Matrix and Tensor Computations For Reconstructing the Pathways of a Cellular System From Genome-Scale Signals

Orly Alter

**Department of Biomedical Engineering,
Institute for Cellular and Molecular Biology and
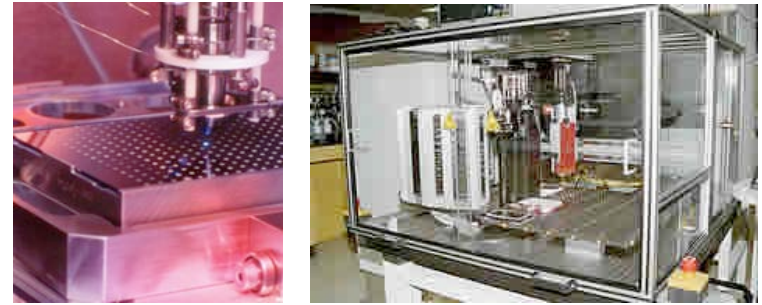Institute for Computational Engineering and Sciences**

University of Texas at Austin

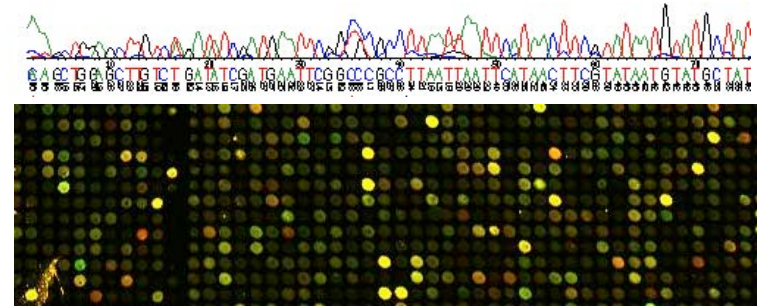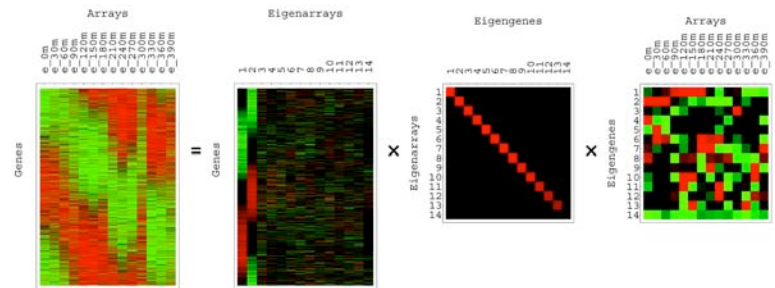|	**Astronomy**	**Molecular Biology**
Technology	**Galileo**	
Large-Scale Data	**Brahe**	
Mathematical Modeling	**Kepler**	
Basic Principles	**Newton**	
Technology	**NASA**	**Control of Cellular Mechanisms**
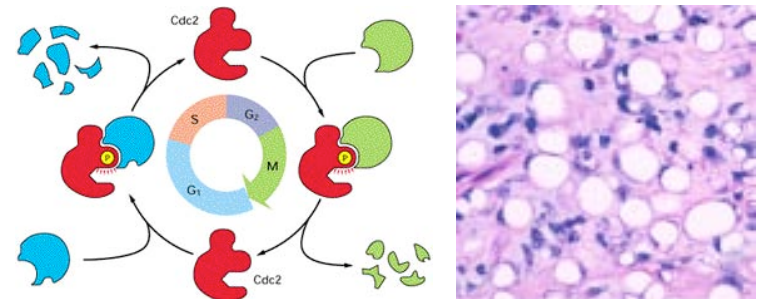
# DNA Microarrays Record Genomic Signals

DNA microarrays rely on hybridization to record the complete genomic signals that guide the progression of cellular processes, such as abundance levels of DNA, RNA and DNA-bound proteins on a genomic scale.
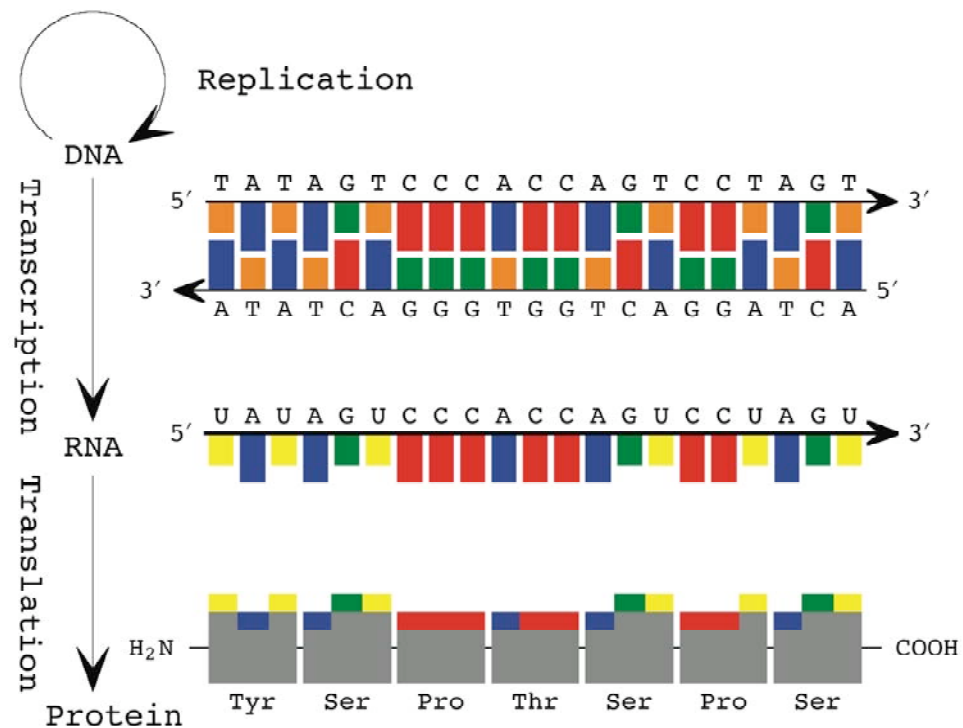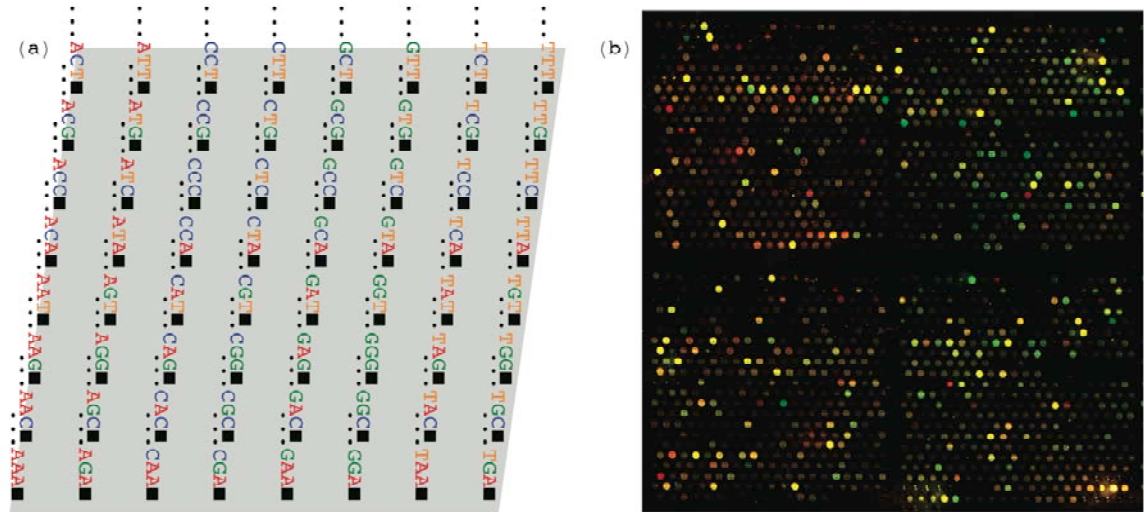
Matrix Models for Genomic Data

Mathematical frameworks for the description of the data, in which the mathematical variables and operations might represent biological reality.
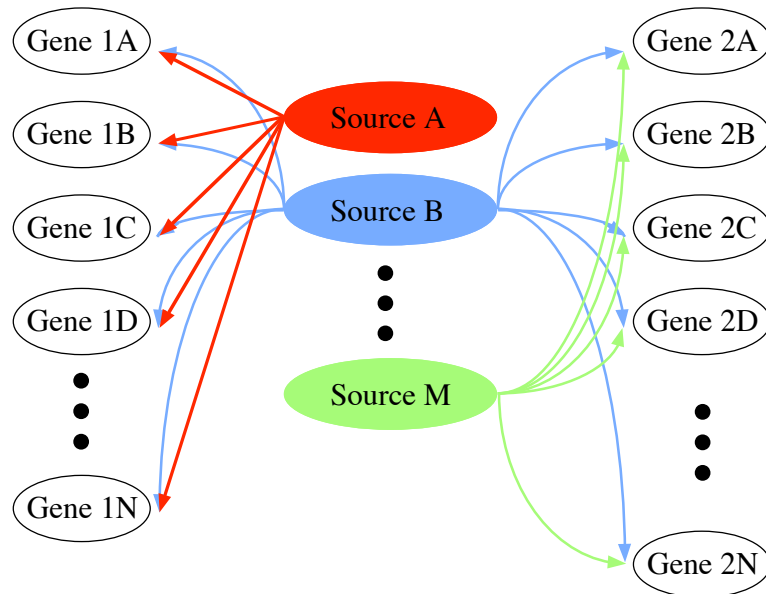
SVD Modeling

GSVD Comparative Modeling

Pseudoinverse Integrative Modeling
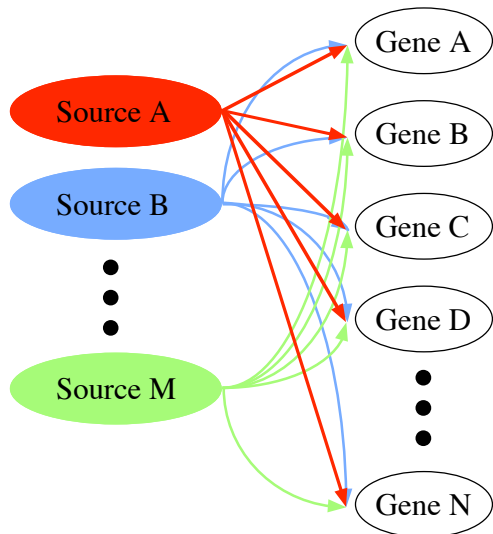
Uncover Cellular Processes and States

Uncover Processes Common or Exclusive Among Two Datasets

Predict a Biological Principle

Alter, Brown & Botstein, *PNAS* 2000;

Alter, Brown & Botstein, *PNAS* 2003;

Alter & Golub, *PNAS* 2004.

Matrix and Tensor Models for Networks of Correlations Computed from Genomic Data

Alter & Golub, *PNAS* <u>102</u>, 17559 (2005);
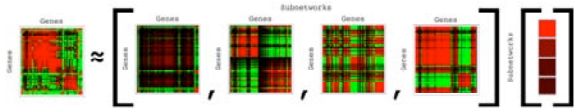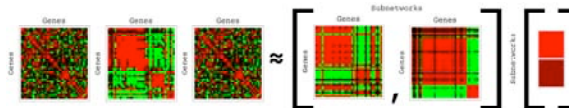http://www.bme.utexas.edu/research/orly/network_decomposition/.
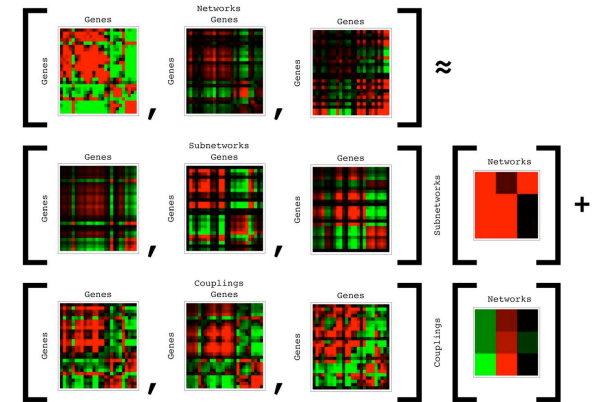
EVD Modeling

Uncover Pathways
in
a Single Network

Pseudoinverse Integrative Modeling
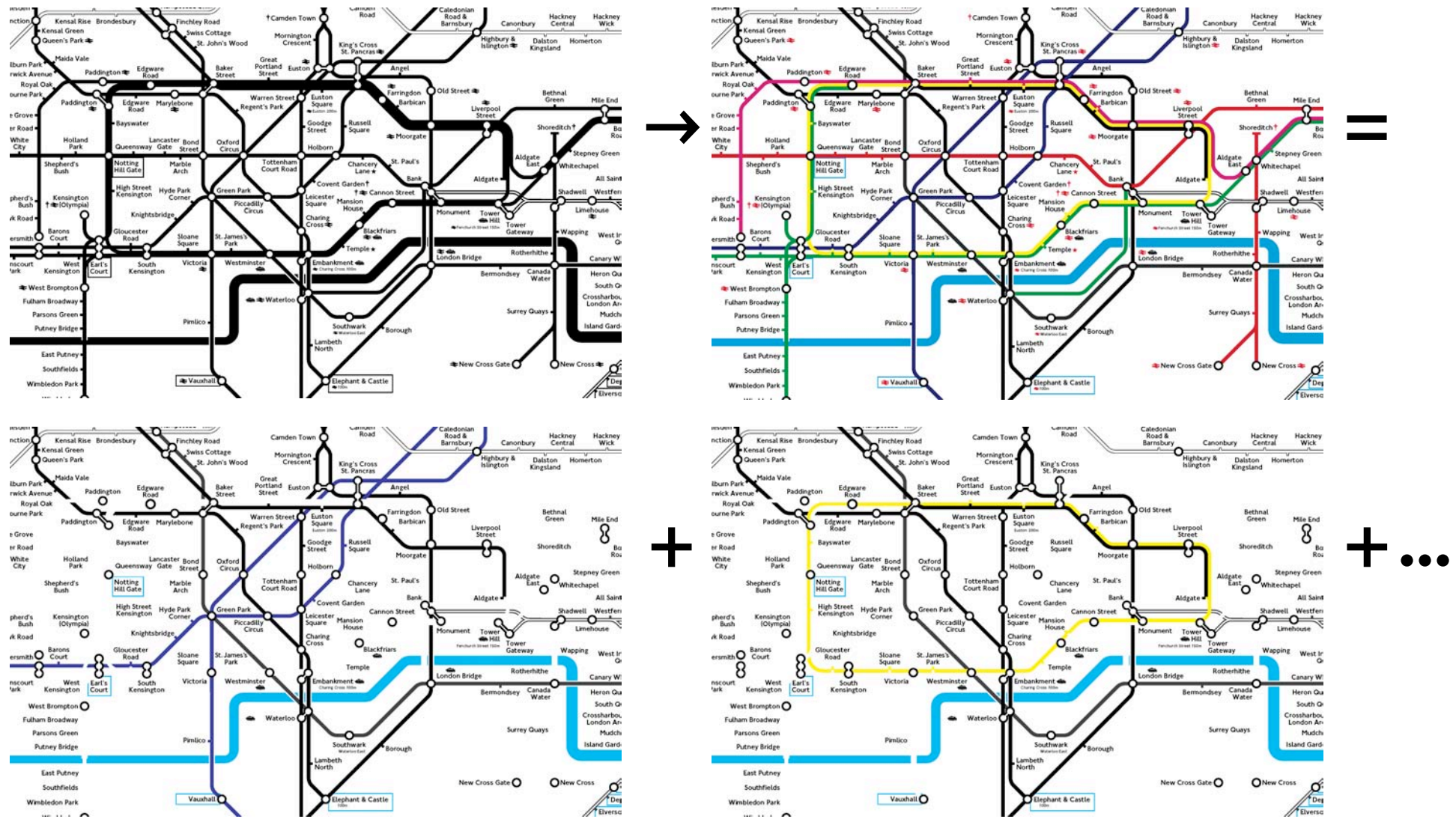

Uncover Pathways
Common to
Two Networks

HOEVD Comparative Modeling

Uncover Pathways
Common or Exclusive
Among Multiple Networks

Networks are Tensors of "Subnetworks"

The 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.
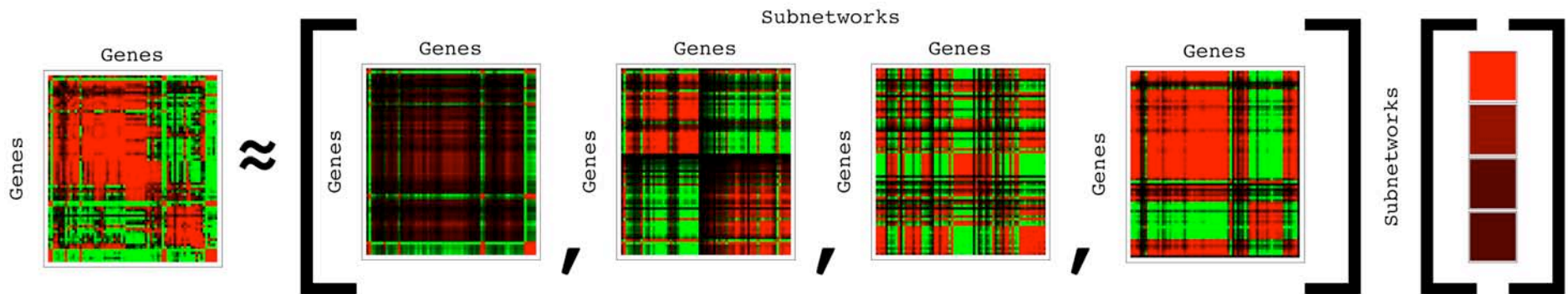
Eigenvalue Decomposition (EVD)

EVD formulates a genes × genes nondirectional network as a linear superposition of genes × genes decorrelated and decoupled rank-1 subnetworks, which can be associated with functionally independent pathways.

EVD of the network $\hat{a}_1$,

$$\hat{a}_1 = \hat{e}_1\hat{e}_1^T = \hat{u}_1\hat{\epsilon}_1^2\hat{u}_1^T = \sum_{m=1}^{M_1} \epsilon_{1,m}^2 |\alpha_{1,m}\rangle\langle\alpha_{1,m}|,$$

is computed from the SVD of the data signal $\hat{e}_1 = \hat{u}_1\hat{\epsilon}_1\hat{v}_1^T$.

Yeast Cell Cycle: Alpha Factor   Spellman et al., *MBC* 1998.

Math Variables → Biology
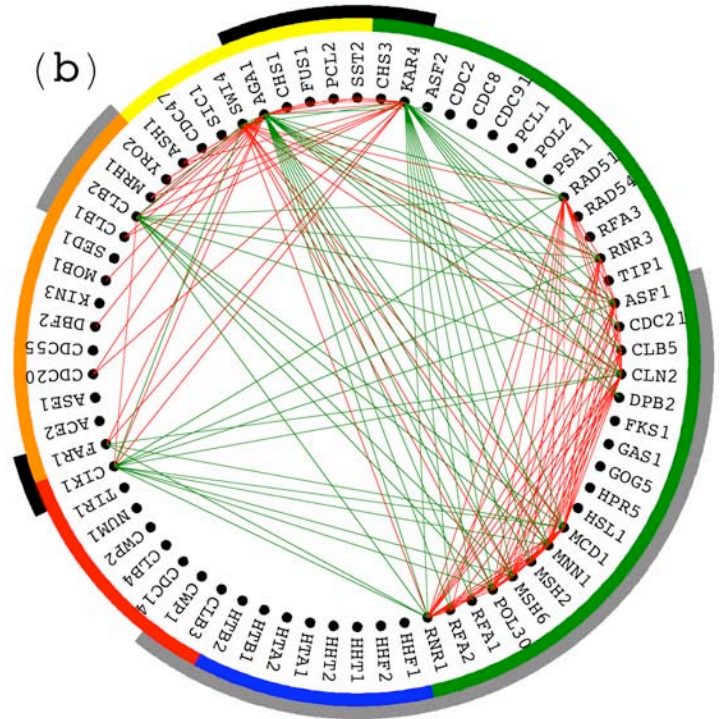## Significant EVD subnetworks →
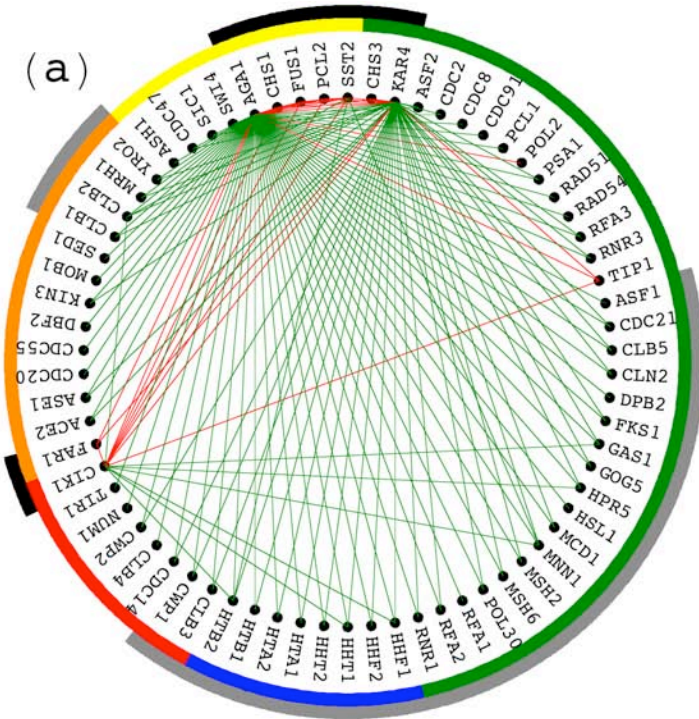### functionally independent pathways:

Pheromone Signaling Pathway
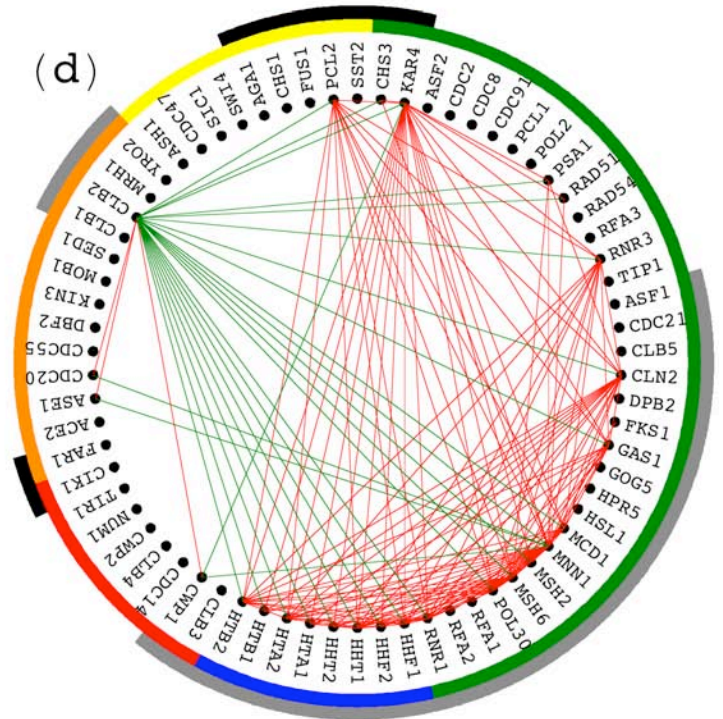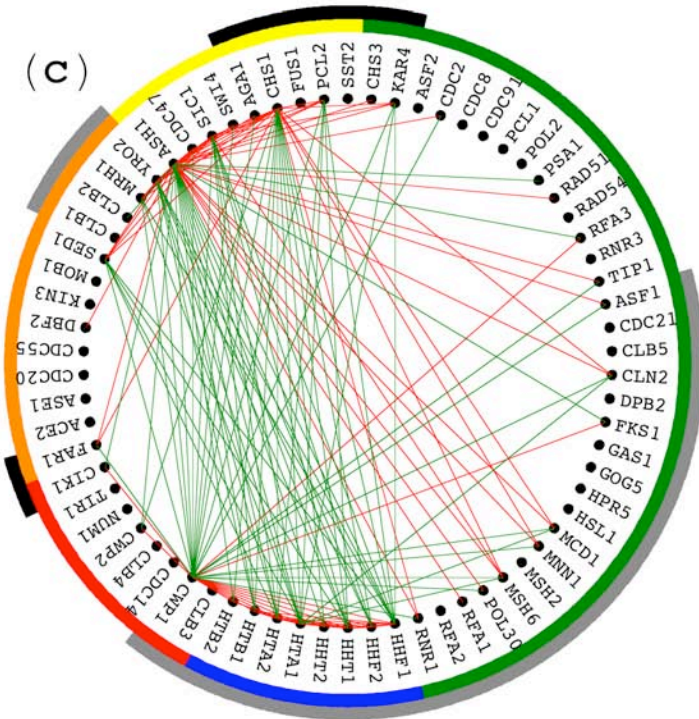## KAR4 ∥ CIK1

Pheromone Arrest Exit & G₁ Entry

Cell Cycle S ↔ M
KAR4 ∥ −CIK1

Cell Cycle G₁ ↔ G₂

Interpretation of the Subnetworks:
Probabilistic Associations by Annotations

Classification	Subnetwork	Most likely parallel association	P value of parallel association	Most likely antiparallel association	P value of antiparallel association
Cell Cycle	1	S S	1.7×10^{-22}	M/G_1 S	5.1×10^{-7}
	2	G_1 G_1	1.3×10^{-29}	G_1 G_2/M	3.2×10^{-11}
	3	S S	2.1×10^{-30}	M/G_1 S	2.6×10^{-25}
	4	G_1 S	2.1×10^{-28}	G_1 G_2/M	5.7×10^{-24}
Pheromone Response	1	Up Up	4.0×10^{-53}	Down Up	2.2×10^{-50}
	2	Down Down	1.6×10^{-11}	Down Up	9.8×10^{-17}
	3	Down Down	6.2×10^{-6}	Down Down	1.6×10^{-11}
	4	Down Down	8.0×10^{-32}	Down Down	2.5×10^{-6}

The P value of a given association by annotation is calculated using combinatorics 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 $x \subseteq X = N(N-1)/2$ pairs of genes with either largest and smallest levels of correlations in the subnetwork
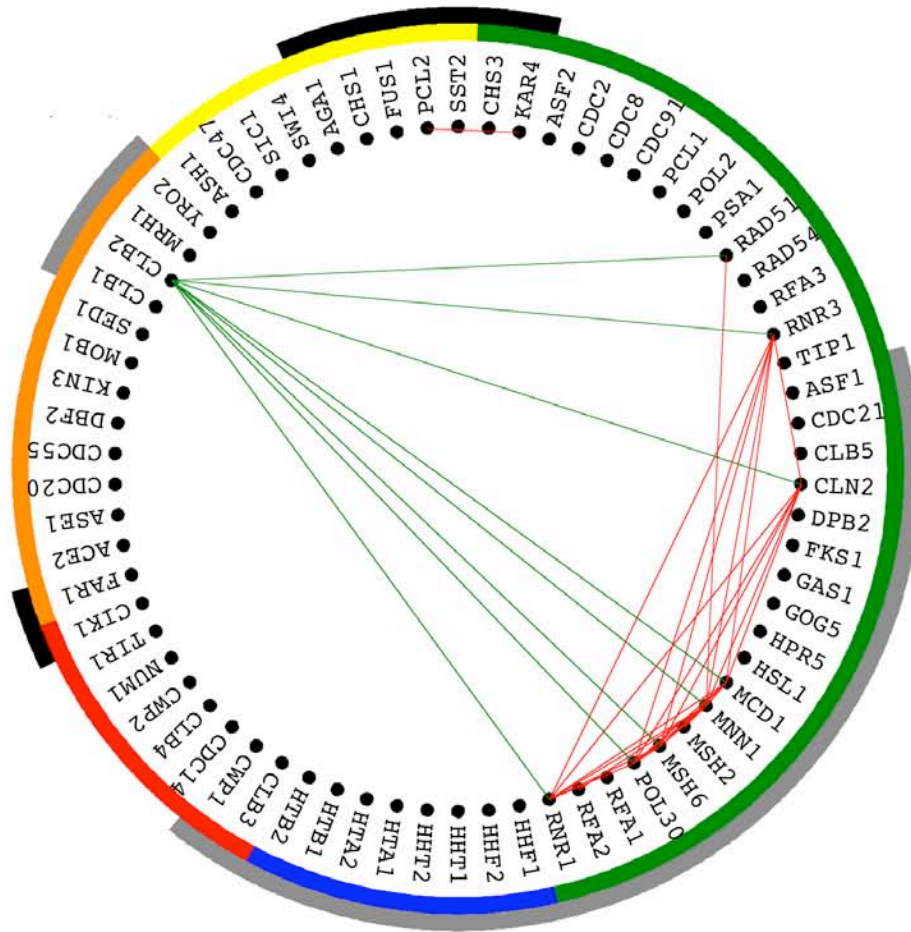
$$P(x; y, Y, X) = \binom{X}{x}^{-1} \sum_{z=y}^{x} \binom{Y}{z}\binom{X-Y}{x-z}.$$

where $\binom{X}{x} = X!x!^{-1}(X-x)!^{-1}$ is the binomial coefficient.

Math Operations → Biology

Boolean functions of subnetworks →
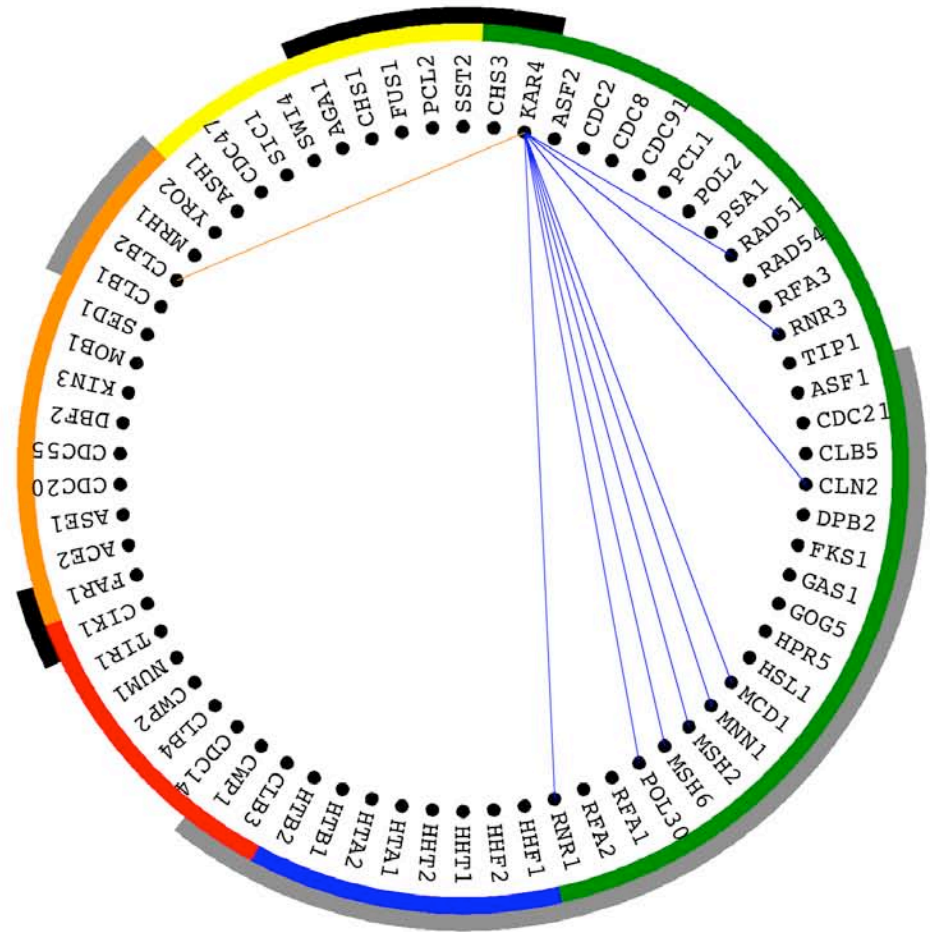
pathway-dependent relations among genes:

Pheromone Arrest Exit & G_1 Entry

AND

Cell Cycle $G_1 \leftrightarrow G_2$

CLB2 ‖ −CLN2

Pheromone Arrest Exit & G_1 Entry

AND NOT

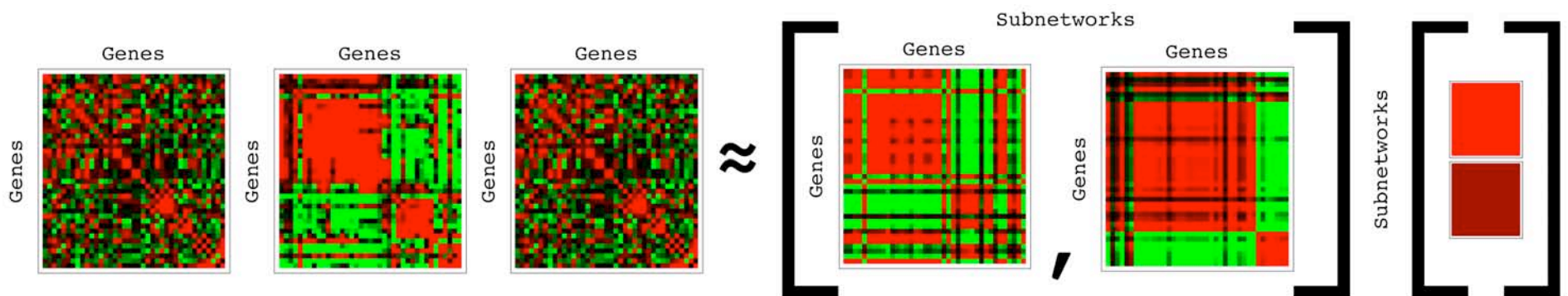Cell Cycle $G_1 \leftrightarrow G_2$

KAR4 ‖ ±CLN2

Integrative Pseudoinverse Projection

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.

EVD of $\hat{a}_2$, the projection of the network $\hat{a}_1$ onto the basis signal $\hat{b}$,

$$\hat{a}_1 \rightarrow \hat{a}_2 \equiv (\hat{b}\hat{b}^\dagger)\hat{a}_1(\hat{b}\hat{b}^\dagger) = \sum_{m=1}^{M_2} \epsilon_{2,m}^2 |\alpha_{2,m}\rangle\langle\alpha_{2,m}|,$$

is computed from the SVD of the projection of the data signal onto the basis signal $\hat{e}_1 \rightarrow \hat{e}_2 \equiv (\hat{b}\hat{b}^\dagger)\hat{e}_1$.
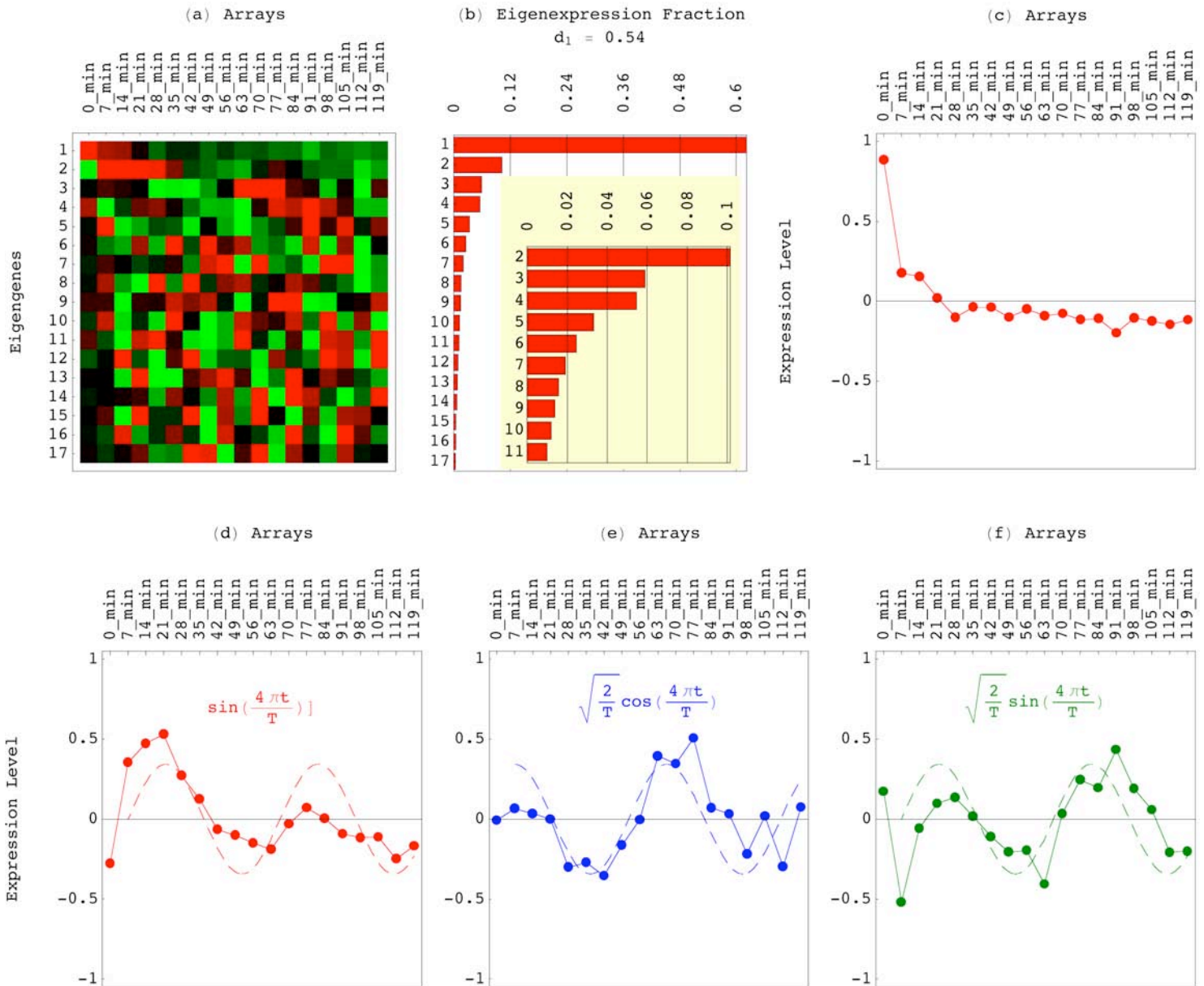

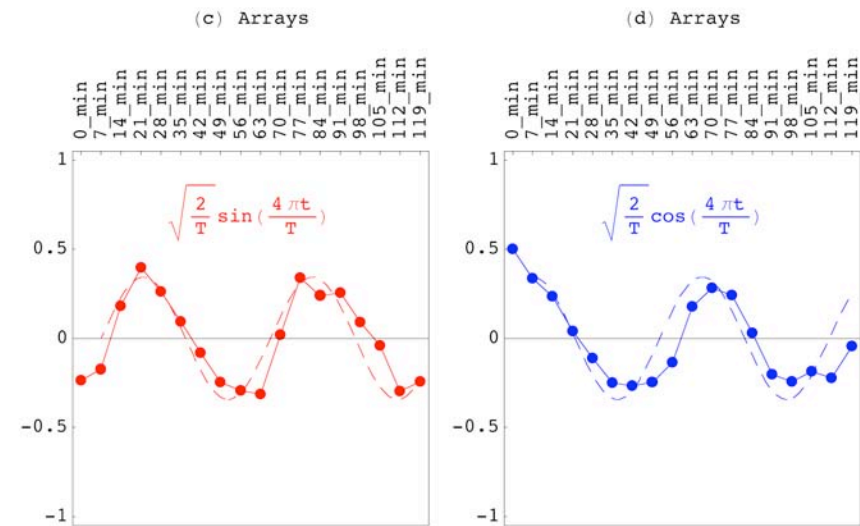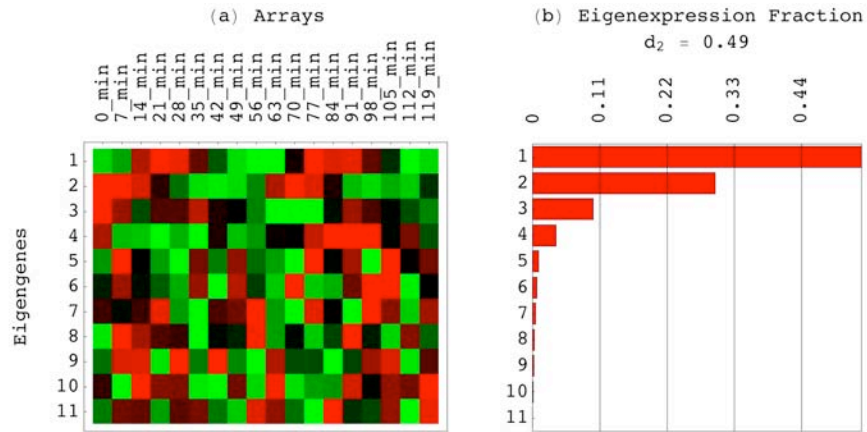Yeast Cell Cycle: Alpha Factor Spellman et al., *MBC* 1998;
Transcription Factors: Cycle & Development Lee et al., *Science* 2002.
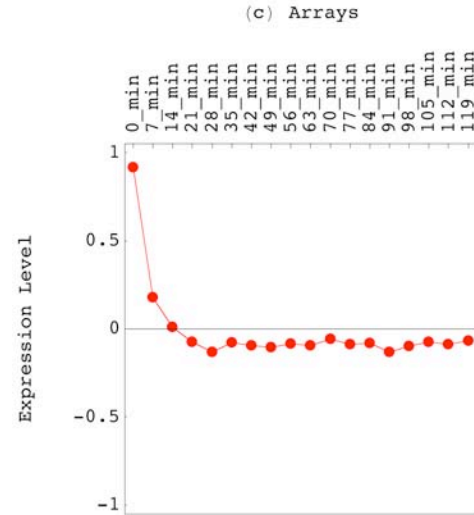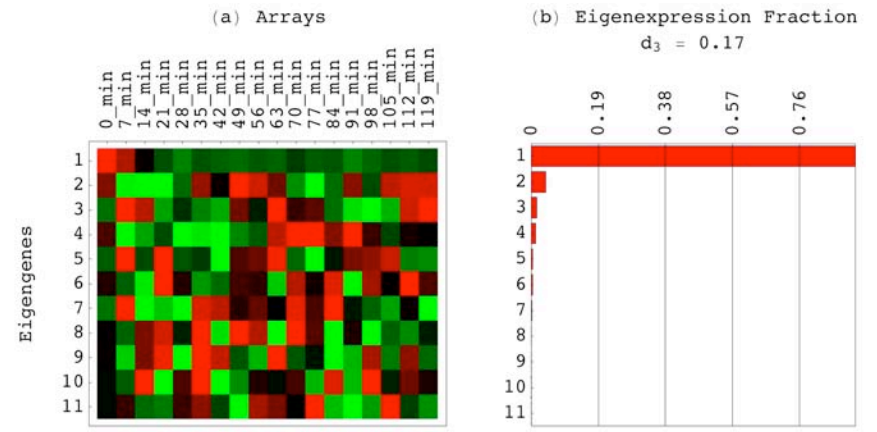
Math Operations & Variables → Biology

Pseudoinverse-projected network → observation of only the pathways manifest in both the data and basis:
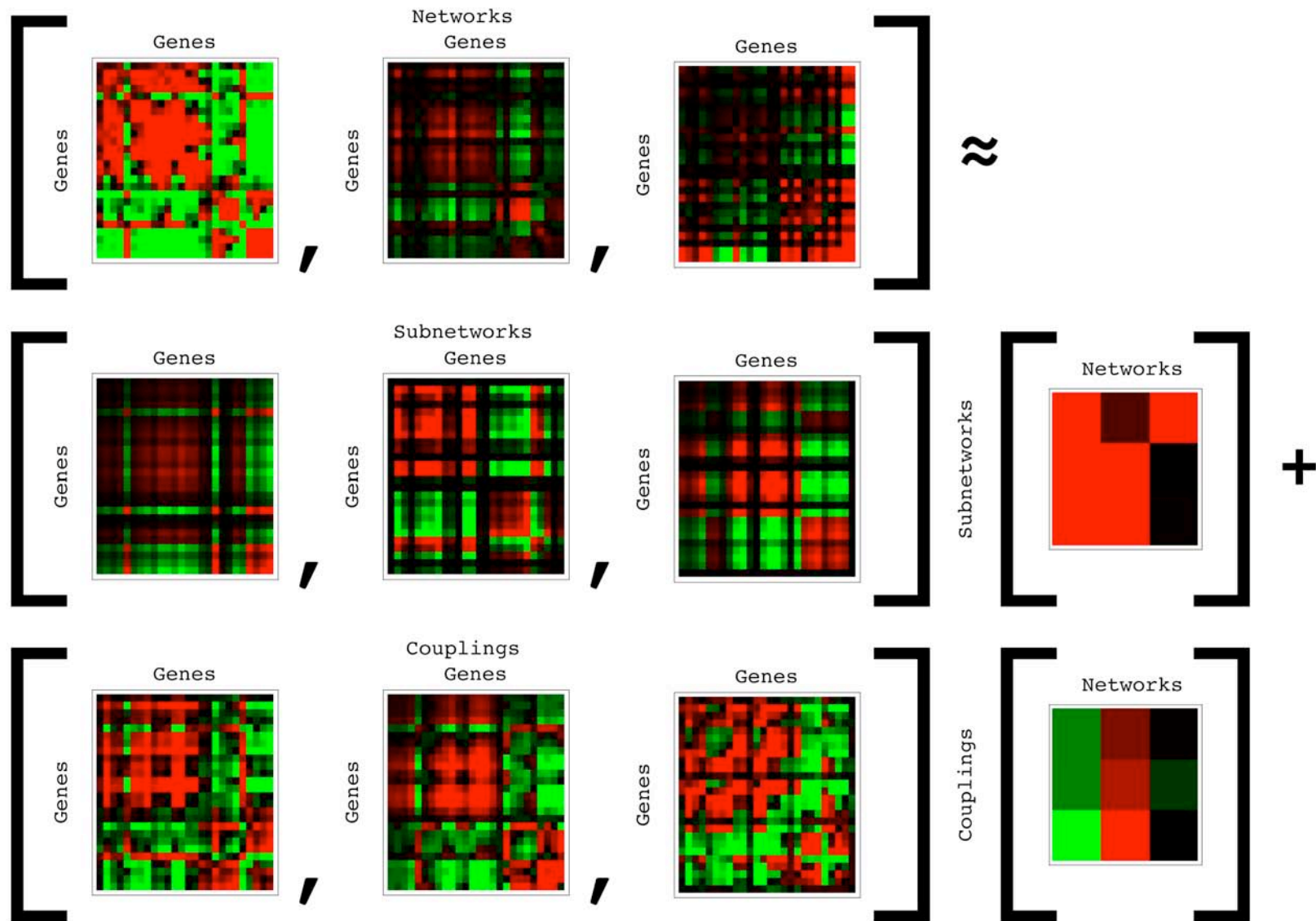
Cycle-Projected

Development-Projected

Comparative Higher-Order EVD (HOEVD)

… formulates a series of networks as linear superpositions of decorrelated rank-1 subnetworks and the rank-2 couplings among them.

Define and compute a higher-order EVD (HOEVD) of the tensor of networks $\{\hat{a}_k\}$,

$$\hat{a} \equiv \sum_{k=1}^{K} \hat{a}_k = \hat{u}\left(\sum_{k=1}^{K} \hat{\epsilon}_k^2\right)\hat{u}^T = \hat{u}\hat{\epsilon}^2\hat{u}^T$$

using the SVD of the appended signals

$$\hat{e} \equiv (\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_K) = \hat{u}\hat{\epsilon}\hat{v}^T$$

This HOEVD formulates each individual network in the tensor $\{\hat{a}_k\}$ as a linear superposition of this series of M rank-1 symmetric decorrelated subnetworks and the series of $M(M-1)/2$ rank-2 symmetric couplings among these subnetworks, such that
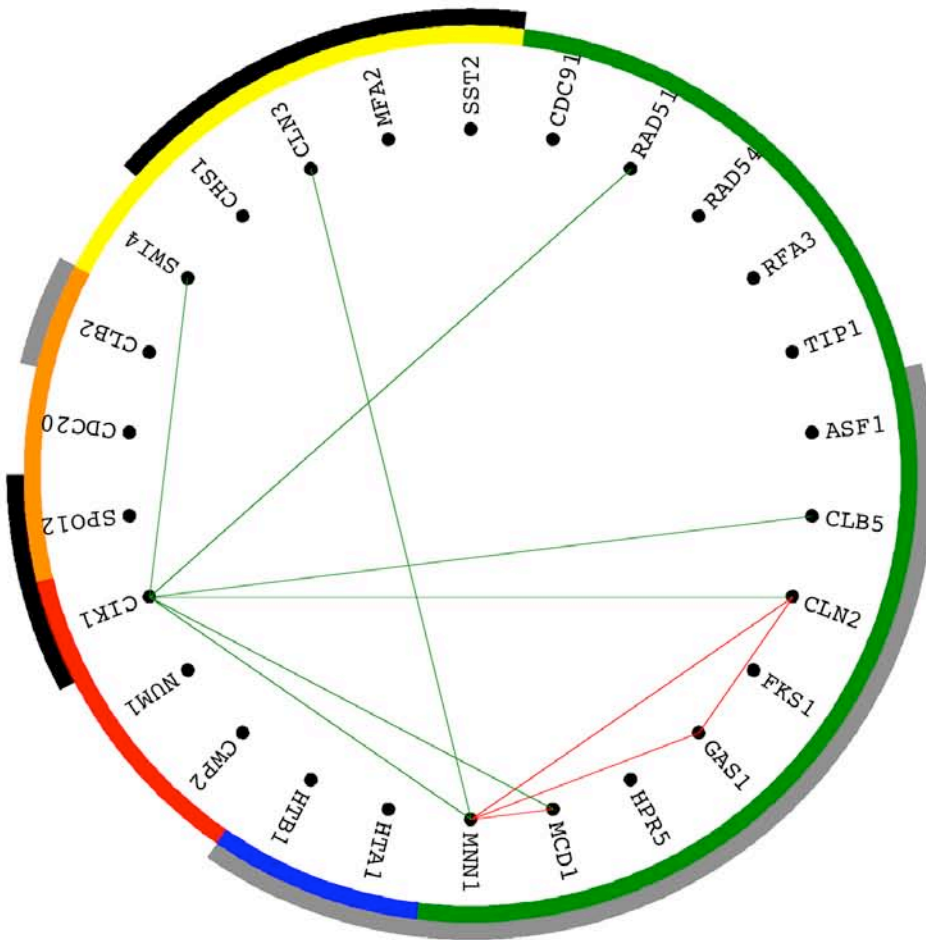
$$\hat{a}_k = \sum_{m=1}^{M} \epsilon_{k,m}^2 |\alpha_m\rangle\langle\alpha_m|$$
$$+ \sum_{m=1}^{M} \sum_{l=m+1}^{M} \epsilon_{k,lm}^2 (|\alpha_l\rangle\langle\alpha_m| + |\alpha_m\rangle\langle\alpha_l|),$$

for all $k = 1, 2, \ldots, K$.

De Lathauwer, De Moor & Vandewalle, *SIAM J. Matrix Anal. Appl.* 2000;
Kolda, *SIAM J. Matrix Anal. Appl.* 2001;
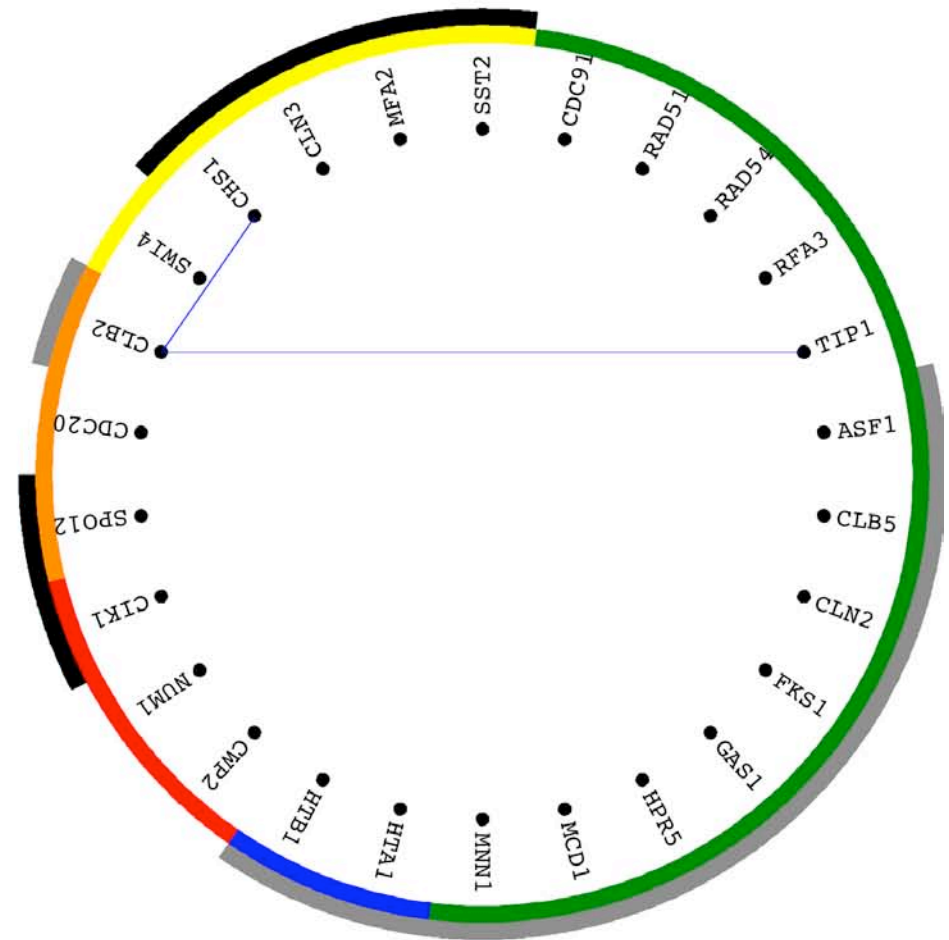Zhang & Golub, *SIAM J. Matrix Anal. Appl.* 2001.

Math Operations & Variables → Biology

HOEVD subnetworks and their couplings → pathways and transitions among them common to the series or exclusive to a subset of networks:

Known Relations:
Pheromone Response **AND** $G_1 \leftrightarrow G_2$
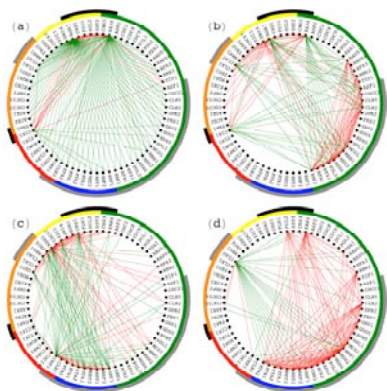**AND** Transition Between These
CLN2 ‖ −*CIK1*

Novel Relations:
Pheromone Response **AND** $G_1 \leftrightarrow G_2$
AND NOT Transition Between These
CLB2 ‖ ±*TIP1*

Uncovering Subnetworks of Conditions
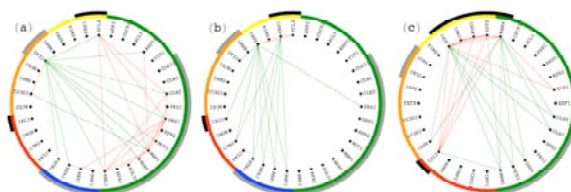# and the Transitions Among Them
From Networks of Correlations

Alter & Golub, *PNAS* 102, 17559 (2005);
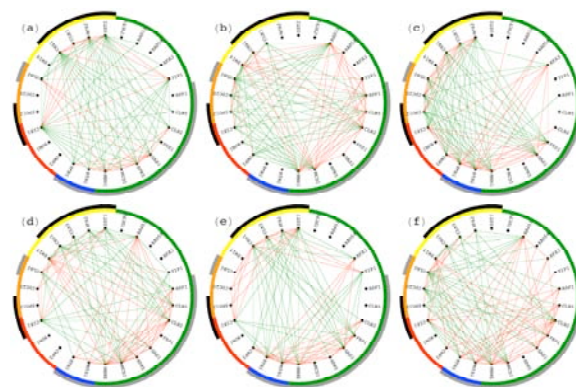http://www.bme.utexas.edu/research/orly/network_decomposition/.

EVD Modeling

Pseudoinverse Integrative Modeling

HOEVD Comparative Modeling

Uncover Pathways
in
a Single Network

Uncover Pathways
Common to
Two Networks
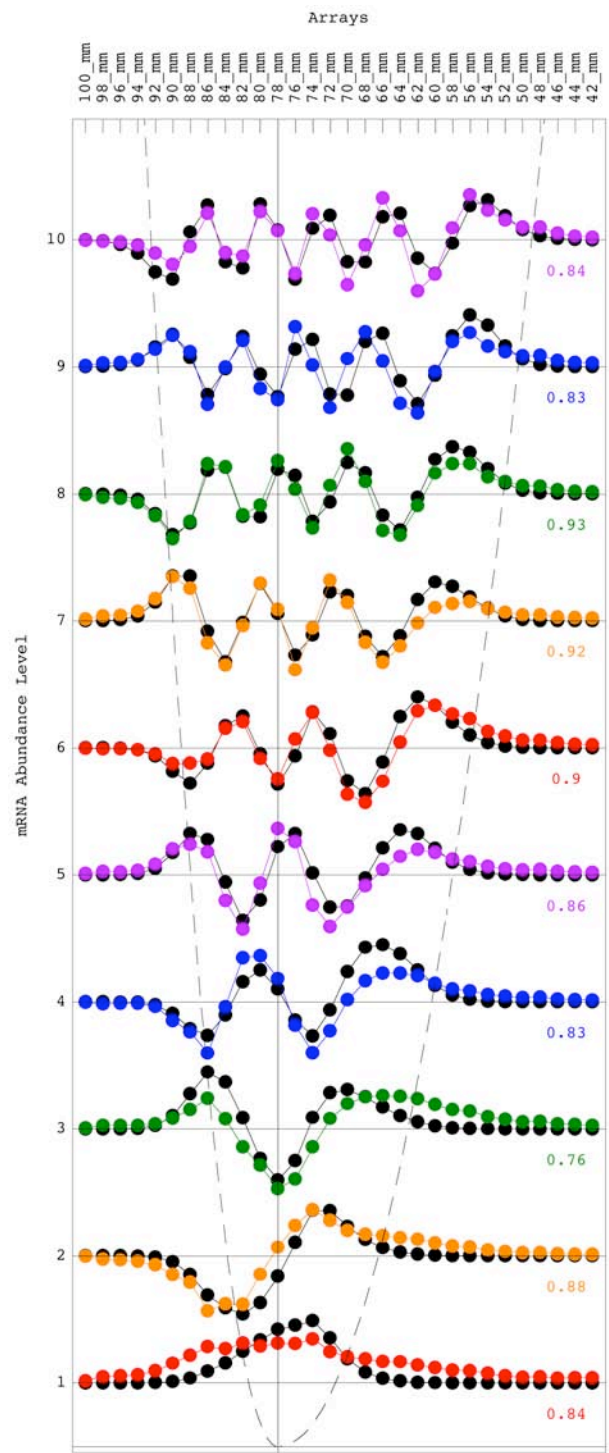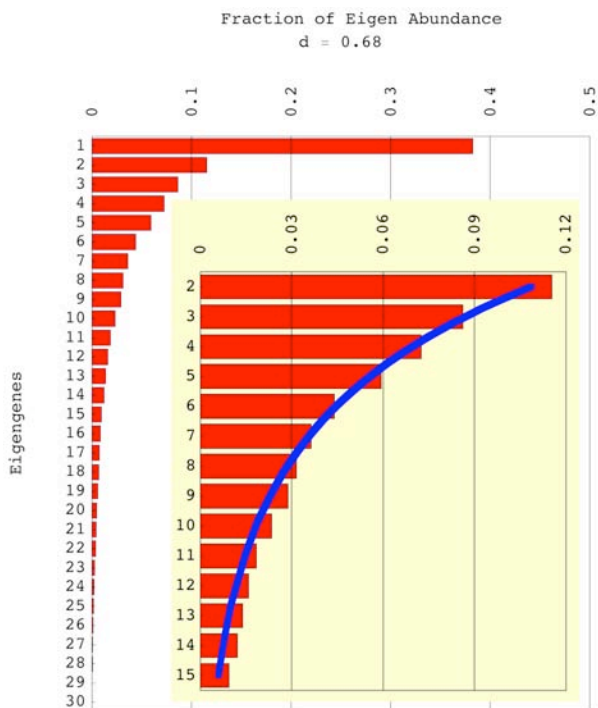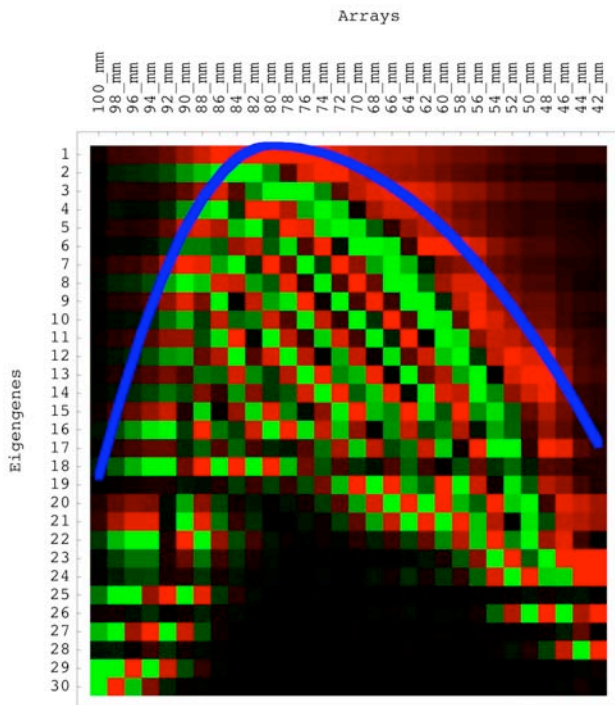
Uncover Pathways
Common or Exclusive
Among Multiple Networks

SVD Modeling of Genome-Wide mRNA Lengths Distribution
Predicts a Physical Principle

Alter & Golub, *PNAS* 103, 11828 (2006);
http://www.bme.utexas.edu/research/orly/harmonic_oscillator/.

Hurowitz & Brown, *Genome Biology* 2003.

Thanks to –

Collaborators:

John F. X. Diffley
Cancer Research UK, London

Gene H. Golub
Computer Science, Stanford

Vishy Iyer
Molecular Genetics, UT

David Botstein
Genomics Institute, Princeton

Patrick O. Brown
Biochemistry, Stanford

Matt van de Rijn
Pathology, Stanford

Students:

Kayta Kobayashi, Pharmacy, UT

Larsson Omberg, Physics, UT

Sri Priya Ponnapalli, ECE, UT

Chaitanya Muralidhara, CMB, UT

Joel Meyerson, BME, UT

Funding:

And, thank you!!!