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 wwwgenome.org. Genomic-expression data have been deposited in the Gene Expression Omnibus database under accession no. GSE2927.]
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 2 mM) 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-activating 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. 1999), 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 IRA1 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 nth mode (of nine total) is defined as the matrix formed by the outer product of the nth eigencondition with the nth eigengene and weighted by the nth 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 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),
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 annotation 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 co-binding 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 S5; 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 S6). This observation suggests that some in-

![Figure 2.](image)

**Table 1.** SVD gene set annotations and transcription factors

<table>
<thead>
<tr>
<th>Set</th>
<th>Genes</th>
<th>GO Annotation</th>
<th>- Log10(P)</th>
<th>Transcription Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-Positive</td>
<td>332</td>
<td>rRNA processing</td>
<td>4.8</td>
<td>Fhl1, Gcn4, Hap5, Leu3, Met31, Rap1, Sfp1</td>
</tr>
<tr>
<td>1-Negative</td>
<td>188</td>
<td>tricarboxylic acid cycle</td>
<td>2.3</td>
<td>Hsf1, Msn2, Msn4</td>
</tr>
<tr>
<td>2-Positive</td>
<td>237</td>
<td>aerobic respiration</td>
<td>7.3</td>
<td>Elp4, Elp6, Hap2, Hap3, Hap4, Hap5, Msn2, Nrg1, Skn7, Sok2, Sto1</td>
</tr>
<tr>
<td>2-Negative</td>
<td>139</td>
<td>Ty element transposition</td>
<td>6.2</td>
<td>Cin5, Gcn4</td>
</tr>
<tr>
<td>3-Positive</td>
<td>164</td>
<td>cell wall</td>
<td>12.3</td>
<td>Aft2, Fkh2, Mbp1, Msn2, Ndd1, Rim1, Skn7, Sok2, Swi4, Swi6</td>
</tr>
<tr>
<td>3-Negative</td>
<td>239</td>
<td>ribosome biogenesis</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>4-Positive</td>
<td>145</td>
<td>ribosome biogenesis</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>4-Negative</td>
<td>259</td>
<td>ribosome biogenesis</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>5-Positive</td>
<td>91</td>
<td>ribosome biogenesis</td>
<td>11.3</td>
<td></td>
</tr>
<tr>
<td>5-Negative</td>
<td>286</td>
<td>ribosome biogenesis</td>
<td>3.6</td>
<td></td>
</tr>
<tr>
<td>6-Positive</td>
<td>160</td>
<td>helicase activity</td>
<td>2.4</td>
<td>Aft2, Cad1</td>
</tr>
<tr>
<td>6-Negative</td>
<td>239</td>
<td>membrane</td>
<td>2.6</td>
<td>Ino2, Ino4, Sto1</td>
</tr>
<tr>
<td>7-Positive</td>
<td>207</td>
<td>membrane</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>7-Negative</td>
<td>152</td>
<td>protein biosynthesis</td>
<td>11.3</td>
<td>Fhl1, Hir1, Hir2, Rap1, Sfp1</td>
</tr>
<tr>
<td>8-Positive</td>
<td>186</td>
<td>protein biosynthesis</td>
<td>11.3</td>
<td></td>
</tr>
<tr>
<td>8-Negative</td>
<td>132</td>
<td>protein biosynthesis</td>
<td>11.3</td>
<td></td>
</tr>
<tr>
<td>9-Positive</td>
<td>177</td>
<td>Ty element transposition</td>
<td>5.2</td>
<td>Cin5, Cin3, Swi4, Tye7, Yap6</td>
</tr>
<tr>
<td>9-Negative</td>
<td>138</td>
<td>nuclear chromatin</td>
<td>1.8</td>
<td>Fhl1, Hir1, Hir2</td>
</tr>
</tbody>
</table>

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

*Bonferroni-corrected - log10 probability (annotation significance).

*Transcription 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 (Stefen 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.

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
transcription factors (Table 1). For example, the 239 genes in the mode represents only 3% of the data (Supplemental Table 2), it conditions is contained in Mode 6 (Fig. 1). Though this RAS2A22 between the dominant-active mutants are evident in multiple modes, a clear anticorrelation be-

Differential expression for RAS2 point mutants

The experimental conditions include dominant-active and domin-

A novel function for the Ira1 protein?
The results allow a comparison of two IRA1 alleles, one produc-

Expression-pattern decomposition and subnetwork mapping achieves a level of network detail greater than previous ex-

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 (Supple-

greatest weight in the data set, they are the dominant feature of the expression patterns of Mode-1 and Mode-2 genes (Supple-

Expression-pattern decomposition and subnetwork mapping achieves a level of network detail greater than previous ex-

Differential expression for RAS2 point mutants

The experimental conditions include dominant-active and domi-

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.

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:

SR95: MATa/ara1::HIS3/ara1::HIS3 ras2A/ara2::pde2::kan8/pde2::kan8 ura3–52/ura3–52 leu2::hisG/leu2::hisG his3::hisG/ his3::hisG TRP1/T::hisG prS315 (Rupp et al. 1999);
SR628: MATa/ira1::HIS3/ira1::LEU2 ura3–52/ura3–52 leu2::hisG/ leu2::hisG his3::hisG/ his3::hisG prS316 (this study);
SR640: MATa/ira1::HIS3/ira1::LEU2 ura3–52/ura3–52 leu2::hisG/ leu2::hisG his3::hisG/ his3::hisG prS316 (this study);
SR 1184: MATa/ira1::HIS3/ira1::LEU2 ura3–52/ura3–52 leu2::hisG/ leu2::hisG his3::hisG/ his3::hisG prS316-Ira1RA (this study);
SR1185: MATa/ura3–52/ura3–52 prS316-RAS2V19 (this study);
SR1186: MATa/ura3–52/ura3–52 prS316-RAS2A22 (this study);
SR1187: MATa/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 OD600 = 1. The culture was split and diluted to OD600 = 0.3 in fresh SC medium with either 0.5, 1.0, or 2 mM cAMP. These cultures were grown to OD600 = 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 OD600 = 0.1–0.2 in SC medium, 2% glucose, without uracil, grown to OD600 = 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 module (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


