Title: Moving Beyond Binary Biomarkers: Machine Learning Model Resolves Concurrent and Molecularly Heterogeneous Mismatch-Repair and Homologous-Recombination Deficiencies in Prostate Cancer

Komal Sharma[1,2], Divin A. Wilson[1,2,#], Yu Wang[3,4,#], Shabi Haider[5*], Vineel Bhatlapenumarthi[5*], Daniel Boiarsky[6], John M. Cipriaso[7], Leo Yermakov[8], Rajanya Dutta[3], Ilsa M. Coleman[9,10], Armand Bankhead III[9,11], Shantanu K. Jadhav[1,2], Bhabishya Neupane [2,12], Bradley W. Taylor[12], Colm Morrissey[13], Michael. T. Schweitzer[14, 15], Robert B. Montgomery[14, 15], Sridhar Rao[16], Marja T. Nevalainen[7], Ariel A. Nelson[5], Emmanuel S. Antonarakis[17], Patrick G. Pilea[18], Jacob E. Berchuk[19], Kevin K. Zarrabi[7], Colin C. Pritchard[8], Anai Kothari[2,4,12,20], Hui-Zi Chen[2,21] , Ben George[22], Razelle Kurzrock [2,21], Gavin Ha[9,10,11], Peter S. Nelson[9,11,14], Paul L. Auer[2, 3, 4], Anjishnu Banerjee[3, 4], Deepak Kilari[2,5], Navonil De Sarkar[1,2,4,23]

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 Comprehensive 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 Center, Seattle, WA
10. Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA
11. Stuart and Molly Sloan Precision Oncology Institute, Fred Hutchinson Cancer Research Center, Seattle, WA
12. Center for Biomedical Informatics, Medical College of Wisconsin, WI
13. Department of Urology, University of Washington, Seattle, WA
14. Clinical Research Division, Fred Hutchinson Cancer Center, Seattle, WA
15. Division of Hematology & Oncology, University of Washington, Seattle, WA
16. Versiti Blood Research Institute, Milwaukee, WI
17. Department of Medicine, Division of Hematology, Oncology, and Transplantation, Masonic Cancer Center, University of Minnesota, Minneapolis, MN
18. Department of Genitourinary Medical Oncology, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX
19. Department of Hematology and Medical Oncology, Emory University School of Medicine, Atlanta, GA
20. Department of Surgery, Medical College of Wisconsin, WI
21. Linda T. and John A. Mellowes Center for Genomic Sciences and Precision Medicine
22. Department of Medical Oncology, Mayo Clinic Comprehensive Cancer Center, MN
23. Department of Pharmacology & Toxicology, Medical College of Wisconsin, WI
#: co-second author, contributed equally

Disclosures: ESA. 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. JEB. is an advisor to Genome Medical, Guardant Health, Oncotect, Precede Biosciences, Tracer Biotechnologies, and Musculo. PSN. has received research support from Janssen and consulting fees from Genentech, Pfizer, and AstraZeneca. NDS. holds patent 22-172-US-PSP2: Cell-free DNA sequence data analysis method to examine nucleosome protection and chromatin accessibility. KS., DAW., YW., AB., PA., NDS. are co-inventors of CHIMERA-DDR -holds provisional patent U.S. Application No. 63/882,124. Other authors do not have any relevant COI to disclose. CM. has received research funds from Genentech, Janssen and Novartis unrelated to the present work. DK reports consulting or advisory role: Pfizer, Aveo Oncology, Bayer, Eisai, Foundation medicine, Exelixis, Janssen, Binaytara foundation, Integra Connect, MJH life sciences, Aptitude; Myovant. Research funding(institutional): Exelixis, Astellas, Medivation, Genentech, SOBI.

Abstract:

Current DNA damage repair (DDR) biomarkers employ binary classifications that fail to capture the molecular complexity of tumors with concurrent repair deficiencies. We used genomics analysis to stratify 672 metastatic prostate cancer patients into 11 DDR subgroups, identifying 51 molecular signatures with weighted roles in class identity. We identified a tumor-mutational-burden very-high subset, characterized by 19 mutations/Mb or more, as a molecularly distinct group characterized by preserved genomic integrity and enhanced immunogenicity. Critically, 2.3 percent of tumors exhibited concurrent TMB-High and HRR mutant phenotypes, while 1.5 percent harbored MMR bi-allelic loss without MMRd (mismatch-repair-deficiency) signatures. Clinical validation in 130 patients demonstrated superior immunotherapy responses in tumors with very high TMB levels. We developed CHIMERA DDR, a probabilistic machine learning tool that integrates these 51 genomic features using a nested Random Forest architecture to infer seven clinically relevant DDR subgroups. CHIMERA DDR achieved exceptional classification performance (AUCs 0.919-0.999) and accurately detects MMRd and HRRm with or without concurrent DDR deficiencies, resolving admixed phenotypes to enable precision therapeutic stratification beyond binary methods.

INTRODUCTION:

Prostate cancer (PC) is a leading cause of cancer-related mortality in the United States, poised to account for over 313,000 new diagnoses and nearly 36,000 deaths in 2025 (1). While early-stage disease may be cured with definitive treatments, metastatic prostate cancer (mPC), represents an incurable stage characterized by substantial morbidity and poor prognosis. Precision medicine has markedly improved outcomes through breakthrough targeted therapies, in part due to the breakthrough success of Poly (ADP-ribose) polymerase inhibitors (PARPi) and immune checkpoint inhibitors (ICIs), which have become foundational targeted therapies for defined molecular subsets (2–7).

PARPi targeting homologous recombination repair-deficient (HRRd) tumors, particularly those with BRCA1/2 mutations, have demonstrated notable efficacy, and treatment guidelines now recommend somatic molecular testing to stratify mPC patients for PARPi therapy. However, approximately 45% of PC patients with BRCA2 mutations fail to respond to PARPi therapy, despite harboring the canonical predictive biomarker, highlighting the inadequacy of genotype-only binary classification (8–10). Recent studies have demonstrated that the anti-tumor effects of PARPi are amplified when combined with androgen receptor pathway inhibitors (ARPI), as clinical studies show enhanced survival even in selected HRR-proficient tumors (2–6). Simultaneously, ICI offers promise for biomarker-defined patients: pembrolizumab is approved for tumors exhibiting high tumor mutational burden (TMB ≥10

mutations/Megabase or mut/Mb), microsatellite instability represented by microsatellite instability high scores (MSI-H) or mismatch repair deficiency (MMRd), which represent distinct immunogenic subtypes of prostate cancer (7,11–15).

Biomarker-selected patients demonstrate improved but variable therapeutic outcomes. Responses to ICI therapy range from no response to durable disease-free survival, even in TMB and other biomarker-selected populations (16–22). Notably, a large study of genomic alterations in mPC reported that only about 16% of TMB-High cancers exhibited MMR deficiency or high MSI (Microsatellite Instability score), and the rest had other molecular associations (23–25). While TMB thresholds ≥20 mutations/Mb correlate with improved objective response rates (~58% vs. ~20% at lower thresholds) (14,23,26,27). The trends are not uniform across all cancers, indicating a need for context-specific optimization beyond static cut-points (23–26). This suggests a need for context-specific optimization of TMB metrics, rather than static cut-points, to improve clinical applicability. In this context, integrating comprehensive DNA and RNA-based next-generation sequencing (NGS) with PD-L1 immunohistochemistry has improved therapeutic outcomes while enabling enhanced detection of clinically actionable biomarkers (23–31).

Despite these multimodal advances, current clinical decision-making still relies on a binary interpretation of DNA repair activity, assuming either a singular or a homogeneously admixed concurrent repair-deficient phenotype within tumors (e.g., concurrent HRR gene mutation and MMRd with TMB-High state). However, tumors even with dual biomarker positivity frequently exhibit non-responsiveness to one or more corresponding therapies, highlighting DNA repair complexity and current biomarker limitations (32–34). Critically, these binary approaches fail to resolve the intratumor heterogeneity and functional interrelationships between concurrent genomic events, a major gap in the field that limits accurate therapeutic stratification. This emphasizes the inadequacy of simplified paradigms that fail to reflect DDR pathway's complex interrelatedness, dynamic nature, phenotypic variability influenced by clonal heterogeneity and variable genomic aberration penetrance (e.g., BRCA1 C61G hypomorphic mutation), and evolving phenotypic landscapes (35–37).

Consequently, there is a critical need for dynamic, multimodal tissue biomarkers that address tumors' molecular and functional heterogeneity by deconvolving functional deficiency spectra within clonally admixed phenotypes of varying degrees, capturing how concurrent DNA repair gene mutations with variable penetrance shape overall tumor phenotype and guide precision treatment strategies in advanced PCs (Supplementary Fig. 1). To address this challenge, we performed deep curation of DNA repair pathway genes. We analyzed genes from various repair pathways, including DNA break sensors (ATM and CHEK2, among others), DNA mismatch repair pathway genes, Homologous recombination

repair pathway genes, and other DDR regulatory genes such as ATR, AR, and RecQ helicases (38). The genomic features were analyzed together with a comprehensive battery of functional signature scores derived from DNA and RNA sequencing data. This resulting integrated approach refined the molecular classification of DNA repair subgroups within established clinical stratification frameworks for advanced PC.

We validated the molecular architecture of refined DNA repair subgroups using an independent clinical case series of PC patients. Finally, we utilized this comprehensive understanding to develop a probabilistic modeling framework that accurately classifies advanced PC patients based on their functional DNA repair activity profiles (Supplementary Fig. 1). This model decomposes DNA repair status into components, representing divergent admixed phenotypes based on their relative importance in overall phenotype, while estimating both the degree of phenotypic heterogeneity and the dominant phenotype within PC tumor samples. The platform leverages multiple orthogonal functional readouts to infer the relative contributions and interactions of co-existing DNA repair deficient phenotypes (e.g., HRRd, MMRd) as well as DNA repair intact compartments. Such multidimensional approaches may be particularly valuable in developing precision biomarkers for context-aware therapeutic stratification.

RESULTS:

Fine-Scale Molecular Stratification of mPC via TMB Thresholds and DDR Alterations

We first assessed the prevalence of tumors with high tumor mutational burden (TMB-H; ≥10 mutations/Mb) in advanced PCs. Whole-exome sequencing of 672 tumors (228 from UW-TAN, 444 from SU2C) yielded 482 high-quality exomes after rigorous filtering (Supplementary Table 1). Of these, 44 tumors (~9.1%) met the TMB-High criterion (≥10 mutations per Mb), an FDA-approved biomarker for selecting pembrolizumab treatment in solid tumors (7). Notably, 33 of the 44 TMB-High tumors lacked detectable mutations in HRR, RecQ, and DSB sensor pathway genes, although a subset of these tumors harbored MMR pathway deficiencies (n=14, ~42.4%). Such an enriched case set provided a robust foundation for characterizing this molecular subtype.

To delineate the substructure of TMB-High tumors, we subclassified the TMB-H tumors into DDR-proficient and DDR deficient mPC. This analysis encompassed 46 curated DDR-related genes across Homologous recombination repair (HRR), Mismatch repair (MMR), Fanconi anemia, Double strand break sensor (DSB Sens.), RecQ helicase pathways, and a few core DDR regulators (Fig. 1A-B). We examined tumor ploidy, percent genome alteration (PGA), genomic scar scores (LOH, LST, TAI, percent LOH), microsatellite instability (MSI), and mutational signature profiles. These analyses revealed

patterns within TMB-High tumors and further subclassification across DDR pathways (Fig. 2A-F, Supplementary Fig. 2A, Supplementary Table 2A-D, Supplementary Table 3A-B). Importantly, as expected, tumors with relatively higher TMB were prominently enriched for MSI-high score and COSMIC MMRd signature positivity (Fig. 1B). To identify the optimal TMB cut-point maximizing MSI-high score enrichment, we applied binary logistic regression with three-fold cross-validation (Supplementary Fig. 2B-E), identifying 18.61 mut/Mb as the most discriminative threshold. The goal of this analysis is to enrich MMRd tumor samples to better characterize this molecular subgroup using multimodal molecular approaches. Tumors exceeding this cutoff were designated as TMB-very high (TMB-vH), distinct from the redefined TMB-High (TMB-H) class encompassing the 10-18.61 mutations/Mb range (Supplementary Fig. 2E). Based on this specific cutoff, approximately 3.9% of this case series qualified as TMB-vH. Overall, 1.2% of the entire case series were characterized as TMB-H with concurrent HRR pathway loss-of-function mutations (TMB-H HRRm), and 1.03% were TMB-vH with HRR loss of function mutation (TMB-vH HRRm; We defined HRRm class -tumors with detected HR pathway gene mutations. Also refer to Supplementary Document 1). Molecular profiling of this relatively rare subset revealed a molecular characteristic dichotomy that splits the samples into two distinct subgroups: i) MMR-deficient-like tumors (MMRd; characterized by very high TMB, MSI-high status, MMRd signature positivity, and low ploidy/PGA/HRD-Scar scores). ii) HR deficient-like tumors (HRRd; exhibiting lower TMB/MSI but elevated ploidy, PGA, and HRD-Scar scores), independent of MMR pathway gene loss status, likely reflecting distinct tumor evolutionary trajectories (Fig. 2A-F, Supplementary Table 2A-D, Supplementary Table 3A-B, also refer to Supplementary Document 1 for HRRm & HRRd definition).

Building on prior findings (39,40), which refer to bi-allelic TP53 or PTEN alterations to elevated HRD-Scar and LOH scores even in HRR-wild-type tumors, we further refined the classification of HRRm and DDR-intact tumors by TP53 and PTEN mutant statuses to better align biological mechanisms with clinical relevance. Ultimately, these analyses precisely defined 11 molecularly distinct DDR axis subgroups within the mPC case series, forming the basis of a novel stratification approach that informs future therapeutic evaluation and tool development. (Fig. 1A-B).

Preserved Genomic Integrity and Heightened Immunogenicity Define TMB-vH Tumors

Comparison of TMB-vH (≥18.61 mut/Mb) and TMB-H (10 -18.61 mut/Mb) tumors revealed extensive molecular divergence (Fig. 2A-F). TMB-vH tumors were enriched for bi-allelic MMR gene mutations, especially in MSH2 (Fig. 1A; Supplementary Fig. 2F-G). 10 out of 14 tumors exhibited the COSMIC MMRd signature (Fig. 1B). Such enrichment was notably absent in TMB-H tumors. Despite their high mutational burden, TMB-vH tumors exhibited lower ploidy ($p = 0.008$) and reduced PGA ($p = 0.01$) and

lacked highly recurrent arm-level alterations (e.g., 8p loss, 8q gain, 17p loss), which are typical of other mPC subtypes, including HRR gene mutant subsets (Fig. 2C, 2F; Supplementary Fig. 3A-B). These findings indicate that preserved genomic structural integrity is preserved despite hypermutation.

Immune infiltrate frequency analyses (Total n=323, SU2C n=236; UW-TAN n=87) using CIBERSORT (v1.02) revealed that the TMB-vH HRRm subset had the highest mean and median $CD8^+$ T cell infiltration, while the TMB-H HRRm group had overall reduced immune cell density (Supplementary Fig. 4A). TMB-vH tumors exhibited a trend toward increased CD274 (PD-L1) mRNA expression (p=0.006, q=0.06) and LAG3 downregulation (p=0.02, q=0.27), though neither reached statistical significance after multiple testing correction. (Fig. 2G-H, Supplementary Fig. 4B-C, Supplementary Table 4A-B) (n=467); further supporting enhanced checkpoint inhibitor sensitivity. Recent clinical case series-based outcome aligns with these findings: patients with TMB > 20 mut/Mb and MMR deficiency have shown improved responses to ICI (14).

Enrichment of AR Pathogenic Mutations and Disrupted Molecular Dependencies in TMB-vH Tumors

To further evaluate alterations in mPC driver events with respect to their frequency in tumors with the relatively intact genomes found in TMB-vH tumors, we investigated androgen receptor (AR) alterations across mutational burden subsets. The AR is commonly altered either by mutation or locus amplification in mPCs following AR pathway targeting. TMB-vH tumors were significantly enriched for pathogenic AR ligand-binding-domain (LBD) mutations associated with ARPI resistance, yet notably lacked AR focal amplifications - a pattern consistent with the genome structural integrity defining this subgroup (Fig. 1A; Supplementary Fig. 5A-B). Among all molecular classes, TMB-vH exhibited the highest frequency of AR pathogenic mutations driving selective resistance to ARPI inhibitors (Supplementary Fig. 5B). Survival analysis & PSA90 objective response analysis showed no significant differences in outcomes between AR mutant and AR wildtype tumors within the TMB-vH subgroup, indicating that conventional AR mutation based prognostic models may not be applicable to this molecularly distinct subset (Supplementary Fig. 5C, Supplementary Fig. 6A-B).

To determine whether TMB-vH tumors exhibit altered gene expression and growth regulatory dependencies, we evaluated established proliferative and aggressive phenotype markers across all 11 DDR axis subgroups, including cell-cycle progression (CCP) score, neuroendocrine activity (NEa) score, MKI67 and BRCA2 mRNA expressions, and androgen receptor activity (ARa) score (41–43)(Fig. 2H. Across the majority of molecular subgroups, we validated established prognostic associations, wherein AR activity exhibited strong inverse correlations with NE activity, CCP score, MKI67 expression, and BRCA1/2 mRNA expression levels (Supplementary Table 5A-C). These patterns were

consistent with prior evidence linking elevated CCP scores to NE-driven aggressive biology and inverse relationships with AR activity (Supplementary Table 5D-F). (42,43). In contrast, the TMB-vH HRRm & TMB-vH subgroup displayed a complete inversion of these canonical relationships in terms of mRNA expression as well Fig. 2H, Supplementary Table 5A-B). Elevated CCP scores were decoupled from NE activity and instead co-occurred with high AR activity, elevated BRCA2 expression, and increased MKI67, representing fundamental disruption of established prognostic marker interdependencies (Supplementary Table 5E-F). This molecular signature analysis indicates that TMB-vH HRRm tumors operate under a distinct regulatory program that diverges entirely from the NE-associated proliferative phenotype defining aggressive behavior in other mPCs.

Comparative analyses of TMB-H versus TMB-vH tumors identified significantly reduced frequency of 8q gain in TMB-vH tumors, with this chromosomal arm alteration being consistent with their compromised genomic structural integrity (Supplementary Fig. 3A-B, Supplementary Table 6A-B). However, within the TMB-vH HRRm subset, we identified a profound disruption of these canonical molecular relationships, which fundamentally challenges the current understanding of proliferative regulation in hypermutated tumors.

We also identified a previously unrecognized molecular subtype defined by biallelic mutations in MMR pathway genes that paradoxically exhibits low TMB (< 10 mut/Mb, MMRm TMB-Low) (44). This subgroup demonstrated a distinctive genomic characteristic, with a very high frequency of 8p loss and 8q gain involving RAD21, along with MYC amplification. Notably, this subgroup also harbors evidence of focal MYC amplification and frequent AR hyper-amplification (Fig. 3G, Supplementary Fig. 3A-B, Supplementary Fig. 5A). Unlike TMB-vH tumors, the median tumor ploidy is approximately 3 (p=0.03), with a very high average LOH score (15)(Fig. 2C-D, Supplementary Table 6C), greater than even the HRRm subset of the mPC group median LOH (9.5) (Supplementary Table 6D). Surprisingly, despite these genomic alterations, this subgroup does not show significant statistical deviation in HRD-Scar signature scores from HRRm or HRRm-tumor suppressor groups (MMRm TMB-L 45 vs HRRm 51.5 (p=0.18) or HRRm tum. supp. 52 (p=0.31)) (Supplementary Table 6E). The key distinguishing feature that differentiates this subgroup from HRRd tumors is the proportion of genome altered (PGA, MMRm TMB-L 0.31 vs HRRm 0.53 (p=0.27) or HRRm tum. supp. 0.59 (p=0.13)), highlighting the complex interplay between different DNA damage repair pathways and their molecular characteristic manifestations (p-values were non-significant due to high PGA score variance across HRRm samples) (Fig. 2A-F, Supplementary Fig. 2A-D, Supplementary Table 6F).

Intragroup Heterogeneity in Molecular and Functional Signatures Leads to Two Distinct Molecular Classes in Concurrent TMB-High and HRR Mutant Tumors

Within the mPC case series, we identified a rare subset (2.3%) harboring both high tumor mutational burden (TMB >10 mut/Mb) and loss-of-function mutations in homologous recombination repair genes (TMB-H HRRm), a phenotype previously described in isolated case reports and one single study using a clinical genome (32–34,38). This dual-alteration group exhibited two clearly separable molecular profiles that paralleled the broader TMB-H population dichotomy. The TMB-vH HRRm subclass (≥18.61 mut/Mb) displayed features consistent with MMRd, including consistent MSI-high status, enrichment for COSMIC MMRd mutational signature scores, and comparatively lower tumor ploidy, percent genome alteration (PGA), and HRD-Scar scores (Fig. 1B, 2C-E). This MMRd-like subclass was enriched for AR ligand-binding domain mutations, mirroring the TMB-vH profile observed in MMRd tumors, and exhibited high $CD8^+$ T-cell infiltration, suggesting an immunologically active phenotype (Supplementary Fig. 4A, Supplementary Fig. 5A).

Critically, the TMB-vH HRRm subset demonstrated the same inversion of canonical prognostic marker relationships observed in TMB-vH tumors: the AR activity score showed a reversed correlation trend with CCP score (ρ =0.53, p=0.28), complementary with the established negative correlations observed across other molecular subgroups (Fig. 2H, Supplementary Fig. 7A, Supplementary Table 5C). In contrast, the TMB-H HRRm subclass (10–18.61 mut/Mb) resembled a functional HRRd phenotype, characterized by lower MSI-high frequency but markedly elevated HRD-Scar scores, PGA, and tumor ploidy, providing functional evidence of HRRd like molecular signature feature (Fig. 2A-F).

Tumors with genomic alterations reflecting both HRRd and MMRd consequently feature biomarkers predicting high response rates to two distinct targeted therapeutic strategies: ICI (driven by high TMB and MMRd-like features in TMB-vH HRRm) and PARPi (driven by functional HRRd in TMB-H HRRm). However, the pronounced variation in genomic and phenotypic features within this dual-alteration subset suggests therapeutic responsiveness may differ substantially depending on the dominant underlying biology, underscoring the critical need for precise molecular stratification to optimize treatment selection.

Clinical Validation Case Series Recapitulates Molecular Features and Reveals Therapeutic Significance of DDR Subtype Stratification

To assess the clinical relevance of the mPC DDR genomic subtypes identified in initial cohort of metastatic tumors, we assembled an independent validation case series comprising genomic data derived from 130 men with mPC and attendant clinical outcomes. Based on available clinical sequencing interpretations and molecular signatures, we defined eight clinical subgroups within a targeted validation case series of 130 patients' data points (Supplementary Table 7). This case series intentionally overrepresented with TMB-vH HRRm (19%) and TMB-vH (19%) tumors, enabling robust

assessment of these rare molecular subtypes using TMB thresholds derived from our discovery analysis and clinical-grade sequencing reports from established diagnostic providers (Fig. 3A-H). The clinical case series recapitulated key molecular features of TMB-vH tumors: significant enrichment for MSH2 mutations and AR ligand binding domain pathogenic mutations, with notable lack of genome wide copy number gains and amplifications consistent with preserved genomic integrity (Fig. 3G-H, Supplementary Fig. 6C). Importantly, we observed significant negative enrichment of TP53/PTEN loss-of-function events in the TMB-vH subgroup (Supplementary Fig. 6C). Unlike the discovery case series, the clinical TMB-H subgroup contained substantial numbers of patients with bi-allelic loss of MMR pathway genes ($p=9.8e^{-07}$) or MSI-high status($p=7.7e^{-05}$), though at frequencies lower than the TMB-vH subset (Fig. 3B-C).

Assessments of the genomic subclassification with clinical outcomes in this cohort revealed divergent therapeutic outcomes that underscore the clinical significance of TMB-vH molecular stratification. While TMB-vH tumors in the discovery case series trended toward poorer prognosis with limited ICI or PARPi exposure, both TMB-vH and TMB-vH HRRm subgroups demonstrated improved overall survival in the validation case series, likely reflecting higher frequency ICI and PARPi utilization (Supplementary Fig. 6D-E). TMB-vH and TMB-vH HRRm tumors showed significantly superior ICI response rates (Kaplan-Meier analysis and PSA50 response) (Fig. 3D-E, Supplementary Fig. 6F). Notably, TMB-vH HRRm patients exhibited prolonged ICI+PARPi combination therapy (Fig. 3F), whereas no differential response to androgen receptor signaling inhibitors (ARSI) was observed across subgroups (clinical case series p=0.74) (45,46) (Supplementary Fig. 6A-B).

Dynamic Prediction of the Relative Composition of DDR Pathway Defects Using CHIMERA-DDR: A Probabilistic Model for Precision DDR Stratification

The genomic analyses of tumors with DDR alterations indicated that current binary DDR classification approaches fail to capture the full molecular complexity of these tumors, particularly MMR deficiency co-occurring with functional HRd signatures. The misclassification of tumors with respect to a particular repair deficiency may impact clinical outcomes. The results also identified a paradoxical relationship between mutational burden and chromosomal instability. Unfortunately, patient cohorts that include comprehensive DDR profiling are small, further limiting supervised machine-learning effectiveness for capturing the full DDR alteration spectrum.

To address these challenges, we developed CHIMERA-DDR (Comprehensive Heterogeneity Informing Molecular Evaluation and Risk Assessment), a probabilistic model for dynamic DDR deficiency prediction from individual tumor samples (Fig. 4A-B, Supplementary Fig. 1). CHIMERA integrates 51 weighted molecular features encompassing DDR pathway gene mutations, functional signatures, large-

scale genomic aberrations, and tumor intrinsic characteristics like ploidy and cellularity, representing a significant advancement over single-biomarker approaches. At its core, this framework integrates genomic aberrations, tumor intrinsic features, and genomic signatures to determine sample molecular subtypes accurately (refer Supplementary Document 2 for CHIMERA-DDR development strategy).

CHIMERA-DDR employs a two-tier nested Random Forest strategy that addresses the fundamental challenge of singleton cutoff approaches across cancer subtypes by readjusting relative signature weights based on tumor intrinsic features and clonality to generate combined probability scores for DDR subclass assignment. Tier 1 decomposes patients into four primary components: TMB-vH, TMB-H, HRRm, and DDR-intact phenotypes. Tier 2 provides refined subcategorization, stratifying the HRRd subset by tumor suppressor gene status (intact versus mutant TP53/PTEN) while subdividing DDR-intact groups into DDR-intact, DDR-intact with tumor suppressor mutations, DSB sensor mutants, and RecQ helicase pathway loss-of-function mutations (Fig. 4A- B). Notably, 77.59% of all HRRm tumors as per CHIMRA prediction also harbor HRRd molecular signatures and are classified as HRRd major or HRRd minor (Based on mutation clonality, mutation penetrance to phenotype, or mutation zygosity) (Supplementary Table 8A-B).

CHIMERA-DDR demonstrated superior classification performance across all DDR phenotype categories (Fig. 5A-B). The model achieved exceptional performance in leave one out cross validation for TMB-vH identification with an area under the curve (AUC) of 0.999. TMB-H subgroup prediction yielded an AUC of 0.998, while Tier 2 HRRd detection achieved 0.982 AUC (Fig. 5A-B). Fine-scale HRRd subset deconvolution based on TP53-PTEN loss status demonstrated an AUC of 0.919 for the tumor suppressor mutant subset, indicating robust performance for stratifying patients with complex overlapping DDR phenotypes. Importantly, in every case, the model also inferred the probability of the tumor behaving as DDR-intact, a critical concept that was not considered in past models. The model showed limitations in RecQ helicase pathway prediction, achieving an AUC of 0.797, representing an area for future improvement (Fig. 5B).

Of the 482 total tumors, 445 were used for CHIMERA-DDR model training and 37 were withheld for model test purposes (5 TMB-vh HRRm (n=5), TMB-H HRRm (n=6), MMR gene mutant TMB-Low (n=7), DSB Sens. gene mutants (n=19) (Fig. 4B-D, Supplementary Table 9A). Despite the limited sample size, this unbiased approach validated CHIMERA's ability to identify overlapping DDR phenotypes and revealed significant heterogeneity within established DDR categories.

Among the 5 TMB-vH HRRd samples, 2 harbored substantial HRRd signature components, suggesting a deeper MMRd-like phenotype with concurrent MMRd and HRRd features (Fig. 4C barplot from the left 1 and 2), while the remaining 3 (Fig. 4C barplot 3 through 5) exhibited predominantly MMRd

phenotypes with HRRd appearing as a passenger event (Fig. 4C-E, Supplementary Table 9A). Similarly, within the 6 TMB-H HRRd samples, 4 demonstrated predominant HRRd phenotypes with substantial TMB-H features, while 2 samples displayed below threshold HRRd signatures, suggesting passenger mutation but predominant TMB-H characteristics, potentially indicating limited MMRd-like features (Fig. 4D-E). To clarify further, these two samples' TMB-H characteristics are not suggestive of an MMRd-like feature. The DSB sensor subset (n=19), consisting primarily of tumors with ATM/CHEK2 mutations, showed one tumor, which is borderline HRRd negative as per the defined confidence threshold. The majority exhibited DDR-intact behavior or low-penetrance HRRd-like phenotypes, suggesting that ATM/CHEK2 mutations alone do not reliably predict HRRd status (Fig. 4D).

Testing of the 7 MMRm-low TMB subset revealed 1 tumor with HRRd-like phenotype scores despite no clear mutational explanation (Supplementary Document 1, refer to MMRm definition). This sample exhibited high LOH scores, borderline but sub-threshold HRD-Scar scores, TP53 bi-allelic loss-of-function mutations, BRCA2 monoallelic copy loss, and a focal copy-neutral LOH event in BRCA2, together indicating a complex genomic state. These features were accurately considered within CHIMERA-DDR and inferred an accurate class identity, underscoring the reliability of the model's prediction (Fig. 4C-D). A second sample within this subset demonstrated borderline MMRd-like features with significant TMB-high characteristics despite a detected TMB of 9.8 mutations/Mb. This tumor harbored MSH2 bi-allelic loss-of-function and MLH1 loss-of-function mutations and tested positive for COSMIC MMRd signatures, potentially indicating functional MMRd despite being below the 10 mutations/Mb mutational burden threshold (47). These findings underscore the clinical utility of probabilistic modeling approaches for capturing the full spectrum of DDR pathway alterations and their complex interactions in prostate cancer genome.

Semi-Independent Test Set Evaluation of CHIMERA-DDR Class Determination Performance Captures Intra- and Inter-Tumor DDR Phenotype Heterogeneity

To evaluate the performance of CHIMERA-DDR, we analyzed whole exome sequences (WES) from 270 mPC tumors derived from 112 patients enrolled into a rapid postmortem tissue collection protocol (UW-TAN) (Supplementary Fig. 7A). Consistent with Kumar et al. 2015, we found limited intratumor heterogeneity in heavily treated multiple metastatic tissues from the mCRPC rapid autopsy series (Supplementary Table 10) (48). Substantial heterogeneity in TMB classification was detected in 4 of 112 patients, where tumors from the same patient included TMB-vH, TMB-H, and TMB-Low categories. Similarly, class switching in MSI scores was observed in tumors from 2 patients (MSI-high vs MSI-low). In contrast, a greater degree of intratumor heterogeneity was seen in LOH and HRD-Scar scores. To determine whether DDR axis subtypes were associated with organ-specific metastatic tropism, we

analyzed site distributions; however, no statistically significant enrichment was observed, indicating that different mechanisms contributing to DDR subtypes do not preferentially drive site-specific metastasis in this dataset (Supplementary Fig. 7B), though limited sample sizes restricted robust inference. (14).

To assess whether CHIMERA-DDR predictive probabilities reflect DNA from admixed normal cells in the tumor sample, we analyzed cellularity fractions in 445 training samples. For DDR-intact tumors, CHIMERA's Tier 2 DDR-intact probability consistently exceeded the inferred normal fraction from Sequenza cellularity (normal fraction = 1 - cellularity) (Fig. 6A). These initial results suggested that CHIMERA's elevated DDR-intact predictions arose from a combination of admixed normal DNA and genuine DDR-intact tumor subclones, indicating that the model was capturing both normal cell contamination and authentic tumor-derived DDR-intact behavior within the sample (Supplementary Fig. 8A-B). However, DDR-deficient tumors, particularly TMB-vH and TMB-H cases, revealed DDR-intact prediction probabilities substantially lower than the corresponding normal DNA content (Supplementary Fig. 8B, Supplementary Fig. 9). If CHIMERA were simply reflecting normal contamination, DDR-intact probabilities would remain consistent across tumor types regardless of their DDR status. Instead, DDR component probabilities lower than tumor DNA content in DDR-deficient tumors strongly suggest that CHIMERA accurately captures the overall DDR phenotype which is not solely driven by tumor cellularity. Of importance, we used cellularity as one of the 51 input features in training CHIMERA (Feature rank # 41).

Among the 58 HRR-mutant tumors in the training set, CHIMERA predicted that 77.59% (n=45) exhibited HRRd-like molecular characteristics, while the remaining 22.41% showed HRRd-like integrative genomic features below the determined threshold (Supplementary Table 8A & Supplementary Table 8B). Detailed curation identified contributing factors, including subclonal loss-of-function events, hypomorphic alleles, and lower-penetrance HRR aberrations that diminished the functional impact on DDR phenotype. Integrative analysis further revealed tumors designated as HRRd that unexpectedly displayed DDR-intact mutational signatures. Together, these findings reconcile the apparent paradox and highlight CHIMERA-DDR's ability to capture the full functional spectrum of DDR pathway dysfunction, moving beyond rigid binary models toward probabilistic resolution of overlapping and heterogeneous phenotypes.

To evaluate CHIMERA's robustness for DDR status inference, we selected 24 patients (52 tissue genomes) from the UW-TAN case series covering 5 DDR subgroups with corresponding whole genome sequencing (WGS) data. (Fig. 6B, Supplementary Table 9B). We noted some discordance in terms of TMB between WES & WGS, potentially suggesting heterogeneous TMB in coding vs noncoding regions in some tumors (Supplementary Table 10). This unique paired dataset enabled evaluation of

CHIMERA's prediction accuracy and ability to capture inter-tumoral heterogeneity while inferring admixed DDR phenotype probabilities. We performed 2-tier DDR subtype prediction and plotted tier 1 probabilities in percent component bar plots to understand intratumor versus inter-tumor heterogeneity, with each bar plot component representing relative prediction probability (Fig. 4E, 6C, Supplementary Fig. 10; also refer to Supplementary Document 3). We queried whether multiple tissues from the same patients clustered tightly. Interestingly, WGS-derived input data revealed substantial intragroup homogeneity in deconvolved DDR deficiency predicted probabilities with notable exceptions; 3 of 24 tumors showed substantial intra-tumor heterogeneity, all associated with variation in HRRd classifications**.**

DISCUSSION:

This study fundamentally reframes DNA damage repair classification in advanced prostate cancer by demonstrating that conventional biomarker approaches inadequately capture therapeutically relevant biology. We observed that tumors with concurrent DDR pathway alterations exhibit variable therapeutic responses to immune checkpoint inhibitors versus PARP inhibitors despite carrying conventional biomarkers predictive for both. Through comprehensive analysis of 672 advanced PC exomes, we present three transformative findings that challenge current precision medicine paradigms in DDR-targeted therapy.

<u>TMB-Very High as a Distinct Molecular Entity:</u> We established that TMB-very high tumors (≥18.61 mut/Mb) represent a molecularly distinct entity from conventional TMB-high tumors, with profound clinical implications for immunotherapy selection. This data-driven threshold emerged from systematic analysis demonstrating that higher TMB cutoffs achieve superior accuracy in identifying MMRd phenotypes, revealing a fundamental specificity-sensitivity trade-off critical for therapeutic decision-making (49–55). Importantly, MMR pathway gene bi-allelic mutants with paradoxically low TMB (<10 mut/Mb) demonstrate that mutation-based detection strategies are insufficient-these tumors fail to exhibit functional MMRd characteristics despite definitive genetic alterations (Bi-allelic loss of MSH2, MSH3).

TMB-very high tumors exhibit MMR-driven biology with preserved genomic integrity, robust CD8+ T-cell infiltration, and PD-L1 upregulation, while paradoxically maintaining AR-driven transcriptional programs. This "inflamed but AR-active" phenotype contradicts prevailing models of AR independence in hypermutated PCs and suggests these tumors may require different therapeutic approaches than conventional TMB-high cases. While conventional TMB ≥10 mut/Mb cutoffs show limited predictive value in PC, our elevated threshold demonstrates improved correlation with immunotherapy response, consistent with observations that patients with TMB above median levels show enhanced checkpoint

inhibitor responses in metastatic castration-resistant PC (23,24).We statistically inferred the TMB-vH subgroup to enrich for definitive MMRd-like tumors, then performed deep molecular characterization to develop our dynamic classification model. This genotype-molecular admixture complexity reflects biological reality and underscores why probabilistic approaches like CHIMERA are essential for precision DDR stratification.

Functional DDR Pathway Dominance in Molecularly Complex Tumors: We discovered that tumors harboring both HRRd and MMRd alterations do not exhibit dual therapeutic vulnerabilities as predicted by conventional biomarkers. Instead, these molecularly complex tumors display functional dominance of one DDR pathway over another, explaining why patients respond to either PARPi or immunotherapy, but not necessarily both. CHIMERA-DDR reveals this biological reality by quantifying relative DDR pathway contributions within individual tumors rather than imposing binary classifications, capturing tumor heterogeneity and functional interrelationship (epistatic) between HRR and MMR pathways. Critically, 24% of HRR-mutant tumors displayed MMRd-predominant phenotypes rather than expected HRRd biology. For example, tumor 00-010 carried both TMB-vH and HRR mutations, yet CHIMERA correctly identified its dominant MMRd phenotype, a finding confirmed by clonality analysis showing major MSH2 loss with only minor BRCA2 alterations (Fig. 4E, Supplementary Fig. 10).

Superior DDRd Classification Beyond Genotype-Only Approaches: CHIMERA-DDR represents a paradigm shift from genotype-only HRRd classification to functional phenotype prediction, addressing the critical gap between genomic alterations and therapeutic vulnerability. Our analysis demonstrates that 77.59% of HRRm tumors exhibited HRRd-like molecular characteristics while 22.41% did not, revealing substantial genotype-phenotype discordance missed by conventional approaches. CHIMERA-DDR identifies TMB-vH and TMB-vH HRRm as clinically distinct molecular subgroups in prostate cancer with unique therapeutic vulnerabilities. CHIMERA's ability to predict DDR-intact-like phenotypes alongside DDR-deficient signatures adds crucial clinical insight, particularly for detecting DDR-intact clones that may require combination targeted approaches.

CHIMERA's biological validation confirmed molecular characteristic inference through multiple lines of evidence. Tumor cellularity analysis revealed that DDR-deficient tumors had substantially lower DDR-intact prediction probabilities compared to normal DNA content (1-cellularity), indicating that CHIMERA probabilities represent weighted molecular characteristics of the tumor composition at the bulk tumor level (Fig. 6A, Supplementary Fig. 8A-B). Semi-independent validation using paired whole genome sequencing data from 52 samples across 24 patients demonstrated substantial intragroup homogeneity in predicted probabilities across metastatic sites, confirming stable, biologically meaningful DDR phenotypes (Fig. 6C).

Clinical Implications for Combination Therapy Strategies: CHIMERA-DDR's findings may offer insights into why PARPi+ICI combination strategies have shown limited success in prostate cancer. Phase 3 KEYLYNK-010 trial (pembrolizumab + olaparib) was stopped for futility (11,12,47,53–56), while phase 2 studies including CheckMate 9KD (nivolumab + rucaparib) and JAVELIN PARP Medley (avelumab + talazoparib) demonstrated only modest activity, primarily in biomarker-selected subgroups. These trials enrolled molecularly heterogeneous populations where genomically similar tumors may have exhibited fundamentally different functional molecular types, potentially diluting treatment effects. Our discovery that concurrent DDR alterations display pathway dominance rather than dual vulnerabilities suggests even tumors with multiple alterations may respond optimally to pathway-specific monotherapy rather than combination approaches.

CHIMERA represents a fundamental shift from binary classification toward dynamic probabilistic inference that captures the biological reality of admixed DDR phenotypes in cancer. By integrating multiple molecular features, CHIMERA better approximates functional DDR status beyond simple mutation presence, enabling accurately stratifying patients across the full spectrum of DDR alterations. This is particularly valuable for tumors with high TMB and concurrent HRR aberrations, where clinicians currently face difficult choices between immunotherapy and PARPi treatments due to conflicting molecular signals. CHIMERA resolves this uncertainty by separating mixed DDR signatures and identifying mutation-positive tumors that actually behave like DDR-intact cancers, providing accurate classification for both HRR and MMR pathways—capabilities impossible with current single-cutoff methods. Such integrated molecular profiling may better identify patients who would benefit from ICI monotherapy versus combination approaches, directly addressing the current clinical challenge of therapeutic selection in molecularly complex DDR-altered tumors.

CHIMERA-DDR's utility will expand alongside emerging therapies. New treatments, including CTLA-4 and PD-L1 combinations, PD-L1-targeted CAR-T cells, and next-generation DNA-damaging agents, will require similar sophisticated patient selection. This framework extends beyond prostate cancer to any malignancy where DNA repair deficiencies guide treatment decisions. However, prospective clinical trials with treatment outcomes are essential to validate whether CHIMERA-predicted phenotypes accurately identify responders to immunotherapies or DNA-damaging therapies, such as PARP inhibitors.

METHODS:

Study Cohorts. We assembled a comprehensive multi-modal genomic dataset from advanced PC patients across multiple independent cohorts. Fundamentally, our framework was designed to integrate genomic variants with genomic signatures for precise molecular subtype classification of individual samples. The primary WES discovery cohort (n=672) comprised samples from two independent sources: the University of Washington rapid autopsy program (UW-TAN, n=270, Supplementary Table 1) and Stand Up To Cancer Dream Team consortium (SU2C, n=444). After quality control filtering to remove low-coverage sequence data, 466 high-quality samples were retained for downstream analysis. RNA sequencing data (n=656) were obtained from overlapping patients across both cohorts: SU2C samples (n=482) sequenced in two batches (n=270 and n=212) and UW-TAN samples (n=174) sequenced as a single batch (Supplementary Table 1). For clinical validation, we assembled an independent cohort of 130 patients with clinical-grade genomic sequencing results from four academic medical centers (Supplementary Table 7). To assess intra-tumoral heterogeneity, we analyzed a test cohort (n=52) comprising multiple metastatic tumor sites with paired whole-genome sequencing data from 24 patients, generating 51 CHIMERA-DDR input features to evaluate model performance across diverse sampling contexts (Supplementary Table 9A-B,10).

DNA Sequence Analysis. WES and WGS reads were aligned to the human reference genome (GRCh38/hg38) using Burrows-Wheeler Aligner-MEM v0.7.17.1 (57). GATK Best Practices workflows were implemented for preprocessing, including alignment quality assessment, read sorting with Picard MarkDuplicates (v2.24.1), and Base Quality Score Recalibration (BQSR v4.1.9.0) to ensure high-quality variant calling (58).

Variant Calling, Annotation & Tumor Mutation Burden estimation.

Germline variants were called using GATK HaplotypeCaller with population database filtering (gnomAD v3.1.1, 1000 Genomes, ESP6500) to exclude common variants (allele frequency >0.1%) (59). Somatic variants were identified using MuTect2 and Strelka v2 in tumor-normal paired mode, requiring minimum coverage of 12 high-quality reads (Q≥40) with at least 6 supporting alternate alleles (60,61). A union set of filtered mutations from both callers maximized detection sensitivity. BEDTools filtration retained variants within target capture regions for accurate TMB inference. All variants underwent manual curation, and TMB was calculated using pyTMB (v1.4.0), incorporating the union variant set, effective genome size, and cellularity (62).

Microsatellite Instability Analysis. Microsatellite instability (MSI) status was determined using MSIsensor-pro v1.2.0 from paired tumor-normal sequencing data, with the same analytical pipeline applied to both whole-exome and whole-genome sequencing datasets (63). MSI-high (MSI-H) tumors were defined using an MSI score threshold ≥3.5, while microsatellite-stable (MSS) tumors had MSI

scores <3.5, following established clinical guidelines for PC. This classification enabled stratification of tumors based on MMRd status for downstream DDR pathway analysis.

Pathogenicity Classification. Somatic pathogenic mutations in 46 DNA repair pathway genes and associated tumor suppressors were identified through comprehensive annotation using ANNOVAR and validated with Mutalyzer 2.0. The curated gene panel encompassed homologous recombination repair (BRCA1, BRCA2, PALB2, RAD51B, RAD51C, RAD51D), DSB sensors (MRE11A, NBN, ATM, CHEK1, CHEK2), mismatch repair (MLH1, MSH2, MSH6, PMS2), Fanconi anemia pathway (FANCA, FANCB, FANCC, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCL, FANCM), RecQ helicases (BLM, WRN, RECQL4, RECQL5, RECQL), and DNA repair-associated genes (TP53, PTEN, RB1, CDK12, RNASEH2B, ATR, AR, SPOP)(36,37,61,64–74). Germline pathogenic variants were annotated following ACMG recommendations (75), while somatic pathogenic mutations were classified based on: (1) nonsense mutations, (2) frameshift insertions/deletions, (3) canonical splice site alterations (±1,2 bp), (4) missense mutations with established functional impact, and (5) previously reported pathogenic variants in clinical databases. Copy number alterations were assessed using Sequenza CNV pipeline. Gene-specific copy number aberrations were performed using a homebrew copy number gene annotation pipeline using pyTHiA-CNA (Supplementary Document 5). Bi-allelic inactivation is defined as two concurrent hits involving pathogenic mutations, complex structural aberrations, and/or copy loss. Classifications were cross-validated against cBioPortal, ExAC, Kaviar, ClinGen, ClinVar, ClinVitae, CIViC, OncoKB, and UniProt databases (69,74,76–79). Identical analytical pipelines and quality control thresholds were applied to whole-genome sequencing analysis.

Mutational Signature Analysis. Cosmic mutational signatures were identified using SigProfilerExtractor (v1.1.23) with the COSMIC v3.4 reference signature set (80) . Analysis employed non-negative matrix factorization (NMF) with 100 replicates for statistical robustness, resampling enabled for stability assessment, and 10,000-1,000,000 iterations with convergence tolerance of 1e-15 using GRCh38 reference and opportunity genomes with default trinucleotide context. Signatures of clinical interest included HRRd (C. signature 3), MMRd (signatures 6, 14, 15, 20, 21, 26, 44), and DNA polymerase epsilon deficiency (POLE; signatures 9, 10a-d, 14, 20).

Copy Number Analysis, LOH & HRD-Scar: Copy number analysis utilized paired tumor and normal depth of coverage and B-allele frequency data. Copy number variations, tumor cellularity, and purity were determined using Sequenza v2.0.0 (78). Tumor ploidy and purity were derived alongside segmented allele-specific copy numbers from allele-specific copy number analysis. Copy number annotation was performed using a homebrew annotator tool referencing GENCODE v45 annotation. Loss of heterozygosity (LOH) scores were calculated from copy number calls using the method

described by Abkevich et al. (81). We counted toward this score events that were greater than 15 Mb in length and defined by a non-zero copy number count with an inferred minor allele count of 0 (81). Chromosomes that were affected by such LOH events over ≥75% of their entire length were excluded from both numerator and denominator, as they typically arise through non-HRRd associated mechanisms. HRRd scores were computed as the sum of LOH, telomeric allelic imbalance (TAI), and large-scale state transitions (LST) using the HRD-Scar tool by Sztupinszky et al (82). Bi-allelic inactivation was defined as concurrent pathogenic mutation and copy number loss (log2 ratio <-0.5 and B-allele frequency indicating loss of heterozygosity) as well as complex structural aberrations.

Subgroup-Specific Consensus Genomic Aberration Analysis. To generate subgroup-specific consensus profiles of genomic aberrations, we established a CNVkit-based analytical framework for comparative copy number analysis across molecular subgroups (83). Discovery cohort samples were stratified into DDR-defined molecular subgroups, and individual BAM files within each subgroup were pooled to create representative datasets. Pooled BAM files were downsampled to 300× depth of coverage using Picard tools to normalize sequencing depth and reduce computational burden while maintaining copy number detection sensitivity (84). CNVkit (v0.9.12) was then applied to the downsampled pooled BAM files to generate consensus copy number profiles, enabling identification of subgroup-specific recurrent genomic alterations and comparative analysis of copy number landscapes across different DDR molecular subtypes (83).

Transcriptome Analysis: RNA sequencing reads were aligned to the hg38 reference genome using STAR v2.7.3a, with gene-level expression quantified using GenomicAlignment Bioconductor package (85). Differential gene expression analysis was performed using edgeR and limma packages with filterByExpr filtering and Benjamin-Hochberg FDR correction (86,87). The RNA-Seq dataset (n=656) comprised SU2C (n=482) and UW-TAN (n=174) cohorts. Only poly(A) capture samples were analyzed to avoid batch effects from different library preparation methods, confirmed by multidimensional scaling analysis (41,43,88).

The 31-gene cell cycle progression (CCP), 10-gene androgen receptor activity(ARa), and 10-gene neuroendocrine activity (NEa) scores were computed using GSVA on $\log_2$ FPKM RNA-seq data with validated gene panels (41,43). All scores were standardized across the cohort for comparative analysis.

Immune Infiltration Analysis: Immune cell infiltration analysis was performed independently on each cohort using CIBERSORT v1.04 with the LM22 signature matrix to estimate the relative abundance of 22 immune cell subsets from RNA-seq data (88). Analysis was conducted with absolute mode enabled and statistical significance evaluated using 100 permutations per cohort. Following individual cohort analysis, mean and median infiltrating immune cell frequencies were calculated for each DDR

molecular subgroup to enable comparative assessment of immune microenvironment profiles across different DNA repair deficiency states.

Clonality and Tumor Evolution Analysis: Tumor clonal evolution and phylogenetic reconstruction were performed using PhyloWGS v1.0-rc2 to infer evolutionary histories from multi-sample tumor data (89). Input preparation utilized the PhyloWGS parser to integrate high-quality VCF variant calls, segmented copy number alterations from Sequenza, and tumor cellularity estimates. Simple somatic mutations in DNA repair pathway genes were prioritized during phylogenetic reconstruction to enhance evolutionary accuracy and focus on DDR-relevant clonal dynamics. PhyloWGS analysis employed Markov Chain Monte Carlo (MCMC) sampling with four independent parallel chains to ensure convergence and robust posterior distribution sampling. Chain convergence was assessed using standard MCMC diagnostics, and posterior probabilities were calculated for phylogenetic tree topologies and clonal frequency estimates. The resulting phylogenetic trees and clonal evolution patterns were visualized using Clonevol v0.1, which generated comprehensive plots displaying subclonal architecture, clonal frequency dynamics across multiple tumor sites, and evolutionary trajectories (90). This analysis enabled assessment of clonal heterogeneity patterns across different DDR molecular subtypes and identification of founder versus acquired DDR alterations during tumor progression.

AUTHORS' CONTRIBUTIONS: K. Sharma: Data curation, formal analysis, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing. D.A. Wilson: data curation, software, formal analysis, visualization, methodology, writing–original draft, writing–review and editing. Y. Wang: software, formal analysis, visualization, methodology, writing–original draft, writing–review and editing. S. Haidar: formal analysis, methodology, investigation, data curation. V. Bhatlapenumarthi: formal analysis, methodology, investigation, data curation. D. Boiarsky: formal analysis, investigation, data curation. J. M. Cipriaso: formal analysis, investigation, Data curation. L. Yermakov: formal analysis, investigation, data curation. R. Dutta: formal analysis, methodology. M. Vashisth: review and editing, I. M. Coleman: formal analysis. A. Bankhead: formal analysis. Shantanu K. Jadhav: data curation, resources, investigation. B. Neupane: formal analysis, resources, investigation, data curation. B. W. Taylor: resources, writing–review and editing. S. Rao: review and editing, M. T. Nevalainen: review and editing, A. A. Nelson: writing–review and editing. C. Morrissey: review and editing, M. T. Schweitzer: writing–review and editing. R. B. Montgomery: writing–review and editing. E. S. Antonarakis: Writing–review and editing. P.G.P: Writing–review and editing. J. E. Berchuk: resources. K. Zarrabi, resources, writing–review, and editing. C. C. Pritchard: resources, writing–review and editing, G. Ha: resources, writing–review and editing, A. Kothari: resources, writing–review and editing, B. Jeorge: review and editing, H. Chen: review and editing, R. Kurzrock: writing–review and editing, P. S. Nelson: resources, writing–review and editing, P. L. Auer: formal analysis, resources, writing–review and editing, A. Banerjee: formal analysis, resources, writing–review and editing. D. Kilari Conceptualization, resources, data curation, supervision, writing–critical review and editing. N. D. Sarkar: Conceptualization, resources, data curation, software, formal analysis, supervision, funding acquisition, validation, investigation, visualization, methodology, writing–original draft, project administration, writing–review and editing.

REFERENCES

1. Cancer Stat Facts: Prostate Cancer [Internet]. Available from: https://seer.cancer.gov/statfacts/html/prost.html
2. de Bono J, Mateo J, Fizazi K, Saad F, Shore N, Sandhu S, et al. Olaparib for Metastatic Castration-Resistant Prostate Cancer. New England Journal of Medicine. 2020;382:2091–102.
3. Agarwal N, Azad AA, Carles J, Fay AP, Matsubara N, Heinrich D, et al. Talazoparib plus enzalutamide in men with first-line metastatic castration-resistant prostate cancer (TALAPRO-2): a randomised, placebo-controlled, phase 3 trial. The Lancet. 2023;402.
4. Abida W, Patnaik A, Campbell D, Shapiro J, Bryce AH, McDermott R, et al. Rucaparib in Men with Metastatic Castration-Resistant Prostate Cancer Harboring a BRCA1 or BRCA2 Gene Alteration. Journal of Clinical Oncology. 2020. page 3763–72.
5. Mateo J, Lord CJ, Serra V, Tutt A, Balmaña J, Castroviejo-Bermejo M, et al. A decade of clinical development of PARP inhibitors in perspective. Annals of Oncology. 2019;30.
6. Marcus L, Fashoyin-Aje LA, Donoghue M, Yuan M, Rodriguez L, Gallagher PS, et al. FDA approval summary: Pembrolizumab for the treatment of tumor mutational burden-high solid tumors. Clinical Cancer Research. 2021;27:4685–9.
7. Marabelle A, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. The Lancet Oncology. 2020;21.
8. Teyssonneau D, Margot H, Cabart M, Anonnay M, Sargos P, Vuong NS, et al. Prostate cancer and PARP inhibitors: progress and challenges. Journal of Hematology and Oncology. 2021.
9. Abida W, Campbell D, Patnaik A, Bryce AH, Shapiro J, Bambury RM, et al. Rucaparib for the Treatment of Metastatic Castration-resistant Prostate Cancer Associated with a DNA Damage Repair Gene Alteration: Final Results from the Phase 2 TRITON2 Study. European Urology. 2023;84.
10. Taza F, Holler AE, Fu W, Wang H, Adra N, Albany C, et al. Differential Activity of PARP Inhibitors in BRCA1 - Versus BRCA2 -Altered Metastatic Castration-Resistant Prostate Cancer . JCO Precision Oncology. 2021;
11. Abida W, Cheng ML, Armenia J, Middha S, Autio KA, Vargas HA, et al. Analysis of the Prevalence of Microsatellite Instability in Prostate Cancer and Response to Immune Checkpoint Blockade. JAMA Oncology. 2019;5:471–8.
12. Marcus L, Lemery SJ, Keegan P, Pazdur R. FDA approval summary: Pembrolizumab for the treatment of microsatellite instability-high solid tumors. Clinical Cancer Research. 2019;25.
13. Melo CM, Vidotto T, Chaves LP, Lautert-Dutra W, Dos Reis RB, Squire JA. The role of somatic mutations on the immune response of the tumor microenvironment in prostate cancer. International Journal of Molecular Sciences. 2021.
14. van Wilpe S, Taha T, Rothmann EC, Altshuler E, Park J, Ledet EM, et al. Efficacy of Anti-PD-(L)1 Immunotherapy in Patients with DNA Mismatch Repair–deficient Metastatic Castration-resistant Prostate Cancer. European Urology Oncology [Internet]. 2025; Available from: https://linkinghub.elsevier.com/retrieve/pii/S2588931125001099
15. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, et al. Tumor mutational

burden as an independent predictor of response to immunotherapy in diverse cancers. Molecular Cancer Therapeutics. 2017;16.

16. Xie N, Shen G, Gao W, Huang Z, Huang C, Fu L. Neoantigens: promising targets for cancer therapy. Signal Transduction and Targeted Therapy. 2023.

17. Wood MA, Weeder BR, David JK, Nellore A, Thompson RF. Burden of tumor mutations, neoepitopes, and other variants are weak predictors of cancer immunotherapy response and overall survival. Genome Medicine. 2020;12.

18. Gupta RG, Li F, Roszik J, Lizée G. Exploiting tumor neoantigens to target cancer evolution: Current challenges and promising therapeutic approaches. Cancer Discovery. 2021.

19. Germano G, Amirouchene-Angelozzi N, Rospo G, Bardelli A. The clinical impact of the genomic landscape of mismatch repair–deficient cancers. Cancer Discovery. 2018.

20. Yang Y, Attwood K, Bshara W, Mohler JL, Guru K, Xu B, et al. High intratumoral CD8+ T-cell infiltration is associated with improved survival in prostate cancer patients undergoing radical prostatectomy. Prostate. 2021;81.

21. Dang B, Liang L, Li Z, Luo J, Zhong S. Bolstering CD8+ T Cells' Antitumor Immunity: A Promising Strategy to Improve the Response to Advanced Prostate Cancer Treatment. Biology. 2025;14.

22. Bibi R, Sarkar K. New approaches and prospects of immunotherapy and gene therapy for prostate cancer. Journal of Translational Genetics and Genomics. 2024;8:119–61.

23. Voutsadakis IA. Microsatellite instability (MSI) and the tumor mutation burden (TMB) as biomarkers of response to immune checkpoint inhibitors in prostate cancer. Translational Cancer Research. 2025;14:2553–7.

24. Jardim DL, Goodman A, de Melo Gagliato D, Kurzrock R. The Challenges of Tumor Mutational Burden as an Immunotherapy Biomarker. Cancer Cell. 2021.

25. Friedman CF, Hainsworth JD, Kurzrock R, Spigel DR, Burris HA, Sweeney CJ, et al. Atezolizumab Treatment of Tumors with High Tumor Mutational Burden from MyPathway, a Multicenter, Open-Label, Phase IIa Multiple Basket Study. Cancer Discovery. 2022;12.

26. van Wilpe S, Kloots ISH, Slootbeek PHJ, den Brok M, Westdorp H, Franken MD, et al. Ipilimumab with nivolumab in molecularly selected patients with castration-resistant prostate cancer: primary analysis of the phase II INSPIRE trial. Annals of Oncology. Elsevier Ltd; 2024;35:1126–37.

27. Goodman AM, Sokol ES, Frampton GM, Lippman SM, Kurzrock R. Microsatellite-stable tumors with high mutational burden benefit from immunotherapy. Cancer Immunology Research. 2019;7.

28. Barnett RM, Jang A, Lanka S, Fu Pi, Bucheit LA, Babiker H, et al. Blood-based tumor mutational burden impacts clinical outcomes of immune checkpoint inhibitor treated breast and prostate cancers. Communications Medicine [Internet]. Springer US; 2024;4:1–7. Available from: http://dx.doi.org/10.1038/s43856-024-00687-5

29. Green MF, Wallen ZD, Ko HC, Strickland KC, Dillard A, Conroy JM, et al. Utilization of a Multi-modal Comprehensive Genomic and Immune Profiling Testing Strategy Results in a High Rate of Test Success and Detection of Clinically Relevant Biomarkers While Optimizing Tissue Usage. Molecular Diagnosis and Therapy [Internet]. Springer International Publishing; 2025; Available from: https://doi.org/10.1007/s40291-025-00793-7

30. Boichard A, Pham T V., Yeerna H, Goodman A, Tamayo P, Lippman S, et al. APOBEC-related

mutagenesis and neo-peptide hydrophobicity: implications for response to immunotherapy. OncoImmunology. 2019;8.

31. Pham T V., Boichard A, Goodman A, Riviere P, Yeerna H, Tamayo P, et al. Role of ultraviolet mutational signature versus tumor mutation burden in predicting response to immunotherapy. Molecular Oncology. 2020;14.

32. Bhatlapenumarthi V, De Sarkar N, Patwari A, Antonarakis ES, Nelson AA, Kilari D. Homologous recombination repair (HRR) alterations may be passenger events in patients with advanced prostate cancer (PC) with high tumor mutational burden (hTMB). Journal of Clinical Oncology. 2024;42.

33. Lenis AT, Ravichandran V, Brown S, Alam SM, Katims A, Truong H, et al. Microsatellite Instability, Tumor Mutational Burden, and Response to Immune Checkpoint Blockade in Patients with Prostate Cancer. Clinical Cancer Research. 2024;30:3894–903.

34. Sokol ES, Jin DX, Fine A, Trabucco SE, Maund S, Frampton G, et al. PARP Inhibitor Insensitivity to BRCA1/2 Monoallelic Mutations in Microsatellite Instability-High Cancers . JCO Precision Oncology. 2022;

35. Krais JJ, Johnson N. BRCA1 mutations in cancer: Coordinating deficiencies in homologous recombination with tumorigenesis. Cancer Research. 2021.

36. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015;161:1215–28.

37. Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I, et al. Genomic correlates of clinical outcome in advanced prostate cancer. Proceedings of the National Academy of Sciences of the United States of America. 2019;166.

38. Rhinehart DP, Lai J, Sanin DE, Vakkala V, Mendes A, Bailey C, et al. Intratumoral heterogeneity drives acquired therapy resistance in a patient with metastatic prostate cancer. npj Precision Oncology [Internet]. Springer US; 2024;8:1–5. Available from: http://dx.doi.org/10.1038/s41698-024-00773-w

39. Lotan TL, Kaur HB, Salles DC, Murali S, Schaeffer EM, Lanchbury JS, et al. Homologous recombination deficiency (HRD) score in germline BRCA2- versus ATM-altered prostate cancer. Modern Pathology. 2021;34.

40. De Sarkar N, Dasgupta S, Chatterjee P, Coleman I, Ha G, Ang LS, et al. Genomic attributes of homology-directed DNA repair deficiency in metastatic prostate cancer. JCI Insight. 2021;6.

41. Cuzick J, Swanson GP, Fisher G, Brothman AR, Berney DM, Reid JE, et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: A retrospective study. The Lancet Oncology. 2011;12.

42. Labrecque MP, Coleman IM, Brown LG, True LD, Kollath L, Lakely B, et al. Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. Journal of Clinical Investigation. 2019;129.

43. Bluemn EG, Coleman IM, Lucas JM, Coleman RT, Hernandez-Lopez S, Tharakan R, et al. Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. Cancer Cell. 2017;32.

44. Holeckova KH, Hives M, Grendar M, Drobkova HB, Kliment J. Genetic Variations in TP53, RB1, and PTEN in a Selected Sample of Slovak Patients With Metastatic Castration-resistant Prostate Cancer. In Vivo. 2024;38:2610–6.

45. Karzai F, Vanderweele D, Madan RA, Owens H, Cordes LM, Hankin A, et al. Activity of

durvalumab plus olaparib in metastatic castration-resistant prostate cancer in men with and without DNA damage repair mutations. Journal for ImmunoTherapy of Cancer. 2018.

46. Sharma P, Pachynski RK, Narayan V, Fléchon A, Gravis G, Galsky MD, et al. Nivolumab Plus Ipilimumab for Metastatic Castration-Resistant Prostate Cancer: Preliminary Analysis of Patients in the CheckMate 650 Trial. Cancer Cell. 2020;38.

47. Subbiah V, Solit DB, Chan TA, Kurzrock R. The FDA approval of pembrolizumab for adult and pediatric patients with tumor mutational burden (TMB) ≥10: a decision centered on empowering patients and their physicians. Annals of Oncology. 2020;31:1115–8.

48. Kumar A, Coleman I, Morrissey C, Zhang X, True LD, Gulati R, et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nature Medicine. 2016;22.

49. Prasad V, Addeo A. The FDA approval of pembrolizumab for patients with TMB >10 mut/Mb: was it a wise decision? No. Annals of Oncology. 2020.

50. Palmeri M, Mehnert J, Silk AW, Jabbour SK, Ganesan S, Popli P, et al. Real-world application of tumor mutational burden-high (TMB-high) and microsatellite instability (MSI) confirms their utility as immunotherapy biomarkers. ESMO Open. 2022;7.

51. McGrail DJ, Pilié PG, Rashid NU, Voorwerk L, Slagter M, Kok M, et al. High tumor mutation burden fails to predict immune checkpoint blockade response across all cancer types. Annals of Oncology. 2021;32.

52. van Wilpe S, Kloots ISH, Slootbeek PHJ, den Brok M, Westdorp H, Franken MD, et al. Ipilimumab with nivolumab in molecularly selected patients with castration-resistant prostate cancer: primary analysis of the phase II INSPIRE trial. Annals of Oncology [Internet]. The Author(s); 2024;35:1126–37. Available from: https://doi.org/10.1016/j.annonc.2024.09.004

53. Antonarakis ES, Park SH, Goh JC, Shin SJ, Lee JL, Mehra N, et al. Pembrolizumab Plus Olaparib for Patients With Previously Treated and Biomarker-Unselected Metastatic Castration-Resistant Prostate Cancer: The Randomized, Open-Label, Phase III KEYLYNK-010 Trial. Journal of Clinical Oncology. 2023;41.

54. Fizazi K, Retz M, Petrylak DP, Goh JC, Perez-Gracia J, Lacombe L, et al. Nivolumab plus rucaparib for metastatic castration-resistant prostate cancer: Results from the phase 2 CheckMate 9KD trial. Journal for ImmunoTherapy of Cancer. 2022;10.

55. Yap TA, Bardia A, Dvorkin M, Galsky MD, Beck JT, Wise DR, et al. Avelumab Plus Talazoparib in Patients with Advanced Solid Tumors: The JAVELIN PARP Medley Nonrandomized Controlled Trial. JAMA Oncology. 2023;9.

56. Thapa B, De Sarkar N, Giri S, Sharma K, Kim M, Kilari D. Integrating PARP Inhibitors in mCRPC Therapy: Current Strategies and Emerging Trends. Cancer Management and Research . Dove Medical Press Ltd; 2024. page 1267–83.

57. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25.

58. BroadInstitue. Picard tools: MarkDuplicates. https://broadinstitute.github.io/picard/. 2018.

59. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, Van der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. bioRxiv. 2017.

60. Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, et al. Strelka2: fast and accurate calling of germline and somatic variants. Nature Methods. 2018;15.

61. Benjamin D, Sato T, Cibulskis K, Getz G, Stewart C, Lichtenstein L. Calling Somatic SNVs and Indels with Mutect2. bioRxiv. 2019;

62. Dupain C, Gutman T, Girard E, Kamoun C, Marret G, Castel-Ajgal Z, et al. Tumor mutational burden assessment and standardized bioinformatics approach using custom NGS panels in clinical routine. BMC Biology. 2024;22.

63. Niu B, Ye K, Zhang Q, Lu C, Xie M, McLellan MD, et al. MSIsensor: Microsatellite instability detection using paired tumor-normal sequence data. Bioinformatics. 2014;30.

64. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: The Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Research. 2019;47.

65. Knudson AG. Mutation and cancer: statistical study of retinoblastoma. Proceedings of the National Academy of Sciences of the United States of America. 1971;68.

66. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature Biotechnology. 2013;31.

67. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581.

68. Li MM, Datto M, Duncavage EJ, Kulkarni S, Lindeman NI, Roy S, et al. Standards and Guidelines for the Interpretation and Reporting of Sequence Variants in Cancer. The Journal of Molecular Diagnostics. 2017;19.

69. Horak P, Griffith M, Danos AM, Pitel BA, Madhavan S, Liu X, et al. Standards for the classification of pathogenicity of somatic variants in cancer (oncogenicity): Joint recommendations of Clinical Genome Resource (ClinGen), Cancer Genomics Consortium (CGC), and Variant Interpretation for Cancer Consortium (VICC). Genetics in Medicine. 2022;24.

70. MacArthur DG, Balasubramanian S, Frankish A, Huang N, Morris J, Walter K, et al. A systematic survey of loss-of-function variants in human protein-coding genes. Science. 2012;335.

71. Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536.

72. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nature Genetics. 2014;46.

73. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: Improvements to accessing data. Nucleic Acids Research. 2020;48.

74. Chakravarty D, Gao J, Phillips S, Kundra R, Zhang H, Wang J, et al. OncoKB: A Precision Oncology Knowledge Base. JCO Precision Oncology. 2017;

75. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine. 2015;17:405–24.

76. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: Public archive of relationships among sequence variation and human phenotype. Nucleic Acids Research. 2014;42.

77. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Research. 2010;38.

78. Favero F, Joshi T, Marquard AM, Birkbak NJ, Krzystanek M, Li Q, et al. Sequenza: Allele-

specific copy number and mutation profiles from tumor sequencing data. Annals of Oncology. 2015;26.

79. Gao J, Mazor T, de Bruijn I, Abeshouse A, Baiceanu D, Erkoc Z, et al. cBioPortal for Cancer Genomics. Cancer Research. 2022.

80. Islam SMA, Díaz-Gay M, Wu Y, Barnes M, Vangara R, Bergstrom EN, et al. Uncovering novel mutational signatures by de novo extraction with SigProfilerExtractor. Cell Genomics. 2022;2.

81. Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, et al. Patterns of genomic loss of heterozygosity predict homologous recombination repair defects in epithelial ovarian cancer. British Journal of Cancer. 2012;107.

82. Sztupinszki Z, Diossy M, Krzystanek M, Reiniger L, Csabai I, Favero F, et al. Migrating the SNP array-based homologous recombination deficiency measures to next generation sequencing data of breast cancer. npj Breast Cancer. 2018;4.

83. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS Computational Biology. 2016;12.

84. Broad Institute. Picard Tools - By Broad Institute. Github. 2009.

85. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29.

86. Robinson MD, McCarthy DJ, Smyth GK. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2009;26.

87. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 2015;43.

88. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nature Methods. 2015;12.

89. Deshwar AG, Vembu S, Yung CK, Jang GH, Stein L, Morris Q. PhyloWGS: Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. Genome Biology. 2015;16.

90. Dang HX, White BS, Foltz SM, Miller CA, Luo J, Fields RC, et al. ClonEvol: Clonal ordering and visualization in cancer sequencing. Annals of Oncology. 2017;28.

Figure Legend:

Fig. 1: Genomic Stratification of Advanced Prostate Cancer Reveals Distinct Molecular Substructures Across DDR Pathways (**A**) Curated genomic calls and tumor mutation burden analysis revealed 14.31% of advanced prostate cancers are HRR bi-allelic mutant, 2.3% harbor both TMB-high (>10 mut/Mb) and HRR mutations, while 71% are DDR-intact. We identified ~3.9% of advanced prostate cancers as TMB-vH using our 18.61 TMB cutoff, with this subset enriched for MMR pathway gene bi-allelic loss-of-function mutations. The HRR subset showed the highest BRCA2 loss-of-function

frequency. (**B**) Using the 18.61 TMB cutoff revealed substantial molecular dichotomy within the TMB-H subgroup, distinguishing TMB-vH tumors with concurrent HRR mutations. All TMB-vH tumors, regardless of concurrent HRR mutation status, were enriched for MSI-high scores, AR pathogenic mutations, and COSMIC MMRd signatures while negatively enriched for AR copy amplification and HRD-Scar signatures(p=0.76). The TMB-high 10-18.61 subgroup was instead enriched for HRD-Scar scores. Based on observed signatures, we detected no distinctive molecular differences between TMB-vH and TMB-vH with concurrent HR mutations. Copy number annotations were performed using the homebrew tool pyTHiA-CNA (Supplementary Document 5).

Fig. 2: Multi-Dimensional Molecular Profiling Enables Fine-Scale Classification into 11 Molecular Subclasses Along the DDR Axis Through Integrated Genomic, Transcriptomic, and Pathway-Specific Signature Analysis (**A-F**) Distinctive intrinsic molecular feature dichotomy within concurrent TMB-high and HR mutant subsets clearly supports the molecular basis for 18.61 TMB cutoff-based subgrouping (Wilcoxon rank test). Analysis revealed molecular dichotomy based on tumor mutation burden, microsatellite instability score, tumor ploidy, HRD-Scar score, and percent genome alterations. LOH scores from Sequenza showed no striking differences. We report significant differences in tumor ploidy, LOH, HRD-Scar score, and PGA between DDR-intact and DDR-intact subgroups with bi-allelic tumor suppressor gene mutations (Wilcoxon rank-sum test for continuous variables like TMB, ploidy, PGA and LOH) (Fisher's exact test for categorical features like MSI and Scar, *::p<0.01, **:: p<0.001, ***::p<0.0001, ****::<0.00001, *****::<0.000001, ******::<0.0000001, *******::<0.00000001). All other comparisons shown as 'ns' stand for statistically non-significant. HRD-Scar). Note, 4 (TMB-vH HRRm, TMB-H HRRm, TMB-vH & TMB-H) of the 10 subgroups are defined by TMB status. Significant P-value is expectedly high or can be ignored. PGA analysis was performed using home brew tool PGA_Calculator [https://github.com/Niel20/PGA_Estimator, Supplementary Document 6] (**G**) Gene expression profiling revealed ~50% of TMB-vH subclass tumors exhibit elevated PD-L1 and LAG3 expression. (**H**) TMB-vH subgroups demonstrated consistently elevated TP53 mRNA expression (corrected p value = 0.005). TMB-vH HRR mutant subgroups displayed elevated AR activity scores with low NE scores but high CCP scores, contrasting with TMB-H subgroups that showed reduced AR activity despite elevated CCP scores. BRCA2 expression exhibited strong correlation with MKI67 and CCP scores across most subgroups, though this relationship was disrupted in TMB-vH-HRR mutant tumors (Pearson's correlation analysis, ρ= -0.351). MSH2 and RAD51B showed notably low expression enrichment in TMB-vH subsets regardless of HRR mutation status, maintaining consistently above-median expression levels in TMB-vH tumors while other subgroups displayed broader dynamic expression ranges.

Fig. 3: Clinical Cohort Analysis Validates TMB >18.61 Cutoff Defining a Distinct MMRd-Like Molecular Subset with Enhanced ICI Response in Advanced Prostate Cancer (**A**) Heatmap demonstrates substantial molecular subtype representation in real-world pooled case series from 4 institutions. This cohort shows overrepresentation of concurrent TMB-H or TMB-vH tumors with or without HRR mutations due to targeted patient selection for robust validation, and substantial underrepresentation of DDR-intact subsets. Since MSI scores and detected mutations are clinical-grade assays that validate our discovery cohort-derived molecular phenotype patterns, this bolsters accuracy of our proposed molecular subgrouping. Heatmap is plotted with home brew tool D-AV plotter (Supplementary Document 7) (**B**) Enrichment of MMR pathway gene bi-allelic mutations in TMB-vH subsets independent of concurrent HRR mutation status validates discovery findings. (**C**) MSI score enrichment in TMB-vH subsets confirms inference accuracy observed in the primary 482-sample discovery cohort. (**D**) Survival curves demonstrate TMB-vH and TMB-vH HRR mutant subsets show significantly superior immune checkpoint inhibitor response, highlighting clinical relevance of accurate TMB-vH subtype determination in prostate cancer. (**E**) Swimmers plot reveals 100% of TMB-vH HRR mutant and 85% of TMB-vH tumors achieved PSA50 response—significant enrichment compared to HRR mutant or DDR-intact subsets ($p<0.005$, Fisher's exact test). (**G**) Consistent with discovery cohort findings, we observed significant AR pathogenic mutation enrichment, though unlike the discovery cohort, negative enrichment of AR copy amplification events was not statistically significant. (**H**) Mutation frequency enrichment analysis revealed borderline enrichment of detected TP53 mutation events in all four TMB >10 mut/Mb groups with and without concurrent HRR mutations compared to DDR-intact subgroups.

Fig. 4: Accurate Dynamic Classification of Advanced Prostate Cancer Using a Multimodal Multilabel Random Forest-Based Nested Classification Model (A) Schematic flowchart illustrating CHIMERA-DDR architecture employing a unique nested classification strategy that first delineates major 4 subclasses, then identifies more nuanced finer subclasses to infer 7 total subgroups. This model integrates 51 genomic features and signature scores from patient tumor genome and transcriptome sequencing data within a probabilistic framework to predict the relative admixture of DDR deficiency phenotypes. **(B)** Schematic illustration of CHIMERA-DDR's 2-tier nested classification architecture enabling hierarchical DDR subtype prediction. **(C & D)** Component barplot percent representation of CHIMERA-predicted probability decomposition for TMB-vH and TMB-H tumors with HRR mutations. Results demonstrate that TMB-vH HRR mutant subgroup tumors exhibit predominantly TMB-vH-like (MMRd-like) characteristics, whereas TMB-H HRR mutant subgroups display prevalently HRRd-like phenotypes. One of the 19 tumors with DSB sensor mutations is HRRd positive in Tier 1 analysis with Tier 2 ambiguity while defining fine-scale HRRd subgrouping, suggesting potential low-

penetrance nature of these genomic events.**(E)** Clonal analysis of TMB-vH-HRR mutant sample 00-010 reveals a truncal MMRd mutation in MSH2 gene likely drives the MMRd-associated hypermutator phenotype, with minor clone S2 harboring additional potentially subclonal copy-loss events in BRCA2 and CHD1 genes. Branch length: Represents the relative mutational burden or evolutionary distance (Relative frequency of somatic events). Node width: Indicates clonal prevalence (cellular fraction that carries those somatic events).

Fig. 5: Leave-one-out cross-validation (LOO-CV) derived AUC-ROC curves showing superior technical performance in both Tier 1 and Tier 2 predictions. CHIMERA-DDR stratifies patients into 4 DDR axis molecular subgroups. Tier 2 analysis sub-stratifies the HRR mutant subgroups into HRRd with loss-of-function of selective tumor suppressor genes and those without tumor suppressor gene loss. Notably, tumor suppressor gene mutations in TP53, RB1, and PTEN impact the manifestation of HRRd-associated signature profiles and disease pathobiology. Similarly, in Tier 2 DDR-intact tumors devided into 3 molecular subgroups based on RecQ helicase pathway activity and tumor suppressor gene mutation status in the setting of DDR-intact molecular phenotype for accurate fine-scale class determination. CHIMERA-DDR demonstrated superior performance in deriving 7 molecular subgroups after Tier 2 analysis (TMB-vH, TMB-H, HRRd, HRRd tumor suppressor, RecQ helicase mutant, DDR-intact tumor suppressor, and DDR-intact without tumor suppressor mutations). Tier 1 + 2 complete analysis infers 7 fine-tuned molecular subgroups demarcated in the figure with solid blue-filled boxes in the top left of the respective AUC-ROC plot areas.

Fig. 6: CHIMERA-DDR Demonstrates Unique Capability to Predict Admixed / Concurrent DDR-Intact Subtype Probabilities and Captures Inter-Tumoral Heterogeneity in Advanced Prostate Cancer (A) Scatter plot comparing CHIMERA's predicted DDR-intact probability versus normal DNA fraction in tumor samples. For likely DDR-intact tumors, CHIMERA's overall Tier 2 DDR-intact prediction probability consistently exceeded the actual normal fraction inferred by Sequenza cellularity (normal fraction = 1-cellularity), initially conceived these probabilities might reflect cumulative contributions of normal DNA fraction and tumor-intrinsic DDR-intact phenotypes. Extended analysis demonstrates that DDR-intact predicted probability is not a function of sample tumor cellularity but rather is primarily guided by the tumor's DDR-intact-like molecular characteristics, tumor’s clonal structure indicating tumor-intrinsic DDR functional states detection; also refer Supplementary Fig. 6-8. **(B)** Distribution of 52 tumor samples from 24 patients across multiple metastatic sites (lung, visceral, soft tissue, lymph node, bone, liver) from the UW-TAN case series covering 5 DDR subgroups with corresponding whole genome sequencing data. This test series enabled evaluation of CHIMERA's robustness for DDR status inference using signature scores derived from multiple metastatic sites. **(C)**

Stacked bar plots showing Tier 1 and Tier 2 DDR subtype prediction probabilities across multiple tumor samples from the same patients. Each bar represents relative prediction probabilities for DDR subtypes (TMB-vH, TMB-H, HRR mutant, DDR-intact) in Tier 1, with refined subcategorization in Tier 2 (DSB sensor/TP53/PTEN mutations, BRCA2 mutations, PTEN/TP53 mutations, and no mutations). Asterisks indicate tumors with CHIMERA-DDR intratumoral heterogeneity. Analysis reveals substantial homogeneity in DDR deficiency prediction probabilities within patient groups, with notable heterogeneity observed in select cases.

Data Availability: The SU2C whole-exome sequencing data and SU2C bulk RNAseq data: available https://github.com/cBioPortal/datahub/tree/master/public/prad_su2c_2019.
The UW-TAN whole-exome sequencing data:
https://github.com/cBioPortal/datahub/tree/master/public/prad_fhcrc
UW- TAN Bulk RNAseq data: Gene Expression Omnibus repository (GEO) IDs: GSE147250, GSE171729, GSE228283.
UW-TAN Whole genome sequence VCF and copy number segments:
https://github.com/Niel20/CHIMERA_DDR_Validation_Cohort_52_WGS

Figure 1

A

TMB (Log2)
Homologous recombination
BRCA2
BRCA1
PALB2
RAD51
RAD51B
RAD51C
RAD51D
BARD1
BRIP1
RAD52
RAD54L2
XRCC2
RBBP8
CHD1
SPOP
Fanconi
FANCB
FANCD2
FANCM
FANCI
FANCL
DSB sensor/ end processor
ATM
CHEK2
MRE11
NBN
DNA2
EXO1
RecQ helicase
WRN
BLM
RECQL5
RECQL
RECQL4
Mismatch Repair
MSH2
MSH6
MLH1
MSH3
PMS2
MLH3
Others
ATRX
ATR
GEN1
RB1
RNASEH2B
POLE2
TP53
PTEN
AR
≥10 TMB-H HRRm
TMB-vH
TMB-H
MMRm TMB-L
HRRm
HRRm tum. supp.
DSB sens.
RecQ
DDR int. tum. supp.
DDR int.
0 20 40 60 Percent
HRRm (14.31%)
DDR intact (71.57%)
Copyloss
Copygain
Amplification
Hyperamplification
Structural Aberration
Truncating
Splice Site
Pathogenic Missense
Germline Pathogenic
Hypomorphic Mutation
TMB 10
TMB 18.61
NA
LOH
Nonpathogenic Missense & Inframe

B

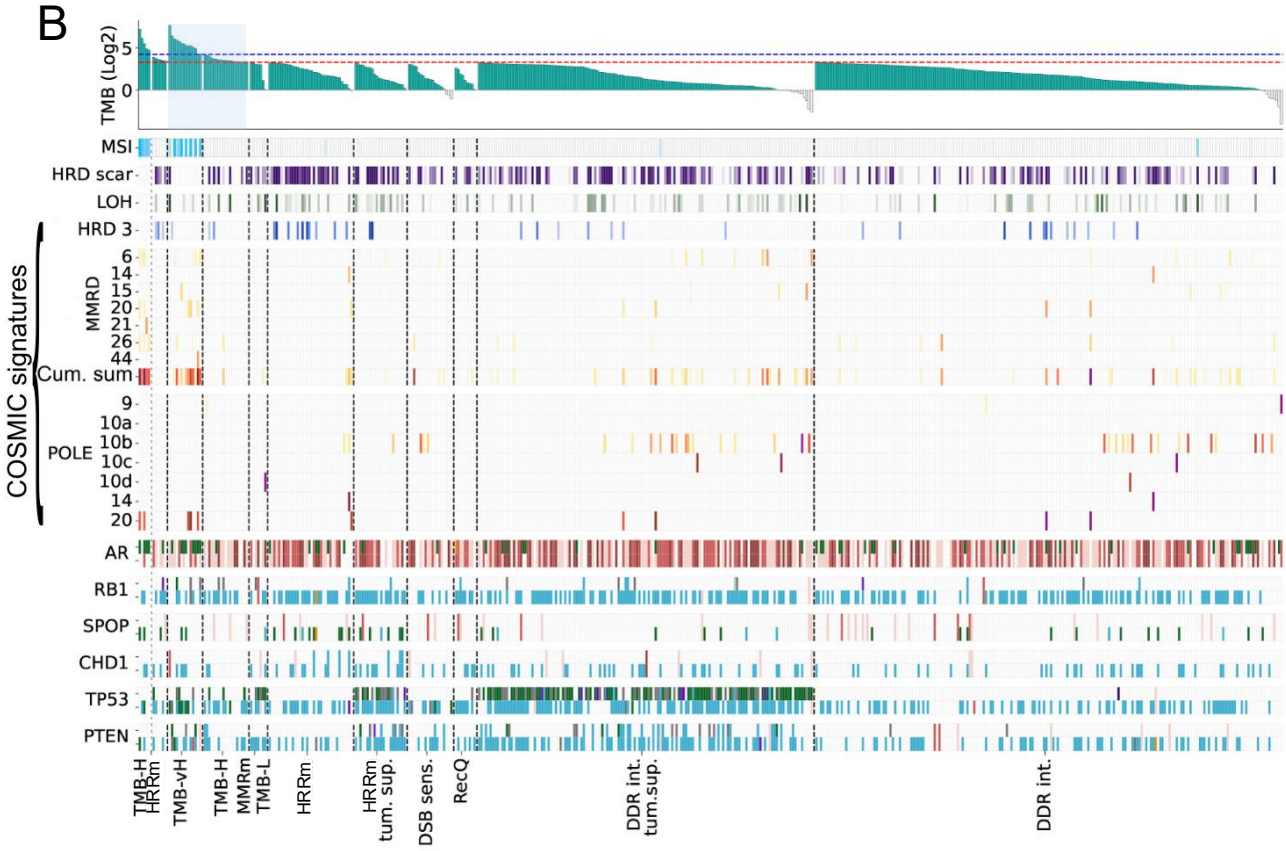

TMB (Log2)
MSI
HRD scar
LOH
HRD 3
COSMIC signatures
MMRD
6
14
15
20
21
26
44
Cum. sum
POLE
9
10a
10b
10c
10d
14
20
AR
RB1
SPOP
CHD1
TP53
PTEN
TMB-H HRRm
TMB-vH
TMB-H
MMRm TMB-L
HRRm
HRRm tum. sup.
DSB sens.
RecQ
DDR int. tum.sup.
DDR int.

Figure 2

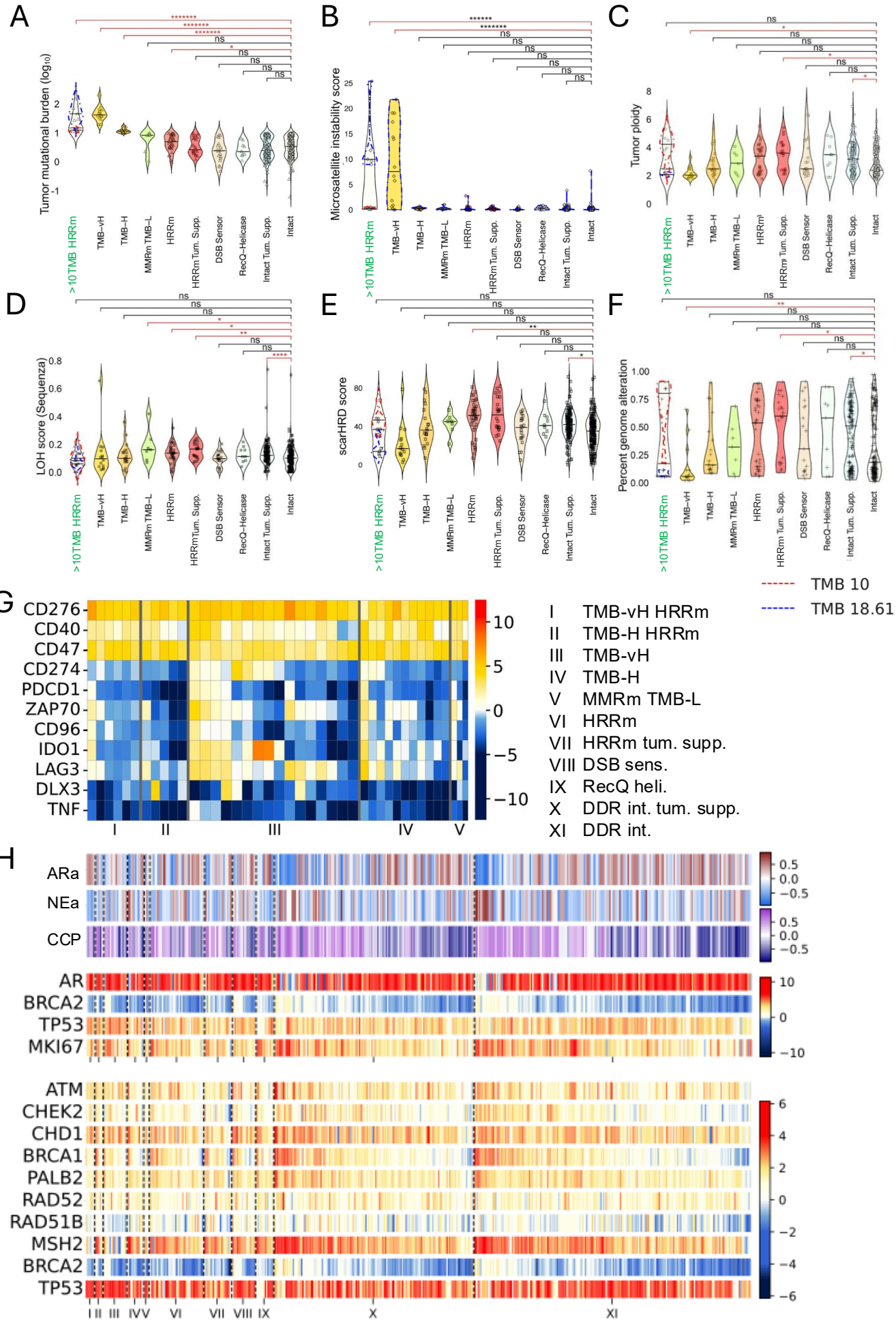
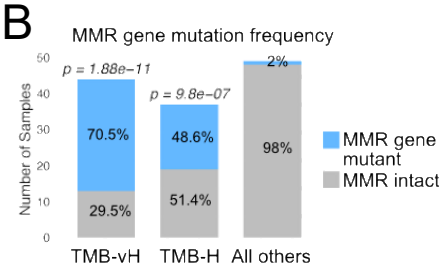
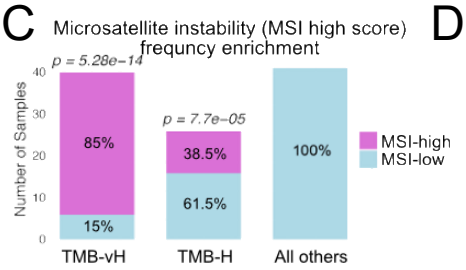
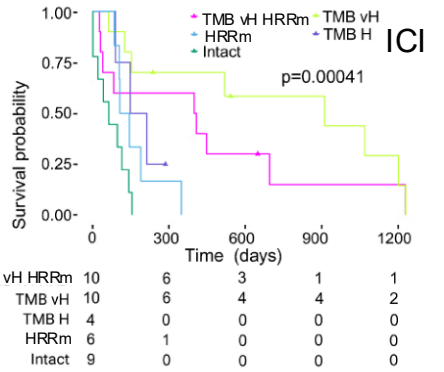
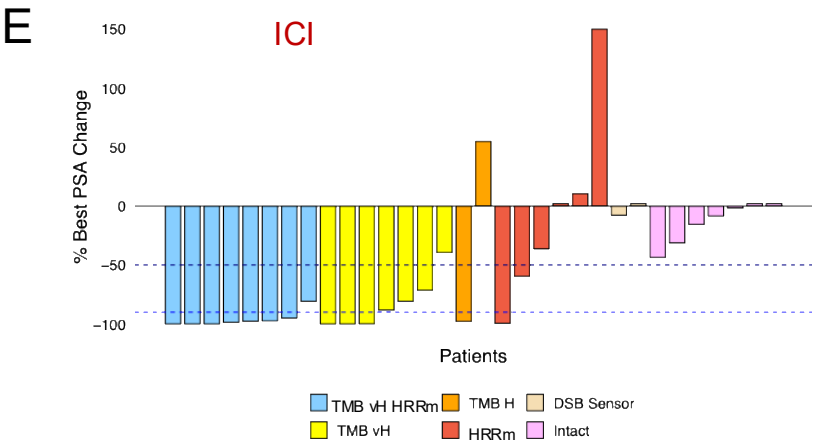

A
Tumor mutational burden (log10)
B
Microsatellite instability score
C
Tumor ploidy
D
LOH score (Sequenza)
E
scarHRD score
F
Percent genome alteration
>10TMB HRRm
TMB–vH
TMB–H
MMRm TMB–L
HRRm
HRRm Tum. Supp.
DSB Sensor
RecQ–Helicase
Intact Tum. Supp.
Intact
TMB 10
TMB 18.61
G
CD276
CD40
CD47
CD274
PDCD1
ZAP70
CD96
IDO1
LAG3
DLX3
TNF
I TMB-vH HRRm
II TMB-H HRRm
III TMB-vH
IV TMB-H
V MMRm TMB-L
VI HRRm
VII HRRm tum. supp.
VIII DSB sens.
IX RecQ heli.
X DDR int. tum. supp.
XI DDR int.
H
ARa
NEa
CCP
AR
BRCA2
TP53
MKI67
ATM
CHEK2
CHD1
BRCA1
PALB2
RAD52
RAD51B
MSH2
BRCA2
TP53

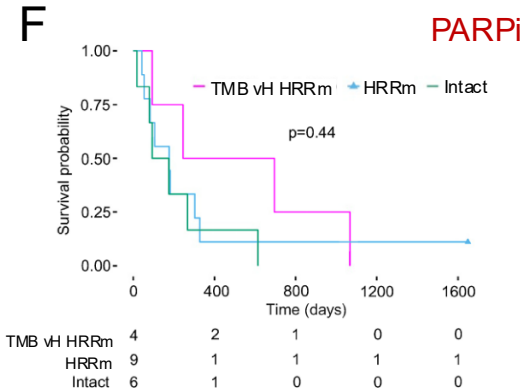
Figure 3











G AR Alteration Frequency (Mutations and Gains) in Comparison with DDR-Intact Tumors

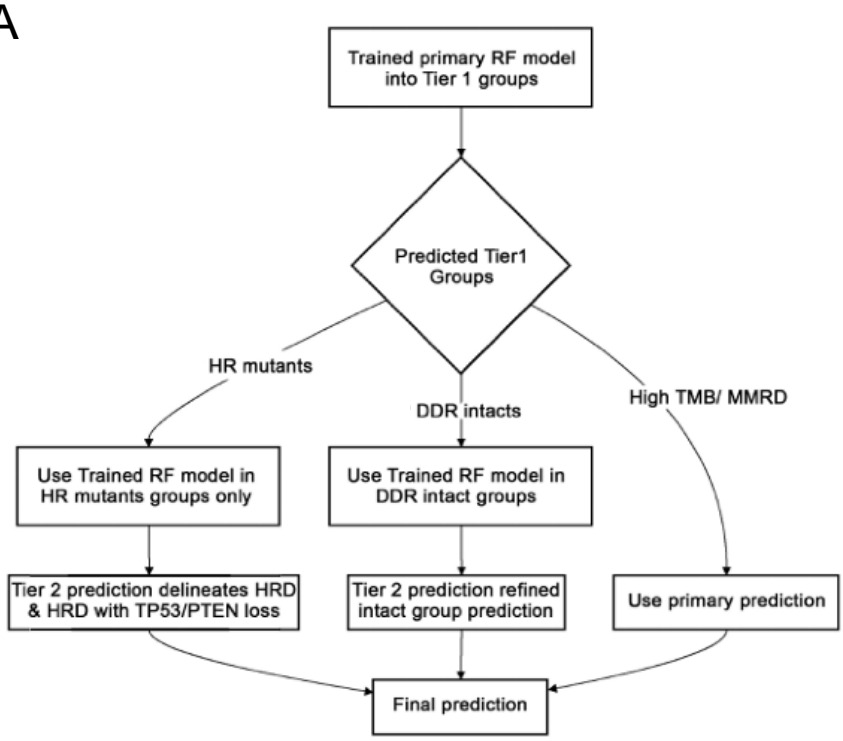
	# Samples	AR_gain	Gain frequency	p_value	Odds_Ratio	AR mutation	Mutation frequency	p_value
TMB vH HRRm	25	2	0.08	0.0738	0.22	9	0.36	**0.002**
TMB H HRRm	8	3	0.375	0.681	1.44	2	0.25	0.057
TMB vH	19	3	0.158	0.47	0.46	5	0.263	**0.012**
TMB H	29	7	0.241	0.76	0.78	4	0.138	0.117
HRRm	17	3	0.176	0.48	0.53	1	0.059	0.415
MMRm TMB-L	1	0	0	1	0	0	0	NA
DDR intact	24	7	0.292	----------	----------	0	0	NA
DSB sens.	7	2	0.286	1	0.97	1	0.143	0.226

H

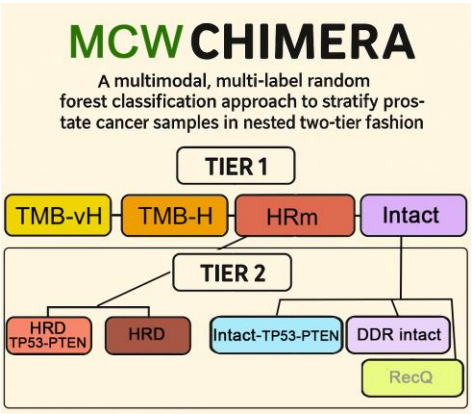
	# Samples	TP53	p_value	Odds_Ratio
TMB vH HRRm	25	5	**0.0186**	0.22
TMB H HRRm	8	0	**0.0104**	0
TMB vH	19	2	**0.0038**	0.11
TMB H	29	5	**0.00813**	0.18
HRRm	17	8	0.756	0.76
MMRm TMB-L	1	1	1	NA
DDR intact	24	13	----------	----------
DSB sens.	7	3	0.685	0.64

Figure 4

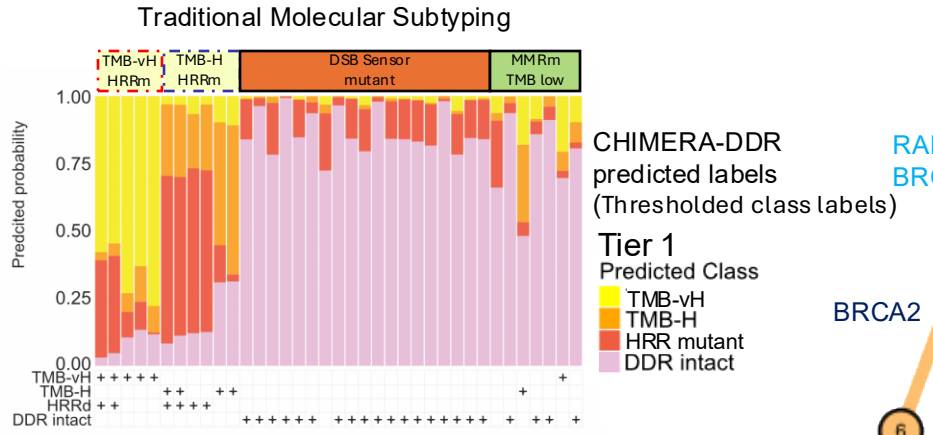
A



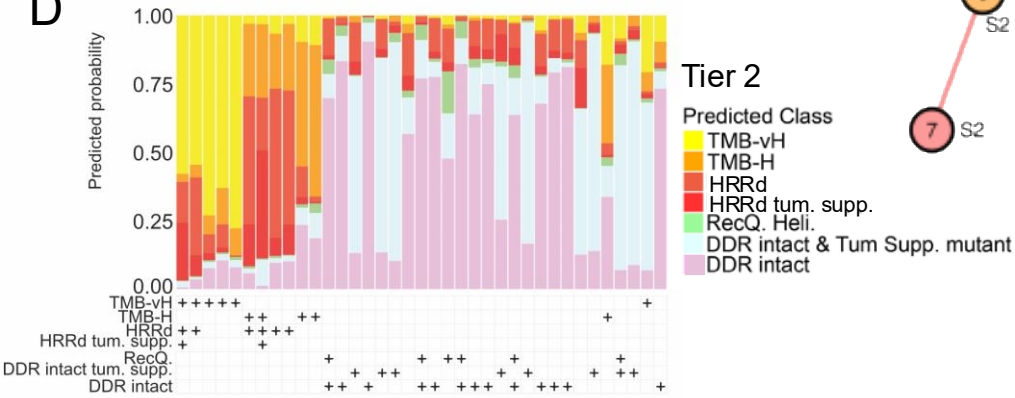
B



C



D



E

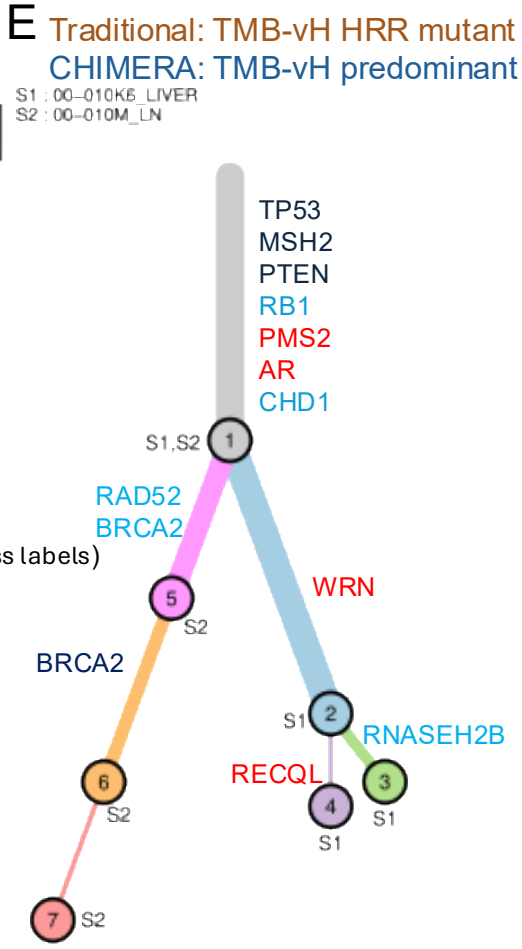


Figure 5

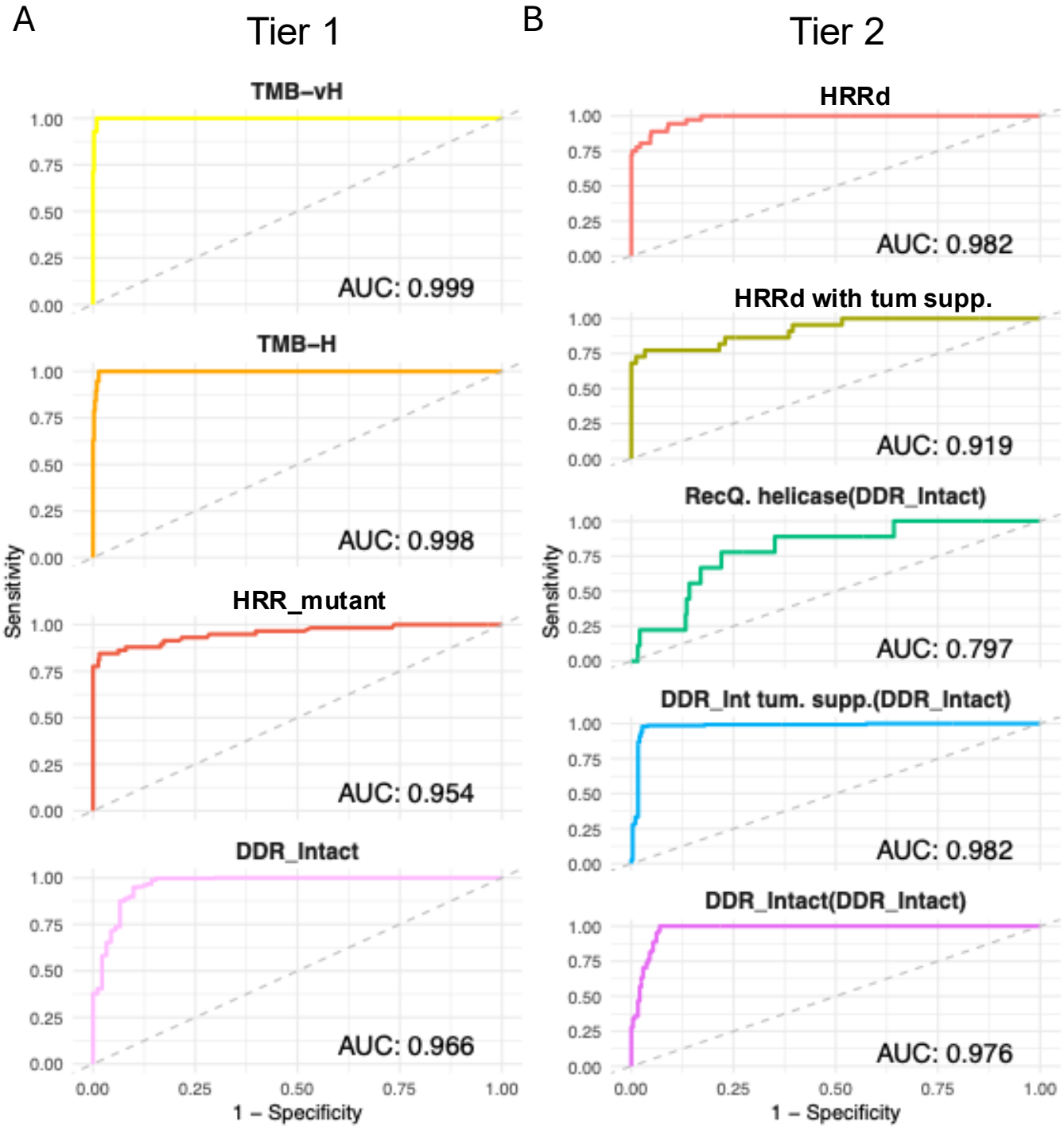
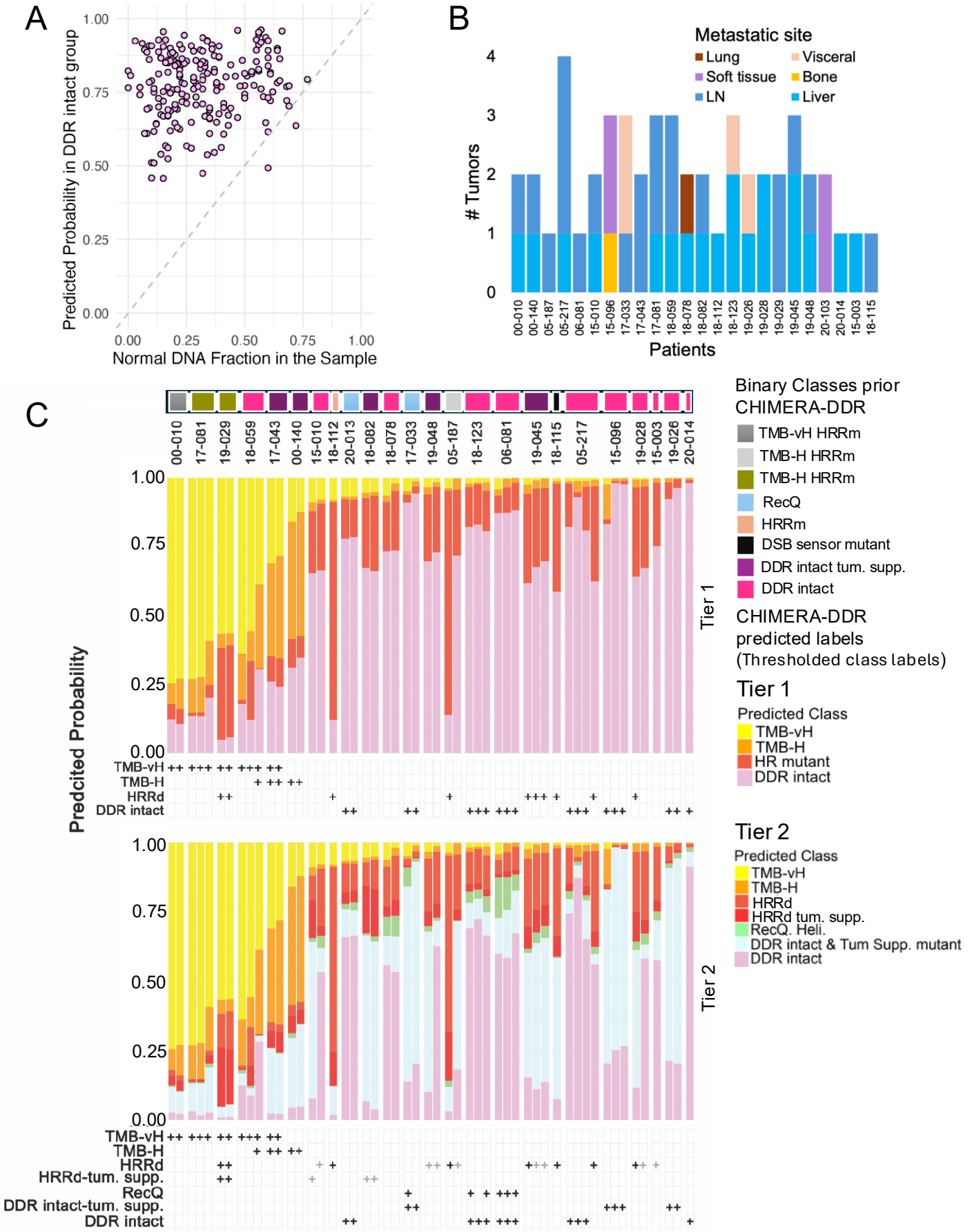

A
Tier 1
TMB–vH
AUC: 0.999
TMB–H
AUC: 0.998
HRR_mutant
AUC: 0.954
DDR_Intact
AUC: 0.966
B
Tier 2
HRRd
AUC: 0.982
HRRd with tum supp.
AUC: 0.919
RecQ. helicase(DDR_Intact)
AUC: 0.797
DDR_Int tum. supp.(DDR_Intact)
AUC: 0.982
DDR_Intact(DDR_Intact)
AUC: 0.976
Sensitivity
1 – Specificity
0.00
0.25
0.50
0.75
1.00

Figure 6


A
Predicted Probability in DDR intact group
Normal DNA Fraction in the Sample
B
Metastatic site
Lung
Visceral
Soft tissue
Bone
LN
Liver
Tumors
Patients
C
Binary Classes prior CHIMERA-DDR
TMB-vH HRRm
TMB-H HRRm
TMB-H HRRm
RecQ
HRRm
DSB sensor mutant
DDR intact tum. supp.
DDR intact
CHIMERA-DDR predicted labels (Thresholded class labels)
Predicted Probability
Tier 1
Predicted Class
TMB-vH
TMB-H
HR mutant
DDR intact
Tier 2
Predicted Class
TMB-vH
TMB-H
HRRd
HRRd tum. supp.
RecQ. Heli.
DDR intact & Tum Supp. mutant
DDR intact
TMB-vH
TMB-H
HRRd
HRRd-tum. supp.
RecQ
DDR intact-tum. supp.
DDR intact

Graphical Abstract

A

Key Insights:

- Tumor Mutation Burden (TMB) ≥ 19 show + COSMIC MMRd signatures and MMR gene mutation, limited aneuploidy, elevated frequency of AR LBD mutations.
- Clonal heterogenity underlines MMRd like vs HRRd like divergence in TMBvH HRRm mCRPC.
- MMRd bi-allelic mutation <10 TMB lack MMRd like genomic signatures
- CHIMERA-DDR probabilistically quantifies DDRd phenotypes on a relative scale, enabling deconvolution of concurrently mixed DDR phenotypes and assessment of individual mutant contributions to DDRd behaviors.

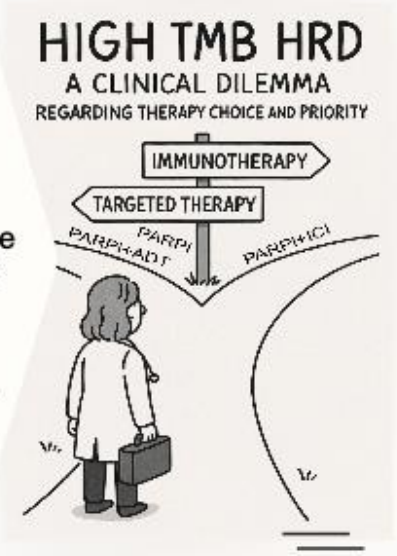
B

Challenges

Multiple DDR Mutations

Mutation positive non responders

Low allele frequency DDR mutations



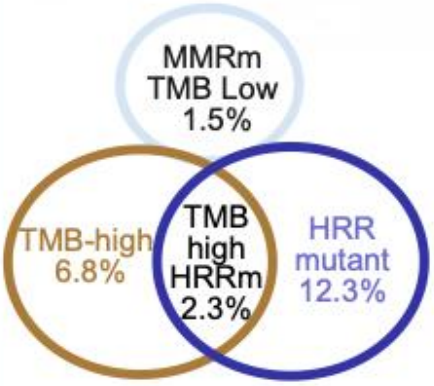
Advancements

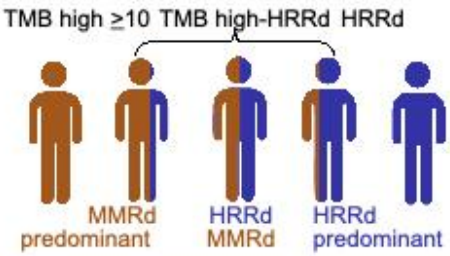
CHIMERA-DDR

Dynamic DDR probabilities

- Prevalent DDRd type
- Resolves admixed/ concurrent DDRd signatures
- Hypomorphism
- Resolves DDR intact like characteristic
- Accurate molecular stratification of patients

C





CHIMERA prediction	Positive		Predominant	
	MMRd	HRRd	MMRd	HRRd
TMB-vH HRRm (n=5)	5	2	5	0
TMB-H HRRm (n=6)	4	6	2	4