Discovery of Principles of Nature From Mathematical Modeling of DNA Microarray Data

Orly Alter

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

University of Texas at Austin

A groundbreaking look at the nature of quantum mechanics

With new technologies permitting the observation and manipulation of single quantum systems, the quantum theory of measurement is fast becoming a subject of experimental investigation in laboratories worldwide. This original new work addresses open fundamental questions in quantum mechanics in light of these experimental developments.

Using a novel analytical approach developed by the authors, *Quantum Measurement of a Single System* provides answers to three long-standing questions that have been debated by such thinkers as Bohr, Einstein, Heisenberg, and Schrödinger. It establishes the quantum theoretical limits to information obtained in the measurement of a single system on the quantum wavefunction of the system, the time evolution of the quantum observables associated with the system, and the classical potentials or forces which shape this time evolution. The technological relevance of the theory is also demonstrated through examples from atomic physics, quantum optics, and mesoscopic physics.

Suitable for professionals, students, or readers with a general interest in quantum mechanics, the book features recent formulations as well as humorous illustrations of the basic concepts of quantum measurement. Researchers in physics and engineering will find *Quantum Measurement of a Single System* a timely guide to one of the most stimulating fields of science today.

ORLY ALTER, PhD, is currently a postdoctoral fellow in the Department of Genetics at Stanford University. YOSHIHISA YAMAMOTO, PhD, is a professor in the Departments of Applied Physics and Electrical Engineering at Stanford University. He is currently the director of the ICORP Quantum Entanglement Project of the Japanese Science and Technology (JST) Corporation. While they collaborated on the research presented in this book, Yamamoto was the director of the ERATO Quantum Fluctuation Project of JST, and Alter was a doctoral student at the Department of Applied Physics at Stanford. She was selected as a finalist for the American Physical Society Award for Outstanding Doctoral Thesis Research in Atomic, Molecular or Optical Physics for 1998 for this work.

Cover Illustration: David B. Oberman

WILEY-INTERSCIENCE
John Wiley & Sons, Inc.
Scientific, Technical, and Medical Division
605 Third Avenue, New York, N.Y. 10158-0012
New York • Chichester • Weinheim
Brisbane • Singapore • Toronto

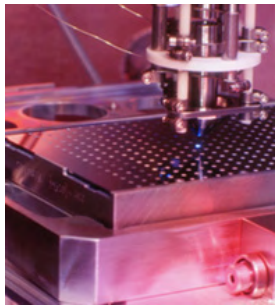
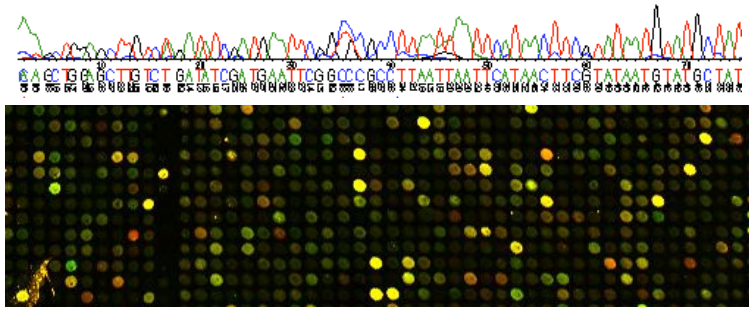
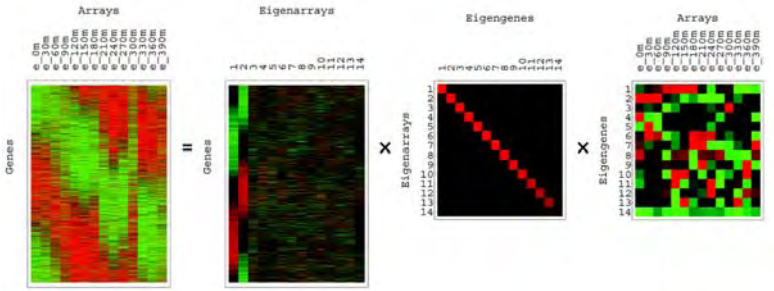
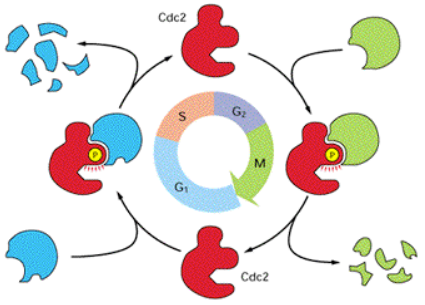
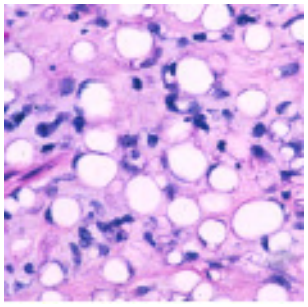
ISBN 0-471-28308-6
90000
9 780471 283089

ALTER
YAMAMOTO

Quantum Measurement of a Single System

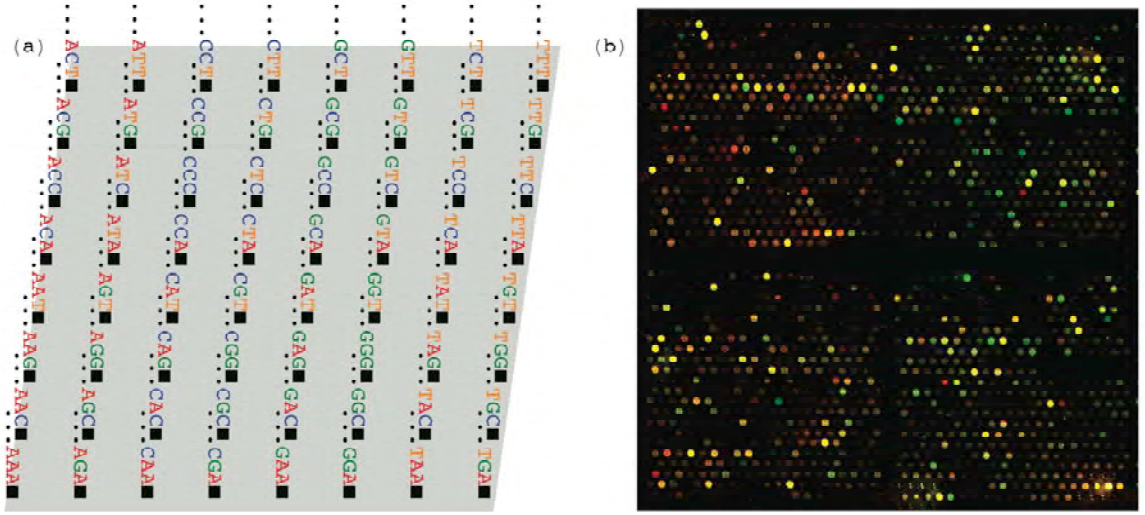
Quantum Measurement of a Single System

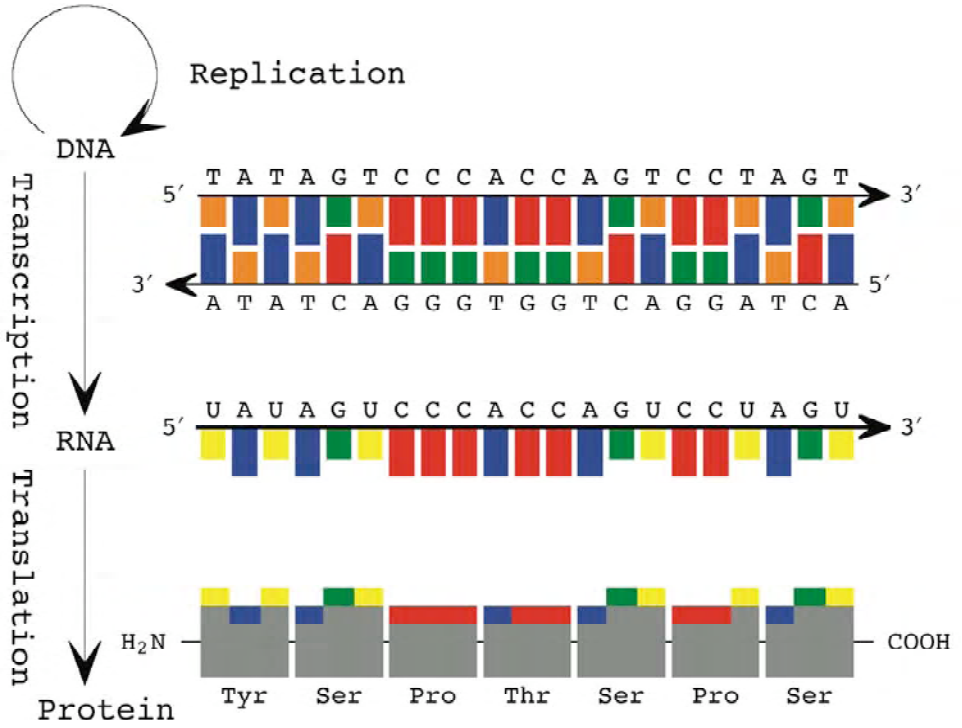
Orly Alter
Yoshihisa Yamamoto

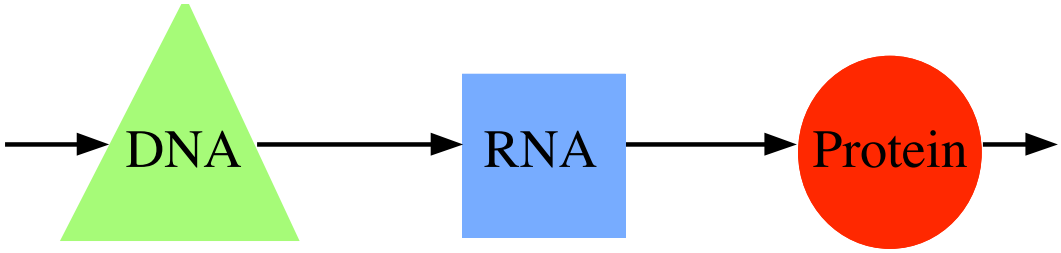
	Astronomy	Molecular Biology
Technology	Galileo	 
Large-Scale Data	Brahe	
Mathematical Modeling	Kepler	
Basic Principles	Newton	 
Technology	NASA	Control of Cellular Mechanisms

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.





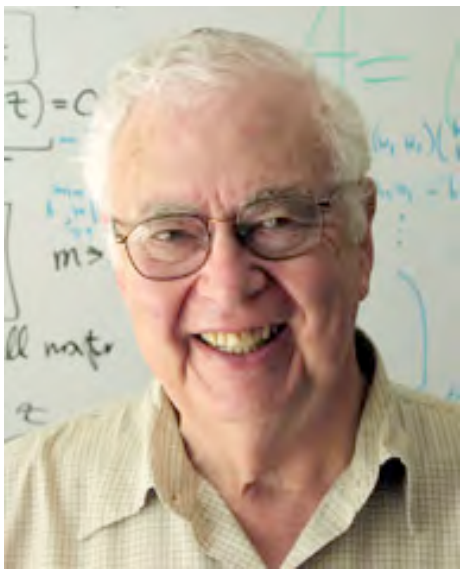


Genomic Signal Processing

The data are in large quantities.
Artifacts are superimposed on the data.
Different types of genome-scale data need to be understood simultaneously.
Existing genetic models applied to genome-wide data appear inconsistent.

Analogy From Machine Vision

Large-scale biological signals are complex, easily understood by the biological system, simple laws may govern the complex signal.

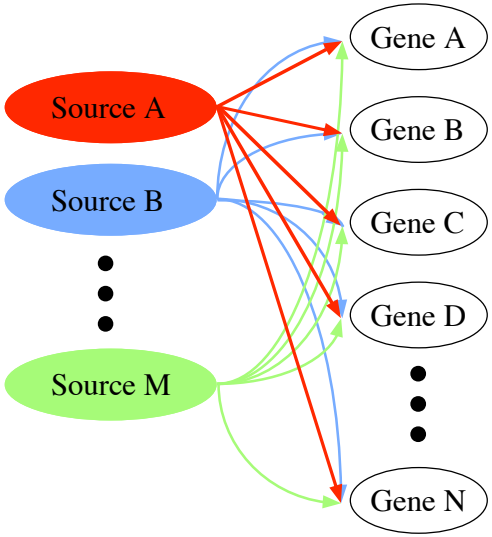


Data-Driven Models for Genomic Data

Alter, *PNAS* 103, 16063 (2006);
Alter, to be published in Microarray Data Analysis: Methods and Applications (Humana Press).

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

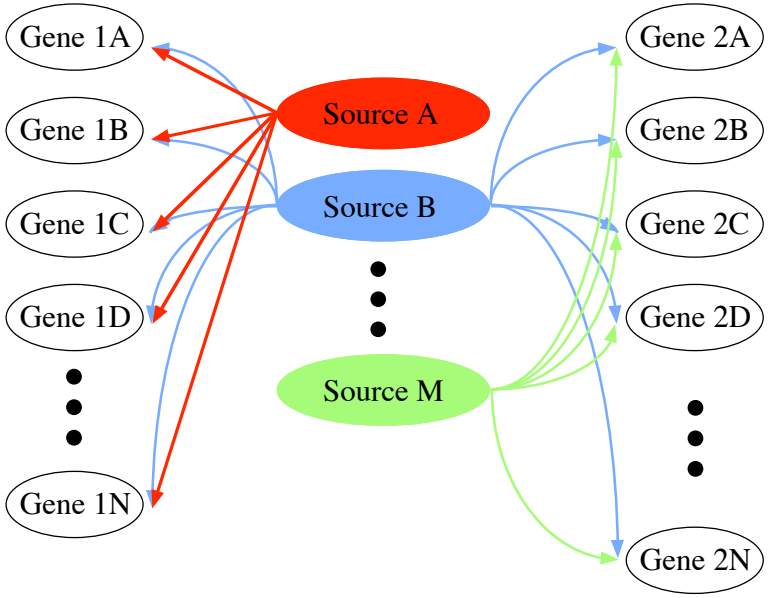
SVD Modeling



Uncover Cellular Processes and States

Eigenvalue Decomposition

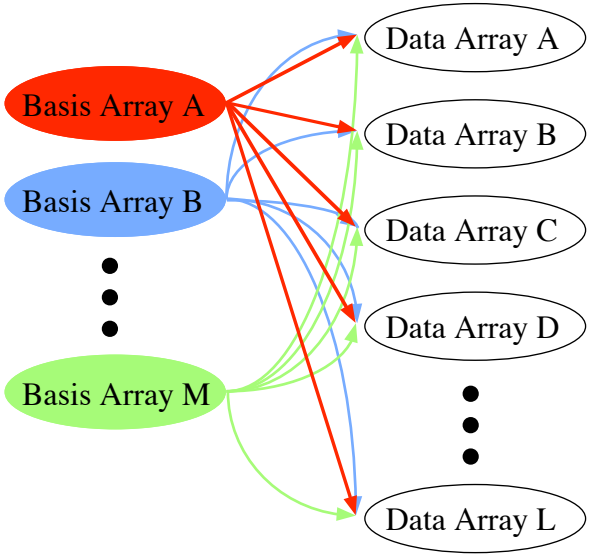
GSVD Comparative Modeling



Uncover Processes Common or Exclusive Among Two Datasets

Generalized Eigenvalue Decomposition

Pseudoinverse Integrative Modeling



Uncover Coordination Among Multiple Sets

Inverse Projection

Predicting a Biological Principle:

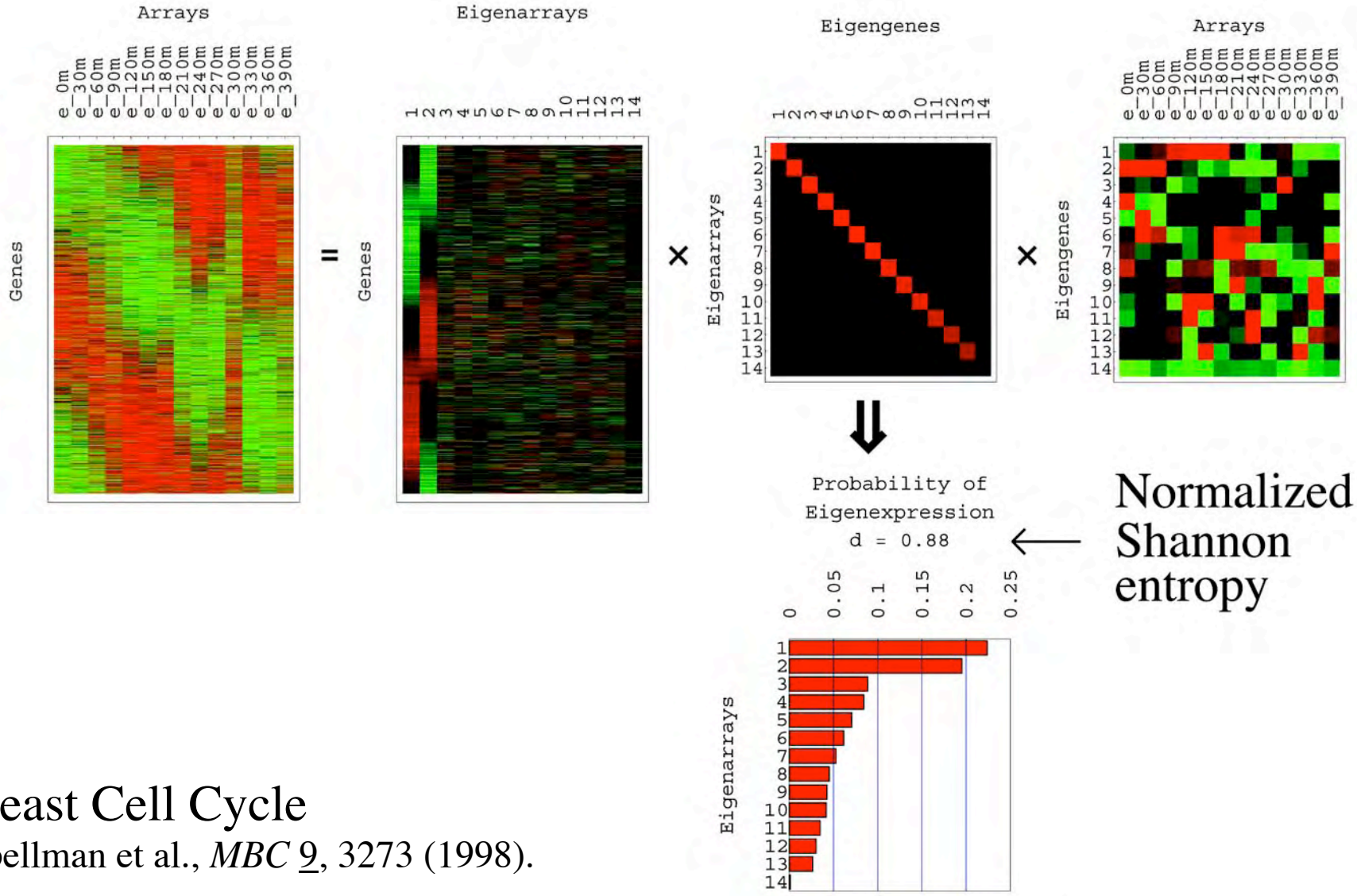
Previously Unknown Correlation Between DNA Replication Initiation and RNA Transcription

Might Be Due to an Undiscovered Mechanism of Regulation

Singular Value Decomposition (SVD)

Alter, Brown & Botstein, *PNAS* 97, 10101 (2000);
http://www.bme.utexas.edu/research/orly/SVD/PNAS_2000/.

Linear transformation of gene expression data from genes × arrays space to reduced diagonalized "eigengenes" × "eigenarrays" space.



Yeast Cell Cycle

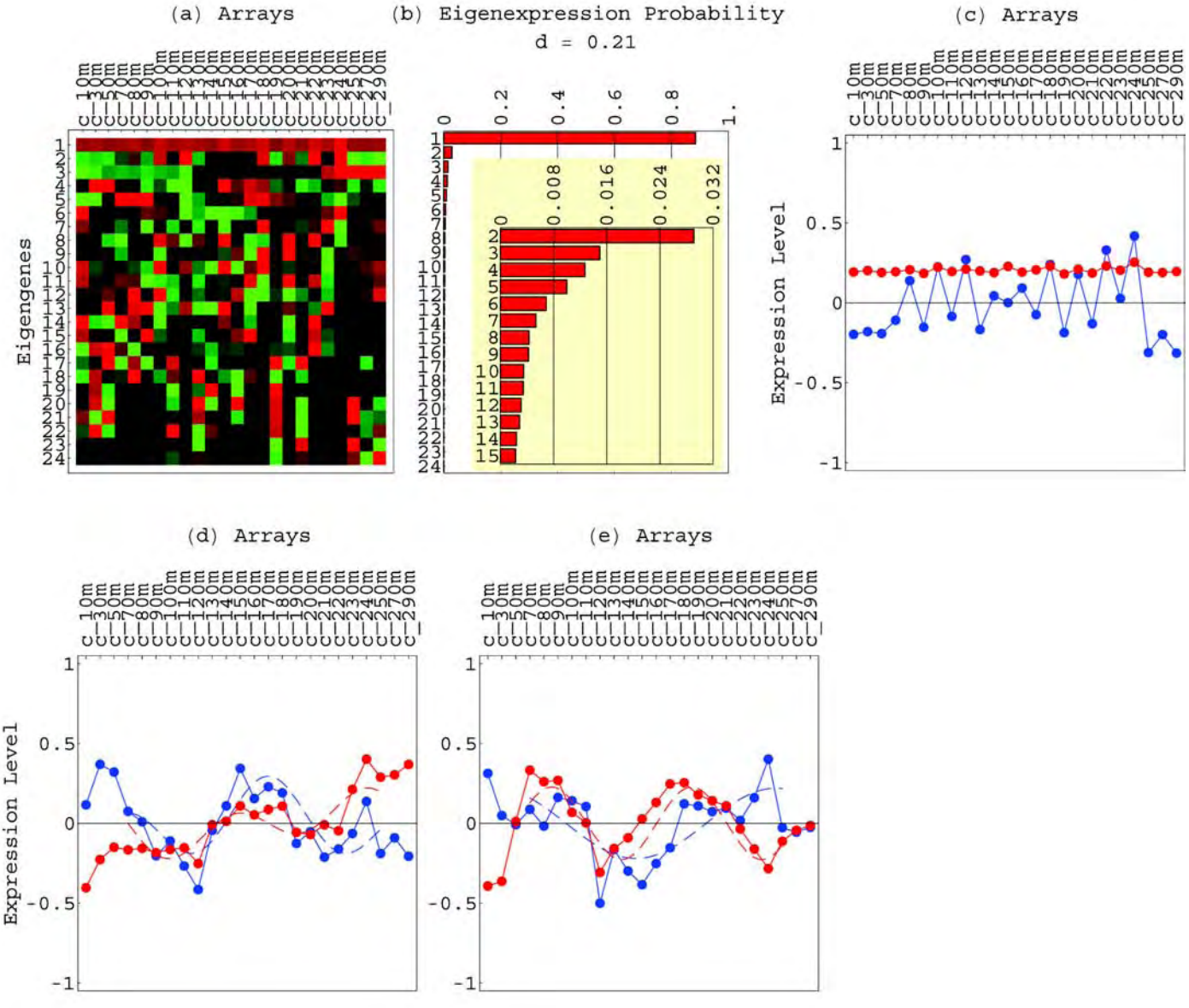
Spellman et al., *MBC* 9, 3273 (1998).

Math Variables → Biology

Significant eigengenes → independent biological processes and experimental artifacts:

90% of expression is steady state,
2.5% is day-of-hybridization artifact,
less than 7.5% is periodic →

Weak Signal Detection



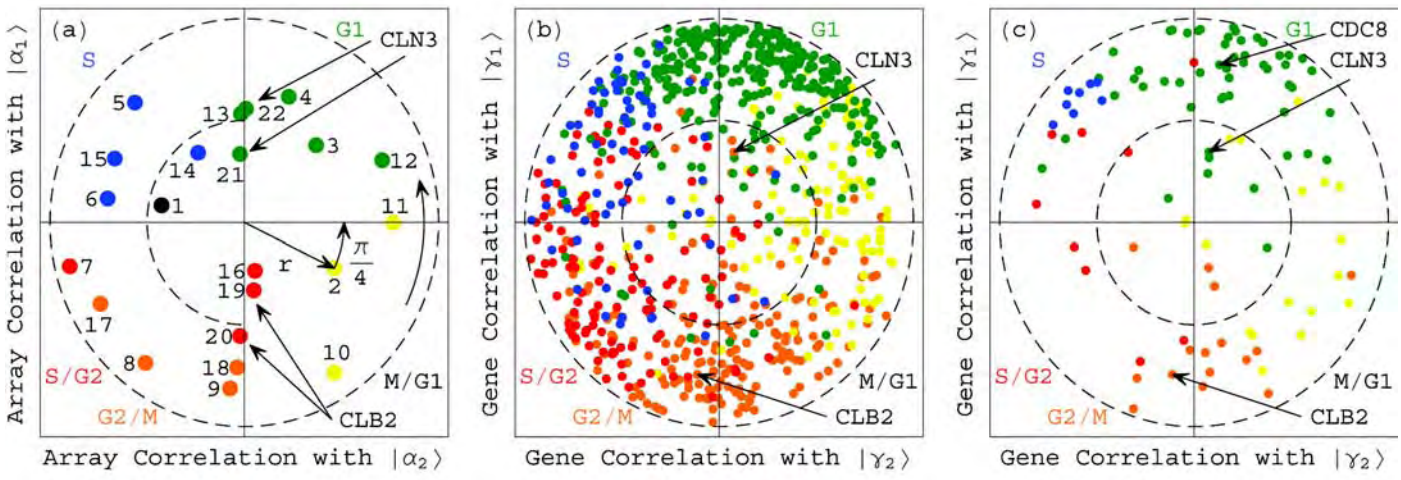
Yeast Cell Cycle: Cdc15 Spellman et al., *MBC* 9, 3273 (1998).

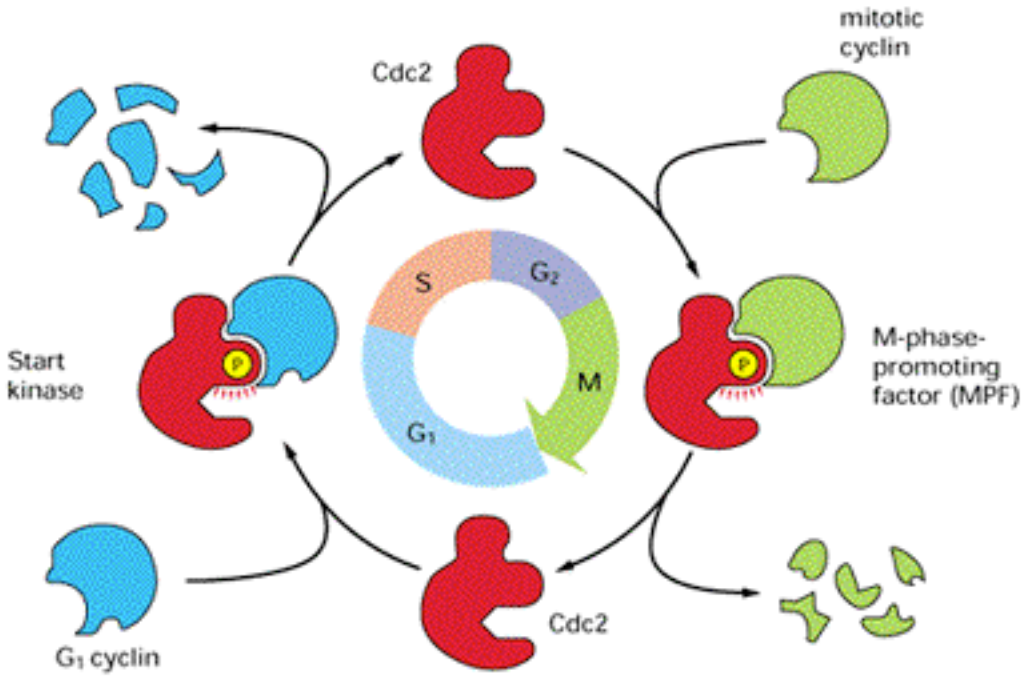
Math Variables → Biology

Significant eigengenes and eigenarrays → genome-wide effects of regulators, and samples in which these regulators are overactive, respectively:

Cln3, Clb2 genome–wide effects = ± first eigengene

Cln3, Clb2 overactive samples = ± first eigenarray

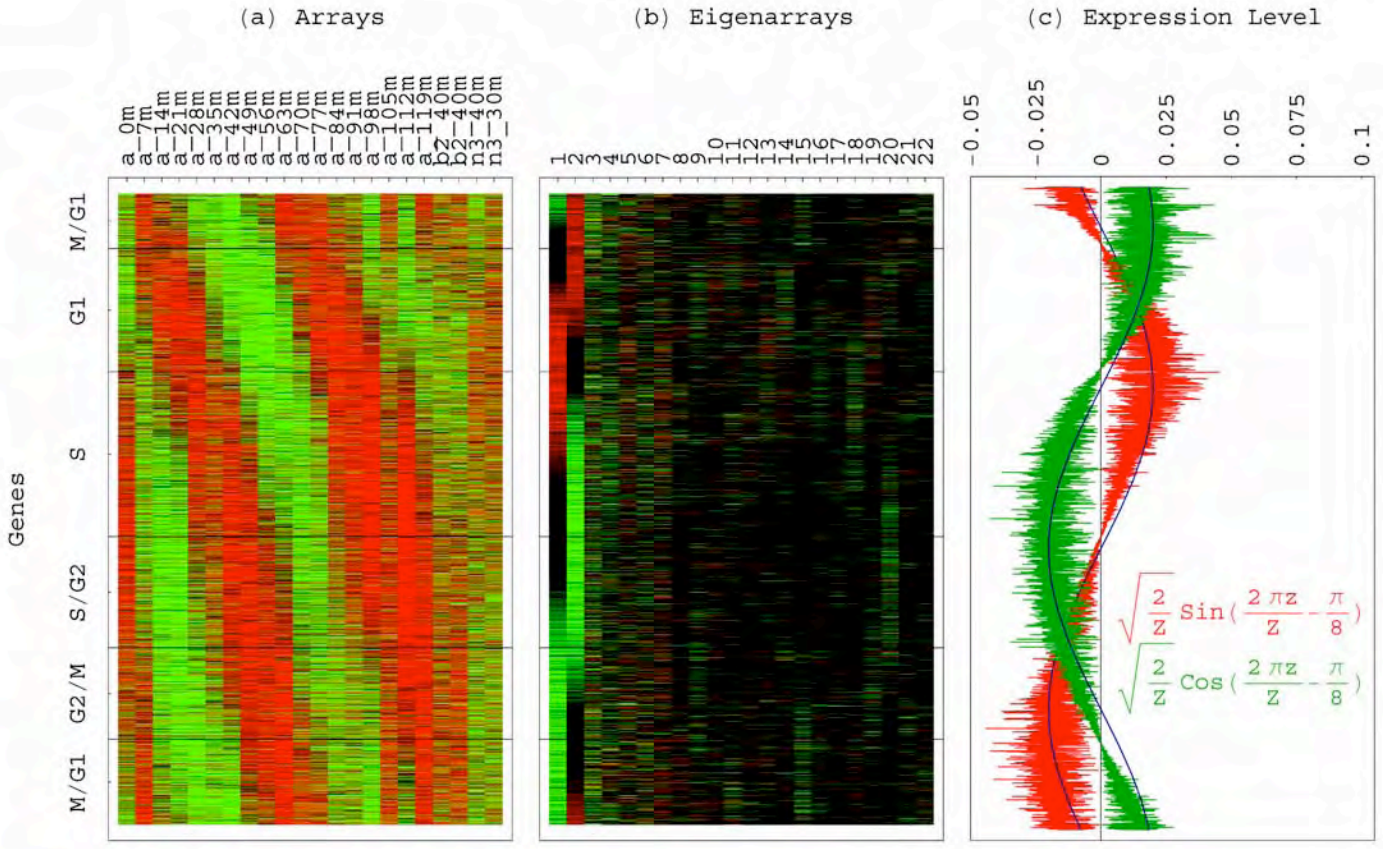




Alberts et al., *Molecular Biology of the Cell* (1994).

Traveling Wave of Expression

Cln3, Clb2 overactive samples = ± first eigenarray



Consistent model for the expression of almost the full yeast genome during cell cycle, in a subspace spanned by only two eigengenes and corresponding eigenarrays.

→ Are there only two cellular elements or modules that drive the yeast cell cycle?

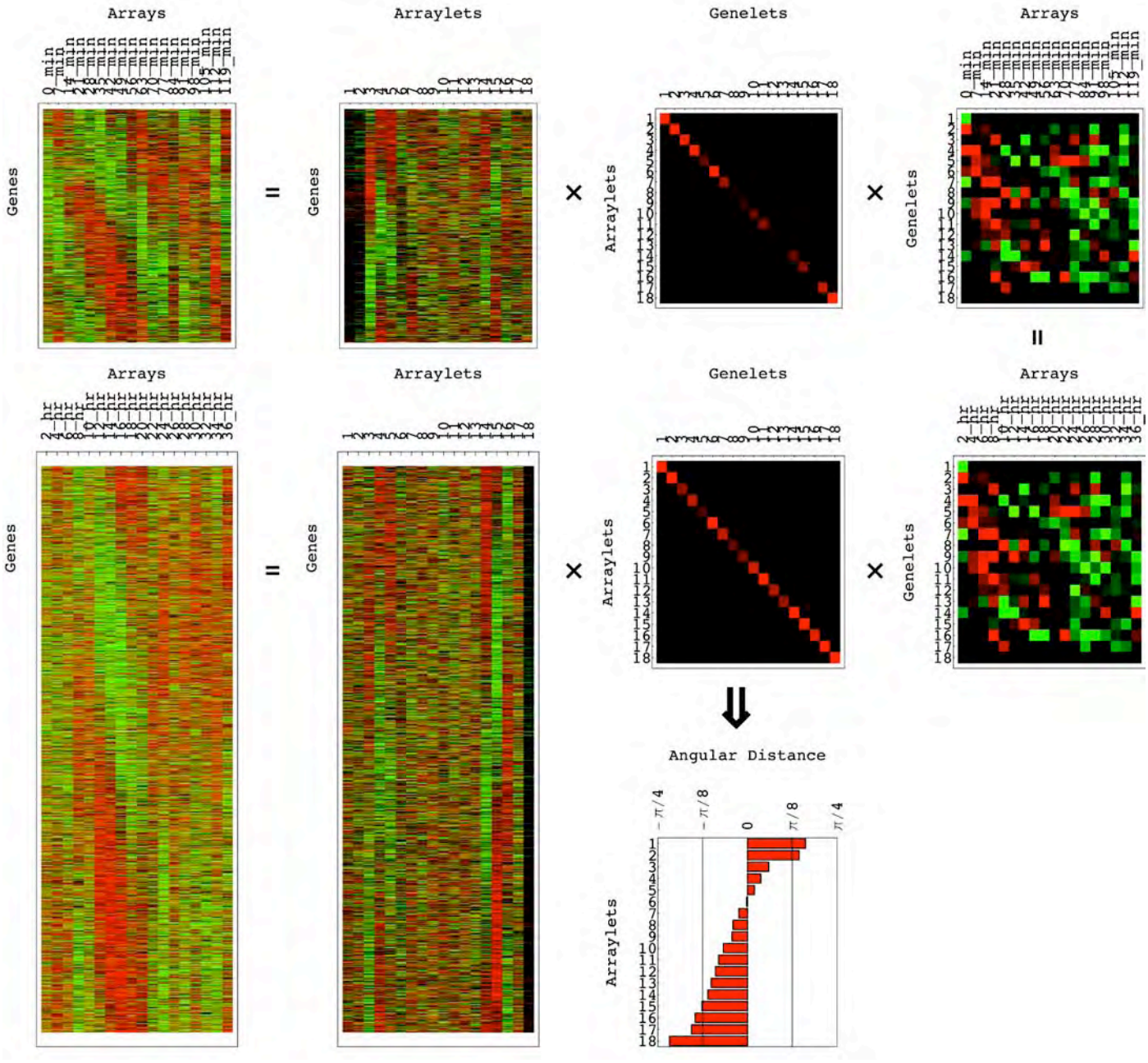
→ Can we design a synthetic genetic network analogous to the analog harmonic oscillator, which would simulate the yeast cell cycle?

GSVD for Comparative Analysis

Alter, Brown & Botstein, *PNAS* 100, 3351 (2003);
http://www.bme.utexas.edu/research/orly/GSVD/.

Linear transformation of two datasets from two genes × arrays spaces to two reduced diagonalized "genelets" × "arraylets" spaces.

Yeast Cell Cycle Spellman et al. *MBC* 9 3273 (1998).

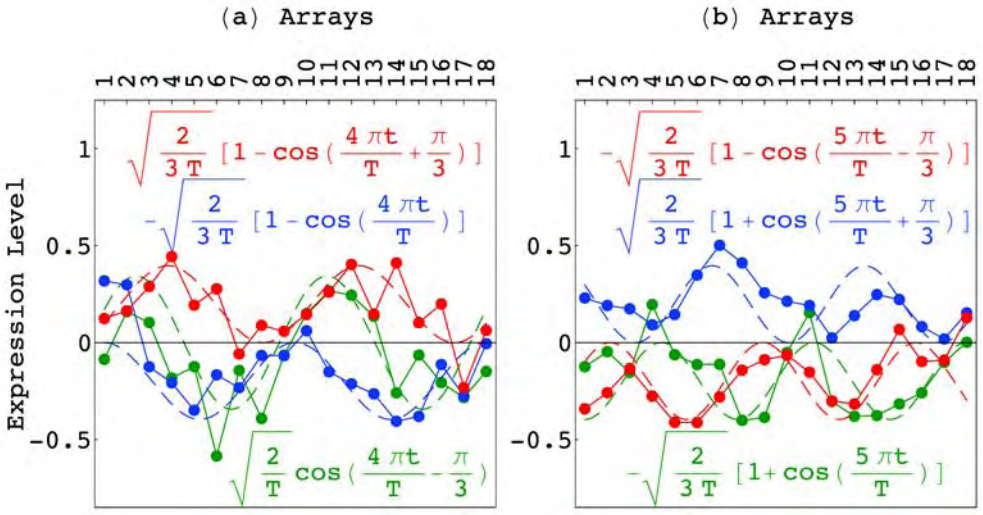


Human Cell Cycle Whitfield et al., *MBC* 13, 1977 (2002).

Math Variables → Biology

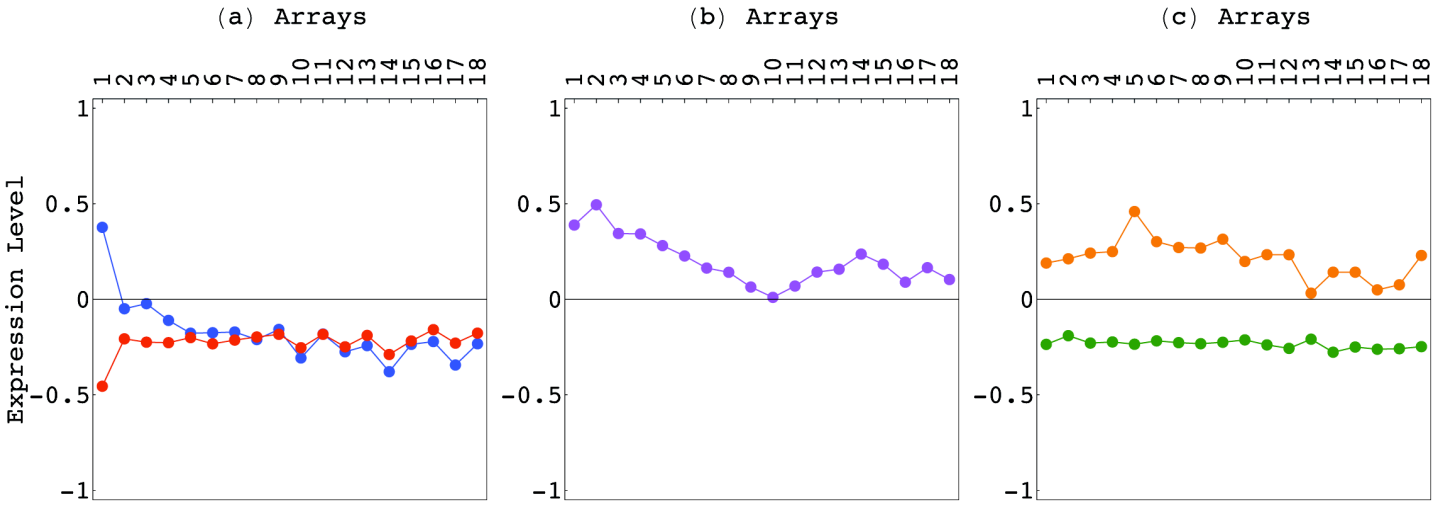
Genelets of almost equal significance in both datasets → processes common to both genomes:

Common Cell Cycle Subspace



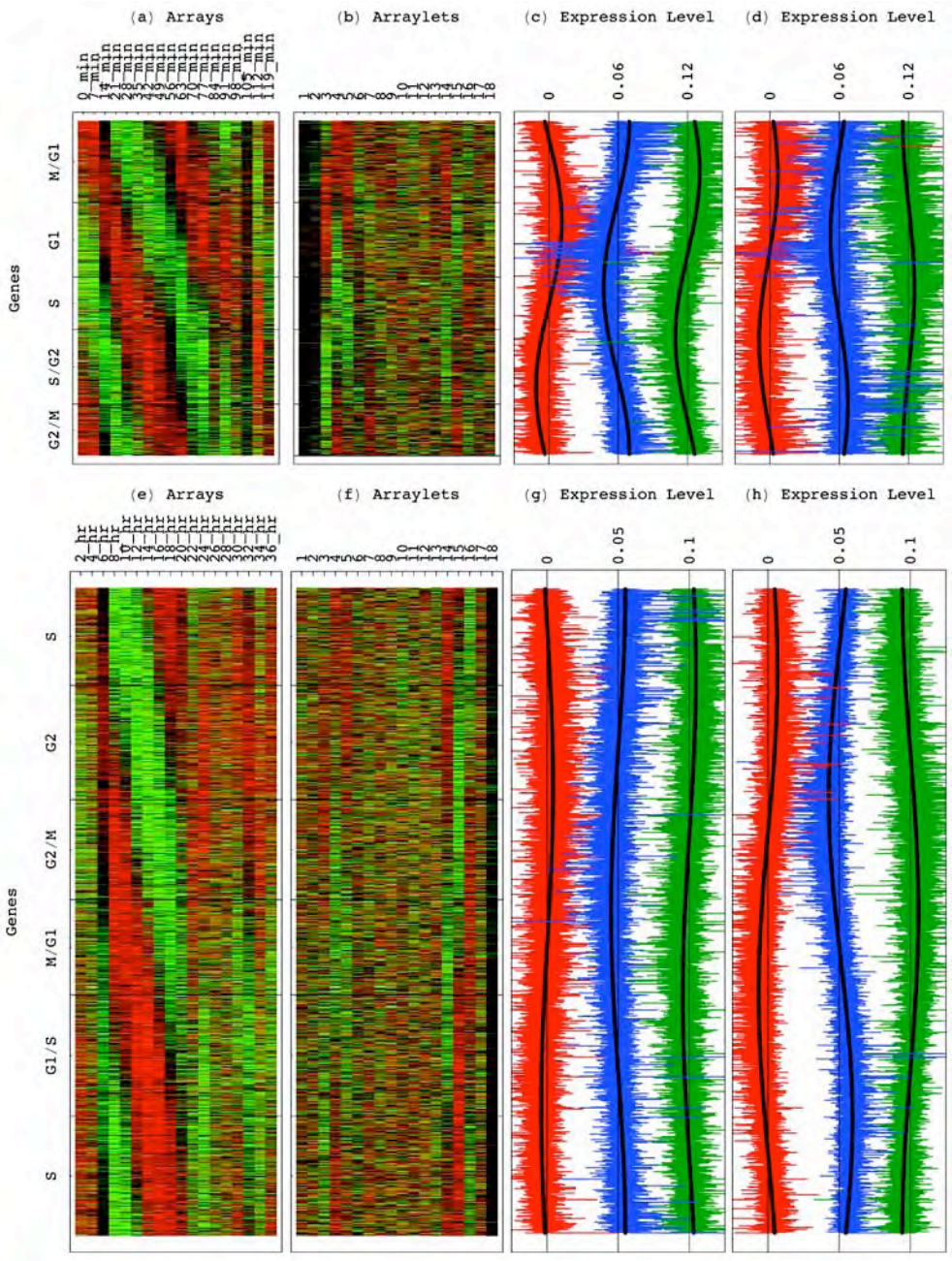
Genelets of almost no significance in one dataset relative to the other → genome exclusive processes:

Exclusive Synchronization Responses Subspaces



← *Saccharomyces cerevisiae* Human →

Simultaneous Classification in Common Cell Cycle Subspace



Saccharomyces cerevisiae

Human

→ Are there only three cellular elements or modules that drive both the yeast and human cell cycles?

→ Can we design a synthetic genetic network analogous to the digital 3-inverter ring oscillator to simulate both yeast and human cell cycles?

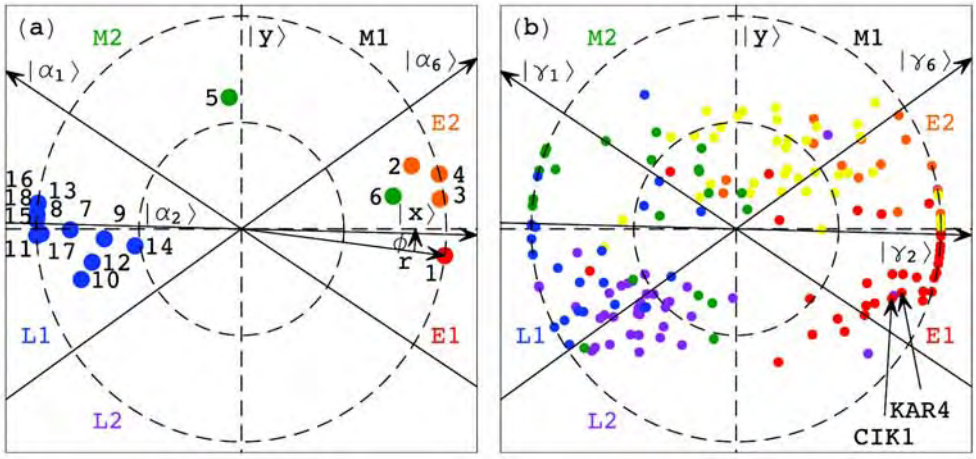
Math Operations → Biology

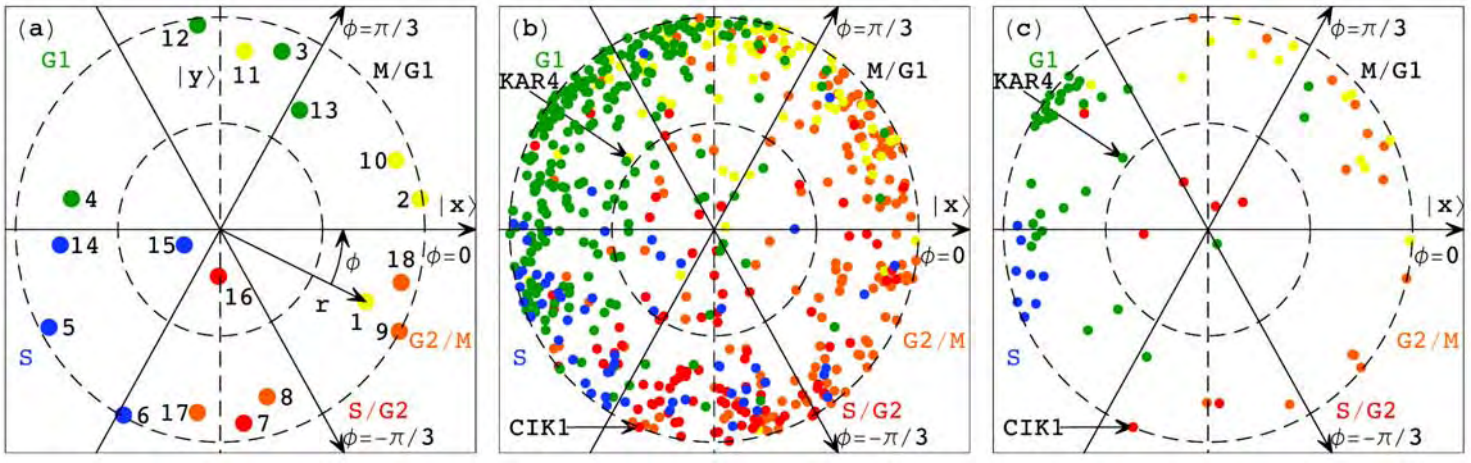
Data reconstruction in two subspaces → experimental observation of differential expression of a genome in the two cellular programs these subspaces represent:

Differential Expression in Yeast During Mating and Cell Cycle

Pheromone Synchronization Response Subspace:

KAR4 is required for CIK1 induction during mating*

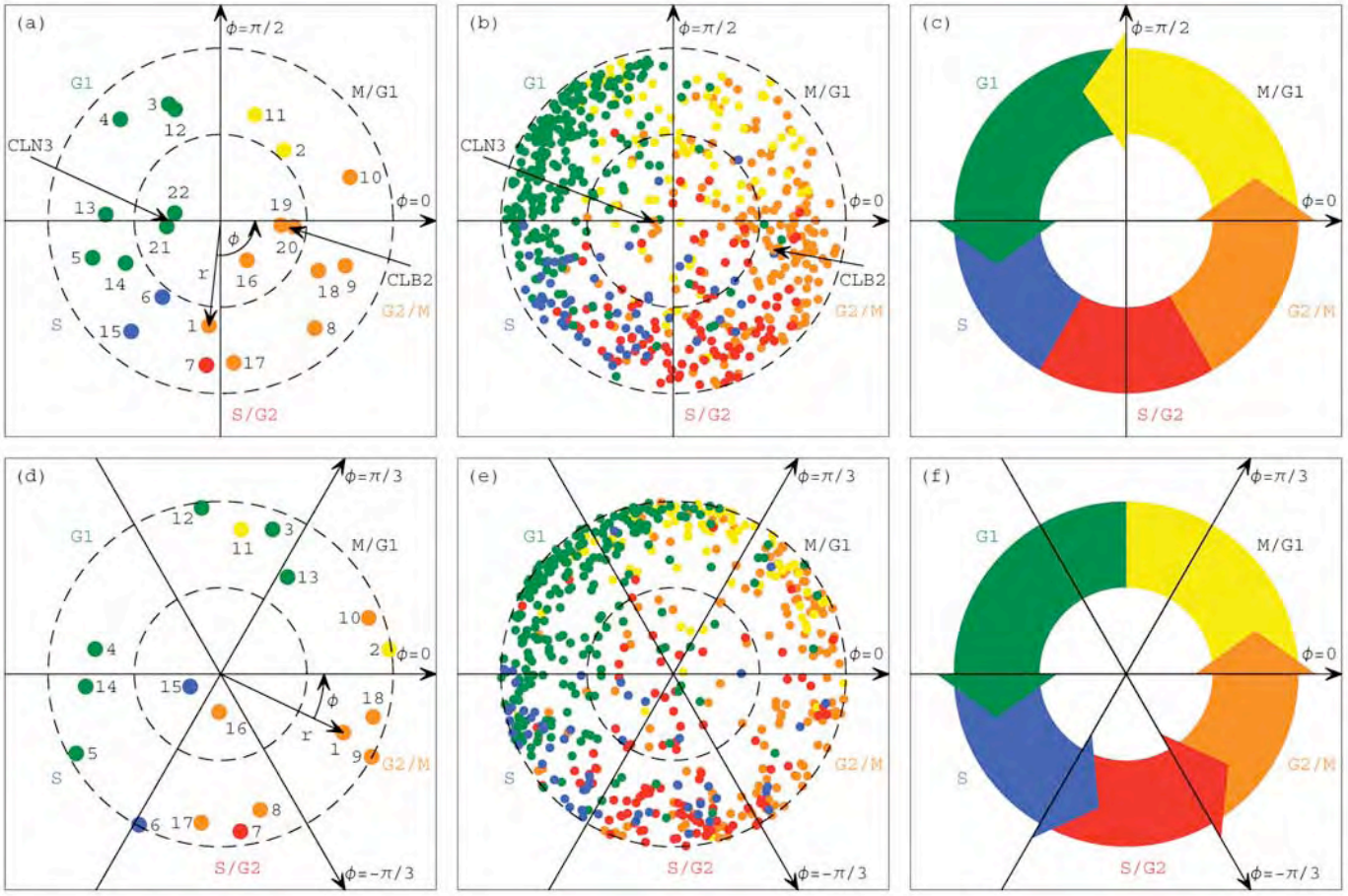




Common Cell Cycle Subspace: Mitotic expression of CIK1 during S/G2 is independent of KAR4*

*Kurihara, Stewart, Gammie & Rose, *MCB* 16, 3990 (1996).

Modeling the Yeast Cell Cycle



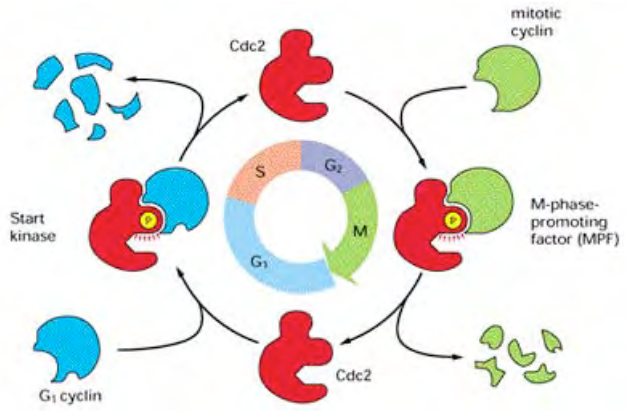
Math Variables → Biology

Eigengenes and genelets correlate with observed genome-wide effects of cell cycle regulators;

Eigenarrays and arraylets correlate with measured samples of the regulated cell cycle stages.

Math Operations → Biology

Classification maps the data onto cell cycle stages

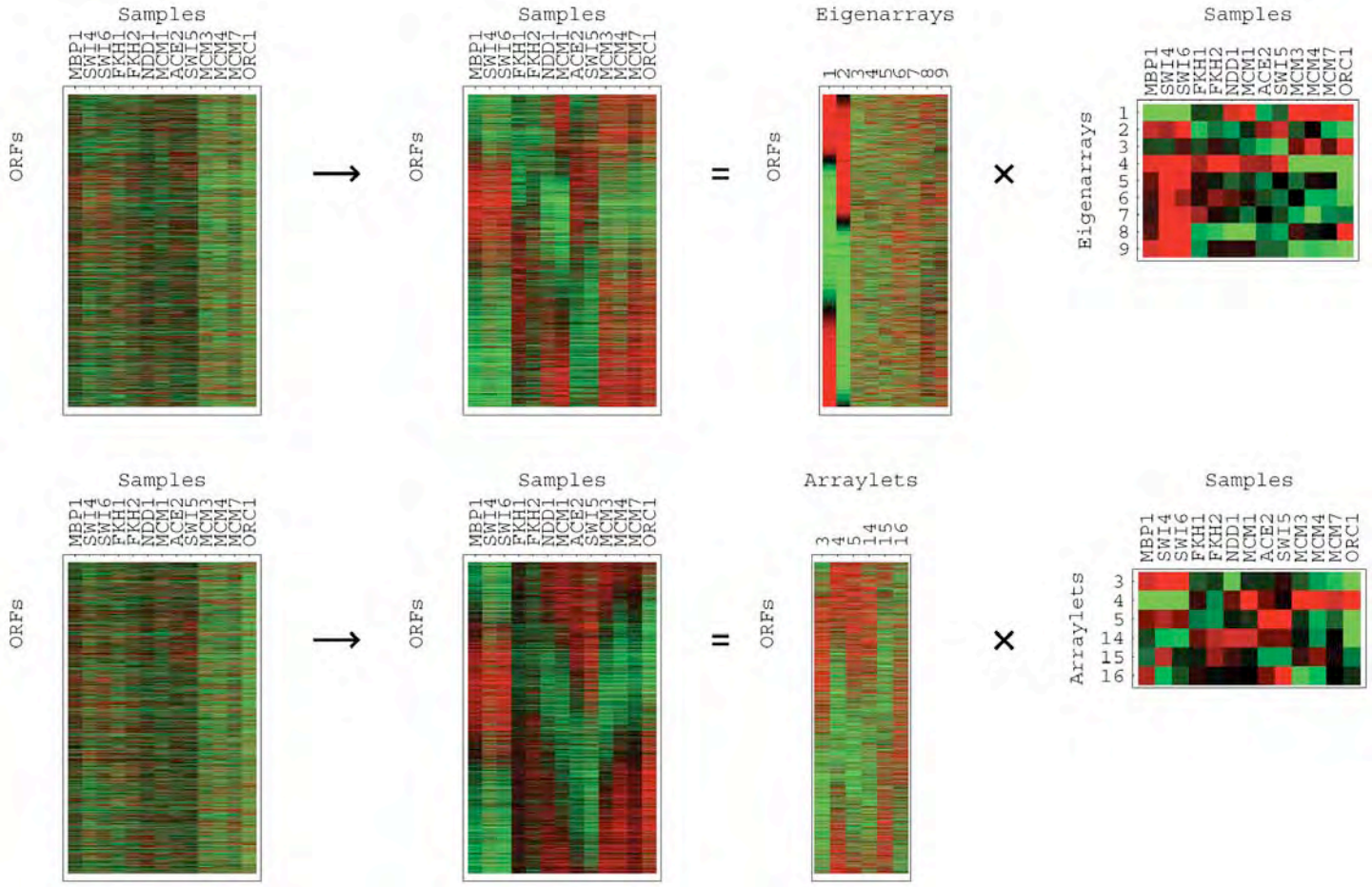


Alberts et al., Molecular Biology of the Cell (1994).

Pseudoinverse Integrative Modeling

Alter & Golub, *PNAS* 101, 16577 (2004); Alter et al., *Proc. MNBWS* 15 (2004);
http://www.bme.utexas.edu/research/orly/pseudoinverse/.

Unique linear transformation of the genome-scale data from ORFs × data arrays space to reduced basis arrays × data arrays space.



Proteins' DNA-Binding Data
Transcription Factors
Replication Initiation Proteins

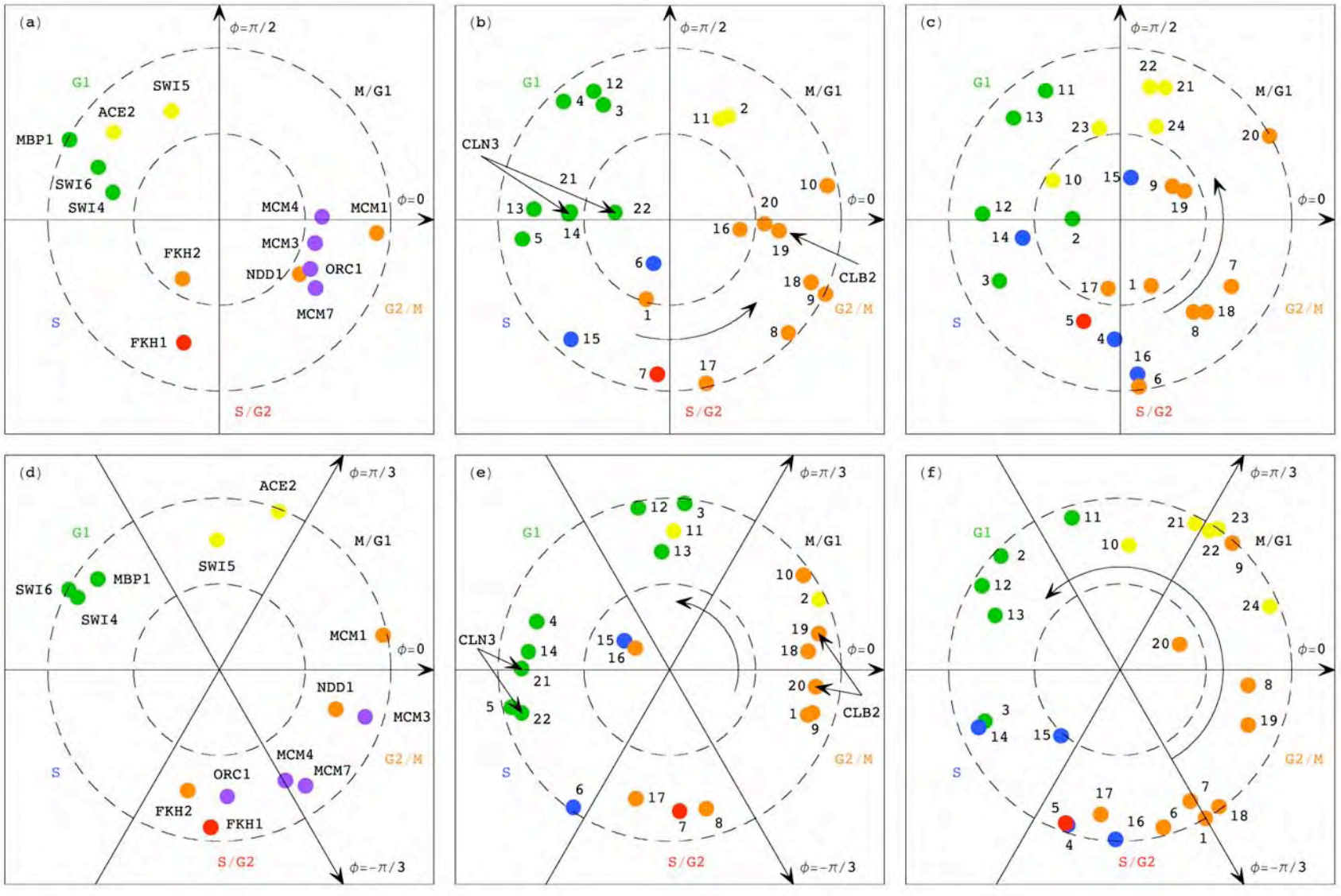
RNA Expression Bases

Simon et al., *Cell* 106, 697 (2001);
Wyrick et al., *Science* 294, 2397 (2001).

Math Operations → Biology

Classification maps reconstructed data states onto those of the basis → global picture of the causal coordination of these two sets of states.

Novel Correlation: DNA ↔ RNA



The genome-scale binding profiles of Mcm3, Mcm4, Mcm7 and Orc1 are correlated with transcription minima during the cell cycle stage G1.

→ Replication initiation requires binding of these proteins at origins of replications across the yeast genome during G1.

Diffley, Cocker, Dowell, & Rowley, *Cell* 78, 303 (1994).

→ They are involved with transcriptional silencing at the yeast mating loci.

Micklem et al., *Nature* 366, 87 (1993).

Either one of two previously unknown mechanisms of regulation may be underlying this correlation:

→ Replication may regulate transcription:
The binding of ORC and MCM proteins, which is known to be required for initiation of replication at origins across the yeast genome, represses, and possibly inhibits the transcription of genes that are located near the origins.

→ Transcription may regulate replication:
The transcription of genes at G1 reduces the efficiency of origins that are located near the transcribed genes.

This is the first time that a data-driven mathematical model has been used to predict a biological principle that is truly on a genome scale.

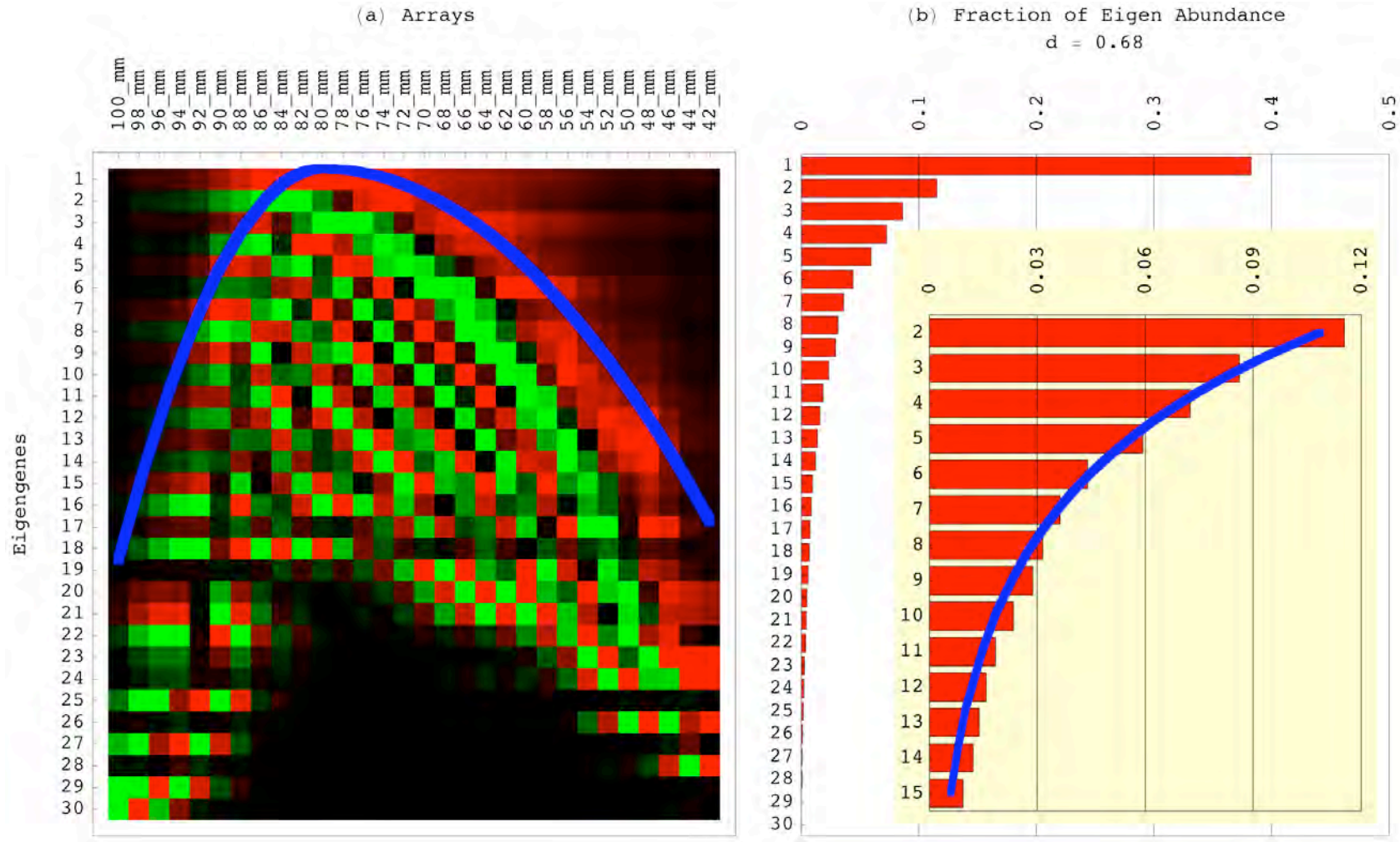
Predicting a Physical Principle:

Previously Unknown Asymmetry in mRNA Abundance Levels Profiles of Genes Across Gel Slices

Might Be Due to a Previously Unknown Asymmetry in the Thermal Broadening of a Moving Band of mRNA Molecules

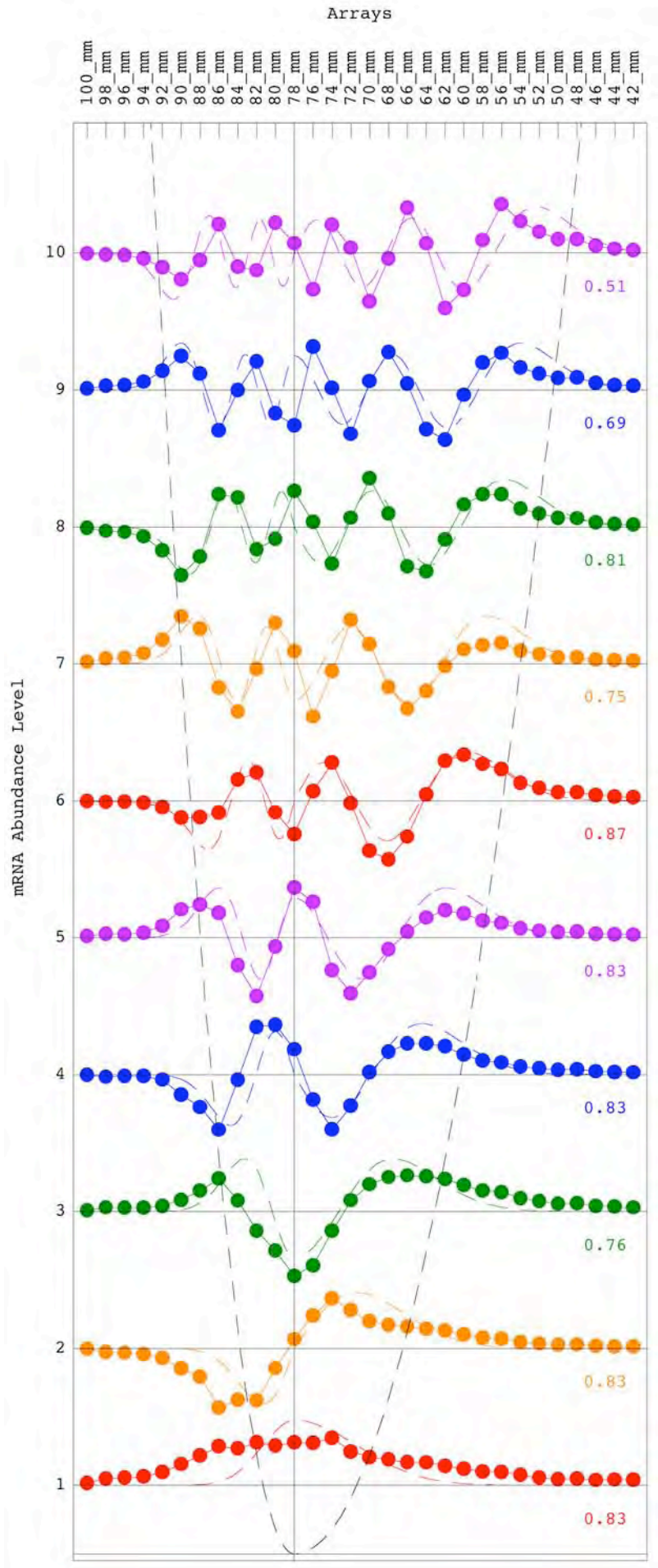
SVD Modeling Reveals Asymmetric Band Broadening in RNA Gel Electrophoresis

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



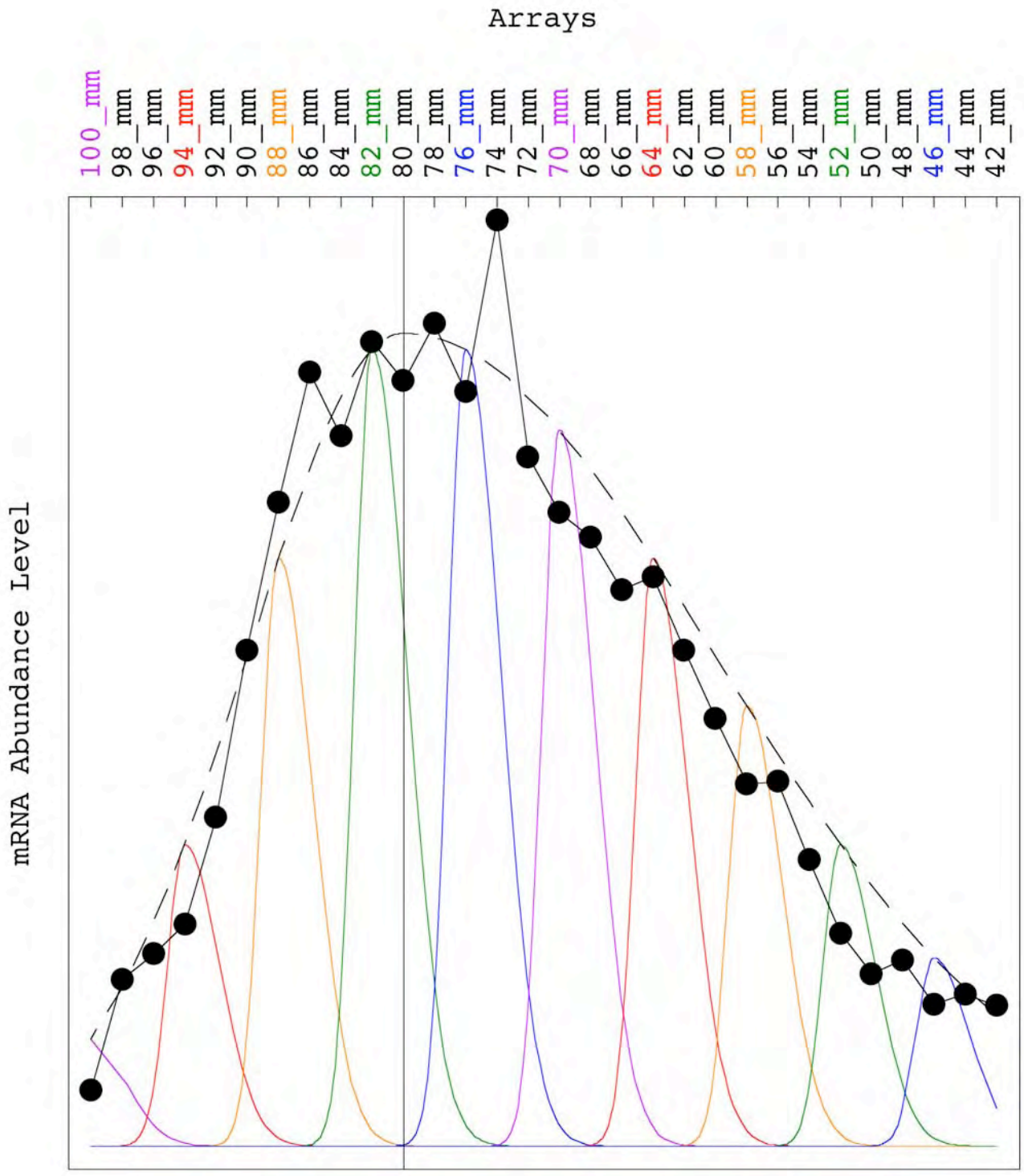
Hurowitz & Brown, *Genome Biology* 5, R2 (2003).

Fractions of Eigenabundance Fit a Geometric Series



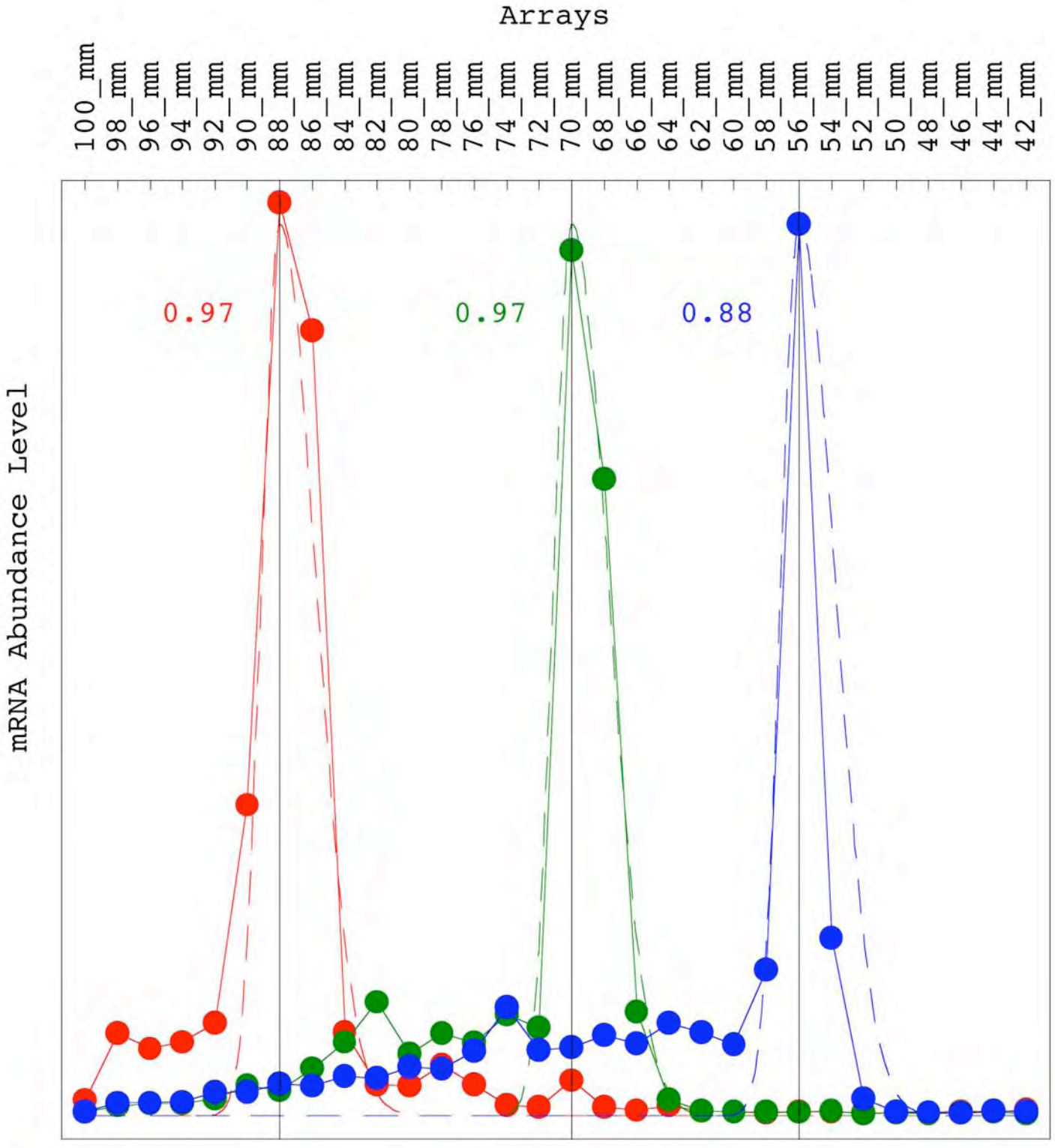
Eigengenes fit "Asymmetric" Hermite Functions

“Asymmetric” Generalized Coherent State Model of Genome-Scale mRNA Lengths Distribution

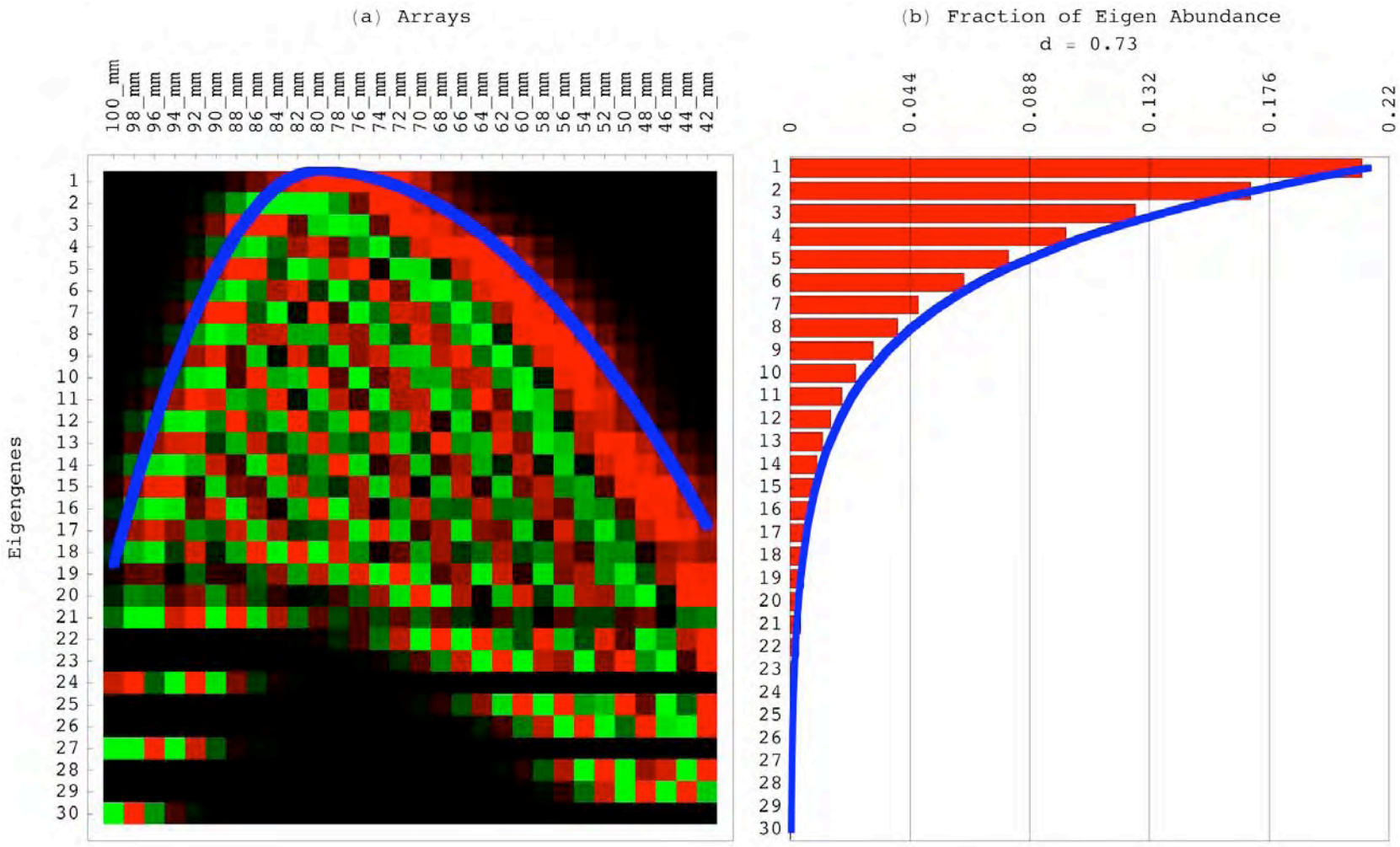


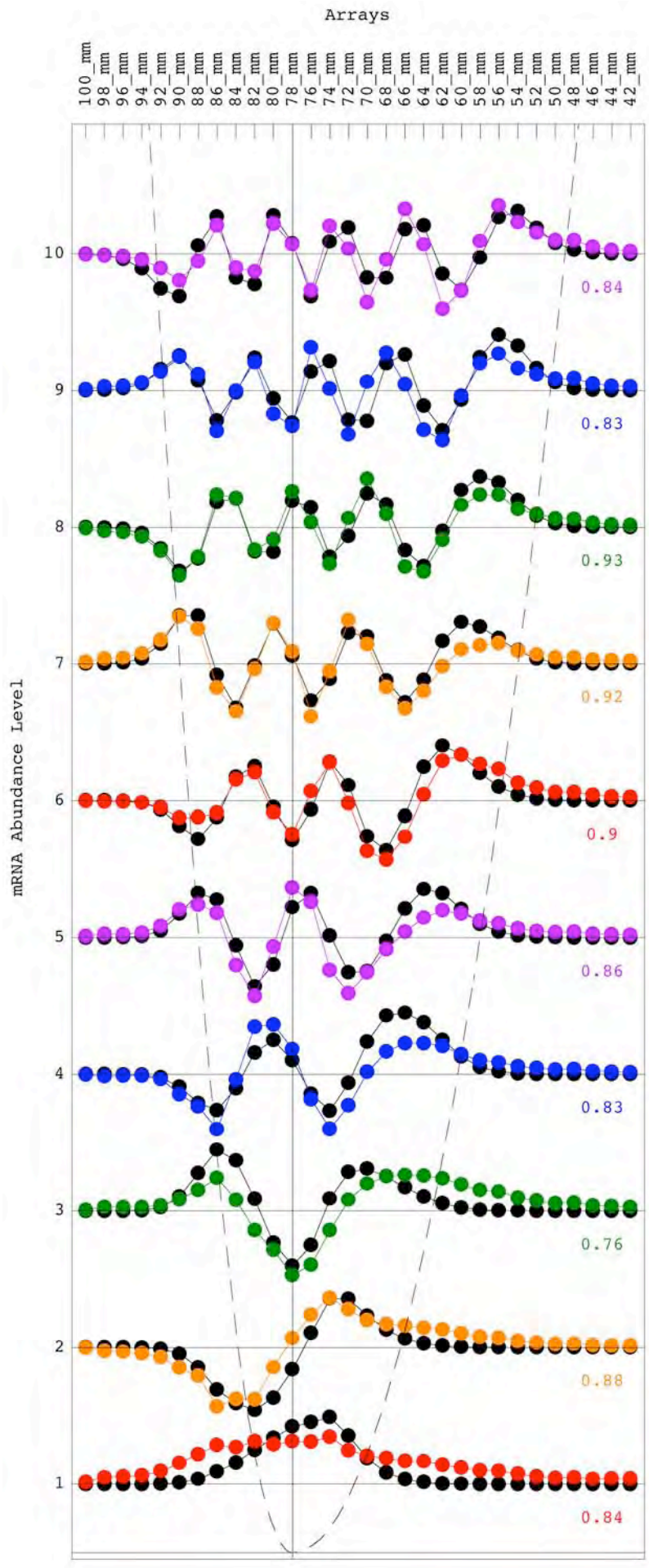
Distribution of the Peaks of the Genes’ Profiles Fits an Asymmetric Gaussian

Profiles of mRNA Abundance Levels of Most Genes Fit Asymmetric Gaussians



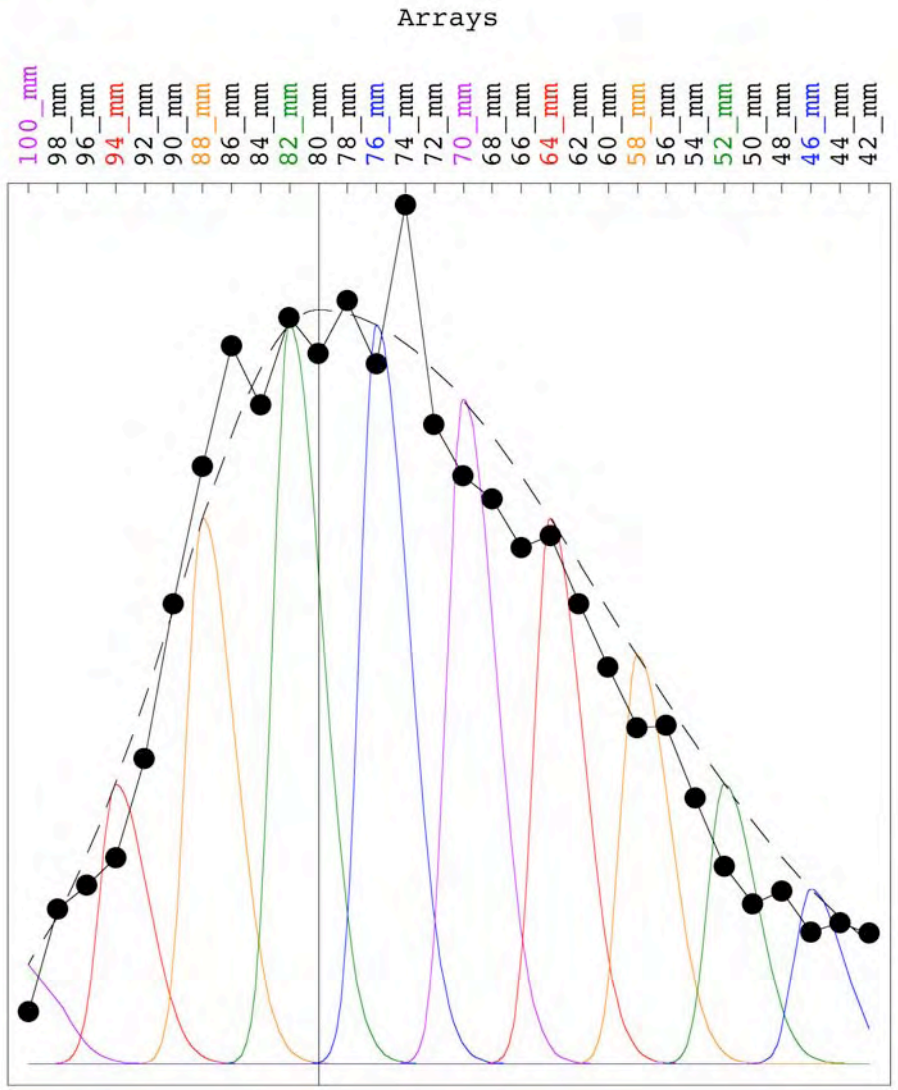
Genome-Scale mRNA Lengths Distribution Fits an Approximated Asymmetric Generalized Coherent State





Genome-Scale mRNA Lengths Distribution Fits an Approximated Asymmetric Generalized Coherent State

Why Does the Distribution of the Peaks of the Genes' Profiles Fit an Asymmetric Gaussian?

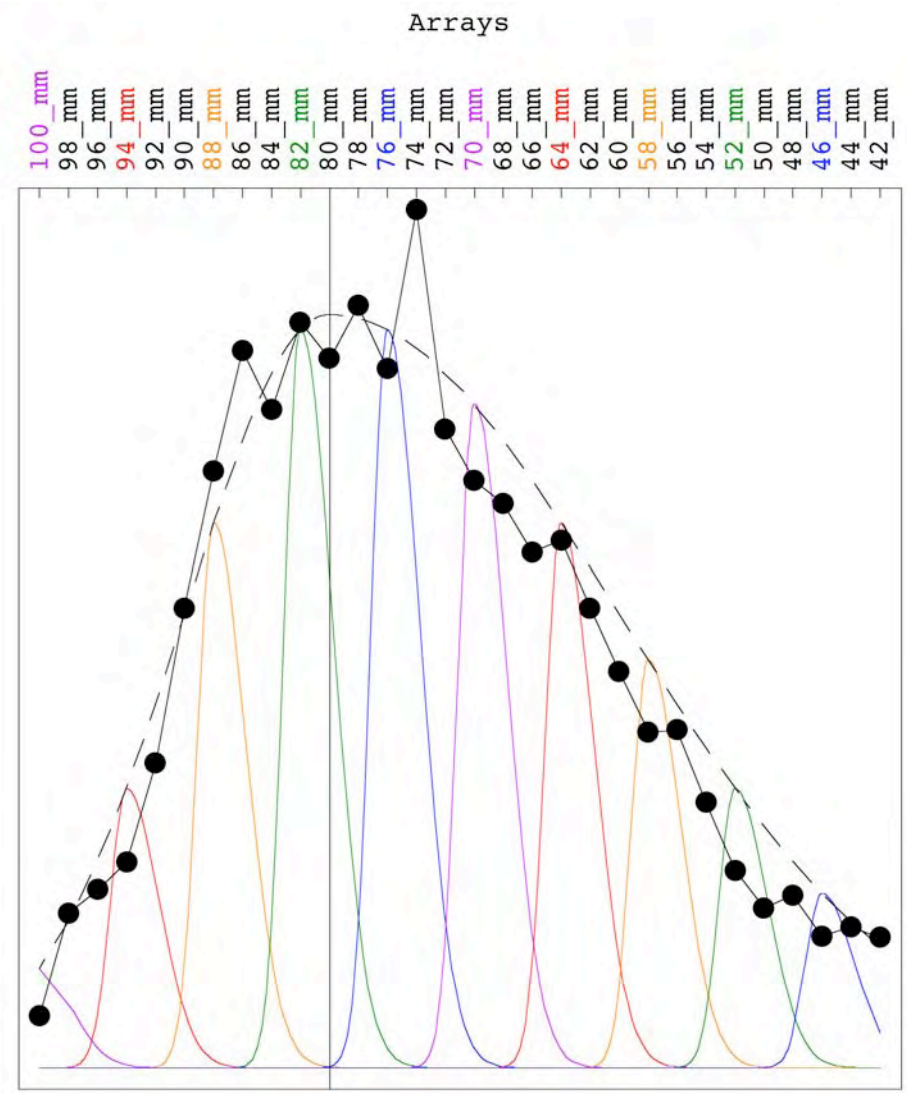


Hypothesis:

Two competing evolutionary forces determine the distribution of mRNA gene transcripts.

These forces are linearly proportional to and oppositely directed to the displacement from the equilibrium gel migration length, in the manner of the restoring force of the harmonic oscillator.

Why Do the Profiles of Most Genes Fit Asymmetric Gaussians?

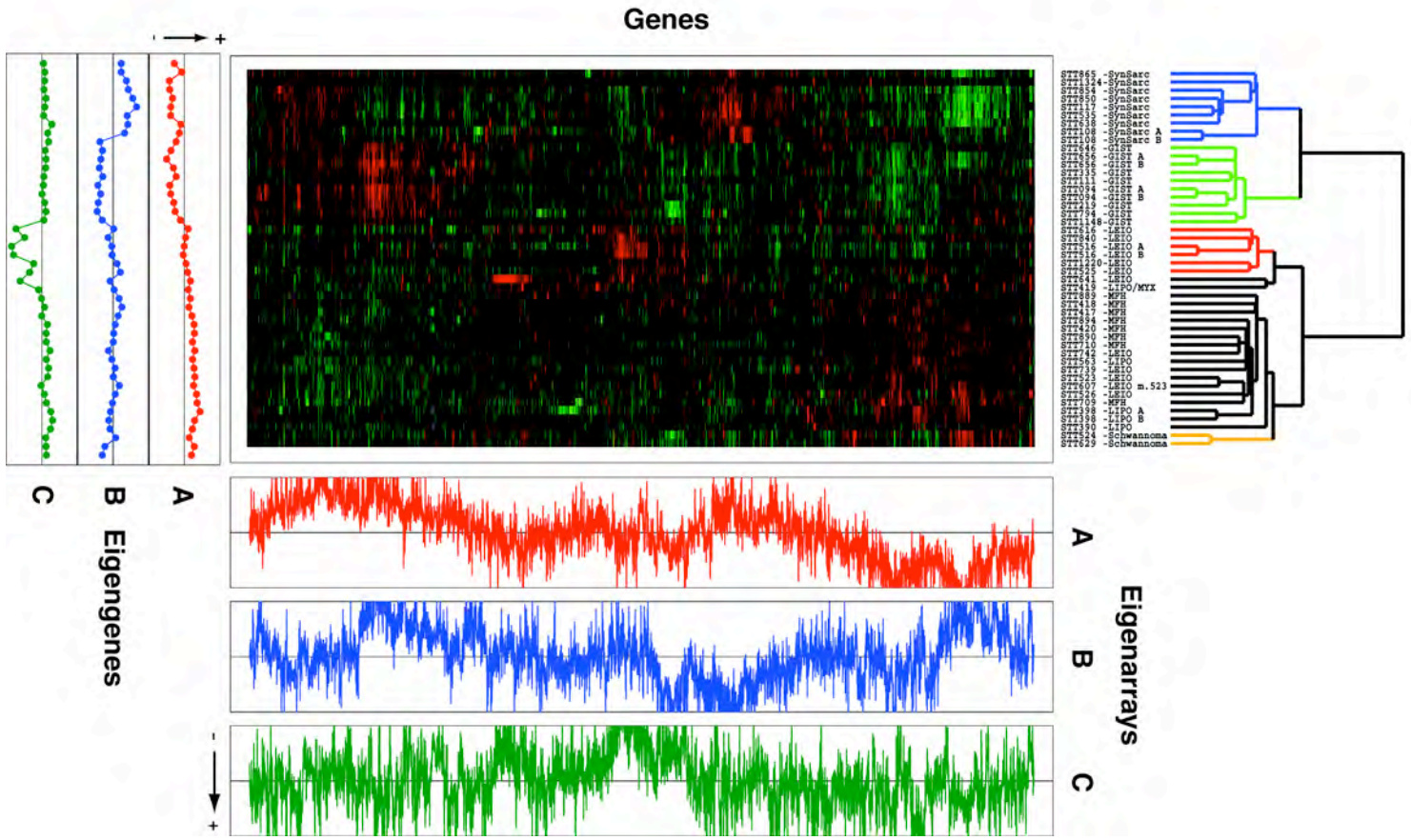


Prediction:

In the thermal broadening of a moving band of RNA molecules the peak of the band is moving toward the front of the band and away from its back.

Previous simulations and measurements of DNA band broadening in gel electrophoresis have shown that the broadening of a moving band can be different from that of a stationary band, but have not suggested asymmetry.

Medical Applications of DNA Microarray Data: Diagnosis, Treatment and Drug Development



SVD normalization and classification of tumor data uncover a novel subtype of leiomyosarcomas that express a group of muscle genes.

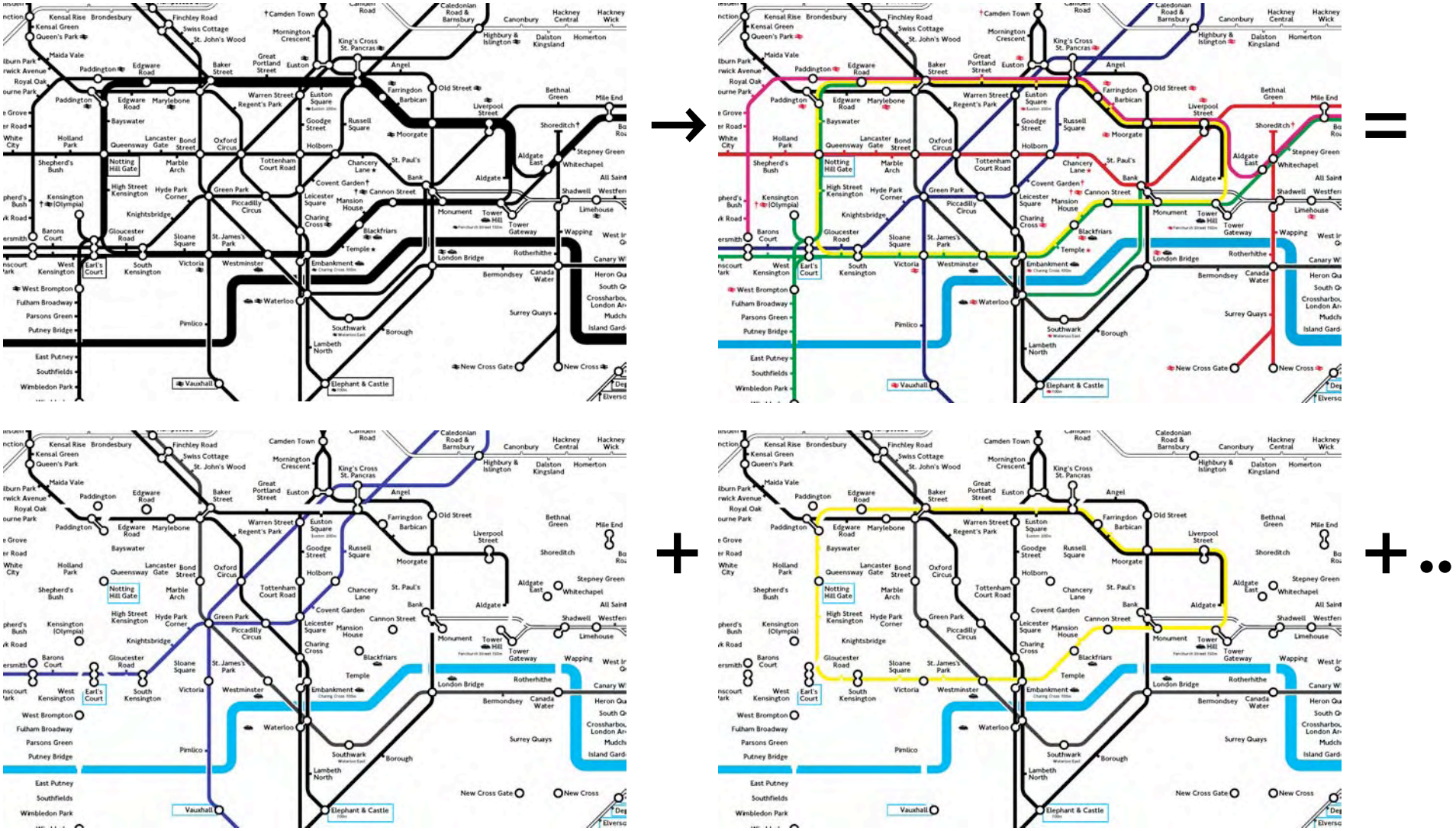
Nielsen, West, Linn, Alter et al., *Lancet* 359, 1301 (2002).

Future Algorithms:

Large-Scale Molecular Biological Data are of Higher Orders

Higher-Order Algorithms Are Needed for Comparative and Integrative Data Modeling

Networks are Tensors of "Subnetworks"

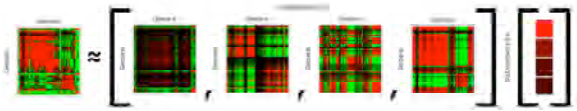


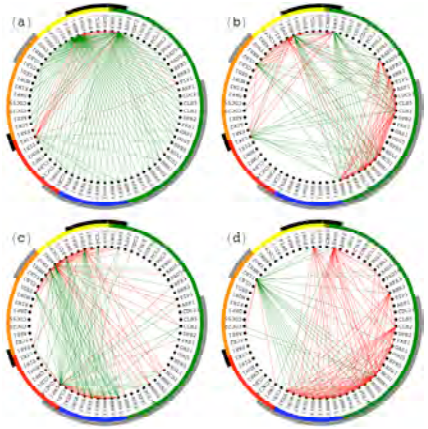
The relations among the activities of genes, not only the activities of the genes alone, are known to be pathway-dependent, i.e., conditioned by the biological and experimental settings in which they are observed.

Tensor Models for Networks of Correlations Computed from Genomic Data

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

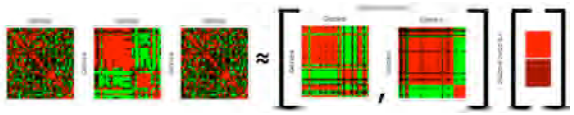
EVD Modeling





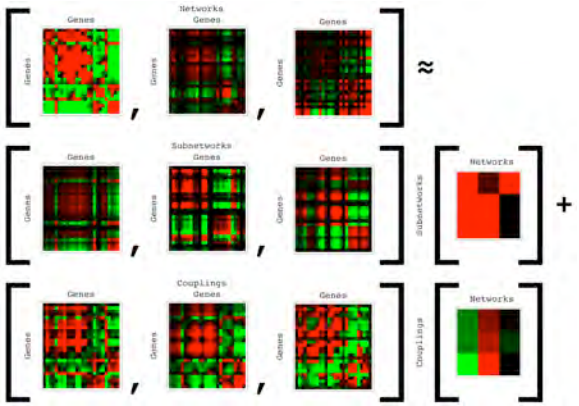
Uncover Pathways in a Single Network

Pseudoinverse Integrative Modeling



Uncover Pathways Common to Two Networks

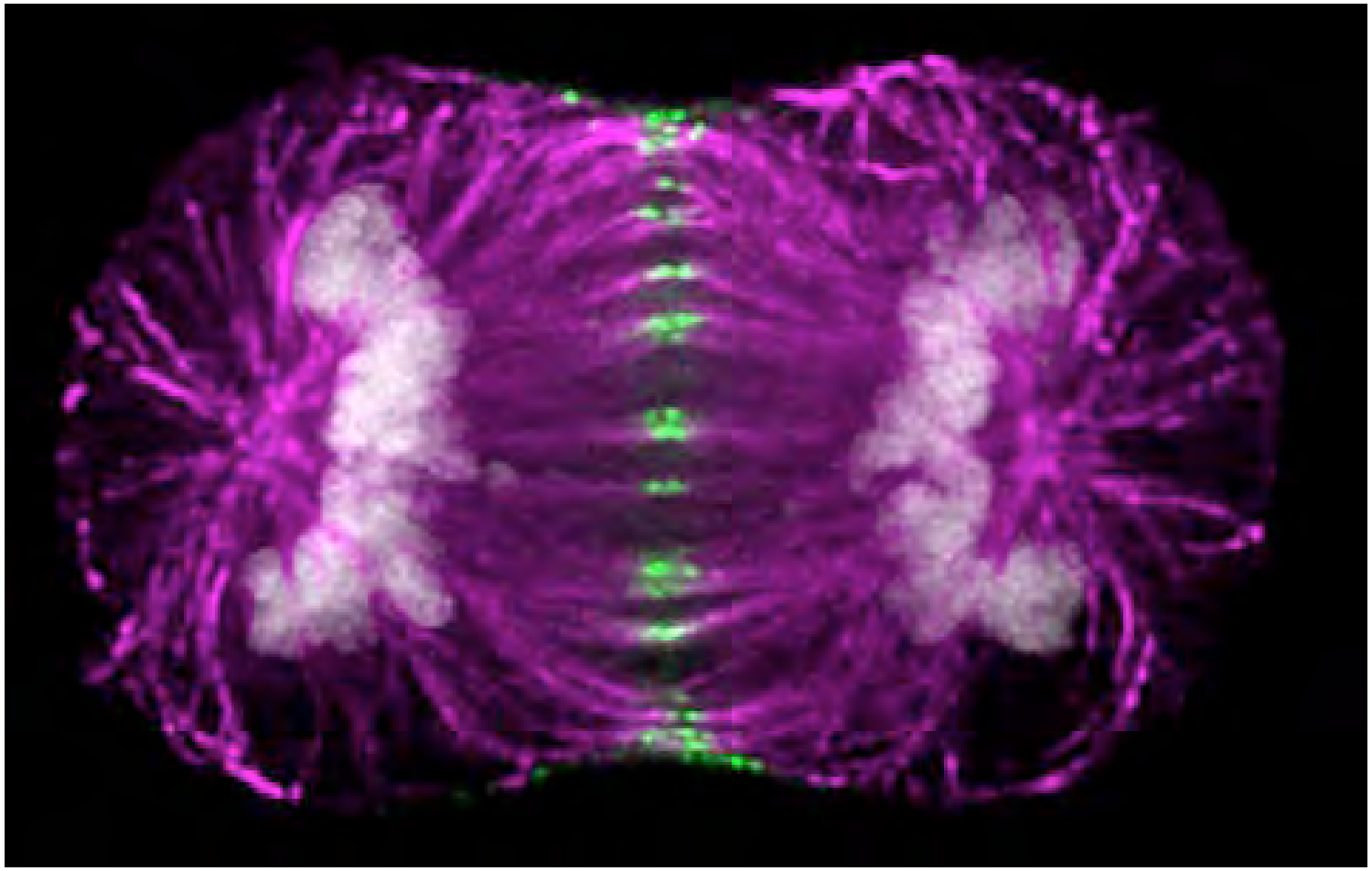
HOEVD Comparative Modeling



Uncover Pathways Common or Exclusive Among Multiple Networks

Future Data Management Tools:

In the future, cellular processes could be controlled in real time and in vivo.

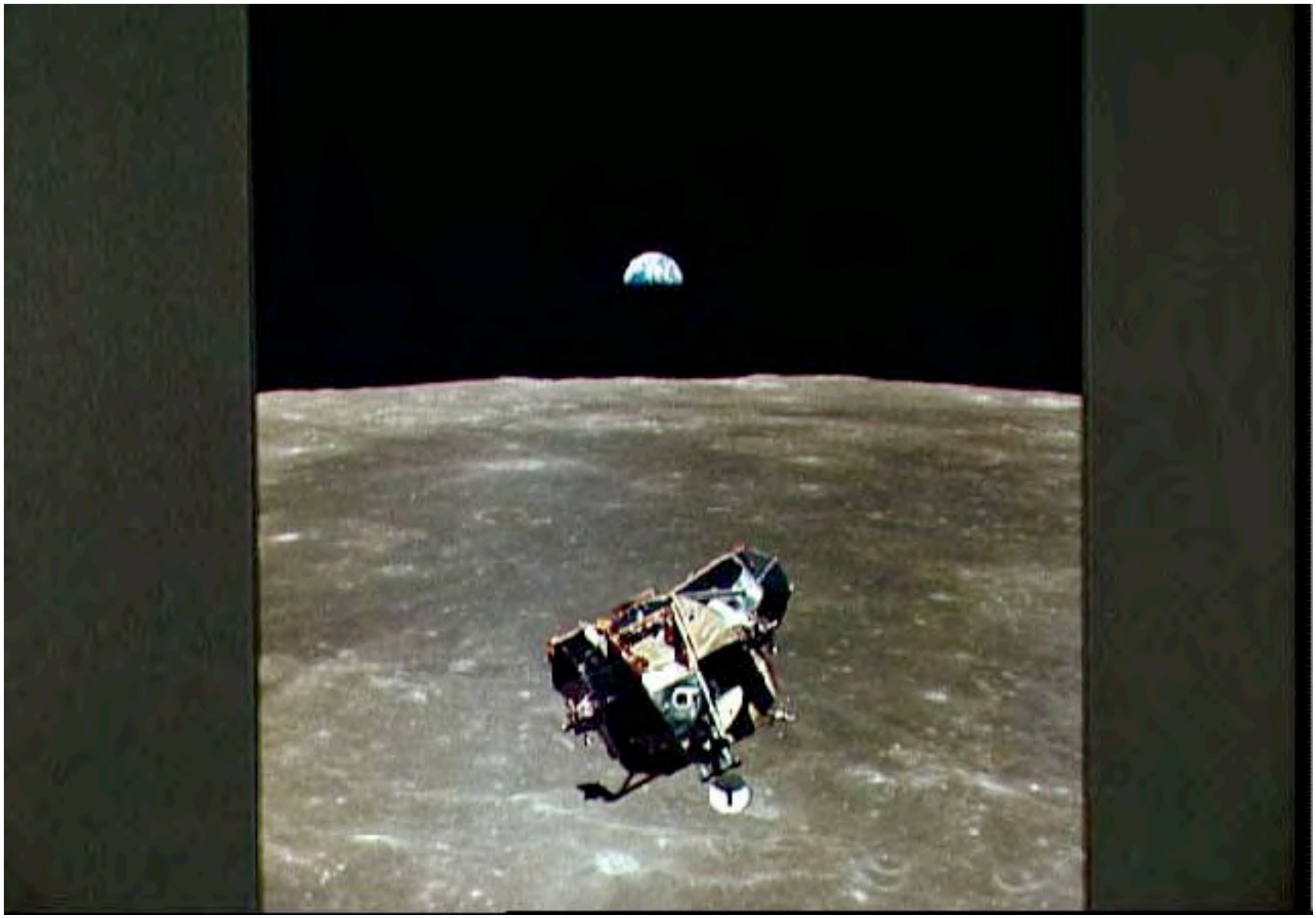


Cancer and disease could be stopped or reversed.

Damaged tissues could be engineered to regenerate.

Aging could be slowed or even halted altogether.

Today, NASA can control the trajectories of its spacecraft…



… because their motion is understood and can be predicted.

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:

NHGRI Individual Development Award in Genomic Research and Analysis

And, thank you!!!