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

Orly Alter

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/
The adoption of omic technologies in the cancer clinic is giving rise to an increasing number of large-scale high-dimensional datasets recording multiple patient-matched aspects of the disease.

Collins & Hamburg, *N Engl J Med* **369**, 2369 (2013); Rehm, et al., Lyon, *Genet Med* **15**, 733 (2013).
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, Ph.D., is currently a postdoctoral fellow in the Department of Genetics at Stanford University. Yoshihisa Yamamoto, Ph.D., is a professor in the Departments of Applied Physics and Electrical Engineering at Stanford University. He is currently the director of the JST/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 JSTO Quantum Entanglement 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.
The singular value decomposition (SVD) underlies the theoretical description of the physical world.

Generalizations of the SVD can be formulated that integrate and compare different data types.

The computations of the SVD and its generalizations scale 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 that underlie, i.e., compose, 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 and 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.
Simultaneously separate, i.e., spectrally decompose, 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**
Ponnapalli, 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 al., 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

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 int., Nelson, *J Natl Cancer Inst* **85**, 704 (1993); Gorlia, et int., Stupp, *Lancet Oncol* **9**, 29 (2008).

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 int., Cremer, *Lab Invest* **74**, 108 (1996); Wiltshire, et int., Bigner, *Neuro Oncol* **2**, 164 (2000); Misra, et int., Feuerstein, *Clin Cancer Res* **11**, 2907 (2005).
Mathematically Universal  
Biologically Consistent  
Genotype-Phenotype

Lee, et al., Alter, *PLoS One* 7, e30098 (2012); Aiello & Alter, *PLoS One* 11, e0164546 (2016); Aiello, Ponnapalli & Alter, *APL Bioeng* 2, Special Topic: Bioengineering of Cancer 031909 (2018).
Invariably Uncovered by and Only by the Generalized SVD (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 Ras, Shh, and Notch

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.

Waldman, et al., Vogelstein, *Nature* 381, 713 (1996); Hahn, et al., Weinberg, *Nature* 400, 464 (1999); Hayden Gephart, et al., Scott, *J Neurooncol* 115, 61 (2013); Irianto, et al., Discher, *Curr Biol* 27, 210 (2017); Mukherjee, et al., Fonkem, *J Neuropathol Exp Neurol* 79, 562 (2020).
Guanine-cytosine (GC) content effects vary in magnitude between batches.

Roberts, Carneiro & Schatz, *Genome Biol* 14, 405 (2013).
WGS with Affymetrix single-nucleotide polymorphism (SNP) and Agilent comparative genomic hybridization (CGH) microarrays represent the main genomic profiling technologies.
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 is Statistically Representative of the U.S. Adult GBM Population

| Phenotype | Group          | Utah Set | CWRU Set | TCGA Set | SEER Set | \( \chi^2 \) P-Value | Utah vs. SEER \( \chi^2 \) P-Value | CWRU vs. SEER \( \chi^2 \) P-Value | TCGA vs. SEER \( \chi^2 \) P-Value |
|-----------|----------------|----------|----------|----------|----------|-----------------------|-----------------------------------|-----------------------------------|-----------------------------------|
| Normal    | Sex            |          |          |          |          |                       |                                   |                                   |                                   |
|           | Female         | 27       | 12       | 169      | 3331     | 1.8 \( \times 10^{-1} \) | 9.0 \( \times 10^{-1} \)         | 1.5 \( \times 10^{-1} \)         |                                   |
|           | Male           | 52       | 16       | 274      | 4670     | 1.2 \( \times 10^{-1} \) | 8.4 \( \times 10^{-1} \)         | 1.1 \( \times 10^{-1} \)         |                                   |
| Race      | Not White      | 74       | 25       | 386      | 7042     | 1.7 \( \times 10^{-1} \) | 1.7 \( \times 10^{-1} \)         | 4.4 \( \times 10^{-3} \)         |                                   |
|           | White          | 5         | 3        | 40       | 959      | 1.2 \( \times 10^{-1} \) | 8.4 \( \times 10^{-1} \)         | 1.1 \( \times 10^{-1} \)         |                                   |
| Ethnicity | Hispanic       | 2         | 0        | 11       | 507      | 1.7 \( \times 10^{-1} \) | 1.7 \( \times 10^{-1} \)         | 4.4 \( \times 10^{-3} \)         |                                   |
|           | Not Hispanic   | 77        | 28       | 382      | 7494     | 1.7 \( \times 10^{-1} \) | 1.7 \( \times 10^{-1} \)         | 4.4 \( \times 10^{-3} \)         |                                   |
| Disease   | Age (Years)    |          |          |          |          |                       |                                   |                                   |                                   |
|           | <50            | 10        | 3        | 117      | 1164     | 6.4 \( \times 10^{-1} \) | 5.7 \( \times 10^{-1} \)         | 1.2 \( \times 10^{-11} \)        |                                   |
|           | \geq 50        | 69        | 25       | 326      | 6837     | 6.4 \( \times 10^{-1} \) | 5.7 \( \times 10^{-1} \)         | 1.2 \( \times 10^{-11} \)        |                                   |
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., Feuk, *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.
**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 int., Ikenaga, *J Neurooncol* 21, 135 (1994); Parsons, et int., Kinzler, *Science* 321, 1807 (2008); Nguyen, et int., Lai, *Neuro Oncol* 19, 394 (2017); Batchelor & Louis, in *UpToDate*, eds. Loeffler & Wen. Waltham, MA: Wolters Kluwer (2018).
Personalized 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 and avoid unnecessary surgery.
  Wen, in *UpToDate*, ed. Loefferl. Waltham, MA: Wolters Kluwer (2019); Patel, et int., Ellingson, *J Neurooncol* 139, 399 (2018); Walter, et int., Czernin, *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 a clinical trial of an *EGFR* inhibitor, the drug’s effect was independent of the tumor’s *EGFR* amplification. Only one new drug has advanced from trials to standard of care over the last 40 years.
  Threadgill, et int., Magnuson, *Science* 269, 230 (1995); Rich, et int., Friedman, *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 int., Wang, *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, i.e., Genetic Background

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 the Multi-Tensor Decompositions are Uniquely Suited for Discovering Accurate and Precise Actionable Genotype-Phenotype Relationships Relevant to the Population Based upon Small Cohorts

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) per 50 SNP associations with disease susceptibility. CNVs overlap ≈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

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

\[ D_i = U_i \Sigma_i V^T, \quad \Sigma_i = \text{diag}(\sigma_i,k), \]

\[ SV = VA, \]

\[ S = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j>i}^{N} (A_iA_j^{-1} + A_jA_i^{-1}), \]

\[ A_i = D_i^T D_i, \quad i = 1, 2, \ldots, N. \]

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

Ponnapalli, 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 $i$ and $j$.

Supplementary Theorem 6 and Conjecture 1:

A role in iterative approximation algorithms.
New tumors, e.g., metastasis, 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 primary tumor’s genotypes predictive of the patient’s overall survival phenotypes, in general as well as following platinum treatment of the primary tumor, and throughout the course of the disease, were discovered by the GSVD and tensor GSVD.
Correlations to Causal Coordination: Global Patterns Underlie Principles of Nature

Alter, *PNAS* 103, 16063 (2006);
Alter, in *Microarray Data Analysis: Methods and Applications*. Berlin, Germany: Springer Nature (2007).

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

Kepler, *Astronomia Nova*. Heidelberg, Germany: Voegelinus (1609).
Collaborators:
Jill S. Barnholtz-Sloan
Population Health Sciences, CWRU
Jay Bowen & 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
Yoshihisa Yamamoto
Applied Physics, Stanford

K07 Research Assistant Professor:
Heidi A. Hanson, UPDB
Senior Research Affiliate & Ph.D. Alumna:
Sri Priya Ponnapalli, 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 & 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).
Ringnér, Nat Biotechnol 26, 303 (2008).

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 int., Alter, 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., in Microarrays: Optical Technologies and Informatics. Bellingham, WA: International Society for Optics and Photonics (SPIE) (2001); https://alterlab.org/SVD/
Fellenberg, et int., Vingron, 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., Van Huffel, *IEEE Trans Biomed Eng* **53**, 2583 (2006);
Rustandi, Just & Mitchell, in *12th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* (London, UK, September 20–24, 2009);
Tenenhaus & Tenenhaus, *Psychometrika* **76**, 257 (2011).