

**Artificial Intelligence and Machine Learning:
From Research to the Cancer Clinic**



Discovering Genome-Scale Predictors of Survival and Response to Treatment with Multi-Tensor Decompositions

Orly Alter, Ph.D.

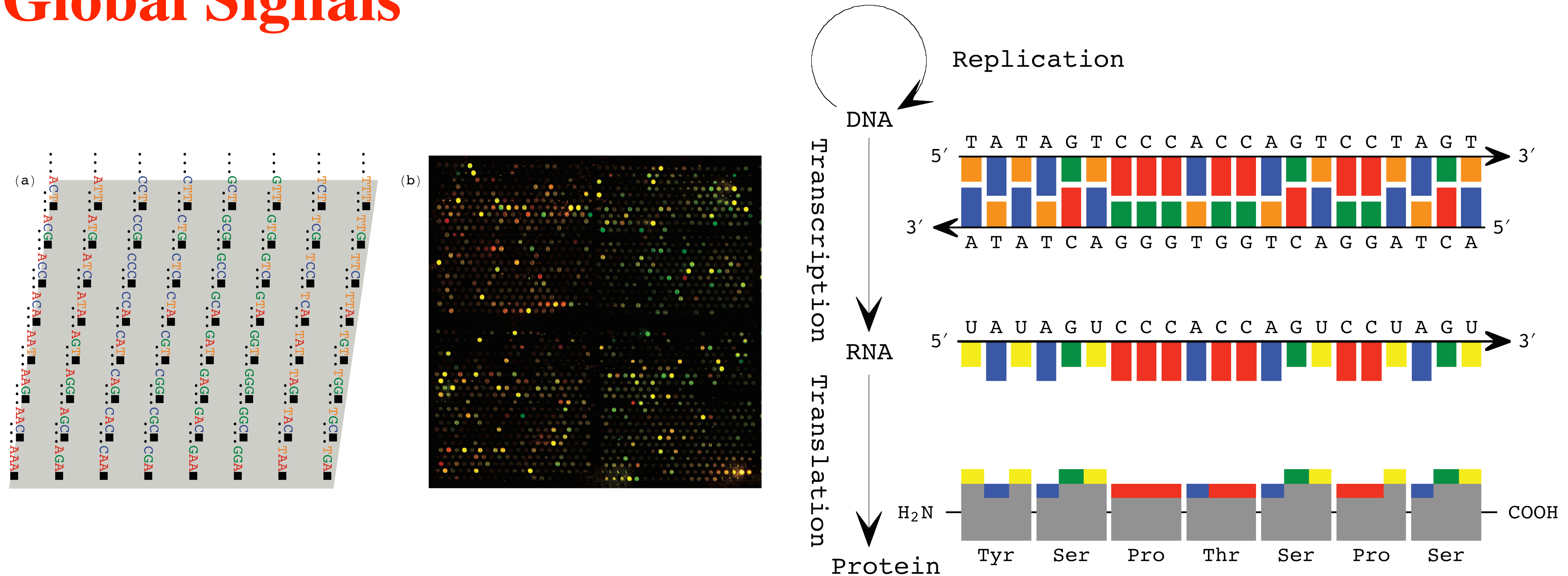
**USTAR Associate Professor of Bioengineering and Human Genetics,
Scientific Computing and Imaging Institute and Huntsman Cancer Institute,
University of Utah
orly@sci.utah.edu
<https://alterlab.org/>**

Disclosures

**I am a co-founder of and an equity holder in Eigengene, Inc.
This does not alter my adherence to the policies of the AACR on
sharing data, code, and materials.**

**Orly Alter, Ph.D.
USTAR Associate Professor of Bioengineering and Human Genetics,
Scientific Computing and Imaging Institute and Huntsman Cancer Institute,
University of Utah
orly@sci.utah.edu
<https://alterlab.org/>**

High-Throughput Biotechnologies Record Global Signals



Replication

DNA

Transcription

RNA

Translation

Protein

5' T A T A G T C C C A C C A G T C C T A G T 3'

3' A T A T C A G G G T G G T C A G G A T C A 5'

5' U A U A G U C C C A C C A G U C C U A G U 3'

H₂N

Tyr Ser Pro Thr Ser Pro Ser

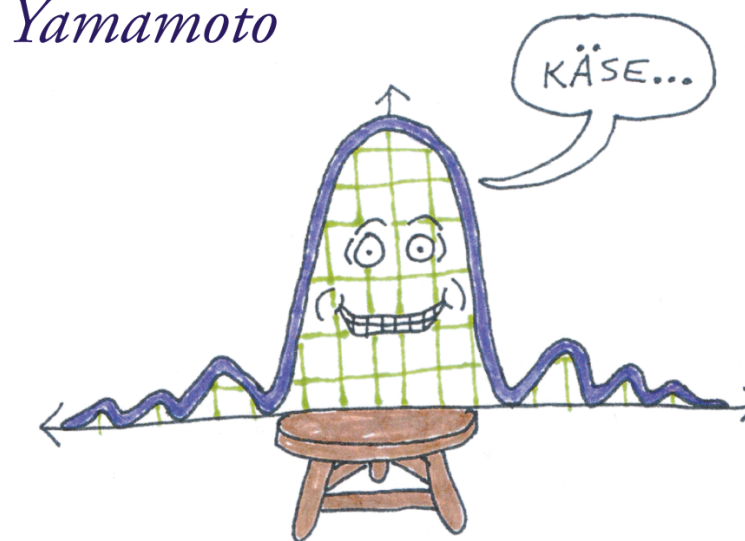
COOH

DNA microarrays, e.g., 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- and RNA-bound proteins**, on genomic scales.

Quantum Measurement of a Single System

Orly Alter

Yoshihisa Yamamoto



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



ALTER
YAMAMOTO

Quantum Measurement of a Single System



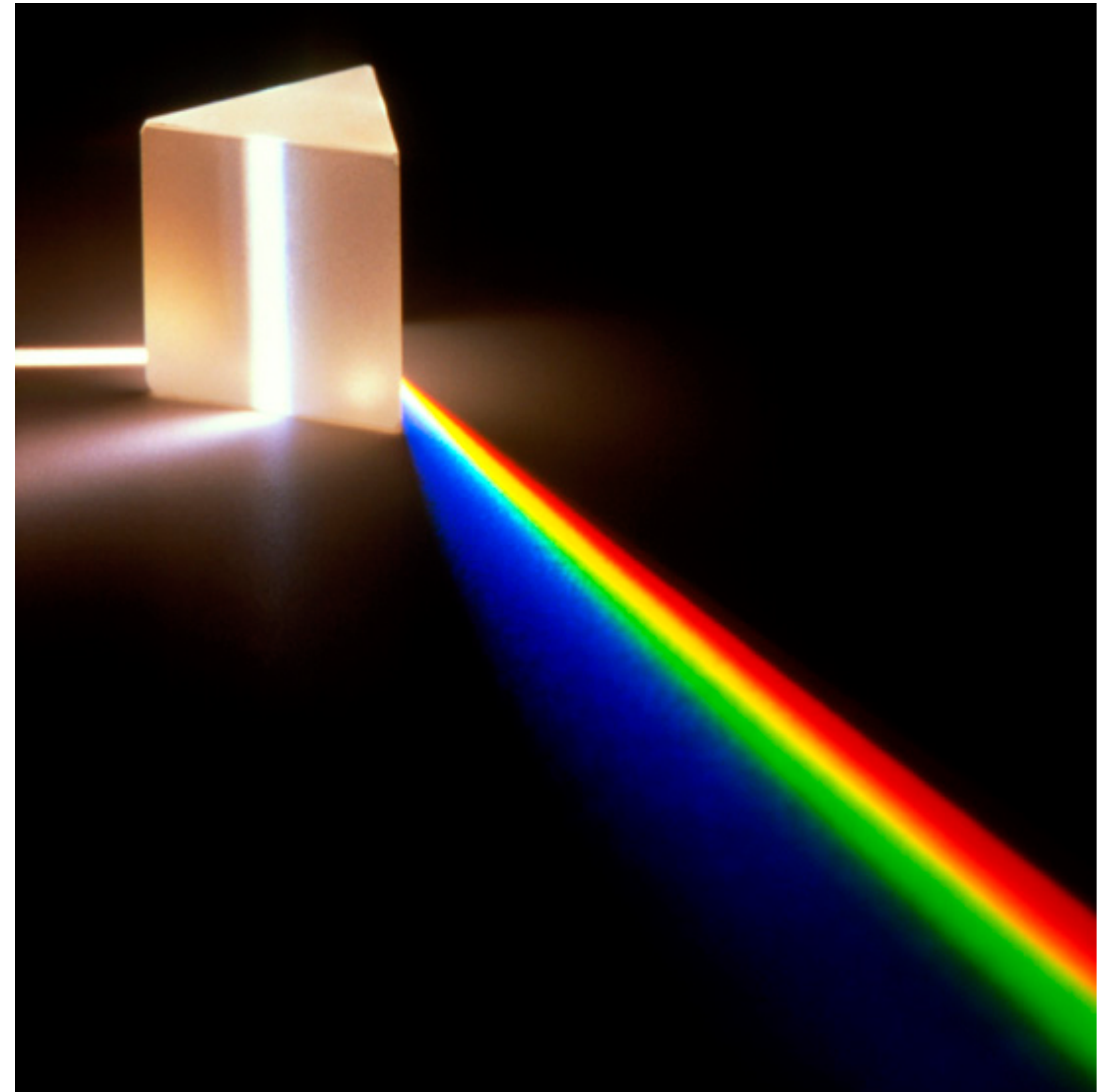
Global Mathematical Vocabulary for Molecular Biological Discovery

The **singular value decomposition (SVD)** underlies the theoretical description of the physical world.

The generalizations of the SVD can integrate and compare different data types.

The computation of the SVD scales with data sizes.

Alter & Yamamoto, *Quantum Measurement of a Single System*. New York, NY: Wiley-Interscience (2001);
<https://doi.org.10.1002/9783527617128>

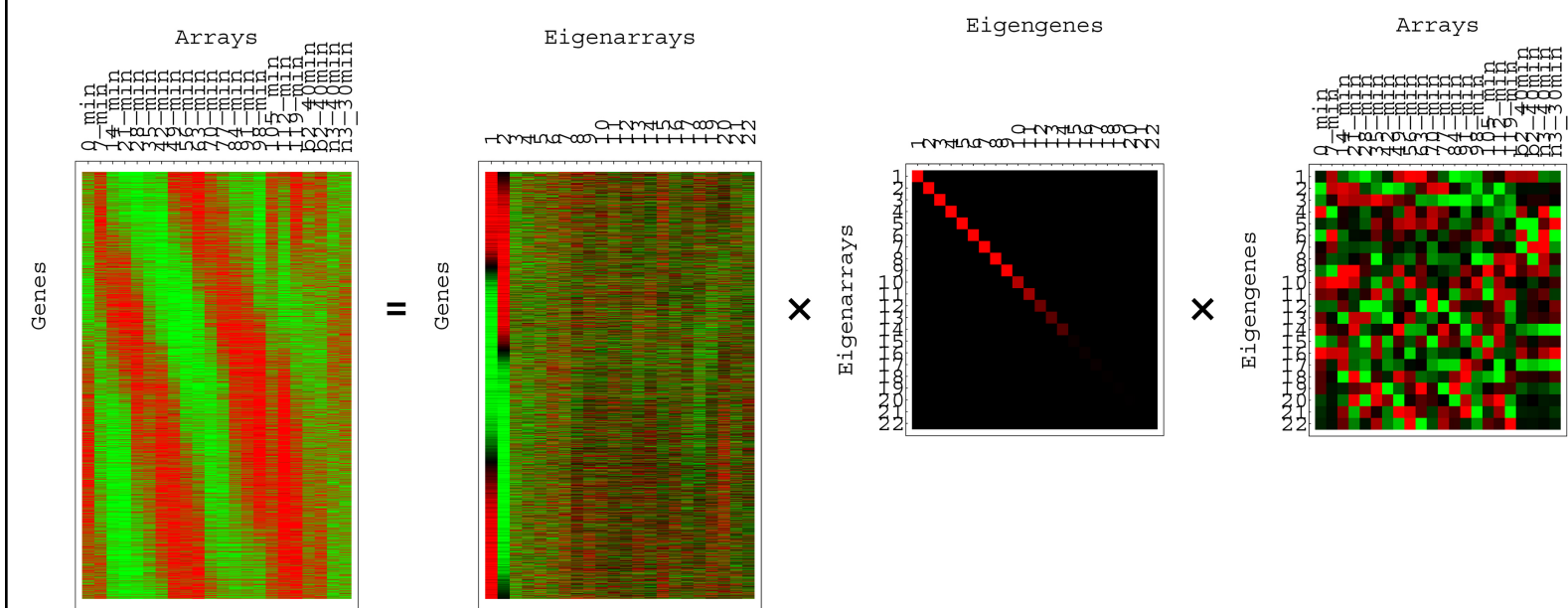


Physics-Inspired Matrix and Tensor Models

The SVD and its generalizations are interpretable in terms of the known biology and batch effects underlying the data.

SVD

Alter, Brown & Botstein,
PNAS 97, 10101 (2000).

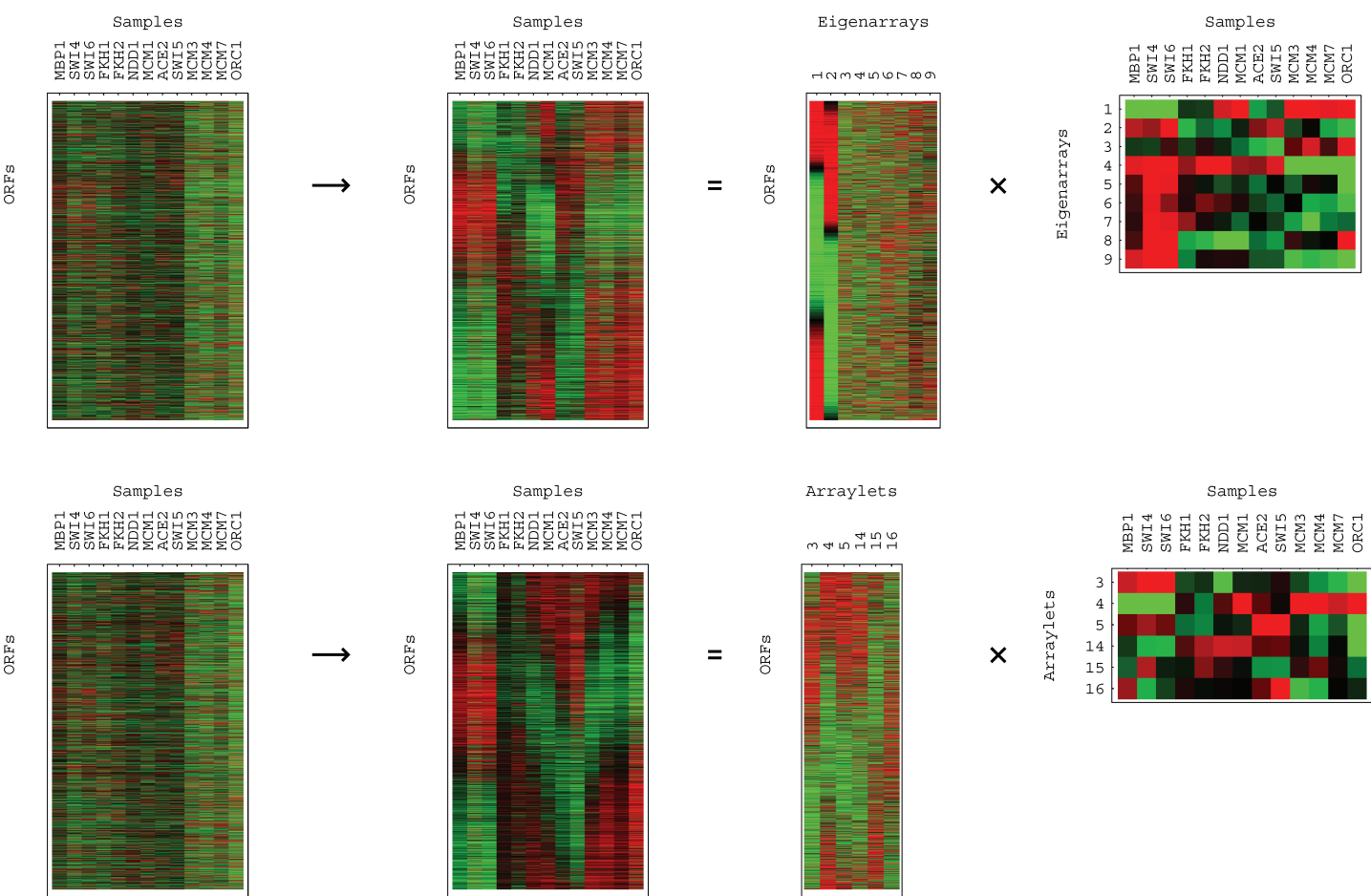


“Eigengenes” and
“eigenarrays” → cellular
processes and states in
one dataset.

Eigenvalue Decomposition

Integrative Pseudoinverse

Alter & Golub,
PNAS 101, 16577 (2004).

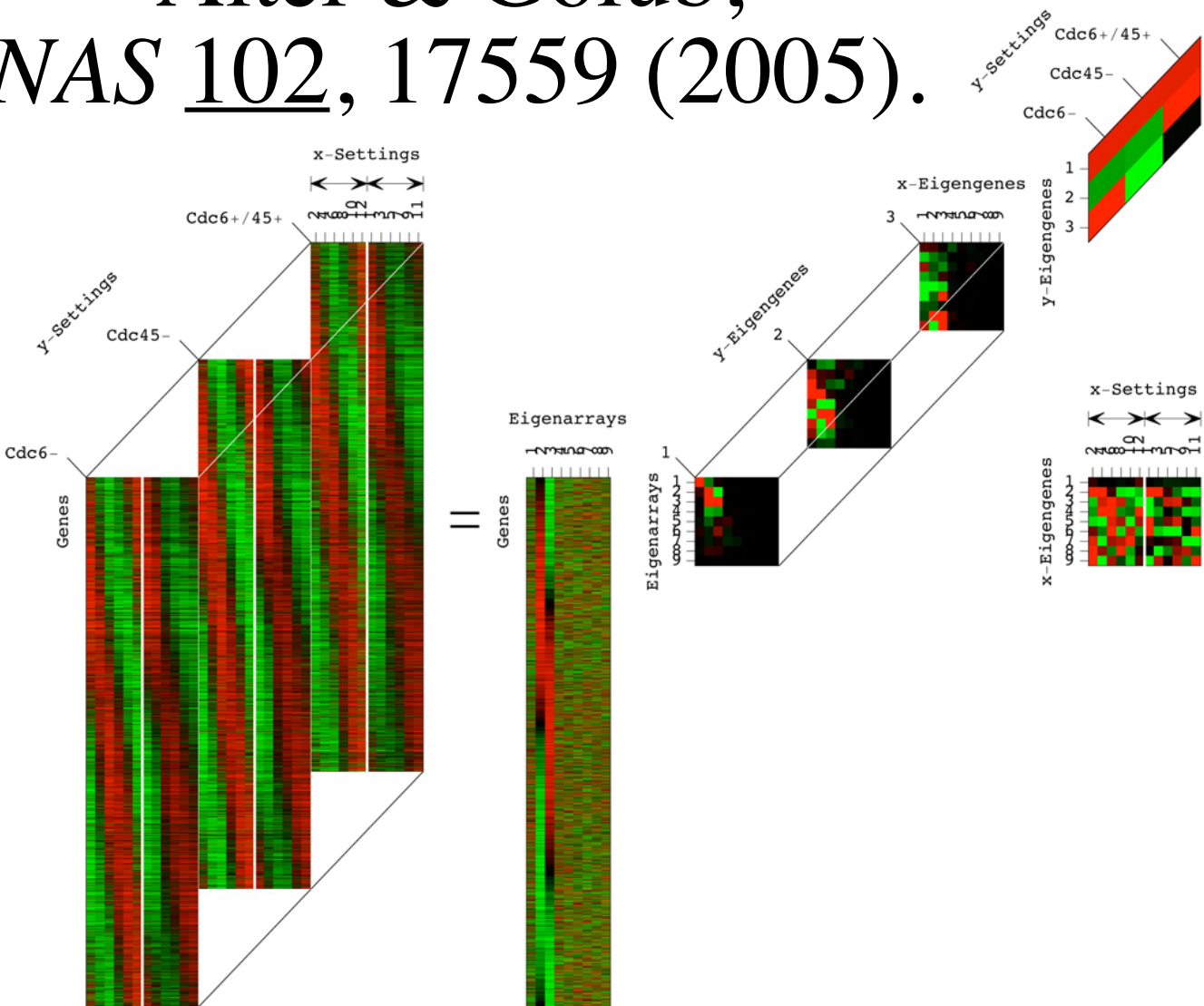


“Pseudoinverse correlation”
→ causal coordination
between two datasets.

Inverse Projection

Integrative Tensor SVD

Alter & Golub,
PNAS 102, 17559 (2005).



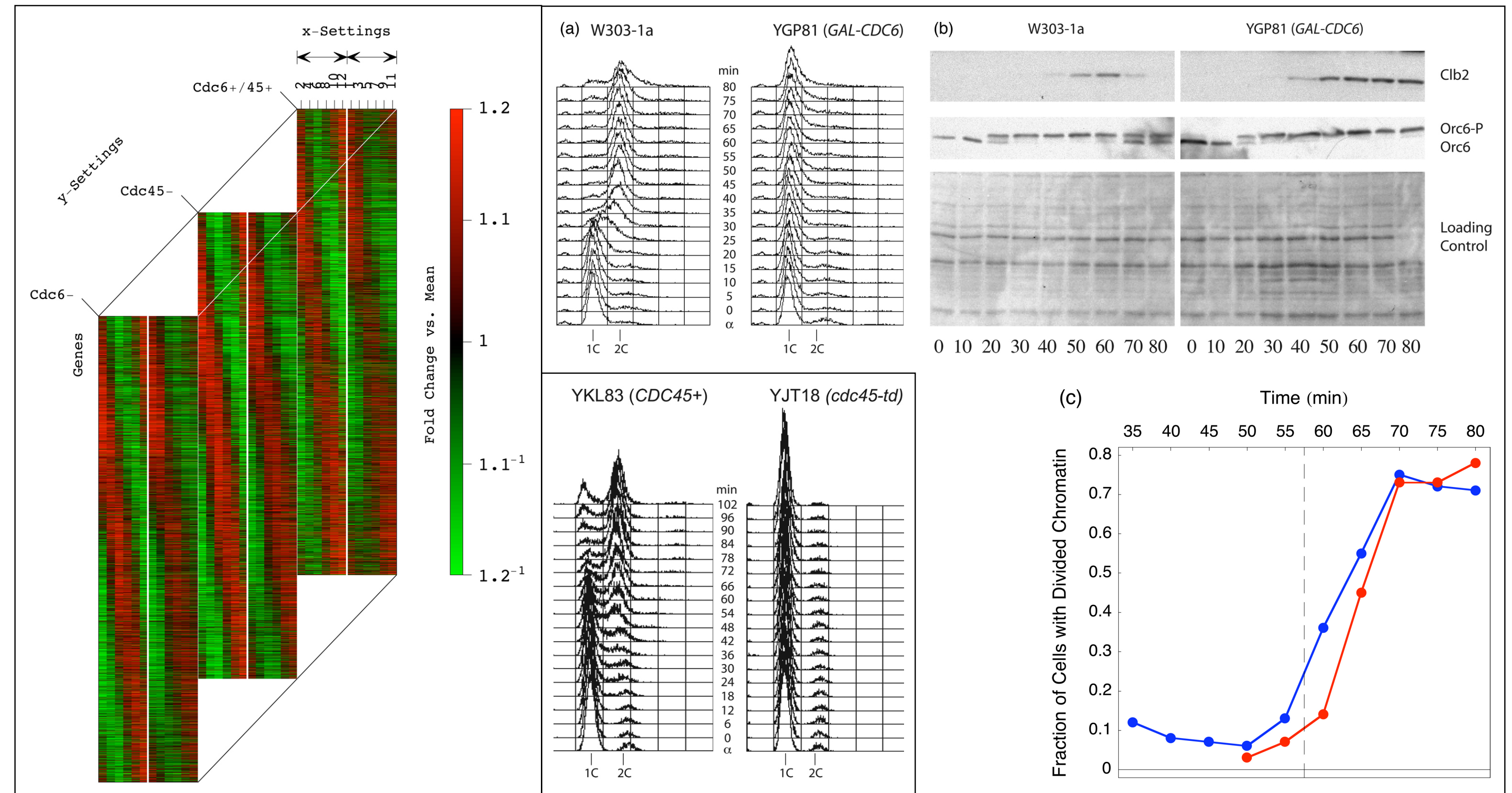
“x-” and “y-Eigengenes” and
“eigenarrays” → interrelations
among the processes and states
of one higher-order dataset.

Experimental Verification of a Computationally Predicted Mechanism for DNA Replication to Affect RNA Expression

Omberg, Meyerson, Kobayashi, Drury, Diffley & Alter, *MSB* 5, 312 (2009);
https://alterlab.org/verification_of_prediction/; Omberg, Golub & Alter, *PNAS* 104, 18371 (2007).

The SVD and its generalizations can correctly predict previously unknown experimentally verifiable global mechanisms.

Replication origin licensing decreases the expression of genes with origins near their 3' ends, revealing that downstream origins can regulate the expression of upstream genes.

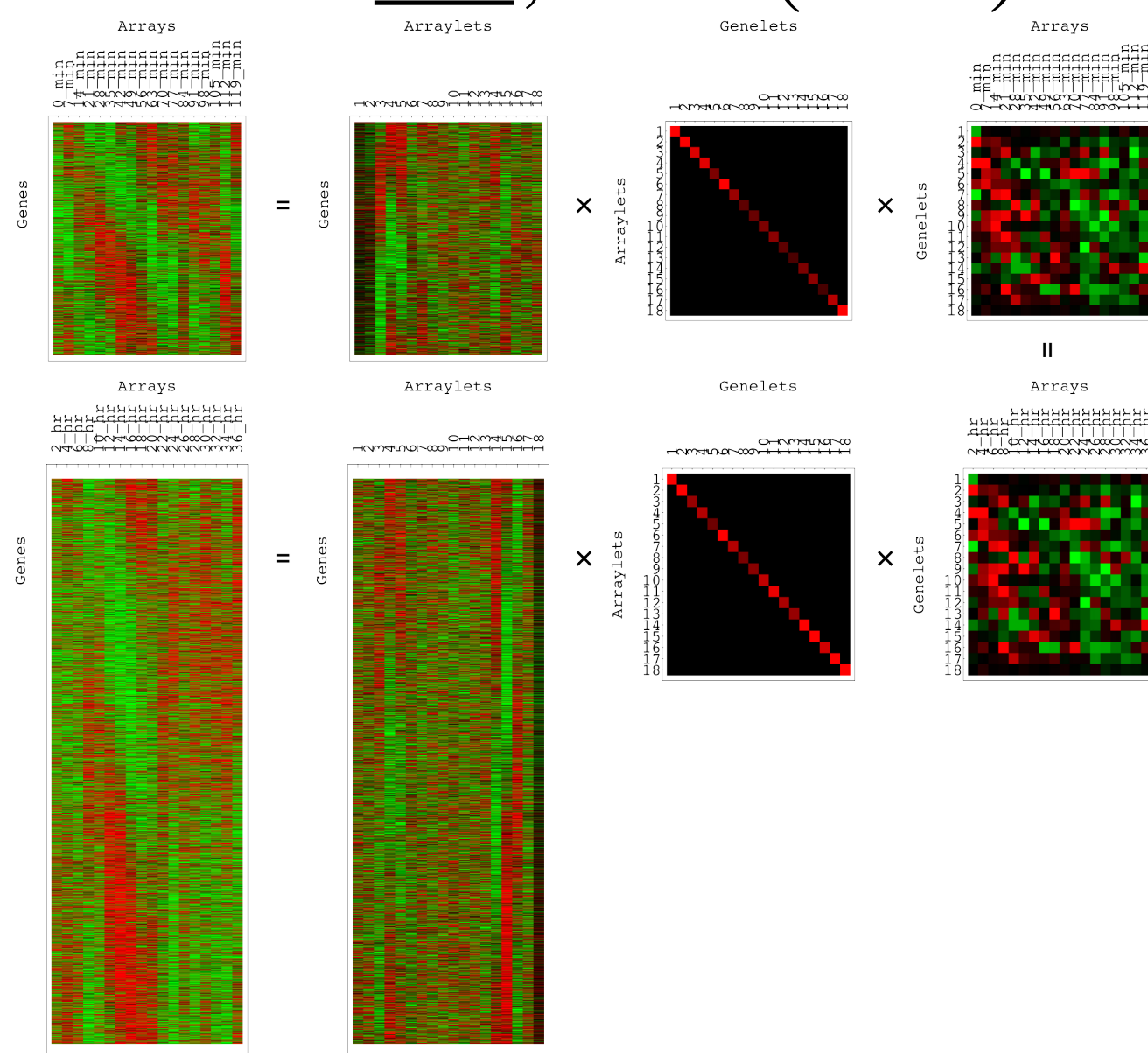


Comparative Spectral Decompositions

Simultaneously separate the similar from the dissimilar among multiple datasets and create a single coherent model.

GSVD

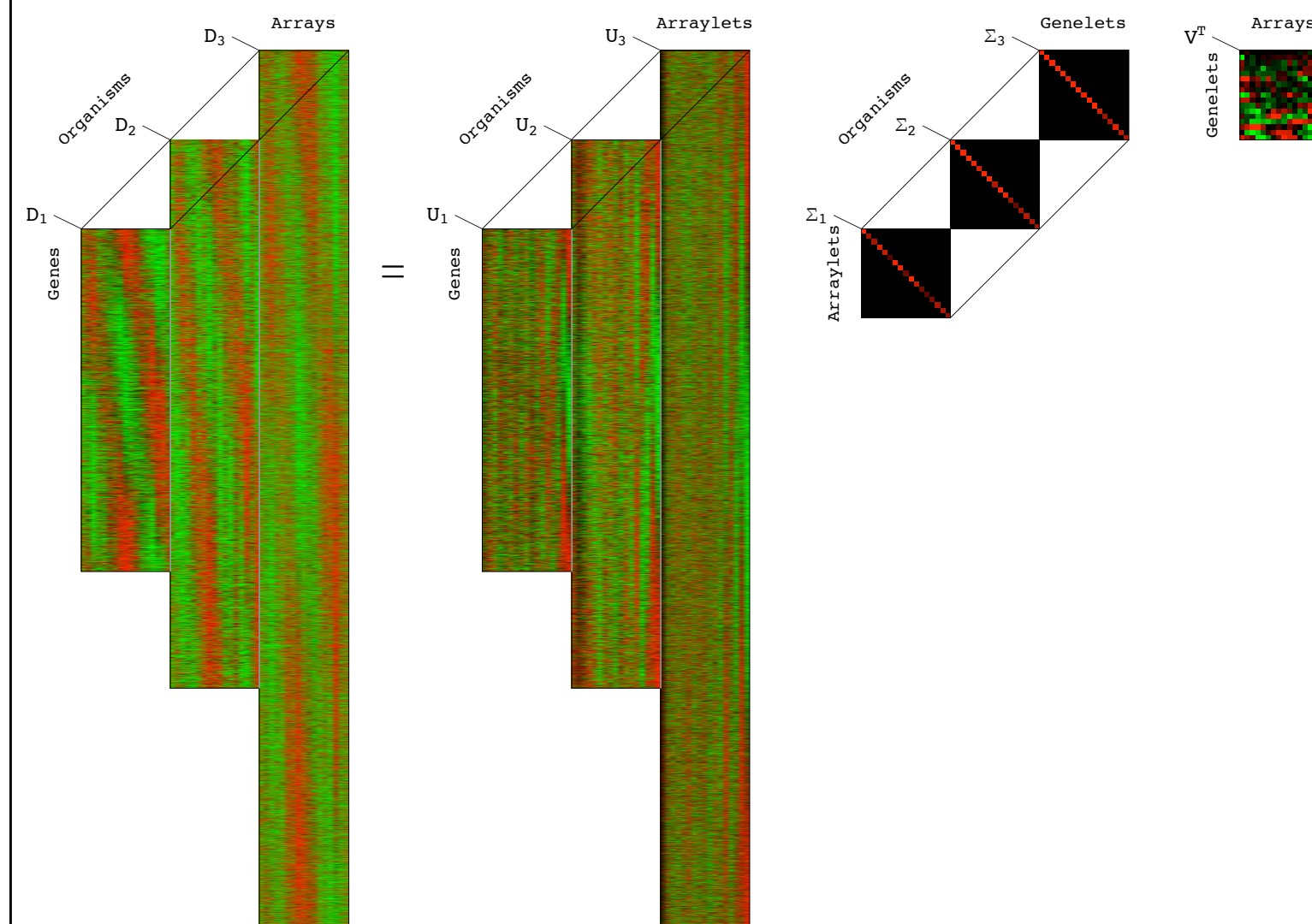
Alter, Brown & Botstein,
PNAS 100, 3351 (2003).



“Genelets” and “arraylets”
→ phenomena
exclusive to one of, or
common to two, datasets.

HO GSVD

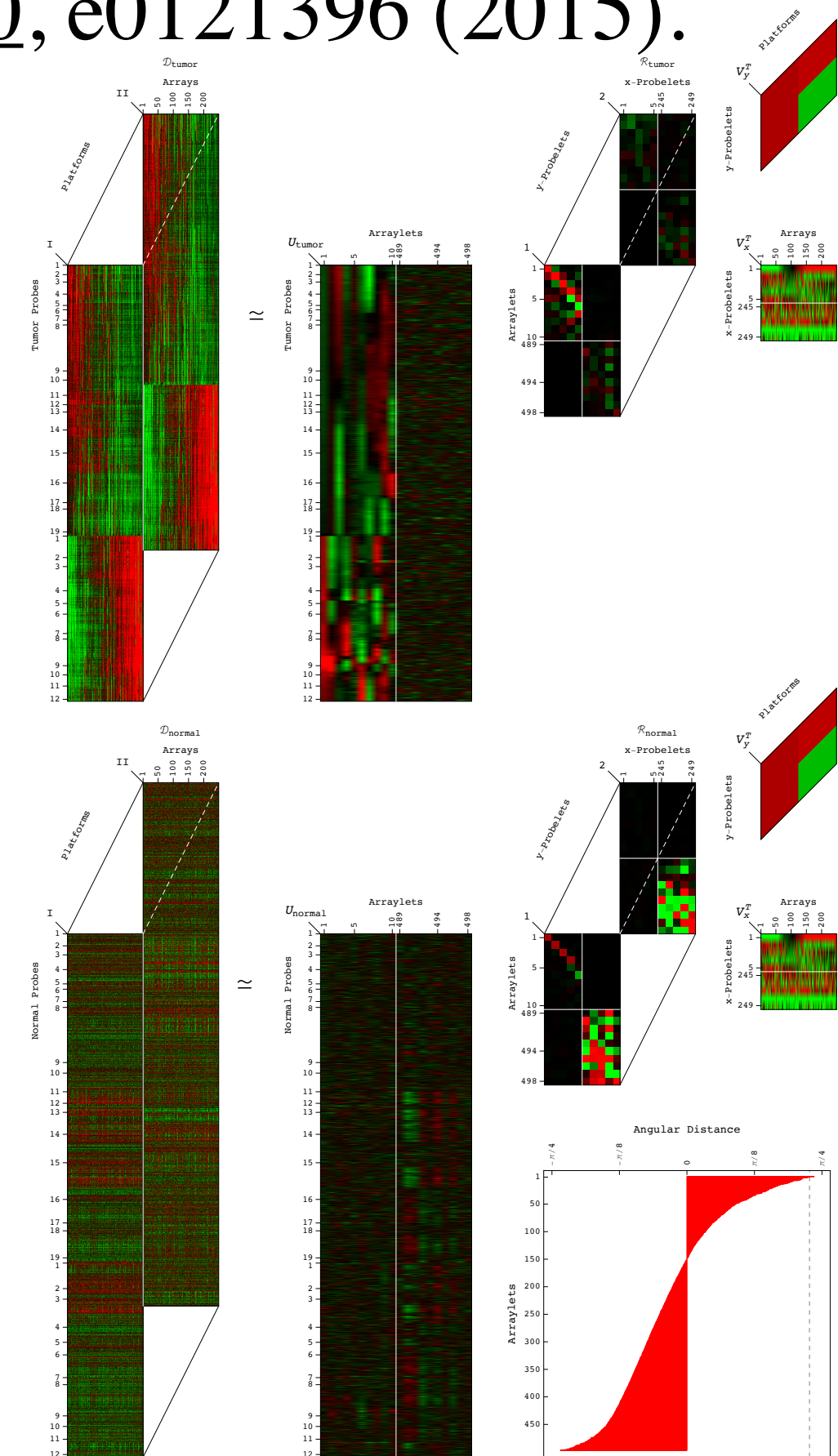
Ponnappalli, Saunders, Van Loan &
Alter, *PLoS One* 6, e28072 (2011).



“Genelets” and “arraylets”
→ phenomena exclusive to
one or more of, or common
to multiple, datasets.

Tensor GSVD

Sankaranarayanan,* et int., Alter,
PLoS One 10, e0121396 (2015).



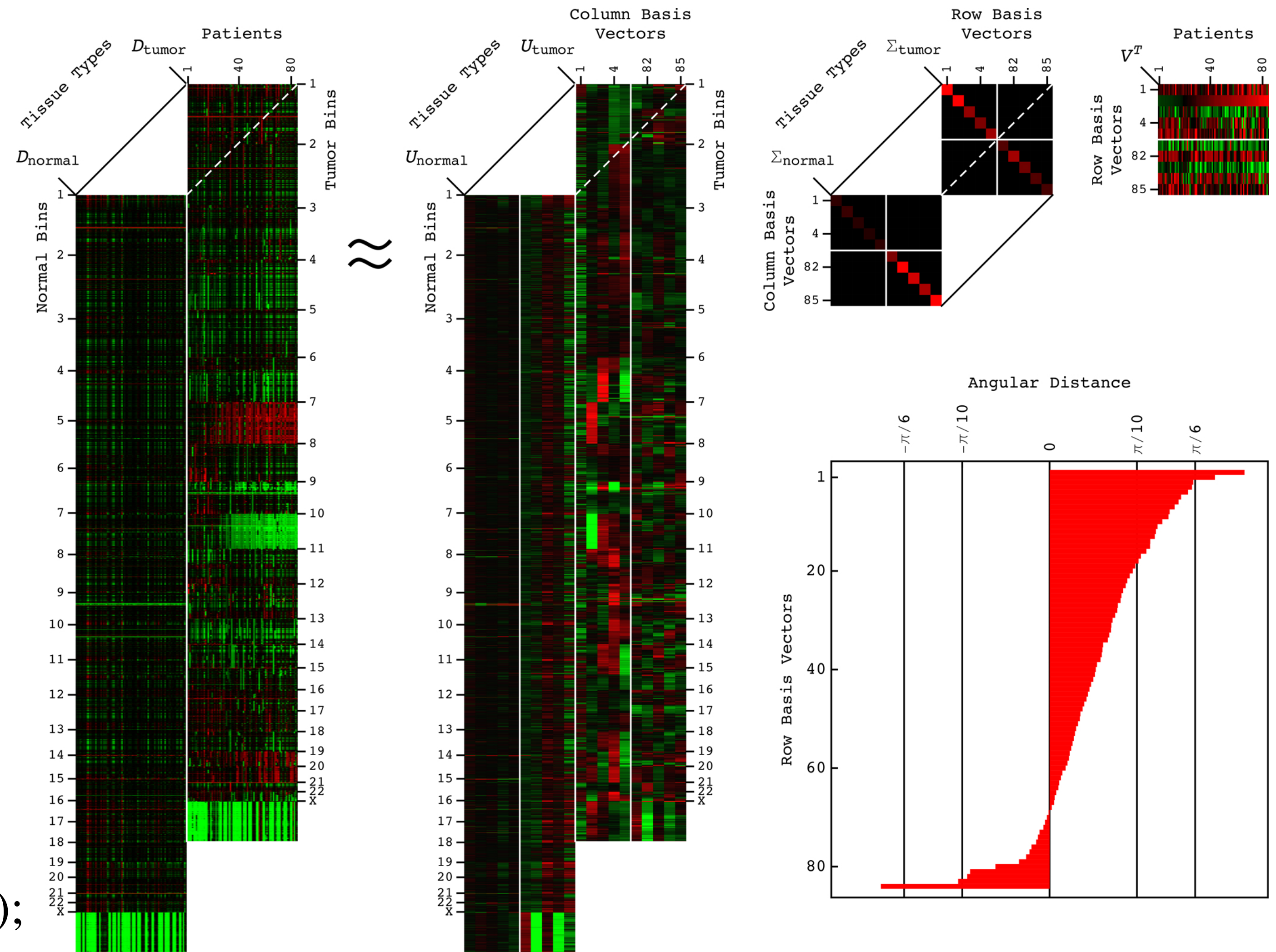
“Subtensors”
→ exclusive
or common
between two
higher-order
datasets,
consistent or
varying
across orders.

Retrospective Clinical Trial Experimentally Validates Glioblastoma Genome-Wide Pattern of DNA Copy-Number Alterations Predictor of Survival

Ponnappalli, Bradley, Devine, Bowen, Coppens, Leraas, Milash, Li, Luo, Qiu, Wu, Yang, Wittwer, Palmer, Jensen, Gastier-Foster, Hanson, Barnholtz-Sloan & Alter, *APL Bioeng* 4, 026106 (2020);
https://alterlab.org/GBM_retrospective_clinical_trial/

- For 70 years, the best indicator of a patient's survival has been age at diagnosis.
Netsky et al., *J Neurosurg* 7, 269 (1950); Curran Jr. et al., *J Natl Cancer Inst* 85, 704 (1993).
- Recurring DNA copy-number alterations (CNAs) have been recognized as a hallmark of cancer for over a century and have been observed in glioblastoma (GBM) tumors.
Boveri, *Concerning the Origin of Malignant Tumours*. Jena, Germany: Gustav Fischer Verlag (1914).
- Repeated previous attempts to associate a GBM tumor's DNA CNAs with a patient's outcome failed, including previous studies of data from the Cancer Genome Atlas that (TCGA) used other methods.
Weber et al., *Lab Invest* 74, 108 (1996); Wiltshire et al., *Neuro Oncol* 2, 164 (2000); Misra et al., *Clin Cancer Res* 11, 2907 (2005).

Mathematically Universal Biologically Consistent Genotype-Phenotype



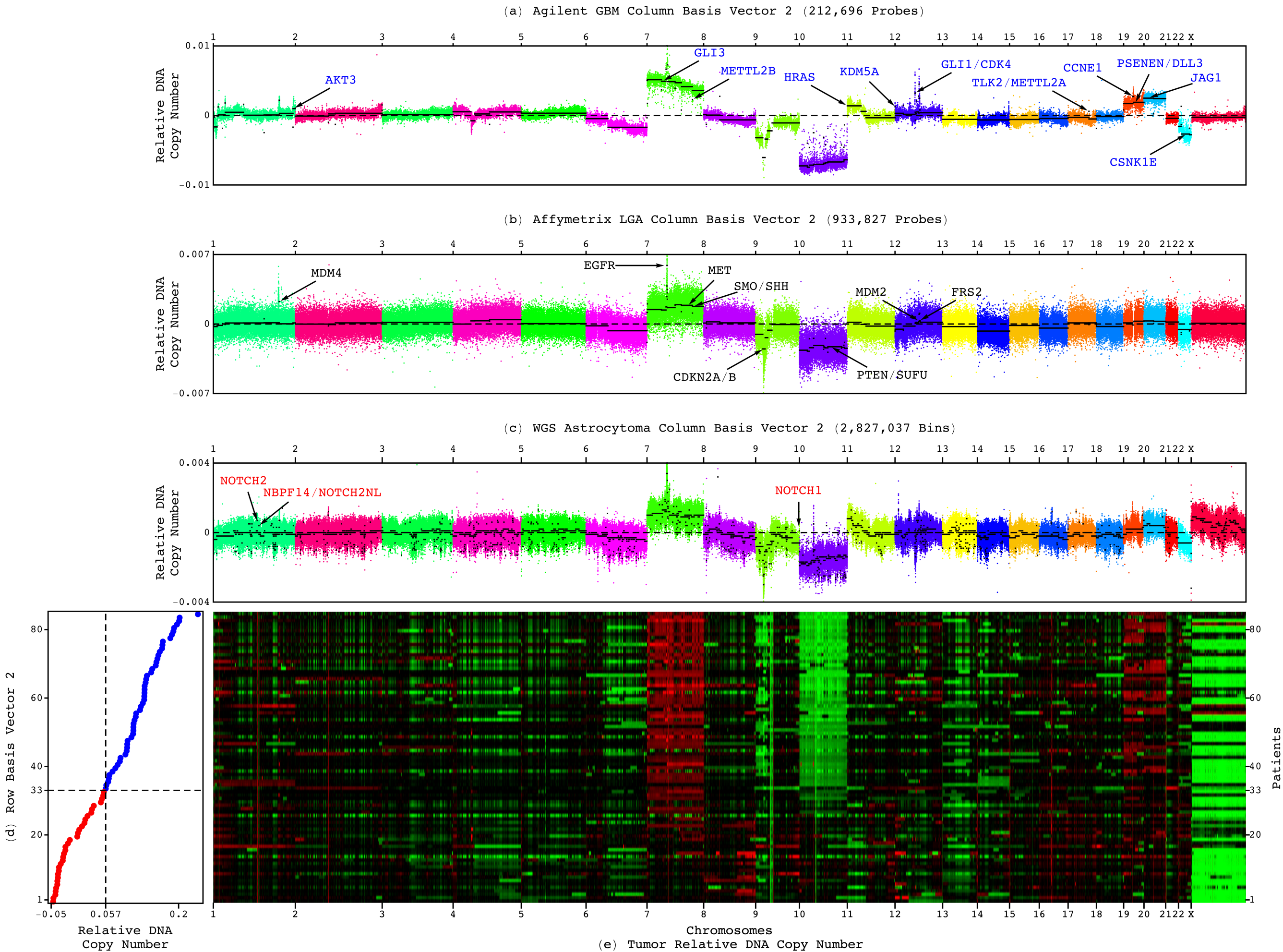
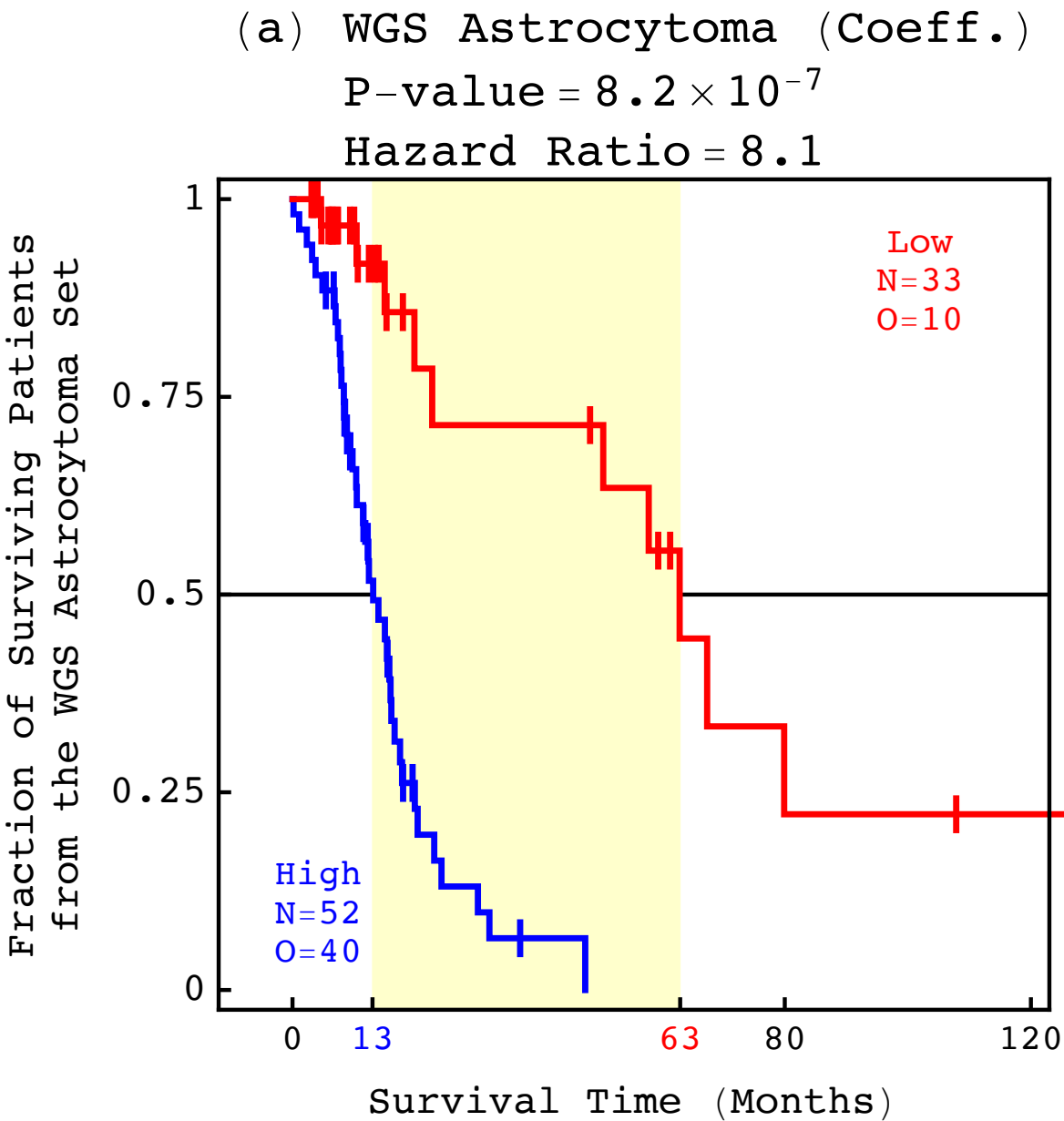
Lee,* et int. & Alter, *PLoS One* 7, e30098 (2012);

Aiello & Alter, *PLoS One* 11, e0164546 (2016);

Aiello, Ponnappalli & Alter, *APL Bioeng* 2, Special Topic: Bioengineering of Cancer 031909 (2018).

Invariably Uncovered by and Only by the GSVD

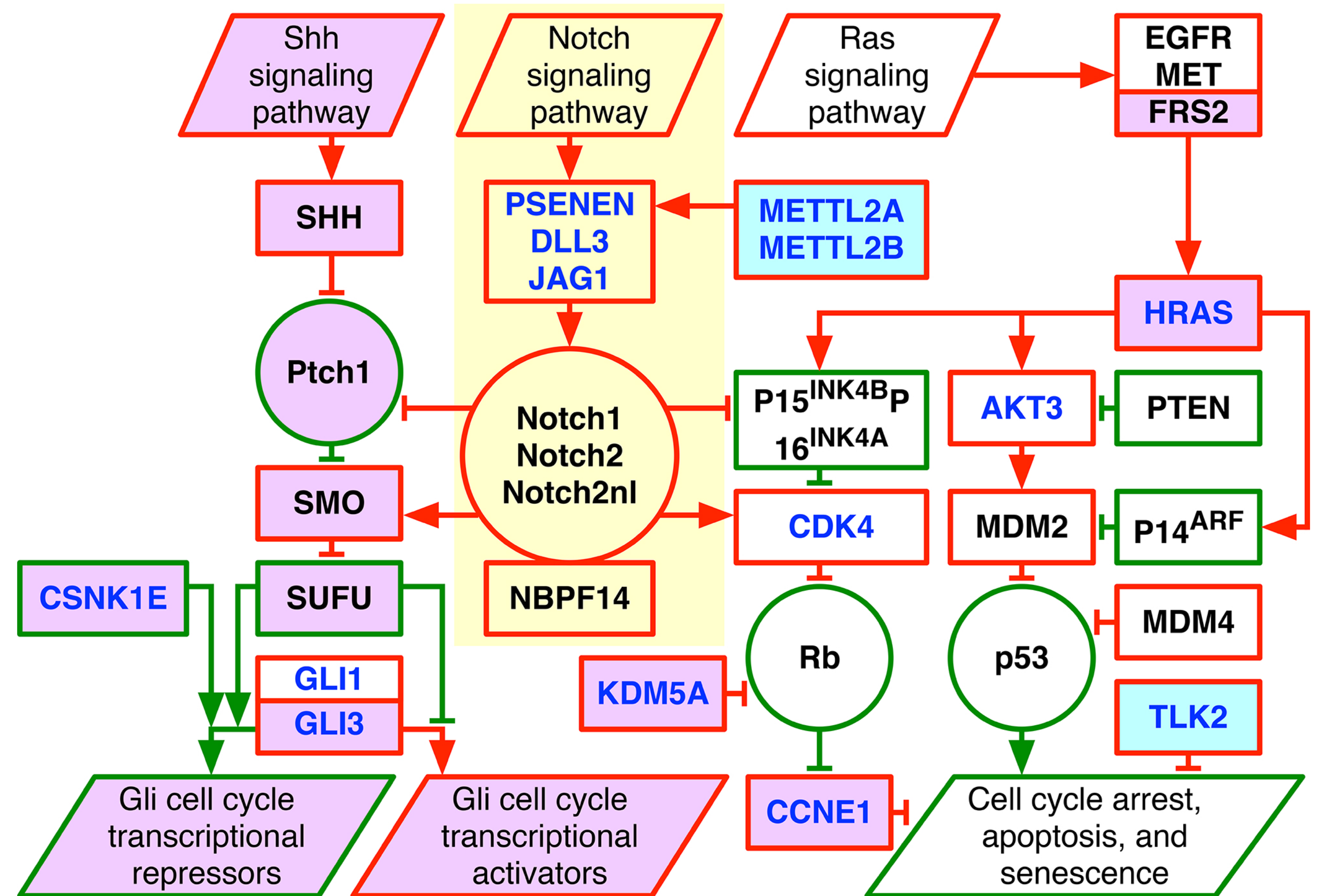
Like the Agilent GBM and Affymetrix lower-grade astrocytoma patterns, the whole-genome sequencing (WGS) astrocytoma pattern is **correlated with a shorter, roughly one-year median survival time**.



TCGA Research Network, *Nature* 455, 1061 (2008); *N Engl J Med* 372, 2481 (2015).

Encodes for Transformation via the Ras, Shh, and Notch Pathways

Includes most CNAs known in GBM, mostly in the rat sarcoma (Ras) pathway, and at least as many previously unrecognized, mostly in the sonic hedgehog (Shh) and Notch pathways that induce medulloblastoma and neuroblastoma. Some of these natural CNAs are analogous to artificial elements that transform human normal into tumor cells with grossly polyploid nuclei.

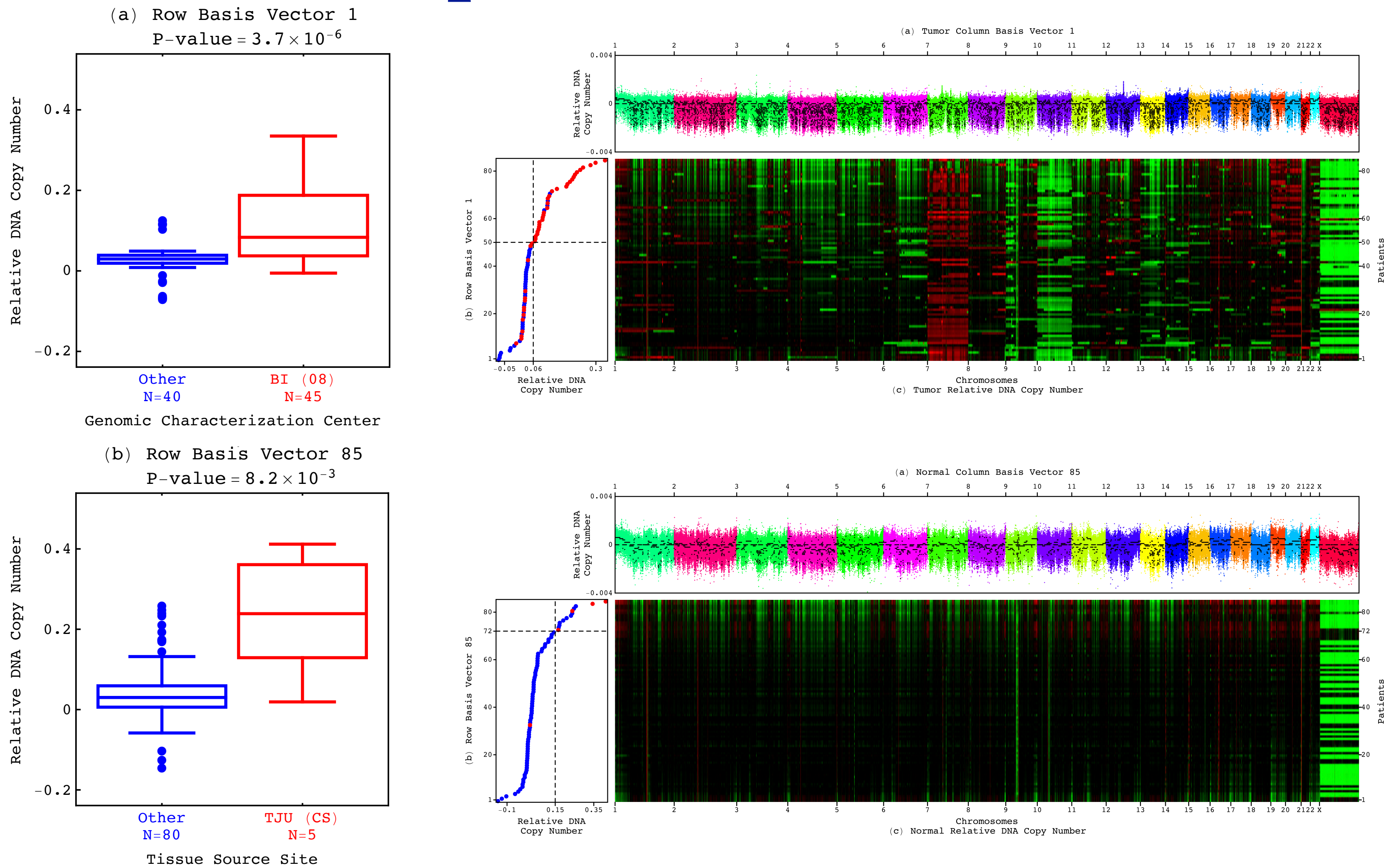


Wen & Kesari, *N Engl J Med* 359, 492 (2008);

Waldman et al., *Nature* 381, 713 (1996);

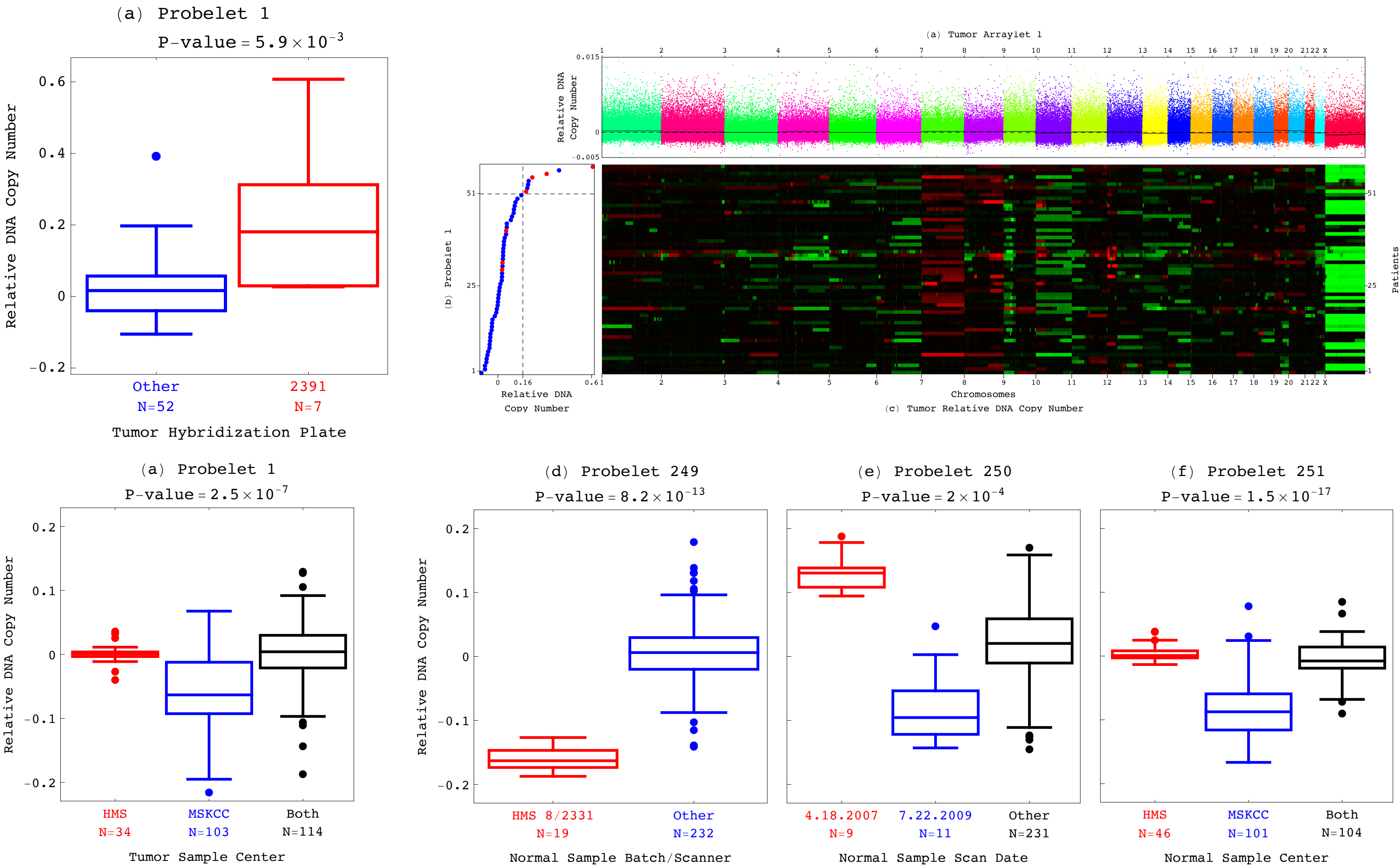
Hahn et al., *Nature* 400, 464 (1999); Irianto et al., *Curr Biol* 27, 210 (2017).

GSVD Blind Separation of WGS Effects



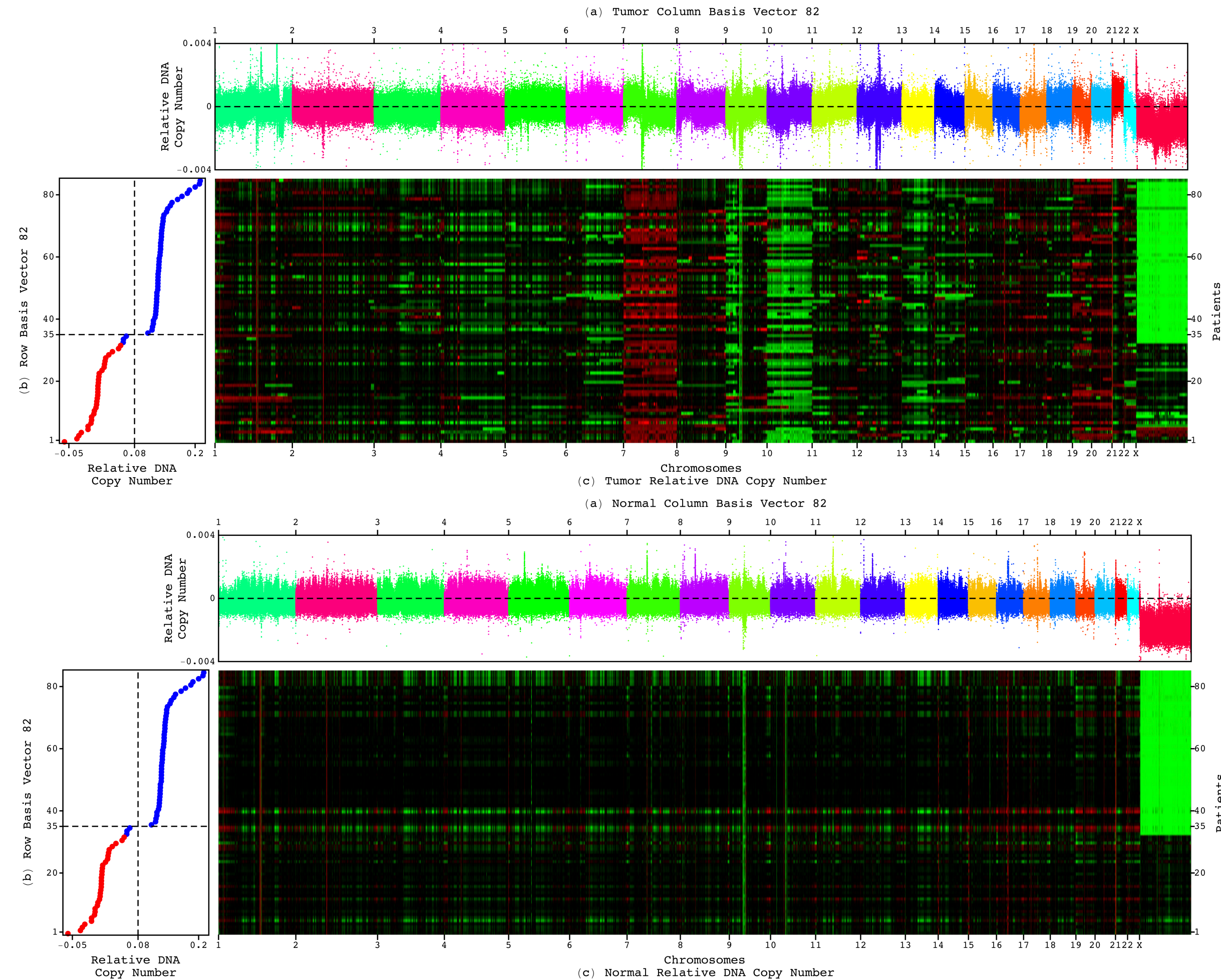
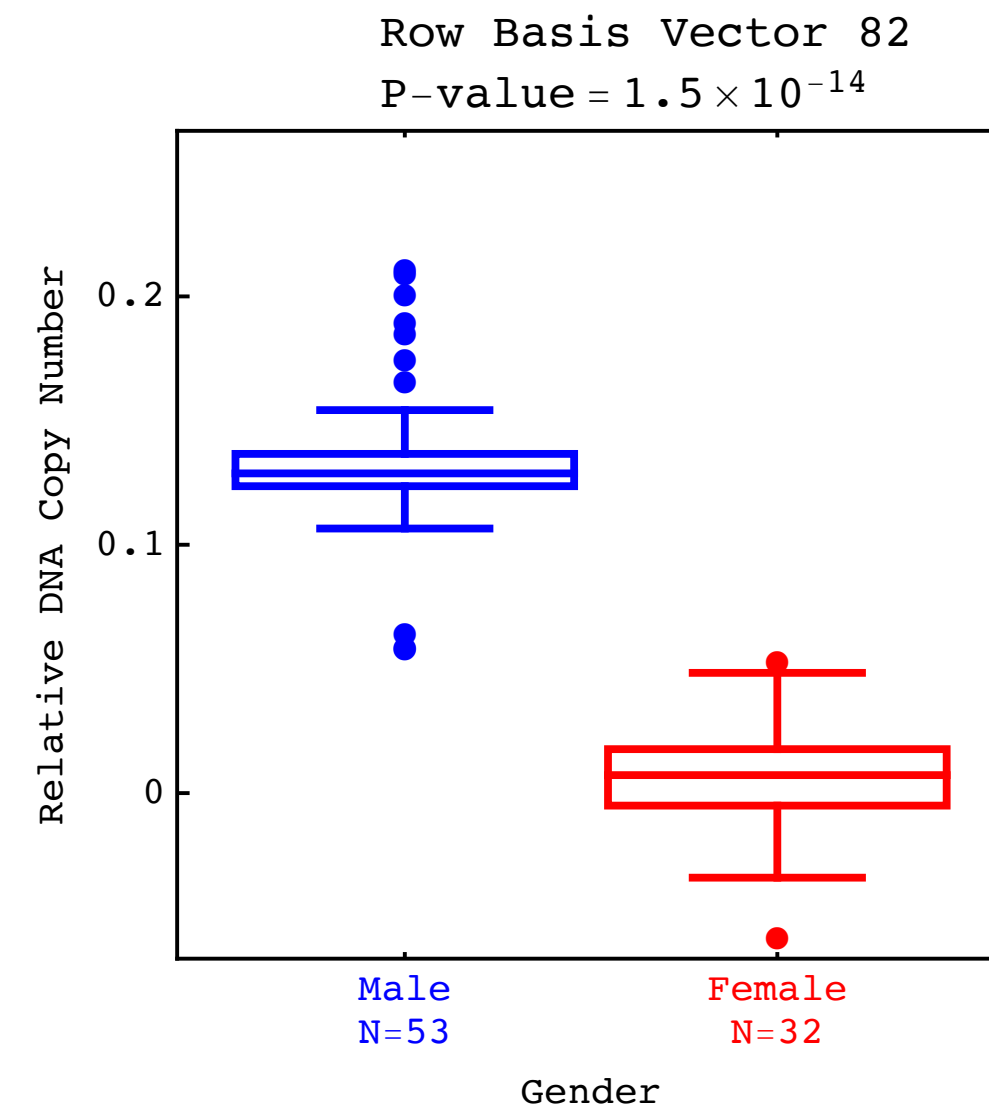
Guanine-cytosine (GC) content effects vary in magnitude between batches.
Roberts, Carneiro & Schatz, *Genome Biol* 14, 405 (2013).

Blind Separation of Microarray Effects



WGS with Affymetrix single-nucleotide polymorphism (SNP) and Agilent comparative genomic hybridization (CGH) microarrays represent the main genomic profiling technologies.

Blind Separation of Normal Variations

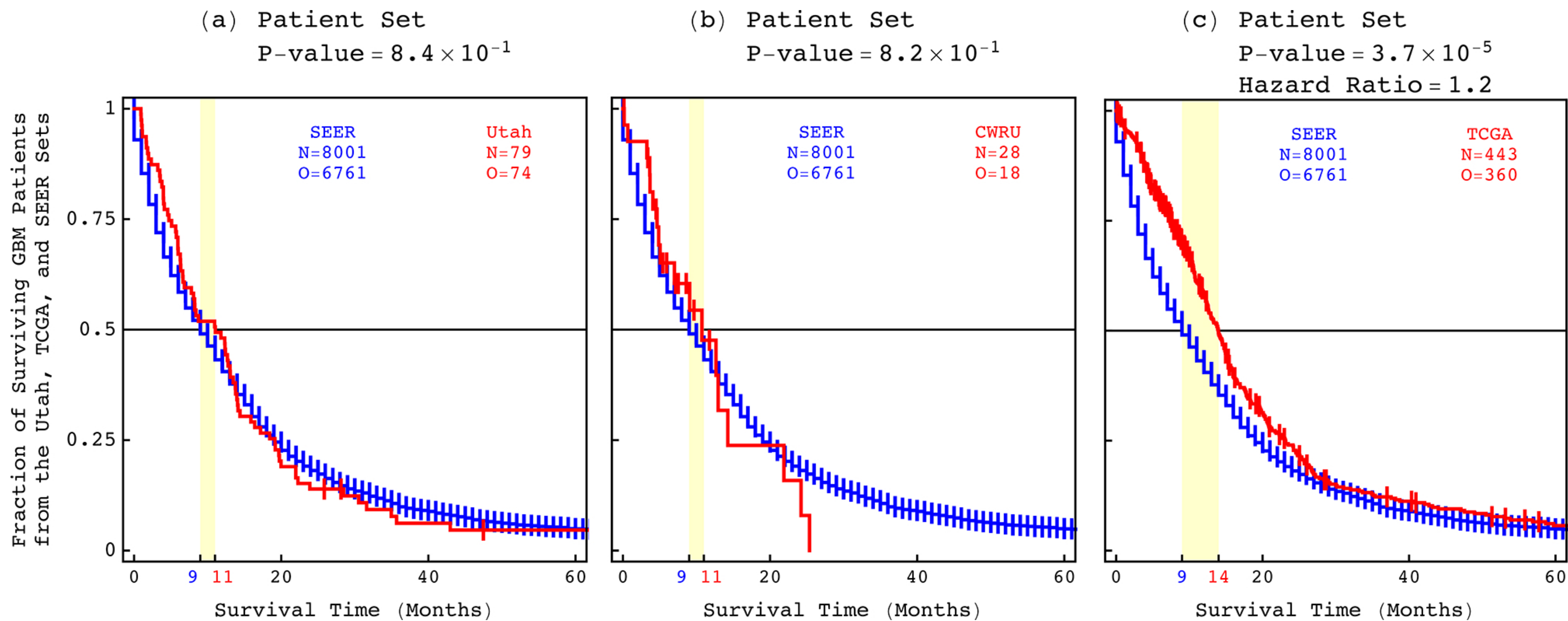


The normal male-specific X chromosome deletion is conserved in the tumors.

TCGA gender labels were corrected.

<https://grants.nih.gov/grants/guide/notice-files/NOT-HG-11-021.html>

The Utah set of 79 Patients Statistically Represents the U.S. Adult GBM Population



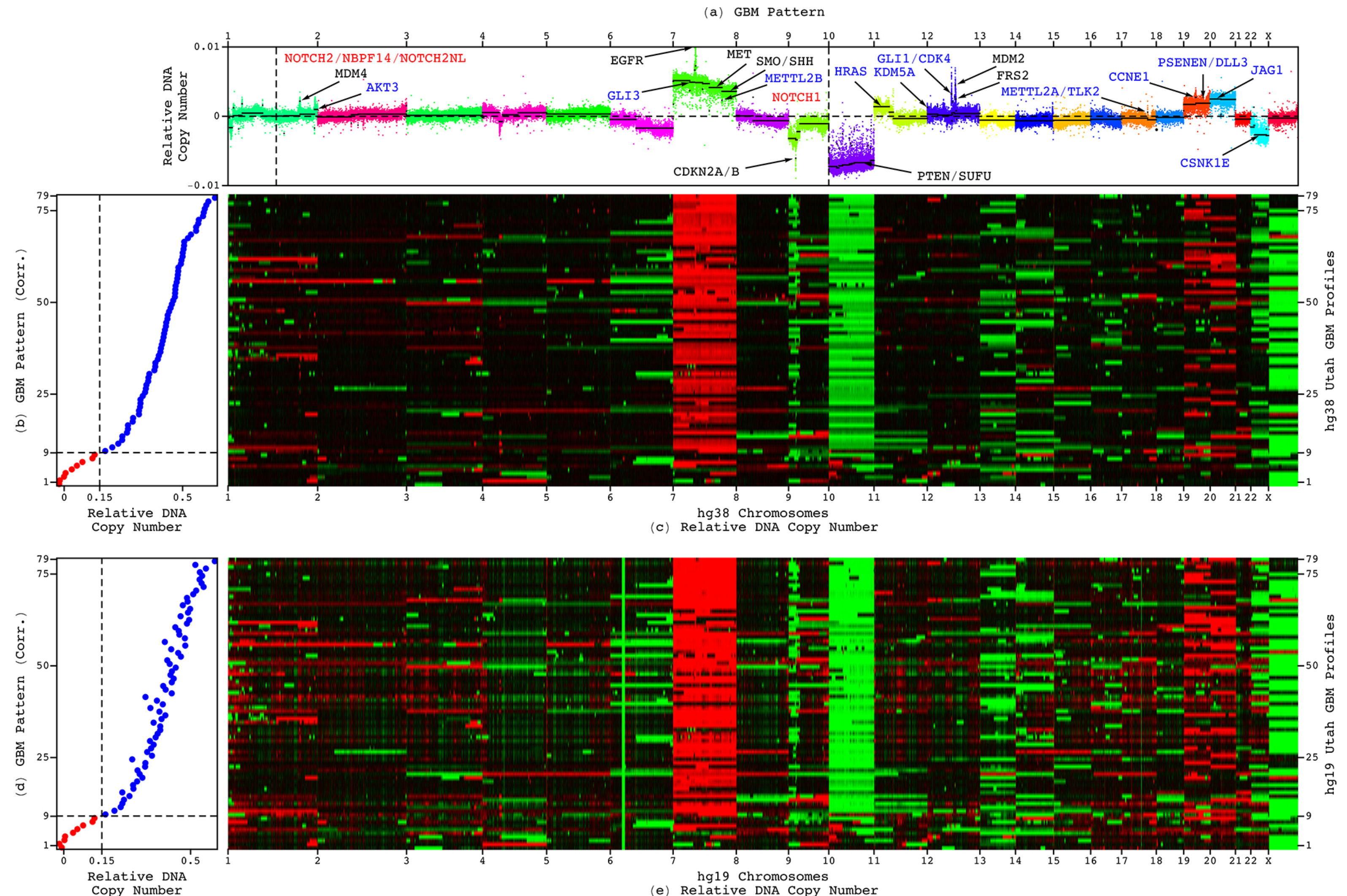
Phenotype			Group	Utah Set	CWRU Set	TCGA Set	SEER Set	Utah vs. SEER χ^2 P-Value	CWRU vs. SEER χ^2 P-Value	TCGA vs. SEER χ^2 P-Value
Normal	Sex	Female		27	12	169	3331	1.8×10^{-1}	9.0×10^{-1}	1.5×10^{-1}
		Male		52	16	274	4670			
	Race	Not White		5	3	40	959	1.2×10^{-1}	8.4×10^{-1}	1.1×10^{-1}
		White		74	25	386	7042			
	Ethnicity	Hispanic		2	0	11	507	1.7×10^{-1}	1.7×10^{-1}	4.4×10^{-3}
		Not Hispanic		77	28	382	7494			
Disease	Age (Years)	<50		10	3	117	1164	6.4×10^{-1}	5.7×10^{-1}	1.2×10^{-11}
		≥ 50		69	25	326	6837			

Profiling Technology and Reference Human Genome Affect <1% of the Classifications

Experimental batch effects normally reduce the reproducibility, i.e., precision, of classifications based upon between one to a few hundred genomic loci by >30%.

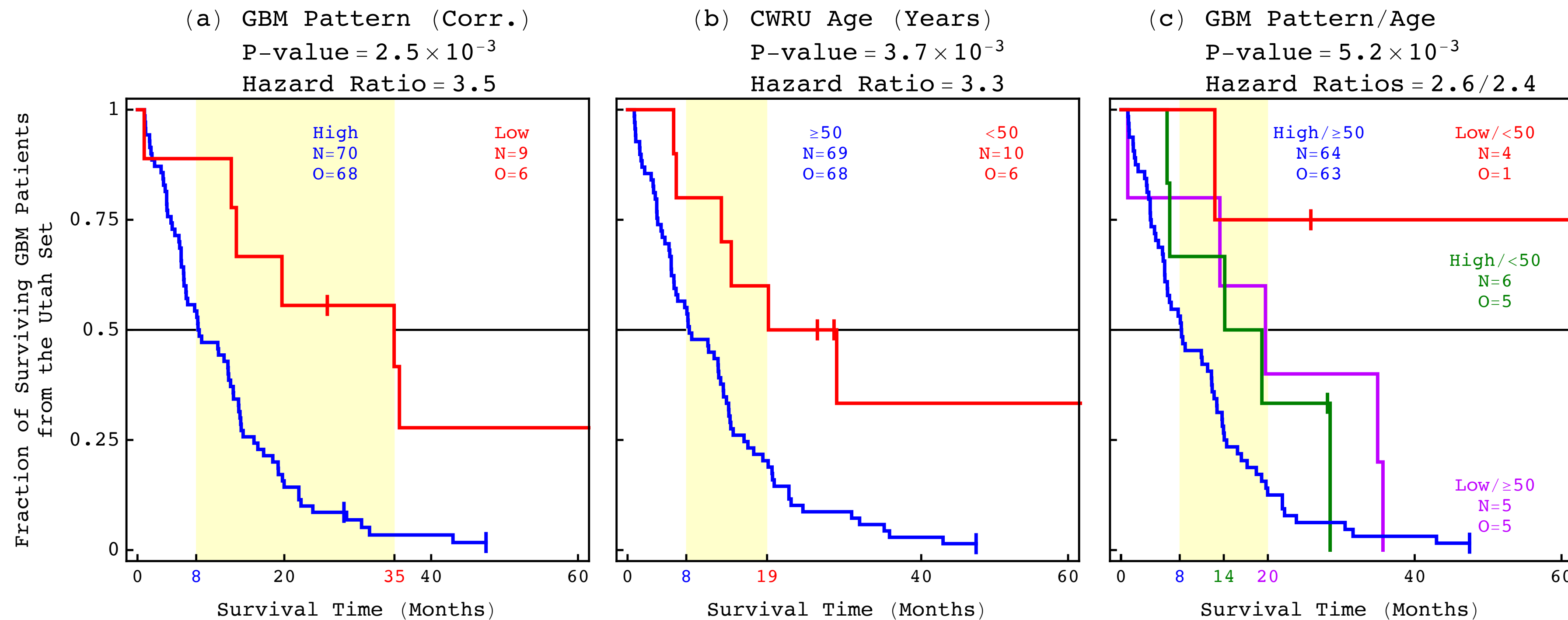
Pinto et al., *Nat Biotechnol* 29, 512 (2011).

Intratumor heterogeneity affects $\approx 11\%$ of the classifications.



With a 2.25-year Kaplan–Meier (KM) median survival difference, a 3.5 univariate Cox hazard ratio, and a 0.78 concordance index, i.e., accuracy.

Statistically Better Than and Independent of the Best Other Indicator, i.e., Age



In general as well as in patients who receive treatments, i.e., chemotherapy and radiation. Independent of chemotherapy and radiation and the post-surgical resection metrics, i.e., the Karnofsky performance score and the percent primary tumor resection.

Statistically Better Than and Independent of the Best Other Indicator, i.e., Age

Greater median survival differences, univariate hazard ratios and concordance indices, i.e., accuracies, and lower log-rank and Wald *P*-values as well as Akaike information criterion (AIC) values.

Bivariate hazard ratios within 95% confidence intervals of univariate ratios.

Patients (Number)	Predictor	KM Group	Median (Months)	Log-Rank <i>P</i> -Value	Cox Model	Hazard Ratio	95% Confidence Interval	Wald <i>P</i> -Value	Concordance Index
79	GBM Pattern (Corr.)	Low High	35 8	2.5×10^{-3}	Univariate	3.5	1.5– 8.2	4.3×10^{-3}	0.78
	CWRU Age (Years)	<50 ≥50	19 8	3.7×10^{-3}		3.3	1.4– 7.5	6.1×10^{-3}	0.78
	GBM Pattern (Corr.)	Low/<50		5.2×10^{-3}	Bivariate	2.6	1.1– 6.3	3.4×10^{-2}	0.76
	CWRU Age (Years)	Low/≥50	20			2.4	1.0– 5.7	5.0×10^{-2}	
		High/<50	14						
		High/≥50	8						
47	GBM Pattern (Corr.)	Low High	36 14	4.8×10^{-3}	Univariate	4.1	1.4– 11.8	8.7×10^{-3}	0.86
	CWRU Age (Years)	<50 ≥50	28 14	7.9×10^{-3}		3.7	1.3– 10.5	1.3×10^{-2}	0.81
59	GBM Pattern (Corr.)	Low High	36 12	2.0×10^{-3}		4.6	1.6– 13.0	4.3×10^{-3}	0.88
	CWRU Age (Years)	<50 ≥50	19 13	1.7×10^{-2}		2.7	1.2– 6.4	2.2×10^{-2}	0.72
75	GBM Pattern (Corr.)	Low High	35 8	2.1×10^{-3}		4.0	1.6– 10.1	4.0×10^{-3}	0.79
	CWRU Chemotherapy	Yes No	14 4	1.0×10^{-12}		6.0	3.5– 10.4	7.0×10^{-11}	0.93
	CWRU Radiation	Yes No	13 2	4.4×10^{-15}		10.8	5.3– 22.1	7.6×10^{-11}	0.93
74/75	GBM Pattern (Corr.)	Low/Yes	36	1.5×10^{-12}	Bivariate	3.8	1.3– 10.8	1.4×10^{-2}	0.91
	CWRU Chemotherapy	High/Yes High/No	14 4			5.0	2.9– 8.6	9.0×10^{-9}	
	GBM Pattern (Corr.)	Low/Yes	36	4.2×10^{-15}		4.6	1.6– 13.0	4.4×10^{-3}	0.92
	CWRU Radiation	High/Yes High/No	12 3			9.3	4.5– 19.3	1.6×10^{-9}	
52	GBM Pattern (Corr.)	Low High	35 12	8.1×10^{-3}	Univariate	3.2	1.3– 7.9	1.1×10^{-2}	0.74
	CWRU Karnofsky Score	≥60 <60	16 5	7.3×10^{-7}		5.8	2.7– 12.4	8.0×10^{-6}	0.87
51/52	GBM Pattern (Corr.)	Low/≥60	35	1.3×10^{-6}	Bivariate	3.5	1.3– 9.4	1.3×10^{-2}	0.83
	CWRU Karnofsky Score	High/≥60 High/<60	14 5			4.5	2.0– 9.9	2.0×10^{-4}	
28	GBM Pattern (Corr.)	Low High	35 6	6.1×10^{-3}	Univariate	10.6	1.4– 80.9	2.3×10^{-2}	0.97
	CWRU Percent Resection	≥30 <30	12 5	3.0×10^{-2}		2.8	1.1– 7.6	3.9×10^{-2}	0.73
	GBM Pattern (Corr.)	Low/≥30	35	4.5×10^{-3}	Bivariate	9.8	1.3– 76.2	2.9×10^{-2}	0.80
	CWRU Percent Resection	High/≥30 High/<30	8 5						

Statistically Better Than and Independent of the Best Other Indicator, i.e., Age in the TCGA Set of 443 Patients

Better than and independent of the existing pathology laboratory tests, i.e., for *MGMT* promoter methylation and *IDH1* mutation, as well as better than *TERT* gene expression.

Before progressing to GBM standard of care, *MGMT*, *IDH1*, and *TERT*, have already been used as indicators of survival and *MGMT* also as an indicator of response to alkalyting agents in other types of cancer.

Hegi et al., *NEJM* 352, 997 (2005); Hotta et al., *J Neurooncol* 21, 135 (1994); Parsons et al., *Science* 321, 1807 (2008); Nguyen et al., *Neuro Oncol* 19, 394 (2017); Batchelor & Louis, in *UpToDate*, eds. Loeffler & Wen. Waltham, MA: Wolters Kluwer (2018).

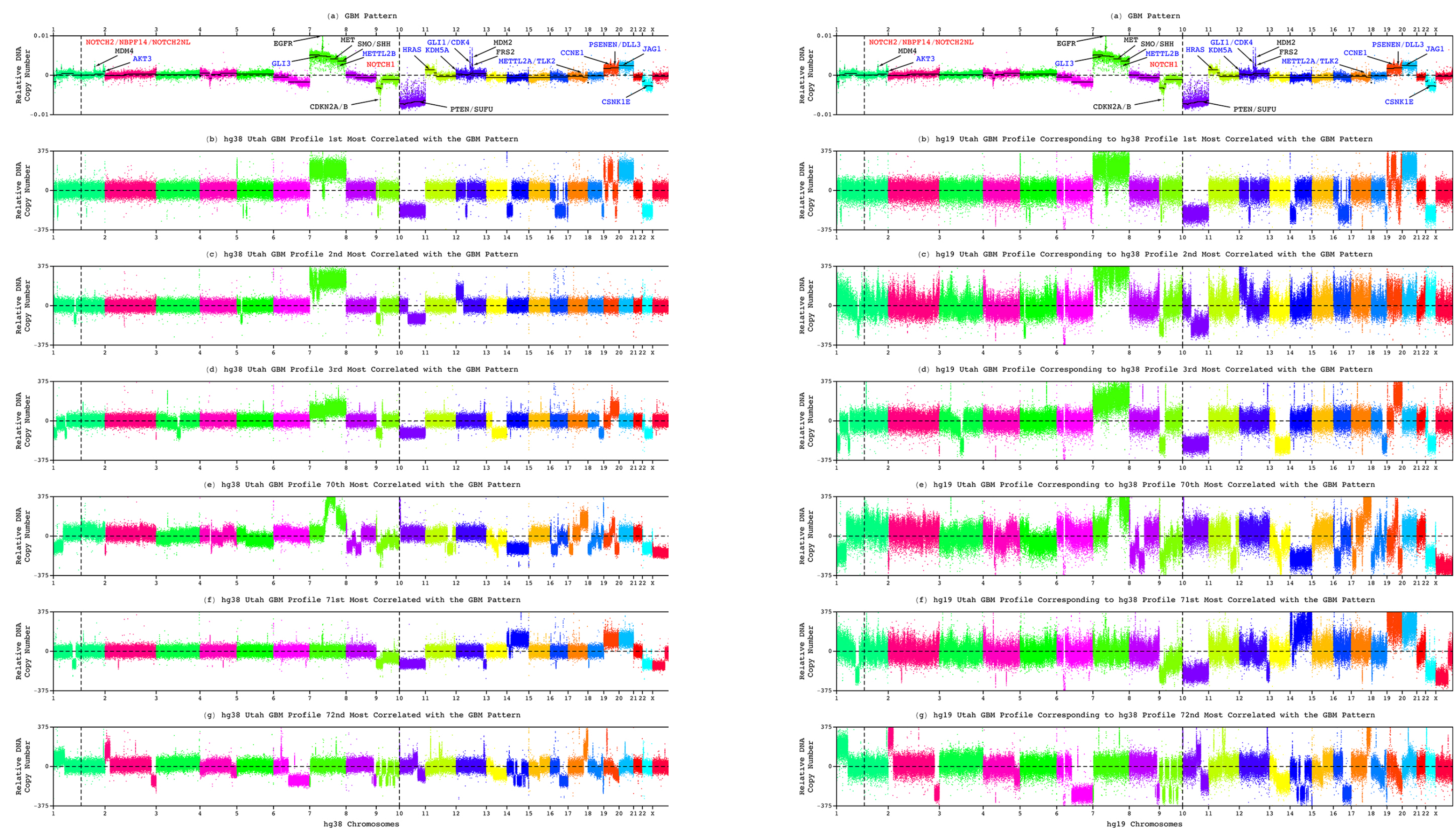
Patients (Number)	Predictor	KM Group	Median (Months)	Log-Rank <i>P</i> -Value	Cox Model	Hazard Ratio	95% Confidence Interval	Wald <i>P</i> -Value	Concordance Index
443	GBM Pattern (Corr.)	Low	29	1.4×10 ⁻⁶	Univariate	2.5	1.7– 3.6	2.9×10 ⁻⁶	0.75
	TCGA Age (Years)	High	13	4.9×10 ⁻¹⁰		2.2	1.7– 2.8	1.1×10 ⁻⁹	0.72
	GBM Pattern (Corr.)	<50	21	2.1×10 ⁻¹⁰	Bivariate	1.9	1.3– 2.8	1.7×10 ⁻³	0.71
	TCGA Age (Years)	≥50	12			1.9	1.4– 2.4	1.9×10 ⁻⁶	
		High/<50	18						
		Low/≥50	16						
284	GBM Pattern (Corr.)	Low	50	3.2×10 ⁻⁴	Univariate	2.5	1.5– 4.1	4.9×10 ⁻⁴	0.73
	TCGA Age (Years)	High	15	3.3×10 ⁻⁵		1.9	1.4– 2.6	4.3×10 ⁻⁵	0.64
	GBM Pattern (Corr.)	<50	22	5.5×10 ⁻⁵	Bivariate	1.8	1.1– 3.2	2.9×10 ⁻²	0.64
	TCGA Age (Years)	≥50	15			1.6	1.1– 2.2	6.8×10 ⁻³	
		High/<50	18						
		Low/≥50	15						
327	GBM Pattern (Corr.)	Low	39	2.8×10 ⁻⁵	Univariate	2.4	1.6– 3.8	4.8×10 ⁻⁵	0.73
	TCGA Age (Years)	High	15	6.7×10 ⁻⁶		1.9	1.4– 2.5	8.9×10 ⁻⁶	0.66
	GBM Pattern (Corr.)	<50	23	2.7×10 ⁻⁶	Bivariate	2.0	1.3– 3.1	3.0×10 ⁻³	0.65
	TCGA Age (Years)	≥50	15			1.6	1.2– 2.1	1.7×10 ⁻³	
		High/<50	20						
		Low/≥50	16						
255	GBM Pattern (Corr.)	Low	34	3.2×10 ⁻⁴	Univariate	2.5	1.5– 4.2	5.0×10 ⁻⁴	0.73
	TCGA <i>MGMT</i> Methylation	High	13	4.7×10 ⁻³		1.5	1.1– 2.0	5.0×10 ⁻³	0.58
	GBM Pattern (Corr.)	Yes	18	3.0×10 ⁻⁴	Bivariate	2.4	1.4– 4.0	1.1×10 ⁻³	0.61
	TCGA <i>MGMT</i> Methylation	No	13			1.4	1.1– 1.9	1.7×10 ⁻²	
		Low/Yes	39						
		Low/No	16						
329	GBM Pattern (Corr.)	High/Yes	16	9.6×10 ⁻⁵	Bivariate	1.9	1.2– 3.1	8.4×10 ⁻³	0.77
	TCGA <i>IDH1</i> Mutation	High/No	12			1.7	1.0– 3.0	4.1×10 ⁻²	
		Low/Yes	25						
		Low/No	24						
	GBM Pattern (Corr.)	High/Yes	24	2.6×10 ⁻⁴	Univariate	2.4	1.5– 3.9	4.0×10 ⁻⁴	0.78
	TCGA <i>IDH1</i> Mutation	High/No	14			2.4	1.5– 3.9	4.0×10 ⁻⁴	
107	GBM Pattern (Corr.)	Low	34	1.8×10 ⁻²	Univariate	3.7	1.2– 12.1	2.7×10 ⁻²	0.90
	TCGA <i>TERT</i> Expression	High	13	9.5×10 ⁻³		2.5	1.2– 5.1	1.2×10 ⁻²	0.77
	GBM Pattern (Corr.)	No	25	2.4×10 ⁻²	Bivariate				0.80
	TCGA <i>TERT</i> Expression	Yes	13						
		Low/No	34						
		Low/Yes	25						
335	GBM Pattern (Corr.)	High/No	24	3.0×10 ⁻⁵	Univariate	2.4	1.6– 3.6	4.9×10 ⁻⁵	0.72
	TCGA Karnofsky Score	High/Yes	12			4.1	2.1– 8.1	4.7×10 ⁻⁵	
		≥60	15						
		<60	5						
	GBM Pattern (Corr.)	Low/≥60	34	1.6×10 ⁻⁸	Bivariate	2.3	1.5– 3.5	6.9×10 ⁻⁵	0.74
	TCGA Karnofsky Score	High/≥60	15			3.8	1.9– 7.6	1.1×10 ⁻⁴	

Improving Prognostics, Diagnostics, and Therapeutics with a Genome-Wide Predictor of Survival

This is the first predictor that encompasses the whole tumor genome.

- The prognostic classification can help to manage GBM pseudoprogression.
Wen, in *UpToDate*, ed. Loeffler. Waltham, MA: Wolters Kluwer (2019);
Walter et al., *J Nucl Med* 53, 393 (2012).
- The diagnostic classification could help drugs progress to regulatory approval. Even if a drug targets just one gene, a patient's response to the drug would depend on the status of the whole genome and not just the one gene. The effect of *EGFR* deficiency in mice, e.g., depends upon the genetic background. In clinical trial of an *EGFR* inhibitor, the drug's effect was independent of the tumor's *EGFR* amplification. Only one drug has advanced from trials to standard of care over the last 40 years.
Threadgill et al., *Science* 269, 230 (1995); Rich et al., *J Clin Oncol* 22, 133 (2004).
Grossman & Ellsworth, *J Clin Oncol* 34, e13522 (2016).
- The therapeutic predictions, of previously unrecognized targets that are correlated with survival, e.g., the druggable *METLL2A/B* and *TLK2*, could lead to new drugs.
Kim et al., *Mol Cancer Res* 14, 920 (2016); Zhang, Guo & Boulianne, *Gene* 280, 135 (2001).

A Patient's Survival is the Outcome of Their Tumor's Whole Genome



Chromosome 10 deletion, chromosome 7 amplification, and chromosome arm 9p deletion, appear in the tumor genomes of some but not all 70 patients with high and, separately, some but not all nine patients with low correlations of their tumor profiles with the pattern.

Proof of Principle that Comparative Spectral Decompositions are Uniquely Suitable for Discovering Clinically Actionable, Accurate, and Precise Genotype–Phenotype Relationships

They have overcome three distinct challenges that other methods had not.

- They found consistent patterns across whole genomes, which have 3B nucleotides.
- They did that across the tumor and the matching normal genomes simultaneously.
- They did so in small cohorts of patients, about 100, that are typical in clinical trials.

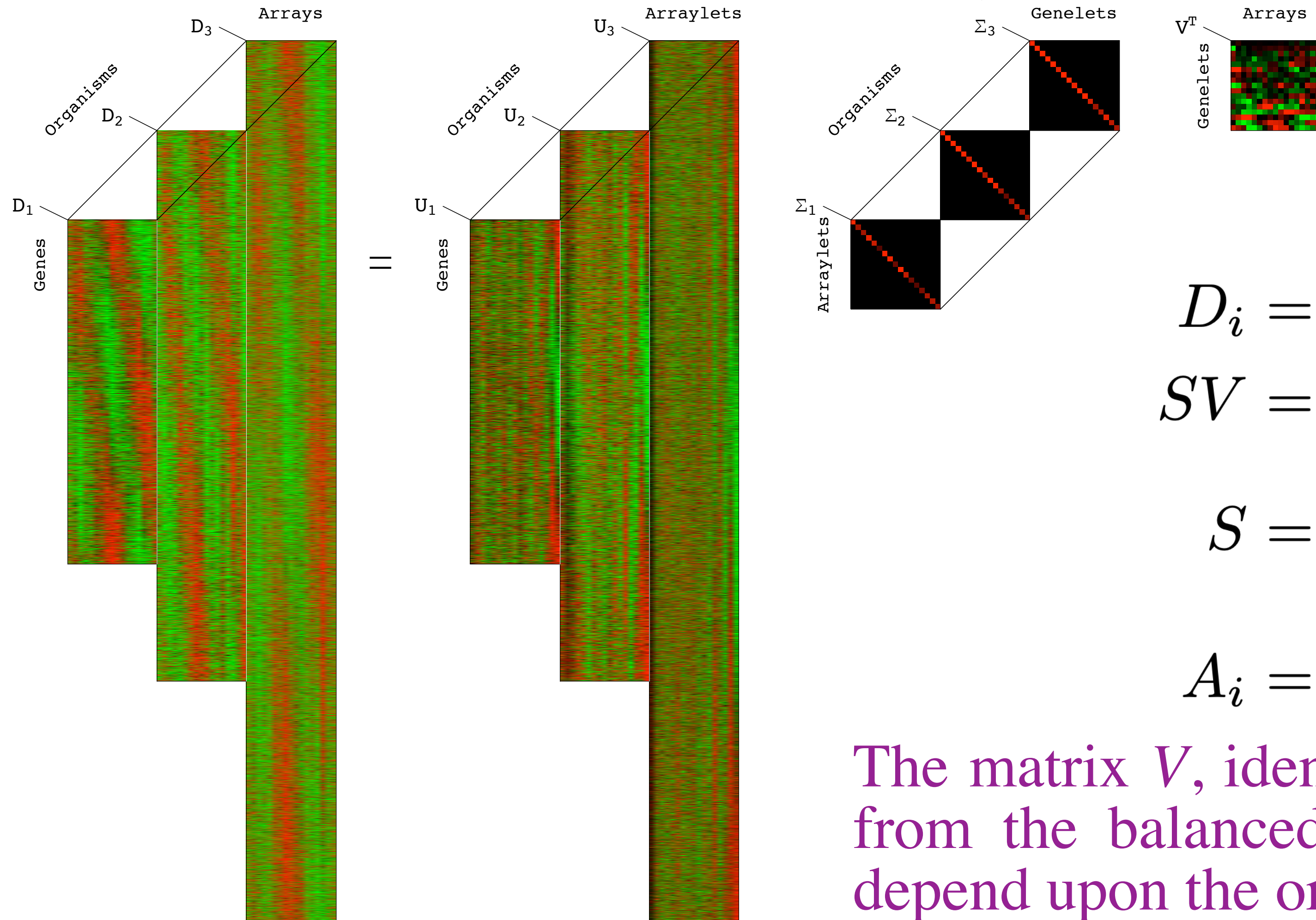
By using the complex structure of the datasets rather than simplifying or standardizing them as is commonly done, they can separate patterns which occur only in the tumor genomes from those that occur in the genomes of normal cells in the body and variations caused by experimental inconsistencies.

Less than one copy-number variation (CNV) to each 50 SNP associations with disease susceptibility. CNVs overlap $\approx 12\%$ of the normal human genome, are 10^2 – 10^4 times more frequent than point mutations, and are implicated in both normal and tumor development.

Conrad et al., *Nature* 464, 704 (2010); Lupski, *Nat Genet* 39, S43 (2007); Diskin et al., *Nature* 459, 987 (2009).

Higher-Order GSVD for Comparison of Multiple Second-Order Datasets

Ponnappalli, Golub & Alter, in *Stanford University and Yahoo! Research Workshop on Algorithms for Modern Massive Datasets* (Stanford, CA, June 21–24, 2006).



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

$$SV = V\Lambda,$$

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

$$A_i = D_i^T D_i, \quad i = 1, 2, \dots, N.$$

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 .

Higher-Order GSVD for Comparison of Multiple Second-Order Datasets

Ponnappalli, Saunders, Van Loan & Alter, *PLoS One* 6, e28072 (2011); https://alterlab.org/HO_GSVD/

The exact HO GSVD directly extends to multiple matrices all mathematical properties of the GSVD except for complete orthogonality of U_i for all i .

Supplementary Theorems 1–5:

For $N=2$, the HO GSVD algebraically leads to the GSVD.

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

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

Theorem 3: **The common HO GSVD subspace.** An eigenvalue satisfies $\lambda_k=1$ 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: $\lambda_k=1$ 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.

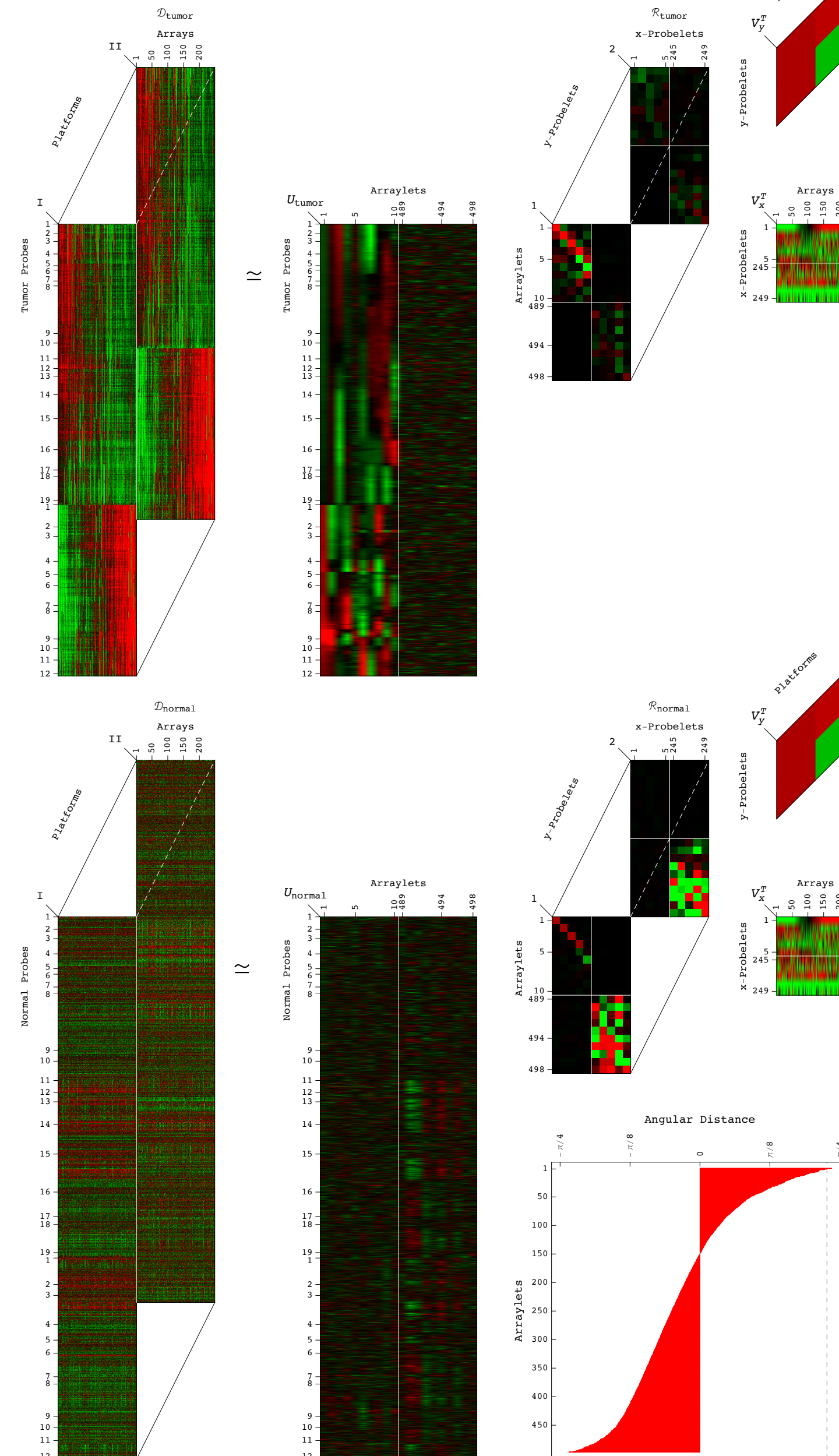
Chromosome Arms-Wide Predictor of Survival Throughout the Course of Adenocarcinomas

Bradley, Aiello, Ponnappalli,* Hanson* & Alter,
APL Bioeng 3, 036104 (2019).

New tumors are the leading cause of death from lung, uterine, and ovarian adenocarcinomas, where most patients experience progression-free survival after the primary treatment.

Yet, no indicator existed that predicts the benefit of platinum in terms of overall survival past the primary treatment.

6p+12p genotypes predictive of survival phenotypes, in general and in response to platinum, were uncovered by the GSVD and tensor GSVD.

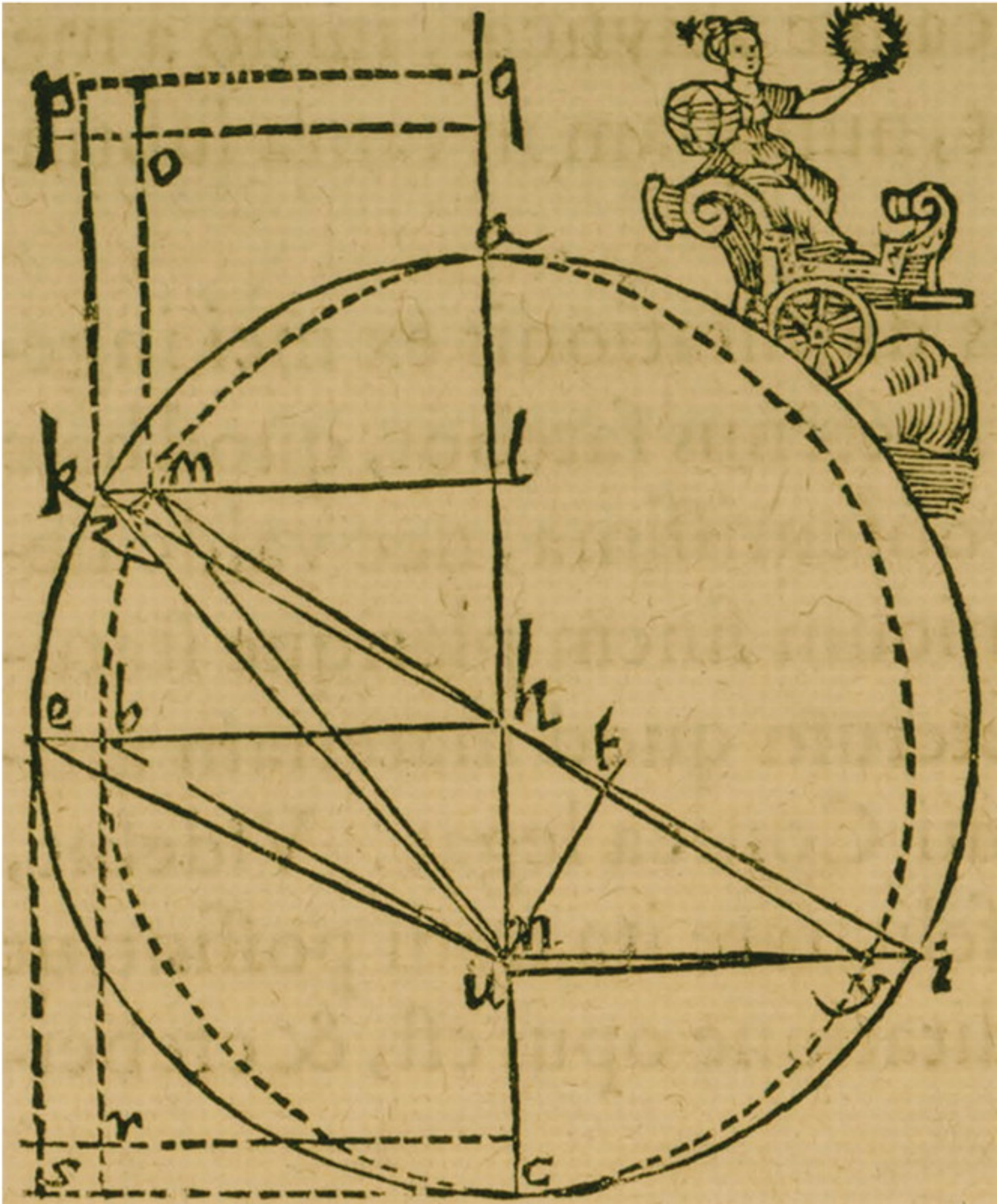


Global Patterns Underlie Principles of Nature

Alter, *PNAS* 103, 16063 (2006);
Alter, in *Microarray Data Analysis: Methods and Applications*. New York, NY: Humana Press (2007).

Kepler's discovery of his first law of planetary motion from mathematical modeling of Brahe's astronomical data.

	Tempus	Locus ☉	Solis a Terra distantia	Martis a Sole distantia
1582.	23 Nove.H. 16. 0	11.41 ♄	98345	158852
	26 Dece.H. 8.30	15. 4 ♄	98226	162104
	30 Dece.H. 8.10	19. 9 ♄	98252	162443
1583.	26 Janua.H. 6.15	16.33 ♄	98624	164421
1584.	21 Dece.H. 14. 0	10.16 ♄	98207	164907
1585.	24 Janua.H. 9. 0	14.53 ♄	98595	166210
	4 Febr.H. 6.40	26.10 ♄	98830	166400
	12 Mart.H. 10.30	2.16 ♄	99858	166170
1587.	25 Janua.H. 17. 0	16. 1 ♄	98611	166232
	4 Mart.H. 13.24	24. 0 ♄	99595	164737
	10 Mart.H. 11.30	29.52 ♄	99780	164382
	21 April.H. 9.30	10.48 ♄	101010	161027
1589.	8 Mart.H. 16.24	28.36 ♄	99736	161000
	13 April.H. 11.15	3.38 ♄	100810	157141
	15 April.H. 12. 5	5.36 ♄	100866	156900
	6 Maji.H. 11.20	25.49 ♄	101366	154326
1591.	13 Maji.H. 14. 0	2.10 ♄	101467	147891
	6 Junii H. 12.20	24.59 ♄	101769	144981
	10 Junii H. 11.50	28.47 ♄	101789	144526
	28 Junii H. 10.24	15.51 ♄	101770	142608
1593.	21 Julii H. 14. 0	8.26 ♄	101498	138376
	22 Aug.H. 12.20	9.11 ♄	100761	138463
	29 Aug.H. 10.20	11.54 ♄	100562	138682
	3 Octo.H. 8. 0	20.15 ♄	99500	140697
1595.	17 Sept.H. 16.45	4.18 ♄	99990	143222
	27 Octo.H. 12.20	13.59 ♄	98851	147890
	3 Nove.H. 12. 0	21. 2 ♄	98694	148773
	18 Dece.H. 8. 0	6.43 ♄	98200	154539



Kepler, *Astronomia Nova*. Heidelberg, Germany: Voegelinus (1609).

Collaborators:

Jill S. Barnholtz-Sloan
Population Health Sciences, CWRU

Jay Bowen
Laboratory Medicine, NCH

Julie M. Gastier-Foster
Laboratory Medicine, NCH

Randy L. Jensen
Neurosurgery, Utah

Cheryl A. Palmer
Pathology, Utah

Carl T. Wittwer
Pathology, Utah

Huanming Yang
Medical Genetics, BGI

John F. X. Diffley
Cancer Research UK, London

Gene H. Golub
Computer Science, Stanford

Michael A. Saunders
Operations Research, Stanford

Charles F. Van Loan
Computer Science, Cornell

K07 Research Assistant Professor:

Heidi A. Hanson, UPDB

**Senior Research Affiliate
& Ph.D. Alumna:**

Sri Priya Ponnappalli, ECE

**K99 Postdoctoral Alumnus:
Jason M. Tennessen, Genetics**

Ph.D. Alumni:
Larsson Omberg, Physics
Kayta Kobayashi, Pharmacy
Chaitanya Muralidhara, CMB

B.Sc. Alumni:
Joel R. Meyerson, BME & Gov
Andrew M. Gross, BME & SSC
Justin A. Drake, BME & SSC
Nicolas M. Bertagnolli, Mathematics

Support:
NHGRI K01 HG-000038
NHGRI R01 HG-004302
NSF CAREER DMS-0847173

NCI U01 CA-202144

<https://physics.cancer.gov/network/UniversityofUtah.aspx>

**Thank
you!**

Physics-Inspired Multi-Tensor Decompositions

Create a single coherent model from multiple high-dimensional diverse datasets at once. By using the complex structure of the datasets, rather than simplifying them as is commonly done, the multi-tensor decompositions can:

- blindly detect and remove experimental artifacts or batch effects;
- blindly identify and separate the biologically similar from the dissimilar;
- discover previously unknown phenomena.

Directly generalize the SVD from a single two-dimensional dataset to multiple three- and higher-dimensional datasets. The SVD underlies:

- theoretical physics;
- recommendation systems, e.g., PageRank and the Netflix challenge.

Physics-Inspired Multi-Tensor Decompositions

Find what other methods miss, and outperform methods that:

- require large amounts of training data (e.g., deep learning);
- require training and are sensitive to imbalanced class representations (e.g., supervised learning);
- require data quantization and are sensitive to cutoff selections (e.g., Bayesian statistics and topological data analysis);
- vary the one-dataset SVD and are, therefore, not exact or unique, rather than use the complex structure of the data (e.g., independent component analysis, sparse and nonnegative factorizations, and randomized decompositions);
- are unsupervised but require data cleaning and are sensitive to artifacts and batch effects (e.g., hierarchical clustering);
- are supervised and require a-priori knowledge (e.g., analysis of variance).

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

The SVD is used for the stable computation of principal component analysis (PCA).

The SVD is Different Than PCA

→ **PCA assumes preprocessing of the data, which limits the data interpretation (e.g., the SVD of a dataset can identify the probability distribution function that is sampled by the dataset with no a-priori assumptions; PCA cannot).**

Alter & Golub, *PNAS* 103, 11828 (2006); https://alterlab.org/harmonic_oscillator/

Cadima & Jolliffe, *Pak J Statist* 25, 473 (2009);

Muralidhara, Gross, Gutell & Alter, *PLoS One* 6, e18768 (2011); <https://alterlab.org/rRNA/>

Bertagnolli et al., *PLoS One* 8, e78913 (2013); https://alterlab.org/GBM_metabolism/

→ **PCA identifies patterns across the columns separately from patterns across the rows; the SVD simultaneously computes the corresponding sets of patterns across the rows and columns, ensuring consistent data interpretation.**

Alter et al., *PNAS* 97, 10101 (2000); https://alterlab.org/singular_value_decomposition/

Alter et al., in *Microarrays: Optical Technologies and Informatics*. Bellingham, WA: International Society for Optics and Photonics (SPIE) (2001); <https://alterlab.org/SVD/>

Fellenberg et al., *PNAS* 98, 10781 (2001).

→ **PCA, as it is programmed in most computational packages, is limited to classifying the data based upon the two or three patterns that capture most of the information in the data (e.g., variance in the case of column centering); the SVD maintains all data patterns, and not just for data classification.**

There are nontrivial connections between the GSVD and canonical correlations analysis (CCA).

The GSVD is Different Than CCA

Ewerbring & Luk, *J Comput Appl Math* 27, 37 (1989).

De Clercq et al., *IEEE Trans Biomed Eng* 53, 2583 (2006);

Rustandi, Just & Mitchell, in *MICCAI 2009* (London, UK, September 20–24, 2009);

Tenenhaus & Tenenhaus, *Psychometrika* 76, 257 (2011).