



Disentangling information flow in the Ras-cAMP signaling network

Gregory W. Carter, Steffen Rupp, Gerald R. Fink, et al.

Genome Res. 2006 16: 520-526

Access the most recent version at doi:[10.1101/gr.4473506](https://doi.org/10.1101/gr.4473506)

References This article cites 33 articles, 12 of which can be accessed free at:
<http://genome.cshlp.org/content/16/4/520.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Disentangling information flow in the Ras-cAMP signaling network

Gregory W. Carter,^{1,4} Steffen Rupp,² Gerald R. Fink,³ and Timothy Galitski¹

¹Institute for Systems Biology, Seattle, Washington 98103, USA; ²Fraunhofer Institute for Interfacial Engineering and Biotechnology IGB, 70569 Stuttgart, Germany; ³Whitehead Institute, Cambridge, Massachusetts 02142, USA

The perturbation of signal-transduction molecules elicits genomic-expression effects that are typically neither restricted to a small set of genes nor uniform. Instead there are broad, varied, and complex changes in expression across the genome. These observations suggest that signal transduction is not mediated by isolated pathways of information flow to distinct groups of genes in the genome. Rather, multiple entangled paths of information flow influence overlapping sets of genes. Using the Ras-cAMP pathway in *Saccharomyces cerevisiae* as a model system, we perturbed key pathway elements and collected genomic-expression data. Singular value decomposition was applied to separate the genome-wide transcriptional response into weighted expression components exhibited by overlapping groups of genes. Molecular interaction data were integrated to connect gene groups to perturbed signaling elements. The resulting series of linked subnetworks maps multiple putative pathways of information flow through a dense signaling network, and provides a set of testable hypotheses for complex gene-expression effects across the genome.

[Supplemental material is available online at www.genome.org. Genomic-expression data have been deposited in the Gene Expression Omnibus database under accession no. GSE2927.]

Biochemical and genetic techniques have led to a picture of intracellular signaling as a sequential cascade or pathway consisting of a limited set of signaling proteins linked by a small number of biochemical interactions. In this “sparse-network” view, signals are propagated via mostly isolated linear sequences of molecular interactions. The collection of high-throughput molecular-interaction data now allows these signaling elements to be mapped in a large dense interaction network. This “dense-network” view suggests that signaling paths are not isolated, but rather form an entangled web of numerous possible signaling avenues. The reconciliation of the sparse-network signaling concept and dense biological networks is a central problem in systems biology (Ideker et al. 2001; Marcotte 2001).

The perturbation of signal-transduction molecules can have distinct regulatory effects of differing magnitudes on overlapping sets of genes. For many genes, if not most, the expression pattern revealed by genomic expression analysis is a composite of overlapping regulatory influences. In other words, the measured expression of a gene in a condition often reflects a summation of separate influences that are prevalent in the genome. Detecting and isolating distinct overlapping expression effects of varying magnitude requires appropriate data-analysis methods. Such methods should be able to (1) decompose the expression pattern of each gene in each condition; (2) detect major expression components as well as minor but biologically informative components that may be difficult to discern; (3) identify overlapping clusters of genes sharing an expression component. Clustering algorithms in common use, for example, hierarchical clustering (Eisen et al. 1998), self-organizing maps (Tamayo et al. 1999), fuzzy k-means clustering (Gasch and Eisen 2002), and biclustering (Cheng and Church 2000), lack one or more of these properties.

⁴Corresponding author.

E-mail gcarte@systemsbiology.org; fax (206) 732-1299.

Article published online ahead of print. Article and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.4473506>.

Distinct influences of varying magnitude within genes and among overlapping gene sets can be discerned using singular value decomposition (SVD) (Weaver et al. 1999; Alter et al. 2000). SVD is an unsupervised algebraic method that mathematically separates a data matrix into a set of “modes” determined by the quantitative composition of the data. Each mode is manifest in the data as a global expression component that influences the expression of each gene to a varying degree. SVD has also proved useful in linear modeling of gene expression (Holter et al. 2000), comparative genomic-expression analysis (Alter et al. 2003), cell sample and gene classification (Ghosh 2002; Anderson et al. 2003), dimensional reduction (Horn and Axel 2002), robust expression-data cleaning (Liu et al. 2003), and network modeling (Yeung et al. 2002).

Here, we propose that a signaling regulatory influence isolated by SVD is delivered by one or a few strands, which we denote an “expression-component subnetwork,” of the dense interwoven signaling network. As a model, we consider the Ras-cAMP signaling pathway in the budding yeast *Saccharomyces cerevisiae*. This pathway is implicated in pseudohyphal growth (Gimeno and Fink 1992; Robertson and Fink 1998; Rupp et al. 1999; Stanhill et al. 1999), cell proliferation, and glycolysis (Thevelein 1992; D’Souza and Heitman 2001; Jones et al. 2003). The pathway centers on the activation of adenylate cyclase (Cyr1) by GTP-bound Ras2 and Ras1 proteins. Ras is negatively regulated by GTPase activating proteins (GAPs) encoded by the *IRA* genes. Activation of Cyr1 protein results in synthesis of cyclic-AMP (cAMP). Increasing concentration of this small-molecule messenger activates Protein Kinase A, which promotes growth and glycolysis while repressing the stress response and gluconeogenesis. This sparse-network view of the Ras-cAMP system is embedded in a dense molecular signaling network. Genetic perturbation of key elements such as the *RAS* and *IRA* genes leads to many effects that can presumably be traced back through the dense network to the perturbed element.

Results

Experimental design

Key elements of the Ras-cAMP network were perturbed in nine genomic-expression profiling experiments (Methods). The first four experimental conditions were designed to directly control the concentration of cAMP. A yeast strain with defective synthesis and defective degradation of cAMP was constructed (Methods). Cellular synthesis of cAMP was prevented by disruption of the major (*RAS2*) and minor (*RAS1*) activators of adenylate cyclase. Cellular degradation of cAMP was prevented by disruption of the cAMP phosphodiesterase gene, *PDE2*. Exogenous cAMP was infused into the cells by adding it at various concentrations (0, 0.5, 1, and 2mM) to the growth medium. This experimental design has been shown to regulate cAMP-pathway activity (Rupp et al. 1999).

The other five experiments used strains altered in their ability to transmit a signal through the cAMP pathway. Two of the strains contained a genetic modification of the Ras2 protein sequence. The *RAS2V19* dominant-active allele locks Ras2 protein in the active, GTP-bound state (Toda et al. 1985), and should constitutively signal cAMP production. The *RAS2A22* dominant-negative allele deactivates signaling activity of the protein, and should fail to up-regulate the signal for cAMP production. These two strains were tested in order to elucidate differences and similarities in the Ras2 GTPase-cycle states. GTPase activity of Ras2 protein is activated by the Ira1 and Ira2 proteins (Tanaka et al. 1990), which have similar sequence (45% identical), and similar roles. Experiments were designed for detailed comparison of Ira protein activity. We constructed strains with deletion alleles of the *IRA* genes as well as an *ira1RA* point mutation that replaces an active-site arginine with alanine. Although similar effects on global gene expression were expected, systematic differences were identified with data decomposition.

Singular value decomposition analysis

The gene-expression data were processed and then analyzed by SVD (Methods; Supplemental text). From 1676 genes showing significant expression change, a set of nine eigenconditions and a similar set of eigengenes were obtained. The n th mode (of nine total) is defined as the matrix formed by the outer product of the n th eigencondition with the n th eigengene and weighted by the n th singular value. Determining the subset of modes that are biologically meaningful requires bioinformatic analysis.

The eigenconditions are plotted in Figure 1. The eigengene matrix is too large for informative display. By inspecting the columns of the raster plot, one can discern the expression component represented in each mode. Comparisons among rows reveal similarities in expression components among conditions. The modes are ordered by their singular values (weights) from highest (Mode 1) to lowest (Mode 9). There is a clear ordering of modes in that their singular values vary widely in magnitude. However, SVD measured a high data set entropy of 0.76 (Methods), indicating genomic expression with multiple substantial genomic expression components rather than dominance by one or a few modes. By perturbing key elements in a major pathway we have apparently affected a diversity of signaling mechanisms and biological processes.

Though all genes and all conditions contribute to each SVD mode, some contributions are significant and others are negligible. To determine which genes and conditions are the most

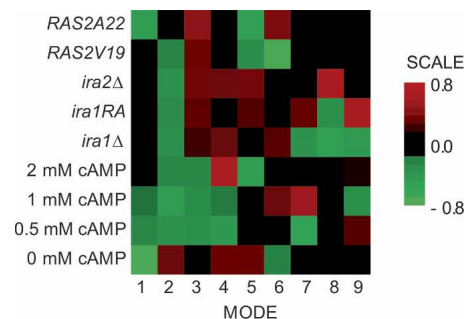


Figure 1. Raster plot of the SVD eigencondition matrix. The plot shows the contributions to each mode from each condition. Contributions are either positive (red) or negative (green). Greater contributions are brighter.

significant contributors in each SVD mode, we extracted those with eigengene and eigencondition matrix entries more than one standard deviation above or below the mean of all modes (similar to Wall et al. 2001; Supplemental text). From this, we obtained four sets for each mode, i.e., a set of positive genes, a set of negative genes, a set of positive conditions, and a set of negative conditions (Supplemental Tables 1, 2). Positive genes are up-regulated under positive conditions and down-regulated under negative conditions, whereas negative genes are regulated conversely. We stress that each sign label is defined relative to the condition set of its respective mode. Because each of the nine modes has positive and negative sets, there are 18 gene sets and 18 condition sets.

Joint membership of any gene in more than one SVD mode is possible. We found intermodal overlaps of 5%–15% (Supplemental Table 3). Statistics on mode memberships of genes are shown in Supplemental Table 4. More than half (52%) of the genes are grouped into more than one mode, and many genes (17%) appear in four or more modes. The joint membership of a gene in more than one mode indicates that the expression pattern of the gene is a weighted composite of the modes of which it is a member. The modes shown in Figure 1 define the orthogonal set of expression components from which the expression pattern of any gene can be composed. This is illustrated in Figure 2 for three genes of increasing expression complexity. Here, “complexity” refers to the number of modes exhibited by the gene. Note that the composite (measured) expression pattern of each gene in each condition is a summation of the contributions of the expression components. Simple expression patterns, such as that of the *ILV6* amino acid biosynthesis gene (Fig. 2A), can be accounted for by a combination of one or two expression components. Genes that have many substantial expression components, like the *ADH5* alcohol dehydrogenase gene (Fig. 2B) and the *MSN4* transcriptional activator gene (Fig. 2C), have a relatively unique composite expression pattern that can be described as a combination of many expression components prevalent across the genome. These examples, and the prevalence of expression-pattern complexity indicated in Supplemental Table 4, demonstrate an essential and advantageous feature of SVD analysis, i.e., the expression data set, and the expression of every gene in each condition, is decomposed into a series of components that are entirely determined by the data itself. In contrast, methods that cluster expression patterns without decomposition are not designed to isolate these overlapping regulatory influences of varying magnitude (Supplemental text; Supplemental Fig. 1),

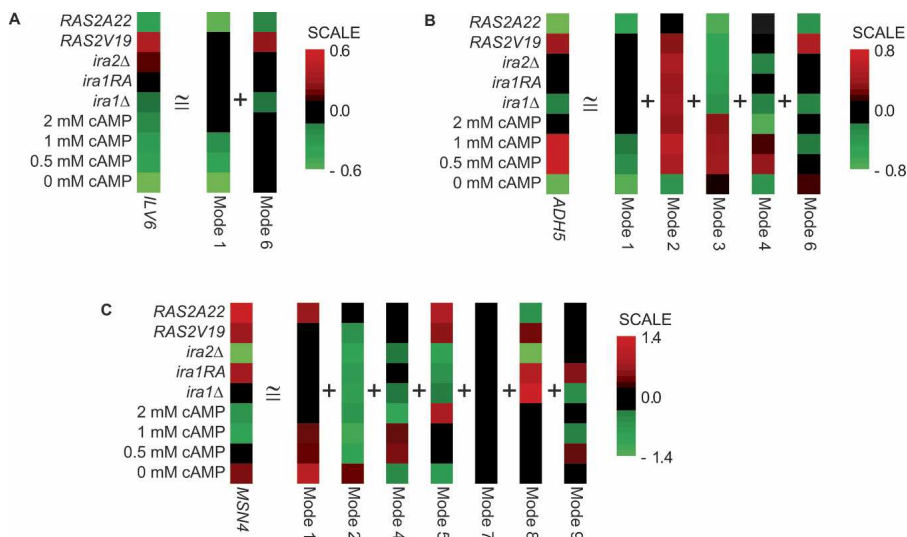


Figure 2. Decomposition of expression patterns of individual genes. The expression pattern (*left*) and equivalent sum of SVD components (*right*) are shown for three representative genes: (A) *ILV6*, (B) *ADH5*, and (C) *MSN4*. Modes shown correspond to gene sets that include the gene; contributions from other modes were not significant. Raster plots are colored with red (positive), green (negative), and brightness proportional to value.

though this is exactly what is sought from genomic-expression data in signaling perturbation studies.

Functional associations of SVD gene sets

To assess the functional relevance of each SVD gene set, we analyzed member genes for overrepresentation of genes with the same Gene Ontology annotations (Table 1; Methods). For a majority of modes there are biological processes, molecular functions, and cellular components associated with the gene sets. Generally, these associations are strongest (i.e., less likely due to chance) for the modes with high singular values. For some gene sets the lack of annotations is likely to be a consequence of expression effects that cut across functional classes, a lack of anno-

tation due to an unknown common function, or nonbiological effects such as systematic error, noise contamination, and data normalization. Nonetheless, most modes, including modes with low singular values, show significant annotations. For example, the highly significant annotation for Mode 7 and the moderately significant annotation of Mode 9 suggest that these modes carry functional information.

Transcription-factor associations with SVD gene sets

The apparent coregulation of the genes in each SVD gene set suggests cobinding of the genes by specific transcription factors. Correspondence of DNA-binding patterns and SVD gene sets would further support the biological significance of SVD modes. For each SVD gene set, we assessed the member genes for a statistical overrepresentation of targets of each of 137 transcription factors (Methods).

Between one and 11 transcription factors were found for 12 of the 18 gene sets (Table 1; Supplemental Table 5; Supplemental Fig. 2). Similar to Gene Ontology annotations previously discussed, transcription factors were more likely to be found for the modes with high singular values. Note, though, that some modes with either low weights or a lack of group annotation show significant enrichment of transcription-factor binding. This lends credence to their transcriptional coregulation. Some transcription factors (e.g., Gcn4) show enrichment for target genes in more than one gene set. A possible explanation is that the same target genes are members of more than one gene set. However, there is generally low overlap among targets of each transcription factor in different gene sets (Supplemental Table 6). This observation suggests that some in-

Table 1. SVD gene set annotations and transcription factors

Set	Genes	GO Annotation ^a	$-\log_{10}(P)^b$	Transcription Factors ^c
1-Positive	332	rRNA processing	4.8	Fhl1, Gcn4, Hap5, Leu3, Met31, Rap1, Sfp1
1-Negative	108	tricarboxylic acid cycle	2.3	Hsf1, Msn2, Msn4
2-Positive	237	aerobic respiration	7.3	Elp4, Elp6, Hap2, Hap3, Hap4, Hap5, Msn2, Nrg1, Skn7, Sok2, Sto1
2-Negative	139	Ty element transposition	6.2	Gcn4
3-Positive	164	cell wall	12.3	Aft2, Fkh2, Mbp1, Msn2, Ndd1, Rim1, Skn7, Sok2, Swi4, Swi6
3-Negative	239	ribosome biogenesis	3.0	
4-Positive	145			
4-Negative	259			Bas1, Cin5, Gcn4, Hap4, Hsf1, Msn2, Msn4, Rlm1, Skn7, Swi4
5-Positive	91			
5-Negative	286	ribosome biogenesis	3.6	
6-Positive	160	helicase activity	2.4	Aft2, Cad1
6-Negative	239	membrane	2.6	Ino2, Ino4, Sto1
7-Positive	207			
7-Negative	152	protein biosynthesis	11.3	Fhl1, Hir1, Hir2, Rap1, Sfp1
8-Positive	186			
8-Negative	132			Abf1, Gcn4, Mac1, Tye7
9-Positive	177	Ty element transposition	5.2	Cin5, Gln3, Swi4, Tye7, Yap6
9-Negative	138	nuclear chromatin	1.8	Fhl1, Hir1, Hir2

^aOverrepresented Gene Ontology annotation of highest significance (Methods). For some gene sets, no overrepresented ($P < 0.05$) annotation was found.

^bBonferroni-corrected $-\log_{10}$ probability (annotation significance).

^cTranscription factors whose target genes are overrepresented in the gene set with Bonferroni-corrected $P < 0.05$.

dividual transcription factors have separate roles in different gene-expression modes.

Expression-component subnetworks

The linkage of SVD gene sets with specific DNA-binding transcription factors enables us to link the regulated genes, via the transcription factors that bind them, to the experimentally perturbed signaling elements to form expression-component subnetworks (Fig. 3; Supplemental Fig. 3). To avoid SVD modes and gene sets that may represent nonbiological effects, expression-component subnetworks were inferred only for the 10 (of 18) gene sets that showed both significant annotation and overrepresented transcription factors (Table 1). We applied subnetwork inclusion criteria based on biological significance rather than magnitudes of singular values or similar approaches, because our experiments were specifically designed to find minor but biologically informative expression influences.

Multiple transcription factors were found for most SVD gene sets. These factors bind not only to target genes in the gene sets; they also have protein–DNA interactions among themselves. Such binding can form transcriptional regulatory loops (e.g., autoregulation, multicomponent loops, feed-forward loops) and regulatory chains and hierarchies (Lee et al. 2003). Indeed, for several gene sets we observe a greater than random degree of

intraconnectivity among the associated transcription factors (Supplemental text; Supplemental Table 7). In these subnetworks, the regulatory connections formed both loops and chains (Fig. 3A,B,E; Supplemental Fig. 3A,C). Other cases suggest parallel regulatory mechanisms (Fig. 3C,D,F; Supplemental Fig. 3B,D). The finding of regulatory intraconnections and established network motifs lends further support for the biological relevance of SVD-derived expression components.

The final step in assembling expression-component subnetworks was connecting transcription factors to the causal perturbations. Public databases were queried (Reiss et al. 2005) for protein–protein and protein–DNA interactions. Each subnetwork was constructed (Methods) using the shortest molecular interaction paths connecting the gene-set transcription factors (Table 1) with the specific signaling elements whose perturbations compose the corresponding condition set (Supplemental Table 2). Although longer pathways could often be found connecting the gene sets to central cAMP-pathway elements, shortest paths were chosen because they are most likely to be biologically active (Steffen et al. 2002). Each resulting subnetwork (Fig. 3; Supplemental Fig. 3) traces a distinct putative information flow from perturbed signaling elements, through specific molecular interactions, to a nonexclusive gene set exhibiting a distinct expression component.

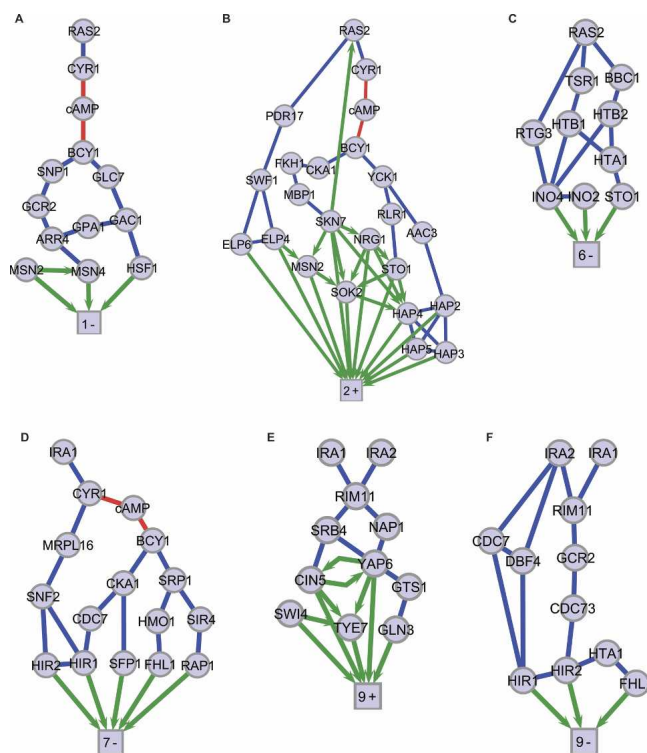


Figure 3. Expression-component subnetworks. Molecular interaction subnetworks are rendered for SVD gene sets with significant gene annotations and transcription factors: (A) 1-Negative, (B) 2-Positive, (C) 6-Negative, (D) 7-Negative, (E) 9-Positive, (F) 9-Negative. Four additional subnetworks for which significant gene annotations and transcription factors were found are shown in Supplemental Figure 3. The square nodes of each subnetwork represent a large number of differentially expressed genes and are labeled by mode number and positive or negative regulation. Protein–protein interaction edges are blue. Protein–DNA interactions are green directional edges. Red edges are interactions with the small molecule cAMP.

Discussion

SVD can isolate large and subtle overlapping effects resulting from signaling perturbations. The experiments in the present study were designed to extract information by comparing the genomic responses elicited by strategic perturbations of Ras-cAMP signaling. Analysis of the experiments by SVD permits: (1) A comparison of the effects of increasing cAMP levels. (2) A comparison of the effects of different *IRA* mutant alleles. (3) A comparison of strains carrying dominant-active and dominant-negative alleles of the major GTPase gene, *RAS2*. By isolating expression changes due to strategic perturbations of key cAMP-pathway elements, the findings directly address questions motivating the experimental design of our genomic-expression analysis of pathway genetics. Because further perturbations of correctly inferred subnetwork elements would induce predictable changes in the expression component mediated by those elements, the results suggest further experimentation to test whether regulatory influences are received through the proposed expression-component subnetworks

Expression responses to cAMP levels

Four of the experimental conditions probed the response of the cell to varying concentrations of cAMP per se. We found that the cell exhibits more than a simple monotonic response to cAMP levels. If there were a simple monotonic response, cAMP-dependent expression across the entire genome would be captured by a single mode encoding this response, and the experimental conditions of varying cAMP concentration would not contribute to any other mode. Instead we find an expression component of proportionality to cAMP levels in Mode 1, and a switch-like expression component activated by cAMP concentrations above some low threshold in Mode 2 (Fig. 1). Together, these two modes capture 63% of the information of the data set in terms of fractional singular values (Supplemental Table 2). Because Modes 1 and 2 represent the expression components of

greatest weight in the data set, they are the dominant feature of the expression patterns of Mode-1 and Mode-2 genes (Supplemental Fig. 4). Their high singular values, plus the significance of their annotations and transcription-factor binding patterns (Table 1), indicate that these regulatory effects are biological, and not due to noise or some other nonbiological cause.

Expression-pattern decomposition and subnetwork mapping achieves a level of network detail greater than previous expression analyses of the Ras-cAMP pathway. For example, Wang and collaborators (Wang et al. 2004) used correlation-based analysis of expression patterns without decomposition to show that Ras2 activation of cAMP synthesis represses genes bound by Msn2/4 and the Hap complex. We found Msn2/4 and the Hap complex to be regulators of the 1-Negative and 2-Positive gene sets, respectively. We are further able to specify that the Msn2/4 target genes are repressed proportionally to cAMP concentration, whereas targets of the Hap complex exhibit a switch-like behavior. Furthermore, the different proteins composing the 1-Negative and 2-Positive expression-component subnetworks (Figs. 3A,B) suggest that two separate mechanisms of signaling are responsible for the two modes of regulation.

A novel function for the Ira1 protein?

The results allow a comparison of two *IRA1* alleles, one producing a protein lacking Ras-GAP activity, the other producing no Ira1 protein. The experiments probe for roles other than the Ras-GAP activity in the very large (351 kD) Ira1 protein. The two *IRA1* mutations (null and GAP-defective point mutation) engender similar expression responses, but differ sharply in Modes 7 and 9 (Fig. 1). This minor difference (2% of the information in the data) is difficult to detect without decomposition of the data (Supplemental Figs. 1 and 5). Although one should note that noise in the data is likely distributed in these low-weight modes along with any information of biological importance, Modes 7 and 9 show significant annotation as well as overrepresented transcription factors (Table 1) enabling the mapping of expression-component subnetworks. These results raise the possibility that there is another Ira1 signaling function distinct from the Ras-GAP activity. Moreover, the results link this putative function with specific biological processes and molecular mediators. For example, the 138 9-Negative genes are overrepresented with genes functioning in DNA metabolism and binding, and chromatin architecture and assembly (Table 1; Supplemental Table 1). Expression of these genes was positively affected by *ira1Δ*, but negatively by *ira1RA* (Fig. 1). We obtained an expression-component subnetwork (Fig. 3F) connecting the perturbed genes to the histone-gene-transcription inhibitors Hir1 and Hir2. One can speculate that these findings are related to *IRA1*-associated epigenetic effects on gene expression (Halme et al. 2004).

Differential expression for *RAS2* point mutants

The experimental conditions include dominant-active and dominant-negative alleles of the major GTPase gene, *RAS2*. The encoded protein variants are locked in different states of the Ras GTPase cycle. Although differences between the *RAS2* point mutants are evident in multiple modes, a clear anticorrelation between the dominant-active *RAS2V19* and dominant-negative *RAS2A22* conditions is contained in Mode 6 (Fig. 1). Though this mode represents only 3% of the data (Supplemental Table 2), it shows significant associations with specific annotations and transcription factors (Table 1). For example, the 239 genes in the

6-Negative gene set are up-regulated in the *RAS2V19* strain and down-regulated in the *RAS2A22* strain. Many of these genes were found to be cell-wall and membrane associated (Table 1; Supplemental Table 1). The expression patterns and Mode-6 expression components of these genes are shown in Supplemental Figure 6. The expression patterns of these genes are dissimilar for most conditions. SVD identifies an underlying expression component (in Mode 6) showing an anticorrelation between the two *RAS2* mutants that is not accounted for in other modes. For example, the expression patterns of two hexose transporter genes, *HXT6* and *HXT7*, do not naively exhibit this anticorrelation. However, after subtracting the contributions from other modes (especially Mode 1), the anticorrelation is revealed. The isolation of the effects in Modes 1 and 6 is also evident in the corresponding expression-component subnetworks. For example, the results allow the hypothesis that separate subnetworks deliver the Mode-1 (Fig. 3A) and Mode-6 (Fig. 3C) influences to the *HXT6* promoter. There is evidence (Harbison et al. 2004) that Hsf1, a regulator of Mode 1, and Ino4, a regulator of Mode 6, both bind the *HXT6* promoter. This analysis illustrates how SVD and subnetwork mapping separates the expression components of genes and identifies otherwise hidden regulatory influences.

Signaling through dense molecular networks

The signaling subnetworks represent testable hypotheses for signal-transduction avenues that deliver regulatory influences to

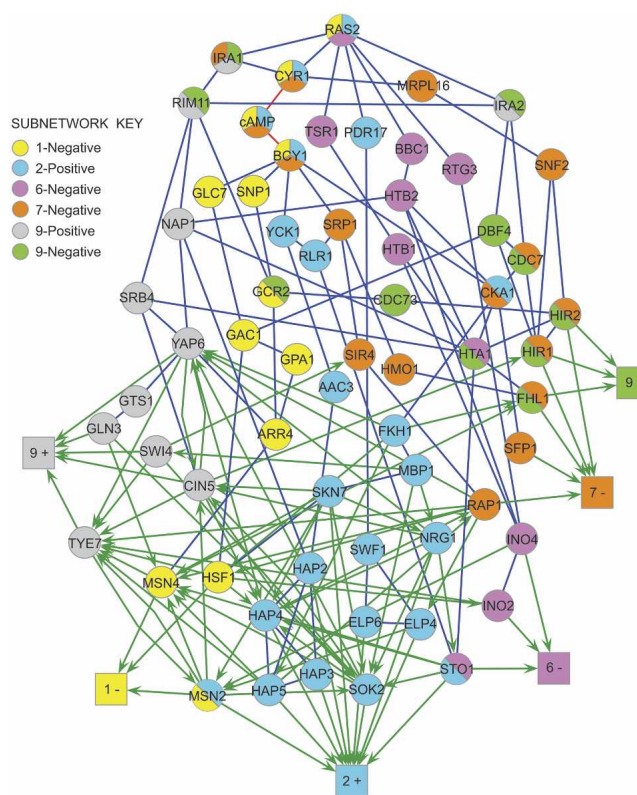


Figure 4. Expression-component subnetworks entangled in a dense composite network. The composite network is the union of all subnetworks in Figure 3. Graph-element representations are as in Figure 3. The composite network for all subnetworks in Figure 3 and Supplemental Figure 3 is shown in Supplemental Figure 7. Nodes are colored based on their subnetwork membership. Sectored node color indicates joint membership.

cofunctioning genes. The union of expression-component subnetworks into a composite network indicates a dense-network organization of entangled signaling paths. A composite network for a subset (those in Fig. 3) of the expression-component subnetworks is shown in Figure 4. A composite network for all (those in Fig. 3 and Supplemental Figure 3) of the expression-component subnetworks is shown in Supplemental Figure 7. The sparse-network view of the Ras-cAMP signaling system as a single, simple, and well-isolated pathway cannot accommodate the results. The multiple subnetworks show considerable overlap that extends beyond the central Ras-cAMP pathway elements that were experimentally perturbed. These observations suggest that the core Ras-cAMP pathway is structurally and functionally embedded in the dense structure of entangled expression-component subnetworks.

Methods

Strains and growth conditions

Standard strain construction methods and growth medium formulations were used (Guthrie and Fink 1991). The following strains were constructed and subjected to genomic expression profiling:

SR959: *MATa/α ras1::HIS3/ras1::HIS3 ras2Δ/ras2Δ pde2::kan^R/pde2::kan^R ura3-52/ura3-52 leu2::hisG/leu2::hisG his3::hisG/his3::hisG TRP1⁺/trp1::hisG* pRS315 (Rupp et al. 1999);
 SR628: *MATa/α ira1::HIS3/ira1::LEU2 ura3-52/ura3-52 leu2::hisG/leu2::hisG his3::hisG/his3::hisG* pRS316 (this study);
 SR640: *MATa/α ira2::HIS3/ira2::LEU2 ura3-52/ura3-52 leu2::hisG/leu2::hisG his3::hisG/his3::hisG* pRS316 (this study);
 SR1184: *MATa/α ira1::HIS3/ira1::LEU2 ura3-52/ura3-52 leu2::hisG/leu2::hisG his3::hisG/his3::hisG* pRS316-IRA1RA (this study);
 SR1185: *MATa/α ura3-52/ura3-52* pRS316-RAS2V19 (this study);
 SR1186: *MATa/α ura3-52/ura3-52* pRS316-RAS2A22 (this study);
 SR1187: *MATa/α ura3-52/ura3-52* pRS316.

For genomic-expression analysis of the response to varying cAMP concentration, strain SR959 was grown in Synthetic Complete (SC) medium, 2% glucose, with 1mM cAMP to $OD_{600} = 1$. The culture was split and diluted to $OD_{600} = 0.3$ in fresh SC medium with either 0, 0.5, 1.0, or 2 mM cAMP. These cultures were grown to $OD_{600} = 1.0$ and harvested by centrifugation. Before and after each experiment, strain SR959 was checked for suppressor mutations by plating on YPD (rich medium) as in Rupp et al. (1999). Experiments containing suppressors (growth on YPD) were discarded. For all other strains, a culture was inoculated to $OD_{600} = 0.1-0.2$ in SC medium, 2% glucose, without uracil, grown to $OD_{600} = 1.0$ and harvested by centrifugation.

Genomic expression data collection and analysis

From cell pellets, total RNA was extracted using a hot-acid phenol preparation. Using the PolyATtract system (Promega), poly(A⁺) RNA was enriched from total RNA. In duplicate for all samples except 0.5 and 1.0 mM cAMP, biotinylated RNA targets were synthesized from poly(A⁺) RNA by reverse transcription followed by in vitro transcription of the resulting cDNAs (Wodicka et al. 1997). Using the Affymetrix GeneChip system, expression levels (trimmed average difference from gene-specific perfect-match and mismatch oligonucleotide probe pairs) were derived from microarray intensity data. The data were normalized using a bulk-signal method. A set of 1676 genes with expression substantially different from wild type (>20% change in signal intensity

for at least one condition) and consistency over replicates (<50% intensity variation) was extracted from the genomic expression data. Although this cutoff accepts relatively small changes in expression, the procedure was designed to minimally bias the data in order to take full advantage of unsupervised analysis. Genes that did not exhibit at least one expression component were not assigned to any gene sets for further analysis; we found 304 such genes (Supplemental Table 4). All expression levels were scaled as log-ratio relative to a wild-type control strain. Singular value decomposition was performed on the data set, a 9×1676 matrix (Supplemental text). Analysis was carried out with the commercial software package Mathematica on a desktop PC, with which most algebraic operations took <1 sec using standard routines.

Functional analysis of SVD-derived gene sets

Each gene set was screened for statistical overrepresentation of biological process, cellular component, and molecular function annotations curated by the Gene Ontology Consortium (www.geneontology.org). The null hypothesis was that genes with a common annotation are distributed randomly throughout the gene sets. The probability of finding the obtained number of identically annotated genes within a random set was computed from a hypergeometric distribution, considering annotation classes with at least five genes. Calculated probabilities were corrected using the Bonferroni method to normalize for the number of tests performed. The most significant annotation was reported, if $P < 0.05$.

Associating transcription factors with SVD gene sets

Specific transcription factors were associated with specific SVD gene sets by finding statistical overrepresentation of target genes for transcription factors. From published high-throughput studies of protein-DNA interactions in yeast (Kellis et al. 2003; Lee et al. 2003; Zeitlinger et al. 2003; Harbison et al. 2004) we assembled a collection of 13,000 interactions involving 137 transcription factors and 4298 target genes. Overrepresentation was defined as having a Bonferroni-corrected probability below 0.05 when compared with the null hypothesis of transcription factor targets randomly distributed across gene sets, which follows a hypergeometric distribution. The incompleteness of existing protein-DNA interaction data is evident in the results. On average, interactions existed for about 75% of the genes.

Construction of expression-component subnetworks

Expression-component subnetworks were assembled by connecting the transcription factor set of a gene set (Table 1) to the perturbed Ras-cAMP pathway elements implicated specifically in the respective mode (Supplemental Table 2) via protein-DNA and protein-protein interactions forming the shortest paths. Connecting physical interactions were found by loading the network elements into the *Cytoscape* software platform (Shannon et al. 2003; www.cytoscape.org) and using the *InteractionFetcher* plugin (Reiss et al. 2005), a tool that searches the public databases DIP (dip.doe-mbi.ucla.edu), BIND (bind.ca), and the data of Harbison et al. (2004) for protein-protein and protein-DNA interactions of specified nodes in a biological network.

Acknowledgments

We thank Hui Ge, Susanne Prinz, David Reiss, James Taylor, and Vesteinn Thorsson for their contributions. T. Galitski is a recipient of a Burroughs Wellcome Fund Career Award in the Biomedical Sciences. G.F.R. was funded by NIH grant GM035010.

References

- Alter, O., Brown, P.O., and Botstein, D. 2000. Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci.* **97**: 10101–10106.
- . 2003. Generalized singular value decomposition for comparative analysis of genome-scale expression data sets of two different organisms. *Proc. Natl. Acad. Sci.* **100**: 3351–3356.
- Anderson, A., Hudson, M., Chen, W., and Zhu, T. 2003. Identification of nutrient partitioning genes participating in rice grain filling by singular value decomposition (SVD) of genome expression data. *BMC Genomics* **4**: 26.
- Cheng, Y. and Church, G.M. 2000. Biclustering of expression data. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **8**: 93–103.
- D'Souza, C.A. and Heitman, J. 2001. Conserved cAMP signaling cascades regulate fungal development and virulence. *FEMS Microbiol. Rev.* **25**: 349–364.
- Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci.* **95**: 14863–14868.
- Gasch, A.P. and Eisen, M.B. 2002. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biol.* **3**: research0059.
- Ghosh, D. 2002. Singular value decomposition regression models for classification of tumors from microarray experiments. *Pac. Symp. Biocomput.* 18–29.
- Gimeno, C.J. and Fink, G.R. 1992. The logic of cell division in the life cycle of yeast. *Science* **257**: 626.
- Guthrie, C. and Fink, G.R. 1991. *Guide to yeast genetics and molecular biology*, Vol. 194. Academic Press, New York.
- Halme, A., Bumgarner, S., Styles, C., and Fink, G.R. 2004. Genetic and epigenetic regulation of the FLO gene family generates cell-surface variation in yeast. *Cell* **116**: 405–415.
- Harbison, C.T., Gordon, D.B., Lee, T.I., Rinaldi, N.J., Macisaac, K.D., Danford, T.W., Hannett, N.M., Tagne, J.B., Reynolds, D.B., Yoo, J., et al. 2004. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**: 99–104.
- Holter, N.S., Mitra, M., Maritan, A., Cieplak, M., Banavar, J.R., and Fedoroff, N.V. 2000. Fundamental patterns underlying gene expression profiles: Simplicity from complexity. *Proc. Natl. Acad. Sci.* **97**: 8409–8414.
- Horn, D. and Axel, I. 2002. Novel clustering algorithm for microarray expression data in a truncated SVD space. *Bioinformatics* **19**: 1110–1115.
- Ideker, T., Galitski, T., and Hood, L. 2001. A new approach to decoding life: Systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**: 343–372.
- Jones, D.L., Petty, J., Hoyle, D.C., Hayes, A., Ragni, E., Popolo, L., Oliver, S.G., and Stateva, L.I. 2003. Transcriptome profiling of a *Saccharomyces cerevisiae* mutant with a constitutively activated Ras/cAMP pathway. *Physiol. Genomics* **16**: 107–118.
- Kellis, M., Patterson, N., Endrizzi, M., Birren, B., and Lander, E.S. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**: 241–254.
- Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. 2003. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**: 799–804.
- Liu, L., Hawkins, D.M., Ghosh, S., and Young, S.S. 2003. Robust singular value decomposition analysis of microarray data. *Proc. Natl. Acad. Sci.* **100**: 13167–13172.
- Marcotte, E.M. 2001. The path not taken. *Nat. Biotechnol.* **19**: 626–627.
- Reiss, D.J., Avila-Campillo, I., Thorsson, V., Schwikowski, B., and Galitski, T. 2005. Tools enabling the elucidation of molecular pathways active in human disease: Application to Hepatitis C Virus infection. *BMC Bioinformatics* **6**: 154.
- Robertson, L.S. and Fink, G.R. 1998. The three yeast A kinases have specific signaling functions in pseudohyphal growth. *Proc. Natl. Acad. Sci.* **95**: 13783–13787.
- Rupp, S., Summers, E., Lo, H.J., Madhani, H., and Fink, G. 1999. MAP kinase and cAMP filamentation signaling pathways converge on the unusually large promoter of the yeast FLO11 gene. *EMBO J.* **18**: 1257–1269.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**: 2498–2504.
- Stanhill, A., Schick, N., and Engelberg, D. 1999. The yeast ras/cyclic AMP pathway induces invasive growth by suppressing the cellular stress response. *Mol. Cell. Biol.* **19**: 7529–7538.
- Steffen, M., Petti, A., Aach, J., D'haeseleer, P., and Church, G. 2002. Automated modelling of signal transduction networks. *BMC Bioinformatics* **3**: 34.
- Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrov, S., Lander, E.S., and Golub, T.R. 1999. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci.* **96**: 2907–2912.
- Tanaka, K., Nakafuku, M., Satoh, T., Marshall, M.S., Gibbs, J.B., Matsumoto, K., Kaziro, Y., and Toh-e, A. 1990. *S. cerevisiae* genes *IRA1* and *IRA2* encode proteins that may be functionally equivalent to mammalian ras GTPase activating protein. *Cell* **60**: 803–807.
- Thevelein, J.M. 1992. The RAS-adenylate cyclase pathway and cell cycle control in *Saccharomyces cerevisiae*. *Antonie Van Leeuwenhoek* **62**: 109–130.
- Toda, T., Uno, I., Ishikawa, T., Powers, S., Kataoka, T., Broek, D., Cameron, S., Broach, J., Matsumoto, K., and Wigler, M. 1985. In yeast, RAS proteins are controlling elements of adenylate cyclase. *Cell* **40**: 27–36.
- Wall, M.E., Dyck, P.A., and Brettin, T.S. 2001. SVDMAN—singular value decomposition analysis of microarray data. *Bioinformatics* **17**: 566–568.
- Wang, Y., Pierce, M., Schnepfer, L., Guldal, C.G., Zhang, X., Tavazoie, S., and Broach, J.R. 2004. Ras and Gpa2 mediate one branch of a redundant glucose signaling pathway in yeast. *PLoS Biol.* **2**: e128.
- Weaver, D.C., Workman, C.T., and Stormo, G.D. 1999. Modeling regulatory networks with weight matrices. *Pac. Symp. Biocomput.* 112–123.
- Wodicka, L., Dong, H., Mittmann, M., Ho, M.H., and Lockhart, D.J. 1997. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **15**: 1359–1367.
- Yeung, M.K.S., Tegner, J., and Collins, J.J. 2002. Reverse engineering gene networks using singular value decomposition and robust regression. *Proc. Natl. Acad. Sci.* **99**: 6163–6168.
- Zeitlinger, J., Simon, I., Harbison, C.T., Hannett, N.M., Volkert, T.L., Fink, G.R., and Young, R.A. 2003. Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell* **113**: 395–404.

Received July 21, 2005; accepted in revised form January 17, 2006.