Machine Learning-based Detection of Concurrent DNA Repair Deficiencies as Dynamic Biomarkers for Prostate Cancer Precision Medicine

Komal Sharma[1,2], Divin A. Wilson[1,2,#], Yu Wang[3,4,#], Shabi Haidar[5], Daniel Boiarsky[6], John M. Cipriaso[7], Leo Yermakov[8], Ilsa M. Coleman[9], Armand Bankhead[9,10,11], Bhabishya Neupane[2,12], Bradley W. Taylor[12], Colm Morrissey[13], Michael. T. Schweitzer[10,14,] , Robert B. Montgomery[10,14], Ariel A. Nelson[5], Emmanuel S. Antonarakis[15], Jacob E. Berchuk[6], Kevin K. Zarrabi[7], Colin C. Pritchard[8], Gavin Ha[9,10,11] , Anai Kothari[2,4,16], Razelle Kurzrock [2,17], Peter S. Nelson[8,9,18], Paul L. Auer[2, 3, 4], Deepak Kilari[2,5,17], Navonil De Sarkar[1,2,4,19]

1: Department of Pathology, Medical College of Wisconsin, WI
2: Cancer Center, Medical College of Wisconsin, WI
3: Division of Biostatistics, Medical College of Wisconsin, WI
4: Data Science Institute, Medical College of Wisconsin, WI
5: Division of Hematology & Oncology, Medical College of Wisconsin, WI
6: Department of Hematology & Oncology, Yale New Haven Hospital, Boston, MA
7. Sidney Kimmel Cancer Center at Thomas Jefferson University, Philadelphia, PA
8: Department of Laboratory Medicine & Pathology, University of Washington, Seattle, WA
9. Division of Human Biology, Fred Hutchinson Cancer Canter, Seattle, WA
10. Clinical Research Division, Fred Hutchinson Cancer Canter, Seattle, WA
11. Public Health Sciences Division, Fred Hutchinson Cancer Canter, Seattle, WA
12. Center for Biomedical Informatics, Medical College of Wisconsin, WI
13. University of Washington, Department of Urology, Seattle, WA
14. Division of Hematology & Oncology, University of Washington, Seattle, WA
15. Division of Hematology, Oncology, and Transplantation, University of Minnesota, MN
16. Department of Surgery, Medical College of Wisconsin, WI
17. Clinical Cancer Center, Froedtert Hospital, WI
18. Stuart and Molly Sloan Precision Oncology Institute, Fred Hutchinson Cancer Canter, Seattle, WA
19. Department of Pharmacology & Toxicology, Medical College of Wisconsin, WI

Abstract

Background: Current DNA damage response (DDR) biomarkers rely on binary classifications that inadequately capture the molecular complexity of advanced prostate cancer, particularly concurrent phenotypes such as mismatch repair deficiency co-occurring with homologous recombination deficiency. Even in biomarker-selected populations, responses to targeted therapies remain highly variable, reflecting the limitations of static cutoff-based approaches that fail to address overlapping DDR pathway alterations and their complex therapeutic implications.

Methods: We analyzed 672 advanced prostate cancer exomes and identified TMB-very high (≥18.61 mutations/Mb) as a molecularly distinct subset from conventional TMB-high tumors through binary logistic regression with three-fold cross-validation. To address DDR complexity, we developed MCW CHIMERA-DDR, a probabilistic machine learning model integrating 51 molecular features using nested Random Forest architecture. CHIMERA employs two-tier classification: Tier 1 decomposes patients into four primary DDR components (TMB-vH, TMB-H, HRRd, DDR-intact), while Tier 2 provides refined subcategorization based on tumor suppressor gene status and specific pathway alterations. Clinical validation was performed using an independent cohort of 130 patients.

Results: CHIMERA-DDR demonstrated exceptional classification performance across all DDR phenotype categories through leave-one-out cross-validation, achieving AUCs of 0.999 for TMB-vH identification, 0.998 for TMB-H prediction, and 0.982 for Tier 2 HRRd detection. Fine-scale homologous recombination-deficient subset deconvolution demonstrated 0.919 AUC, indicating robust performance for stratifying patients with complex overlapping DDR phenotypes. TMB-very high tumors were characterized by preserved genomic integrity, enhanced immunogenicity, and inverted prognostic marker relationships.

Concurrent TMB-high and homologous recombination-mutant tumors bifurcated into MMRd-predominant or HRd-predominant subclasses with distinct molecular characteristics. Biological validation using tumor cellularity analysis confirmed that CHIMERA's probabilistic outputs reflect genuine tumor-intrinsic DDR functional states rather than normal tissue contamination. Semi-independent validation using paired whole-exome and whole-genome sequencing data from 52 samples across 24 patients demonstrated substantial intragroup homogeneity in deconvolved DDR predictions across multiple metastatic sites.

Clinical Significance: Clinical validation demonstrated superior immunotherapy responses in TMB-very high tumors, revealing the therapeutic significance of fine-scale DDR stratification. CHIMERA enables precision therapeutic stratification where patients with concurrent molecular features can be guided by dominant molecular phenotypes rather than mutation status alone. Tumors with predominant MMRd-like features may benefit from immunotherapy monotherapy, while those with concurrent characteristics represent optimal candidates for rationally designed combination approaches. This probabilistic framework addresses the clinical challenge of variable treatment responses in molecularly complex DDR-altered tumors, establishing a foundation for optimized precision medicine through dynamic biomarker assessment beyond binary classifications.

Funding Acknowledgement: This research was supported by the MCW Department of Pathology and Cancer Center startup grant provided to Navonil De Sarkar and ACS IRG Award (ACS-IRG #22-151-37). This research was executed in part with computational resources and technical support provided by the Research Computing Center at the Medical College of Wisconsin and supported in part by the Medical College of Wisconsin Cancer Center Biostatistics Shared Resource. We extend special thanks to William Branson and Matthew Flister from the MCW Research Computing Core, and Jason Erb from MCW IS for facilitating the setup of computational terminals, platforms, and environments, and for their consistent support throughout this data-intensive research. CCP is supported by NIH SPORE CA097186 supports CCP.

COI: E.S.A. reports grants and personal fees from Janssen, Johnson & Johnson, Sanofi, Bayer, Bristol Myers Squibb, Convergent Therapeutics, Curium, MacroGenics, Merck, Pfizer, and AstraZeneca; personal fees from Aadi Bioscience, Abeona Therapeutics, Aikido Pharma, Astellas, Amgen, Blue Earth, Boundless Bio, Corcept Therapeutics, Duality Bio, Exact Sciences, Hookipa Pharma, Invitae, Eli Lilly, Foundation Medicine, Menarini-Silicon Biosystems, Tango Therapeutics, Tempus, Tolmar Scientific, VIR Biotechnology, and Z-alpha; grants from Novartis, Celgene, and Orion; and has a patent for an AR-V7 biomarker technology that has been licensed to Qiagen. J.E.B. is an advisor to Genome Medical, Guardant Health, Oncotect, Precede Biosciences, Tracer. Biotechnologies, and Musculo. NDS: has one patent 22-172-US-PSP2: Cell-free DNA sequence data analysis method to examine nucleosome protection and chromatin accessibility. *MCW-CHIMERA-DDR holds provisional protection under PCT.* *.* Other authors do not have any relevant COI to disclose.