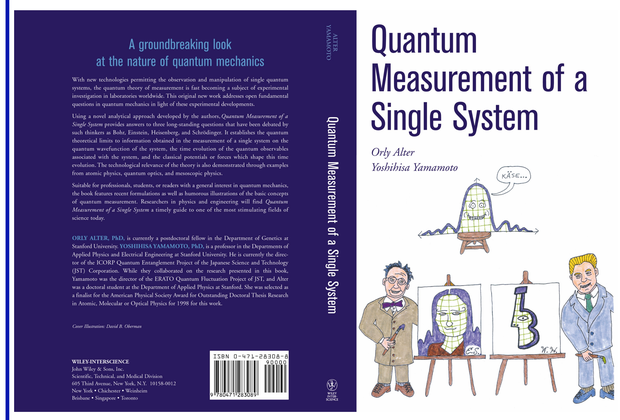
AI/ML-Derived Mechanistically-Interpretable Whole-Genome Biomarkers of Patient Survival in Pre-Treatment Primary Neuroblastoma Tumors and Whole Blood

Orly Alter,[1]* Elizabeth Newman,[2] Sri Priya Ponnapalli[1] and Jessica W. Tsai[3]

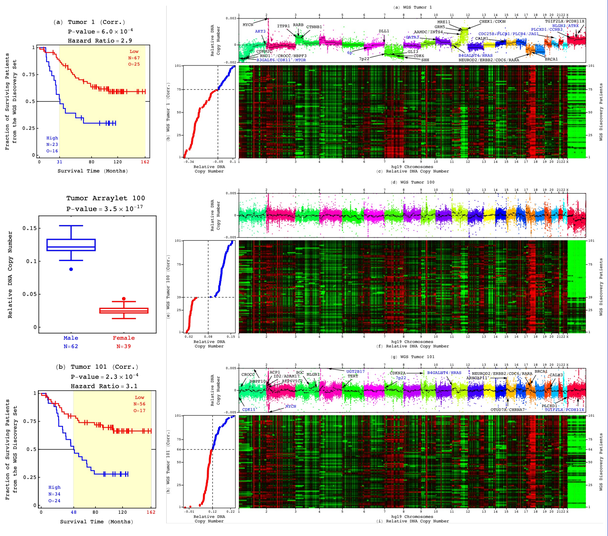
[1]University of Utah, Salt Lake City, UT, [2]Emory University, Atlanta, GA and [3]Children's Hospital of Los Angeles and University of Southern California, Los Angeles, CA; ***orly@sci.utah.edu**

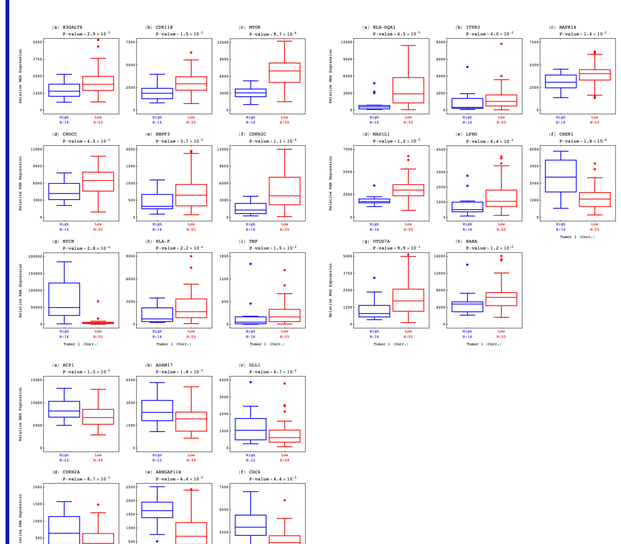
Background: Prediction, understanding, and management, of neuroblastoma (NBL) outcomes, from spontaneous regression to relapse and death, remain limited, and rely mostly on the International Neuroblastoma Staging System (INSS) stage, age, and the one-gene test for *MYCN* amplification. The entire multi-ome affects every aspect of the disease. But typical artificial intelligence and machine learning (AI/ML) cannot discover effective biomarkers from real-life, small-cohort, multi-omic, and noisy data.



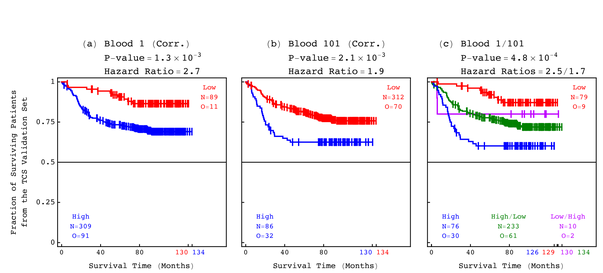
Quantum Measurement of a Single System

Orly Alter
Yoshihisa Yamamoto



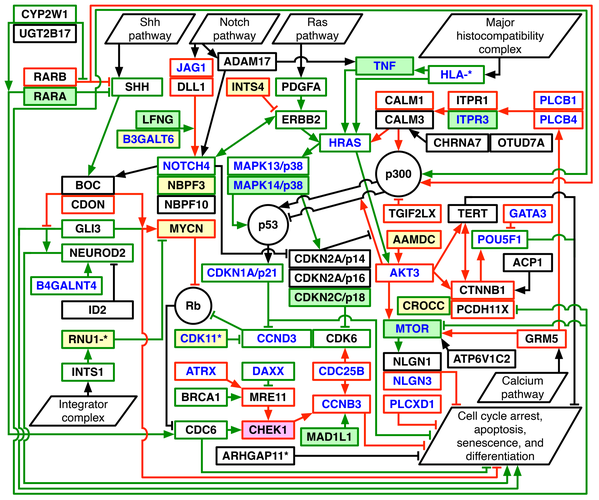


Methods: We use our AI/ML to identify two whole-genome biomarkers from open-source pre-treatment NBL patient-matched primary tumor and whole blood DNA copy numbers, and primary tumor mRNA expression. Our data-agnostic, unsupervised algorithms extend the mathematics that underlies quantum mechanics to overcome the limitations of typical AI/ML [Bradley, et int., Alter, *APL Bioeng* 2019; Alter et al., *PNAS* 2003]. Our platform- and reference genome-agnostic biomarkers in astrocytoma, including glioblastoma, and lung, ovarian, and uterine adenocarcinomas, were mathematically discovered, and computationally and experimentally validated, repeatedly, in federated, imbalanced datasets from as few as 50–100 patients [Ponnapalli, et int. Alter, *APL Bioeng* 2020].



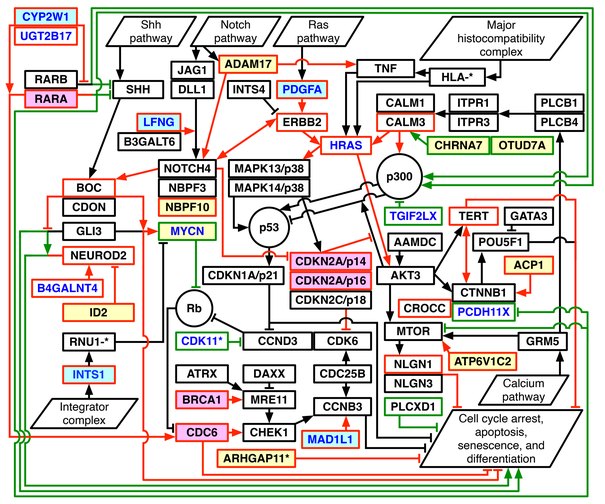
Results: We discover the biomarkers in "skinny" datasets of minimally preprocessed ≈3M-bin whole-genome sequences from tumors and blood of a set of 101 patients. Blindly, i.e., label-free, we separate both biomarkers from the normal X chromosome-number variation, demonstrating sex-agnostic learning. Combined, the biomarkers are statistically independent of all standard-of-care indicators, with the univariate Cox hazard ratio of 4.0 (log-rank *P*-value $=2.3\times10^{-5}$) within the 95% confidence intervals of the bivariate ratios. The risk that the tumor's whole genome confers upon outcome, as is reflected by the bivariate ratios, is greater than that conferred by all standard-of-care indicators, including DNA ploidy, the Children's Oncology Group (COG) risk, and the mitosis-karyorrhexis index (MKI), except for histopathology.

We validate the biomarkers in ≈10K-bin target-capture sequences of a mutually-exclusive set of 419 patients, demonstrating generalizability as well as site-, platform-, and protocol-agnostic transfer learning. At 73–80% concordance, in both the tumor and blood profiles of both the discovery and validation sets, the biomarkers combined are more accurate than *MYCN*.



We show that the biomarkers identify known and previously unrecognized disease mechanisms and druggable gene alterations, including co-amplification of *MYCN* with genes encoding for extra-embryonic transcripts, as well as the hijacking of embryonic development toward aneuploidy, which can spontaneously regress via embryonic self-correction.

Conclusions: This is a proof of principle that our AI/ML is uniquely suited for personalized medicine.



Acknowledgements: This work was funded by National Cancer Institute (NCI) grant U01 CA-202144 and by Utah Science, Technology, and Research (USTAR) support. Some of the work was done at the Center for High-Performance Computing (CHPC) at the University of Utah. O. A. is a co-founder of and an equity holder in Prism AI. This does not alter our adherence to the policies of the American Society of Clinical Oncology (ASCO) on sharing data, code, and materials.

THE UNIVERSITY OF UTAH

2024 ASCO® ANNUAL MEETING

SC22

www.sci.utah.edu