Reconstructing the pathways of a cellular system from genome-scale signals by using matrix and tensor computations

Orly Alter* and Gene H. Golub‡§

*Department of Biomedical Engineering and Institute for Cellular and Molecular Biology, University of Texas, Austin, TX 78712; and ‡Scientific Computing and Computational Mathematics Program and Department of Computer Science, Stanford University, Stanford, CA 94305

Contributed by Gene H. Golub, October 17, 2005

We describe the use of the matrix eigenvalue decomposition (EVD) and pseudoinverse projection and a tensor higher-order EVD (HOEVD) in reconstructing the pathways that compose a cellular system from genome-scale nondirectional networks of correlations among the genes of the system. The EVD formulates a genes × genes network as a linear superposition of genes × genes decorrelated and decoupled rank-1 subnetworks, which can be associated with functionally independent pathways. The integrative pseudoinverse projection of a network computed from a “data” signal onto a designated “basis” signal approximates the network as a linear superposition of only the subnetworks that are common to both signals and simulates observation of only the pathways that are manifest in both experiments. We define a comparative HOEVD that formulates a series of networks as linear superpositions of decorrelated rank-1 subnetworks and the rank-2 couplings among these subnetworks, which can be associated with independent pathways and the transitions among them common to all networks in the series or exclusive to a subset of the networks. Boolean functions of the discretized subnetworks and couplings highlight differential, i.e., pathway-dependent, relations among genes. We illustrate the EVD, pseudoinverse projection, and HOEVD of genome-scale networks with analyses of yeast DNA microarray data.

DNA microarrays | eigenvalue decomposition | higher-order eigenvalue decomposition | pseudoinverse projection | yeast Saccharomyces cerevisiae cell cycle and mating

DNA microarrays make it possible to record the complete genomic signals, such as mRNA expression (e.g., refs. 1 and 2) and DNA-bound proteins’ occupancy levels (e.g., ref. 3), that are generated and sensed by cellular systems. The underlying genome-scale networks of relations among all genes of the cellular systems can be computed from these signals (e.g., refs. 4–6). These relations among the activities of genes, not only the activities of the genes alone, are known to be pathway-dependent, i.e., conditioned by the biological and experimental settings in which they are observed (e.g., ref. 7). For example, the mRNA expression patterns of the yeast Saccharomyces cerevisiae genes KAR4 and CIK1 are correlated during mating yet anti-correlated during cell-cycle progression (8). A single genome-scale nondirectional network of correlations cannot describe the pathway-dependent differences in relations, such as those between the expression patterns of KAR4 and CIK1.

Recently, we showed that the matrix singular-value decomposition (SVD), generalized SVD, and pseudoinverse projection separate genome-scale signals, i.e., gene and array patterns of, e.g., mRNA expression and proteins’ DNA binding, into mathematically defined patterns that correlate with the independent biological and experimental processes and cellular states that compose the signals (9–12). For example, the comparative generalized SVD of yeast and human mRNA expression during their cell cycles formulates the yeast expression as a linear superposition of cell-cycle oscillations, which are common to the yeast and human, and response to synchronization by the mating pheromone, which is exclusive to the yeast, and describes a differential relation in the expression of genes such as KAR4 and CIK1 that is in agreement with their pathway-dependent activities (11).

Now, we describe the use of the matrix eigenvalue decomposition (EVD) and pseudoinverse projection and a tensor higher-order EVD (HOEVD) in reconstructing the pathways, or genome-scale pathway-dependent relations among the genes of a cellular system, from nondirectional networks of correlations, which are computed from measured genomic signals and tabulated in symmetric matrices. The EVD formulates a genes × genes network, which is computed from a “data” signal, as a linear superposition of genes × genes decorrelated and decoupled rank-1 subnetworks. The integrative pseudoinverse projection of a network, computed from a data signal, onto a designated “basis” signal approximates the network as a linear superposition of only the subnetworks that are common to both signals, i.e., pseudoinverse projection filters off the network the subnetworks that are exclusive to the data signal. We show that the pseudoinverse-projected network simulates observation of only the pathways that are manifest under both sets of the biological and experimental conditions where the data and basis signals are measured. We define a comparative HOEVD that formulates a series of networks computed from a series of signals as linear superpositions of decorrelated rank-1 subnetworks and the rank-2 couplings among these subnetworks. We show that significant HOEVD subnetworks and couplings might represent independent pathways or transitions among them common to all or exclusive to a subset of the signals. Boolean functions of the discretized subnetworks and couplings highlight known as well as previously unknown differential, i.e., pathway-dependent relations between genes. We illustrate the EVD, pseudoinverse projection, and HOEVD of genome-scale networks with analyses of mRNA expression data from the yeast Saccharomyces cerevisiae during its cell cycle (1) and DNA-binding data of yeast transcription factors that are involved in cell-cycle, development, and biosynthesis programs (3).

Mathematical Methods: EVD, Pseudoinverse Projection, and HOEVD of Networks

Eigenvalue Decomposition. Let the symmetric matrix \( A \) of size \( N \)-genes × \( N \)-genes tabulate the genome-scale nondirectional

Conflict of interest statement: No conflicts declared.

Abbreviations: EVD, eigenvalue decomposition; HOEVD, higher-order EVD; SVD, singular-value decomposition.

© 2005 by The National Academy of Sciences of the USA
network of correlations among the genes of a cellular system. The network \( \hat{a}_1 \) is computed from a genome-scale signal, designated the data signal, of, e.g., mRNA expression levels measured in a set of \( M_1 \) samples of the cellular system using \( M_1 \) DNA microarrays and tabulated in the \( N \)-genes \( \times M_1 \)-arrays matrix \( \bar{e}_1 \), such that \( \hat{a}_1 = \hat{e}_1 \bar{e}_1^T \). We compute the EVD of the network \( \hat{a}_1 \),

\[
\hat{a}_1 = \bar{u}_1 \hat{e}_1 \bar{e}_1^T \bar{u}_1^T,
\]

from the SVD of the data signal \( \hat{e}_1 = \bar{u}_1 \hat{e}_1 \bar{e}_1^T \). The \( M_1 \)-“eigenarrays” \( \times M_1 \)-“eigengenes” diagonal matrix \( \hat{e}_1 \) defines the \( M_1 \) nonnegative “eigenexpression” levels, such that the expression of the \( m \)th eigengene in the \( m \)th eigenarray is the \( m \)th eigenexpression level of \( \hat{e}_1, e_{1,m} = \langle \bar{m} | \hat{e}_1 | \bar{m} \rangle \geq 0 \). The orthogonal transformation matrices \( \bar{u}_1 \) and \( \bar{e}_1^T \) define the \( N \)-genes \( \times M_1 \)-eigenarrays and the \( M_1 \)-eigengenes \( \times M_1 \)-arrays subspaces, respectively. The \( m \)th column of \( \hat{a}_1, [\alpha_{1,m}] = [\bar{u}_1(m) \bar{m}] \), lists the genome-scale expression of the \( m \)th eigenarray of \( \hat{e}_1 \). The \( m \)th row of \( \bar{e}_1^T, \langle \gamma_{1,m} | = \langle n | \bar{e}_1^T(n) \rangle \), lists the expression of the \( m \)th eigengene.

EVD formulates the network \( \hat{a}_1 \) as a linear superposition of a series of \( M_1 \) rank-1 symmetric “subnetworks” of size \( N \)-genes \( \times N \)-genes each, where the \( m \)th subnetwork is the outer product of the \( m \)th eigenarray with its transpose \( [\alpha_{1,m}]^T [\alpha_{1,m}] \) (Fig. 5 in Supporting Appendix, which is published as supporting information on the PNAS web site),

\[
\hat{a}_1 = \sum_{m=1}^{M_1} e_{1,m}^2 [\alpha_{1,m}]^T [\alpha_{1,m}].
\]

The significance of the \( m \)th subnetwork is indicated by the \( m \)th “fraction of eigenexpression” \( p_{1,m} = e_{1,m}^2 / (\sum_{m=1}^{M_1} e_{1,m}^2) \), i.e., the expression correlation captured by the \( m \)th subnetwork relative to that captured by all subnetworks. Each subnetwork is decorrelated of all other subnetworks, i.e., \( [\alpha_{1,m}]^T [\alpha_{1,n}] = 0 \) for all \( m \neq n \), since \( \bar{u}_1 \) is orthogonal. Each subnetwork is also decoupled of all other subnetworks, such that there are no contributions to the network \( \hat{a}_1 \) from the \( M_1(M_1 - 1)/2 \) rank-2 symmetric “couplings” among the subnetworks, i.e., \( [\alpha_{1,m}]^T [\alpha_{1,n}] = 0 \) for all \( m \neq n \), since \( \bar{e}_1 \) is diagonal. For a real data signal \( \hat{e}_1 \), the eigenarrays are unique up to phase factors of \( \pm 1 \), and therefore the subnetworks are also unique, i.e., data-driven, except in degenerate subspaces defined by subsets of equal eigenexpression levels.

**Pseudoinverse Projection.** Let the matrix \( \hat{b} \) of size \( N \)-genes \( \times L \)-arrays tabulate the genome-scale signal, designated the “basis” signal, of, e.g., proteins’ DNA-binding occupancy levels measured in a set of \( L \) samples of the cellular system using \( L \) arrays. We compute the pseudoinverse projection (12, 13) of the \( N \)-genes \( \times M \)-arrays matrix \( \hat{e}_1 \), such that \( \hat{a}_1 = \hat{e}_1 \hat{b}^\dagger \). We compute the EVD of the projected network \( \hat{a}_2 \),

\[
\hat{a}_2 = \bar{u}_2 \hat{e}_2 \tilde{\bar{e}}_2 \tilde{\bar{u}}_2^T = \sum_{m=1}^{M_2} e_{2,m}^2 [\alpha_{2,m}]^T [\alpha_{2,m}],
\]

where \( M_2 = \min\{L, M_1 \} \). From the SVD of the projected signal \( \hat{e}_2 = \hat{a}_2 \hat{b}^\dagger \), the \( m \)th column of \( \hat{a}_2, [\alpha_{2,m}] = [\bar{u}_2(m) \bar{m}] \), lists the genome-scale expression of the \( m \)th eigenarray of \( \hat{e}_2 \). In reconstructing \( \hat{a}_2 \), the pseudoinverse projection filters out of \( \hat{a}_1 \) each of its subnetworks \( [\alpha_{1,m}]^T [\alpha_{1,m}] \), which is decorrelated of the series of \( L \) rank-1 symmetric subnetworks \( [\beta_1 \beta_2 \ldots \beta_{L-1}] \) that compose the network \( \hat{b} \hat{b}^T \) computed from the basis signal \( \hat{b} \), such that \( \langle [\beta_1 \beta_2 \ldots \beta_{L-1}] | [\alpha_{1,m}]^T [\alpha_{1,m}] = 0 \) for all \( l = 1, 2, \ldots, L \) (Fig. 6 in Supporting Appendix).

**Higher-Order EVD (HOEVD).** Let the third-order tensor \( \{\hat{a}_k\} \) of size \( K \)-networks \( \times N \)-genes \( \times N \)-genes tabulate a series of \( K \) genome-scale networks computed from a series of \( K \) genome-scale signals \( \{\hat{e}_k\} \), of size \( N \)-genes \( \times M_2 \)-arrays each, such that \( \hat{a}_k = \hat{e}_k \hat{b}^\dagger \) for all \( k = 1, 2, \ldots, K \). We define and compute a HOEVD of the tensor of networks \( \{\hat{a}_k\} \),

\[
\hat{a} = \sum_{k=1}^{K} \hat{a}_k = \hat{u} \left( \sum_{k=1}^{K} \bar{e}_k \right)^T = \hat{u} \hat{e} \hat{b}^T,
\]

using the SVD of the appended signals \( \hat{e} = (\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_K) = \hat{u} \hat{e} \hat{b}^T \), where the \( m \)th column of \( \hat{u}, [\alpha_{k,m}] = [\bar{u}_2(m) \bar{m}] \), lists the genome-scale expression of the \( m \)th eigenarray of \( \hat{e}_k \). Whereas the EVD is equivalent to the matrix SVD for a symmetric nonnegative matrix, this tensor HOEVD is different from the tensor higher-order SVD (14–16) for the series of symmetric nonnegative matrices \( \{\hat{a}_k\} \), where the higher-order SVD is computed from the SVD of the appended networks \( \{\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_K\} \) rather than the appended signals. This HOEVD formulates the overall network computed from the appended signals \( \hat{a} = \hat{e} \hat{b}^T \) as a linear superposition of a series of \( M = \sum_{k=1}^{K} M_k \) rank-1 symmetric “subnetworks” that are decorrelated of each other, \( \hat{a} = \sum_{k=1}^{K} \hat{e}_k \hat{b}_k^\dagger \), where \( \{\hat{e}_k\} \) are symmetric but not diagonal, such that \( \langle [\hat{e}_k] \hat{b}_k^\dagger [\hat{b}_k \hat{e}_k] = (m | \hat{e}_k | m) \rangle \neq 0 \). The significance of the \( m \)th subnetwork in the \( k \)th network is indicated by the \( m \)th fraction of eigenexpression of the \( k \)th network \( \hat{p}_{k,m} = e_{k,m}^2 / (\sum_{m=1}^{M_2} e_{k,m}^2) \), where, for all \( m \leq M \) in all \( K \) networks, \( \sum_{m=1}^{M_2} e_{k,m}^2 = 1 \). For \( k = 1 \), this expression correlation captured by the \( m \)th subnetwork in the \( k \)th network relative to that captured by all subnetworks (and all couplings among them), where \( \sum_{m=1}^{M_2} e_{k,m}^2 = 1 \) indicates the significance of the coupling between the \( l \)th and \( m \)th subnetworks in the \( k \)th network. The sign of this fraction indicates the direction of the coupling, such that \( p_{2,1} > 0 \) corresponds to a transition from the \( l \)th to the \( m \)th subnetwork and \( p_{1,2} < 0 \) corresponds to the transition from the \( m \)th to the \( l \)th.
For real signals \( \hat{e}_k \), the subnetworks are unique, and the
couplings among them are unique up to phase factors of \( e^{i\theta} \),
except in degenerate subspaces of \( \hat{e} \).

Interpretation of the Subnetworks and Their Couplings. We parallel-
and antiparallel-associate each subnetwork or coupling with most
likely expression correlations, or none thereof, according to the
annotations of the two groups of \( x \) pairs of genes each, with largest
and smallest levels of correlations in this subnetwork or coupling
among all \( X = N(N-1)/2 \) pairs of genes, respectively. The \( P \) value
of a given association by annotation is calculated by using combi-
natorics and assuming hypergeometric probability distribution of
the \( Y \) pairs of annotations among the \( X \) pairs of genes, and of
the subset of \( y \subseteq Y \) pairs of annotations among the subset of \( x \subseteq X \)
pairs of genes, \( P(x; y, X, Y) = \binom{Y}{y} / \binom{X}{x} \). where \( \binom{X}{y} = X!/(y!(X-y)) \)
is the binomial coefficient (17). The most likely association
of a subnetwork with a pathway or of a coupling between two
subnetworks with a transition between two pathways is that
which corresponds to the smallest \( P \) value. Independently, we
also parallel- and antiparallel-associate each eigenarray with
most likely cellular states, or none thereof, assuming hypergeo-
metric distribution of the annotations among the \( N \)-genes and
the subsets of \( n \subseteq N \) genes with largest and smallest levels of
expression in this eigenarray. The corresponding eigengene
might be inferred to represent the corresponding biological
process from its pattern of expression.

For visualization, we set the \( x \) correlations among the \( X \) pairs
of genes largest in amplitude in each subnetwork and coupling
equal to \( \pm 1 \), i.e., correlated or anticorrelated, respectively,
according to their signs. The remaining correlations are set equal
to 0, i.e., decorrelated. We compare the discretized subnetworks
and couplings using Boolean functions (6).

Biological Results: Yeast Pathways from mRNA Expression and
Proteins’ DNA-Binding Signals

Significant EVD Subnetworks Are Associated with Functionally
Independent Pathways. We compute the network \( \hat{a}_1 \) from the data
signal \( \hat{e}_1 \), which tabulates relative mRNA expression levels of \( n =
4,153 \) yeast genes with valid data in at least 15 of the \( M = 18 \)
samples of a cell cycle time course of a culture synchronized by
the mating pheromone \( \alpha \) factor (1). The relative expression level
of the \( n \)th gene in the \( m \)th sample is presumed valid when the

![Fig. 1](image-url) Discretized significant EVD subnetworks of the network \( \hat{a}_1 \) in the subsets of 150 correlations (red) and anticorrelations (green) largest in amplitude among all traditionally classified cell-cycle genes of \( \hat{a}_1 \), color-coded according to their cell-cycle classifications, M/G1 (yellow), G1 (green), S (blue), S/G2 (red), and G2/M (orange), and separately also according to their pheromone-response classifications, up-regulated (black) and down-regulated (gray). (a) The first subnetwork shows pheromone-response-dependent and cell-cycle-independent relations among the genes. (b) The second subnetwork shows pheromone response- and cell-cycle-dependent relations. (c and d) The third and fourth subnetworks show cell-cycle-dependent relations that are orthogonal to each other.
ratio of the measured expression to the background signal is >1.5 for both the synchronized culture and asynchronous reference. Before computing \( \hat{a}_1 \), we use SVD to estimate the missing data in \( \hat{e}_1 \) (10, 18) and to approximately center the expression pattern of each gene in \( \hat{e}_1 \) at its time-invariant level (Supporting Appendix).

EVD of the network \( \hat{a}_1 \) uncovers four significant subnetworks, which capture >60%, 10%, 5%, and 5%, respectively, of the expression correlation of \( \hat{a}_1 \). These subnetworks are associated with the independent pathways manifest in the data signal \( \hat{e}_1 \), following the P values for the distribution of the \( Y = 1,035 \) pairs of the 46 genes that were microarray-classified as pheromone-regulated (2) among all \( X = 2,926 \) pairs of the 77 genes that were traditionally classified as cell-cycle-regulated (1), and among each of the subsets of \( x = 150 \) pairs of genes with largest and smallest levels, respectively, of expression correlation (Table 2 in Supporting Appendix). The associations of the EVD subnetworks of \( \hat{a}_1 \) are consistent with those of the corresponding SVD eigenvectors of \( \hat{e}_1 \) following the P values for the distribution of the 284 pheromone-regulated genes and that of the 574 genes, which were traditionally or microarray-classified as cell-cycle-regulated, among all \( X = 4,153 \) genes and among each of the subsets of \( x = 150 \) genes with largest and smallest levels, respectively, of expression (Table 1 in Supporting Appendix). The associations of the EVD subnetworks of \( \hat{a}_1 \) are also consistent with the patterns of expressions of the corresponding SVD eigengenes of \( \hat{e}_1 \) (Fig. 8 in Supporting Appendix). We visualize the discretized four subnetworks and their Boolean functions in the subset of 70 genes that constitute the \( x = 150 \) correlations in each subnetwork that are largest in amplitude among the \( X = 2,926 \) pairs of traditionally classified cell-cycle-regulated genes.

The first and most significant subnetwork is associated with the \( \alpha \) factor signal-transduction pathway, where the relations among the genes depend only on their pheromone-response classifications. Genes that are up-regulated in response to pheromone, and separately also genes that are down-regulated, are correlated, even when these genes are classified into antipodal cell-cycle stages. Genes that are up-regulated in response to pheromone are anticorrelated with genes that are down-regulated, even when these genes are classified into the same cell-cycle stages. For example, \( KAR4 \), which is up-regulated in response to pheromone, is correlated with \( CIK1 \), which is also up-regulated, and anticorrelated with \( CLN2 \), which is down-regulated (Fig. 1a), even though the expression of both \( KAR4 \) and \( CLN2 \) peaks at the cell-cycle stage G1 while the expression of \( CIK1 \) peaks at the antipodal stage S/G2. In the second subnetwork, which is associated with the exit from the \( \alpha \) factor-induced cell-cycle arrest in \( M/G1 \) and the entry into cell-cycle progression at \( G1 \), genes that are up-regulated in response to pheromone are correlated, independent of their cell-cycle classification. The relations among genes that are down-regulated, however, depend on their cell-cycle, rather than their pheromone-response, classification. For example, \( CLN2 \) and \( CLB2 \), which encode cyclins of the antipodal stages \( G1 \) and \( G2/M \), respectively, are anticorrelated, even though both are down-regulated in response to pheromone; and \( SWI4 \), which encodes a G1 transcription factor, is correlated with \( CLN2 \) and anticorrelated with \( CLB2 \) (Fig. 1b). In the third and fourth subnetworks, which are associated with the two pathways of antipodal cell-cycle-expression oscillations that are orthogonal, i.e., \( \pi/2 \) out of phase relative to one another, the relations among genes depend only on their cell-cycle classifications. For example, in the third subnetwork, which is associated with the cell-cycle-expression oscillations at S vs. those at M, \( KAR4 \) is anticorrelated with \( CIK1 \), where \( KAR4 \) is correlated, and \( CIK1 \) is anticorrelated with \( ASH1 \) (Fig. 1c). In the fourth subnetwork, which is associated with expression at \( G1 \) vs. that at \( G2 \), \( KAR4 \) is correlated with \( CLN2 \) (Fig. 1d).

Boolean functions of the discretized subnetworks highlight known pathway-dependent relations among genes, common to a subset of the subnetworks or antipodal across the subnetworks (Fig. 9 in Supporting Appendix).

**Integrative Pseudoinverse-Projected Networks Simulate Observation of only the Pathways Manifest in both the Data and Basis Signals.** We compute the network \( \hat{a}_2 \) by pseudoinverse-projecting the network \( \hat{a}_1 \) onto the basis signal, which tabulates the relative DNA-bound protein occupancy levels of the 2,120 genes with at least one valid data point in any one of \( L = 12 \) samples that correspond to 12 yeast-cell-cycle transcription factors (3). The relative binding occupancy level of the \( n \)th gene in the \( l \)th sample is presumed valid when the associated P value is <0.1. Similarly, \( \hat{a}_3 \) is computed by projecting \( \hat{a}_1 \) onto the basis signal, which tabulates the occupancy levels of 2,476 genes in 12 samples of transcription factors involved in developmental programs, such as mating; and \( \hat{a}_4 \) is computed by projecting \( \hat{a}_1 \) onto the basis signal, which tabulates the occupancy levels of 2,943 genes in eight samples of factors involved in biosynthesis, such as DNA replication. Before computing \( \hat{a}_2, \hat{a}_3, \) and \( \hat{a}_4 \) for the 1,588, 1,827, and 2,254 genes at the intersections of \( \hat{a}_1 \) and the proteins’ DNA-binding basis signals, we divide each gene measurement in each basis signal by the arithmetic mean of the measurements for that gene in that signal, thus converting the

---

**Fig. 2.** Boolean AND intersections of the discretized EVD subnetworks of the pseudoinverse-projected \( \hat{a}_2 \) and \( \hat{a}_3 \) in the subsets of 200 correlations largest in amplitude among all traditionally classified cell-cycle genes of \( \hat{a}_2 \) and \( \hat{a}_3 \), respectively, with those of \( \hat{a}_1 \). (a) The first subnetwork of \( \hat{a}_2 \) AND \( \hat{a}_3 \). (b) The second subnetwork of \( \hat{a}_2 \) AND third subnetwork of \( \hat{a}_1 \). (c) The subnetwork of \( \hat{a}_2 \) AND fourth subnetwork of \( \hat{a}_1 \).
signals to DNA-binding levels of each transcription factor relative to those of all other factors. We also approximately center the binding pattern of each gene at its transcription factor-invariant level using SVD (Supporting Appendix).

EVD of the cell-cycle-projected network $\hat{a}_2$ uncovers only two significant subnetworks, which capture $\approx 55\%$ and $30\%$ of the expression correlation of $\hat{a}_2$, respectively, and are associated with the two pathways of antipodal cell-cycle-expression oscillations at G1 vs. those at G2, and at S vs. M, respectively [Table 4 (row a) in Supporting Appendix]. Boolean AND intersection of the discretized first subnetwork of $\hat{a}_2$, in the subset of 200 correlations largest in amplitude among all traditionally classified cell-cycle genes of $\hat{a}_2$, with the discretized fourth subnetwork of $\hat{a}_1$ highlights correlations among traditionally classified M/G1, G1, and S genes, and anticorrelations among these genes and G2/M genes, independent of their responses to pheromone (Fig. 2a). Boolean AND of the second

![Fig. 3. Discretized significant HOEVD subnetworks of the series of networks ($\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$) and their couplings, in the subsets of 100 correlations largest in amplitude among all traditionally classified cell-cycle genes of ($\hat{a}_1$, $\hat{a}_2$, $\hat{a}_3$). (a) The first subnetwork shows pheromone-response-dependent only relations among the genes. (b and c) The second and third subnetworks show orthogonal cell-cycle-dependent relations. (d and e) The couplings between the first and second, and first and third subnetworks, respectively, both show pheromone-response- and cell-cycle-dependent relations. (f) The coupling between the second and third subnetworks shows cell-cycle-dependent only relations.]

![Fig. 4. Fractions of eigenexpression of the HOEVD subnetworks (a) and their couplings (b) in the individual networks $\hat{a}_1$ (red), $\hat{a}_2$ (blue), and $\hat{a}_3$ (green). The contributions of each coupling in each individual network cancel out in the overall network $\hat{a} = \hat{a}_1 + \hat{a}_2 + \hat{a}_3$.]

Alter and Golub PNAS December 6, 2005 vol. 102 no. 49 17563
APPLIED MATHEMATICS GENETICS
subnetwork of $\hat{a}_2$ with the third subnetwork of $\hat{a}_1$ highlights correlations among M/G1 genes and their anticorrelations with S and S/G2 genes (Fig. 2b). The $\alpha$ factor signal-transduction pathway that is manifest in the data but not in the basis signal is not associated with either one of the subnetworks of $\hat{a}_2$. Similarly, EVD of the development-projected network $\hat{d}_3$ uncovers only one significant subnetwork, which captures $\geq90\%$ of the expression correlation of $\hat{d}_3$ and is associated with the $\alpha$ factor signal-transduction pathway [Table 4 (row b) in Supporting Appendix]. Boolean AND of the subnetwork of $\hat{d}_3$ with the first subnetwork of $\hat{d}_1$ highlights correlations among genes that are up-regulated in response to pheromone and their anticorrelations with down-regulated genes, independent of their cell-cycle classifications (Fig. 2c). The cell-cycle-expression oscillation pathways that are manifest in the data but not in the basis signal are not associated with either one of the subnetworks of $\hat{d}_3$. EVD of the biosynthesis-projected network $\hat{a}_4$ uncovers three significant subnetworks, which capture together $\geq90\%$ of the expression correlation of $\hat{a}_4$, all of which are associated with the activity of histones that peaks during DNA replication at the cell-cycle stage S [Table 4 (row c) and Fig. 13 in Supporting Appendix].

The associations of the EVD subnetworks of the projected networks $\hat{a}_2$, $\hat{a}_3$, and $\hat{a}_4$ are consistent with the associations of the corresponding SVD eigenarrays (Table 3 in Supporting Appendix) and eigefuncs (Figs. 10–12 in Supporting Appendix) of the projected signals $\hat{e}_2$, $\hat{e}_3$, and $\hat{e}_4$, respectively.

**Comparative HOEVD Subnetworks and Their Couplings Are Associated with Pathways and the Transitions Among Them Common to the Series or Exclusive to a Subset of Networks.** HOEVD of the series of networks $\{\hat{a}_1, \hat{a}_2, \hat{a}_3, \hat{a}_4\}$ uncovers three significant subnetworks, which capture $\approx40\%$, $15\%$, and $9\%$ of the expression correlation of the overall network $\hat{a} = \hat{a}_1 + \hat{a}_2 + \hat{a}_3$, respectively, and the three couplings among these subnetworks, which capture expression correlations only in the individual networks. The associations of the HOEVD subnetworks and couplings of $\{\hat{a}_1, \hat{a}_2, \hat{a}_3\}$ (Table 6 in Supporting Appendix) are consistent with the associations of the corresponding SVD eigenarrays (Table 5 in Supporting Appendix) and eigefuncs (Fig. 14 in Supporting Appendix) of the appended signals $\hat{e} = (\hat{e}_1, \hat{e}_2, \hat{e}_3)$, computed for the 868 genes at the intersection of $\hat{e}_1$, $\hat{e}_2$, and $\hat{e}_3$.

The subnetworks are associated with the independent pathways that are manifest in the overall network as well as the individual networks. The first subnetwork, which is associated with the $\alpha$ factor signal-transduction pathway (Fig. 3a), contributes to the expression correlations of the network $\hat{a}_1$ as well as to the development-projected network $\hat{a}_2$, but its contribution to the cell-cycle-projected network $\hat{a}_3$ is negligible (Fig. 4a). The second and third subnetworks, which are associated with the two pathways of antipodal cell-cycle-expression oscillations at G1 vs. that at G2 and at S vs. that at M, respectively (Fig. 3 b and c), contribute to $\hat{a}_1$ and $\hat{a}_2$ but not to $\hat{a}_3$. The couplings are associated with the transitions among these independent pathways that are manifest in the individual networks only. The coupling between the first and second subnetworks is associated with the transition between the two pathways of response to pheromone and cell-cycle expression at G1 vs. that at G2, i.e., the exit from pheromone-induced arrest and entry into cell-cycle progression (Fig. 3d). The coupling between the first and third subnetworks is associated with cell-cycle-expression oscillations at the two antipodal cell-cycle checkpoints of G1/S vs. G2/M (Fig. 3f). All these couplings contribute to the expression correlation of $\hat{a}_2$. Their contributions to the expression correlations of $\hat{a}_1$ and $\hat{a}_3$ are negligible (Fig. 4b).

Boolean functions of the discretized subnetworks and couplings highlight known as well as previously unknown pathway-dependent relations among genes that are in agreement with current understanding of the cellular system of yeast (Fig. 15 in Supporting Appendix) (19).

**Discussion**

We have shown that the matrix EVD and pseudoinverse projection and a tensor HOEVD can separate genome-scale nondirectional networks of, e.g., mRNA expression and proteins’ DNA-binding relations among genes into mathematically defined subnetworks and their couplings that can be associated with functionally independent pathways and the transitions among them. In analyses of genome-scale yeast networks, these subnetworks and couplings uncover coordinated differential relations among cell-cycle- and pheromone-regulated genes that are in agreement with reported pathway-dependent activities of these genes. Possible additional applications of EVD, pseudoinverse projection, and HOEVD include reconstruction of pathways and transitions among these pathways from nondirectional networks of correlations among sets of orthologous genes, which are computed from genome-scale signals of different types and from different organisms to elucidate organism, as well as pathway, dependence of relations among genes (e.g., refs. 6, 11, 20, and 21).

We thank T. G. Kolda and T. O. Yeates for thoughtful reviews of this manuscript; J. F. X. Diffley, V. R. Iyer, E. M. Marcotte, and B. K. Tye for helpful comments; and the American Institute of Mathematics in Palo Alto for hosting the 2004 Workshop on Tensor Decompositions where some of this work was done. This work was supported by National Science Foundation Grant CCR-0430617 (to G.H.G.) and National Human Genome Research Institute Individual Mentored Research Scientist Development Award in Genomic Research and Analysis 5K01 HG00038 (to O.A.).

Fig. 5. The eigenvalue decomposition (EVD) [13] of the symmetric yeast network $\hat{a}_1$ of 4,153-genes × 4,153-genes correlations [4–6], computed from the genome-scale mRNA expression data signal $\hat{e}_1$ of 4,153-genes × 18-samples of a cell cycle time course of an $\alpha$ factor-synchronized culture [1] (Mathematica Notebook 1 and Data Set 1). Raster display of $\hat{a}_1 \approx \sum_{m=1}^{4} \epsilon_{1,m}^2 |\alpha_{1,m}\rangle\langle \alpha_{1,m}|$, with correlation (red), decorrelation (black), and anticorrelation (green) in expression, visualizing the network as an approximate superposition of only its four most significant EVD subnetworks, in the subset of 70 genes which constitute the 150 correlations in each subnetwork that are largest in amplitude among the 2,926 correlations of 77 traditionally-classified cell cycle-regulated genes [1] (Data Set 2). The subnetworks are associated with the functionally independent pathways that are manifest in the signal $\hat{e}_1$. The first and most significant subnetwork is associated with the $\alpha$ factor signal transduction pathway. The second subnetwork is associated with the exit from the $\alpha$ factor-induced cell cycle arrest in the cell cycle stage M/G$_1$ and the entry into the cell cycle stage G$_1$. The third and fourth subnetworks, which are of similar significance, are associated with the two pathways of antipodal cell cycle expression oscillations, which are orthogonal, i.e., $\pi/2$ out of phase relative to one another, at the cell cycle stage S vs. those at M, and at G$_1$ vs. G$_2$, respectively.

Fig. 6. The pseudoinverse projection [12, 13] of the network $\hat{a}_1$ onto a genome-scale proteins’ DNA-binding basis signal of 2,120-genes × 12-samples of cell cycle transcription factors [3] (Mathematica Notebook 2 and Data Set 3), computed for the 1,588 genes at the intersection of $\hat{a}_1$ and the basis signal. Raster display of the pseudoinverse-projected network $\hat{a}_2 \equiv (\hat{b}\hat{b}^\dagger)\hat{a}_1(\hat{b}\hat{b}^\dagger) \approx \sum_{m=1}^{2} \epsilon_{2,m}^2 |\alpha_{2,m}\rangle\langle \alpha_{2,m}|$, visualizing this network as an approximate superposition of only its two most significant EVD subnetworks, in the subset of 39 genes which constitute the 200 correlations in each subnetwork that are largest in amplitude among the 1,128 correlations of 48 traditionally-classified cell cycle-regulated genes. The subnetworks are associated with the functionally independent pathways that are manifest in both the data and basis signals. The two most significant subnetworks are associated with the two pathways of antipodal cell cycle expression oscillations, which are orthogonal, i.e., $\pi/2$ out of phase relative to one another, at the cell cycle stage G$_1$ vs. those at G$_2$, and at S vs. M, respectively. The $\alpha$ factor signal transduction pathway and the transition from the $\alpha$ factor-induced cell cycle arrest into the cell cycle progression, that are manifest in the data but not in the basis signal, are not associated with either one of the significant subnetworks of $\hat{a}_2$. 
Fig. 7. A higher-order EVD (HOEVD) of the third-order series of the three networks \( \{ \hat{a}_1, \hat{a}_2, \hat{a}_3 \} \). The network \( \hat{a}_3 \) is the pseudoinverse projection of the network \( \hat{a}_1 \) onto a genome-scale proteins’ DNA-binding basis signal of 2,476-genes × 12-samples of development transcription factors [3] (Mathematica Notebook 3 and Data Set 4), computed for the 1,827 genes at the intersection of \( \hat{a}_1 \) and the basis signal. The HOEVD is computed for the 868 genes at the intersection of \( \hat{a}_1, \hat{a}_2 \) and \( \hat{a}_3 \). Raster display of \( \hat{a}_2 \approx \sum_{m=1}^{3} \epsilon_{k,m}^{2} |\alpha_{m}\rangle \langle |\alpha_{m}| + \sum_{m=1}^{3} \sum_{l=m+1}^{3} \epsilon_{k,l}^{2} |\alpha_{l}\rangle \langle |\alpha_{l}| + |\alpha_{m}\rangle \langle |\alpha_{m}| \rangle \), for all \( k = 1, 2, 3 \), visualizing each of the three networks as an approximate superposition of only the three most significant HOEVD subnetworks and the three couplings among them, in the subset of 26 genes which constitute the 100 correlations in each subnetwork and coupling that are largest in amplitude among the 435 correlations of 30 traditionally-classified cell cycle-regulated genes. This tensor HOEVD is different from the tensor higher-order SVD [14–16] for the series of symmetric nonnegative matrices \( \{ \hat{a}_1, \hat{a}_2, \hat{a}_3 \} \). The subnetworks correlate with the genomic pathways that are manifest in the series of networks. The most significant subnetwork correlates with the response to the pheromone. This subnetwork does not contribute to the expression correlations of the cell cycle-projected network \( \hat{a}_2 \), where \( \epsilon_{2,1}^2 \approx 0 \). The second and third subnetworks correlate with the two pathways of antipodal cell cycle expression oscillations, at the cell cycle stage \( G_1 \) vs. those at \( G_2 \), and at \( S \) vs. \( M \), respectively. These subnetworks do not contribute to the expression correlations of the development-projected network \( \hat{a}_3 \), where \( \epsilon_{3,2}^2 \approx \epsilon_{3,3}^2 \approx 0 \). The couplings correlate with the transitions among these independent pathways that are manifest in the individual networks only. The coupling between the first and second subnetworks is associated with the transition between the two pathways of response to pheromone and cell cycle expression oscillations at \( G_1 \) vs. those \( G_2 \), i.e., the exit from pheromone-induced arrest and entry into cell cycle progression. The coupling between the first and third subnetworks is associated with the transition between the response to pheromone and cell cycle expression oscillations at \( S \) vs. those at \( M \), i.e., cell cycle expression oscillations at \( G_1/S \) vs. those at \( G_2/M \). The coupling between the second and third subnetworks is associated with the transition between the orthogonal cell cycle expression oscillations at \( G_1 \) vs. those at \( G_2 \) and at \( S \) vs. \( M \), i.e., cell cycle expression oscillations at the two antipodal cell cycle checkpoints of \( G_1/S \) vs. \( G_2/M \). All these couplings add to the expression correlation of the cell cycle-projected \( \hat{a}_2 \), where \( \epsilon_{2,12}^2, \epsilon_{2,13}^2, \epsilon_{2,23}^2 > 0 \); their contributions to the expression correlations of \( \hat{a}_1 \) and the development-projected \( \hat{a}_3 \) are negligible (see also Fig. 4).

**Significant EVD Subnetworks are Associated with Functionally Independent Pathways.** The data signal \( \epsilon_1 \) we analyze tabulates relative mRNA expression levels of \( N = 4,153 \) yeast genes with valid data in at least 15 of the \( M_1 = 18 \) samples of a cell cycle time course of an \( \alpha \) factor-synchronized culture monitored by Spellman et al. [1]. The relative expression level of the \( n \)th gene in the \( m \)th sample is presumed valid when the
The ratio of the measured expression to the background signal is > 1.5 for both the synchronized culture and asynchronous reference (Data Set 1). We use SVD to estimate these missing data [10] (and see also [18]) and to approximately center the expression pattern of each gene at its sample-, i.e. time-invariant level [9] (Mathematica Notebook 1).

For the nth gene $|g_{1,n}|$ with missing data in $M_1' < M_1$ of the arrays, we estimate the missing expression level in the nth array $(m|g_{1,n}|)$ to be a superposition of the $L' < M_1$ significant eigengenes $\{\gamma_{1,i}\}$ in the nth array as computed for the subset of $N' < N$ genes with no missing data in any of the $M_1$ arrays. The coefficients of this superposition are determined by the expansion of the expression of the nth gene across all $M_1 - M_1'$ arrays with no missing data, $(g_{1,n})_{M_1'}$, in the subspace spanned by the significant eigengenes across the same $M_1 - M_1'$ arrays, $\{\gamma_{1,i,M_1'}\}$, such that $(m|g_{1,n}|) \rightarrow \sum_{i=1}^{L'}(m|\gamma_{1,i}|)M_1';(\beta_{1,i})_{M_1'}$, where $\{\beta_{1,i}\}_{M_1'}$ span the $L' \times (M - M_1')$ subspace $(\beta_{1,i,M_1'})^\top$ that is pseudoinverse to the $(M - M_1') \times L'$ subspace $\beta_{1,i,M_1'}$, which is spanned by $\{\gamma_{1,i,M_1'}\}$.  

Fig. 8. Eigengenes of the data signal $\hat{\gamma}_1$ as computed for the 4,153 genes after missing data estimation and approximate centering.  (a) Raster display.  (b) Bar chart of the fractions of eigenexpressions.  (c-f) Line-joined graphs of the first (red), second (red), third (blue) and fourth (green) eigengenes, respectively, describe expression patterns across the arrays that are consistent with the associations of the corresponding subnetworks. The first eigengene describes an initial transient increase in expression superimposed on time-invariant expression (as well as the antiparallel pattern of initial decrease in expression). The second eigengene describes an initial transient increase superimposed on periodic expression oscillations that fit a dashed graph of a normalized sine of two periods (red). The third and fourth eigengenes describe periodic expression oscillations that fit dashed graphs of normalized cosine (blue) and sine (green) functions, respectively, of two periods.
We use the 5 most significant eigengenes as computed for the subset of 2,493 genes with no missing data in the 18 arrays in order to estimate the missing data in the remaining 1,660 genes. We find that these eigengenes and corresponding fractions of eigenexpression are similar to those computed for the full set of 4,153 genes after the missing data are estimated suggesting that the five most significant eigengenes, as computed for the 2,493 genes with no missing data, are meaningful patterns for estimating the missing data in the data signal. This also illustrates the robustness of the significant SVD eigenexpression levels as well as eigengenes to perturbations in the genes that compose the data signal $\hat{e}_1$.

After missing data estimation the first eigengene, which captures $\approx$90% of the expression information and describes expression that is approximately invariant across the samples, i.e., in time, is inferred to represent steady-state expression. We filter out this eigengene and the corresponding eigenarray without eliminating genes or arrays from the data signal by setting the corresponding eigenexpression level in $\hat{e}_1$ to zero and reconstructing the data signal from $\hat{u}_1 \hat{e}_1 \hat{v}_1^T$. After filtering out the first eigengene, the expression pattern of each gene is approximately centered at its sample-, i.e. time-invariant level (Fig. 8).

<table>
<thead>
<tr>
<th>Classification</th>
<th>Eigenarray</th>
<th>Most likely parallel association</th>
<th>$P$ value of parallel association</th>
<th>Most likely antiparallel association</th>
<th>$P$ value of antiparallel association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Cycle</td>
<td>1</td>
<td>M/G$_1$</td>
<td>$9.3 \times 10^{-10}$</td>
<td>G$_1$</td>
<td>$7.8 \times 10^{-18}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>G$_1$</td>
<td>$3.4 \times 10^{-55}$</td>
<td>G$_2$/M</td>
<td>$8.2 \times 10^{-30}$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>M/G$_1$</td>
<td>$3.9 \times 10^{-39}$</td>
<td>S/G$_2$</td>
<td>$1.9 \times 10^{-23}$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>G$_1$</td>
<td>$1.1 \times 10^{-69}$</td>
<td>G$_2$/M</td>
<td>$1.5 \times 10^{-33}$</td>
</tr>
<tr>
<td>Pheromone</td>
<td>1</td>
<td>Up</td>
<td>$3.4 \times 10^{-65}$</td>
<td>Down</td>
<td>$5.7 \times 10^{-34}$</td>
</tr>
<tr>
<td>Response</td>
<td>2</td>
<td>Down</td>
<td>$4.3 \times 10^{-12}$</td>
<td>Up</td>
<td>$1.5 \times 10^{-16}$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Down</td>
<td>$5.4 \times 10^{-9}$</td>
<td>Down</td>
<td>$8.9 \times 10^{-14}$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Down</td>
<td>$5.6 \times 10^{-29}$</td>
<td>Down</td>
<td>$4.2 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

Table 1. Most likely parallel and antiparallel associations of the significant SVD eigenarrays of the data signal $\hat{e}_1$ according to the traditional and microarray classifications of cell cycle-regulated [1] and pheromone-regulated [2] yeast genes, are consistent with the associations of the corresponding subnetworks.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Subnetwork</th>
<th>Most likely parallel association</th>
<th>$P$ value of parallel association</th>
<th>Most likely antiparallel association</th>
<th>$P$ value of antiparallel association</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell Cycle</td>
<td>1</td>
<td>S      S</td>
<td>$1.7 \times 10^{-22}$</td>
<td>M/G$_1$ S</td>
<td>$5.1 \times 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>G$_1$ G$_1$</td>
<td>$1.3 \times 10^{-29}$</td>
<td>G$_1$ G$_2$/M</td>
<td>$3.2 \times 10^{-11}$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>S      S</td>
<td>$2.1 \times 10^{-30}$</td>
<td>M/G$_1$ S</td>
<td>$2.6 \times 10^{-25}$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>G$_1$ S</td>
<td>$2.1 \times 10^{-28}$</td>
<td>G$_1$ G$_2$/M</td>
<td>$5.7 \times 10^{-24}$</td>
</tr>
<tr>
<td>Pheromone</td>
<td>1</td>
<td>Up     Up</td>
<td>$4.0 \times 10^{-63}$</td>
<td>Down     Up</td>
<td>$2.2 \times 10^{-30}$</td>
</tr>
<tr>
<td>Response</td>
<td>2</td>
<td>Down Down</td>
<td>$1.6 \times 10^{-11}$</td>
<td>Down     Up</td>
<td>$9.8 \times 10^{-17}$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Down Down</td>
<td>$6.2 \times 10^{-6}$</td>
<td>Down     Down</td>
<td>$1.6 \times 10^{-11}$</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Down Down</td>
<td>$8.0 \times 10^{-32}$</td>
<td>Down     Down</td>
<td>$2.5 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

Table 2. Most likely parallel and antiparallel associations of the significant EVD subnetworks of the network $\tilde{a}_1$ according to the traditional and microarray classifications of cell cycle- and pheromone-regulated yeast genes.
subnetwork. For example, KAR4 and a to their cell cycle classifications, M/G1 (yellow), G1 (green), S (blue), S/G2 (red) and G2/M (orange), and separately also their pheromone response classifications, up-regulated (black) and down-regulated (gray). (a) Intersection of the first AND second subnetworks of a highlights pheromone response-dependent correlations (red), such as that between KAR4 and CIK1 [8], as well as anticorrelations (green), such as that between KAR4 and CLN2. (b) Intersection of the first AND NOT second subnetworks of a highlights relations (blue) among genes that are antipodal in the pathway of pheromone signal transduction vs. the pathway of exit from pheromone-induced arrest and entry into cell cycle progression. Highlighted relations correspond to anticorrelations in the first subnetwork and correlations in the second subnetwork. For example, AGA1 and CLB2 are anticorrelated in the pheromone signal transduction pathway but are correlated in the pathway of entry into cell cycle progression at the cell cycle stage G1. Also included are the relations between the pheromone-response up-regulated KAR4, which cell cycle expression peaks at G1 and pheromone-response down-regulated CLB1 and CLB2, which cell cycle expression peaks at G2/M. (c) Intersection of the second
AND fourth subnetworks highlights correlations (red) among G1 genes and their anticorrelations (green) with CLB2 that encodes a G2/M cyclin. (d) Intersection of the second AND NOT fourth subnetworks highlights correlations in the second subnetwork that correspond to anticorrelations in the fourth (orange) and anticorrelations in the second that correspond to correlations in the fourth (blue). All relations that are antipodal in the pathway of exit from pheromone-induced arrest and entry into G1 vs. that of cell cycle expression oscillations at G1 vs. G2, are between the pheromone-response up-regulated KAR4 and down-regulated genes. KAR4 and CLB2, for example, are anticorrelated during the exit from the pheromone-induced arrest but correlated during the cell cycle stage G1. Similarly, KAR4 and CLB2 are correlated during the exit from the pheromone-induced arrest but anticorrelated during G1. (e) Intersection of the third AND fourth subnetworks shows very few relations that are common to the two orthogonal pathways of antipodal cell cycle expression oscillations, at S vs. M, and at G1 vs. G2, respectively. These common relations include correlations (red) among S genes and an anticorrelation (green) between MNN1 and CWPI, which encode glycoproteins that are active at the antipodal stages G1 and S/G2, respectively [19]. (f) Intersection of the third AND NOT fourth subnetworks shows very few anticorrelations in the third subnetwork that correspond to correlations in the fourth (blue). All relations that are antipodal in the pathway of cell cycle expression oscillations at G1 vs. G2 vs. the pathway of oscillations at S vs. M are between G1 and M/G1 genes, such as CLN2, and S stage histones.

**Integrative Pseudoinverse-Projected Networks Simulate Observation of Only the Pathways Manifest in Both the Data and Basis Signals.** The basis signals we analyze tabulate relative DNA-bound occupancy levels of (i) 2,120 genes in L = 12 samples of the 12 cell cycle transcription factors Ace2, Fkh1, Fkh2, Mbp1, Mcm1, Ndd1, Rme1, Skn7, Stb, Swi4, Swi5 and Swi6 (Data Set 3); (ii) 2,476 genes in 12 samples of the development transcription factors Ash1, Dlg1, Hms1, Ime4, MATa1, Mot3, Phd1, Rim101, Rim1, Sok2, Ste12 and Sum1 (Data Set 4); and (iii) 2,943 genes in eight samples of the biosynthesis factors Abf1, Dot6, Fhl1, Hir1, Hir2, Rap1, Reb1 and Rgm1 (Data Set 5) measured by Lee et al. [3]. We use SVD to approximately center the pattern of binding of each gene [9] (Mathematica Notebook 2).

The most significant eigengene of each one of the basis signals is approximately invariant across the samples, and is inferred to represent steady-state binding. We filter out these eigengenes and the corresponding eigenarrays without eliminating genes or arrays from either basis signal by setting the corresponding eigenbinding level in the SVD of each basis signal to zero. After filtering out the first eigengene, the binding pattern of each gene is approximately centered at its sample-, i.e. transcription factor-invariant level.

**Fig. 10.** Eigengenes of ê2, i.e., the data signal ê1 pseudoinverse-projected onto the cell cycle transcription factors’ DNA-binding basis signal. (a) Raster display. (b) Bar chart of the fractions of eigenexpressions. (c and d) Line-joined graphs of the first (red) and second (blue) eigengenes fit dashed graphs of normalized sine (red) and cosine (blue) functions, respectively, of two periods, and are consistent with the associations of the corresponding subnetworks with the two orthogonal pathways of cell cycle expression oscillations.
Fig. 11 (left). Eigengenes of \( \hat{e}_3 \), i.e., \( \hat{e}_1 \) pseudoinverse-projected onto the development basis signal. (a) Raster display. (b) Bar chart of the fractions of eigengen- expressions. (c) Line-joined graph of the first (red) eigengene describes an initial transient increase in expression superimposed on time-invariant expression, and is consistent with the association of the corresponding subnetwork with the pheromone signal transduction pathway.

Fig. 12 (below). Eigengenes of \( \hat{e}_4 \), i.e., \( \hat{e}_1 \) pseudoinverse-projected onto the biosynthesis basis signal. (a) Raster display. (b) Bar chart of the fractions of eigengen- expressions. (c–e) Line-joined graphs of the first (red), second (blue) and third (green) eigengenes fit dashed graphs of normalized sine (red) and cosine (blue, green) functions, respectively, of two periods, and are consistent with the associations of the corresponding subnetworks with the expression of histones during DNA replication at the cell cycle stage S.
Table 3. Most likely parallel and antiparallel associations of the significant SVD eigenarrays of the pseudoinverse-projected signals $\hat{e}_2$, $\hat{e}_3$, and $\hat{e}_4$, i.e., $\hat{e}_1$ pseudoinverse-projected onto the (a) cell cycle, (b) development and (c) biosynthesis basis signals, according to the traditional and microarray classifications of cell cycle- and pheromone-regulated yeast genes.

<table>
<thead>
<tr>
<th>Transcription Factors</th>
<th>Classification</th>
<th>Subnetwork</th>
<th>Most likely parallel association</th>
<th>$P$ value of parallel association</th>
<th>Most likely antiparallel association</th>
<th>$P$ value of antiparallel association</th>
</tr>
</thead>
<tbody>
<tr>
<td>a Cell Cycle</td>
<td>Cell Cycle</td>
<td>1</td>
<td>$G_1$</td>
<td>$9.4 \times 10^{-19}$</td>
<td>$G_2/M$</td>
<td>$1.4 \times 10^{-19}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>$M/G_1$</td>
<td>$2.2 \times 10^{-12}$</td>
<td>$G_2/M$</td>
<td>$1.5 \times 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td>Pheromone</td>
<td>1</td>
<td>Down</td>
<td>$1.0 \times 10^{-14}$</td>
<td>Up</td>
<td>$2.3 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Up</td>
<td>$2.3 \times 10^{-4}$</td>
<td>Down</td>
<td>$1.7 \times 10^{-17}$</td>
</tr>
<tr>
<td>b Development</td>
<td>Cell Cycle</td>
<td>1</td>
<td>$M/G_1$</td>
<td>$6.9 \times 10^{-3}$</td>
<td>None</td>
<td>$9.8 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>Pheromone</td>
<td>1</td>
<td>Up</td>
<td>$1.0 \times 10^{-13}$</td>
<td>None</td>
<td>$2.9 \times 10^{-3}$</td>
</tr>
<tr>
<td>c Biosynthesis</td>
<td>Cell Cycle</td>
<td>1</td>
<td>$G_1$</td>
<td>$9.0 \times 10^{-8}$</td>
<td>None</td>
<td>$1.9 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>$S$</td>
<td>$5.3 \times 10^{-4}$</td>
<td>None</td>
<td>$2.1 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>$G_2/M$</td>
<td>$1.9 \times 10^{-5}$</td>
<td>$G_1$</td>
<td>$3.0 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>Pheromone</td>
<td>1</td>
<td>Down</td>
<td>$2.6 \times 10^{-4}$</td>
<td>None</td>
<td>$1.2 \times 10^{-1}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2</td>
<td>Down</td>
<td>$7.6 \times 10^{-2}$</td>
<td>None</td>
<td>$7.6 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3</td>
<td>Down</td>
<td>$6.3 \times 10^{-5}$</td>
<td>None</td>
<td>$7.6 \times 10^{-2}$</td>
</tr>
</tbody>
</table>

Table 4. Most likely parallel and antiparallel associations of the significant EVD subnetworks of the pseudoinverse-projected networks $\hat{a}_2$, $\hat{a}_3$ and $\hat{a}_4$, i.e., $\hat{a}_1$ pseudoinverse-projected onto the (a) cell cycle, (b) development and (c) biosynthesis basis signals, according to the traditional and microarray classifications of cell cycle- and pheromone-regulated yeast genes.

Fig. 13. Boolean functions of the discretized significant EVD subnetworks of the pseudoinverse-projected network $\hat{a}_4$ in the subsets of 200 relations largest in amplitude among all traditionally-classified cell cycle genes of $\hat{a}_4$, also with the third subnetwork of $\hat{a}_1$. (a) Intersection of the first AND second AND third subnetworks of $\hat{a}_4$ highlights correlations among the histones and other genes that are involved in biosynthesis of nuclear components, such as MCDI that is involved in sister chromatid cohesion. (b) Intersection of the first AND second AND third subnetworks of $\hat{a}_4$ AND the third subnetwork of $\hat{a}_1$, which is associated with expression at S vs. M, highlights correlations among histones.
Comparative HOEVD Subnetworks and Their Couplings are Associated with Pathways and the Transitions Among Them Common to the Series or Exclusive to a Subset of the Networks.

Fig. 14. Eigengenes of the appended signals \( \hat{\epsilon} \equiv (\hat{\epsilon}_1, \hat{\epsilon}_2, \hat{\epsilon}_3) \) computed for the 868 genes at the intersection of \( \hat{\epsilon}_1 \), \( \hat{\epsilon}_2 \) and \( \hat{\epsilon}_3 \). (a) Raster display. (b) Bar chart of the fractions of eigenexpressions. (c) Line-joined graphs of the first (red), second (blue) and third (green) eigengenes, respectively, describe expression patterns across the arrays that are consistent with the associations of the corresponding subnetworks and their couplings with the independent pathways that are manifest in the overall network as well as the individual networks, and the transitions among these independent pathways that are manifest in the individual networks only. The first eigengene describes an initial transient increase in expression superimposed on time-invariant expression. The second and third eigengenes describe periodic expression oscillations that fit dashed graphs of normalized sine (blue) and cosine (green) functions, respectively, of two periods.
<table>
<thead>
<tr>
<th>Classification</th>
<th>Eigenarray</th>
<th>Most likely parallel association</th>
<th>$P$ value of parallel association</th>
<th>Most likely antiparallel association</th>
<th>$P$ value of antiparallel association</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cell Cycle</strong></td>
<td>1</td>
<td>M/G₈</td>
<td>$8.0 \times 10^{-7}$</td>
<td>G₅</td>
<td>$8.3 \times 10^{-20}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>G₅</td>
<td>$1.4 \times 10^{-36}$</td>
<td>G₂/M</td>
<td>$5.5 \times 10^{-18}$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>M/G₈</td>
<td>$3.1 \times 10^{-16}$</td>
<td>G₅</td>
<td>$4.8 \times 10^{-12}$</td>
</tr>
<tr>
<td><strong>Pheromone Response</strong></td>
<td>1</td>
<td>Up</td>
<td>$5.3 \times 10^{-18}$</td>
<td>Down</td>
<td>$5.5 \times 10^{-23}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Down</td>
<td>$8.0 \times 10^{-4}$</td>
<td>Down</td>
<td>$2.7 \times 10^{-3}$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Up</td>
<td>$1.8 \times 10^{-3}$</td>
<td>Down</td>
<td>$1.6 \times 10^{-9}$</td>
</tr>
</tbody>
</table>

Table 5. Most likely parallel and antiparallel associations of the significant SVD eigenarrays of the appended signals $\hat{e} \equiv (\hat{e}_1, \hat{e}_2, \hat{e}_3)$ according to the traditional and microarray classifications of cell cycle- and pheromone-regulated yeast genes.

<table>
<thead>
<tr>
<th>Classification</th>
<th>Subnetwork or Coupling Between Subnetworks</th>
<th>Most likely parallel association</th>
<th>$P$ value of parallel association</th>
<th>Most likely antiparallel association</th>
<th>$P$ value of antiparallel association</th>
</tr>
</thead>
<tbody>
<tr>
<td>a</td>
<td><strong>Cell Cycle</strong></td>
<td>M/G₈ M/G₈</td>
<td>$2.2 \times 10^{-5}$</td>
<td>M/G₈ S</td>
<td>$3.1 \times 10^{-5}$</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>G₅ G₅</td>
<td>$1.8 \times 10^{-7}$</td>
<td>G₅ G₂/M</td>
<td>$2.4 \times 10^{-8}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>G₁ S</td>
<td>$1.4 \times 10^{-6}$</td>
<td>M/G₈ S</td>
<td>$1.2 \times 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Down Down</td>
<td>$7.5 \times 10^{-16}$</td>
<td>Down Up</td>
<td>$2.0 \times 10^{-27}$</td>
</tr>
<tr>
<td></td>
<td><strong>Pheromone Response</strong></td>
<td>Down None</td>
<td>$1.6 \times 10^{-2}$</td>
<td>Down Up</td>
<td>$2.6 \times 10^{-2}$</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Down Down</td>
<td>$1.4 \times 10^{-2}$</td>
<td>Down Up</td>
<td>$2.1 \times 10^{-6}$</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>Up</td>
<td>$2.6 \times 10^{-10}$</td>
<td>Up</td>
<td>$3.6 \times 10^{-14}$</td>
</tr>
<tr>
<td>b</td>
<td><strong>Cell Cycle</strong></td>
<td>G₁ G₁</td>
<td>$1.8 \times 10^{-5}$</td>
<td>G₁ M/G₈</td>
<td>$6.2 \times 10^{-9}$</td>
</tr>
<tr>
<td></td>
<td>1 ↔ 2</td>
<td>G₁ S</td>
<td>$1.4 \times 10^{-6}$</td>
<td>M/G₁ S</td>
<td>$1.2 \times 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td>1 ↔ 3</td>
<td>G₁ S</td>
<td>$1.1 \times 10^{-5}$</td>
<td>G₁ G₂/M</td>
<td>$1.6 \times 10^{-7}$</td>
</tr>
<tr>
<td></td>
<td>2 ↔ 3</td>
<td>Down Down</td>
<td>$2.3 \times 10^{-10}$</td>
<td>Down Up</td>
<td>$2.7 \times 10^{-10}$</td>
</tr>
<tr>
<td></td>
<td><strong>Pheromone Response</strong></td>
<td>Down Down</td>
<td>$1.1 \times 10^{-7}$</td>
<td>Down Up</td>
<td>$3.6 \times 10^{-14}$</td>
</tr>
<tr>
<td></td>
<td>1 ↔ 2</td>
<td>Down Down</td>
<td>$1.6 \times 10^{-8}$</td>
<td>Down Up</td>
<td>$3.9 \times 10^{-5}$</td>
</tr>
</tbody>
</table>

Table 6. Most likely parallel and antiparallel associations of the significant HOEVD subnetworks of the series of the three networks $\{\hat{a}_1, \hat{a}_2, \hat{a}_3\}$ and their couplings, according to the traditional and microarray classifications of cell cycle- and pheromone-regulated yeast genes.
follow the same pathway-dependence as that of experimental conditions of Spellman et al. [1] (see also [2]), and that, e.g., the relation of a Fig. 1–3). Our analyses, therefore, predict that TIP1 is common to all subnetworks and couplings, and is not limited to those that represent cell cycle pathways (see also Figs. 1–3). Our analyses, therefore, predict that TIP1 is up-regulated in response to pheromone under the experimental conditions of Spellman et al. [1] (see also [2]), and that, e.g., the relation of TIP1 with CIK1 would follow the same pathway-dependence as that of KAR4 with CIK1. The glycoproteins encoding CWP1 and MNN1 are classified as pheromone-regulated, suggesting further that additional cell cycle regulated glycoproteins might also be regulated by pheromone. (a) Intersection of the first AND second subnetworks AND the coupling between them highlights correlations among G1 genes (red) and their anticorrelations (green) with S/G2, G2/M and M/G1 genes. (b) Intersection of the first AND second subnetworks AND NOT the coupling between them highlights a couple of

Fig. 15. Boolean functions of the discretized significant HOEVD subnetworks of the series of the three networks \( \{\tilde{a}_1, \tilde{a}_2, \tilde{a}_3\} \) and their couplings in the subsets of 100 relations largest in amplitude among all traditionally-classified cell cycle genes of the series highlight known as well as previously unknown pathway-dependent relations that are in agreement with current understanding of the cellular system of yeast. For example, TIP1 that encodes a G1 glycoprotein [19] is not reported to be regulated by pheromone. Yet, the correlation between KAR4 and TIP1 is common to all subnetworks and couplings, and is not limited to those that represent cell cycle pathways (see also Figs. 1–3). Our analyses, therefore, predict that TIP1 is up-regulated in response to pheromone under the experimental conditions of Spellman et al. [1] (see also [2]), and that, e.g., the relation of TIP1 with CIK1 would follow the same pathway-dependence as that of KAR4 with CIK1. The glycoproteins encoding CWP1 and MNN1 are classified as pheromone-regulated, suggesting further that additional cell cycle regulated glycoproteins might also be regulated by pheromone. (a) Intersection of the first AND second subnetworks AND the coupling between them highlights correlations among G1 genes (red) and their anticorrelations (green) with S/G2, G2/M and M/G1 genes. (b) Intersection of the first AND second subnetworks AND NOT the coupling between them highlights a couple of
relations (blue) among genes that correspond to anticorrelations in the pathway of pheromone signal transduction as well as the pathway of cell cycle expression oscillations at G\textsubscript{1} vs. those at G\textsubscript{2}, but correspond to correlations in the transition between these two pathways, i.e., the exit from pheromone-induced arrest and entry into cell cycle progression at the cell cycle stage G\textsubscript{1}. For example, CLB2 and TIP1 are classified into the antipodal cell cycle stages of G\textsubscript{2}/M and G\textsubscript{1}, respectively. While CLB2 is classified as pheromone-response down-regulated, TIP1 appears to be up-regulated in response to pheromone. These are consistent with their anticorrelations in the two pathways that the first and second subnetworks represent. During the transition between these two pathways, from pheromone-induced arrest to cell cycle progression, expression of both CLB2 and TIP1 is suppressed and therefore they are correlated in the transition between the pathways that the coupling between the first and second subnetworks represents. (c) Intersection of the first AND third subnetworks AND the coupling between them highlights correlations among G\textsubscript{1} and S genes, and also separately among M/G\textsubscript{1} genes (red) and anticorrelations among these two subsets of genes (green). (d) Intersection of the first AND third subnetworks AND NOT the coupling between them highlights a single relation (blue) among CLB2 and CIK1 that corresponds to an anticorrelation in the pathway of pheromone signal transduction as well as the pathway of cell cycle expression oscillations at S vs. those at M, but corresponds to a correlation in the transition between these two pathways, i.e., the exit from pheromone-induced arrest and entry into cell cycle progression at the cell cycle stage G\textsubscript{1}/S. While CLB2 is down-regulated by pheromone, CIK1 is up-regulated. While CLB2 encodes a G\textsubscript{2}/M cyclin, CIK1 peaks in expression at the stage S/G\textsubscript{2}. These are consistent with their anticorrelations in the two pathways that the first and third subnetworks represent. During the transition between these two pathways, from pheromone-induced arrest to cell cycle progression, expression of both CLB2 and CIK1 is suppressed and therefore they are correlated in the transition between the pathways that the coupling between the first and third subnetworks represents. (e) Intersection of the second AND third subnetworks AND the coupling between them highlights correlations among G\textsubscript{2}/M and M/G\textsubscript{1} genes (red) and their anticorrelations with MNN1 that encodes a G\textsubscript{1} glycoprotein (green).