PBS. Adherent cells were lysed in ice-cold lysis buffer (as described in Ref. 7) and combined with the dead cells pelleted from the culture medium. An aliquot of the whole-cell homogenate was immediately frozen for use in immunoblot analysis. The lysates were clarified by centrifugation (10 min at 15,000 rpm) and aliquoted for antibody array and kinase assay. Total protein concentration was determined using the bicinchoninic acid assay (Pierce) for both the clarified lysates and the whole-cell homogenate following addition of SDS-based lysis buffer to final concentrations of 112.5 mM Tris-Cl, pH 7.5, 4% SDS, 10 mM β-glycerophosphate, 10 mM NaF, 10 mM sodium pyrophosphate, 1 mM Na₃VO₄, 1 μg/ml leupeptin, 1 μg/ml pepstatin, and 1 μg/ml chymostatin.

**Kinase Assays**—Kinase assays were performed as described previously (7). All signaling measurements were normalized to total protein concentration in each lysate. For kinase assays, we also corrected for variations in ATP specific activity; all measurements were normalized to total protein concentration. Relative units were calculated by comparing the day-specific ratio of t = 0 min (untreated) kinase activity/background activity to a compendium-wide average. The data were normalized so that one relative unit corresponded to the compendium-averaged t = 0 min activity for the kinase.

**Antibody Microarray**—Antibody microarrays were processed as described previously (24). For EGFR, the same antibodies were used (24), for Akt array measurements anti-Akt (Upstate Biotechnologies) was used as a capture antibody, and anti-phospho-Akt and anti-pan-specific Akt (catalog numbers 4051 and 9272, Cell Signaling Technologies) were used as detection antibodies. We obtained three measurements for EGFR and Akt, phosphoprotein levels, total protein level, and phosphorylated to total protein ratio, from trilicate spots for each lysate. The ratio metric measurement was used to account for spot-to-spot variation in printing. All three measures were normalized to total protein concentration and to the t = 0 min average for each time course and are therefore expressed as fold activation over the base-line signals.

**Immunoblots**—For immunoblots, 75 μg of total cell homogenates were separated by SDS-PAGE and transferred to nitrocellulose. Reference samples of HT-29 cells treated with 200 units/ml interferon-γ for 48 h and then 500 ng/ml insulin for 30 min or 100 ng/ml TNF for 40 h were run on each gel for normalization purposes. Membranes were cut at specific molecular weights to allow probing with several antibodies for each gel. Immunopробing was performed as described previously (25). Primary antibodies used were anti-phospho-Akt-Ser⁷⁷² (catalog number 9271), anti-caspase-8 1C12 (catalog number 9746), anti-phospho-FKHR-Ser⁹⁶⁸ (catalog number 9461), anti-phospho-IRS-Ser⁷⁸⁰ (catalog number 2388), and anti-phospho-MEK1/2-Ser²⁷⁷/²⁷² (catalog number 9121) from Cell Signaling Technologies; anti-phospho-IRS1-Tyr⁹⁸⁰ (catalog number GF1003) from Calbiochem; and anti-caspase-3 (1:250, catalog number SC-7272) from Santa Cruz Biotechnologies. The fluorescence was assessed using a Fluormaster 495 scanner (Amersham Biosciences) and quantitated using ImageQuant (Amersham Biosciences). Signal intensities were normalized to intensities in the reference samples to correct for differences in transfer and probing efficiencies. Signals where the t = 0 min average was consistently above background level were normalized to this basal level for each time course and expressed in fold activation units.

**Flow Cytometry**—Cells were grown in 24-well plates and treated with the indicated cytokines as described above. For each experiment, triplicate reference samples (untreated, treated with 100 ng/ml TNF, and treated with 100 ng/ml TNF + the prosurvival factor used in that experiment) were collected for normalization purposes and to verify the bioactivity of the cytokines. The supernatant and rinse were saved and combined with the trypsinated cells to assay both floating and adherent cells. Live cells were stained with Alexa 488-labeled annexin V and 1 μg/ml propidium iodide (Molecular Probes). Fixed cells (4% formaldehyde for 10 min followed by 100% methanol) were stained with 50 μg/ml propidium iodide for sub-G₁ DNA content or with M30 monoclonal antibody against cleaved cytokeratin (Roche Applied Science) and anti-cleaved caspase-3 (Cell Signaling Technologies, BD Biosciences) followed by Alexa 488-conjugated donkey anti-mouse IgG and Alexa 647-conjugated donkey anti-rabbit IgG (Molecular Probes). Samples were analyzed on a BD Biosciences FACScan or FACScan. Apoptosis measurements were normalized to average values found in contemporaneously treated and processed reference samples to account for variations in staining efficiency: 100 ng/ml TNF-treated reference samples were used for propidium iodide permeability and cleaved caspase-3 and cytokeratin staining assays, and untreated reference samples were used for annexin V binding and sub-G₁ DNA content assays.

**Partial Least Squares Model**—The descriptors of signaling dynamics and partial least squares models were generated using MATLAB® and SIMCA-P, respectively, as described elsewhere.³ The fitness of values predicted by the model to the measured values for cell death was evaluated according to the equation,

$$\text{Fitness} = 1 - \frac{\sum_{i=1}^{n} (\text{Predicted}_i - \text{Measured}_i)^2}{\sum_{i=1}^{n} \text{Predicted}_i^2}$$  \hspace{1cm} (Eq. 1)

where n is the number of response measurements. The non-parametric 90% confidence intervals on the fitness to measured values were evaluated, based on the Fisher inversion, on the average $R^2$ value for seven models from cross-validation,

$$x = \frac{(n + 1) - y} {n + 1}$$

$$y = \frac{1}{2} \left( 1 + \frac{R^2}{1 - R^2} \right)$$

$$\pm 1.645 \sqrt{\frac{n}{n - 3}}$$  \hspace{1cm} (Eq. 2)

The variable importance in the projection (VIP) was calculated using

$$\text{VIP}_k = \frac{1}{k} \sum_{a=1}^{A} w_{ak} SS_a$$  \hspace{1cm} (Eq. 3)

where k is the number of variables, $w_{ak}$ is the weight of the $k$th variable for principal component $a$, A is the total number of principal components, and $SS_a$ is the sum of squares explained by principal component $a$.  

**Molecular & Cellular Proteomics 4.10**

![Image](https://via.placeholder.com/150)
To mimic measurement error, random noise was added to the signaling measurements in the 5 ng/ml TNF + 1 ng/ml EGF treatment from a normal distribution centered at zero with a standard deviation of 0.1, 0.2, 0.3, 0.4, or 0.5 times the range of values for the time course in that signal. Time course descriptors were then extracted as described elsewhere.

RESULTS

Data Generation

To explore systematically the relationships between cytokine-receptor interaction, activation of intracellular signaling cascades, and apoptosis-survival cell fate decisions, HT-29 human colon adenocarcinoma cells were exposed to a set of 10 cytokine treatments and monitored over a 48-h period. Each treatment consisted of a combination of TNF and either EGF or insulin (Fig. 1A). HT-29 cells respond to TNF, EGF, and insulin in a dose-dependent manner (26, 27), and all three cytokines were therefore examined at subsaturating concentrations designed to mimic physiological conditions and at saturating concentrations at which essentially all receptors were ligand-bound. At 13 time points after cytokine addition, three replicate dishes of cells (six for the zero time point) were harvested to measure kinase activities, changes in protein phosphorylation, caspase cleavage, and changes in protein abundance. Altogether, 19 protein distinct signals were extracted as time course descriptors. Time course measurements in this initial dataset was 8,340, collected over a period of 18 months by four scientists working in collaboration.

To characterize the responses of cells to cytokine combinations, four measures of apoptosis were performed on triplicate plates 12, 24, and 48 h after stimulation (Fig. 1B). These four measures represent distinct stages or aspects of programmed cell death (Fig. 1C). Both adherent (viable) and floating (dying) cells were examined in all cases. In TNF-treated cells, procaspase-8, an initiator caspase, is activated by cleavage and then cleaves and activates the executioner caspase, caspase-3, leading to the degradation of a wide variety of essential proteins. The cleavage of procaspase-3 and -8 measured by immunoblots constituted signals in our dataset, whereas downstream events, measured by flow cytometry, were considered to be responses. The cleavage of caspase-3 and -8, as well as its target, cytochrome c, was measured using cleavage-specific antibodies (Table I, response R1). DNA digestion by endonucleases and subsequent nuclear fragmentation were quantified by propidium iodide staining to detect cells with sub-G1 DNA content (R2). Loss of plasma membrane asymmetry was assayed by the binding of annexin V to exposed phosphatidylserine, a phospholipid normally restricted to the inner membrane leaflet (R3). Finally, the complete loss of plasma membrane integrity was assayed using a vital stain (R4). Most of the apoptotic response measures are end points that increase steadily over time. However, phosphatidylserine exposure (R3) represents a transient state fairly early in apoptosis in which membrane structure is partially but not yet completely disrupted. The final product of the signal and response measurements was a set of 10 time courses, each describing the activities of 19 protein signals measured at 13 time points and four response measures at three time points, all in triplicate biological samples. The total number of measurements in this initial dataset was 8,340, collected over a period of 18 months by four scientists working in collaboration.
Creation of a Cytokine-Signal-Response Data Compendium by Data Normalization, Fusion, and Validation

Our overall aim was to merge time course data collected at different times into a single self-consistent compendium. This involved a variety of experimental and data analysis techniques, some of which were obvious, whereas others were more subtle. One obvious step was to limit day-to-day and sample-to-sample variation by standardizing assay protocols, limiting cells to a narrow range of passage number, and ensuring uniformity in cell culture and treatment conditions through the use of lot-controlled sera, cytokines, and antibodies (supplemental material, Table S1). Raw data were corrected for other sources of variation by signal-specific normalization. As described in greater detail under “Experimental Procedures,” normalization included the following: 1) correcting each signaling measurement for the total amount of cellular protein in the sample, 2) adjusting kinase activity measurements for changes in the specific activity of \[^{32}P\]ATP, and 3) comparing immunoblot signals to contemporaneous positive controls to correct for variations in transfer efficiency and antibody binding. Signals with \(t_{100}\) normalized averages that were consistently above background were divided by the \(t_{100}\) average for that treatment to yield a measure of “-fold activation.” Other signals were expressed in activity units referenced to a base line of zero. Apoptosis measurements were compared with contemporaneous positive and negative control cell populations to yield a normalized cell death fraction. We found that careful normalization was essential for data fusion: by minimizing the contributions of experimental variability on signal measurements, it was possible to reduce the median coefficient of variation among biological replicates to \(11\%\).

As one means to judge how reliably data from different assays had been fused in the CSR compendium, we examined correlations among three measures of Akt protein kinase activation. Immunoblots and antibody microarrays were used to monitor the levels of phospho-Ser473, a modification that is associated with Akt activation (31, 32) (Table I and Fig. 2A). An in vitro immunocomplex kinase assay was used to monitor Akt-catalyzed phosphorylation of a sequence-optimized peptide substrate (Table I and Fig. 2A). After normalization, we found that all three measures of Akt were correlated across the full panel of cytokine combinations with pairwise Pearson coefficients of 0.80–0.87 (Fig. 2, B–D). Although fully activated only when doubly phosphorylated, Akt

**Table I**

| No. | Assay     | Protein   | Name                      | Function            | Marker         | Abbreviation |
|-----|-----------|-----------|---------------------------|---------------------|----------------|--------------|
| 1   | Kinase    | Akt       | Akt/PKB                   | Ser-Thr kinase      | Activity       | Akt*         |
| 2   | ERK       | Extracellular signal-regulated kinase | Ser-Thr kinase | Activity | ERK         |
| 3   | IKK       | Inhibitor of nuclear factor-\(\alpha B\) kinase | Ser kinase       | Activity | IKK         |
| 4   | JNK1      | c-Jun N-terminal kinase | Ser-Thr kinase | Activity | JNK1         |
| 5   | MK2       | MAPK-associated protein kinase | Ser-Thr kinase | Activity | MK2         |
| 6   | Immunoblot| Akt       | Akt/PKB                   | Cysteine protease   | Zymogen level  | ProC3        |
| 7   | Casp.3    | Caspase-3 | Caspase-3                 | Cysteine protease   | Zymogen level  | ProC8        |
| 8   | Casp.8    | Caspase-8 | Caspase-8                 | Cleaved level       | Cleaved        | ClvC8        |
| 9   | IRS1      | Forkhead receptor substrate | Transcription factor | Adaptor-scaffold | Phospho-Ser\[^{256}\]pFKHR |
| 10  | MEK1/2    | MAPK and ERK kinase | Dual specificity kinase | Phospho-Ser\[^{217,221}\]pMEK |
| 11  | Ab array  | Akt       | Akt/PKB                   | Ser-Thr kinase      | Total level    | Akt          |
| 12  | EGFR      | Epidermal growth factor receptor | Receptor Tyr kinase | Total level | tEGFR        |
| 13  | R1        | Flow cytometry | Cleaved caspase-3 and cytokeratin | Nuclear fragmentation | Cleaved levels | CC3/CCK |
| 14  | R2        | Sub-G\(_1\), DNA content | Low PI staining | Sub-G\(_1\) |
| 15  | R3        | Phosphatidylserine exposure | Recognition by macrophages | Annexin V binding | PS exposure |
| 16  | R4        | Membrane permeabilization | Secondary necrosis | Memb. perm. |

*Signals used to validate measurements across assays.

Creation of a Cytokine-Signal-Response Data Compendium by Data Normalization, Fusion, and Validation

Our overall aim was to merge time course data collected at different times into a single self-consistent compendium. This involved a variety of experimental and data analysis techniques, some of which were obvious, whereas others were more subtle. One obvious step was to limit day-to-day and sample-to-sample variation by standardizing assay protocols, limiting cells to a narrow range of passage number, and ensuring uniformity in cell culture and treatment conditions through the use of lot-controlled sera, cytokines, and antibodies (supplemental material, Table S1). Raw data were corrected for other sources of variation by signal-specific normalization. As described in greater detail under “Experimental Procedures,” normalization included the following: 1) correcting each signaling measurement for the total amount of cellular protein in the sample, 2) adjusting kinase activity measurements for changes in the specific activity of \[^{32}P\]ATP, and 3) comparing immunoblot signals to contemporaneous positive controls to correct for variations in transfer efficiency and antibody binding. Signals with \(t_{100}\) min normalized averages that were consistently above background were divided by the \(t_{100}\) min average for that treatment to yield a measure of “-fold activation.” Other signals were expressed in activity units referenced to a base line of zero. Apoptosis
is at least partially active with only one modification (35, 36). We might therefore expect to see situations in which Akt activity was relatively high but Ser473 phosphorylation was low; this is exactly what was observed in cells treated with saturating TNF+$\alpha$ plus insulin at $t$ = 2–12 h (Fig. 4A). Other discrepancies in measures of Akt activation could similarly be explained by current understanding of its regulation. In summary, the strong correlation among three different measures of Akt kinase activity argues that data from different biochemical assays can be fused into one compendium. It is important to note, however, that correlations among different measures of protein function are critically dependent on understanding the mechanisms of signal activation and inhibition.

Signals and responses were acquired as time course experiments, one cytokine combination at a time over a period of 18 months. To enable direct comparison of data collected non-contemporaneously, it was essential that the CSR compendium be self-consistent. Reproducibility over time is the most obvious criterion for self-consistency. To address the issue of reproducibility, we generated a validation dataset by using an experimental scheme orthogonal to that of the primary dataset. The validation data consisted of the 19 protein signals measured in triplicate at a single time point ($t$ = 15 min) following exposure to eight different cytokine treatments; this corresponds to a single-time slice through the CSR data space (Fig. 3A). Apoptotic responses were also assayed at 12, 24, and 48 h in this experiment. We observed good correlations between values in the validation dataset and the initial CSR compendium: most signals and responses had correlation coefficients of 0.9 or higher. However, a subset of signals were poorly correlated. We noticed that these signals did not vary substantially from one treatment to the next, and thus the Pearson correlation coefficient was dominated by experimental noise. To distinguish high and low variance signals, we calculated their “dynamic range” by taking the variation in a signal (measured across treatments) and dividing by the measurement noise (from triplicate biological samples; see Fig. 3 legend for details). A signal such as phospho-Ser473 Akt (measured by immunoblotting) had a high dynamic range (7.5) at $t$ = 15 min and was characterized by a high correlation between the single treatment and single time experiments ($R$ = 0.98; Fig. 3B). IKK varied little among treatments (dynamic range, 1.5), and the correlation between compendium and validation data was low, although the absence of observable differences between the two datasets emphasizes that the issue was primarily one of noise and not inconsistency (Fig. 3C). Overall the great majority of signals and responses had a dynamic range above 2.5 and were well correlated between the compendium and the validating single time dataset (median Pearson coefficient of 0.95 for responses and 0.91 for signals; Fig. 3D). Monte Carlo simulations indicated that a Pearson coefficient of 0.9 corresponds to a $\approx$14% coefficient of variation between measurements. This is only slightly larger than the 11% median coefficient of variation observed among biological replicates assayed at the same time, suggesting that day-to-day variation in compendium data was similar in magnitude to biological variation and measurement noise. We conclude that our experimental methods and data fusion techniques had therefore generated a self-consistent compendium in which data collected at different times could be compared with confidence.

Fig. 2. Heterogeneous Akt measurements are quantitatively comparable in the fused compendium. A, schematic showing three measures of Akt activation: kinase activity assays (KA), phospho-Ser473 levels by immunoblot (IB), and ratio of phospho-Ser473 to total levels by antibody array (AA). B–D, scatter plots of pairwise comparisons of heterogeneous Akt activation measurements: immunoblot phospho-Akt-Ser473 assay compared with kinase activity (B), pAkt ratio from antibody arrays compared with kinase activity (C), and immunoblot phospho-Akt-Ser473 assay compared with pAkt ratio from antibody array (D). Pearson correlation coefficients are indicated above each plot ($R$). The pAkt measurements by antibody microarray are the most distant because they are an aggregate measure of phospho-to-total levels of Akt. Values in B–D are means ± S.E. of biological triplicate samples normalized to the maximal value in the compendium for that signal. a.u., arbitrary units.
Compendium of Signals and Responses to TNF, EGF, and Insulin

A

Intracellular signals
Cell death markers
Treatments
Signals
Responses
Time course
48 h

B

pAkt-S473 (IB) at t = 15 min
Dynamic range = 7.5, R = 0.98

C

IKK activity at t = 15 min
Dynamic range = 1.5, R = 0.00

D

From compendium
From single-time experiment

| Signal          | From compendium | From single-time experiment | Dynamic Pearson range coefficient |
|-----------------|-----------------|----------------------------|----------------------------------|
| JNK1 (KA)       | 57.7            | 0                          | 0.95                             |
| MK2 (KA)        | 27.0            | 0                          | 0.94                             |
| CC3/CCK (48h)   | 15.2            | 0                          | 0.98                             |
| Memb. perm. (48h) | 13.2          | 0                          | 0.96                             |
| pIRS1-636 (IB)  | 10.9            | 0                          | 0.73                             |
| ERK (KA)        | 10.8            | 0                          | 0.85                             |
| Memb. perm. (24h) | 8.6            | 0                          | 0.96                             |
| CC3/CCK (24h)   | 8.3             | 0                          | 0.98                             |
| pMEK (IB)       | 7.6             | 0                          | 0.68                             |
| pFKHR (IB)      | 7.5             | 0                          | 0.98                             |
| pAkt (IB)       | 7.5             | 0                          | 0.98                             |
| CC3/CCK (12h)   | 7.4             | 0                          | 0.97                             |
| PS exposure (12h)| 7.0            | 0                          | 0.95                             |
| Akt (KA)        | 6.2             | 0                          | 0.92                             |
| pFKHR (IB)      | 5.3             | 0                          | 0.40                             |
| pAkt (AA)       | 4.9             | 0                          | 0.91                             |
| pAkt-896 (IB)   | 4.9             | 0                          | 0.98                             |
| Memb. perm. (12h)| 4.8             | 0                          | 0.79                             |
| SubG1 DNA (12h)| 4.3             | 0                          | 0.93                             |
| PS exposure (24h)| 3.5            | 0                          | 0.90                             |
| SubG1 DNA (48h)| 2.5             | N.A.                       | 0.93                             |
| p1EGFR (AA)     | 2.4             | 0                          | N.A.                             |
| IKK (KA)        | 1.5             | N.A.                       | 0.9                 |
| C1cB (IB)       | 1.5             | N.A.                       | 0.9                 |
| 1-ProC3 (IB)    | 0.7             | N.A.                       | 0.9                 |
Key Features of the Signaling Compendium Identified by Cytokine-Signal Covariates

To visualize the entire CSR compendium in a compact form, a heat map was constructed (Fig. 4, A and B). The error analysis described above (Fig. 3D) suggested that quantitative differences larger than ~30% between observed signals or responses were likely to be biologically significant. Most variation in the compendium was much larger than this, ranging from 3-fold changes in pTEGFR to over 50-fold changes in JNK and MK2 activity. Many of the patterns that could be discerned from the heat map were expected (e.g. EGF and insulin significantly inhibited TNF-induced caspase-8 cleavage at t = 8–24 h; p < 10^{-4}; Fig. 4A), but several were unanticipated (see supplemental material, Figs. S1–S4). In these cases the primary value of the CSR compendium was its ability to put each measured signal within the broader biological context of other signals, treatments, and time points.

In addition to visual inspection, the quality of the CSR data enabled us to examine quantitative correlations between cytokines and signals. We applied a previously described statistical mapping technique based on eigenvalue decomposition and rotation (25).2 Briefly each protein signal in the compendium was integrated over its 24-h time course and then projected onto a set of classifiers describing the TNF and EGF or insulin treatment (see Ref. 25 for details). This projection generated a map in which the positions of 19 protein signals were plotted relative to the TNF, EGF, and insulin stimuli. The closer a protein signal is to a cytokine, the more strongly and specifically the signal is activated by that cytokine. For example, TNF strongly induces caspase-8 cleavage and JNK1 and MK2 activation (14, 15), and these signals were closest to TNF on the cytokine-signal map (Table II). Of particular interest were unanticipated map positions. We illustrate this point with two examples: one in which an unexpectedly close correlation was observed between EGF and pIRS1-Tyr896 and a second in which a lack of correlation was strongly observed between the Akt kinase and its substrate, pFKHR.

**IRS1-Tyr896 as a Candidate EGFR Phosphorylation Site**—The signal that mapped closest to EGF was not a measure of EGF receptor activation but IRS1 phosphorylation at Tyr896 (Table II). IRS1 is a key adaptor protein associated with insulin signaling and is known to contain at least 12 functionally important phosphorylation sites (37, 38). IRS1 tyrosine phosphorylation, on sites such as Tyr612, Tyr941, and Tyr989, is strongly induced by insulin and insulin-like growth factor receptor tyrosine kinases (37, 39). Whereas IRS1 tyrosine phosphorylation stimulates downstream signaling, serine phosphorylation is inhibitory (40, 41). The kinase that mediates IRS1-Tyr896 phosphorylation has not been identified, but the modification is known to create an Src homology 2 binding site for Grb2 (growth factor receptor-bound 2) and thus to serve as a positive regulator of signaling (39, 42) (Fig. 1C).

In agreement with results from the cytokine-signal map, strong IRS1-Tyr896 phosphorylation was observed in time courses of EGF-treated cells, but the modification was undetectable in the presence of insulin or TNF alone (Fig. 4, A and C). EGF-induced IRS1 modification was not unique to HT-29 cells: EGFR-dependent IRS1-Tyr896 phosphorylation was also observed in A431 epidermoid carcinoma cells.4 EGF has previously been reported to promote tyrosine phosphorylation of IRS1, but specific sites of phosphorylation have not been identified (43–45). To search for candidate IRS1-Tyr896 kinases, we used the Scansite motif-based searching algorithm (46), which compares Swiss-Prot protein sequences to experimentally determined consensus sites for dozens of protein kinases. Strikingly the amino acid sequence surrounding IRS1-Tyr896 (EPKSPGEYVNFEGS) conformed well to the EY(F/V/I) consensus site for the EGFR tyrosine kinase (47, 48) and ranked in the top 0.113% of potential EGFR substrates overall (Table III). This score is considerably better than the 0.2% cutoff recommended to avoid false positives with Scansite (49). Moreover only one known EGFR autophosphorylation site scored better than IRS1-Tyr896 (EGFR-Tyr1197 at 0.041%); conversely other well characterized EGFR sites had poorer scores than IRS1-Tyr896 (Table III). Several sequences within IRS1 were correctly identified as substrates for the insulin receptor (IR), but IRS1-Tyr896 scored only within the

---

4 S. Gaudet and P. K. Sorger, unpublished observations.
FIG. 4. Compendium of signaling and cell death measurements in TNF-, EGF-, and insulin-treated HT-29 cells. A, heat map of time-dependent signals in cells treated with 10 cytokine combinations. For each treatment, the average signal intensities were normalized to the maximal value obtained for that signal (0, green; 0.5, black; 1, red) and are plotted for the 13 time points. The signals were measured by kinase assay (KA), immunoblot (IB), or antibody microarray (AA). B, heat map of cell death responses for the 10 treatments. The average values from triplicate samples were normalized to the maximal value for that assay and are plotted for 12, 24, and 48 h of treatment (0, green; 0.5, red).
The fractions of cells doubly positive for cleaved caspase-3 and cleaved cytokeratin (\textit{CC3}/\textit{CCK} red, black) were measured by flow cytometry. In vivo IR kinase substrate.

Intriguingly, IRS1-Tyr 896 also scored highly as a potential IR kinase substrate (Table III), therefore less likely to be an Akt substrate, and is important for IR-dependent modification (48) and is therefore less likely to be an \textit{in vivo} IR kinase substrate. Intriguingly, IRS1-Tyr 896 also scored highly as a potential substrate for the fibroblast growth factor receptor and platelet-derived growth factor receptor tyrosine kinases (data not shown). This raises the possibility that IRS1 functions downstream of EGFR and other growth factor receptors. Directed experiments \textit{in vivo} and \textit{in vitro} will be required to test this hypothesis and unravel the role played by IRS1 in cross-talk between EGFR, fibroblast growth factor receptor, platelet-derived growth factor receptor, and insulin receptors.

FKHR Phosphorylation Does Not Correlate with Akt Activation—The closest signals to insulin on the cytokine-signal map were the three measures of Akt activation (Table II) (50). One might expect that Akt substrates, like FKHR (51, 52) would also map closely to insulin. However, pFKHR-Ser\textsuperscript{256} was closer to TNF and EGF (map distances of 0.50 for TNF and 0.58 for EGF versus 0.68 for insulin), and there was no positive correlation with insulin treatment (Table II). When the dynamics of Akt activation and FKHR-Ser\textsuperscript{256} phosphorylation were examined, we found that EGF induced a sharp, transient Akt signal at \( t = 5 \text{ min} \), consistent with previous studies in myocytes (53), and an overlapping peak in pFKHR-Ser\textsuperscript{256} was observed (Fig. 4, A, D, and E). TNF treatment also caused transient (but weaker and delayed) Akt activation, probably as a result of a TNF-inducible TGF-\( \alpha \) autocrine loop,\textsuperscript{2} and a transient pFKHR-Ser\textsuperscript{256} peak (Fig. 4, A, D, and E). In contrast, Akt was strongly induced in a sustained fashion by insulin, but no significant increase in pFKHR-Ser\textsuperscript{256} was observed (Fig. 4, A, D, and E). Thus, within the CSR compendium, Akt activation and FKHR-Ser\textsuperscript{256} phosphorylation are not always correlated (\(-0.01 < R < 0.02\) overall with each of the three measures of Akt activation). The absence of a correlation between signals thought to be biochemically coupled is potentially as valuable as a strong correlation because it implies that mechanistic models are incomplete.

One simple explanation for the discordance between Akt activation and pFKHR-Ser\textsuperscript{256} modification is that, in TNF- and EGF-treated cells, FKHR is phosphorylated by a basophilic kinase other than Akt. Indeed, we found that EGF induced a sharp, transient Akt signal at \( t = 5 \text{ min} \), consistent with previous studies in myocytes (53), and an overlapping peak in pFKHR-Ser\textsuperscript{256} (Fig. 4, A, D, and E). Thus, within the CSR compendium, Akt activation and FKHR-Ser\textsuperscript{256} phosphorylation are not always correlated (\(-0.01 < R < 0.02\) overall with each of the three measures of Akt activation). The absence of a correlation between signals thought to be biochemically coupled is potentially as valuable as a strong correlation because it implies that mechanistic models are incomplete.

In one simple explanation for the discordance between Akt activation and pFKHR-Ser\textsuperscript{256} modification is that, in TNF- and EGF-treated cells, FKHR is phosphorylated by a basophilic kinase other than Akt. Indeed, we found that EGF induced a sharp, transient Akt signal at \( t = 5 \text{ min} \), consistent with previous studies in myocytes (53), and an overlapping peak in pFKHR-Ser\textsuperscript{256} (Fig. 4, A, D, and E). Thus, within the CSR compendium, Akt activation and FKHR-Ser\textsuperscript{256} phosphorylation are not always correlated (\(-0.01 < R < 0.02\) overall with each of the three measures of Akt activation). The absence of a correlation between signals thought to be biochemically coupled is potentially as valuable as a strong correlation because it implies that mechanistic models are incomplete.

One simple explanation for the discordance between Akt activation and pFKHR-Ser\textsuperscript{256} modification is that, in TNF- and EGF-treated cells, FKHR is phosphorylated by a basophilic kinase other than Akt. Indeed, we found that EGF induced a sharp, transient Akt signal at \( t = 5 \text{ min} \), consistent with previous studies in myocytes (53), and an overlapping peak in pFKHR-Ser\textsuperscript{256} (Fig. 4, A, D, and E). Thus, within the CSR compendium, Akt activation and FKHR-Ser\textsuperscript{256} phosphorylation are not always correlated (\(-0.01 < R < 0.02\) overall with each of the three measures of Akt activation). The absence of a correlation between signals thought to be biochemically coupled is potentially as valuable as a strong correlation because it implies that mechanistic models are incomplete.
described in detail elsewhere. Briefly, signal measurements based on measured intracellular signaling profiles, will be designed to generate predictions of apoptotic responses. To evaluate rigorously our design strategy for the CSR compendium (e.g. selection of which signals to measure and time points to sample, assembly of multiple data types, choice of cytokine treatments, and approaches to data processing), we used a data-driven PLSR model. This model, designed to generate predictions of apoptotic responses based on measured intracellular signaling profiles, will be described in detail elsewhere. Briefly, signal measurements were cast as independent (predictor) variables, and apoptotic responses were cast as dependent (predicted) variables. To maximize the information extracted from compendium data, the set of independent variables was expanded to include not only the actual measurements of each of the 19 signals at 13 time points (247 independent variables) but also 22–25 metrics describing the time-dependent dynamics of that signal. These derived metrics included “local descriptors” such as the instantaneous derivative of the signal at each time point, the area under the curve, and the activation and down-regulation rates for each peak in the time course. “Global descriptors” included the total area under the curve over a 24-h period and the global maximum, mean, and steady-state values of the signal. Together the time point measurements and the derived metrics for each signal constituted a set of 660 independent variables. Changes in the independent variables were related to changes in the dependent variables (apoptosis measures) through regression coefficients that the PLSR algorithm calculates for the full range of cytokine treatments in the compendium. Cross-validation showed that the PLSR model could predict the apoptotic responses for any single treatment withheld from the training set with a squared Pearson correlation ($R^2$) of 0.94. Furthermore this model could predict data not in the training set with $R^2 = 0.91$.3

Maximizing Information Content in Models by Unit-Variance Scaling

PLSR models weigh independent and dependent variables with large covariance most heavily (57); this biases the models toward variables with the largest dynamic range. In the compendium, JNK1 activity at $t = 15$ min varied more than 200-fold across all treatments, whereas Akt activity never varied more than 5-fold at any time point. A priori it is not obvious that large changes in JNK1 activity are more important than more modest but sustained changes in Akt activity, yet regression-based models emphasize the former. To give all variables an equal likelihood of contributing to the PLSR model, we applied unit-variance scaling, a common preprocessing technique in regression analysis (57). Each variable (JNK1 activity at $t = 15$ min, for example) was divided by the square root of its variance calculated across all cytokine treatments. This maintained the relative variation in the CSR compendium (e.g. JNK1 activity at $t = 15$ min was still higher for TNF treatment when compared with insulin treatment), but the dataset was scaled so that all variables had the same “spread” of values. Some time courses were altered quite dramatically by unit-variance scaling. For example, high variance in JNK1 activity at $t = 5$, 15, and 30 min and lower (but potentially informative) variance across treatments at later time points led to a scaled TNF-induced time course in which the second wave of activation was emphasized (Fig. 5A). For caspase-8 activation, easily overlooked differences in the extent of cleavage at early times were amplified relative to cleavage at later times (Fig. 5B). As a general rule, we found that scaling enhanced the importance of signals at late time points relative to those at early times (Fig. 5D). However, a few signals with relatively uniform variance were minimally affected by scaling; raw and scaled phospho-Akt time courses were very similar for example (Fig. 5C).

Does unit-variance scaling increase the amount of information extracted from the compendium as measured by the number of independent variables incorporated into the model? To answer this, we constructed PLSR models from either scaled or unscaled data and then evaluated the contribution of each variable (i.e. measurements and derived metrics such as maximum, area under the curve, etc.) to the two models by examining the VIP. The VIP is a normalized expression for the magnitude of the regression coefficients for each variable (see Ref. 58 and “Experimental Procedures”). Variables with a high VIP value (VIP > 1) play an important role in the PLSR model, whereas variables with a low VIP value (VIP <= 1) do not. For a model with unscaled variables (the “unscaled” model), only a small number of independent variables had a high VIP with most of the signals contributing negligibly to predictions (Fig. 5E). By contrast, a “scaled” model built from unit-variance scaled data incorporated far more signaling variables into the response prediction, and the signals were weighed more equally (Fig. 5F). This illustrates that the scaled model drew from data across the CSR compendium, whereas the unscaled model relied entirely on the highest variance signals.

Both scaled and unscaled models predicted cell death responses with good accuracy for single treatments withheld from the training set (squared Pearson correlation of 0.94 with measured values). As a more stringent test of performance, we examined how well the two models predicted death responses under conditions that were different from those used to generate training data. A first set of test data was obtained from cells treated with TNF in the presence of the EGFR-
blocking antibody C225 (59), and the second was obtained from cells treated with TNF in the presence of interleukin-1 receptor (IL-1R) antagonist (IL-1ra) (60). We show elsewhere that, at early times, the signaling and cell death response of cells to TNF is mediated in part by an TGF-α/EGFR autocrine loop and, at later times, by an IL-1α/IL-1R autocrine loop;
C225 and IL-1ra disrupt the TGF-α and IL-1α autocrine loops, respectively. Their importance for the current discussion is that C225 caused fairly dramatic changes in signaling as early as \( t = 15 \) min but only small changes in responses. Conversely IL-1ra elicited more modest changes in signaling (mostly at \( t > 4 \) h) but caused substantial changes in TNF-induced cell death responses (Fig. 6, A–C). When test data from cells treated with TNF + C225 were used as independent variables in the scaled PLSR model, cell death response predictions had a least squares fitness of 0.81 to experimental values (a perfect prediction would yield a fitness of 1; see “Experimental Procedures”). In contrast, the unscaled model performed very poorly, exhibiting no significant correlation between prediction and experiment (Fig. 6D). Although the impact of scaling was less dramatic for TNF + IL-1ra test data, we again found that the scaled model performed significantly better than the unscaled model (0.94 versus 0.81 least squares fit; Fig. 6D). We conclude that important information for predicting cell fate responses is contained in the small, consistent variations of low dynamic range signals and intermediate-to-late time points. Unit-variance scaling extracts this information from the CSR compendium and enables accurate prediction of cell death in treatments that are different from those within the training set data.

### Aspects of the Compendium That Contribute to Prediction Accuracy and Model Robustness

The example above suggested a general approach for evaluating the importance of different aspects of the CSR compendium design using PLSR modeling. Scaled models lacking one or more data types were built and then compared with the
full scaled model to judge their ability to predict apoptotic responses in TNF + C225- and TNF + IL-1ra-treated cells.

The Value of Heterogeneous Signals Measured by Distinct Assays—To test our assumption that it was valuable to sample multiple features of a signaling network (Fig. 1C) we built PLSR models from subsets of the data. As expected, models built on measurements on a single protein performed poorly (Fig. 7A). More surprisingly, models built from multiple signals gathered using a single type of assay (e.g., immunoblots) were also inferior to the model based on the full compendium (Fig. 7B). For example, kinase activity data were as good as the full model at predicting apoptosis following TNF + IL-1ra treatment (0.94 least squares fit) but very poor at predicting the outcome of TNF + C225 treatment (no significant fit; Fig. 7B). Overall, immunoblot data yielded the best single assay model, but this model was still significantly less predictive than the full model (0.84 versus 0.94 for TNF + IL-1ra and 0.65 versus 0.81 for TNF + C225; Fig. 7B). We do not yet know whether the limitations of single-assay data can be overcome simply by measuring many more signals, using the same assay or whether something more fundamental is at work. However, given current technology, cell-signaling measurements are most powerful when different techniques are used in combination.

Efficient Data Collection by Combination Cytokine Treatments—in the design of a typical experiment, parameters are changed one at a time. In contrast, the CSR compendium included many treatments combining prodeath and prosurvival cytokines. To assess the value of combinatorial stimuli, we identified two models trained on few treatments that were still predictive of cell death in the test datasets: 1) a TNF-only model containing subsaturating and saturating TNF-alone treatments as well as the mock treatment and 2) a TNF-EGF-insulin model built from the saturating TNF + EGF and saturating TNF + insulin treatments (Fig. 7C). We then expanded the training data for the two models by including the low TNF + low EGF dataset. Both models predicted cell death in test data as accurately as the full model (fitness of 0.90 for model 1 and 0.94 for model 2; Fig. 7D). However, we reasoned that models trained on smaller datasets might be less robust to measurement noise in the training data. When random noise was added to the low TNF + low EGF signaling data and the models were recalculated, the TNF-only model performed poorly, whereas the TNF-EGF-insulin model and the full model remained reliable (Fig. 7D). We therefore conclude that inclusion of signaling data from cells exposed to multiple cytokines in combination yields models that are less sensitive to experimental noise.

Contributions from Derived Metrics—The full PLSR model was constructed using both signal measurements and derived metrics as independent variables. Was it important to include these derived metrics? Remarkably a model that included derived metrics but excluded actual signal measurements performed as well as the full model on both sets of test data (Fig. 7E, “No time points”). In contrast, a model built from signal measurements alone was unable to predict cell death in the C225 test dataset (Fig. 7E, “Time points”). Moreover activation rates and global descriptors of signaling dynamics (area under the curve, mean, maximal, and steady-state values) were among the most heavily weighted variables in the full model. Why are derived metrics so important for model construction? In the PLSR approach described here, each time point is treated as a separate independent variable, and thus, derived metrics are the only variables that contain information on how measurements were ordered in time. To assess the relative value of each derived metric, we constructed models from single metric types. Several of the models accurately predicted apoptotic responses in the IL-1ra test data, but only one (based on activation rate, or on-slope) yielded predictions with a significant fitness to the measured apoptosis values in the C225 test data (Fig. 7E). Combining several derived metrics therefore seems to yield the most predictive power. Because derived metrics are more predictive than measurements themselves, we conclude that information encoded in time-dependent signaling is a critical aspect of the CSR compendium.

Predictive Information in Early and Late Signals—Can early signals predict late responses? Surprisingly, early time points between 5 and 90 min were predictive of cell death in the combined test datasets 12–48 h later (Fig. 7F), although they were significantly worse at predicting the C225 than IL-1ra test data (data not shown). This suggests that protein activities at early times encode much of the information needed to specify an apoptosis-survival cell fate decision. Late time points (t > 12 h) also had good overall predictive power (Fig. 7F), but this is to be expected because caspase activation at these late time points is mechanistically linked to cell death. Strikingly, a clear drop was observed in the predictive ability of the protein signals that were measured at t = 2, 4, and 8 h (Fig. 7F). It seems unlikely that information is really lost between 2 and 8 h; rather it may be encoded in transcriptional responses absent from our dataset (61). We hypothesize that the inclusion of transcriptional data in the compendium will reestablish the predictive power of models based on these time points. Taken together, compendium-based modeling results suggest that early signals immediately downstream of receptors contain much of the information needed to specify cell fate decisions up to 24 h later.

DISCUSSION

In this article, we describe the construction and validation of a cytokine-signal-response compendium with which to investigate the regulation of cell fate in cells exposed to combinations of the prodeath cytokine TNF and the prosurvival cytokines EGF and insulin. The compendium contains more than 10,000 biochemical measurements on the states and activities of cell-signaling proteins and apoptotic responses in human cells. At the outset of the work described here it was unclear whether heterogeneous data from immunoblots, an-
The PLSR model requires multiple, heterogeneous, time-dependent protein signals to predict apoptosis robustly. Shown are bar graphs of the least-squares fitness of predicted-to-measured cell death responses in test datasets (IL-1ra, green; C225, orange) for models built from data subsets. Light green and orange horizontal bands mark the 90% confidence interval of the prediction fitness for the full compendium for the IL-1ra and C225 test datasets, respectively. Least-squares fitness values less than zero are marked as non-significant (N.S.). Error bars show the 90% confidence interval in the estimate. A and B, fitness values of predicted-to-observed cell death responses for models built from time course data on single protein signals (A) or from signaling data generated by a single assay (antibody arrays (AA), kinase assays (KA), or immunoblots (IB)) (B). Datasets that included cleaved caspase-8 (ClvC8 in A, and immunoblots in B) performed best because this signal is the most directly linked, biochemically, to the outcomes. C, three-dimensional representation of the cytokine combinations used in models with subsets of treatment conditions. Model 1 used data from mock, 5 ng/ml TNF, and 100 ng/ml TNF treatments, whereas model 2 was built from data on 100 ng/ml TNF with either 100 ng/ml EGF or 500 ng/ml insulin. D, comparison of the fitness of predicted-to-measured apoptosis values in the IL-1ra and C225 test datasets combined for models with (gray) or without (white) simulated random normally distributed noise added to the 5 ng/ml TNF + 1 ng/ml EGF signal measurements. The fitness of models with noise represents the average predictive ability of three instances each at five noise levels (10, 20, 30, 40, or 50% of the range of values for each signal). E, fitness values of predicted-to-
tibody microarrays, kinase activity assays, and flow cytometry could be effectively assembled into a single dataset. However, we have now established that a carefully assembled CSR compendium enables the construction of informative statistical models and yields new hypotheses about the regulation of cell fate by cytokines. Elsewhere we describe some specific biological conclusions that can be inferred by mining compendium data; here we concentrate on more general issues associated with the construction, validation, and mining of such datasets.

A key step in the construction of self-consistent datasets is data fusion. Heterogeneous measurements acquired at different times must be combined in such a way that they can be compared quantitatively across the full dataset. Experimental databases are common in genomics largely because sequence data are homogeneous and structured with a clear beginning and end, and data fusion is therefore straightforward. In contrast, cell-signaling data are heterogeneous and unstructured, lack an obvious completion point, and depend heavily on biological context. These features make fusion of signaling data challenging. Our approach to building a data compendium started with the definition of an experimental template that was applied repeatedly to the analysis of different cytokine combinations. The template specified standardized experimental protocols and optimized normalization procedures for determining the time-dependent activities of 19 protein signals and four cell death responses over a 48-h period. The template was applied to data collected from cells treated with 10 cytokine combinations and two combinations of a cytokine and a receptor inhibitor. Overall the raw cytokine-signal-response dataset contained ~10,000 quantitative measurements assembled from 12 independent experiments.

**Accuracy and Self-consistency of the CSR Compendium**—To support reliable hypothesis generation, data fusion must maintain measurement accuracy and self-consistency. A number of procedures were used to ensure the accuracy of the data, and these procedures were validated experimentally. First, the linearity of each assay was verified, and measurements were shown to be directly proportional to signal strength. Experimental protocols, cell treatments, and reagents were standardized, and measurements were adjusted for assay- and sample-specific fluctuations in factors such as reagent activities and lysate concentrations. For each treatment and time point, at least three replicate biological samples were measured, allowing formal statistical testing and assessment of the magnitude of biological variation and measurement noise.

Second, to rule out fluctuations in measurements during the 18-month data collection period, measurements in the primary compendium were compared with those in a validation dataset. The validation dataset replicated a subset of the compendium measurements but with an orthogonal experimental design. Whereas the primary compendium consisted of single treatments analyzed at different time points, validation data were collected at a single time from cells treated with different cytokine combinations. For most signals and responses, validation data matched primary data very well. A few signals could not be validated quantitatively because they lacked sufficient dynamic range to distinguish real variation from noise. Nonetheless we conclude that most signals and responses were consistent throughout the data collection process; validation experiments must be designed to maximize the number of signals showing significant variation across treatments.

As a third self-consistency check we asked whether different measures of the same biochemical process yielded similar data. In DNA sequencing this type of quality control is straightforward and involves checking the noncoding and coding sequences for complementarity. However, verifying the congruence of protein signals is more complex in part because a mechanistic model relating the different biochemical properties measured by each assay is necessary. For example, we assayed the activation of Akt both by kinase activity assay and by tracking its phosphorylation state at Ser\(^{473}\), an activating site, using immunoblots and antibody arrays. Perfect correlations between these assays would require that phospho-Ser\(^{473}\) be essential for kinase activity and that all forms of phospho-Akt-Ser\(^{473}\) be fully active. Although kinase activity and phospho-Ser\(^{473}\) assays of Akt were generally well correlated, subtle discrepancies were observed. These discrepancies are consistent with the known mechanisms of Akt regulation (bisphosphorylated Akt is more active than Akt singly phosphorylated at Ser\(^{473}\) or Thr\(^{308}\) (35, 36)), but they also highlighted the challenges involved in validating the merger of data from heterogeneous assays into a single coherent representation.

**Context-dependent Signaling Dynamics**—The CSR compendium allowed us to compare signaling dynamics within a range of cytokine contexts. We used two approaches to mine compendium data for interesting signaling patterns: heat maps (see Supplementary Material) and cytokine-signal maps.\(^2\) One surprising finding was that the treatment of cells with EGF, but not with insulin, was strongly associated with
phosphorylation of insulin receptor substrate 1 on Tyr^896. The sequence flanking IRS1-Tyr^896 is a remarkably good match to an EGF receptor consensus phosphoacceptor site, leading us to hypothesize that EGF may be an important pIRS1-Tyr^896 inducer and that EGFR may modify IRS1 directly. This hypothesis is now being tested experimentally. A second unexpected discovery was a lack of correlation between Akt activation and Ser^256 phosphorylation of the Forkhead transcription factor, an Akt substrate (51, 52). The lack of correlation leads us to hypothesize that FKHR-Ser^256 may be targeted by a kinase other than Akt in EGF- and TNF-treated cells. Consistent with this idea, the basophilic kinase SGK has recently been shown to modify the FKHR homologue FOXO3 on two phosphoacceptor sites previously thought to be Akt substrates (56). These two findings illustrate the value of comparing signal strengths quantitatively across a set of different biological conditions. Unexpected statistical correlations observed among signals are suggestive of new mechanistic links, whereas the absence of an expected correlation suggests that existing links may require reinvestigation.

Experimental Requirements for Signaling Compendia—The construction and validation of computational models of cell signaling was our primary goal in generating the CSR compendium. As a first step, we built a regression (PLSR) model that relates protein signals to cell death responses. PLSR models built from data scaled to unit variance extracted the most information from the compendium and performed much better than models built from unscaled variables. The best PLSR data-driven model was surprisingly powerful: given signaling information from test data not included in the training set, the model predicted apoptotic responses to within the range of error for experimental measurements. PLSR models were used to analyze the design and assembly of CSR data with the goal of developing optimized strategies for the construction of future compendia.

Would a simpler experimental scheme with fewer signals, treatments, or time points have yielded a CSR compendium with a similar ability to support PLSR modeling? Although the answer appears to be “yes” because some data scored low in the VIP (Fig. 5F), statistical modeling clearly benefited from heterogeneous measurements, information on signal dynamics, and the analysis of cells treated with combinations of cytokines. Models built from single signals were ineffective at predicting apoptosis, consistent with the idea that cells weigh the contributions from multiple signals when committing to survival or death (62). More interestingly, models built from any single type of assay were significantly less predictive than models based on the full compendium. Among the single-assay models, immunoblot data performed best probably because immunoblots were used to assay two distinct biochemical processes: protein phosphorylation and protein cleavage. Most proteomic efforts to date (63) have centered on the use of a single technology such as protein arrays (64) or mass spectrometry (10). However, our results suggest that it is highly advantageous to combine several different assay methods to capture the diversity of protein signals that mediate cell fate decisions.

If the analysis of single proteins is insufficient to predict cell fate, then how much coverage of a signaling network is required? The success of our CSR compendium shows that it is not necessary to strive for complete coverage: apoptosis was predicted within 94% accuracy given signaling information from only 11 of the roughly 75 protein nodes shown in Fig. 1C (15% coverage). However, we suspect that a sparse sampling of a network will only yield predictive models when nodes are repeatedly sampled under a variety of conditions. In a typical experiment one usually decides to vary one parameter at a time, but we have found that data from cells treated with a combination of cytokines were more informative than data from single cytokine treatments particularly when measurements were noisy. Combinations of prodeath and prosurvival cytokines evoke more complex signaling patterns than individual stimuli, and it seems likely that this provides a broader context within which to assess the role of each signal in cell death responses.

Which signals should be measured in the protein network? Although the answer clearly depends on the biological system, we found that upstream signals (e.g. receptors and adaptors) were easiest to relate to cytokine stimuli in statistical maps. For instance, IRS1-Tyr^896 phosphorylation was prominent as an EGF-dependent signal on the cytokine-signal map. With the PLSR model, however, downstream signals like ERK, JNK1, and MK2 were most predictive of apoptotic responses. We also found that measuring connected nodes in the network, initially thought to represent somewhat redundant assays, provided useful insight into the propagation of signals through particular pathways. Akt-FKHR is one example of a densely sampled region of the network; another is the dual measurements of ERK activity and phospho-MEK level (Fig. 8B). The correlation between these two signals was reasonable (Pearson coefficient of 0.68 overall), but more in-depth analysis revealed that ERK and phospho-MEK were best correlated at the earliest time point after cytokine addition, implying that their activation was coincident (Fig. 8A, 5 min). At later time points, MEK was inactivated more rapidly than ERK, and the correlation was lost probably as a result of differential dephosphorylation (Fig. 8, A and B). The ERK-MEK disparity was exacerbated under conditions of saturating TNF treatment (Fig. 8A, circles). Thus, TNF may activate the Ser-Thr phosphatases that act on MEK or alternatively inactivate the dual specificity and Tyr phosphatases that act on ERK (Fig. 8B). Directed phosphomapping experiments and phosphatase assays should be able to test this prediction. More generally, the variability in MEK-ERK induction highlights how proteins that interact directly and function in the same pathway can exhibit time-dependent changes in relative activities.

Propagation of Signaling Information with Time—In the CSR
Compendium of Signals and Responses to TNF, EGF, and Insulin

compendium, protein signals were encoded by two types of variables: discrete measurements collected at 13 points over a 24-h period and derived metrics describing the local and global dynamics of signaling. We found that these derived metrics were essential for constructing predictive PLSR models and that they could not be replaced by the measurements from which they were derived. PLSR treats each time point independently, and derived metrics were therefore the primary means by which time dependence was encoded. We presume that this explains the importance of the metrics for the PLSR model. Unexpectedly, however, when the CSR compendium was modeled using a multilinear regression technique (66) that constrains the sequence of time points by explicitly ordering signaling network “snapshots” in time, we were unable to construct predictive models.5 One explanation for the success of derived metrics in PLSR and failure of multilinear regression is that critical signaling information is imbedded in the cumulative activities and rates of change of key signals, calculated as derived metrics, rather than in the ordering of signaling events. For example, the cumulative activity of caspase-8 is likely to be more important than the time at which activity is first detected. Indeed molecular integrators and differentiators have been previously identified in both prokaryotic and eukaryotic signaling networks (21, 23, 67).

PLSR models based on single time point snapshots between 5 and 90 min or 12 and 24 h were able to predict apoptotic responses with reasonable accuracy although not as well as the full model. A clear loss of information was observed with protein signals at time points greater than 2 h. In HT-29 cells, TNF-induced apoptosis is not apparent until 8 h after stimulus addition, implying that cells must still be processing apoptosis-survival cues at those time points. One explanation for the loss of predictive power from 2 to 8 h is that informa-

5 K. A. Janes and P. K. Sorger, unpublished observations.

Fig. 8. Differential regulation of MEK and ERK by phosphatases is context-dependent. A, scatter plots of pairwise comparisons of ERK activity and phospho-MEK (pMEK) levels at 5, 15, and 30 min and at time points greater than 30 min as indicated. Values are means ± S.E. of biological triplicate samples normalized to the maximal value in the compendium for that signal. Pearson correlations are noted in the upper left corner when the dynamic range is greater than 2.5. B, schematic diagram showing the two measurement types, ERK kinase activity assays (KA) and phospho-MEK levels by immunoblot (IB), the differential regulation of ERK and MEK by phosphatases. MEK is targeted only by Ser-Thr phosphatases (S/T-Pase) that may be activated by TNF treatment. ERK is additionally targeted by dual specificity phosphatases (DS-Pase) and Tyr phosphatases (Y-Pase) that may be inhibited by TNF treatment. a.u., arbitrary units.

Molecular & Cellular Proteomics 4.10 1587
tion processing involves changes in gene transcription, which were not measured in our experiments. Microarray experiments are currently underway to test this idea. Nonetheless it is clear that protein signals alone, when sampled over a time course, are sufficient for predicting HT-29 apoptotic responses induced by cytokines. It will be interesting to determine whether PLSR models based on the current CSR compendium will also be effective in predicting apoptosis induced by other stimuli or in different cell lines.

**Conclusion**—The data in this article show that cytokine-signal-response compendia should be constructed using measurements that are well distributed across a signaling network, although sparse coverage of the network is acceptable. Experimental validation of measurements is best carried out under conditions in which all signals have a sufficiently large dynamic range for correlation coefficients to be meaningful. In some cases this will involve two validation experiments, for example one at an early time point and one later. Measurement of signaling dynamics is clearly critical for CSR compendia, and uneven spacing of time points makes it possible to sample adequately both transient and sustained signals. Within the context of the regression-based models used here, it is important to rely not only on the data themselves but also to compute metrics that describe signal dynamics. Moreover scaling data to unit variance is necessary to optimize the extraction of information from compendium data.

In summary, we have described methods for compiling validated, self-consistent data compendia describing time-varying signals induced by cytokines and have demonstrated the value of compendium data in the study of mammalian signal transduction. We believe that similar CSR compendia coupled with both regression-based and mechanistic models will be valuable in the systematic analysis of other complex biological networks.

**Acknowledgments**—We thank members of the Sorger and Lauffenburger laboratories for helpful discussions.

*This work was supported by National Institutes of Health Grant P50-GM68762 and by the Whitaker Foundation (to K. A. J.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.*

**The on-line version of this article (available at http://www.mcponline.org) contains supplemental material.**

**To whom correspondence should be addressed: Dept. of Biology, Rm. 68-371, Massachusetts Inst. of Technology, 77 Massachusetts Ave., Cambridge, MA 02139. Tel.: 617-252-1648; E-mail: psorger@mit.edu.**

**REFERENCES**

1. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwen, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Graham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, C., Minna, J., Mungall, A., Plumley, R., Ross, M., Shawcross, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Narra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W., R. Chaisson, S. L., Wendel, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscombe, E., Peddik, P., Richardson, P., Wening, S., Slezak, T., Doggett, N., Cheng, J., Oki, K., Livak, S. J., Quackenbush, J., Frazer, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bock, J. D., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissbach, J., Heilig, R., Saurin, W., Antignaveanu, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D. R., Doucette-Stamm, L., Rubenfield, M. L., Dubinot, K., Lee, B., Plancon, E., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R. W., Federespiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Ols, M. V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R. O., Bae, A., Chen, F., Fan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dadhia, N., Blobek, H., Hornischer, K., Nordsieck, G., Agarwala, R., Aravind, L., Bailey, J. A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burke, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eicher, E. E., Furey, T. S., Galagan, J., Gilbert, J. G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jiang, W., Johnson, L. S., Jones, T. A., Kasif, S., Kasparyzak, A., Kennedy, S., Kent, W. J., Kitts, P., Koons, N., Kozich, J. D., Lancet, D., Lander, E. L., McLaughlin, A., Miikkelin, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schulier, G., Schultz, J., Slater, G., Smit, A. F., Stupka, E., Suzastokowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yen, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patirosian, A., Morgan, M. J., de Jong, P., Catanese, J. J., Osoegawa, K., Shizuya, H., Choi, S., and Chen, Y. H. (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921

2. Botstein, D., and Risch, N. (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat. Genet.* 33, (suppl.) 228–237

3. Irish, J. M., Hovland, R., Krutzik, P. O., Perez, O. D., Brusserud, O., Gjertsen, B. T., and Nolan, G. P. (2004) Single cell profiling of potentiated phospho-protein networks in cancer cells. *Cell* 118, 217–228

4. Blagoev, B., Ong, S. E., Kratchmarova, I., and Mann, M. (2004) Temporal analysis of phosphotyrosine-dependent signaling networks by quantitative proteomics. *Nat. Biotechnol.* 22, 1139–1145

5. Gilman, A. G., Simon, M. I., Bourne, H. R., Harris, B. A., Long, R., Ross, E. M., Stull, J. T., Taussig, R., Arkin, A. P., Cobb, M. H., Oyster, J. G., Devreotes, P. N., Ferrel, J. E., Fruman, D., Gold, M., Weiss, A., Berend, M. J., Cantley, L. C., Catterall, W. A., Coughlin, S. R., Olson, E. N., Smith, D. J., Buggia, J. S., Bos, J. D., Bredesen, D., Lehrka, L., Pawson, A. J., Stemberg, P. W., Varmus, H., Subramaniam, S., Sinkovits, R. S., Li, J., Mock, D., Ning, Y., Saunders, B., Strenwic, P. C., Hilgeman, M., Scheurermann, R. H., DeCamp, D., Hsieh, R., Lin, K. M., Ni, Y., Seanman, W. E., Simpson, P. C., O’Connell, T. D., Roach, T., Choi, S., Eversole-Cire, P., Fraser, I., Mumbty, M. C., Zhao, Y., Brekken, D., Shu, H., Meyer, T., Chandy, G., Heo, W. D., Liou, J. O., O’Rourke, N., Verghese, M., Mumber, S. M., Han, H., Broxin, H. A., Forrester, J. W., Milne, S. B., Casey, P. J., Hadden, T. K., Joye, D., Grey, M. L., Michnick, S., Schmidt, M. A., Toner, M., Tsien, R. Y., Natarajan, M., Ranganathan, R., and Sambrano, G. R. (2002) Overview of the Alliance for Cellular Signaling. *Nature* 420, 703–706

6. Papin, J. A., Hunter, T., Palisson, B. O., and Subramaniam, S. (2005) Reconstruction of cellular signaling networks and analysis of their prop-
16. Karin, M., and Ben-Neriah, Y. (2000) Phosphorylation meets ubiquitination.
15. Kyriakis, J. M., and Avruch, J. (2001) Mammalian mitogen-activated protein
13. Micheau, O., and Tschopp, J. (2003) Induction of TNF receptor I-mediated
12. Wajant, H., Pfizenmaier, K., and Scheurich, P. (2003) Tumor necrosis factor
10. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics.
31. Scheid, M. P., Marignani, P. A., and Woodgett, J. R. (2002) Multiple phos-
29. Datta, S. R., Brunet, A., and Greenberg, M. E. (1999) Cellular survival: a play
27. Remacle-Bonnet, M. M., Garrouste, F. L., Heller, S., Andre, F., Marvaldi,
23. Alon, U., Surette, M. G., Barkai, N., and Leibler, S. (1999) Robustness in
19. Schlessinger, J. (2004) Common and distinct elements in cellular signaling
17. Yarden, Y., and Sliwkowski, M. X. (2001) Untangling the ErbB signalling
15. Kyriakis, J. M., and Avruch, J. (2001) Mammalian mitogen-activated protein
13. Micheau, O., and Tschopp, J. (2003) Induction of TNF receptor I-mediated
12. Wajant, H., Pfizenmaier, K., and Scheurich, P. (2003) Tumor necrosis factor
10. Aebersold, R., and Mann, M. (2003) Mass spectrometry-based proteomics.
8. Sambrano, G. R., Chandy, G., Choi, S., Decamp, D., Hsueh, R., Lin, K. M.,
7. Janes, K. A., Albeck, J. G., Peng, L. X., Serper, P. K., Lauffenburger, D. A.,
5. Brach, A., Burgers, J. T. M., and Bourne, H. (2002) Unravelling the signal-transduction network in B
4. Datta, S. R., Brunet, A., and Greenberg, M. E. (1999) Cellular survival: a play
3. Scheid, M. P., Marignani, P. A., and Woodgett, J. R. (2002) Multiple phos-
2. Sambrano, G. R., Chandy, G., Choi, S., Decamp, D., Hsueh, R., Lin, K. M.,
1. Schlessinger, J. (2004) Common and distinct elements in cellular signaling

Molecular & Cellular Proteomics 4.10
1589
Holder, J. C., and Cohen, P. (1997) Insulin activates protein kinase B, inhibits glycogen synthase kinase-3 and activates glycogen synthase by rapamycin-insensitive pathways in skeletal muscle and adipose tissue. FEBS Lett. 406, 211–215

54. Nakae, J., Park, B. C., and Accili, D. (1999) Insulin stimulates phosphorylation of the forkhead transcription factor FKHR on serine 253 through a Wortmannin-sensitive pathway. J. Biol. Chem. 274, 15882–15885

55. Nakae, J., Kitamura, T., Ogawa, W., Kasuga, M., and Accili, D. (2001) Insulin regulation of gene expression through the forkhead transcription factor Foxo1 (Fkhrl) requires kinases distinct from Akt. Biochemistry 40, 11768–11776

56. Brunet, A., Park, J., Tran, H., Hu, L. S., Hemmings, B. A., and Greenberg, M. E. (2001) Protein kinase SGK mediates survival signals by phosphorylating the forkhead transcription factor FKHRL1 (FOXO3a). Mol. Cell. Biol. 21, 952–965

57. Geladi, P., and Kowalski, B. R. (1986) Partial least-squares regression—a tutorial. Anal. Chim. Acta 185, 1–17

58. Wold, S. (1994) Exponentially weighted moving principal components—analysis and projections to latent structures. Chemom. Intell. Lab. Syst. 23, 149–161

59. Sato, J. D., Kawamoto, T., Le, A. D., Mendelsohn, J., Polikoff, J., and Sato, G. H. (1983) Biological effects in vitro of monoclonal antibodies to human epidermal growth factor receptors. Mol. Biol. Med. 1, 511–529

60. Dinarello, C. A. (2000) The role of the interleukin-1 receptor antagonist in blocking inflammation mediated by interleukin-1. N. Engl. J. Med. 343, 732–734

61. Manos, E. J., and Jones, D. A. (2001) Assessment of tumor necrosis factor receptor and Fas signaling pathways by transcriptional profiling. Cancer Res. 61, 433–438

62. Lowe, S. W., Cepero, E., and Evan, G. (2004) Intrinsic tumour suppression. Nature 432, 307–315

63. Service, R. F. (2003) Proteomics. Public projects gear up to chart the protein landscape. Science 302, 1316–1318

64. Melton, L. (2004) Protein arrays: proteomics in multiplex. Nature 429, 101–107

65. Tamura, S., Hanada, M., Ohnishi, M., Katsura, K., Sasaki, M., and Kobayashi, T. (2002) Regulation of stress-activated protein kinase signaling pathways by protein phosphatases. Eur. J. Biochem. 269, 1060–1066

66. Bro, R. (1996) Multiway calibration. Multilinear PLS. J. Chemom. 10, 47–81

67. Yi, T. M., Huang, Y., Simon, M. I., and Doyle, J. (2000) Robust perfect adaptation in bacterial chemotaxis through integral feedback control. Proc. Natl. Acad. Sci. U. S. A. 97, 4649–4653