INTRODUCTION

A goal of systems biology is to identify regulatory properties of biological networks by integrating experiments and kinetic models around a defined biological “parts list” (1). For signal transduction, such approaches have clarified the molecular pathways and regulatory events that can explain signaling thresholds (2–4), adaptation (5–8), and oscillations (9–15), as well as signal-induced cell fate transitions (16–18). A complementary use of systems methods is for large-scale profiling and analysis of cellular signals in pursuit of integrative understanding or new molecular mechanisms (19–22). These applications favor data-driven models based on statistics or bioinformatics, which propose new links within the network that can be tested experimentally (23–28). Testing of a data-driven hypothesis about mechanism requires reductionist experimental approaches, even when the hypothesis itself emerges from network-level data (29, 30). Consequently, systems-level studies can yield findings that are immediately relevant to signal transduction biology (31).

Receptor-mediated signaling pathways are densely connected and exhibit nonadditive behaviors when two ligands activating different receptors are applied simultaneously (32–34). Downstream of receptor activation and signal transduction, complexity increases even further when considering the consequences of signaling on gene regulation (35–37). Fortunately, new signaling synergy or antagonism does not typically emerge with more than two inputs (38–41), suggesting that stimulus pairs are sufficient to assess potential crosstalk. For systems-level discovery, this type of in-depth cellular profiling must be complemented by data-driven modeling and extensive mechanistic follow-up.

Here, we examined crosstalk between a proinflammatory stimulus, tumor necrosis factor (TNF), and one of two growth factors, insulin or epidermal growth factor (EGF), with a focus on signaling network activation and downstream regulation of the transcriptome. To examine the interface between signal transduction and transcription networks at the systems level, we built upon a preexisting compendium of signaling events collected in HT-29 colon adenocarcinoma cells, which were initially exposed to interferon-γ (IFNγ) and then stimulated with TNF, EGF, and insulin in various combinations (42, 43). Because TNF, EGF, and insulin trigger widespread and time-dependent changes in transcript abundance (44–47), we collected dynamic mRNA profiles after stimulation with single or paired stimuli and integrated these data with the previously measured intracellular signaling events (42, 43).

Because each signal and transcript was measured at every time point and stimulus condition, we could organize the signaling and transcriptomic data cubes as three-way tensors to build a model constrained by the structure of the data. Model analysis together with bioinformatics predicted a link between AKT–glycogen synthase kinase 3 (GSK3) signaling and the endodermal transcription factor GATA6. We found that GSK3 phosphorylation of Ser37 on the long form of GATA6 (GATA6L) accelerated its degradation in cells. The increased turnover reduced transcriptional repression by GATA6L of genes induced by TNF. Collectively, our results showed that GATA6L integrated growth factor–induced signaling activity and inflammatory transcriptional regulation.
RESULTS

Stimulus combinations elicit complex changes in signaling and transcript abundance

For the signaling pathway measurements, we used data from our previous studies (42, 43), which could be combined with the transcriptomic data generated here under the same experimental conditions. In the previous studies, HT-29 cells were exposed to IFNy, which primes HT-29 cells to apoptose upon stimulation with TNF (48), and subsequently stimulated with saturating or subsaturating doses of TNF, EGF, or insulin alone or in combination. Lysates were profiled at 13 time points over 24 hours for 19 intracellular signaling events measured by kinase assay, immunoblot, or antibody array (Fig. 1, A and B) (49–51). These data provided quantitative, systematically collected information on phosphorylation-mediated regulatory events, changes in protein abundance, and cleavage-dependent protein activation.

To determine how the signaling events altered gene expression, we complemented the signaling compendium with a matched set of transcriptomic profiles (Fig. 1A). IFNy-pretreated HT-29 cells were exposed to the same combinations of TNF-EGF-insulin and analyzed at a subset of the time points from the previous signaling studies (Fig. 1B). We collected transcriptomic profiles by microarray at intermediate-to-late times after cytokine stimulation, thereby avoiding the bursts of immediate-early transcripts that are already well characterized (46, 47, 52). We selected the time points

![Diagram of signaling and transcript abundance](diagram.png)

**Fig. 1.** A compendium of ligand-induced signals and transcriptional responses. (A) Overview of the experimental design. HT-29 cells were pre-treated with IFNy, stimulated with various combinations and concentrations of TNF, EGF, and insulin, and profiled for the indicated signaling receptors, adaptors, and effectors by kinase assay (KA), immunoblot (IB), or antibody array (AA) and for the associated transcriptomic signatures by microarray. The goal is to determine whether global ligand-induced mRNA regulatory states (Y) can be predicted from the upstream signaling network activation (X). (B) Hierarchical clustering of the signaling compendium for saturating (High) and subsaturating (Low) concentrations of TNF, EGF, and insulin (42, 43). Data are means of n = 3 independent biological replicates. (C) Hierarchical clustering of the dynamic transcriptomic responses resulting from the ligand combinations in (B). Data are means of n = 2 independent biological replicates.
Each data point thus contains information about the various “modes” of its acquisition, for example, which stimulus was added (mode 1), the time after stimulus (mode 2), and the signal or transcript measured (mode 3). Additional modes are possible if multiple pharmacologic perturbations (56, 57) or cell types (58) are profiled systematically along with the modes listed above.

For systematic measurements, the acquisition modes create a data structure that is very powerful mathematically because it conveys how different data points are related to one another. This structure vanishes when, for example, a data cube is sliced along one of its modes and “unfolded” end-to-end as a series of matrices (Fig. 2A). When matrix-based algorithms are applied to unfolded data, each unfolded measurement variable is treated independently, and modes 2 and greater are lost. Using Fig. 2A as an example, AKT measurements at 2 and 4 hours after stimulation (same signal, two time points) are not handled any differently than AKT and epidermal growth factor receptor (EGFR) measurements at 2 hours after stimulation (two signals, same time point). The result of unfolding is a model that is less interpretable because of too many fitted regression coefficients (59).

The alternative to unfolding is to retain data sets as cubes (three modes) or hypercubes (four or more modes) in the form of data “tensors,” which are the higher-dimensional generalization of vectors (one mode) and matrices (two modes). For example, the TNF-EGF-insulin signaling compendium naturally organizes as a third-order tensor defined by stimulus, time point, and measured signaling event (Fig. 2B). The transcriptomic profiles likewise arrange as a third-order tensor according to stimulus, time point, and transcript or cluster of transcripts (Fig. 2C). Tensors reduce the parameterization of a data-driven model because free regression coefficients remain fixed across the other acquisition modes of each tensor (Fig. 2B) (59, 60). In this instance, the stimulus–time point–signaling tensor (the “regressor” tensor) is linked to the stimulus–time point–transcript tensor (the “regressand”) tensor by the regression coefficients.

Biological data tensors have been used successfully for supervised purposes, such as singular value decomposition (27), to analyze transcriptional kinetics during DNA replication origin firing (28) and to identify consistent copy number changes across different array-based comparative genomic hybridization platforms (61). Here, we sought a supervised method that could connect the signaling tensor to the transcriptomic tensor and predict gene expression patterns from signaling network dynamics. This application is ideal for the tensor generalization of partial least squares regression (PLSR), a matrix implementation that has been used widely to model signaling networks (20, 25, 58, 62–76).

Tensor PLSR (equivalently, “multilinear PLS” (59)) is an established method that creates a data-driven model by jointly factorizing an independent “predictor” tensor (X; here, the signaling tensor) and a dependent “predicted” tensor (Y; the transcriptomic tensor). X and Y are factorized as an element-by-element product of vectors, where the number of vector elements multiplied is equal to the number of dimensions in the data tensor (text S1). Thus, if X is a third-order tensor, then \( X_{(1,1,1)} \) [the tensor element in X occupying the first position in mode 1 (stimulus), the first position in mode 2 (time point), and the first position in mode 3 (signal)] is factorized as: \( X_{(1,1,1)} = \kappa(1)w_1(1)p_1(1) \) (Fig. 2D, purple). In the factorization, \( \kappa(1) \) is the first element of a “scores” vector (\( t \)) that relates to the stimulus conditions that are shared with the Y tensor: \( w_1(1) \) and \( p_1(1) \) are the first elements of two “weight” vectors (\( w_1 \) and \( w_q \)) that relate to modes 2 and 3 of the tensor (here, time and signal). A similar calculation is performed for Y by factorizing it into its own scores (u) and weight (\( q_1 \) and \( q_m \)) vectors. X and Y are linked by an “inner relationship” between their respective scores vectors: \( u = bt \), where b is a linear regression coefficient determined by the model. The inner relationship implies that how a stimulus projects on \( t \) [through the signaling (\( w_1 \))] that occurs over time \( (w_1) \) is directly proportional to its projection on \( u \) and thus how that stimulus changes gene expression (\( q_m \) with time \((q_t)\).
The factorization of the two tensors is posed as a numerical optimization that seeks to capture as much of the inner relationship between $X$ and $Y$ as possible (details on the covariance maximization algorithm are described in text S1). The best first set of scores and weight vectors defines the first “latent variable” of the tensor PLSR model. Residual information (covariation) in $X$ and $Y$ not captured by the first latent variable is then subjected to a second factorization, which is optimized to capture as much covariance in the residual as possible (Fig. 2D). By repeating the algorithm, latent variables are iteratively calculated until there are no predictive inner relationships remaining between the $X$ and $Y$ data tensors (20, 62).

Predictions with a tensor PLSR model use $w_1$ and $w_2$ from each latent variable to project an $X$-like observation onto $t$ (Fig. 2E). Then, the $u = bt$ inner relationship is used to predict $u$, which is backprojected with $q_1$ and $q_m$ to yield a predicted set of values in the form of $Y$ (time-dependent gene expression). The project-predict-backproject sequence is important for making independent predictions with new data and for cross-validation of the model to identify the optimum number of latent variables (20, 31, 62, 77).

**Tensor PLSR modeling identifies predictive links between signaling and transcriptional dynamics**

We first constructed a tensor PLSR model of three latent variables that predicted the 14,541 probe set fluorescence intensities of the transcriptomic data set. Although cross-validated predictions of the model were 99% accurate (fig. S1A), the model was strongly biased toward the differences in fluorescence intensities across probe sets. Consequently, changes in probe set intensities across treatment conditions were overlooked, and the resulting components of the model were uninterpretable (fig. S1, B to E). To focus on recurrent stimulus-dependent changes in transcript abundance shared by multiple genes, we condensed the transcriptomic data set by using the unbiased CLuster Identification via Connectivity Kernels (CLICK) algorithm (78). Among the transcripts profiled, CLICK identified nine separable clusters composed of dozens to hundreds of genes (file S2), the mean trajectories of which were organized as the $Y$ data tensor (Fig. 2C). Using the entire signaling compendium as $X$, we constructed a tensor PLSR model of four latent variables that predicted gene clusters to within 72% (Fig. 3A and fig. S2A). Although the model did not predict certain cytokine-induced changes for some gene clusters (Fig. 3A, see EGF and insulin stimuli of cluster #3), we considered the overall accuracy of predictions remarkable considering that the model involved ~10-fold fewer parameters than previous PLSR models of TNF-induced apoptosis (42, 62, 64).

Our principal motivation for building the tensor PLSR model was to use the model to reveal undiscovered mechanisms of how signaling alters gene expression. To identify which latent variables captured both signaling and gene cluster dynamics, we analyzed the time weights ($w_1$ and $q_l$) and inner regression coefficients for $X$ and $Y$ (Fig. 3B). The leading two latent variables (LV#1 and LV#2) harbored signaling time weights ($w_j$ and $w_j$) that were nearly constant from 0 to 24 hours, indicating that time-dependent changes in signaling did not determine the projection of $X$ along LV#1 or LV#2. Accordingly, time weights for the gene clusters ($q_l$ and $q_l$) were time-variant and derived from the stimulus-independent transcriptional changes of clusters #1, #2, and #7 (Fig. 3, A and B), presumably resulting from IFNγ pretreatment. Because latent variables are calculated iteratively (Fig. 2D), LV#1 and LV#2 eliminated the TNF-, EGF-, and insulin-independent transcriptional changes, revealing paired signaling and gene cluster dynamics in the third and fourth latent variables (LV#3 and LV#4). LV#3 harbored time weights of late-phase signaling ($w_k$) and sustained transcriptional activation ($q_4$). Conversely, LV#4 was weighted with early-phase signaling ($w_k$) and late-phase transcriptional regulation ($q_4$). The inner regression coefficient for LV#4 was negative (Fig. 3B, orange), implying a link between early-phase signaling and downstream transcriptional repression.

Focusing on LV#3 and LV#4, we evaluated the relationship between the treatment scores ($t_1$ and $t_2$, Fig. 3C). Relative to mock treatment, saturating TNF stimulation projected almost entirely along LV#3, suggesting that LV#3 represented a TNF “axis.” In contrast to TNF, we found that EGF and insulin projected in opposite directions along LV#4, indicating that this latent variable distinguished between the two growth factor stimuli. Combinatorial stimulations exhibited intermediate scores that approximately averaged the scores of the individual stimuli. For example, TNF + EGF projected positively along LV#3 (like TNF) and negatively along LV#4 (like EGF). The interpolated response observed here for gene regulation contrasts with previous work on apoptosis in which EGF and insulin each nonlinearly antagonized TNF-induced cell death (62).

To connect specific signals and gene clusters with the prevalent cytokine-induced dynamics, we evaluated the signaling and gene cluster weights along LV#3 (TNF; late-phase signaling axis: $w_3$ and $q_m$) and LV#4 (EGF; early-phase signaling axis: $w_4$ and $q_m$) (Fig. 3, D and E). Multiple signals, such as Ser1068-phosphorylated insulin receptor substrate-1 [pIRS1 (Ser1068)], c-Jun N-terminal kinase (JNK), and mitogen-activated protein kinase (MAPK)-activated protein kinase-2 (MK2) activity, were negligibly weighted (Fig. 3D), implying that these early-phase TNF-induced signals (Fig. 1B) were statistically uninformative for predicting $Y$ (the transcriptional response). Gene cluster #1 was also unweighted along LV#3 and LV#4, because its dynamics were almost entirely captured by LV#1 and LV#2 (Fig. 3, A, B, and E, and fig. S2B). To filter the weight vectors further, we randomly shuffled the signaling, gene cluster, and time information within each cytokine stimulation (mode 1 slice) of $X$ and $Y$ (79). With hundreds of shuffled tensor PLSR models, we estimated a null projection for the weight vectors of LV#3 and LV#4 (Fig. 3, D and E, gray line). We considered signals and gene clusters outside 1 SD (σ, gray dashed lines) of the null projection as weighted strongly enough to warrant further analysis (Fig. 3, D and E).

Among the strongest signaling weights, we found clear agreement with known mechanisms of signal transduction (Fig. 3D). For example, cleavage of apoptotic caspases (negative weighting for procaspase-8 and procaspase-3 and positive weighting of cleaved caspase-8) was strongly aligned along LV#3, which is consistent with late-phase caspase activation triggered by TNF (48, 80). Along LV#4, insulin stimulation coincided with positive weights for three complementary measures of AKT activation, a recognized effector pathway (81). Likewise, multiple measures of EGFR phosphorylation were weighted in a direction that corresponded to EGF stimulation. Also strongly associated with EGF signaling was phosphorylated insulin receptor substrate-1 [pIRS1 (Tyr1165)], consistent with reports that this site may be directly phosphorylated by active EGFR (42, 82). Not all signaling events were associated with individual stimuli. For instance, the weights associated with inhibitor of nuclear factor-κB kinase (IKK) activation mapped not only to TNF but also to insulin stimulation, possibly because AKT signaling can activate IKK in certain contexts (83, 84). Similarly, phospho-EGFR (Tyr1068) projected with early-phase EGF and also late-phase TNF signaling. The latter is probably due to autocrine signaling by transforming growth factor-α, an EGFR ligand that is released after TNF stimulation (43, 85). Together, the weights of LV#3 and LV#4 provided a condensed map of the signaling compendium that was optimized for predicting the observed transcriptomic profiles.

The inner regression coefficient ($b_3$) connecting $X$ and $Y$ along LV#3 was a positive value, indicating gene activation, whereas the inner regression coefficient $b_4$ was negative, indicating that signaling along LV#4 resulted in gene repression (Fig. 3, B and E). Contrary to that of the signaling compendium, the projection of gene clusters along LV#3 and LV#4 was
Fig. 3. A tensor PLSR model linking ligand-induced signaling and changes in transcript abundance. (A) Time-unfolded measurements of transcriptional clusters (blue) compared to cross-validated predictions of the tensor PLSR model (brown). Standardized z-scores of measured transcriptional clusters are means ± SD of n = 897 (#1), 841 (#2), 119 (#3), 106 (#4), 66 (#5), 49 (#6), 42 (#7), 33 (#8), and 26 (#9) probe sets (file S2). High (H) indicates saturating concentration of ligand, 0 indicates absence of ligand, and low (L) indicates subsaturating concentration of ligand. (B) Latent variable time weights for the signaling and transcriptomic tensors. LV#4 has a negative inner relationship (orange), indicating that LV#4 signaling is anticorrelated with LV#4 transcription. (C to E) Projections of the indicated stimulus conditions (C), signals (D), and transcriptional clusters (E) onto LV#3 and LV#4. For (D) and (E), the null projections of shuffled data tensors are means (solid gray) ± SD (dashed gray) of n = 500 randomizations (79). In (D), the type of assay used to measure the signaling protein is indicated in parentheses (see Fig. 1A for details). ClvC8, cleaved caspase 8; ProC3, procaspase 3; ProC8, procaspase 8; lowercase p prefix represents phosphorylated protein; lowercase t prefix represents total protein; lowercase pt prefix represents the ratio of phosphorylated protein to total protein.

Amidst thousands of time-dependent transcriptional changes, few clusters were weighted toward specific stimuli. Clusters #3 and #6 were primarily weighted along LV#3, indicating an association with TNF stimulation. Accordingly, promoter analysis (86) of the transcripts in these two clusters revealed a strong overrepresentation of binding sites for nuclear factor κB (NF-κB) (fig. S2C). Cluster #7 mapped along LV#4 because of the mild suppression of transcripts observed with saturating insulin alone (Fig. 3, A and E). Only cluster #9 projected strongly along both latent variables, indicating that the transcripts in this cluster were induced by TNF and repressed by insulin (fig. S3). TNF antagonism of insulin function has been well documented in adipocytes (87, 88), but there are few reports of insulin antagonizing TNF (89). Given this predicted TNF and insulin “crosstalk cluster,” we used the tensor PLSR model to investigate its mechanism of regulation by the upstream signaling network.

With respect to its latent variable projections, cluster #9 was cartographically similar to IKK (Fig. 3, D and E). If these shared projections were indicative of mechanism, however, it would imply that early-phase IKK activity (downstream of TNF and insulin signaling) represses transcription of the crosstalk cluster, whereas late-phase IKK (downstream of TNF signaling) promotes it. Repress-then-activate kinetics are opposite of the prevailing view of IKK signaling (90). Accordingly, we found that TNF-induced responses of 85% of transcripts in cluster #9 were not significantly affected when a phosphorylation- and degradation-resistant mutant
of Ixβx was ectopically expressed in HT-29 cells (log-transformed Welch’s t test, false discovery rate = 15%; fig. S4). We therefore considered alternatives that were consistent with the tensor PLSR model.

One possible explanation was that the crosstalk cluster integrated two distinct signaling inputs. An activating input could arise from a TNF-specific signal that was either not measured or not projected strongly on LV#3 and LV#4. In parallel, the cluster could be transcriptionally inhibited by an insulin-specific signal, such as AKT (Fig. 3D) or a downstream effector pathway of AKT (91).

**Promoter and signaling bioinformatics suggest a link between GSK3 and the crosstalk cluster through GATA6.**

First, we determined the reliability and generality of TNF-insulin crosstalk among transcripts in cluster #9. We repeated the stimulation experiments with an independently obtained vial of HT-29 cells and assayed individual transcripts by quantitative reverse transcription polymerase chain reaction (qRT-PCR). This analysis confirmed the expression of more than 90% of the 22 cluster #9 transcripts (SPRR1B and PPARD from fig. S3 were false positives), and we detected an antagonistic interaction between TNF and insulin from 2 to 8 hours after stimulation (fig. S5). At individual time points for specific genes, we observed instances of antagonism represented by significant interaction P values (Pint < 0.05, two-way ANOVA; Fig. 4, A and B), indicating that TNF and insulin have nonadditive effects on the expression of those genes. We also observed additive differences in gene expression for other transcripts when Pint was not significantly different (Fig. 4, C and D). The qRT-PCR data thus confirmed the microarray results and the tensor PLSR model, showing an early-phase suppression of TNF-induced cluster #9 genes by insulin (Fig. 3, B to E). Furthermore, some of the TNF-insulin crosstalks cannot be predicted by adding the effect of insulin to the TNF response.

To identify candidate mediators of TNF-insulin crosstalk, we analyzed the expression-verified transcripts of cluster #9 with three orthogonal promoter analysis algorithms (86, 92, 93). Only two transcription factors were suggested as candidate regulators by all three algorithms: T cell

---

**Fig. 4.** Multipronged bioinformatics of TNF-insulin crosstalk suggests post-translational regulation from GSK3 to GATA6. (A to D) qRT-PCR validation of selected cluster #9 transcripts upon pretreatment of HT-29 cells with IFNy and stimulation with TNF with or without insulin for 2 hours (A and B) or 6 hours (C and D). Data are geometric means ± log-transformed SEM of n = 4 or 8 biological replicates. Full cluster #9 data are shown in fig. S5. (E) Promoter bioinformatics (86, 92, 93) suggests GATA and TCF4 as candidate regulators of TNF-insulin crosstalk. (F) Relative copy number estimates (96, 97) for the six GATA isoforms in HT-29 cells. Data are medians ± range of n = 3 biological replicates. n.d., not detected. (G to I) Transcriptional dynamics of GATA isoforms in response to TNF and insulin. Data are geometric means ± log-transformed SEM of n = 4 or 8 biological replicates. (J) ScanSite (98) identification of candidate GSK3 phosphorylation sites (red). Each site’s percentile rank is averaged across the indicated sequences. (K) Mass spectrometry identifies 11 phosphorylation sites on GATA6. Previously unreported sites (New) are shown below the primary sequence, those consistent with reports in the literature (104) (Reported) are shown above, and reported sites not detected in this study are gray. Start methionines (arrows) for the long and short forms are indicated along with the conserved GATA core and zinc finger (ZnF) domains.
factor 4 (TCF4) and GATA (Fig. 4E). HT-29 cells harbor a truncating mutation in APC, encoding a protein that inhibits the β-catenin pathway, and this truncation renders β-catenin and its transcriptional partner TCF4 constitutively active (94). Moreover, no changes in β-catenin localization were observed upon TNF simulation with or without insulin (fig. S6). Therefore, we focused on GATA, a family of six transcription factors that are important for development and differentiation (95). Using qRT-PCR (96, 97), we quantified the relative copy numbers of the GATA family and found that GATA6 was the most abundant isoform (Fig. 4F and fig. S7). Copies of GATA6 transcript also remained high during the early phase of TNF-only and TNF + insulin stimulation, whereas GATA2 and GATA3 were reduced two- to fourfold (Fig. 4, G to I). Notably, bioinformatic analysis of TNF-only and TNF + insulin stimulation, whereas (98) of the full-length GATA6 protein sequence uncovered a cluster of highly conserved serine residues that were candidate phosphorylation sites for GSK3 (Fig. 4J). Because GSK3 is a recognized substrate of AKT (81), these results suggested that GATA6 could be an insulin-dependent regulator of the crosstalk cluster.

Full-length GATA6 (GATA6L) is distinguished by a long N-terminal extension that is conserved across vertebrates but missing in other GATA family members (fig. S8). In addition, leaky ribosome scanning (99) onto an in-frame methionine (Met1-45) gives rise to a short form of GATA6 (GATA6S) that is comparable in size (~45 kD) to the other GATA isoforms. GATA6S lacks the candidate GSK3 phosphorylation sites that reside at the N terminus of GATA6L (Fig. 4I), raising the possibility of selective regulation of GATA6L by insulin.

Compared to GATA6S, GATA6L has been understudied because of early misannotations of nonhuman genomes and a lack of suitable reagents. Common plasmid repositories, including Addgene (100), the human ORFeome (101), and the Mammalian Gene Collection (102), have only the short form or lack the gene entirely. The unusually slow electrophoretic mobility of GATA6 has created additional confusion because commercial antibody vendors mistakenly label GATA6S as “GATA6L,” implying that the detected protein is the full-length form. As a result, the GATA6 literature is incredibly ambiguous, with many papers inadvertently focusing on GATA6S.

To determine whether the predicted GSK3 phosphorylation sites of GATA6S could be phosphorylated in HT-29 cells, we cloned the full-length gene with N-terminal FLAG and C-terminal AU1 tags into a doxycycline-inducible lentivector (see Materials and Methods) (103). Stable HT-29 transductants were induced with doxycycline for 24 hours before lysis, and FLAG immunoprecipitates were subjected to phosphorylation analysis by mass spectrometry. We achieved 92% coverage of the GATA6L sequence and identified 11 phosphorylation sites, including 7 that had not been previously reported (Fig. 4K) (104). Among sites in the N-terminal extension specific to GATA6L, two (Ser33 and Ser37) were consistent with the bioinformatic predictions of GSK3 phosphorylation (Fig. 4J). Two N-terminal sites (Thr34 and Ser37) were also corroborated by a proteomics study of proline-directed phosphorylation in 293T cells (105). We concluded that Ser33, Thr34, and Ser37 were the leading candidates for GATA6L phosphorylation-mediated regulation by GSK3.
(3xSA) into 293T cells, cells in which GATA6L phosphorylation has been detected previously (105). Compared to the wild-type allele, we noted a pronounced electrophoretic downshift in the major FLAG-immunoreactive band of the Ser37 and 3xSA mutants (Fig. 6A). The mobility shift was larger than that expected for a single phosphorylation site, suggesting that Ser37 phosphorylation was required for other phosphorylation events within GATA6L. Iterative phosphorylation-dependent phosphorylation is characteristic of many GSK3 substrates, such as glycogen synthase (GS) (111).

Careful inspection of endogenous GATA6L immunoreactivity in HT-29 extracts revealed a slower migrating species that was similar to the electrophoretic mobility of the densitometric traces (see Materials and Methods). Acute GATA6L through Phos-tag electrophoresis (C) and quantification (D) of the upper and lower bands of the Ser37 and 3xSA mutants (Fig. 6A). The mobility shift was larger than that expected for a single phosphorylation site, suggesting that Ser37 phosphorylation was required for other phosphorylation events within GATA6L. Iterative phosphorylation-dependent phosphorylation is characteristic of many GSK3 substrates, such as glycogen synthase (GS) (111).

Our next goal was to evaluate the specific impact of Ser37 phosphorylation on GATA6L in HT-29 cells. One challenge was that the endogenous abundance of GATA6L was high compared to GATA6L (Fig. 5A), which could confound interpretations of ectopically expressed proteins. Therefore, we inducibly knocked down endogenous GATA6L in HT-29 cells with short hairpin RNA and added back epitope-tagged versions of wild-type or S37A GATA6L so that the abundance was comparable to total endogenous GATA6 (Fig. 6, E and F; see Materials and Methods). Doxycycline-induced addback in HT-29 cells recapitulated the electrophoretic mobilities of GATA6L that were observed in transfected 293T cells. The data suggested that the GATA6L modifications were not artifacts of overexpression, enabling use of the addback cells to examine the consequences of Ser37 phosphorylation.

Because GSK3 phosphorylation often accelerates substrate turnover (114), we combined the Phos-tag analysis with the HT-29 addback lines to estimate half-lives of wild-type and S37A GATA6L. We combined inhibition of protein synthesis with TNF stimulation to enrich for the lower migrating form of wild-type GATA6L in the addback cells. Under these conditions, we found that the half-life of more phosphorylated GATA6L was more than twice that of the wild-type GATA6L form that was less phosphorylated (Fig. 6, G and H). Surprisingly, the half-life of the S37A mutant was comparable to that of the more phosphorylated form of GATA6L, suggesting that phosphorylation of Ser37 without subsequent additional phosphorylation renders GATA6L unstable. Ser37 resides in the middle of a proline-glutamate-serine-threonine (PEST) degradation motif of GATA6L, and this motif scores more strongly as a PEST motif than those in other well-known unstable proteins (Fig. 4J and table S1) (115, 116). Ser37 phosphorylation might activate or expose the PEST sequence for rapid proteolytic degradation of GATA6L, whereas additional phosphorylation at other sites could inhibit substrate recognition (115).

To monitor Ser37 phosphorylation specifically, we raised and affinity-purified a phospho-specific antibody against a monophosphorylated peptide fragment of the PEST sequence in GATA6L (see Materials and Methods; fig. S11). If Ser37 modification were a prerequisite for subsequent phosphorylation, then the antibody would capture this initial phosphorylation event of GATA6L, with the caveat that phosphorylation on Ser37 and Thr34 would ultimately disrupt the antibody epitope. To evaluate the
phospho-Ser\(^{37}\) antibody, we reassessed the immunoreactivity of total GATA6. The predicted molecular weights of GATA6s and GATA6L are 45.4 and 60 kD, respectively. However, extensive posttranslational modifications (Fig. 4K) cause most GATA6s and GATA6L to run at an apparent molecular weight of \(~54\) kD and \(~69\) to 75 kD depending on electrophoresis conditions (Fig. 7A). Multiply phosphorylated GATA6s (~54 kD) can be misinterpreted as unmodified GATA6L (60 kD). Upon long exposure with a total GATA6 antibody, we revealed an additional immunoreactive band at \(~60\) kD that was eliminated by GATA6 knockdown and reconstituted with addback of wild-type GATA6l and the S37A mutant (Fig. 7A). In contrast to the 75-kD form (Fig. 6E), the ~60-kD form of wild-type GATA6s was significantly less abundant than the S37A mutant (Fig. 7B), consistent with decreased stability. We interpreted the ~60-kD band as the unmodified form of GATA6L.

Endogenous Ser\(^{37}\) phosphorylation of the 75- and 60-kD GATA6s forms was not detectably altered in response to TNF treatment for 1 hour or CT99021 treatment for 6 hours (Fig. 7, C and D). However, the endogenous 60-kD phospho-GATA6s signal was barely above the detection limit and thus highly variable (coefficient of variation, ~40%), yielding only ~50% statistical power for detecting a 1.5-fold change. To evaluate phosphorylation of GATA6s at Ser\(^{37}\), we reconstituted the HT-29 cells with 20 \(\mu\)M CT99021 pretreatment for 1 hour and reconstituted with addback of wild-type GATA6s and the S37A mutant (Fig. 7A). knockdown and reconstituted with addback of wild-type GATA6s and the S37A

Fig. 7. Phosphorylation and destabilization of endogenous GATA6s at 60 kD. (A) Knockdown (shGATA6) and FLAG-tagged addback of GATA6s at ~60 kD (red). Samples were immunoblotted for total (modified and unmodified) GATA6s (upper), FLAG (lower), and the indicated loading controls. (B) Destabilization of the 60-kD form of WT GATA6s compared to the S37A-mutant addback cells. (C and D) Endogenous phospho-GATA6s (Ser\(^{37}\)) immunoreactivity is not detectably affected by stimulation with TNF for 1 hour, inhibition with 20 \(\mu\)M CT99021 for 6 hours, or both. (E and F) Phosphorylation and destabilization of the 60-kD form of GATA6s upon serum starvation. Specificity was confirmed by preincubation of cells with 20 \(\mu\)M CT99021 for 1 hour before serum starvation. (G and H) Phospho-GATA6s (Ser\(^{37}\)) immunoprecipitation and total GATA6 immunoblot of HT-29 cells pretreated with IFN\(\gamma\) and stimulated with TNF \(\pm\) insulin for 1 hour. The gamma of the immunoprecipitation image is set to 1.5 to minimize background from the immunoprecipitating antibody heavy chain. Input (0.5%) of each immunoprecipitate was immunoblotted for total and total GATA6 immunoblot of HT-29 cells pretreated with IFN\(\gamma\) and stimulated with TNF \(\pm\) insulin for 1 hour. The gamma of the immunoprecipitation image is set to 1.5 to minimize background from the immunoprecipitating antibody heavy chain. Input (0.5%) of each immunoprecipitate was immunoblotted for total and total GATA6 immunoblot of HT-29 cells pretreated with IFN\(\gamma\) and stimulated with TNF \(\pm\) insulin for 1 hour.
correlated with changes in 60-kD GATA6L phosphorylation at Ser37 (Fig. 7E and F). These experiments provide further evidence that Ser37 is a site phosphorylated by GSK3 and that phosphorylation at this site promotes turnover of GATA6L in the absence of phosphorylation at additional sites.

With greater confidence in the phosphoryso-GATA6L (Ser37) antibody, we revisited the original biological context of IFNγ-pretreated HT-29 cells stimulated with TNF and insulin. To enable detection, we immunoprecipitated cell extracts with phospho-GATA6L (Ser37) antisera and immunoblotted for total GATA6 (Fig. 7G). An extended electrophoresis was required to separate the 60-kD form from the heavy chain of the immunoprecipitating antibody, causing a smear of immunoreactivity rather than a discrete band (see Materials and Methods). In response to TNF alone, we repeatedly observed a drop in 75-kD, but not 60-kD, GATA6L phosphorylation (Fig. 7H), corroborating the dephosphorylation previously noted by Phos-tag electrophoresis (Fig. 6B). Moreover, the reduction in 75-kD phospho-GATA6L (Ser37) was blocked by insulin costimulation, indicating a specific point of crosstalk between TNF and insulin. Insulin, by contrast, independently elevated the abundance of 60-kD GATA6L phosphorylation, suggesting that insulin delays the turnover of this form beyond its effect on Ser37 phosphorylation. We conclude that the phosphorylation of endogenous GATA6L at Ser37 is consistent with the antagonism and linear superposition in abundance of cluster #9 transcripts observed upon TNF + insulin stimulation (Fig. 4, A to D, and fig. S5).

**GSK3-dependent phosphorylation of Ser37 alleviates GATA6L repression of transcripts in the crosstalk cluster**

We investigated the role of GATA6L phosphorylation in the regulation of transcripts exhibiting TNF-insulin crosstalk. We inducibly overexpressed wild-type GATA6L in HT-29 cells (Fig. 8A) before stimulation with TNF for 2 hours and transcriptomic profiling by microarray (see Materials and Methods). GATA6 can act as either a transcriptional activator or repressor ([117]), but we found that, in the GATA6L-overexpressing HT-29 cells, the effects of GATA6L were predominantly repressive. Without stimulation, GATA6L overexpression induced 51 transcripts and repressed 136 transcripts at a 5% false discovery rate ($P < 10^{-10}$, binomial test) (file S4). TNF stimulation increased the number of genes affected by GATA6L overexpression, but the bias toward repression persisted (317 induced transcripts versus 438 repressed transcripts; $P < 10^{-5}$, binomial test) (file S4). Notably, the same repressive bias was observed for transcripts in the crosstalk cluster (Fig. 8B), and those with the strongest GATA6L-associated repression were among the clearest examples of TNF-insulin crosstalk (Fig. 4, A to C). Using chromatin immunoprecipitation (ChIP), we confirmed binding of GATA6L to consensus sites within the promoters of many TNF-insulin crosstalk genes (fig. S13), suggesting that repression is direct.

If GATA6L mediates the insulin-stimulated repression of the crosstalk cluster and is stabilized by inhibition of the GSK3 pathway, then phosphorylation of Ser37 would provide a mechanism for TNF-insulin crosstalk. Furthermore, the S37A mutant of GATA6L should mimic the effect of insulin on TNF-stimulated gene expression for transcripts in the crosstalk cluster and dampen the crosstalk observed when insulin is added to S37A mutant cells. We tested this prediction with the wild-type and S37A addback lines (Fig. 6, E and F), inducing GATA6L and then stimulating with TNF, insulin, or both. By qRT-PCR, we identified multiple instances in which S37A addback reduced transcript abundance similar to that observed in wild-type addback cells stimulated with insulin (Fig. 8C, green). We also found many examples in which TNF + insulin–induced transcript abundance was higher in S37A addback cells compared to wild-type addback cells (Fig. 8C, purple) and more similar to TNF-treated cells, suggesting reduced crosstalk. Analysis of the entire cluster #9 data set revealed significant interactions between GATA6 and TNF or insulin (interaction $P < 10^{-3}$, five-way ANOVA), indicating that the Ser37 genotype alters the transcriptional response to both stimuli. Together, our data support a model whereby TNF promotes and insulin inhibits the formation and degradation of GATA6L monophosphorylated on Ser37 (Fig. 8D). This phosphoregulation is ultimately reflected by the abundance of transcripts in the crosstalk cluster.

**Ser37 phosphorylation of different GATA6L forms is observed in diverse cell types**

The model of GATA6L phosphorylation–mediated regulation (Fig. 8D) may be specific to HT-29 cells or could occur in other cell types. We immunoblotted various cell lines with the phospho-Ser37 antibody in comparison to affinity-purified antisera binding the nonphosphorylated peptide surrounding Ser37 (see Materials and Methods) and to other commercial GATA6 antibodies (100, 118). In HCT-8 and DLD-1 colorectal cancer lines, with doxycycline (1g/ml) for 24 hours. (A) Doxycycline-inducible overexpression of WT GATA6L in HT-29 cells. Cells were treated with doxycycline (1 μg/ml) for 24 hours. (B) Ratio of TNF-induced transcript abundance for crosstalk cluster genes in the presence or absence of GATA6L overexpression. Data are mean ratios of expression for transcripts in the crosstalk cluster and dampen the crosstalk observed when insulin is added to S37A mutant cells. We tested this prediction with the wild-type and S37A addback lines (Fig. 6, E and F), inducing GATA6L and then stimulating with TNF, insulin, or both. By qRT-PCR, we identified multiple instances in which S37A addback reduced transcript abundance similar to that observed in wild-type addback cells stimulated with insulin (Fig. 8C, green). We also found many examples in which TNF + insulin–induced transcript abundance was higher in S37A addback cells compared to wild-type addback cells (Fig. 8C, purple) and more similar to TNF-treated cells, suggesting reduced crosstalk. Analysis of the entire cluster #9 data set revealed significant interactions between GATA6 and TNF or insulin (interaction $P < 10^{-3}$, five-way ANOVA), indicating that the Ser37 genotype alters the transcriptional response to both stimuli. Together, our data support a model whereby TNF promotes and insulin inhibits the formation and degradation of GATA6L monophosphorylated on Ser37 (Fig. 8D). This phosphoregulation is ultimately reflected by the abundance of transcripts in the crosstalk cluster.

![Fig. 8. S37A mutation of GATA6L mimics and competes with the repression of transcript abundance in the TNF-insulin crosstalk cluster.](https://www.sciencesignal.org/doi/fig)
lines, AC16 cardiomyocytes (110), and MCF10A-5E breast epithelial cells (119), we observed the 60- and 75-kD forms of GATA6L, which were phosphorylated to variable extents according to phospho-Ser37 immunoreactivity (Fig. 9, A and B). Multiple GATA6 antibodies also recognized another species at ~100 kD (Fig. 9, A to D), suggesting that an even more phosphorylated form of GATA6L may remain to be characterized. The aggregate number of reported phosphorylation sites on GATA6L now exceeds 20 (Fig. 4K).

**DISCUSSION**

Our study here introduces and implements tensor PLSR as an approach for structured biological data sets. Considering that signaling dynamics often occur in discrete temporal phases (49, 80, 120, 121), tensor PLSR provides an attractive means to deconstruct time course data in a systematic manner. Although the mathematics has been established for decades (59), data types that can exploit the tensor framework are relatively new to cell signaling. For tensor generation, a multiplex technique that simply measures many genes or proteins is insufficient. The method must also be cost-effective, reproducible, and scalable for repeated use across multiple treatments, time points, and perturbations. The newest technologies rarely meet these criteria, prompting our use of long-established methods at a scale not typically considered.

We applied tensor PLSR with the goal of discovering molecular mechanisms that connect signaling to transcriptional regulation. Ideally, the mechanisms would involve proteins not originally included in the systematic data set. Systems-level studies rarely uncover these “hidden nodes” and validate them experimentally like we achieved here for GATA6L (25, 122). Testing model- or bioinformatics-derived predictions requires a skill set entirely different from the one needed to perform the analysis. Our findings argue for the benefits of dual training, where computationalists work at the bench and experimentalists use quantitative models, gaining an appreciation for the thematic similarities in each approach. For example, just as modeling assumptions should be subjected to falsification (31), we sought to challenge the prevailing biological assumptions about GATA6 and its different forms.

The deceptive electrophoretic mobilities of GATA6S and GATA6L have important implications for biological function. Although GATA6S is generally less abundant than GATA6L in most cell types, there is evidence that GATA6L is the more potent transcriptional regulator (99). GATA6 promotes the expression of the stem cell marker LGR5 in colorectal cancer (118, 123). Neither paper clarified whether the regulation occurs through GATA6L, GATA6S, or both. However, insulin-like growth factor inhibits GSK3 and promotes expansion of Lgr5+ stem cells in mice (81, 124). Our results indicate that one mechanism for this expansion is the stabilization of GATA6L.

The phosphorylation of Ser37 adds a GATA6L-specific mode of regulation to reports of posttranslational modifications that would presumably target both long and short forms (107–109). Although we were unable to reproduce the mechanism exactly (109), modification of Ser436 by AKT2 should coincide with the loss of Ser37 phosphorylation to stabilize GATA6L synergistically in contexts where both pathways operate. GATA6L phosphorylation–mediated regulation could prove important in endothelial cells, a TNF- and insulin-responsive cell type in which GSK3 and GATA6 interact as a complex (125). Our mass spectrometry study also uncovered other GATA6L-specific proline-directed modification sites (Thr62 and Ser137) that could function with the Ser266 site phosphorylated by extracellular signal–regulated kinase (ERK) (107, 108). Such complex layers of regulation should be expected of a transcription factor that is central to embryonic development and cell specification (95).

By coupling systematic experiments with statistical modeling approaches, such as tensor PLSR, one can identify relationships that would otherwise go unnoticed. Although it remains atypical to collect transcriptomic data as tensors, we expect widespread systematization of transcriptomics as expression profiling costs drop. A model is just the first step, however, because the most surprising

---

Fig. 9. Diversity of GATA6L forms across different cell lineages. (A to D) Arrows indicate the GATA6 forms confirmed earlier by knockdown or observed with multiple antibodies. Red asterisks indicate nonspecific bands. The gap represents a deleted lane in each blot. All samples were analyzed together on the same blots. Data are representative of n ≥ 3 independent experiments.
data-derived connections will require the identification of previously unrecognized mechanisms to explain them. These, in turn, require hypothesis-driven experiments with the best molecular genetic and pharmacologic perturbations available. For understanding how gene expression is controlled by complex stimuli, the integration of molecular biology and systems biology has yet to be fully exploited.

MATERIALS AND METHODS

Cell culture
HT-29, 293T, DLD-1, and HCT-8 cells were obtained from the American Type Culture Collection (ATCC) and cultured according to their recommendations. The 5E clone of MCF10A cells was cultured as described previously (119). AC16 cells (110) were purchased from M. Davidson (Columbia University) and cultured in Dulbecco's modified Eagle's medium/F-12 medium (Life Technologies) with 12.5% tetracycline-free fetal bovine serum (Clontech) and penicillin-streptomycin (Gibco).

Cell stimulation
HT-29 cells were plated at 50,000 cells/cm² for 24 hours, sensitized with human IFNγ (200 U/ml; Roche), and then treated with TNF (100 ng/ml; PeproTech), insulin (500 ng/ml; Sigma), or both for the indicated times. HT-29 cells engineered to express GATA6L inducibly were treated with doxycycline (1 μg/ml) for 24 hours (overexpression) or 48 hours (addback) before cytokine stimulation.

Plasmids
Wild-type GATA6L was amplified by PCR from HT-29 RNA that had been reverse-transcribed with a GATA6-specific primer (CAAAAGCAGAGACTGTTGAGA). A N-terminal 3×FLAG tag and a C-terminal 3×AU1 tag were added by PCR before cloning into the Bam HI and Sal I sites of pBabe puro (126) or the Mfe I and Spe I sites of pEN_TTmiRc2 (103). The pEN_TT donor vector containing GATA6L was then recombined with the pSLIK-Neo destination vector (103) by using LR Clonase (Invitrogen). The shGATA6 sequence (CCCAGACCACTTGCTATGAAA; #TRCN0000005390 from the RNAi Consortium) was cloned into tet-pLKO-puro (127) as described previously (97). S33A, T34A, S37A, and 3×SA point mutants were prepared by site-directed mutagenesis (QuikChange II XL, Agilent). RNA interference–resistant mutants of wild-type and S37A GATA6L were prepared by introducing four silent mutations into the sequence targeted by shGATA6, which were replaced with rare mammalian codons that would minimize ectopic expression. The phosphorylation–degradation-resistant 1kB3a super-repressor plasmid has been previously described (128). All DNA constructs were verified by sequencing.

Production and purification of phospho-GATA6L (Ser27) antibody
The peptide sequence Ac-CREPSTPPpsSPIS-amide was conjugated to keyhole limpet hemocyanin and used to immunize rabbits according to the manufacturer’s recommendations (Covance). Serum samples were tested by immunoblotting with positive and negative controls for phospho-GATA6L (Ser27). Serum pooled from the production and terminal bleeds was negatively selected on a CRESTP TPPSIP peptide–conjugated N-hydroxysuccinimide (NHS)–Sepharose column. The bound immunoglobulin G (IgG) was eluted as the non–phospho-GATA6L custom antibody while the flow through was exposed to a second CRESTP TPPSIP peptide–conjugated NHS–Sepharose column. The bound IgG was eluted as the phospho-GATA6L (Ser27) antibody and used for detection by immunoblotting.

Lentiviral packaging and transduction
Lentiviruses were prepared in human embryonic kidney (HEK) 293T cells (ATCC) by calcium phosphate transfection of the lentivector together with psPAX2 and pMD.2G (Addgene). Lentiviral transduction of HT-29 cells was performed as described previously (96). Transduced cells were selected in growth medium containing puromycin (2 μg/ml) or G418 (600 μg/ml) until control plates had cleared. For addback experiments, viral titers were reduced to ensure single-virion transductants that matched the endogenous protein abundance as closely as possible.

Microarray profiling
HT-29 cells were plated at 50,000 cells/cm² for 24 hours and sensitized with IFNγ (200 U/ml; Roche) for 24 hours before stimulation with TNF (0, 5, or 100 ng/ml), EGF (0, 1, or 100 ng/ml), and insulin (0, 5, or 500 ng/ml) for 4, 8, or 16 hours. RNA isolation was performed with the RNeasy Mini Kit (Qiagen), and integrity of purified RNA was confirmed on a Bioanalyzer (Agilent). Preparation of labeled complementary RNA, hybridization to GeneChip Human Genome U133A Arrays (Affymetrix), microarray scanning, and microarray processing were performed as previously described (129).

For inducible GATA6L overexpression, stably transduced HT-29 cells were plated at 50,000 cells/cm² for 24 hours, induced with doxycycline (1 μg/ml), and sensitized with IFNγ (200 U/ml; Roche) for 24 hours before stimulation with TNF (100 ng/ml) for 2 hours. RNA was purified as described above and amplified with the Illumina TotalPrep-96 RNA Amplification Kit (Life Technologies) before hybridization to a HumanHT-12 v4 Expression BeadChip.

Hierarchical and CLICK clustering
One-way hierarchical clustering of the signaling and transcriptomic compendia was performed in MATLAB with the clustergram function using Euclidean distance and Ward’s linkage after row standardization. CLICK clustering was performed as described (78) with the default homogeneity parameter.

Tensor PLSR
Tensor PLSR was performed in MATLAB with version 2.02 of the NPLS Toolbox (130). The signaling compendium was structured by stimulus condition (mode 1), time point (mode 2), and measured signal (mode 3). The transcriptomic profiles were structured by stimulus condition (mode 1), time point (mode 2), and CLICK gene cluster (mode 3). Both data tensors were mean-centered along mode 1 and variance-scaled along modes 2 and 3 before calculation of latent variables (131). The scores and time weights of the LVs4 signaling tensor were both multiplied by –1 to improve model interpretability. Randomized models were constructed in MATLAB with the shufflematrix function applied within each stimulus condition before pre-processing and calculation of latent variables. The code for the tensor PLSR models is included in file S3.

Bioinformatic analyses of crosstalk cluster
The 20 transcripts from cluster #9 confirmed present by qRT-PCR were submitted to three promoter analysis algorithms. First, the proximal promoter of each transcript [defined as 2000 base pairs (bp) upstream and 500 bp downstream of the transcription start site] was collected from the National Center for Biotechnology Information (NCBI) and used as an input set for MEME, which uses expectation maximization to define recurrent motifs in a set of sequences (132). The top five enriched motifs were searched against a database of 843 binding specificities (133) using TOMTOM (134) to identify known transcription factor recognition sequences. A GATA motif was also enriched when using 2000 bp of upstream sequence alone or 1500 bp of upstream sequence and 500 bp of downstream sequence.
Expression-verified cluster #9 transcripts were additionally analyzed with DiRE (86), which uses interspecies sequence conservation to define motifs that are searched against the TRANSFAC 10.2 database of roughly 400 transcription factor binding motifs. In DiRE, the occurrence metric reflects the overall frequency of a conserved binding motif in the input data set, whereas the importance metric reflects the specificity of the binding motif to the input data set compared to a background data set of 5000 randomly selected genes. The top 20 motifs based on occurrence were used as the DiRE predictions.

Last, X2K (92) was used to identify bioinformatic connections between cluster #9 transcripts and signaling pathways. X2K integrates the ChEA database (135) of transcription factor binding sites detected by ChIP, the JASPAR and TRANSFAC position weight matrices, as well as various protein-protein interaction and kinase-substrate databases to connect kinase signaling events to gene expression patterns. The top 20 transcription factors linked to signaling and cluster #9 transcripts in a 2011 analysis were used as the X2K predictions.

**Quantitative RT-PCR**

RNA from cultured cells was isolated with the RNeasy Plus Mini Kit (Qiagen) according to the manufacturer’s protocol. First-strand complementary DNA synthesis and qRT-PCR were performed as described (63). Parental HT-29 samples were normalized to the geometric mean of GAPDH, HINT1, PPIA, and PRDX6. GATA6L addback samples were normalized to the geometric mean of GAPDH, HINT1, PPIA, PRDX6, B2M, and GUSB. Primer sequences are available in table S2.

**Mass spectrometry**

HT-29 cells stably expressing doxycycline-inducible 3×FLAG-GATA6L were induced with doxycycline (1 µg/ml) for 24 hours and lysed in NP-40 lysis buffer plus protease and phosphatase inhibitors (51). Protein extract (60 mg) in 6-ml volume was first cleared with 50 µl of mouse IgG–agarose beads (Sigma) for 1 hour at 4°C on a nutator. The cleared lysates were subjected to immunopurification using 80 µl of anti-FLAG M2 affinity gel (Sigma) for 3 to 4 hours followed by two washes with NP-40 lysis buffer, one wash with 500 mM NaCl, and one wash with tris-buffered saline. Immunoprecipitates were eluted with 3×FLAG peptide (500 ng/ml; Sigma) in 100 µl for 30 min at 4°C on a nutator. The eluate was concentrated using an Amicon ultra centrifugal filter (Millipore), and samples were prepared in dithiothreitol-containing Laemmli sample buffer and separated by SDS-polyacrylamide gel electrophoresis on an 8% polyacrylamide gel followed by Coomassie brilliant blue staining. The stained gels were incubated with 1 mM EDTA in modified Towbin sample buffer (136). Membrane blocking, antibody probing, and near-infrared fluorescence detection were then performed as described (51).

For Gaussian mixture modeling of GATA6L forms, raw 16-bit pixel intensities were integrated horizontally across each lane and then plotted along the vertical dimension. Using the fit function in MATLAB, the vertical trace was fit by nonlinear least-squares to the following function:

\[
 f(x) = b + w_1 \exp\left( - \frac{x - \mu_1}{2\sigma^2} \right) + w_2 \exp\left( - \frac{x - \mu_2}{2\sigma^2} \right)
\]

where \(f(x)\) is height of the vertical trace as a function of the vertical position \(x\), \(b\) is a fixed background, \(w_1\) and \(w_2\) are the relative weights of the two bands, \(\mu_1\) and \(\mu_2\) are the mean vertical positions of the two bands, and \(\sigma^2\) is a shared variance for the two bands. Normalized versions of \(w_1\) and \(w_2\) were taken as the relative band densities for the two forms.

**Phospho-GATA6L (Ser\(^{37}\)) immunoprecipitation**

HT-29 cells were plated at 75,000 cells/cm², pretreated with human IFNγ for 24 hours (Roche), and then treated with TNF (100 ng/ml; PeproTech), insulin (500 ng/ml; Sigma), or both for 1 hour. Cells were lysed in NP-40 lysis buffer (51) supplemented with 10 mM sodium pyrophosphate and 30 mM sodium fluoride, and ~4 mg of cellular extract (adjusted according to total GATA6L abundance based on immunoblotting) was incubated with 10 µl of phospho-GATA6L (Ser\(^{37}\)) antiserum overnight on a nutator at 4°C. The following day, 30 µl of Protein A/G Plus UltraLink resin (Thermo) was added to the immune complexes for 1 hour on a nutator at 4°C. Beads were washed twice with ice-cold supplemented NP-40 lysis buffer and twice with ice-cold phosphate-buffered saline (PBS) before elution in Laemmli sample buffer (136). Samples were electrophoresed on a 10% polyacrylamide gel for 3 to 5 hours (until the 50-kD marker reached the bottom of the gel) before proceeding with quantitative immunoblotting as described (51). Densitometry was performed by integrating the pixel intensity (without lane-specific background subtraction) of the 75-kD band and, for the 60-kD form, the intensity between the 75-kD band and the upper shoulder of the immunoprecipitating antibody heavy chain (file S5).

**Chromatin immunoprecipitation**

Five million wild-type and S37A GATA6L-addback HT-29 cells were seeded in 10-cm culture plates for 24 hours before inducing knockdown-addback with doxycycline (1 µg/ml) for 48 hours. Cells were fixed for 7 to 10 min by adding a 37% formaldehyde stock to the culture medium to a final concentration of 1%. Fixation was quenched with 1/20 volume of

\[
\Delta x = x_0 + \alpha_1 \left[ \frac{w_1}{X_1} + \frac{w_2}{X_2} \right] 
\]
2.5 M glycine for 7 to 10 min at room temperature. Cells were washed twice with cold PBS, scraped into 1 ml of PBS, and centrifuged at 400 relative centrifugal force (rcf) for 3 min. The cell pellets from four 10-cm plates were combined and lysed in ChIP lysis buffer (96) to a final volume of 1.5 ml. Lysates were incubated on ice for 10 min and then sonicated using a Branson digital sonifier for 5 min at 40% amplitude with 0.7-s “on” and 1.3-s “off” pulse cycles. After centrifugation at 14,000 rcf for 20 min, the supernatant was collected, and 20 μl of the soluble chromatin was retained as the input fraction. Soluble chromatin was diluted 10-fold in dilution buffer (96), pre cleared with 100 μl of mouse IgG–conjugated agarose beads (Sigma) for 4 hours at 4°C with constant agitation, and then incubated with 100 μl of anti-FLAG M2 affinity gel (Sigma) or mouse IgG–conjugated agarose beads overnight at 4°C with constant agitation. Agarose beads were collected and washed as previously described (96). DNA from the beads and the input fraction was eluted by reversing methylene cross-links with 500 μl of elution buffer (96) at 65°C for 5 hours. Samples were then treated with ribonuclease A (100 μg/ml; Sigma) for 30 min at 37°C and protease K (200 μg/ml) for 90 min at 50°C, followed by phenol–chloroform extraction. The aqueous fraction was ethanol-precipitated, washed once in 70% ethanol, air-dried, and dissolved in nuclease-free water. The samples were diluted 10-fold in nuclease-free water and quantified by PCR with primers designed for proximal promoter regions of selected crosstalk genes (123). Primer sequences are available in table S3.

Statistical analysis

Microarray data were analyzed by four-way ANOVA (factors: TNF, EGF, insulin, and time) in MATLAB with the anovan function by four-way ANOVA (factors: transcript, TNF, insulin, and time) or, for crosstalk cluster.

REFERENCES AND NOTES

www.sciencesignaling.org/cgi/content/full/9/431/ra59/DC1

SUPPLEMENTARY MATERIALS

www.sciencesignaling.org/cgi/content/full/9/431/ra59/DC1

References


Acknowledgments: We thank M. Luo of the BiobioCenter and J. Davis of the Center for Public Health Genomics for assistance with the microarray experiments. E. Jeffrey and N. Sherman of the W.M. Keck Biomedical Mass Spectrometry Laboratory for assistance with the mass spectrometry experiments, the laboratory of M. Adi for support with the ChIP experiments, M. Mayo for plasmids, C. Borgman for copyediting, and I. Schulman, D. Brautigan, and T. Harris for comments on the manuscript. Funding: This work was supported by the NIH (#1-DP2-OD00644), the Pew Charitable Trusts (#2008-000410-006), and The David and Lucile Packard Foundation (#2005-34710). The W.M. Keck Biomedical Mass Spectrometry Laboratory and The University of Virginia Biomedical Research Facility are funded by a grant from the University of Virginia’s School of Medicine. Author contributions: Z.C. performed the qRT-PCR experiments and the promoter bioinformatics analyses, prepared the mass spectrometry samples, cloned and tested the GATA6 mutants, performed the Phos-tag and standard immunoblotting as well as immunoprecipitation, engineered the GATA6 addback lines, prepared the illumina microarray experiments, and performed the ChIP-qPCR experiments. Z.Y. performed the GATA copy number analysis and assisted with the ChIP-qPCR experiments. Z.S. built the Gaussian mixture model of immunoblot data and analyzed the immunoblotts relating to GATA6 half-life. S.L. assisted with cloning of GATA6, and testing shGATA6 hairpins. R.C.F. processed the Affymetrix microarray data and performed the CLICK clustering analysis. D.A.L. supervised the tensor PLSR model construction and supported the Affymetrix microarray study. K.A.J. performed the Affymetrix microarray samples, built and analyzed the tensor PLSR model, performed the rapamycin, serum starvation, and immunofluorescence experiments, and was responsible for all statistical analysis. K.A.J. and Z.C. wrote the manuscript with input from R.C.F. and D.A.L. Competing interests: K.A.J. is a paid consultant for Cell Signaling Technology and LI-COR Biosciences.

Data and materials availability: Microarray data are available through the NCBI Gene Expression Omnibus (GSE69639 and GSE72079).

Submitted 30 August 2015
Accepted 19 May 2016

Final Publication 7 June 2016
10.1126/scisignal.aad3373

TNF-insulin crosstalk at the transcription factor GATA6 is revealed by a model that links signaling and transcriptomic data tensors

Zeinab Chitforoushzadeh, Zi Ye, Ziran Sheng, Silvia LaRue, Rebecca C. Fry, Douglas A. Lauffenburger and Kevin A. Janes

Sci. Signal. 9 (431), ra59.
DOI: 10.1126/scisignal.aad3373

Identifying integrators of multiple signals

Cells are never exposed to only one signal at a time. They are bathed in a complex, dynamically changing milieu of growth factors, nutrients, cytokines, and hormones, which creates enormous complexity for studying cellular regulation. Chitforoushzadeh et al. applied a statistical modeling approach called “tensor partial least squares regression,” which maintains data structures as multidimensional elements called tensors. Application of tensor modeling to proteomic signaling data and transcriptomic data revealed a specific phosphorylation event on the long form of the transcription factor GATA6 that enabled the growth factor insulin to inhibit the expression of genes targeted by the proinflammatory cytokine TNF. The computational analysis revealed information not readily obvious in the large data sets and provided a molecular explanation for the specific patterns of gene expression that occurred when the cells experienced growth factors in the presence or absence of a proinflammatory cytokine.