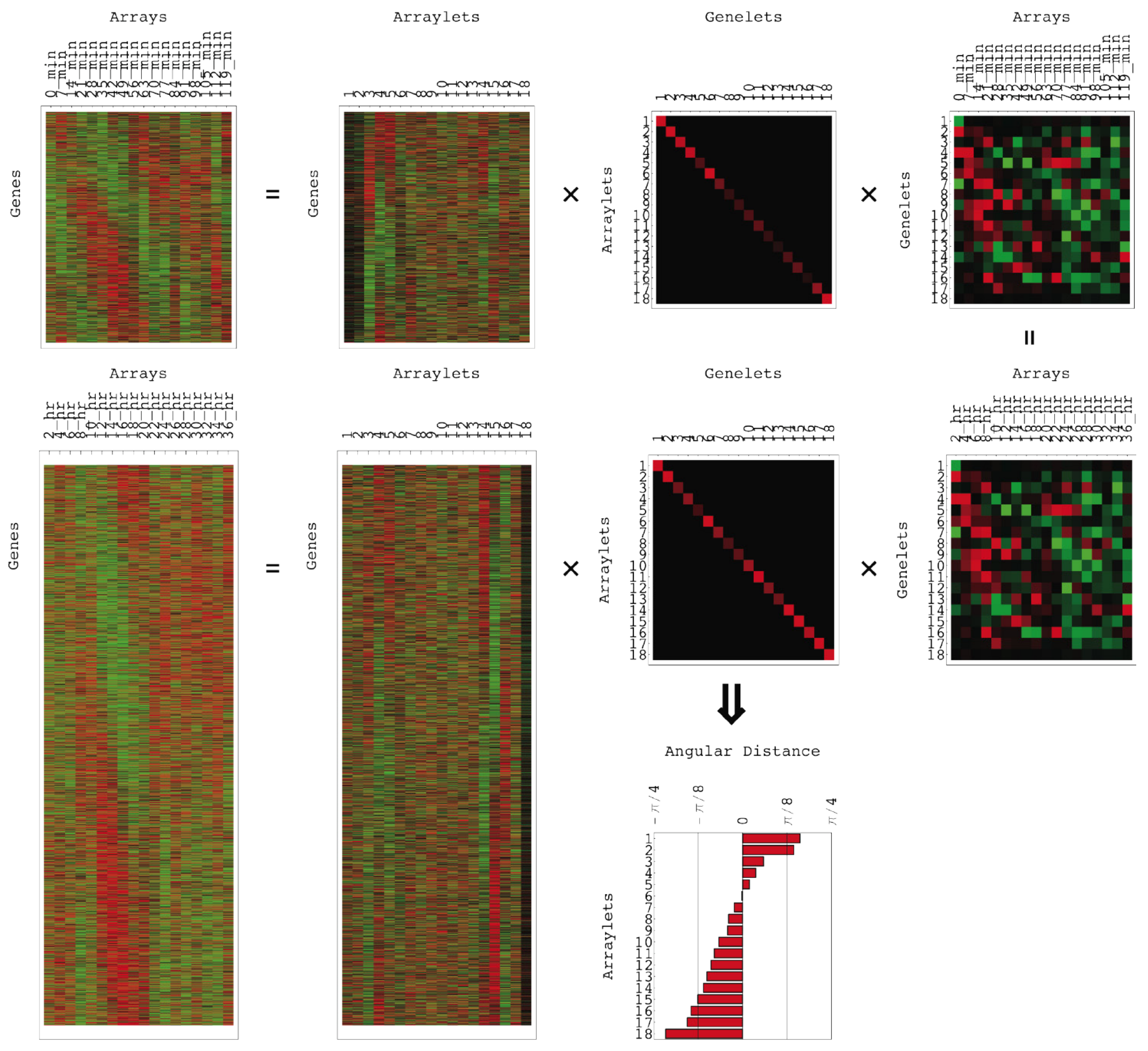
A Higher-Order Decomposition for Comparison of Multiple Large-Scale Datasets

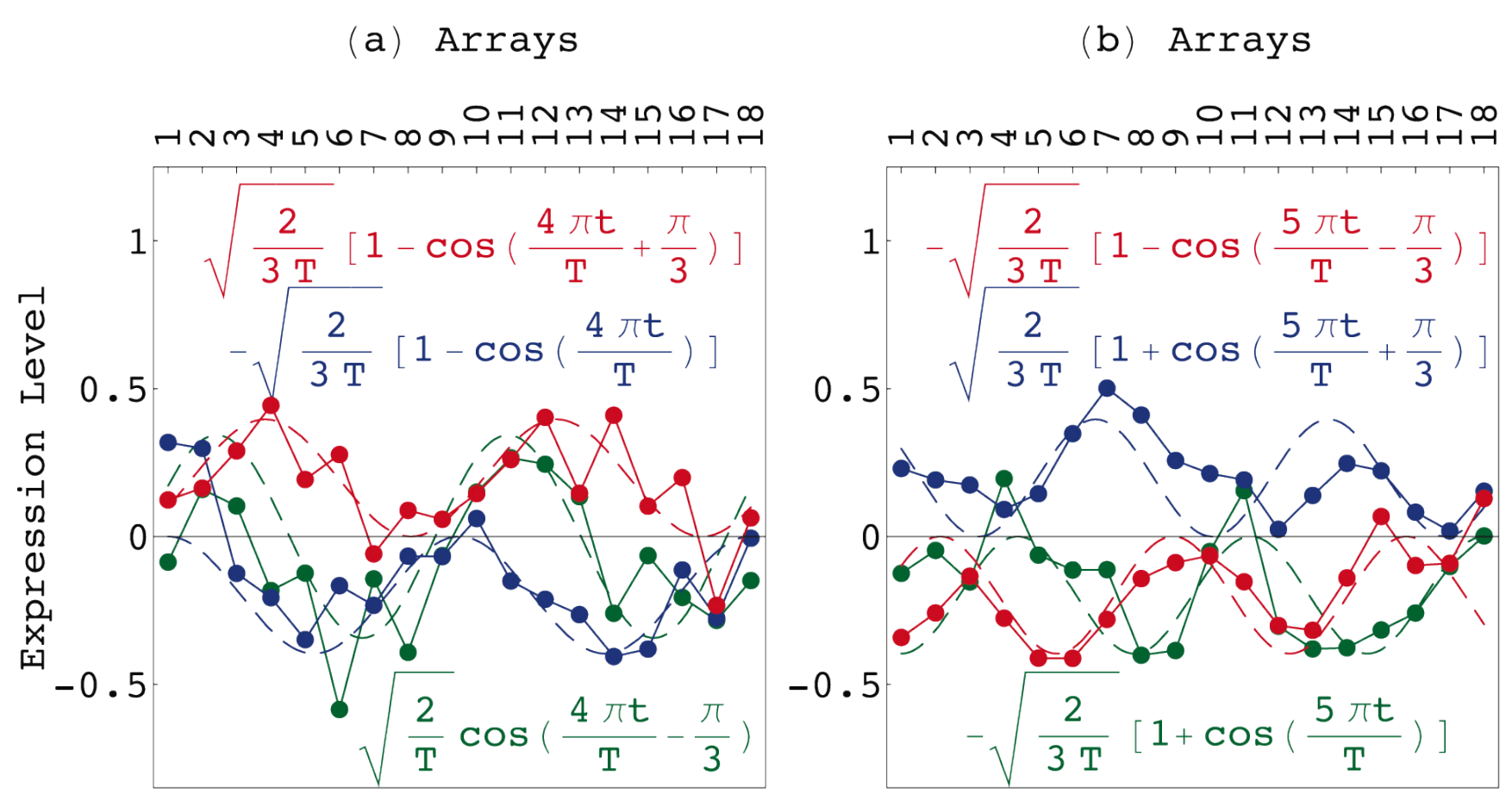
Sri Priya Ponnapalli *(University of Texas at Austin)*, Michael A. Saunders *(Stanford University)*,
Charles F. Van Loan *(Cornell University)* and Orly Alter *(University of Utah)*

The Problem

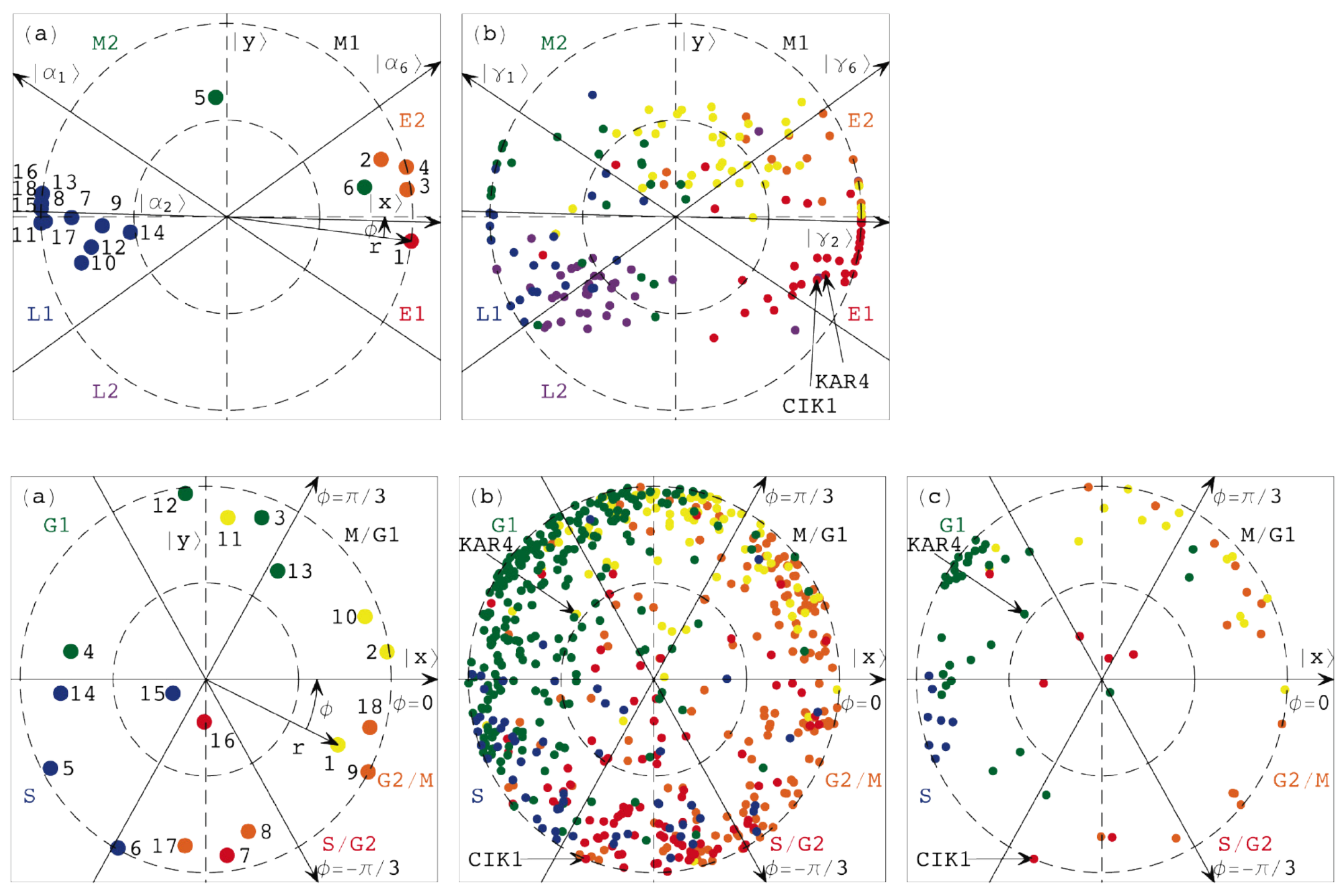
The number of high-dimensional datasets recording multiple aspects of a single phenomenon is increasing in many areas of science, accompanied by a need for mathematical frameworks that can compare multiple large-scale matrices with different row dimensions. The only such framework to date, the generalized singular value decomposition (GSVD), is limited to two matrices.



It was shown that the GSVD can be formulated as a mathematical framework for sequence-independent comparison of DNA microarray data from two organisms, where the mathematical variables and operations represent experimental and biological reality.



The variables, significant subspaces that are common to both or exclusive to either one of the datasets, correlate with cellular programs that are conserved in both or unique to either one of the organisms, respectively.

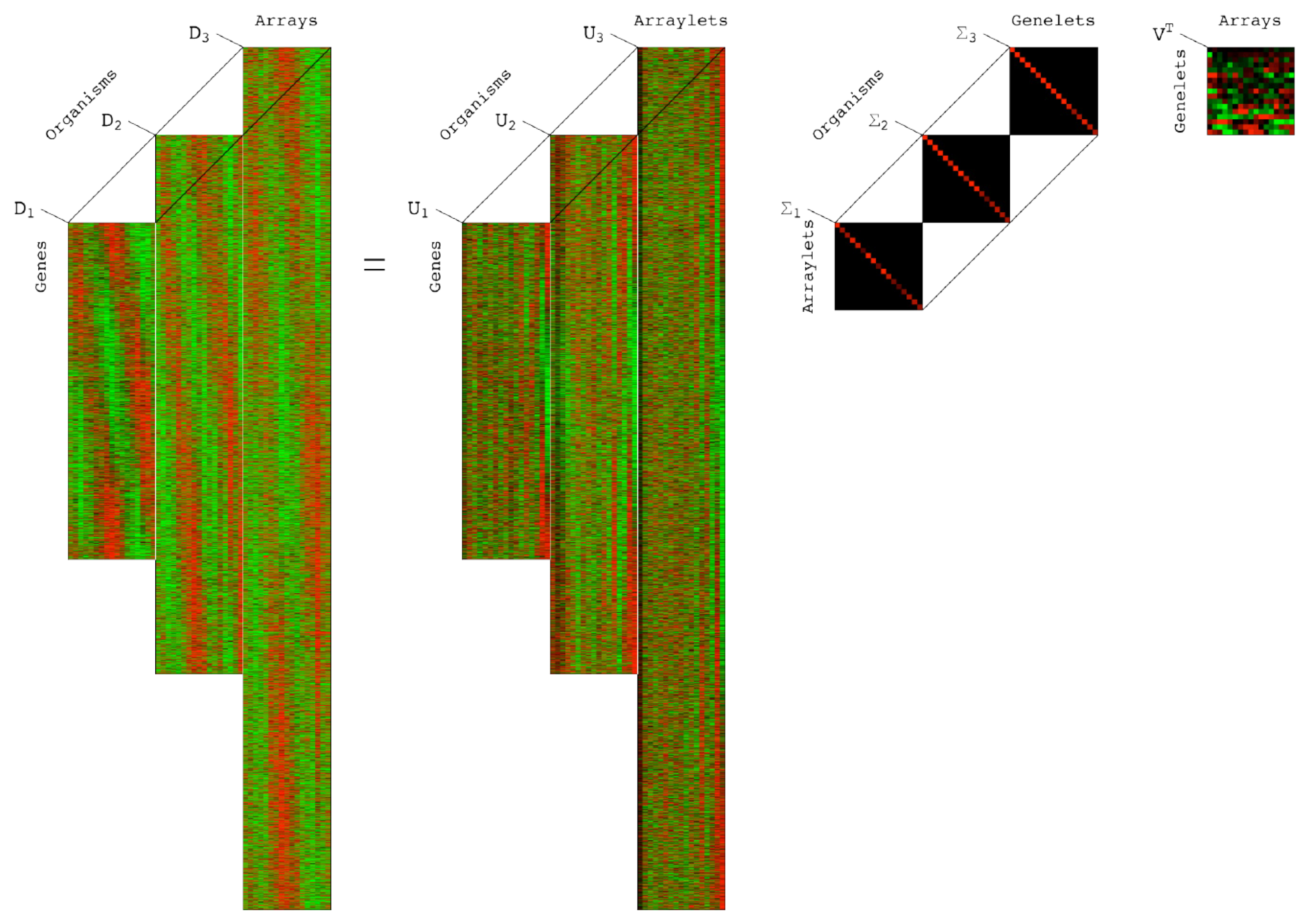


Reconstruction in the common and exclusive subspaces of either dataset outlines the differential regulation of the conserved relative to the unique programs in the corresponding organism.

Alter, Brown & Botstein *PNAS* 103, 3351 (2003).

The Solution: HO GSVD

We mathematically define a higher-order GSVD (HO GSVD).



$$D_i = U_i \Sigma_i V^T, \qquad \Sigma_i = \mathrm{diag}(\sigma_{i,k})$$

The matrix V, identical in all factorizations, is obtained from the balanced eigensystem of S, which does not depend upon the ordering of D_i.

$$SV = V\Lambda$$

$$S \equiv \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j>i}^{N} (A_i A_j^{-1} + A_j A_i^{-1})$$

$$A_i = D_i^T D_i$$

The matrices D_i are assumed to be with full column rank.

We prove that this exact decomposition extends to higher orders almost all of the mathematical properties of the GSVD:

Supplementary Theorems 1–5:
For N=2, our HO GSVD leads algebraically to the GSVD.

Theorem 1:
S has n independent eigenvectors, and the eigenvectors and eigenvalues of S are real.

Theorem 2:
The eigenvalues of S satisfy $\lambda_k \geq 1$.

Theorem 3:
The common HO GSVD subspace: An eigenvalue satisfies λ_k=1 if and only if the corresponding right basis vector v_k is of equal significance in all matrices D_i and D_j, i.e., $\sigma_{i,k}/\sigma_{j,k}$=1 for all i and j, and the corresponding left basis vector $u_{i,k}$ is orthonormal to all other left basis vectors in U_i for all i.

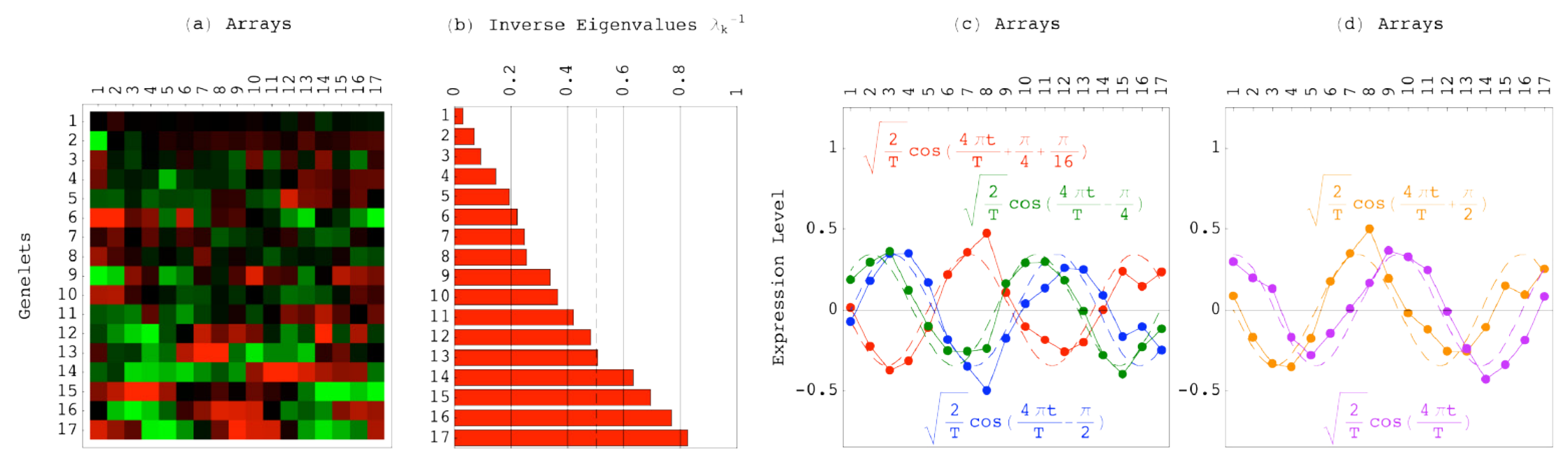
Corollary 1:
An eigenvalue satisfies λ_k=1 if and only if the corresponding right basis vector v_k is a generalized singular vector of all pairwise GSVD factorizations of the matrices D_i and D_j with equal corresponding generalized singular values for all for all i and j.

Supplementary Theorem 6 and Conjecture 1:
A role in iterative approximation algorithms.

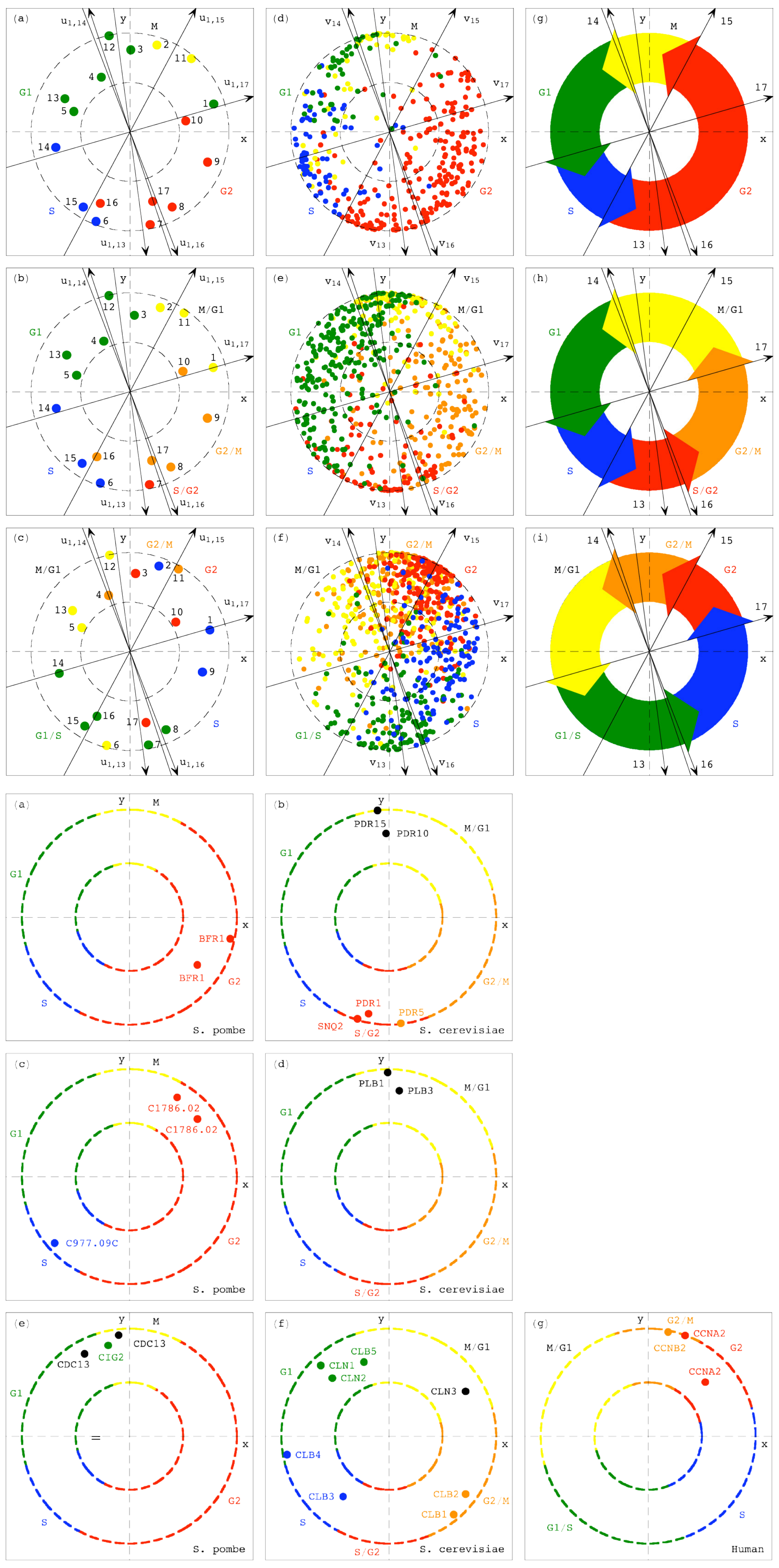
Ponnapalli, Saunders, Van Loan & Alter, under review.

Comparison of DNA Microarray Data from Multiple Organisms

We illustrate the HO GSVD with a comparison of genome-scale cell-cycle mRNA expression from *S. pombe*, *S. cerevisiae* and human. Unlike existing algorithms, a mapping among the genes of these disparate organisms is not required. We find that the approximately common HO GSVD subspace represents the cell-cycle mRNA expression oscillations, which are similar among the datasets.



Simultaneous reconstruction in the common subspace removes the experimental artifacts, which are dissimilar, from the datasets.



In the simultaneous sequence-independent classification in this common subspace, genes of highly conserved sequences across the organism but significantly different cell-cycle peak times are correctly classified.

This research was supported by the Utah Science Technology and Research (USTAR) Initiative, National Human Genome Research Institute R01 Grant HG-004302 and National Science Foundation CAREER Award DMS-0847173.

SCI INSTITUTE • EXHIBIT • EXPLORE • EXCITE • EXPERIENCE • EXCHANGE

THE UNIVERSITY OF UTAH

www.sci.utah.edu