Dissertation

Multimodal liquid biopsy analysis for the detection of homologous recombination deficiency in prostate cancer

submitted by

Georgios VLACHOS, BSc Mpharm MSc

for the Academic Degree of

Doctor of Philosophy (PhD)

at the

Medical University of Graz Diagnostic & Research Institute of Human Genetics

under the Supervision of

Univ.-Prof. Mag. Dr. rer. nat.

Ellen HEITZER

2026

STATUTORY DECLARATION

I hereby confirm that the present diploma thesis is the result of my own independent scholarly work. I also confirm that in all cases, where material from the work of others (in books, articles, essays, dissertations, and on the internet) is acknowledged, quotations and paraphrases are clearly indicated. No material other than that cited in the reference list has been used. I have read and understood the Medical University's regulations and procedures concerning plagiarism. Furthermore, I hereby declare that if artificial intelligence (AI) tools were used for the generation and/or correction of certain text passages in the creation of this work, such employment was conducted in compliance with ethical principles, academic integrity, and the regulations of my university. Additionally, it was ensured that this usage was transparently disclosed and appropriately attributed.

Georgios Vlachos; January 2026

DISCLOSURES

Part of this thesis has been published in:

Georgios Vlachos[1,2], Tina Moser[1], Isaac Lazzeri[1,2], Matthias J. Moser[1], Lisa Glawitch[1], Emil Thomas Bauernhofer[1], Anna Eberhard[1], Christine Beichler[1], Hanieh Sadeghi[1], Jasmin Blatterer[1], Stefan Kühberger[1], Nina Monsberger[1], Angelika Terbuch[3], Karl Kashofer[4], Jochen B. Geigl[1], Thomas Bauernhofer[3], Ellen Heitzer[1,2]. Functional footprints of homologous recombination deficiency in prostate cancer revealed by ctDNA fragmentation and transcription factor accessibility. British Journal of Cancer. 2026. doi:10.1038/s41416-025-03301-0.

[1]Institute of Human Genetics, Diagnostic & Research Center for Molecular BioMedicine, Medical University of Graz, Austria

[2]Christian Doppler Laboratory for Liquid Biopsies for Early Detection of Cancer, Medical University of Graz, Neue Stiftingtalstrasse 6, 8010 Graz, Austria

[3]Department of Internal Medicine Graz, Division of Oncology, Medical University of Graz

[4]Diagnostic and Research Institute of Pathology, Medical University Graz, Graz, Austria

All co-authors have agreed to the inclusion of their published data in the dissertation and permission to reproduce illustrations and figures from own or third-party publications has been granted.

Acknowledgements

It is difficult to adequately acknowledge everyone who contributed to this journey; the list of people I would like to thank is far longer than what can reasonably fit on a single page.

My deepest gratitude goes to **Prof. Ellen Heitzer**, my supervisor. Thank you for your guidance through all these years, for the time you consistently made available, and for the patience with which you taught me how to approach research questions with rigor and curiosity. I am especially grateful for the trust you placed in me and the freedom you gave me to pursue ideas that mattered to me, as well as for the flexibility and understanding you showed when I needed space to choose the direction, I believed was most scientifically sound. Your exceptional work ethic has shaped the environment of our group and, through it, has enabled opportunities and collaborations that strengthened this work and my development as a researcher. I also deeply appreciate your constant effort to secure the resources that make our research possible in practice, especially the support for conference participation, which allowed us to disseminate our work, build new collaborations, and share many memorable adventures that expanded both our scientific and human horizons.

I would like to thank all past and present members of **Humangenetik** for the supportive environment that made this work possible. In particular, I would like to thank **Tina, Isaac, Benjamin, Nina, Silvia, Hanieh, Kathi, Stefan, Anna, Samantha, Leandra, Christine, Ricarda, Raul, Maris, Faruk, Matthias and Maida** for the many everyday moments of help; scientific, practical, and personal. **Tina**, I am especially grateful for teaching me the wet-lab work with remarkable patience and care, and for the times you stayed late to help me troubleshoot and finish experiments - once even at the expense of a Pilates appointment, which I

have not forgotten. To the whole group; I will also always cherish our lunch/coffee breaks (as well as our Wednesday evenings) as a group, which made this long process very enjoyable.

I would also like to express my gratitude to the late **Prof. Michael R. Speicher** and **Prof. Jochen B. Geigl** for their guidance and support, particularly during the earlier phases of my PhD. I am grateful for their scientific perspective, encouragement, and the standards they upheld; standards that helped shape the foundations of our Institute and continue to push me to uphold the same.

I would like to thank **Prof. Karl Kashofer** and **Prof. Thomas Bauernhofer** for serving on my scientific committee. I am grateful for their time, their thoughtful input, and their commitment to evaluating and strengthening this work. In particular, I would like to thank **Prof. Bauernhofer** for providing an essential connection to the clinical perspective, ensuring that the research remained anchored to real-world needs and meaningful translational context.

Finally, I want to thank my family and friends, without whom this journey would not have been possible. To my parents, **Kosmas** and **Katerina**, thank you for raising me to see education not as an obligation, but as a privilege. To my sister, **Fotini**, thank you for always being there for me, in the way that matters most. I am grateful to **Nathalie** for her patience and kindness, and for enduring far too many rounds of presentation practice. Lastly, to my **friends** back home, thank you for being there all this time.

This work received funding from the Medical University of Graz through the PhD Program in Molecular Medicine (MolMed). In addition, this work was supported by the Austrian Federal Ministry for Digital and Economic Affairs through the Christian Doppler Research Fund for Liquid Biopsies for Early Detection of Cancer.

Table of Contents

Abbreviations

ADT	Androgen Deprivation Therapy
AUC	Area Under the Curve
ATM	Ataxia Telangiectasia Mutated
ATR	Ataxia Telangiectasia and Rad3-related protein
BARD1	BRCA1 Associated RING Domain protein 1
BAM	Binary Alignment Map
BND	Breakend
bp	Base Pair
BRCA1 / BRCA2	Breast Cancer susceptibility gene 1 / 2
CBP	CREB-Binding Protein
CDK12	Cyclin-Dependent Kinase 12
cfDNA	Cell-free DNA
CHIP	Clonal Hematopoiesis of Indeterminate Potential
CRPC	Castration-Resistant PCa
ctDNA	Circulating tumor DNA
DBS	Doublet Base Substitution signature
DDR	DNA Damage Repair
DNA	Deoxyribonucleic acid
DSB	Double-Strand Break
dsDNA	Double-Stranded DNA
FANC	Fanconi Anemia Complementation group
FFPE	Formalin-Fixed, Paraffin-Embedded
GIS	Genomic Instability Score
GTRD	Gene Transcription Regulation Database
H&E	Haematoxylin and Eosin Staining
HR	Homologous Recombination
HRD	Homologous Recombination Deficiency
HRR	Homologous Recombination Repair
HSPC	Hormone Sensitive PCa
ICI	Immune Checkpoint Inhibitors
ID	Insertion/Deletion signature
kb	Kilobase

LGA	Large-scale Genomic Alteration
LINE-1	Long Interspersed Nuclear Element-1
LOH	Loss Of Heterozygosity
LST	Large-scale State Transitions
mCRPC	Metastatic Castration-Resistant PCa
mFAST-SeqS	Modified Fast Aneuploidy Screening Test Sequencing System
mHSPC	Metastatic Hormone Sensitive PCa
MMR	Mismatch Repair
MSH2	MutS Homolog 2
MSI	Microsatellite Instability
NEPC	Neuroendocrine PCa
NHEJ	Non Homologous End Joining
NGS	Next Generation sequencing
OS	Overall Survival
PALB2	Partner And Localizer of BRCA2
PARP	Poly(ADP-ribose) Polymerase
PARPi	Poly(ADP-ribose) Polymerase inhibitor
PCa	Prostate Cancer
PFS	Progression-Free Survival
PTEN	Phosphatase and Tensin homolog
RB1	Retinoblastoma 1
RPA	Replication Protein A
sHRD	Shallow HRD score
sWGS	Shallow Whole-Genome Sequencing
SBS	Single Base Substitution signature
SNV	Single Nucleotide Variant
SCNA	Somatic Copy Number Alteration
SSA	Single Strand Annealing
SSB	Single Strand Break
t-SNE	t-Distributed Stochastic Neighbor Embedding
TAI	Telomeric Allelic Imbalance
TCGA	The Cancer Genome Atlas
TF	Tumor Fraction
TFBS	Transcription Factor Binding Site
UMI	Unique Molecular Identifier

VAF	Variant Allele Frequency
VUS	Variant of Uncertain Significance
WES	Whole Exome Sequencing
WGS	Whole Genome Sequencing
ZNF	Zinc Finger Protein

List of Figures

List of Tables

Zusammenfassung

Die homologe Rekombinationsdefizienz (HRD) ist ein wesentliches molekulares Merkmal, das mit einer erhöhten Ansprechrate auf PARP-Inhibitoren und platinbasierte Chemotherapien beim Prostatakarzinom assoziiert ist. Aktuelle diagnostische Ansätze erfassen HRD jedoch nur unvollständig, da sie sich primär auf den Nachweis von *BRCA1*- oder *BRCA2*-Mutationen oder auf gewebebasierte Analyse sog. „genomic Scars" (genomische Narben) stützen. Darüber hinaus wird die gewebebasierte Testung häufig durch die ausgeprägte intratumorale Heterogenität sowie durch die Schwierigkeit, Gewebeproben von Patienten mit metastasierendem, knochendominantem Krankheitsverlauf zu gewinnen, eingeschränkt. Ziel dieser Arbeit war es, ein minimalinvasives und umfassendes Konzept zur Charakterisierung von HRD anhand von zirkulierender Tumor-DNA (ctDNA) aus Plasmaproben zu entwickeln.

Dazu wurde ctDNA aus Plasma von 106 Männern mit metastasiertem Prostatakarzinom mit einem multimodalen Analyseansatz untersucht, der die gezielte Sequenzierung von Genen der homologen Rekombinationsreparatur (HRR), eine „shallow whole-genome sequencing" (sWGS)-Analyse zur Quantifizierung der genomischen Instabilität, eine tumor-normal-abgestimmte Whole-Exome-Sequenzierung zur Erfassung von Mutationssignaturen sowie eine fragmentomische Analyse der cfDNA-Größenverteilung und der Chromatinzugänglichkeit kombinierte.

Unter den HRR-Genen waren *BRCA2*-Veränderungen am häufigsten und traten häufig gemeinsam mit einem *PTEN*-Verlust auf. Erhöhte genomische Instabilitätswerte korrelierten mit kombiniertem *BRCA2*- und *RB1*-Verlust, einer hohen Anzahl somatischer Kopienzahlveränderungen und einer verkürzten Gesamtüberlebenszeit. HRD-positive Tumoren zeigten eine Anreicherung der charakteristischen Mutationssignaturen SBS3 und ID6, eine relative Zunahme von dinukleosomalen cf DNA-Fragmenten sowie eine verminderte Zugänglichkeit an Zinkfinger-Transkriptionsfaktor-Bindungsstellen, was auf Veränderungen der Nukleosomenorganisation und der Chromatinstruktur hinweist. Basierend auf diesen Merkmalen konnte ein auf fragmentomischen Daten trainierter maschineller Lernklassifikator HRD-positive Fälle zuverlässig identifizieren.

Diese Ergebnisse belegen, dass die multimodale Analyse von ctDNA ein leistungsfähiges, nichtinvasives Verfahren zur Erkennung von HRD beim Prostatakarzinom darstellt. Die Kombination

aus mutationalen, strukturellen und fragmentomischen Informationen ermöglicht ein umfassenderes und funktionelleres Verständnis der HRD-Biologie und kann die Identifizierung von Patienten verbessern, die am ehesten von zielgerichteten Therapien profitieren.

Abstract

Homologous recombination deficiency (HRD) is a key molecular hallmark that predicts response to PARP inhibitors and platinum-based chemotherapy in PCa. Nevertheless, routine diagnostic strategies, which mostly rely on identifying *BRCA1* or *BRCA2* mutations or on genomic scar assays from tissue biopsies, fail to capture the full spectrum of HRD-related alterations. In addition, biopsy-based testing is often limited by intratumoral heterogeneity and by the difficulty of obtaining tissue from patients with metastatic bone disease. The work presented here sought to establish a minimally invasive and comprehensive framework for HRD characterization using circulating tumor DNA (ctDNA) from plasma samples.

Plasma-derived ctDNA from 106 men with metastatic PCa was examined using a multimodal analytical pipeline that combined targeted sequencing of genes involved in homologous recombination repair (HRR), shallow whole-genome sequencing (sWGS) to quantify the Genomic Instability Score (GIS), matched tumor-normal whole-exome sequencing for mutational signature profiling, and fragmentomic assessment of cfDNA fragmentation and chromatin accessibility patterns.

Among HRR genes, *BRCA2* alterations were most frequent and commonly co-occurred with *PTEN* loss. Elevated GIS correlated with combined *BRCA2* and *RB1* loss, a high burden of somatic copy-number changes, and reduced overall survival. Tumors with HRD exhibited enrichment of the characteristic mutational signatures SBS3 and ID6, a relative increase in dinucleosome-sized cfDNA fragments, and decreased accessibility at zinc-finger transcription factor binding regions, indicating altered nucleosome organization and chromatin structure. Using these features, a machine-learning classifier based on fragmentomic data accurately distinguished HRD cases.

Taken together, these results demonstrate that multimodal ctDNA profiling can serve as a powerful, noninvasive tool to detect HRD in PCa. The integration of mutational, structural, and fragmentomic information provides a more comprehensive and functional picture of HRD biology and may improve the identification of patients who are most likely to benefit from targeted therapies.

1 Introduction

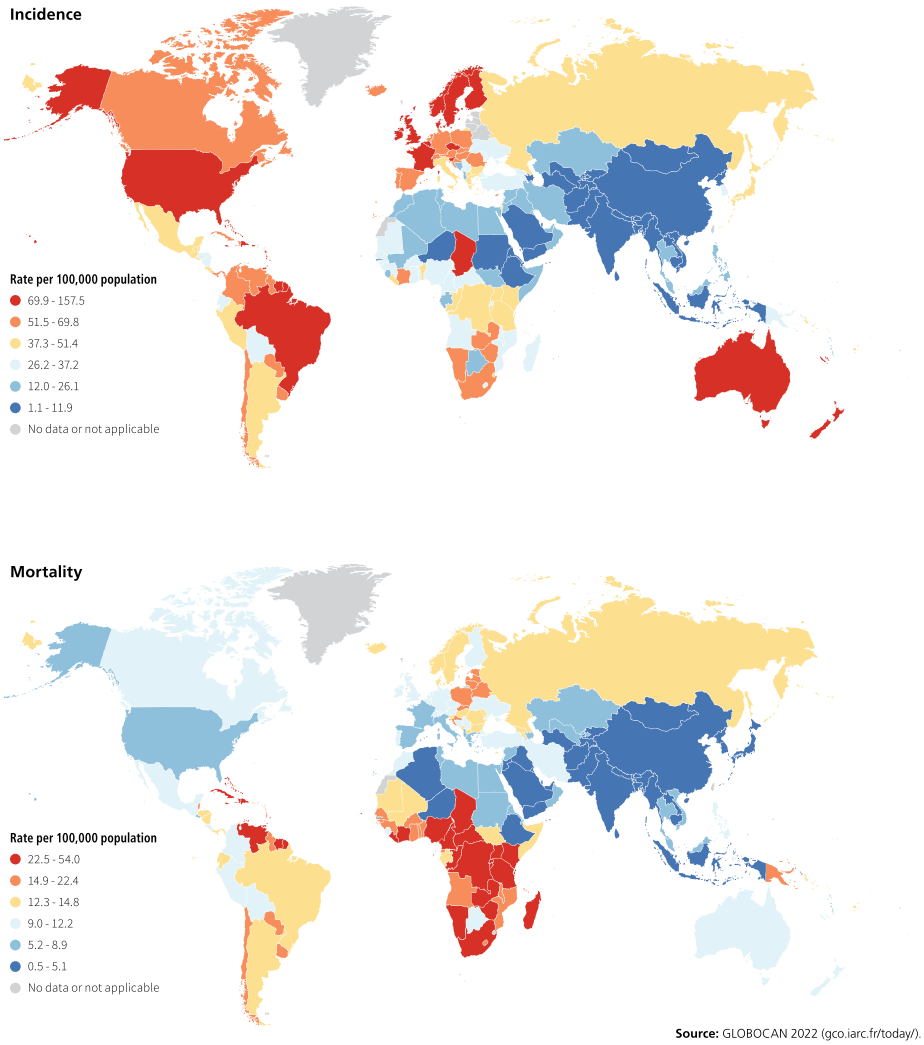
1.1 Prostate cancer epidemiology

Prostate cancer (PCa) is one of the most common malignancies in males and a major contributor to cancer-related mortality worldwide (1). In the United States alone, approximately 299,010 new cases and 35,250 deaths are projected in 2024, reflecting its substantial public health burden. Globally, incidence is expected to rise from 1.4 million new cases in 2020 to 2.4 million by 2040 due to aging populations.

Incidence and mortality vary widely across regions, influenced by differences in healthcare access, screening practices, and population genetics **(Figure 1)**. Incidence is more than 50-fold higher in regions with widespread PSA testing, such as Northern Europe and the Americas, compared to parts of Asia and the Middle East. However, mortality is highest in sub-Saharan Africa and the Caribbean, where limited access to early detection and treatment exacerbates outcomes. Men of Western African ancestry also face higher biological risk, while migration studies highlight the interplay between genetic and environmental factors in shaping disease susceptibility.

Despite the high global burden of PCa, mortality trends do not always mirror incidence patterns. Countries with well-developed healthcare systems, such as the United States and Northern Europe, demonstrate relatively low mortality despite high incidence rates, reflecting the impact of early detection and effective treatment strategies. In contrast, countries in sub-Saharan Africa and the Caribbean continue to experience disproportionately high mortality, largely due to limited healthcare infrastructure, delayed diagnoses, and restricted access to curative therapies (2).

Environmental and epigenetic factors also contribute to regional disparities in PCa risk and outcomes. Migration studies, for example, show that individuals relocating from low-incidence regions to higher-incidence settings experience an increased risk of developing PCa, though typically not to the same level as native populations. This observation underscores the complex interaction between inherited predisposition and environmental exposures in shaping disease risk (3).



Source: GLOBOCAN 2022 (gco.iarc.fr/today/).

Figure 1: Global distribution of PCa incidence and mortality (age-standardised rates, 2022).

The top panel shows the incidence rates of PCa per 100,000 men, highlighting high-incidence regions such as North America, Northern and Western Europe, and Oceania. These patterns are largely influenced by widespread prostate-specific antigen (PSA) testing and increased healthcare access. The bottom panel depicts mortality, which is highest in sub-Saharan Africa and parts of the Caribbean, reflecting disparities in early detection and treatment availability. In contrast, mortality rates remain relatively low in high-incidence but resource-rich settings. Together, these maps underscore the global heterogeneity in PCa burden, shaped by both biological and socioeconomic factors. Data adapted from GLOBOCAN 2022; maps were created using MapChart.net and are shared under the Creative Commons Attribution–ShareAlike 4.0 (CC BY-SA 4.0) license.

1.2 Histopathology of PCa

PCa assessment relies primarily on histopathological evaluation of tissue obtained through transrectal or transperineal needle biopsy. This remains the gold standard for diagnosis and provides critical information regarding tumor grade, extent, and potential aggressiveness. Following biopsy, hematoxylin and eosin (H&E) staining is used to evaluate glandular architecture and identify malignant features, including loss of normal gland structure, prominent nucleoli, and invasive growth. The degree of differentiation is quantified using the Gleason grading system **(Figure 2)**. Developed to enhance accuracy and clinical relevance, the system evaluates the architectural patterns and degree of differentiation of malignant prostate cells (4). The two most common histologic patterns observed in a tumor are graded on a scale of 1 to 5, and their scores are summed to generate the Gleason score, ranging from 2 to 10. Higher scores indicate more aggressive disease and poorer outcomes. Importantly, tumors with the same score can behave differently depending on which pattern is dominant; for example, a Gleason score of 3 + 4 = 7 generally carries a more favorable prognosis than 4 + 3 = 7, reflecting the greater proportion of higher-grade pattern 4 disease (4,5). Heterogeneity within the tumor can result in different biopsy cores receiving different grades. In such cases, the highest Gleason score is typically reported for risk stratification, although some variability in clinical practice persists (6).

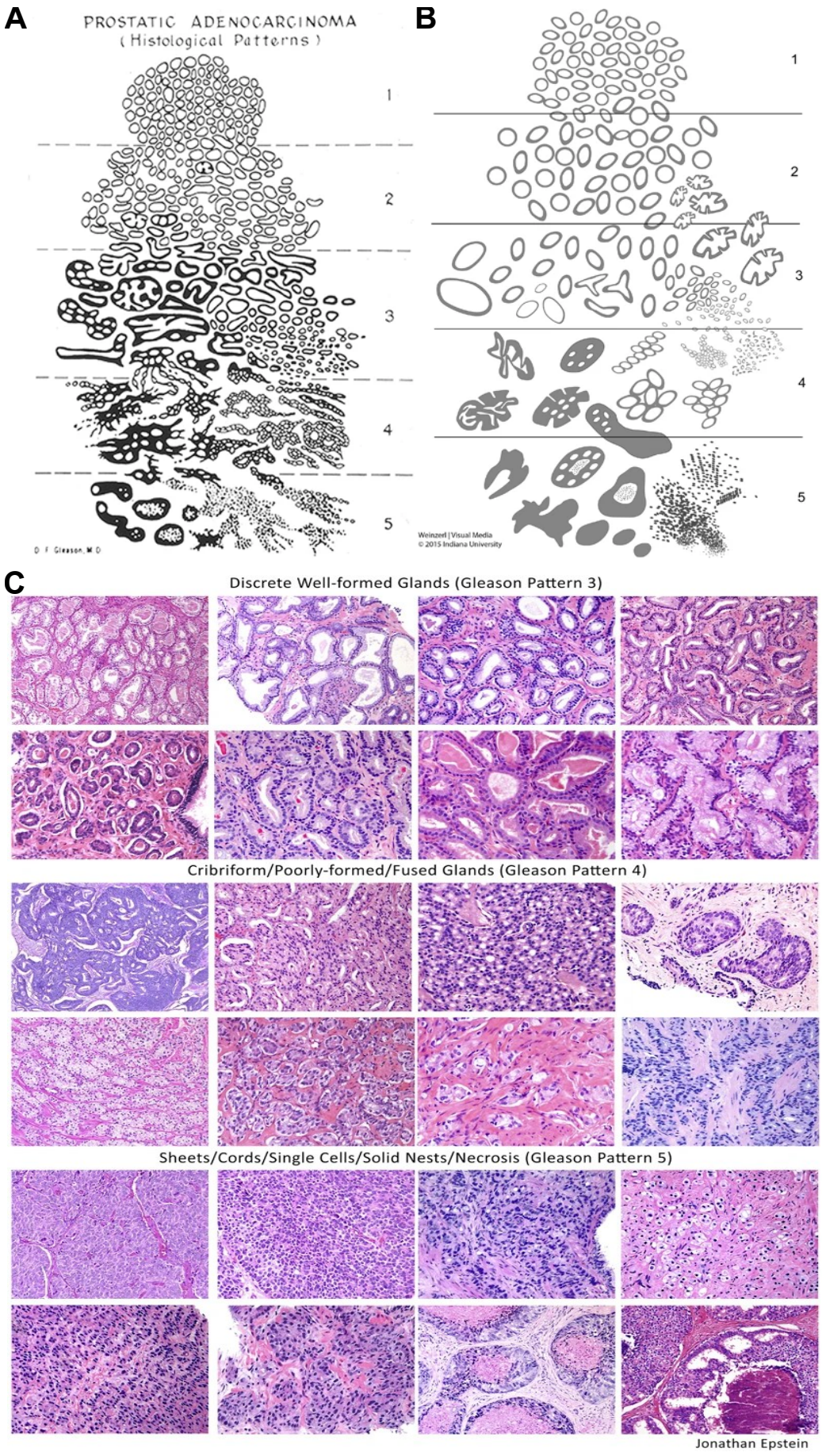

A
PROSTATIC ADENOCARCINOMA
(Histological Patterns)
1
2
3
4
5
D. F. Gleason, M.D.
B
1
2
3
4
5
Weinzerl | Visual Media
© 2015 Indiana University
C
Discrete Well-formed Glands (Gleason Pattern 3)
Cribriform/Poorly-formed/Fused Glands (Gleason Pattern 4)
Sheets/Cords/Single Cells/Solid Nests/Necrosis (Gleason Pattern 5)
Jonathan Epstein

Figure 2: Original (A) and modified (B) Gleason grading diagrams, with representative histologic examples of Gleason patterns in panel (C).

The Gleason grading system classifies prostate adenocarcinoma based on glandular architecture and degree of tumor differentiation, assigning patterns from 1 (well-formed glands, well-differentiated) to 5 (minimal or no glandular formation, poorly differentiated). Higher Gleason patterns correspond to increasingly aggressive, poorly differentiated tumors. Panels A, C adapted under the terms of the Creative Commons CC BY 4.0 license (6). Panel B is reproduced with permission from Elsevier (RightsLink license no. 6180811109578) (7).

1.3 PCa genetics

PCa development is driven by the accumulation of genetic and epigenetic alterations that disrupt normal cellular growth, differentiation, and survival. The genetic landscape of PCa is complex, adding layers of complexity to disease prediction and prognosis (7). While most PCa are sporadic, inherited genetic susceptibility plays a significant role in approximately 5–10% of cases, with familial clustering observed in up to 20% of patients (8–11). High-penetrance germline mutations are relatively rare, but among these, *BRCA2* is the most significant predisposition gene, associated with increased incidence, earlier onset, and poor clinical outcomes (12–14). Inherited or somatic alterations in DNA damage repair (DDR) genes - such as *ATM, CHEK2, BRCA1*, and mismatch repair genes (MMR) including *MSH2* - further contribute to aggressive disease behavior. These biological vulnerabilities have enabled the adoption of precision therapies, including immune checkpoint inhibitors (ICI) for microsatellite-instability (MSI)–high tumors and PARP inhibitors (PARPi) for homologous recombination–deficient (HRD) disease (15,16).

PCa is characterized by marked spatial, cellular, and genetic heterogeneity (17). The most prevalent form is acinar adenocarcinoma, arising from premalignant epithelial lesions that evolve through key molecular alterations such as *TP53* and *PTEN* loss, *TMPRSS2–ETS* gene fusions, and *MYC* amplification (18). *ETS* gene fusions, most notably *TMPRSS2–ERG*, are present in ~50% of PCas and drive aberrant transcriptional programs.

Early tumor growth is largely dependent on androgen receptor (AR) signaling, and most patients initially present with hormone-sensitive PCa (HSPC) that responds to androgen deprivation therapy (ADT). However, AR signaling is often amplified or mutated in advanced disease, promoting continued tumor growth despite androgen deprivation. Therefore, over time, selective pressure from AR-targeted therapies leads to the development of castration-resistant PCa (CRPC), a more aggressive and potentially lethal state (18–20) **(Figure 3)**. The presence of rare histologic subtypes - such as ductal adenocarcinoma, basal-cell carcinoma, and neuroendocrine PCa (NEPC) - further reflects the disease's biological diversity (21).

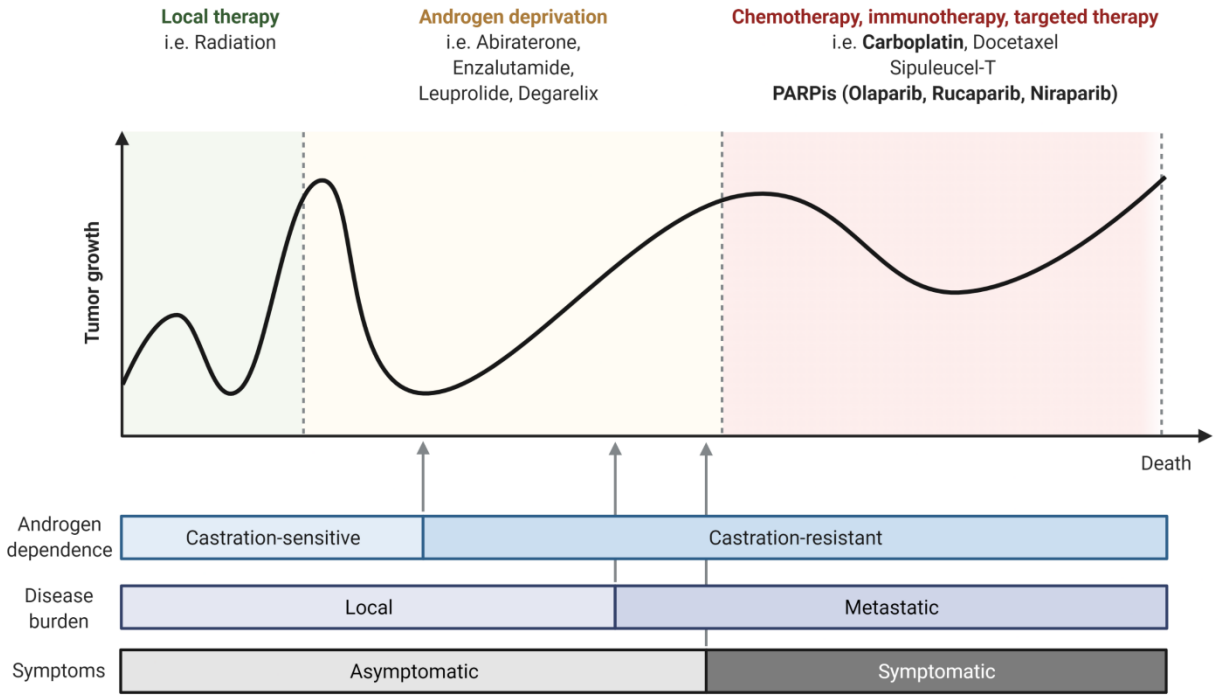


Figure 3: Schematic overview of PCa progression and corresponding therapeutic interventions.

PCa typically progresses from localized disease (confined to the prostate gland) to locally advanced and ultimately metastatic castration-resistant PCa (mCRPC). Standard treatments for localized stages include active surveillance, radical prostatectomy, and radiotherapy. In advanced stages, androgen deprivation therapy (ADT) is commonly used, often followed by next-generation hormonal therapies, chemotherapy, or radiopharmaceuticals in mCRPC. Molecular stratification, including testing for homologous recombination deficiency (HRD) and other biomarkers, is increasingly integrated into clinical decision-making to guide precision therapies such as poly (adenosine diphosphate-ribose) polymerase inhibitors (PARPi). Created with BioRender.

Somatic genetic alterations acquired during tumor evolution are diverse and include point mutations, somatic copy number alterations (SCNA), genomic rearrangements, and chromosomal instability. Large-scale sequencing efforts, such as The Cancer Genome Atlas (TCGA) and the International PCa Genomics Consortium, have substantially refined molecular classification frameworks and provided critical insights into tumor evolution, intratumoral heterogeneity, and therapeutic vulnerabilities (22–24). Among recurrent genomic events in PCa, *TP53*, *PTEN*, and *RB1* stand out as frequently inactivated tumor suppressor genes and represent hallmarks of the disease's genomic landscape. Alterations in these genes are already detectable in primary prostate tumors but become significantly more prevalent in advanced disease stages, with marked enrichment in metastatic castration-resistant PCa (mCRPC) (25). Large-scale sequencing studies indicate that TP53 is mutated in approximately 50% of mCRPC tumors, *PTEN* is functionally lost in roughly 40-50%, and *RB1* is inactivated in approximately 10–20%, frequencies that are substantially higher than those observed in localized disease (26). Clinically, the inactivation of these tumor suppressors is strongly associated with aggressive tumor behavior and poor

patient outcomes. Combined loss of *TP53, PTEN*, and *RB1* is frequently observed in the most aggressive PCa subtypes and correlates with early disease progression, treatment resistance, and reduced survival (25). Specifically, *PTEN* deletions have been linked to increased risks of recurrence and metastasis, *TP53* mutations are associated with resistance to androgen receptor-targeted therapies, and *RB1* loss is a defining feature of therapy-related neuroendocrine PCa, an AR-independent and highly lethal variant of the disease (27–29). Together, these findings underscore the crucial role of *TP53*, *PTEN*, and *RB1* alterations in driving PCa progression and lethal disease, especially in the context of mCRPC.

In summary, PCa evolves from an initially androgen-driven disease into a genomically unstable and treatment-resistant malignancy. A deeper understanding of this molecular evolution is essential to guide biomarker development and the implementation of personalized therapeutic strategies.

1.4 Homologous recombination repair deficiency

Homologous recombination repair (HRR) is a high-fidelity DNA double-strand break (DSB) repair pathway essential for maintaining genomic stability. Homologous recombination (HR) is a genetic process whereby two homologous DNA molecules exchange genetic information, and HRR exploits this mechanism to accurately repair DNA damage, primarily DSB (30). If HRR is intact, DSBs are repaired accurately and the cell cycle continues normally. In contrast, HRR deficiency forces cells to rely on alternative, error-prone repair mechanisms (31), particularly poly(ADP-ribose) polymerase (PARP)-mediated single-strand break repair. This results in the accumulation of genomic alterations that can promote tumorigenesis (32). More specifically, HRD leads to increased genomic instability score (GIS), accumulation of chromosomal aberrations, and accelerated tumor evolution.

Pharmacological inhibition of PARP in HRD tumors exploits the principle of synthetic lethality, whereby the simultaneous loss of homologous recombination repair and PARP-dependent repair leads to the accumulation of unrepaired DNA damage, replication fork collapse, and ultimately tumor-selective cell death. This therapeutic vulnerability forms the biological basis for the clinical efficacy of PARPi in HRD-positive cancers (33). Furthermore, recent evidence suggests that the HRD phenotype could be used to predict response to pan-cancer patients' responses to Programmed cell death protein 1 immunotherapy (34). Recent reports have highlighted that the difference of efficacy between PARPi and the emergence of resistance are not explained adequately by the DSB hypothesis. PARP trapping (35) and replication gaps (36) have been proposed to explain this discrepancy. Recent advances in massive parallel sequencing have provided the opportunity to predict the occurrence of the HRD phenotype by examining the underlying cause in the form of pathogenic genomic alterations in the genes encoding proteins involved in the pathway. Furthermore, the results of the phenotype in the

genome can be observed in the form of GIS and mutational signatures (37–40). Collectively, the increasing prevalence of HRD in advanced disease, its prognostic significance, and its role as a predictive biomarker reinforce the importance of developing accurate, minimally invasive methods to detect these alterations - particularly as tumors evolve under therapeutic pressure.

1.4.1 HRD in PCa

In PCa, HRD is primarily driven by pathogenic alterations in key HRR genes, most notably *BRCA2*, which is the most frequently disrupted HRR gene in PCa and strongly associated with aggressive clinical behavior (17–19). Other commonly altered genes include *BRCA1, ATM, CHEK2, PALB2*, as well as components of the Fanconi anemia network. Goodall et al. reported that 20-30% of lethal PCa harbor deletions in genes related to HRD including *BRCA2*, *BRCA1*, *ATM*, *PALB2*, *FANCA*, *CHEK2*, and *CDK12* (41).

In HSPC, HRR gene mutations are relatively uncommon, reported in approximately 8–12% of tumors (30). However, the prevalence increases markedly with disease progression, with up to 25–30% of mCRPC harboring deleterious HRR alterations (42,43). This enrichment reflects strong selective pressure favoring tumor clones capable of sustaining GIS and adapting to therapy. HR-deficient tumors typically exhibit characteristic genomic scars, including loss of heterozygosity (LOH), large-scale state transitions, and telomeric allelic imbalance (44), which can be quantified through composite HRD scoring methods developed primarily in breast and ovarian cancer (45). However, the clinical utility of these assays in PCa remains under active investigation, as HRD biology appears more complex and less uniformly driven by *BRCA*-associated mechanisms compared with other tumor types (46,47). Nevertheless, PCa with HRD demonstrated increased sensitivity to PARPi, leading to regulatory approval for mCRPC patients harboring actionable HRR mutations (48) **(Figure 4)**. Yet, response variability indicates that improved biomarkers are needed to identify patients most likely to benefit from HRD-targeted therapies.

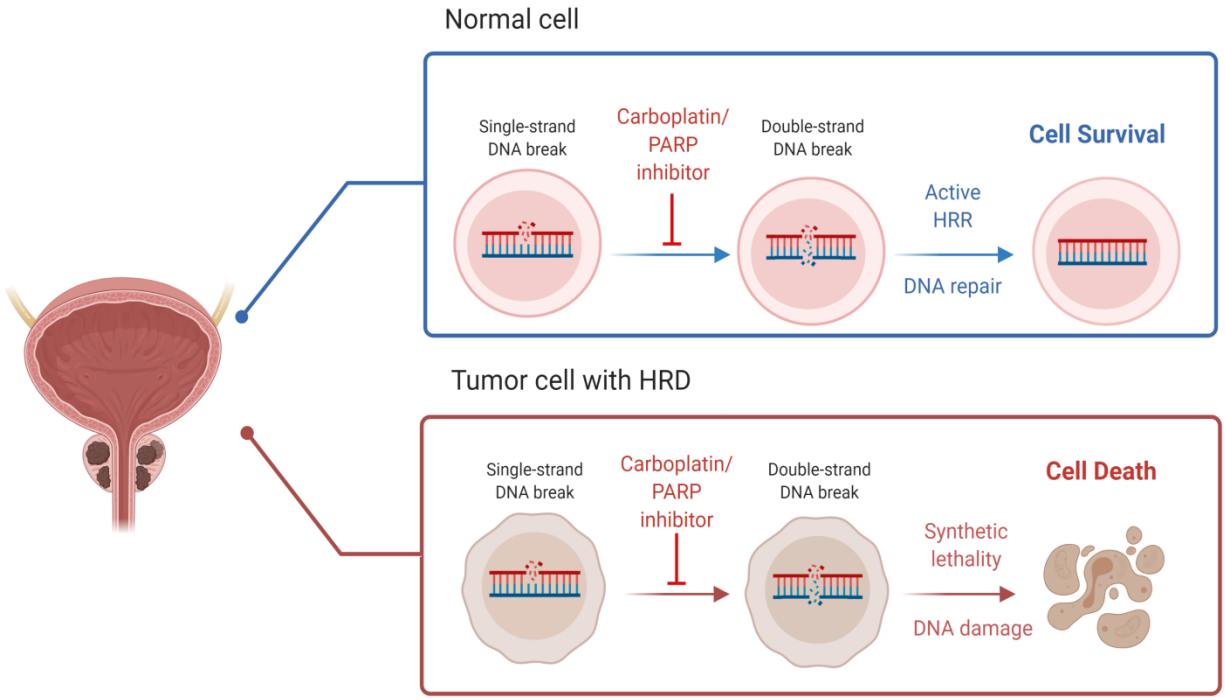


Figure 4: Synthetic lethality in homologous recombination-deficient (HRD) tumors treated with PARP or platinum-based agents.

In normal cells, DNA damage from single-strand breaks is effectively repaired by homologous recombination repair (HRR), allowing cell survival despite PARP inhibition. In contrast, HRD tumor cells lack functional HRR, and treatment with PARP inhibitors or carboplatin induces accumulation of DNA damage, leading to double-strand breaks and cell death via synthetic lethality. Created with BioRender

1.5 Overview of the HRR mechanism

HRR is a multistep, high-fidelity DNA repair pathway that operates primarily during the S and G2 phases of the cell cycle, when a sister chromatid is available as a homologous template. HRR uses HR to repair DSBs, which are among the most mutagenic and cytotoxic forms of DNA damage. DSBs can arise from various sources, including ionizing radiation, reactive oxygen species, mechanical stress, and the failure to repair single-strand breaks (SSBs) **(Figure 5)** (49).

In general, cells rely on two major pathways to repair DSBs: HRR, which uses a sister chromatid as a template and therefore is restricted to S/G2 phases, and non-homologous end-joining (NHEJ), which is faster but error-prone. The latter category includes canonical NHEJ (cNHEJ), alternative NHEJ (aNHEJ), and single-strand annealing (SSA) (31). When HRR is impaired, reliance on NHEJ leads to GIS through chromosomal rearrangements, loss of tumor suppressors, and oncogene amplification.

HRD is further associated with catastrophic events such as chromothripsis and clustered hypermutation (50,51).

DSB recognition and signaling constitute the first step of HRR. DSBs are sensed primarily by the MRN complex (MRE11–RAD50–NBS1) (52,53), which recruits and activates DNA damage signaling kinases, most notably ATM and ATR. These kinases initiate checkpoint signaling through downstream effectors such as CHK1 and CHK2, inducing cell-cycle arrest and coordinating DNA repair (31).

Following damage recognition, DNA end resection is initiated. Short-range resection by the MRN complex and CtIP generates 3′ single-stranded DNA (ssDNA) overhangs, which are subsequently extended by long-range resection enzymes such as EXO1 and DNA2, assisted by helicases including BLM or WRN. Resection suppresses non-homologous end joining and commits the break to HRR (54). However, recent evidence suggests some resection activity may also occur during the cNHEJ pathway, warranting further research to fully contextualize its relevance to cancer manifestation and treatment (55).

The exposed ssDNA is rapidly coated by replication protein A (RPA), which stabilizes the DNA and prevents secondary structure formation (56). In the next step, RPA is replaced by RAD51, forming the presynaptic nucleoprotein filament. This exchange is mediated by BRCA2, in cooperation with PALB2 and the BRCA1–BARD1 complex, which are essential for efficient RAD51 loading and filament stabilization (57,58). The RAD51 filament conducts the homology search and strand invasion, probing duplex DNA for homologous sequences and forming a displacement loop (D-loop) (59,60). Accessory proteins such as RAD54 facilitate chromatin remodeling, stabilize strand invasion, and promote DNA synthesis.

In the postsynaptic phase, DNA synthesis is initiated from the invading 3′ end, followed by resolution of recombination intermediates. HRR can proceed through multiple subpathways, including synthesis-dependent strand annealing (SDSA), canonical double-strand break repair (DSBR), or break-induced replication (BIR), yielding crossover or non-crossover products (61,62). Proper resolution, primarily via dissolution mechanisms involving the BLM–TOP3α–RMI complex, is critical to preserve genomic integrity.

Defects at any stage of this tightly regulated process result in HRD, characterized by GIS, accumulation of structural alterations, and increased reliance on compensatory DNA repair pathways.

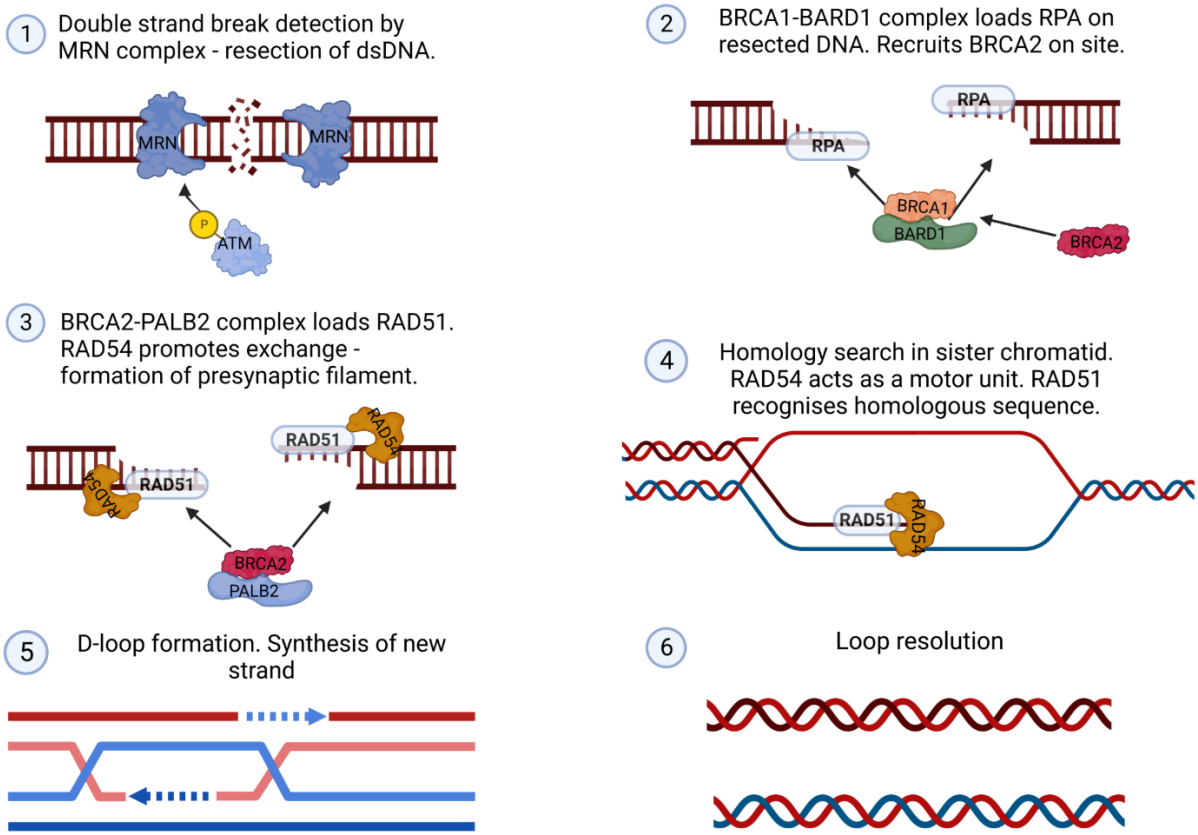


Figure 5: Mechanistic overview of homologous recombination repair (HRR) following double-strand DNA breaks.

(1) The MRN complex detects the break and initiates DNA end resection. (2) RPA binds single-stranded DNA, while the BRCA1-BARD1 complex facilitates BRCA2 recruitment. (3) BRCA2 and PALB2 load RAD51 onto ssDNA to form the presynaptic filament, aided by RAD54. (4,5) RAD51 mediates homology search and strand invasion into the sister chromatid, followed by D-loop formation and DNA synthesis (6). The repair is completed through resolution of recombination intermediates. This high-fidelity process is crucial for genome stability and its disruption defines the HRD phenotype. Created with BioRender.

1.6 HRD biomarkers and detection strategies

As the clinical impact of HRD grows - particularly in enabling PARPi therapy - there is a critical need for accurate and accessible biomarkers. Multiple genomic, epigenomic, and functional assays are now used to detect HRD, although their performance varies significantly across tumor types, especially in PCa and liquid biopsy settings.

1.6.1 *BRCA1* and *BRCA2*: From discovery to HRD biomarkers

The discovery of *BRCA1* and *BRCA2* marked a pivotal moment in cancer genetics, particularly in the understanding of hereditary breast and ovarian cancer. *BRCA1* was mapped to chromosome 17q21 in 1990 and cloned in 1994, followed by identification of *BRCA2* on 13q12–13 in 1995 (63–65).

BRCA1 and *BRCA2* encode tumor suppressor proteins that are essential for the maintenance of genomic stability, primarily through their roles in the HRR pathway. *BRCA1* acts early in the HRR cascade by sensing DSB and facilitating the end-resection process, whereas BRCA2 plays a direct role in loading RAD51 onto single-stranded DNA to promote strand invasion and repair (66,67).

Germline mutations in *BRCA1* and *BRCA2* confer a significantly increased lifetime risk of developing breast, ovarian, prostate, and pancreatic cancers. Women carrying pathogenic variants in *BRCA1* face a lifetime breast cancer risk of approximately 60–70% and an ovarian cancer risk of up to 40%, while *BRCA2* mutations confer risks of approximately 45% and 20%, respectively (68).

Loss of BRCA function causes HRD, forcing reliance on error-prone repair pathways and driving GIS. Importantly, this deficiency forms the biological rationale for the use of PARPi, which exploit the concept of synthetic lethality. PARPis inhibit the base excision repair pathway, and in HRD-positive cells-particularly those harboring BRCA1/2 mutations-this leads to the accumulation of DNA damage and subsequent cell death. This therapeutic vulnerability has led to the approval of PARPi in *BRCA*-mutant breast, ovarian, pancreatic, and PCas (15).

Despite the strong predictive value of *BRCA1/2* loss-of-function mutations for PARP inhibitor sensitivity, resistance can emerge through secondary or "reversion" mutations. These mutations restore the open reading frame of *BRCA* genes-particularly *BRCA2*, thereby re-establishing functional HRR and negating synthetic lethality. Such reversions are a well-characterized mechanism of acquired resistance to both PARPi and platinum-based chemotherapy (41).

1.6.2 Non-*BRCA* HRR genes: Expanding the HRD landscape

Mutations in other HRR genes, such as *ATM, PALB2, RAD51C/D, CHEK2, BARD1,* and *FANCA,* can also impair HR and predict PARPi or platinum sensitivity in some settings (34). For example, *PALB2* inactivation has shown PARPi response rates approaching BRCA-mutated tumors (69), and RAD51C/D alterations demonstrated benefit in subgroups of the PROfound trial (70).

However, not all HRR gene mutations produce comparable functional effects. *ATM*-mutated PCa often lacks classical HRD features and responds poorly to PARPi (70,71). *CDK12*-mutated tumors, although associated with GIS, do not typically confer PARPi sensitivity and may instead require immunotherapy-based approaches (72).

The divergent responses observed across non-BRCA HRR genes highlight a broader challenge in HRD biomarker development: not all mutations confer equivalent functional deficiency. Differences in mutation type (e.g., monoallelic vs. biallelic, pathogenicity status), gene function within the HRR pathway, and tumor lineage all influence the penetrance of HRD and subsequent therapeutic response.

As such, a more refined stratification of HRR mutations is needed, based on functional assays or HRD-associated genomic signatures rather than relying solely on gene-level mutation status. This could include integrating mutational signatures (e.g., SBS3, ID6), GIS, other genomic and epigenomic features (methylation, fragment profiling) and functional assays such as RAD51 foci formation, as potential complementary biomarkers.

These variable responses emphasize that gene mutation status alone is insufficient to define HRD in PCa - supporting the integration of functional or GIS-based biomarkers.

1.6.3 *BRCA1* promoter methylation and epigenetic HRD

In addition to genetic mutations, epigenetic silencing of HRR genes, most notably *BRCA1*, represents a key alternative mechanism leading to HRD. Promoter hypermethylation of *BRCA1* suppresses gene transcription without altering the DNA sequence. Tumors with *BRCA1* promoter methylation often exhibit similar features to *BRCA1*-mutant cancers, including GIS, HRD-specific mutational signatures and sensitivity to PARPi and platinum agents (73,74). This functional mimicry has led to the classification of *BRCA1*-methylated tumors within the broader "BRCAness" phenotype (75).

BRCA1 promoter methylation has been robustly associated with HRD phenotypes in breast and ovarian cancers, including high GIS and PARPi responsiveness (76). Importantly, BRCA1 methylation has also been detected in circulating tumor DNA (ctDNA), suggesting potential utility in non-invasive liquid biopsy assays (77). However, intratumoral heterogeneity and the potential reversibility of methylation under therapeutic pressure may restore BRCA1 expression and drive acquired resistance, limiting its reliability as a standalone biomarker.

1.6.4 GIS and the HRD “scar”

HRD gives rise to characteristic patterns of GIS, as cells increasingly rely on error-prone DNA repair mechanisms in the absence of effective HR. This GIS accumulates over time, leaving a detectable footprint known as a “genomic scar.” Three key features of this scarring have emerged as quantifiable markers of HRD: LOH, telomeric allelic imbalance (TAI), and large-scale state transitions (LST) **(Figure 6)** (37–39).

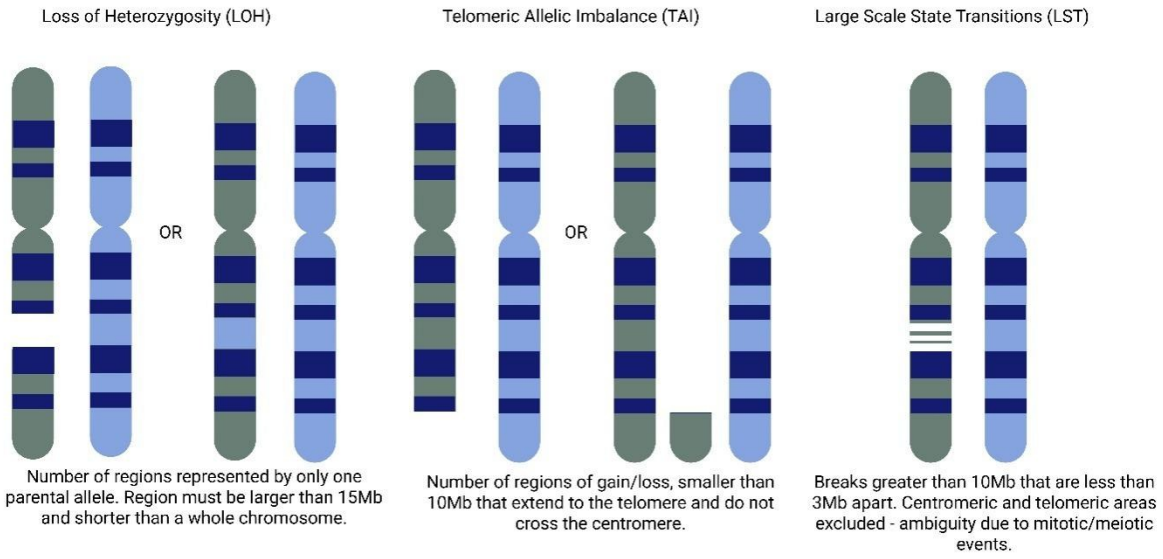


Figure 6: Genomic scar components contributing to homologous recombination deficiency (HRD) scores.

Schematic representation of the three key metrics used to quantify genomic scarring: Loss of heterozygosity (LOH): regions larger than 15 Mb and shorter than a full chromosome represented by only one parental allele; Telomeric allelic imbalance (TAI): allelic imbalances extending to the telomere but not crossing the centromere; and Large-scale state transitions (LST): chromosomal breaks greater than 10 Mb and less than 3 Mb apart, excluding centromeric and telomeric regions. Created with BioRender.

LOH refers to chromosomal regions in which one parental allele is lost, typically through deletions or unbalanced recombination events, and is a frequent consequence of defective homologous recombination repair leading to tumor suppressor gene inactivation. TAI describes subchromosomal regions with allelic imbalance that extend to the telomeres but do not cross centromeres, reflecting errors during replication and segregation that are hallmarks of HRD. LST represent chromosomal breakpoints between adjacent regions of at least 10 megabases, and are thought to arise from misrepair of double-strand breaks. Each of these features individually captures a distinct aspect of genomic disruption, but when combined, they offer a more robust reflection of HRD than any single marker alone.

These metrics are often aggregated into a composite GIS, a concept validated in multiple clinical contexts, including triple-negative breast cancer and high-grade serous ovarian carcinoma (38,39). A GIS threshold (e.g., ≥42) is frequently used to designate HRD positivity, and has been incorporated into assays such as Myriad's myChoice® CDx and FoundationOne's CDx HRD panel. Studies have shown that tumors with high GIS are more likely to respond to PARPi and platinum-based chemotherapy, even in the absence of *BRCA1/2* mutations (37).

However, it is important to recognize the limitations of genomic scar-based assays. These scores reflect a tumor's historical HRD status but cannot capture real-time restoration of homologous recombination function. For instance, tumors that acquire secondary (reversion) mutations restoring BRCA2 function may regain DNA repair capacity and become resistant to therapy, despite retaining high GIS. Consequently, tumors may be misclassified as HRD-positive based on their scar profile, even though they have functionally reverted to an HR-proficient state (41). This underscores the importance of integrating scar-based metrics with dynamic biomarkers, such as appropriate liquid biopsy testing or functional assays, for accurate therapeutic stratification.

1.6.5 Mutational signatures associated with HRD

Mutational signatures are reproducible patterns of somatic mutations reflecting specific DNA damage and repair processes during tumor evolution. Defined by characteristic combinations of base substitutions, insertion–deletions, and structural or copy number alterations, these patterns are shaped by endogenous defects and exogenous exposures. Large-scale efforts such as TCGA and COSMIC have systematically catalogued mutational signatures across diverse cancer types (78). These signatures are typically derived through matched tumor-normal whole-genome or whole-exome sequencing using deconvolution algorithms that match observed mutational patterns to reference catalogues, such as COSMIC.

In the context of HRD, the most well-characterized somatic signatures are Single Base Substitution Signature 3 (SBS3) and Insertion-deletion Signature 6 (ID6), both strongly associated with *BRCA1/2* inactivation. SBS3 is characterized by an elevated frequency of T>G substitutions in a specific trinucleotide context. ID6 reflects microhomology-mediated deletions, consistent with reliance on error-prone repair pathways in the absence of effective HRR (40). These signatures have been associated with improved response to platinum-based chemotherapy and PARPi in breast cancer, ovarian cancer, and PCa (79). Importantly, mutational signatures reflect a tumor's evolutionary history and provide orthogonal evidence of HRD that can complement structural or gene-based assays.

While mutational signatures provide complementary, evolutionary evidence of HRD beyond gene- or scar-based assays, their detection depends on adequate sequencing depth, tumor purity, and germline filtering. Moreover, they reflect historical rather than real-time HRD status. Nonetheless, when applied appropriately, mutational signature analysis remains a valuable tool for HRD classification, particularly when *BRCA1/2* alterations or genomic scars are absent or ambiguous.

1.6.6 Functional assays for HRD detection

Functional assays are pivotal in directly assessing HR activity within tumors, thereby identifying HRD beyond what genomic analyses can reveal. These assays evaluate the tumor's current DNA repair capacity, offering real-time insights into HR functionality (80).

The most widely used functional assays for HRD detection is the RAD51 foci formation test, which measures the tumor's ability to form RAD51 protein foci at sites of DNA double-strand breaks, a critical intermediate step in the HR repair pathway. Tumors with functional HR repair exhibit robust RAD51 foci formation in proliferating cells following DNA damage, while HR-deficient tumors show reduced or absent foci. The REpair CAPacity (RECAP) test is a standardized implementation of this assay, typically performed on fresh tumor samples. It involves ex vivo irradiation of tumor tissue to induce double-strand breaks, followed by immunofluorescent detection of RAD51 foci in S/G2 phase cells. This controlled induction of DNA damage allows for consistent evaluation of HR functionality and has shown strong predictive value for PARPi sensitivity across multiple cancer types (81,82).

Additionally, a virus-based functional assay has been developed to quantify real-time HR activity directly. This method establishes an activity threshold for identifying HRD across various cancer types and has been validated by correlating with PARPi sensitivity and *BRCA* status (80).

While these functional assays provide valuable, real-time assessments of HR competency, their clinical implementation faces challenges. Standardization across laboratories, the need for viable tumor tissue, and integration with existing diagnostic workflows are areas requiring further development. Nonetheless, functional assays hold significant promise for enhancing personalized treatment strategies by accurately identifying patients who may benefit from targeted therapies like PARPi.

1.7 Clinically approved HRD assays

Several commercial assays are currently available for the assessment of HRD, with two of them receiving FDA approval for specific clinical indications. The most widely used and clinically validated test is myChoice® CDx (Myriad Genetics) (83), an FDA-approved companion diagnostic for PARPi use in ovarian cancer. This test integrates three genomic scar metrics (LOH, TAI, and LST) into a composite GIS and additionally assesses *BRCA1/2* mutation status.

Another clinically approved assay is FoundationOne® CDx (Foundation Medicine) (84), which utilizes NGS to provide comprehensive genomic profiling. It reports *BRCA1/2* mutations and other HRR gene alterations, along with a genomic LOH score. FoundationOne® CDx is FDA-approved as a companion diagnostic for rucaparib in ovarian cancer, and for olaparib in both breast and PCas. Its cfDNA-based counterpart, FoundationOne® Liquid CDx, is also FDA-approved for the detection of *BRCA1/2* and

other HRR alterations in metastatic PCa, enabling non-invasive testing when tissue biopsy is not feasible (85). While its clinical utility is well established, limitations remain in detecting certain genomic alterations such as homozygous deletions from cfDNA.

In contrast, BRACAnalysis CDx® (Myriad Genetics) is a germline *BRCA1/2* test based on Sanger sequencing and multiplex PCR performed on DNA extracted from whole blood (not plasma cfDNA) (86). Although FDA-approved for identifying patients eligible for PARPi therapy across multiple cancers including breast, ovarian, pancreatic, and prostate, this assay detects only germline *BRCA1/2* mutations and is not designed to capture somatic alterations in tumors (86). Therefore, a negative result does not exclude the presence of a tumor-specific mutation, which limits its standalone utility in mCRPC.

Additional platforms, including Tempus xT/xR and Caris Molecular Intelligence®, assess BRCA1/2 and other HRR genes using targeted or exome sequencing and are widely used as laboratory-developed tests within CLIA-certified laboratories, despite lacking FDA approval (87,88).

Most currently available HRD assays rely on FFPE tumor tissue, which poses a significant limitation in metastatic PCa due to biopsy challenges, particularly in bone metastases where decalcification degrades nucleic acid quality. Tumor heterogeneity further complicates tissue-based testing, underscoring the need for robust liquid biopsy–compatible HRD assays (89). This clinical reality highlights the urgent need for liquid biopsy-compatible HRD tests.

1.8 Computational techniques for HRD assessment

1.8.1 Current computational pipelines for HRD detection

A broad array of computational tools and algorithms has been developed to identify HRD based on genomic features extracted from tumor sequencing data. These methods predominantly focus on somatic GIS patterns, such as the aforementioned mutational signatures (SBS3, ID6) and large-scale copy number alterations such as the genomic scars (LOH, TAI, LST).

Among the most widely used HRD prediction tools is HRDetect, a supervised lasso logistic regression model trained on whole-genome sequencing WGS data from *BRCA1/2*-deficient breast cancers. HRDetect integrates six mutational features, including substitution and rearrangement signatures, microhomology-mediated indels, and copy number–based GIS, and can identify HRD even in the absence of detectable *BRCA* mutations (79).

For applications using whole-exome or targeted sequencing data, SigMA offers a probabilistic approach to estimate HRD by modeling the activity of SBS3 from sparse mutation profiles. It uses panel-specific simulations and a Bayesian framework to determine whether observed single nucleotide variants are consistent with HRD-associated mutational processes (90).

Additionally, shallowHRD and shallowHRD v2 apply genomic scar-based methods to shallow whole-genome sequencing (sWGS) data, quantifying allele-specific copy number alterations such as LOH, LST, and telomeric allelic imbalance (TAI) to infer HRD status at ultra-low sequencing coverage (91,92). These models are particularly useful for retrospective or cfDNA-based studies where high coverage is not feasible. Tools like scarHRD and GIInger follow similar logic, aggregating discrete chromosomal aberration metrics derived from SNP arrays, sWGS, or targeted sequencing to produce a composite HRD score (93,94). These methods are widely adopted in both tissue and liquid biopsy contexts, though their performance may vary across tumor types and platforms.

Machine learning has also been applied to structural and mutational patterns to expand HRD classification capabilities. CHORD is a random forest classifier trained on WGS data that uses the frequency of deletions with microhomology and tandem duplications to distinguish BRCA1- and BRCA2-like HRD profiles (95). Unlike signature-based tools, CHORD emphasizes rearrangement-based features and structural variation. In targeted sequencing applications, HRD-MILN offers a framework for identifying HRD from panel sequencing data by learning patterns of clustered LOH events or scoring high-impact genomic scars using multi-instance learning approaches (96). This allows for HRD status inference even when only limited genomic regions are available.

In the liquid biopsy domain, several exploratory tools aim to identify HRD phenotypes directly from ctDNA. DirectHRD is a scar-based classifier that infers homologous recombination deficiency by detecting microhomology-mediated deletions (mhDels) from whole-genome sequencing data. It uses a specialized caller to identify mhDels and quantifies their similarity to COSMIC Indel Signature ID6 through a multinomial mixture model. By focusing on this single, biologically specific signature of HRD, DirectHRD enables classification from both tumor tissue and cell-free DNA, even at low tumor fractions (TF) (97). Recently, DARC Sign (DnA Repair Classification SIGNatures) has introduced an XGBoost-based classifier trained on cfDNA whole-exome sequencing (WES) data to classify BRCA2, CDK12, and mismatch repair (MMR) deficiencies in metastatic PCa (98). By integrating 224 somatic features-including mutational contexts and copy number segmentation metrics-DARC Sign addresses common limitations in ctDNA assays, such as low TF and sparse variant calls, and is designed to be directly applicable in clinical liquid biopsy workflows.

Deep learning approaches based on histopathological imaging have also emerged. Tools such as DeepHRD, DeepSMILE, and attMIL use convolutional or attention-based neural networks trained on hematoxylin and eosin (H&E) stained slides to predict HRD status directly from digital pathology (99–101). These models provide a complementary modality to DNA-based testing and may help stratify patients when genomic data are unavailable.

1.8.2 Foundation models in genomic sequence analysis and HRD prediction

The emergence of foundation models in genomics represents a transformative advance in computational biology, building on innovations originally developed in natural language processing such as Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformer (GPT) (102,103). These large-scale neural networks are pre-trained on massive unlabeled genomic datasets using self-supervised learning, enabling them to learn directly from raw DNA sequences without relying on manually engineered features or curated annotations (104). By solving surrogate tasks such as masked token prediction or next-token generation, foundation models capture complex sequence dependencies and biological patterns embedded in the genome. This paradigm is particularly well suited to genomics, where vast amounts of sequencing data exist but only a small fraction is experimentally annotated (105–107). The representations learned through this process can then be fine-tuned for diverse downstream tasks, including regulatory element classification, variant effect prediction, and splice site recognition (108).

Central to their success is the transformer architecture, which eschews traditional recurrent or convolutional designs in favor of multi-head self-attention (109). This mechanism allows the model to dynamically weigh dependencies between distant sequence elements, making it particularly well-suited to biological sequences where regulatory logic often spans kilobases, such as enhancer-promoter interactions relevant to DNA repair gene regulation. Models like Enformer, DNABERT, and the Nucleotide Transformer exemplify how such architectures can learn complex patterns of transcriptional control, even in the absence of explicit labels (106,110,111).

The representation of genomic data in these models hinges on the choice of tokenization scheme. Most approaches convert nucleotide sequences into fixed-length overlapping substrings known as k-mers (e.g., 3-mers or 6-mers), which serve as the analog of words in natural language (111). Alternative strategies, such as byte-pair encoding (BPE), segment sequences into variable-length tokens based on frequency, offering a compact yet expressive vocabulary and helping mitigate class imbalance issues common in fixed k-mer spaces (105). Once tokenized, the sequences are embedded into high-dimensional vector spaces, where each token's representation incorporates both its identity and its relative position, enabling the model to reconstruct biological syntax and grammar (106). These embedding layers are either learned directly during training or enhanced with positional encodings (e.g., sinusoidal or rotary schemes), allowing the model to preserve sequential context over long genomic distances (112).

The biological rationale for applying transformers to DNA lies in the structural similarities between genome and language. Like sentences, genomes are composed of sequential elements governed by

rules; some local, others distal, and many context dependent. Just as words in a sentence derive meaning from surrounding syntax, nucleotide sequences often depend on long-range chromatin interactions and regulatory logic that span kilobases. Transformer-based models are uniquely capable of modeling such hierarchical and context-dependent relationships due to their self-attention mechanisms, which enable global context integration without being limited by fixed receptive fields (109). This architectural advantage makes them ideally suited to identifying and interpreting sequence variants that disrupt regulatory elements, including enhancers, promoters, and insulators. These capabilities are particularly relevant in the context of HRD, where regulatory perturbations in genes such as *BRCA1*, *BRCA2*, and *PALB2* can result in impaired DNA repair and contribute to tumorigenesis (113). Recent studies have demonstrated that foundation models like Enformer and the Nucleotide Transformer can learn non-linear, cross-distance regulatory dependencies, providing a path forward for modeling subtle sequence variations that underlie HRD-related phenotypes (114). Although foundation models have not yet been explicitly trained for HRD detection, they offer significant potential in this domain. Foundation models could be applied to non-coding variant interpretation, predicting the effect of regulatory disruptions that may silence key homologous recombination repair genes. Other protein-level models, including AlphaMissense and PrimateAI-3D, incorporate structural and evolutionary priors to classify missense variants, helping to reclassify variants of uncertain significance in HRD-associated loci (115,116). Furthermore, emerging multimodal models-currently in early development-are capable of integrating genomic, epigenomic, and fragmentomic features, opening the possibility of learning complex GIS signatures directly from cfDNA. Importantly, foundation models are also robust to sparse and noisy input data, making them especially well-suited to applications in liquid biopsy where coverage is often limited.

In summary, foundation models for genomics could represent a paradigm shift in how we learn from biological sequences. Their scalability, contextual awareness, and flexibility make them promising tools for modeling the regulatory and mutational landscapes underlying HRD, with the potential to enable more accurate, interpretable, and integrative assays for clinical diagnostics and personalized therapy.

1.8.3 Unmet needs and translational gap

While tissue-based assays such as myChoice® CDx and FoundationOne® CDx have become established tools for HRD detection in ovarian, breast, and PCas, their clinical utility is constrained by sample accessibility and quality. These limitations are particularly pronounced in PCa, where metastatic disease frequently involves the bone, rendering tissue acquisition technically challenging, invasive, and often associated with poor nucleic acid quality due to decalcification procedures. As a result, many patients with advanced or metastatic PCa are ineligible for reliable tissue-based genomic profiling, creating a substantial diagnostic gap. Importantly, emerging evidence suggests that HRD in PCa may

not be an early, truncal event, but instead can arise during disease progression under therapeutic selective pressure, particularly in the metastatic castration-resistant setting. Consequently, archival primary tumor tissue may fail to capture clinically relevant, late-emerging HRD alterations that drive treatment response or resistance. This temporal and spatial heterogeneity further undermines the reliability of single-site tissue biopsies for HRD assessment in advanced PCa.

1.9 Non-invasive means of HRD assessment

Accurate assessment of HRD is essential for guiding precision therapy in advanced prostate cancer, yet – as mentioned above - current diagnostic strategies rely predominantly on tumor tissue. In metastatic disease, tissue acquisition is often limited by poor sample quality, tumor heterogeneity, and the impracticality of repeated biopsies. These challenges have driven increasing interest in non-invasive approaches, particularly circulating tumor DNA (ctDNA)-based liquid biopsies, which offer the ability to capture genomic, structural, and epigenomic features of HRD dynamically and longitudinally from plasma.

The development and validation of ctDNA-based assays-leveraging GIS, mutational signatures, epigenomic alterations, fragmentomic features, or multimodal combinations thereof - represent a critical translational frontier. As sequencing technologies advance and computational methods mature, multi-layered liquid biopsy approaches are expected to complement or potentially replace tissue-based assays, enabling dynamic, minimally invasive, and clinically actionable HRD detection throughout disease evolution. Notably, no ctDNA-based HRD assay has received FDA approval, and current guidelines lack standardized approaches for liquid biopsy-based detection.

1.9.1 Biology of cell-free DNA

cfDNA refers to extracellular DNA fragments that are released into the bloodstream and other bodily fluids under physiological and pathological conditions. In healthy individuals, cfDNA primarily originates from apoptotic and, to a lesser extent, necrotic hematopoietic cells **(Figure 7A)**. These DNA fragments are characteristically short, with modal lengths around 166 base pairs, corresponding to the length of DNA wrapped around a nucleosome plus linker DNA. The predominant mechanisms of cfDNA release include apoptotic cell turnover, where endonucleases generate mono- and oligonucleosomal DNA fragments, and necrosis, which results in more heterogeneous and longer fragments due to less controlled cleavage events (117,118).

In cancer patients, a subset of cfDNA originates from tumor cells and is referred to as ctDNA. This fraction is typically more fragmented than cfDNA from healthy tissues and is shaped by the unique chromatin structure, transcriptional activity, and nucleosome positioning of the tumor cells from which

it derives (119). ctDNA can carry somatic mutations, structural variations, copy number alterations, and epigenetic features reflective of the tumor genome, thus providing a non-invasive proxy for tumor profiling. Notably, prostate cancer is generally considered a relatively low-shedding tumor type for ctDNA, meaning that a substantial subset of patients may have low or intermittently detectable tumor fractions in plasma, which can limit analytical sensitivity and increase false-negative rates (120).

The abundance and composition of ctDNA are influenced by multiple factors, including tumor burden, anatomical location, vascularization, and biological characteristics such as proliferation and apoptosis rates **(Figure 7B)**. In general, rapidly growing or necrotic tumors tend to shed more DNA into circulation (121). Moreover, physiological factors such as renal clearance and liver metabolism, as well as biological phenomena like clonal hematopoiesis of indeterminate potential (CHIP), may confound the interpretation of cfDNA signals (122).

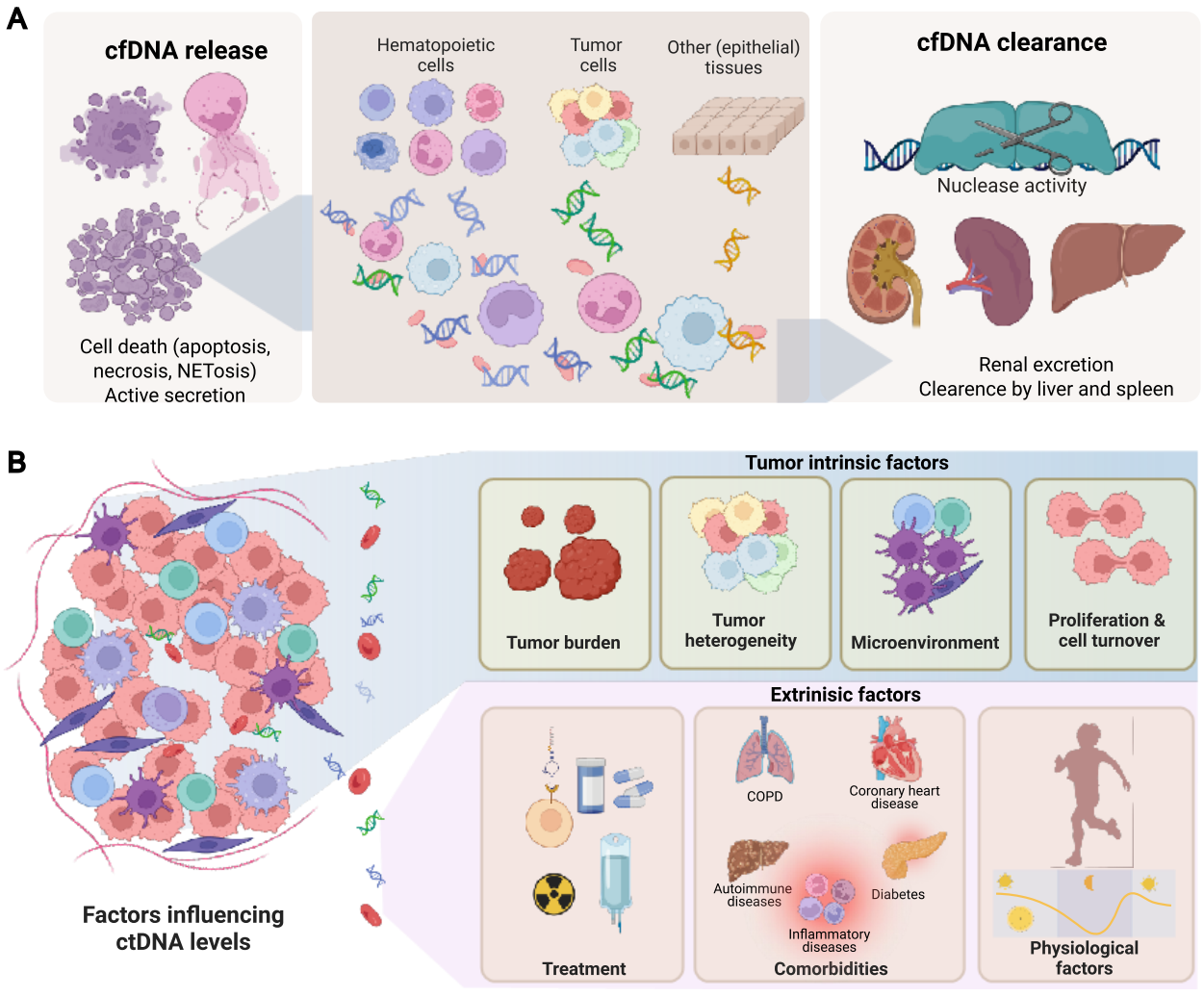


Figure 7: Evolution of liquid biopsy and cfDNA-based biomarker approaches.

(A) The cfDNA pool originates mainly from hematopoietic cells, with minor contributions from solid tissues. In cancer patients, tumor cells contribute circulating tumor DNA (ctDNA), whose levels reflect the balance between DNA release and clearance via nucleases, renal excretion, and hepatic/splenic uptake. (B) ctDNA abundance depends on tumor-intrinsic factors, such as burden, heterogeneity, and proliferation, as well as extrinsic physiological and treatment-related factors. Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license. Figure and legend adapted from Moser et. al (123). Created with BioRender.com

Recent investigations have provided detailed insights into the structural and biophysical properties of cfDNA. Accumulating evidence indicates that cfDNA fragmentation patterns are shaped by the nucleosomal organization and chromatin accessibility of the cells of origin, thereby enabling inference of tissue-specific contributions through analyses of fragment size distributions, end-motif composition, and nucleosome phasing (117). Tumor-derived cfDNA is characterized by increased fragmentation heterogeneity, systematic shifts in end-motif frequencies, and non-random cleavage at preferred genomic loci, reflecting tumor-specific chromatin architecture and epigenetic dysregulation (118,119). These reproducible, non-stochastic features constitute the biological basis for emerging fragmentomics-based analytical frameworks.

Critically, these insights into cfDNA and ctDNA biology have direct translational relevance. They underpin the conceptual and methodological foundations of liquid biopsy technologies and inform the development of assays for early cancer detection, molecular characterization, and longitudinal disease monitoring. Accordingly, a mechanistic understanding of cfDNA generation, fragmentation, and clearance is essential for accurate signal interpretation and for the rational design of robust analytical pipelines capable of distinguishing tumor-derived signals from the abundant background of non-malignant cfDNA (124).

1.9.2 Analytical approaches to ctDNA detection

Analytical strategies for the detection and quantification of ctDNA have evolved substantially, progressing from highly targeted mutation assays to comprehensive, multi-modal platforms that integrate genomic, epigenomic, and fragmentomic features. These approaches can be broadly categorized into targeted, genome-wide, and integrated or multimodal methodologies, each tailored to specific clinical use cases and varying levels of ctDNA abundance. Selection of the appropriate analytical framework depends on factors such as disease stage, tumor burden, and the intended application, including early detection, molecular profiling, or longitudinal disease monitoring.

Targeted mutation-based approaches

Early ctDNA assays focused on ultra-sensitive detection of somatic mutations using tumor-informed methods, including digital PCR, BEAMing, and hybrid-capture next-generation sequencing (NGS). With deep sequencing and molecular error correction, these approaches achieve sensitivities down to ~0.01% variant allele frequency (125). CAPP-Seq exemplifies this strategy, enabling patient-specific mutation tracking based on tumor exome data (122).

Tumor-naive (non-informed) mutation-based assays detect recurrent cancer-associated mutations without prior tumor profiling, expanding clinical applicability (126). However, both approaches are inherently limited by the low prevalence of clonal mutations in early-stage disease and the restricted genomic scope of mutation panels.

Several analytical innovations have also expanded cfDNA's somatic mutation diagnostic toolkit. CODEC (Concatenating Original Duplex for Error Correction) improves mutation detection by linking Watson and Crick strands, allowing ultra-sensitive sequencing with reduced error rates and lower sequencing depth (127).

Copy number and structural variation analysis

Somatic copy number alterations (SCNAs) and structural variants (SVs) are pervasive features of cancer and can be detected using ultra-low-pass or shallow whole-genome sequencing. SCNAs, which involve large genomic gains or losses, and SVs, including translocations and inversions, often arise early in tumorigenesis and provide informative biomarkers for detection and classification (128,129). These alterations are common across many cancer types and often represent early events in tumorigenesis, making them valuable for both detection and classification.

While short-read sequencing has limited resolution for complex SVs, long-read platforms substantially improve breakpoint detection and structural characterization. Computational tools such as ichorCNA, plasma-Seq, and PARE support tumor-naïve, genome-wide SCNA detection (130), whereas Delly, Manta, GRIDSS, and alignment-free methods like Aperture enable sensitive SV identification from ctDNA (131–134).

Long-read sequencing offers an orthogonal route for ctDNA structural-variant profiling because reads can span repetitive regions and complex rearrangement junctions, improving breakpoint localization and interpretation compared with short-read data (135) Recent nanopore-based approaches have shown that long-read workflows can be adapted for ctDNA SV detection and patient-specific monitoring, supporting their potential utility even at low tumor fractions (136). In practice, long-read ctDNA assays typically rely on preserving or enriching longer cfDNA molecules and using library preparation compatible with long inserts (often alongside single-stranded library preparation strategies that broaden

cfDNA fragment recovery), followed by alignment and SV calling tailored to long-read error profiles (137).

Overall, integrating genome-wide copy number and structural analyses - particularly with emerging long-read technologies - enhances ctDNA detection in low-burden disease and supports non-invasive cancer profiling, monitoring, and early detection.

1.9.3 Mutational signatures in ctDNA

Extracting mutational signatures from ctDNA can provide insight into a tumor's origin, evolutionary history, and therapeutic vulnerabilities. Traditionally, mutational signature analysis has required matched tumor-normal WES or WGS, which enables accurate somatic mutation calling by filtering out germline variants. However, in clinical practice, particularly in liquid biopsy settings, matched normal samples are often unavailable, and sequencing depth is limited. This has led to the development of two classes of computational tools: those for signature attribution, and those for mutation calling directly from tumor-only or cfDNA data.

For signature attribution, several computational tools have been validated for performance across various input sizes and mutation burdens. SigProfilerSingleSample, part of the broader SigProfiler framework, is optimized for small somatic mutation catalogs and allows precise assignment of known COSMIC signatures (138). MuSiCal (Mutational Signature Calculator) is a recent tool capable of both discovering novel signatures and fitting known ones with high accuracy, even in sparse datasets (139). Meanwhile, PASA (Presence Attribute Signature Activity) introduces a binary feature-based method that identifies signature activity based on the presence or absence of diagnostic mutations, making it particularly suited for samples with low variant counts (140).

To complement these, tools such as MisMatchFinder and Pointy have been developed to improve upstream mutation calling in tumor-only or low-pass ctDNA contexts. MisMatchFinder uses a read-level approach to detect mismatches and infer somatic mutations from low-coverage ctDNA WGS data, circumventing the need for a matched germline reference (141). Pointy, on the other hand, employs a machine learning model trained on ULP-WGS (0.3–1.5×) to identify somatic mutations and recover signature activity with high sensitivity and specificity, enabling signature analysis directly from cfDNA in early-stage cancers (142).

1.9.4 Fragmentomic signatures of ctDNA

ctDNA carries distinctive biophysical signatures that arise from tissue-specific chromatin organization and nuclease-mediated DNA processing. These fragmentomic features include fragment size distributions, nucleosome phasing, and DNA end motifs, which collectively reflect the transcriptional

activity and chromatin accessibility of the tissue of origin (143). Recent analytical advances have repurposed these mutation-agnostic signals for ctDNA detection, classification, and quantification.

Early work with DELFI (DNA Evaluation of Fragments for Early Interception) demonstrated that genome-wide fragmentation profiles derived from shallow whole-genome sequencing (sWGS) can distinguish cancer from non-cancer states across tumor types with high sensitivity (121). Subsequent approaches incorporated higher-resolution features, including end motifs, cleavage positioning relative to nucleosomes, and fragment jaggedness. Tools such as Griffin map fragmentation patterns to regulatory regions to enable molecular subtyping (e.g., estrogen receptor status in breast cancer) (144), while EPIC-seq links promoter fragmentation entropy to gene expression states and tissue-of-origin inference (145). Orientation-aware analyses further exploit strand-specific cleavage biases around open chromatin to predict tissue sources (146), and measures of cfDNA jaggedness capture aberrant nuclease activity associated with malignancy (147).

Recent innovations have further expanded fragmentomics into high-resolution and machine-learning-driven territory. FRAGMAXR evaluates nucleosome-level methylation-associated fragmentation patterns across ±800 bp windows, revealing how higher-order chromatin structure impacts cfDNA degradation. It improves detection across multiple cancer types, particularly in early stages (148). FRAGLE, a novel deep learning framework trained on large cfDNA cohorts, directly estimates ctDNA fraction from fragment length distributions with ~1% sensitivity and is compatible with both whole-genome and targeted sequencing data (149).

Beyond canonical mono- and dinucleosomal cfDNA, recent studies have uncovered distinct cfDNA populations, ultrashort (<70 bp, often single-stranded) and long (>500 bp) fragments, enabled by single-stranded library preparation and long-read sequencing (118). These newly characterized populations reveal tumor-specific nuclease activity and chromatin accessibility signatures, thereby expanding the biological resolution of fragmentomics. Advances such as FRAGHA (150), which integrates fragment size and 5′ end motifs to infer histone modification landscapes, and GALYFRE (151), which quantifies aberrant end positioning through information-weighted metrics, demonstrate how high-dimensional fragmentation data can capture functional epigenetic states. In parallel, deep learning approaches, including transformer-based models that learn cfDNA end-motif “languages” and convolutional networks capable of inferring CpG methylation directly from fragmentation patterns, are redefining computational fragmentomics (152,153). Collectively, these developments mark a shift toward multimodal, AI-integrated fragmentomics, bridging nucleosomal biology with clinical detection, functional interpretation, and precision biomarker discovery.

Recent studies have shown that cfDNA fragmentation features, including size profiles, end-motifs, and nucleosome positioning carry significant diagnostic value across cancers (154–156). Given their

elevated GIS, HRD phenotypes likely extend beyond canonical mutations and HRD tumors may also display distinct fragmentation and chromatin accessibility patterns in cfDNA, informing both molecular subtyping and HRD detection. Therefore, single-modality assays may underestimate HRD prevalence, while multimodal approaches integrating mutations, SCNAs, genomic scars, and fragmentomics features may offer superior detection of intrinsic tumor biology (157–159).

1.9.5 Multi-modal and ensemble approaches

Recognizing that single-feature models, relying on mutations, methylation, or fragmentomic patterns individually, have intrinsic limitations, recent efforts have converged on multi-modal liquid biopsy platforms. These approaches combine orthogonal cfDNA features to enhance detection sensitivity and specificity, especially in early-stage cancers where signal strength from any single modality may be weak **(Figure 8)**.

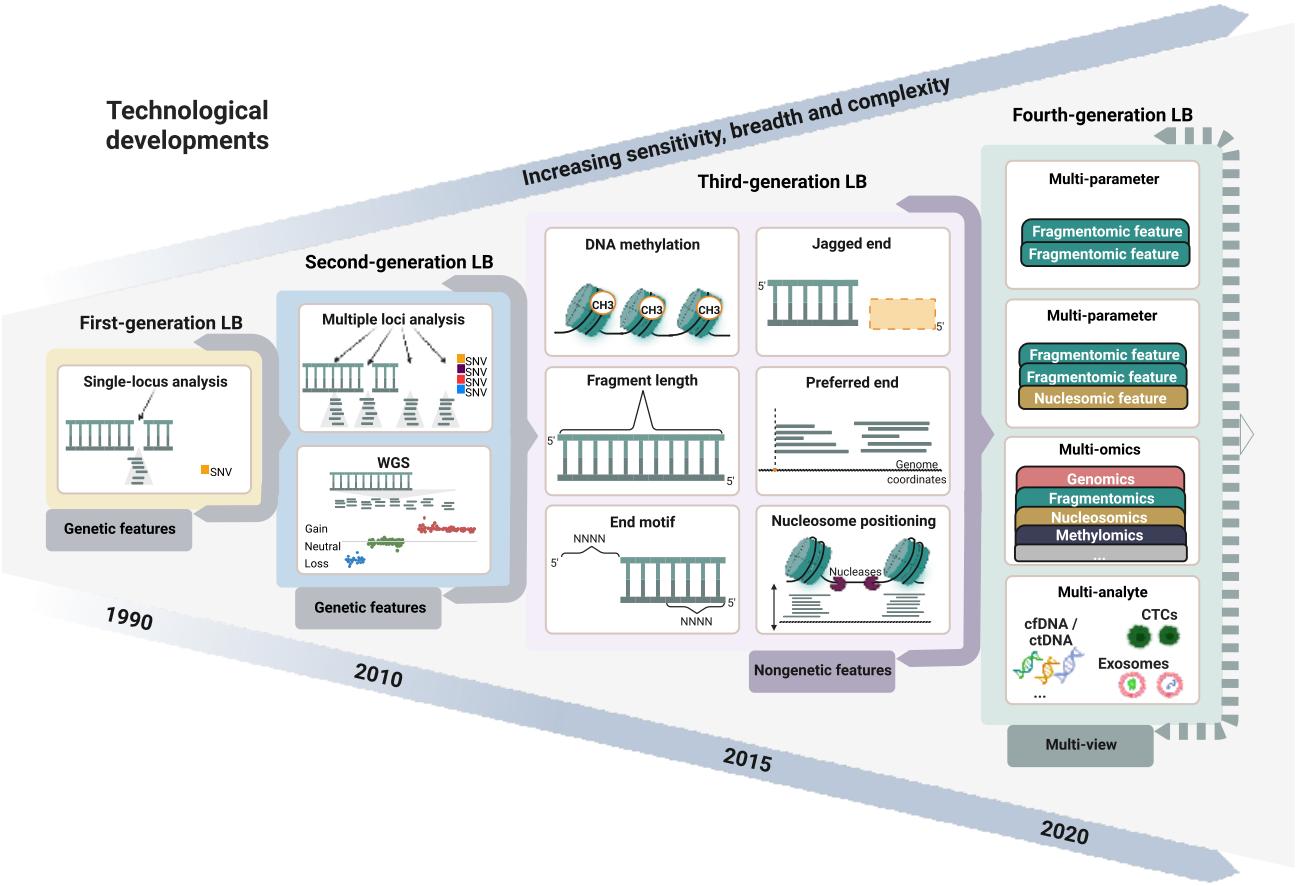


Figure 8: Evolution of liquid biopsy (LB) technologies and cfDNA analysis across generations.

First-generation assays targeted single or few recurrent cancer mutations (yellow panel). Second-generation methods expanded genomic breadth through multi-gene panels and whole-genome sequencing (blue panel). Third-generation approaches incorporated nongenetic cfDNA features such as methylation, fragment length, and end-motif profiles (purple panel). The emerging fourth-generation LB models (green panel) integrate multi-omic and multi-analyte data, including genomics, fragmentomics, nucleosomics, and methylomics, to enhance sensitivity and specificity. This evolution increases analytical complexity, necessitating advanced machine learning algorithms for data integration and interpretation. Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license. Created with BioRender.com. Figure and legend adapted from Moser et al. (122).

SPOT-MAS, integrates genome-wide and targeted methylation, CNAs, fragment length distributions, and end motif frequencies into a unified diagnostic pipeline. Using a two-stage machine learning framework and low-depth sequencing (~0.55×), SPOT-MAS achieved 72.4% sensitivity at 97.0% specificity across multiple cancers, including breast, colorectal, liver, and lung, demonstrating its potential for population-scale early cancer screening (159).

Similarly, EMMA (Expanded Multi-Modal Analysis) analyzes whole-genome bisulfite sequencing (WGBS) data to simultaneously detect differentially methylated regions (DMRs), SCNAs, and fragment

size ratios. This integrative method enabled sensitive detection of early and even pre-neoplastic esophageal squamous cell carcinoma, outperforming single-signal classifiers and highlighting its potential in early interception strategies (157).

The Cancer Signature Ensemble (CSE) classifier, which integrates average methylation fraction, copy number ratio, and fragment size ratio. In a multi-cancer validation study, the CSE achieved 88.9% sensitivity at 95.2% specificity-substantially higher than any single-modality classifier-and showed strong performance in traditionally hard-to-detect cancers such as prostate and lung (160).

Together, these multi-modal and ensemble approaches mark a shift in liquid biopsy toward maximizing biological signal through integration. By simultaneously capturing genetic, epigenetic, and structural alterations in cfDNA, these methods substantially enhance the sensitivity, robustness, and utility of blood-based cancer detection platforms.

1.9.6 ctDNA in PCa

PCa has some disease specific complications when it comes to genetic testing of the tumor. Tissue biopsy is usually performed with a core needle biopsy (161). However, due to the relatively high spatial heterogeneity of PCa, it is not uncommon to underdiagnose, misclassify (162), or miss subclonal tissue that could harbour treatable mutations (163). Emerging imaging techniques such as multiparametric or fusion Magnetic Resonance Imaging (MRI) and Transrectal Ultrasound (TRUS), or their combination, allow for guided biopsies with increased detection rates of malignant disease (164,165). However, these systematic biopsies still pose limitations and are prone to the aforementioned problems or miss metastatic heterogeneity. Furthermore, bone metastasis is very common in PCa, complicating tissue testing (166). Bone lesion biopsies are challenging to perform successfully. The amount of tissue is often insufficient and requires decalcification, a procedure that severely reduces the quality of DNA and therefore the results of genetic testing (167). Additionally, serious patient discomfort and the related morbidities limit their use and repeatability.

In recent years, ctDNA has emerged as a minimally invasive biomarker for disease detection, molecular profiling, and treatment monitoring in PCa. ctDNA levels correlate with tumor burden and disease stage, with the highest detectability observed in mCRPC, while lower fractions are typically present in localized disease. Despite these limitations, ctDNA analysis provides a dynamic and repeatable means of capturing tumor heterogeneity and clonal evolution under therapeutic pressure.

A unique advantage compared to tissue biopsies is that ctDNA can confer information about genomic intra-tumor heterogeneity that traditional biopsies could miss due to spatial reasons, and metastatic tumors (168,169). Multiregion sampling of tissue has been shown to improve genotyping, but it inadvertently increases patient discomfort (170). This marker can provide a rapid, non-invasive,

longitudinal analysis that covers the full extend of tumor genomics found in tumor clones. These characteristics make ctDNA sequencing an enticing alternative to tissue biopsies in PCa, which is a disease prone to challenging tissue biopsies and high levels of heterogeneity. Furthermore, in the context of PARPi resistance, subclonal *BRCA2* alterations have been reported as an important mechanism of resistance (171). Acquiring information from the different subclonal lesions in a manner that is easy to repeat simultaneously would therefore be particularly advantageous in this setting.

Despite these advantages, there are numerous challenges to be addressed concerning this biomarker, such as frequently low levels in the analyte, limited sensitivity in early stage cancers, DNA contribution from non-tumor tissues and ageing events like CHIP, over or under-representation of certain cell populations, possible overdiagnosis leading to overtreatment of indolent tumors, and lack of clinical validation (172–174). Despite the rapidly growing knowledge in the field, ctDNA testing remains in its infancy, particularly when it comes to clinical testing (175).

Emerging approaches in artificial intelligence and machine learning offer promising avenues to address some of these limitations. AI-based models are uniquely suited to handle the high-dimensional, complex, and often noisy data generated from ctDNA analyses (123,176). These methods can enhance analytical sensitivity by identifying subtle patterns in fragment length, end motifs, jagged ends, and methylation signatures-features often overlooked by traditional pipelines. For example, convolutional neural networks (CNNs) and deep learning models have been employed to decode fragmentation profiles and infer tissue-of-origin at high accuracy, even at low sequencing depth.

In this evolving landscape, AI-enhanced ctDNA analysis represents a critical step forward in overcoming biological and technical limitations. As these models mature and become clinically validated, they are poised to play a key role in refining early cancer detection, tracking minimal residual disease, and personalizing therapeutic decisions in PCa and beyond.

2 Rationale of the study

HRD is a clinically actionable biomarker with growing therapeutic relevance in PCa. While its predictive value for response to PARPi is well established, accumulating evidence suggests that HRD may also confer sensitivity to immune checkpoint inhibition, underscoring its broader relevance in precision oncology (177).

Despite this potential, HRD assessment in PCa remains largely tissue based, which poses substantial limitations in the metastatic castration-resistant setting. Metastatic biopsies are invasive and frequently yield insufficient or non-representative material due to spatial heterogeneity and prior therapeutic exposure. In addition, archival tissue often fails to capture the evolving genomic landscape characteristic of mCRPC, where clonal dynamics are rapid and treatment driven.

ctDNA-based liquid biopsy offers a compelling alternative by enabling non-invasive, longitudinal profiling of tumor genomes, including subclonal alterations and resistance mechanisms. However, the clinical implementation of ctDNA-based HRD detection is challenged by limited sensitivity at low TF, inter-assay variability, and the lack of standardized methods to reliably detect complex genomic events such as homozygous deletions or biallelic gene inactivation (178). Moreover, the predictive value of HRD-associated genomic features, such LOH, LST, TAI, and mutational signatures, has not been fully validated in liquid biopsy settings for PCa (179). Finally, fragmentomic and nucleosomic features of ctDNA, which reflect chromatin organization, nuclease activity, and nucleosome positioning, remain largely unexplored in the context of HRD, representing a potentially informative but underinvestigated dimension of HRD biology in liquid biopsy. **Therefore, the aim of this study was to systematically evaluate ctDNA-based HRD detection in PCa by integrating genomic scar metrics, mutational signatures, and fragmentomic features, with the goal of improving the sensitivity and clinical interpretability of HRD assessment in liquid biopsy.**

2.1 Research questions

This study aims to address the following key questions:

- **Feasibility:** Can the HRD phenotype be reliably detected using circulating cell-free DNA (cfDNA) in metastatic PCa, and what proportion of patients exhibit sufficient TF to enable informative liquid biopsy–based analysis?

- **Prevalence:** Among patients with adequate cfDNA tumor content, what fraction displays genomic or phenotypic evidence of HRD?
- **Non-BRCA HRR alterations:** What is the predictive value of non-BRCA HRR gene alterations (e.g., *PALB2, ATM, RAD51C/D*) for identifying HRD and for predicting sensitivity or resistance to PARP inhibitor therapy?
- **Modifier genes:** How do alterations in classically non-HRR genes (e.g., *TP53, PTEN, RB1*) influence the emergence, modulation, or masking of HRD in metastatic PCa?
- **Mutational signatures:** Can mutational signatures derived from cfDNA serve as robust biomarkers of HRD, providing a liquid biopsy–based alternative or complement to tissue genotyping?
- **Epigenomic features:** Do epigenomic alterations, including disruptions of transcription factor binding sites (TFBS), provide HRD-associated signals detectable in cfDNA?
- **Fragmentomics:** Are cfDNA fragmentomic features, such as fragment length distributions and end-motif patterns, associated with the HRD phenotype and informative for HRD classification?

2.2 Aims and hypotheses

The primary aim of this study was to investigate novel, clinically relevant strategies for detecting the HRD phenotype in mCRPC using circulating cfDNA. Given the well-recognized limitations of tissue-based testing in PCa - particularly the challenges posed by bone-predominant metastases, spatial heterogeneity, and limited tissue availability - this study evaluates liquid biopsy as a non-invasive, repeatable, and clinically informative alternative. By integrating multiple sequencing modalities and computational frameworks, we sought to improve the sensitivity, accessibility, and clinical interpretability of HRD diagnostics in real-world settings **(Figure 9)**.

A central hypothesis was that HRD can be reliably detected through cfDNA in a subset of patients with mCRPC, provided that sufficient tumor-derived DNA is present. Accordingly, we first quantified the proportion of patients with adequate cfDNA TF to support high-confidence genomic analysis and subsequently estimate HRD prevalence within this evaluable cohort using established GIS metrics and HRD-associated alterations. We further hypothesized that mutational signatures reflective of HRD can be extracted from cfDNA using both ultra-deep matched tumor-normal WES and be used to reliably distinguish HRD tumors.

Beyond mutation-centric approaches, this study investigated emerging epigenomic and structural indicators of HRD. We explored whether TFBS activity, inferred from coverage patterns in hcWGS data, differs in HRD tumors. Furthermore, we assessed whether fragmentomic features, specifically cfDNA

fragment lengths and end-motif profiles, are altered in HRD-positive samples, and whether these patterns can serve as surrogate markers for the HRD phenotype.

Through this multi-dimensional approach, we aimed to define a set of robust, accessible, and scalable biomarkers for HRD detection that extend beyond current tissue-based standards. This work may provide a foundation for integrating liquid biopsy into HRD-focused therapeutic decision-making, including for PARPi and potentially for immunotherapy.

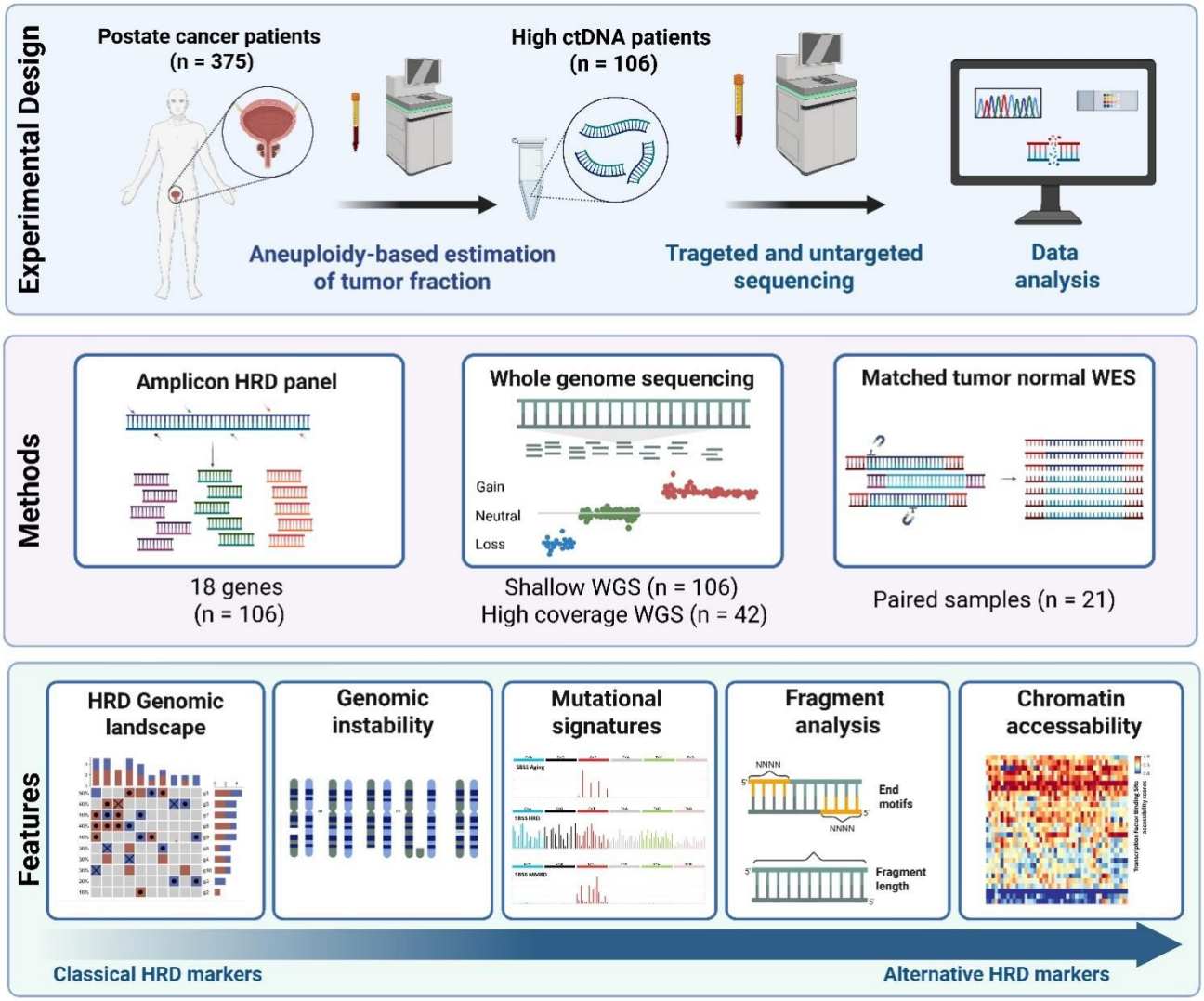


Figure 9: Multimodal framework for HRD detection from cfDNA in advanced PCa.

From an initial cohort of 375 patients with PCa with multiple sampling time points, we selected samples with elevated ctDNA levels from 106 patients for in-depth analysis. A tiered approach was applied to characterize both the causes and consequences of homologous recombination deficiency (HRD). Targeted panel sequencing was used to assess pathogenic variants in 18 HRR-related genes. Shallow and high-coverage whole-genome sequencing (WGS) enabled the quantification of GIS and extraction of fragmentomics and chromatin accessibility features. Whole-exome sequencing (WES) of matched tumor-normal cfDNA samples was performed in a subset to characterize HRD-associated mutational signatures. Downstream analyses integrated genomic alterations,

GIS, mutational signatures, fragment length patterns, end motifs, and transcription factor binding site (TFBS) accessibility to derive classical and alternative HRD biomarkers from plasma. Created in BioRender. Heitzer, E. (2025) https://BioRender.com/zhwwhzq Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.

This study addresses a critical unmet need by developing and validating ctDNA-based strategies for HRD detection in mCRPC. Through the integration of mutation analysis, copy number profiling, fragmentomic patterns, and epigenomic signals, it aims to improve the identification of HRD-positive patients who may benefit from targeted or immune-based therapies. In doing so, this work contributes to expanding the utility of liquid biopsy, advancing biomarker-driven precision oncology, and enhancing access to personalized cancer care.

3 Methods

3.1 Patient cohort and sample collection

Between 2012 and 2023, a total of 898 plasma samples were obtained from 375 individuals diagnosed with mPCa at the Division of Oncology, Medical University of Graz. Blood samples were collected by trained hospital staff and shipped to the Institute of Human Genetics on the same or the following day. Peripheral blood was drawn into either Streck Cell-Free DNA BCT® tubes (Streck, La Vista, NE, USA) or PAXgene Blood ccfDNA tubes (PreAnalytiX, Hombrechtikon, Switzerland) and processed by a two-step centrifugation protocol (1900 × g for 10 minutes, followed by 16,000 × g for 10 minutes). The buffy coat was retained after the first spin for matched germline DNA. All specimens were stored at −80 °C until further processing. All participants provided written informed consent, and ethical approval was granted by the institutional ethics committee (EK Nr ex 21-228).

We collected detailed clinical data for a subset of 106 patients selected for downstream sequencing, of whom 102 had complete records between June 19, 2018, and April 24, 2024. These patients comprised the analytical cohort used for targeted plasma-based mutation analysis and molecular characterization. The cohort included patients with metastatic castration-resistant PCa (mCRPC, n = 58), non-metastatic castration-resistant PCa (nmCRPC, n = 2), metastatic hormone-sensitive PCa (mHSPC, n = 30), non-metastatic hormone-sensitive PCa (nmHSPC, n = 3), and neuroendocrine PCa (nePC, n = 7) and Unknown (Unk, n = 6).

Each plasma sample represented a distinct biological replicate corresponding to a unique patient and collection time point. Technical replicates were not generated, as every cfDNA library was independently prepared and sequenced. Grouping of samples was guided by predefined molecular and clinical parameters rather than random assignment or blinding. All analyses were conducted retrospectively following completion of molecular profiling, with sample categorization such as HRD status, established on the basis of objective genomic criteria.

3.2 cfDNA and gDNA isolation

After the first centrifugation, the plasma fraction was carefully removed without disturbing the cellular interface, and the buffy coat (leukocyte fraction) was collected into a separate tube and stored at −80 °C for subsequent germline DNA isolation. The recovered plasma was then centrifuged again to remove residual cellular debris.

cfDNA was isolated from 2 mL of plasma using the QIAamp Circulating Nucleic Acid Kit (QIAGEN, Hilden, Germany) and quantified using the Qubit™ dsDNA High Sensitivity Assay Kit on a Qubit 4 Fluorometer

(Thermo Fisher Scientific, Waltham, MA, USA; RRID: SCR_026883), according to the manufacturer's instructions. If cfDNA yield was <40 ng, cfDNA was extracted from 4–8 mL plasma (using multiple aliquots as required). Germline DNA from white blood cells was isolated with the QIAamp DNA Mini and Blood Mini kit (QIAGEN, Hilden, Germany), according to the manufacturer's instructions.

3.3 Clinical data collection and processing

Baseline clinical and demographic information-including age, height, smoking status, date of diagnosis, and Gleason score was extracted from electronic medical records. Tumor burden and disease progression were characterized through TNM staging, biopsy assessments, and metastatic site documentation. Performance status was assessed using ECOG and Karnofsky indices, while comorbidity burden was quantified with the Charlson Comorbidity Index.

Treatment-related variables included the number and type of systemic therapies, details of ADT, bone-targeting agents, radiation therapy, and reasons for therapy modification or discontinuation. Clinical responses were assessed using serial PSA measurements, radiographic evaluation (RECIST criteria), and clinical progress notes. Survival outcomes, progression-free survival (PFS) and overall survival (OS), were calculated from the initiation of systemic therapy to the date of documented progression or death.

Baseline laboratory values included PSA, hemoglobin, neutrophil and lymphocyte counts, CRP, albumin, LDH, alkaline phosphatase, and neuroendocrine markers (NSE, chromogranin A, DHEA-S). Inflammatory and nutritional indices, such as the neutrophil-to-lymphocyte ratio (NLR) and CRP/albumin ratio, were derived accordingly. Data processing and statistical analyses were performed using Python (RRID: SCR_008394, v3.11.7), and patients with missing values for the respective analyses were excluded.

3.4 Tumor fraction estimation using mFAST-SeqS

To assess tumor-derived cfDNA content, the mFAST-SeqS protocol was utilized. LINE-1 repetitive elements were selectively amplified from 1 ng of cfDNA with specific primers (5 cycles), followed by indexing PCR (18 cycles). Purified libraries were sequenced using paired-end Illumina chemistry (Illumina, San Diego, CA, USA), targeting approximately 100,000 reads per sample. Alignments to the hg19 reference genome were performed using BWA-MEM2 (RRID: SCR_022192, v0.7.4), and high-quality reads (mapping quality >15) were used to calculate normalized Z-scores per chromosomal arm. A global Z-score was computed by summing the squared normalized values across all arms,

benchmarked against a panel of 35 healthy controls. The method was described in detail previously (181).

Based on the initial screening, a total of 106 plasma samples fulfilled the quality and tumor-fraction requirements and were subsequently selected for targeted sequencing to enable detailed profiling of HRR gene alterations. Given the exploratory and observational nature of this study, which aimed to describe cfDNA features associated with HRD rather than to evaluate a predefined effect size, no formal power calculation was conducted. The final cohort size (n = 106) was therefore determined by the number of high-quality plasma samples that met both sequencing and analytical inclusion criteria.

3.5 Targeted sequencing of DNA repair and tumor suppressor genes

Targeted amplicon sequencing was performed using the QIAseq HRR panel (QIAGEN, Hilden, Germany), covering a core set of HRR genes and custom-expanded to include *RB1*, *TP53*, and *PTEN*. Input cfDNA ranged from 10 to 40 ng. Libraries were prepared with enzymatic fragmentation, ligation of UMI-containing adapters, and dual PCR enrichment (8 cycles target-specific, followed by 20 cycles universal). Library quality was verified on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA; RRID: SCR_018043) and quantified via qPCR (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA; RRID: SCR_023455). Sequencing was carried out in 2×150 bp mode on either Illumina NextSeq 550 System (Illumina, San Diego, CA, USA; RRID:SCR_016381) or Illumina NovaSeq 6000 Sequencing System (Illumina, San Diego, CA, USA; RRID:SCR_016387), achieving a minimum coverage of 25,600×.

Raw reads were processed using the QIAGEN GeneGlobe platform (QIAGEN, Hilden, Germany; RRID: SCR_021211), which performed adapter trimming, UMI clustering, and alignment to hg19 using BWA-MEM2 (RRID: SCR_022192, v0.7.9a). Variants were called using smCounter2 and annotated in Golden Helix VarSeq (Golden Helix, Bozeman, MT, USA; RRID: SCR_001285, v2.2.0) using Genome Aggregation Database (gnomAD RRID:SCR_014964) and dbNSFP (RRID:SCR_005178) databases. Variant classification followed ACMG-AMP criteria (RRID: SCR_005769), with additional manual review of low variant allele frequency (VAF) (<0.5%) calls.

To assess the performance of the HRR assay, we used Seraseq ctDNA Complete™ Mutation Mix standards (SeraCare Life Sciences, Milford, MA, USA) with different VAF, including 5%, 2.5%, 1%, 0.5%, 0.1% and a wild type sample. These reference materials included synthetic oligos carrying common mutations in a synthetic plasma-like matrix. The extended panel (including *RB1*, *PTEN*, and *TP53*) was validated using the Seraseq ctDNA™ Mutation Mix v2 standards (1%, 0.5%, 0.125% VAF and WT). The reference material was suspended in a plasma-like matrix, which allowed us to test the

whole procedure, from DNA isolation, quality and quantity control, library preparation, sequencing and analysis.

3.6 Clonality assignment of variants

Variants detected from the targeted cfDNA panel were further classified by variant origin (somatic vs germline) and (ii) clonality (clonal vs subclonal) to support downstream analyses of tumor heterogeneity. Variants were annotated as germline when their VAF was consistent with heterozygous inheritance (i.e., clustering around ~50% VAF), particularly in samples with low TF. For putative somatic variants, clonality was inferred using the expected relationship between TF and VAF. Variants were considered clonal when the observed VAF was close to the expected clonal VAF (operationally, within ±10% of the TF value), and subclonal when the observed VAF was substantially lower than expected, compatible with presence in a tumor subpopulation (heterogeneity) or late-emerging clones. Borderline/low-VAF calls were reviewed to ensure that the assigned label was consistent with assay noise limits and the overall TF context. In cases where cfDNA VAF alone was insufficient to distinguish somatic from germline origin (e.g., at intermediate tumor fractions of ~40–60%, where VAF patterns can overlap), variants were orthogonally assessed by bidirectional capillary Sanger sequencing in matched WBC genomic DNA to establish germline status.

3.7 Sanger sequencing

Target loci were PCR-amplified from WBC genomic DNA and subjected to bidirectional capillary Sanger sequencing on an Applied Biosystems AB 3730 platform (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA; RRID:SCR_018052). Sequencing trace files (.ab1) were analysed in SeqScape 3 (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, USA) using gene-/region-specific project templates and reference data groups. Where applicable, M13-tailed primer designs were used to enable standardized sequencing priming. Chromatograms were reviewed manually and a variant was considered confirmed when the expected heterozygous signal was observed in both forward and reverse reads; runs with insufficient quality or discordant traces were repeated or classified as non-evaluable. Confirmed variants were annotated as germline for downstream analyses.

3.8 Whole-exome sequencing

Matched tumor and germline WES was conducted for 21 patients. cfDNA libraries were constructed using the Illumina DNA Prep with Twist Exome 2.5 Enrichment Kit (Illumina, San Diego, CA, USA), while germline libraries were prepared using the Twist Human Exome 2.0 (Twist Bioscience, South San Francisco, CA, USA). Sequencing was performed on Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA; RRID:SCR_016387) in 2×150 bp format. Tumor libraries achieved >1500× mean coverage;

germline samples averaged at 150×. Performance metrics were validated using the Twist cfDNA Pan-Cancer Reference Standard (Twist Bioscience, South San Francisco, CA, USA). FASTQ files were aligned to hg38 with BWA-MEM2 (RRID: SCR_022192, v0.7.9a), and somatic mutations were identified using GATK Mutect2 (RRID: SCR_026692). Signature extraction utilized the MutationalPatterns R package (RRID: SCR_024247) with COSMIC v3.4 (RRID: SCR_002260). Copy number signature analysis was performed using Sequenza (RRID: SCR_016662) and SigProfilerExtractor (RRID:SCR_023121)

3.9 Whole-genome sequencing

cfDNA libraries were prepared using the TruSeq DNA Nano Kit (Illumina, San Diego, CA, USA) with 10 ng input. Low-pass WGS (~0.1–0.2×) was conducted on NovaSeq 6000 SP flow cells; high-coverage WGS (~30×) was carried out on S4 flow cells (Illumina, San Diego, CA, USA). Read alignment to hg19 was performed using BWA-MEM2 (RRID: SCR_022192, v0.7.9a) (182), and duplicates were marked with Picard (RRID: SCR_006525). Low-pass WGS data were analyzed with ichorCNA (RRID: SCR_024768) (130), and large-scale genomic alterations were identified using shallowHRD (v1.0) (91). Thresholds for HRD classification followed published LGA cutoffs.

3.10 Structural variant analysis

Structural variant analysis was performed on high-coverage cfDNA whole-genome sequencing data using Manta (v1.6.0; RRID:SCR_022997) configured in tumor-only mode (183). Sequence alignment and variant calling were conducted against the hg38 human reference genome. The resulting VCF files were subjected to post-processing to retain only high-confidence structural events spanning ≥50 bp, encompassing deletions, duplications, inversions, and translocations. Functional annotation of SVs was carried out with the Ensembl Variant Effect Predictor (VEP v115; RRID:SCR_007931) in offline mode using the Ensembl release 115 cache and the --everything parameter to ensure full functional and regulatory annotation (184). The annotated dataset was subsequently filtered to retain structural alterations intersecting HRR genes.

3.11 Fragmentomic analysis

After alignment to the hg38 reference genome, PCR duplicates were marked using Picard's MarkDuplicates (RRID: SCR_006525) and removed together with unmapped reads, reads with mapping quality below 20, and secondary or supplementary alignments (185,186). Fragment lengths were directly extracted from the filtered BAM files. For end-motif analysis, the four-nucleotide (4-mer) sequence at the 5′ end of each cfDNA fragment was retrieved and used for motif-based feature construction.

Fragment length analysis was restricted to cfDNA fragments ranging from 80 to 400 base pairs (bp). For each sample, fragment length counts were converted to relative frequencies and $\log_{10}$-transformed. A 32 bp rolling average was applied to smooth the frequency distributions, and the resulting values were Z-score standardized using the mean and standard deviation derived from an independent cohort of 14 healthy, age- and sex-matched control samples. The smoothed Z-score profiles were divided into nine non-overlapping 32 bp bins, and the standard deviation of Z-scores within each bin was used to represent fragment length variability. These metrics were retained as the final set of fragment length–based features.

For end-motif analysis, all 256 possible 4-mer sequences were normalized using centered log-ratio (CLR) transformation to correct for compositional dependencies. Spectral clustering was then applied with a target of 25 clusters, followed by exclusion of clusters predominantly composed of low-frequency motifs (mean frequency < 0.008), yielding a refined set of 19 representative end-motif features. Both the fragment length and motif-derived features were computed genome-wide, without restricting the analysis to regions affected by copy number alterations.

Given the limited number of HRD-positive cases in the dataset, a conventional training/test split was avoided to prevent variance inflation. Instead, model performance was assessed using a 100-fold repeated stratified cross-validation framework. In each iteration, a logistic regression classifier with L2 regularization (C = 1) was trained using the combined fragment length and end-motif features. A late-fusion strategy was employed, whereby the positive-class probabilities across all iterations were averaged to generate final prediction scores per sample.

To minimize bias and improve interpretability in this small and imbalanced cohort, features were engineered using control data alone, ensuring that model construction did not leak information from HRD-positive samples. This approach emphasized biologically meaningful deviations in cfDNA fragmentation patterns rather than aggregate statistical metrics. By integrating localized variability in fragment length and clustered end-motif structures, the analysis captured subtle, yet interpretable, signals associated with HRD.

Overall, this modeling strategy was optimized for rare-event detection, balancing robustness with biological interpretability. Repeated cross-validation and model averaging provided stable performance estimates and reduced overfitting, supporting reliable discrimination of HRD-positive cases based on cfDNA fragmentomic signatures.

3.12 Chromatin accessibility at transcription factor binding sites

Chromatin accessibility was inferred from cfDNA whole-genome sequencing data using LBFextract (v0.1.0a1), which quantifies nucleosomal coverage around TFBS (187). For each TFBS locus, fragment

coverage was computed within ±2,000 bp windows (BED format) and normalized to account for sequencing depth and local variation. Outlier removal and signal centering were performed automatically within the LBFextract workflow.

To minimize GC-content bias at the fragment level, cfDNA reads were processed with GCparagon, a computational method specifically optimized to correct fragment length–dependent GC biases in cell-free DNA. GCparagon assigned bias-correction weights to individual reads, which were incorporated during coverage estimation by LBFextract, ensuring accurate and GC-adjusted fragment coverage across samples (188).

For the transcription factor-specific analysis, binding site coordinates were obtained from the Gene Regulation Databases (GTRD, RRID:SCR_008033 version 21.12). The *Homo_sapiens_meta_clusters* dataset was parsed by the *tfTitle* field, and transcription factors were ranked by *peak.count*, corresponding to the number of ChIP-seq experiments supporting each metacluster. To avoid artefactual regions, all TFBS (±2,000 bp around each center) overlapping the ENCODE Blacklist v2 (hg38) were excluded. From the remaining high-confidence sites, the top 1000 TFBS per factor were retained, yielding 1,048 transcription factors with sufficient genomic representation.

Coverage extraction and normalization proceeded in three stages:

1) *In silico* normalization: To harmonize TF and sequencing depth across samples, patient-derived reads were computationally mixed with randomly selected reads from healthy donor cfDNA pools, as described by Lazzeri et al. (187), generating standardized synthetic datasets.

2) Local signal normalization: Within LBFextract, mean signal normalization was applied to the flanking regions (−2,000 to −1,000 bp and +1,000 to +2,000 bp relative to each TFBS center) to correct for regional coverage variation.

3) Differential accessibility analysis: Accessibility scores were calculated for each transcription factor, and group-wise comparisons (e.g., HRD vs. HRR-proficient) were performed using the Mann–Whitney U test with Benjamini–Hochberg FDR adjustment. Transcription factors with an adjusted $p < 0.05$ were considered significantly differentially accessible.

3.13 Statistical analysis

All statistical analyses were conducted using Python (RRID: SCR_008394, v3.11.7) and R (RRID: SCR_001905, v4.3.3). Data distributions were evaluated for normality using the Shapiro-Wilk test. When normality was confirmed, homogeneity of variances across groups was verified by examining residual plots. For data meeting these assumptions, two-sided parametric tests (Student's *t*-test or ANOVA) were applied; otherwise, non-parametric tests (two-sided Mann–Whitney *U*) were used.

Associations between variables were evaluated using Spearman's rank correlation. Adjustment for multiple testing was performed using the Benjamini-Hochberg false discovery rate (FDR) procedure.

The choice of statistical test was guided by the nature and distribution of the data, and all relevant details, including the test type, *p*-values, and effect size estimates (e.g., hazard ratios with 95% confidence intervals), were provided in the corresponding figure legends and Methods subsections. Unless stated otherwise, all tests were two-sided and a threshold of $p < 0.05$ was considered statistically significant.

4 Results

4.1 Patient cohort characteristics and ctDNA tumor burden

To ensure sufficient analytical depth for genomic and fragmentomic profiling, samples were enriched for high ctDNA content, irrespective of their known HRD status at the time of selection. TF was primarily estimated using the mFAST-SeqS, which provided a quantitative, aneuploidy-based metric of ctDNA abundance. Samples with a Z-score >5, corresponding to an estimated TF of approximately ≥ 5–10% (181), were considered to have adequate tumor-derived signal for downstream analyses **(Figure 10A-C)**. Of the full dataset, 257 samples met this threshold, from which a final subset of 106 high-TF samples was selected based on plasma volume, sequencing quality, and availability of clinical follow-up data.

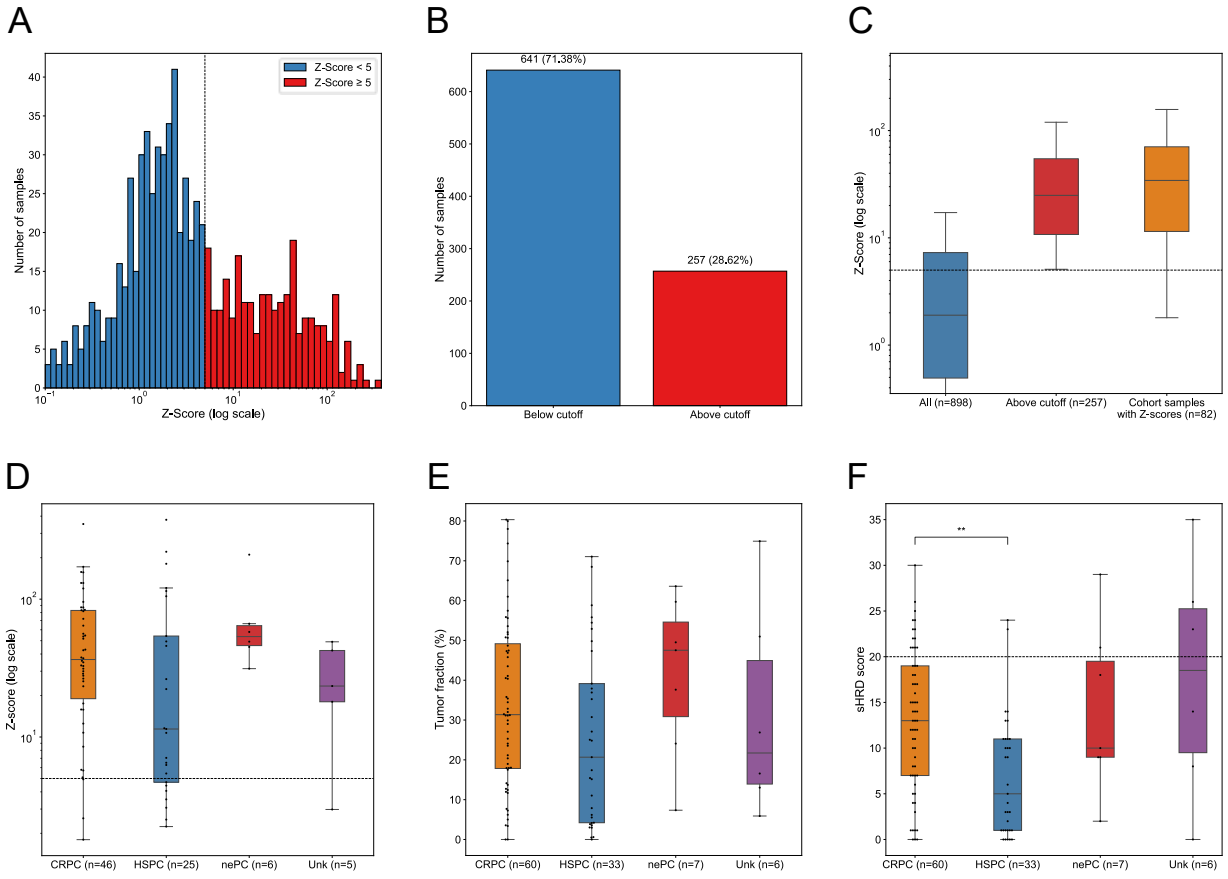


Figure 10: Summary of tumor fractions assessed by mFAST-SeqS and sWGS.

A) *Histogram of Z-scores derived from mFAST-SeqS across 898 plasma samples from 375 PCa patients. A threshold of Z > 5 was used to enrich for samples with estimated TF ≥ 5–10%.* ***B)*** *Proportion of samples above and below the Z-score cutoff.* ***C)*** *Boxplot of Z-scores (log scale) in all samples, those above the cutoff, and the final high-TF subset used for genomic analysis.* ***D)*** *Z-score distribution across four clinical subgroups: castration-resistant PCa (CRPC), hormone-sensitive PCa (HSPC), neuroendocrine PCa (nePC), and unclassified (Unk).* ***E)*** *Tumor fraction (TF) distribution across across the same subgroups.* ***F)*** *sHRD scores across the same subgroups. A significant difference in sHRD was observed between mCRPC and mHSPC groups (p < 0.01, Mann–Whitney U test). In all boxplots, the center line denotes the median, box edges indicate the interquartile range (IQR), and whiskers extend to 1.5× IQR. Individual data points are overlaid as black dots. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

As a secondary validation, TF was also inferred using ichorCNA on sWGS data, showing high concordance with the mFAST-SeqS estimates (Spearman ρ = 0.89, p < 0.0001; **(Figure 11A)**.

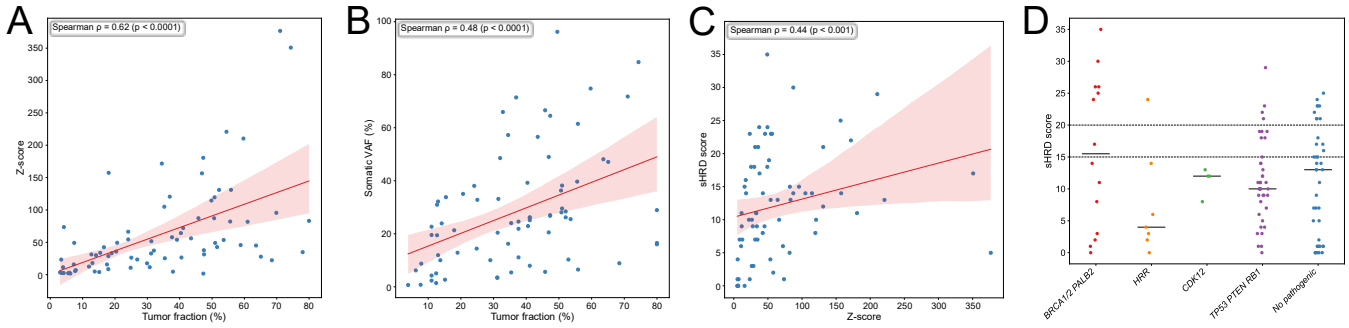


Figure 11: Correlation between tumor fraction, variant allele frequency, and GIS scores in plasma cfDNA.

A) *Scatterplot comparing tumor fraction (TF) estimated by sWGS with Z-scores.* ***B)*** *Scatterplot of TF versus the median variant allele frequency (VAF) of somatic mutations detected in each sample.* ***C)*** *Scatterplot comparing Z-scores with corresponding sHRD scores derived from sWGS, reflecting large-scale GIS. Spearman correlation coefficients and p-values are shown in panels A–C.* ***D)*** *Strip plot showing the distribution of sHRD scores across five mutation-defined subgroups: BRCA1/2 or PALB2, other HRR genes, CDK12, TP53/PTEN/RB1, and samples without pathogenic mutations. Horizontal dashed lines indicate sHRD thresholds of 15 and 20. Median values for each group are indicated with horizontal black lines. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

Clinical data were available for 102/106 patients (96.2%). The cohort was comprised predominantly of patients with mCRPC (*n* = 60), mHSPC (*n* = 33), and nePC (*n* = 7). As expected, the median TF was highest in mCRPC and nePC cases **(Figure 10D–E)**, reflecting the increased tumor burden and

aggressive biology of these subtypes. Neuroendocrine PCa cases exhibited elevated TFs values, with a median of 47.5%, compared to 31.4% in CRPC and 20.7% in HSPC. The differences between the groups did not reach statistical significance (Kruskal–Wallis, p=0.145).

CRPC samples showed significantly higher sHRD scores than HSPC ($p < 0.01$, Mann–Whitney U test). As sHRD reflects genome-wide 'genomic scar' features (copy-number–derived instability associated with HRD), this difference is consistent with increased chromosomal instability in the castration-resistant disease state. Genomic instability is a major feature of progression to CRPC and can be recovered non-invasively from cfDNA profiling, supporting the biological plausibility of higher sHRD in advanced disease (189,190). Notably, this pattern may also be influenced by higher ctDNA fractions typically observed in advanced disease, which improves sensitivity for copy-number and genome-wide scar detection in plasma.

Patient demographics and clinical variables were summarized in **Table 1**. The median age at sampling was 63 years (range: 47–88). The majority of patients had preserved performance status, as measured by ECOG and Karnofsky indices, despite advanced disease. Laboratory variables such as PSA, CRP, albumin, LDH, and hematologic markers were recorded to support subsequent correlative analyses. Survival analyses revealed that neuroendocrine PCa was associated with the poorest OS (median OS = 5.2 months), consistent with its aggressive clinical phenotype **(Figure 12A)**. In contrast, hormone-sensitive patients exhibited longer survival, with median OS exceeding 18 months in some cases. PFS analyses mirrored these findings **(Figure 12B)**. These differences underscored the heterogeneity of advanced PCa and highlighted the need for molecular biomarkers to stratify patients beyond traditional clinical classifications.

Table 1: Patient characteristics and clinical data

Reproduced by Vlachos et al. (180) with permission from the original authors, under the terms of the Creative Commons CC BY license.

Patient Characteristic	CRPC (n=60)	HSPC (n=33)	Unknown (n=6)	nePC (n=7)	Total (n=106)
Age	63.0 (47.0–88.0)	68.0 (55.0–84.0)	73.0 (66.0–80.0)	62.0 (47.0–65.0)	65.0 (47.0–88.0)
Gleason score	7 (6-9)	8 (6-8)	7 (6-8)	8 (8-8)	7 (6-9)
Metastatic diagnosis	Metachronous: 34	Metachronous: 10	NA	Metachronous: 3	Metachronous: 47
	De novo: 24	De novo: 20	NA	De novo: 4	De novo: 48
	No: 2	No: 3	NA	NA	No: 5
Radiotherapy as primary treatment	Yes: 14	Yes: 7	Yes: 0	Yes: 0	Yes: 21
	No: 46	No: 26	No: 6	No: 7	No: 85
	NA	NA	NA	NA	NA
Palliative radiotherapy	Yes: 34	Yes: 19	Yes: 0	Yes: 5	Yes: 58
	No: 24	No: 14	No: 2	No: 2	No: 42
	NA: 2	NA	NA: 4	NA	NA: 6
Neoadjuvant hormone therapy	Yes: 12	Yes: 2	Yes: 1	Yes: 0	Yes: 15
	No: 48	No: 31	No: 1	No: 7	No: 87
	NA	NA	NA:4	NA	NA: 4
PSA doubling time (Months)	2.2 (0.25–25.3)	1.85 (1.1–13.7)	NA	1.8 (1.1–3.8)	2.05 (0.25–25.3)
Karnofsky performance index	80.0 (50.0–100.0)	80.0 (50.0–100.0)	NA	90.0 (60.0–100.0)	80.0 (50.0–100.0)
Alkaline phosphatase (U/L)	221.0 (44.0–1721.0)	131.0 (63.0–1626.0)	582.5 (534.0–631.0)	130.0 (69–272)	206.0 (44.0–1721.0)
Lactate dehydrogenase (U/L)	309.0 (165.0–2411.0)	246.5 (163.0–1106.0)	584.0 (584.0–584.0)	657.0 (227–1789)	292.0 (163.0–2411.0)
Hemoglobin (g/L)	11.55 (7.5–15.4)	12.65 (9.1–15.8)	10.8 (10.8–10.8)	10.8 (8.8–13.5)	11.7 (7.5–15.8)
PSA (ng/mL, plasma)	162.35 (0.01–10000.0)	16.89 (0.01–2397.34)	229.0 (4.61–453.39)	22.75 (3.2–204.51)	55.52 (0.01–10000.0)
ctDNA fraction	31.35 (0.0–80.34)	20.67 (0.0–71.05)	21.73 (5.9–74.92)	47.53 (7.35–63.6)	30.32 (0.0–80.34)
sHRD score	13.0 (0–30)	5.0 (0–24)	18.5 (0–35)	10.0 (2–29)	11.0 (0–35)
Z-score	36.48 (1.8–350.87)	11.43 (2.24–376.65)	23.43 (2.98–49.06)	53.51 (31.33–210.35)	33.93 (1.8–376.65)
Follow-up for OS (months)	76.5 (4.0–255.0)	56.0 (2.0–237.0)	16.5 (2.0–31.0)	52.0 (20.0–192.0)	61.5 (2.0–255.0)

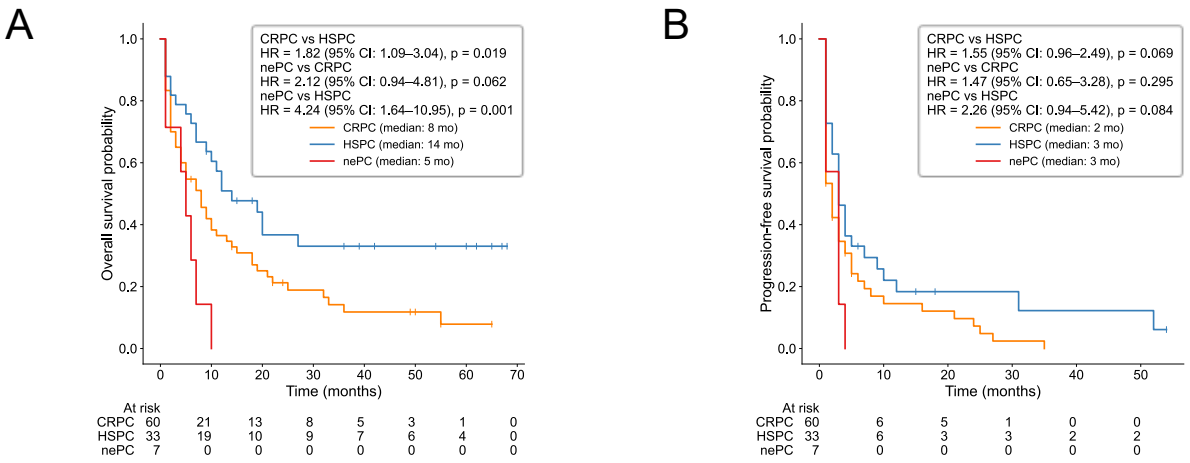


Figure 12: Kaplan Meier plots stratifying the patients by PCa subtype.

A) *Overall survival (OS) and* ***B)*** *progression-free survival (PFS) stratified by clinical group: castration-resistant PCa (CRPC), hormone-sensitive PCa (HSPC), and treatment-emergent neuroendocrine PCa (nePC). Median survival times are indicated in the legend. Hazard ratios (HR), 95% confidence intervals (CI), and log-rank p-values are shown for pairwise comparisons between groups. Censoring is indicated by vertical tick marks. At-risk tables are shown below each plot. Survival curves were estimated using the Kaplan–Meier method; HRs were derived from Cox proportional hazards models. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

4.2 Genomic landscape of HRD-related gene alterations

To investigate the genomic alterations contributing to HRD in ctDNA, we applied a targeted amplicon sequencing panel covering 18 HRR genes, including *BRCA1*, *BRCA2*, and *PALB2* as well as *TP53*, *RB1*, *PTEN*, and *CDK12*, as key tumor suppressors and modulators of PCa progression and therapeutic resistance.

Among the 106 high TF samples, at least one pathogenic or likely pathogenic alteration was identified in 65 patients (61.3%). Mutations in canonical HRD associated genes (*BRCA1*, *BRCA2*, or *PALB2*) were observed in 14 cases (13.2%), with an additional 13 patients (12.3%) harboring alterations in other HRR genes such as *ATM*, *RAD51C*, *RAD51D*, *CHEK2*, and *BRIP1*. Notably, mutations in non-HRR tumor suppressors were more frequent: *TP53* (43.4%), *RB1* (43.4%), *PTEN* (9.4%), and *CDK12* (3.8%). Overall, 26.4% of patients had biallelic or multi-gene alterations, suggesting functional redundancy or convergent pathway inactivation.

Analysis of VAFs revealed distinct patterns across gene categories. Mutations in *TP53*, *RB1*, and *CDK12* were typically high-VAF and clonal, consistent with early driver events or subclonal expansions under therapeutic pressure. In contrast, HRR gene alterations displayed broader VAF distributions, including numerous subclonal or borderline-detectable events **(Figure 13A)**. This was particularly evident for *ATM*, *CHEK2*, and *PALB2*, where lower VAFs suggested either spatial heterogeneity, clonal hematopoiesis, or late-emerging clones. Subclonal variants frequently occurred below the expected VAF for diploid heterozygous mutations, indicating allelic imbalance or subpopulation confinement **(Figure 13C-D)**.

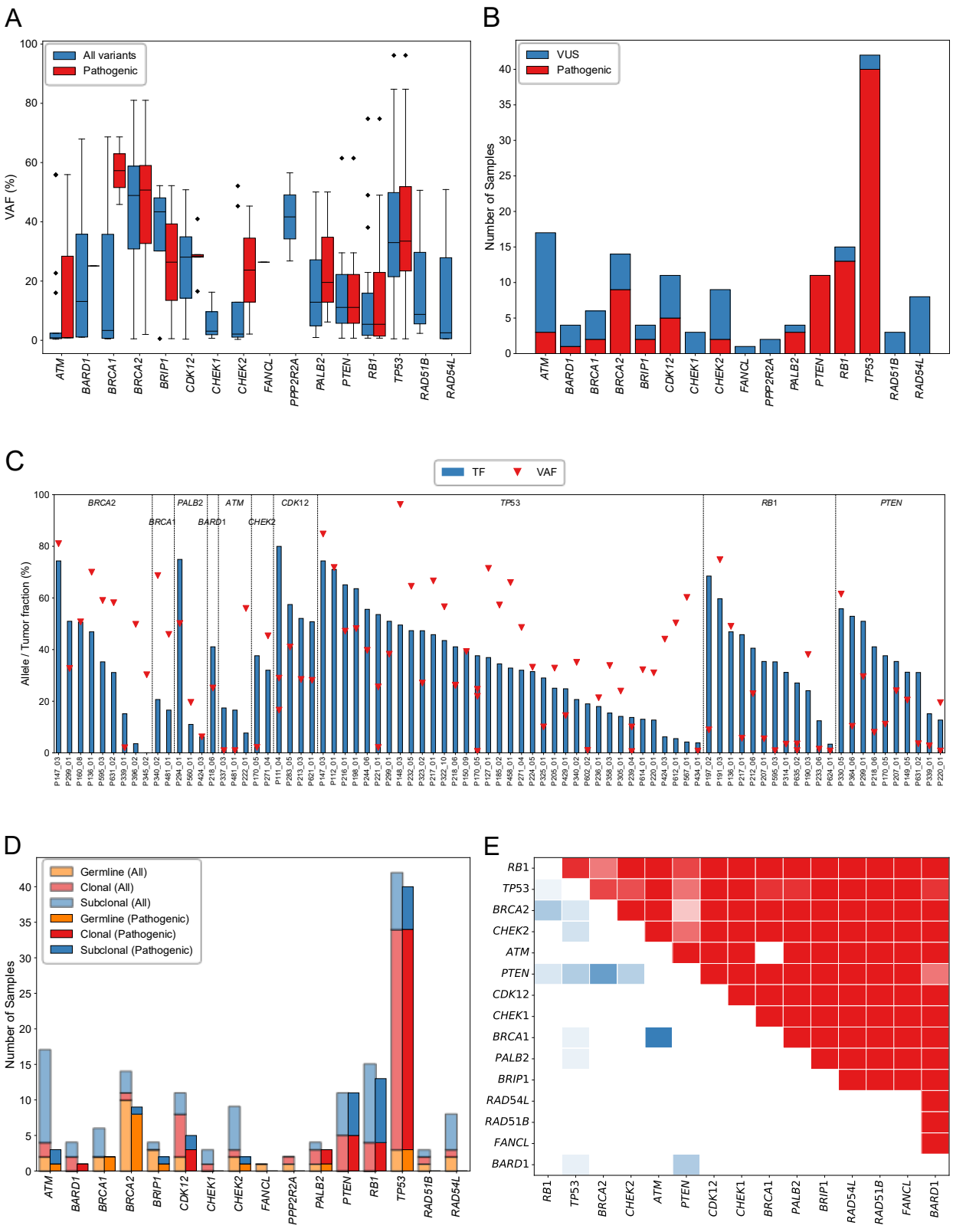

A
All variants
Pathogenic
VAF (%)
ATM
BARD1
BRCA1
BRCA2
BRIP1
CDK12
CHEK1
CHEK2
FANCL
PPP2R2A
PALB2
PTEN
RB1
TP53
RAD51B
RAD54L
B
VUS
Pathogenic
Number of Samples
C
TF
VAF
Allele / Tumor fraction (%)
BRCA2
PALB2
ATM
CDK12
TP53
RB1
PTEN
BRCA1
BARD1
CHEK2
D
Germline (All)
Clonal (All)
Subclonal (All)
Germline (Pathogenic)
Clonal (Pathogenic)
Subclonal (Pathogenic)
Number of Samples
E
RB1
TP53
BRCA2
CHEK2
ATM
PTEN
CDK12
CHEK1
BRCA1
PALB2
BRIP1
RAD54L
RAD51B
FANCL
BARD1

Figure 13: Patterns of pathogenicity, clonality, and co-occurrence of HRD-associated alterations in cfDNA.

A) *Variant allele frequencies (VAFs) of mutations detected in HRR genes and key tumor suppressors (TP53, RB1, PTEN)* ***B)*** *Distribution of variant classifications across genes* ***C)*** *Shown are tumor fractions (TF, blue bars) and VAF of detected mutations for clonality assessment of mutations* ***D)*** *Distribution of mutational origin* ***E)*** *Co-occurrence matrix of identified mutations; positive co-occurrence (blue): significantly co-occur; negative co-occurrence (red). Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

Pathogenicity assessment revealed a striking imbalance between gene classes: while nearly all mutations in non-HRR genes were classified as pathogenic or likely pathogenic (94.1%), only one-third of HRR gene variants met these criteria ($p < 0.0001$), with the remainder designated as variants of uncertain significance (VUS) **(Figure 13B)**. This uncertainty was especially pronounced for *BRCA2*, highlighting a key limitation of mutation-based HRD classification from ctDNA alone. Structural mapping showed that many HRR variants localized to protein interaction domains **(Figure 14B)**, suggesting functional relevance despite uncertain annotation and underscoring the need for integrative genomic and functional approaches to resolve HRD status.

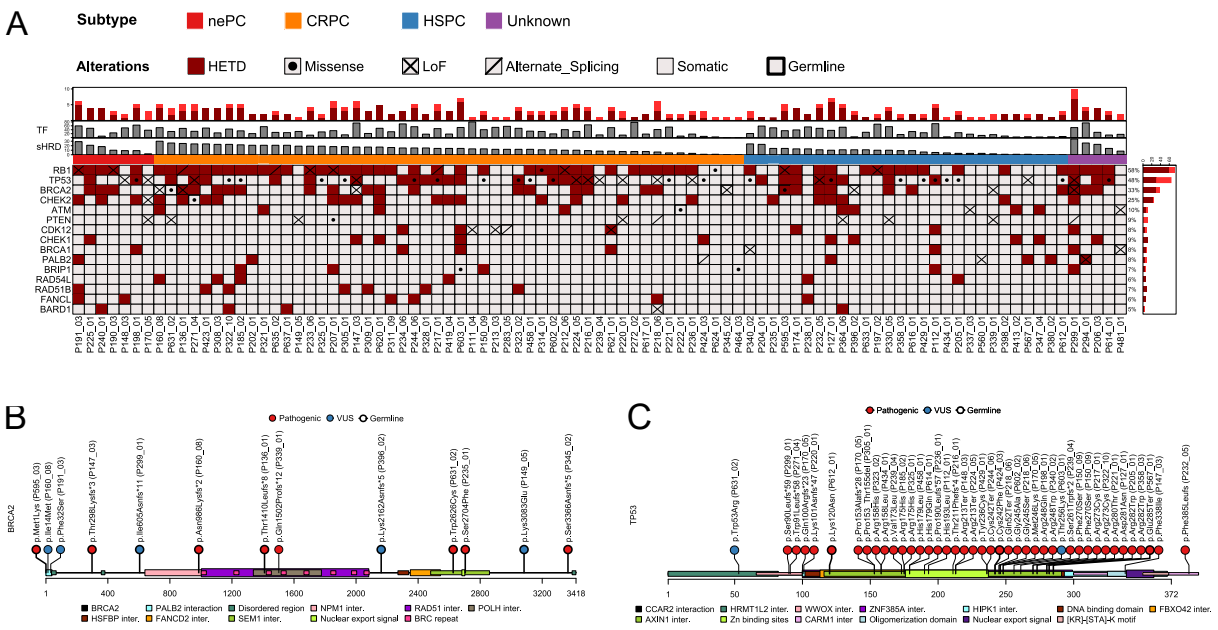


Figure 14: Somatic landscape and mutation distribution of HRD-associated genes in plasma cfDNA.

A) Oncoprint showing pathogenic mutations and copy number alterations across selected HRD-related genes in cfDNA from PCa patients. Mutation types are annotated by shape (e.g., missense, Loss-of-function (LoF), splicing), and heterozygous deletion (HETD) is annotated by background color. Bold outline indicates a germline variant. Samples are annotated by tumor fraction (TF), sHRD score, and clinical subtype. **B)** Lollipop plot of *BRCA2* mutations annotated by protein domain. Each lollipop represents a unique somatic or germline variant, colored by pathogenicity (pathogenic: red; VUS: blue), with germline variants indicated by bold outlines. Protein interaction domains and functional regions are shown along the BRCA2 protein axis. **C)** Lollipop plot of *TP53* mutations visualized along the protein structure. Pathogenic and VUS variants are colored and labeled as in B, with structural and interaction domains annotated. Germline variants are shown with bold outlines. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.

Clonality assessment further distinguished the nature of alterations: somatic events in *TP53*, *RB1*, and *CDK12* were predominantly clonal, whereas subclonal variants were disproportionately enriched in HRR genes and often required manual review due to borderline VAFs **(Figure 13C-D)**. Clonal and subclonal calls were assigned by comparing each variant's VAF to the sample TF; clonal variants had VAFs within ±10% of the TF-scaled expectation, whereas subclonal variants had markedly lower VAFs relative to TF.

Variants clustering around ~50% VAF were considered supportive of heterozygous germline status only in samples with low TF, whereas variants observed in samples with intermediate TF were treated as potentially ambiguous and therefore prioritised for orthogonal assessment in WBC DNA with Sanger sequencing. Germline origin was common among HRR genes, particularly *BRCA2*, for which 70.6% of variants were germline, followed by *ATM*, *BRCA1*, *BRIP1*, and *RAD54L*. In contrast, alterations in *TP53*, *RB1*, and *PTEN* were almost exclusively somatic **(Figure 13C-D)**. This divergence highlights

fundamentally different biological roles: stable, early driver mutations in core tumor suppressors versus inherited or late-emerging alterations affecting DNA repair capacity.

To extend the HRR gene analysis, WES data were examined for mutations in Fanconi anemia (FANC) pathway genes **(*Figure 15).*** Pathogenic and likely pathogenic variants were identified in several core members, including *FANCA*, *FANCL*, *FANCD2*, and *FANCM*, occasionally co-occurring with *BRCA2* or *PALB2* alterations. These findings suggested that disruption of the Fanconi complex contributes to HRD in PCa and highlighted the potential value of including FANC genes in cfDNA panels to improve HRD detection and therapeutic stratification.

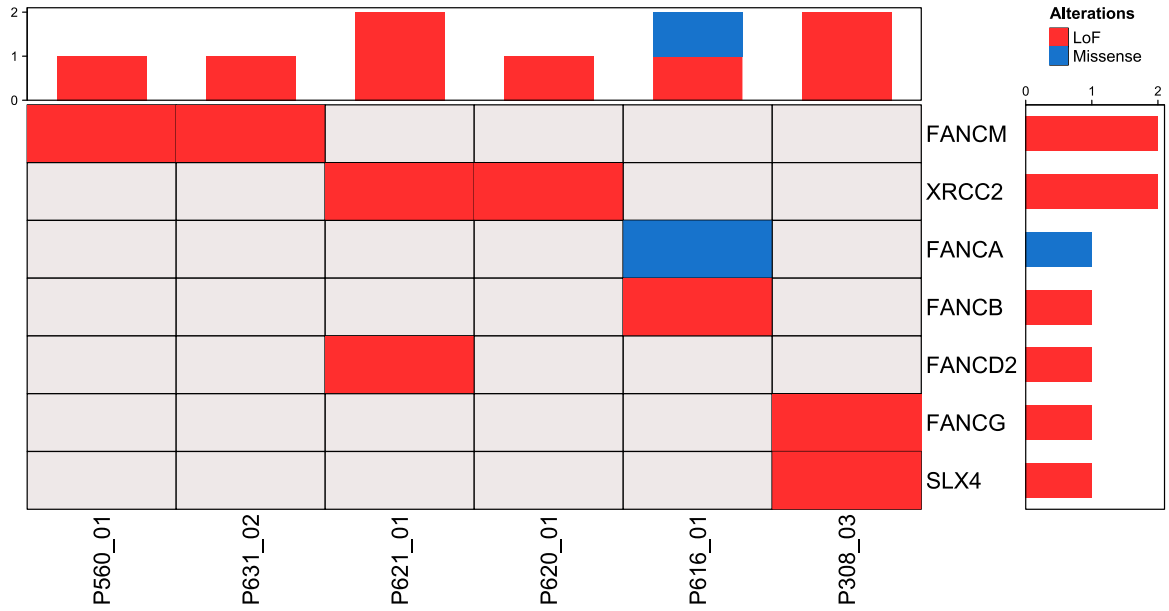


Figure 15: Somatic landscape of Fanconi anemia (FANC) pathway genes in PCa WES data.

Oncoprint summarizing pathogenic and likely pathogenic mutations in Fanconi pathway genes across the WES cohort. Each column represents a patient, and each row corresponds to a gene. The top barplot indicates the number of mutated genes per patient, while the right barplot shows the number of patients affected per gene. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.

Table 2: Co-occurrence matrix of pathogenic mutations

Reproduced by Vlachos et al. (180) with permission from the original authors, under the terms of the Creative Commons CC BY license.

	RB1	TP53	BRCA2	CHEK2	ATM	PTEN	CDK12	CHEK1	BRCA1	PALB2	BRIP1	RAD54L	RAD51B	FANCL	BARD1
RB1	0.00	0.02	0.11	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
TP53	0.02	0.00	0.05	0.06	0.00	0.10	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.00	0.03
BRCA2	0.11	0.05	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHEK2	0.00	0.06	0.00	0.00	0.00	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
ATM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00
PTEN	0.05	0.10	0.19	0.09	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.10
CDK12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
CHEK1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BRCA1	0.00	0.03	0.00	0.00	0.25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
PALB2	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BRIP1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RAD54L	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
RAD51B	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
FANCL	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
BARD1	0.00	0.03	0.00	0.00	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

*Jaccard indices for co-occurrence of pathogenic mutations (intersection/union across samples); non-pathogenic entries were treated as 0 and the diagonal is suppressed. For example, a value of 0.25 indicates substantial co-occurrence (25% of samples with a pathogenic alteration in either gene carried alterations in both), whereas a value of 0.1 indicates weak overlap (most alterations occurred in different samples). Values near 0 reflect little-to-no co-occurrence.

Molecular co-occurrence analysis **(Figure 13E, Table 2)** revealed distinct molecular architectures underlying GIS. *BRCA2* mutations frequently co-occurred with *RB1* and *PTEN* deletions, suggesting a multi-hit, high-genomic-instability subtype. Conversely, *TP53* mutations were inversely associated with both *RB1* and *BRCA2* alterations, pointing to distinct molecular or evolutionary trajectories. *CDK12* mutations were mutually exclusive with all other major events. Loss of *CDK12* is associated with focal tandem duplications rather than classical homologous recombination deficiency, generating a unique GIS pattern and immunogenic landscape that is mechanistically and clinically distinct from *BRCA*-driven HRD. These results supported a model in which *BRCA*-driven HRD represented one of several non-overlapping molecular routes to GIS in advanced disease.

4.3 Copy number alterations and the GIS phenotype

Large-scale SCNAs are a defining feature of HRD and can be robustly quantified in cfDNA using WGS. Genome-wide $\log_2$ ratio profiles revealed a pronounced contrast between HRD and HR-proficient tumors. In a representative *BRCA2*-mutant case, extensive arm-level gains and losses spanned nearly the entire genome, reflecting cumulative DNA repair failure and ongoing chromosomal missegregation **(Figure 16A)**. In contrast, an HR-proficient tumor that lacked alterations in HRR or *TP53* pathway genes exhibited a largely flat copy number profile with only a few SCNA **(Figure 16B)**. These divergent SCNA landscapes underscored the value of cfDNA WGS as an orthogonal readout of HRD that complements mutation-based approaches. By capturing the integrated consequences of defective DNA repair across the genome, CNA profiling enabled functional inference of HRD even when individual gene alterations were ambiguous or present at low allele fractions, supporting its utility for plasma-based patient stratification.

Beyond canonical HRD signatures, distinct focal tandem duplications (FTDs) were observed in tumors harboring CDK12 mutations (**Figure 16*C*).** These FTDs manifested as narrow, high-amplitude copy number gains on specific chromosomes, consistent with previous reports linking CDK12 loss to a tandem duplication–rich phenotype. Although the overall shallowHRD (sHRD) scores in CDK12-mutant samples were moderate, the SCNA pattern was qualitatively unique, distinguishing these cases from BRCA- or TP53-driven instability.

To systematically compare GIS across molecular contexts, the cohort was stratified into five subgroups based on mutation status: (1) *BRCA1/2* or *PALB2*, (2) other HRR genes, (3) *CDK12*, (4) *TP53/PTEN/RB1*, and (5) samples without known pathogenic mutations. TF estimates varied substantially across groups **(Figure 16D)**, with *TP53/PTEN/RB1*mutant cases showing the highest median TF%, consistent with increased ctDNA shedding and tumor burden.

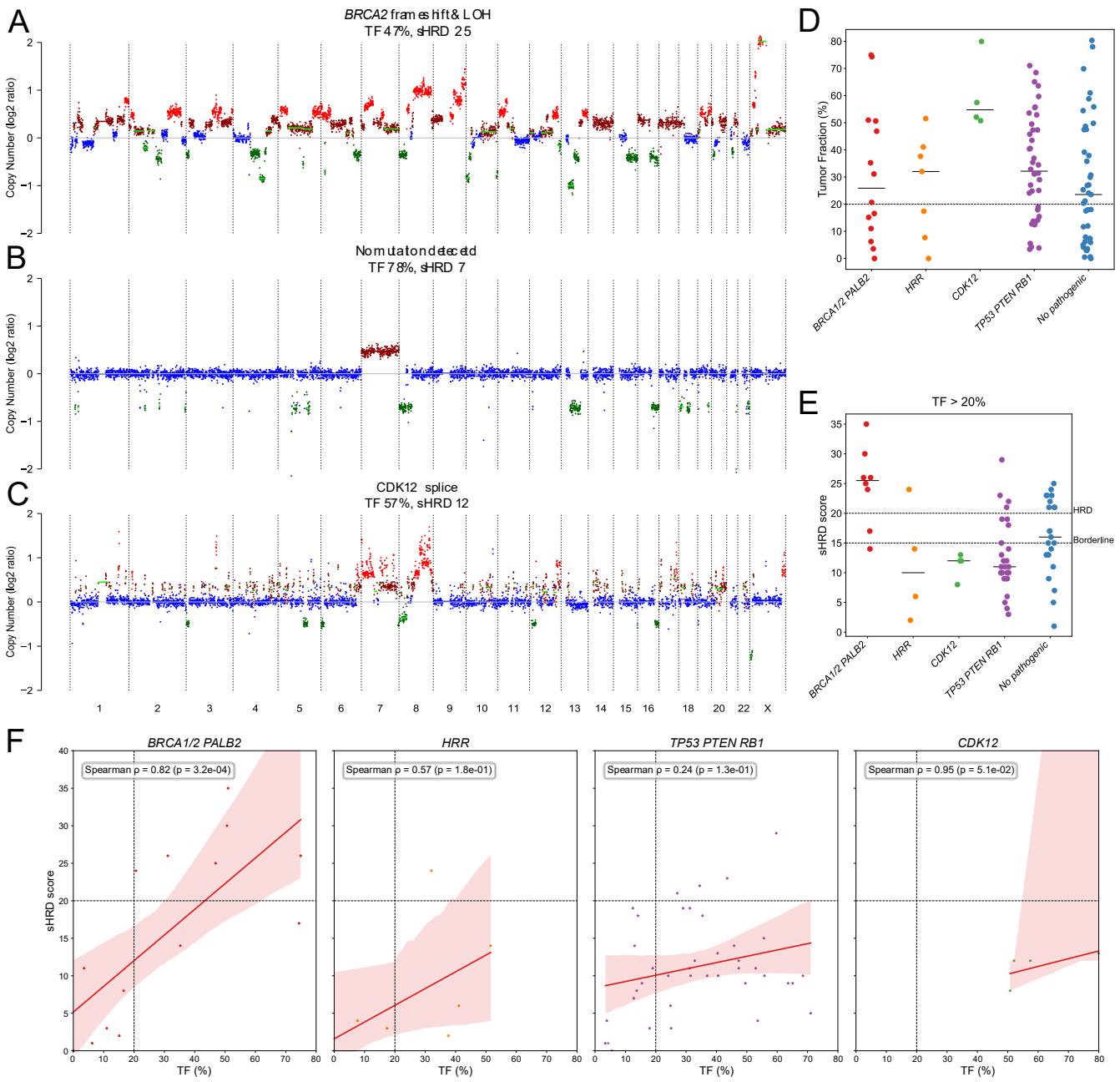


Figure 16: GIS and copy number features associated with HRD in cfDNA.

A) *Representative copy number profile of a BRCA2-mutated (HRD) sample* ***B)*** *Representative copy number profile of an HR-proficient PCa samples* ***C)*** *Distinct focal tandem duplications (FTDs), characteristic of CDK12-mutated tumors* ***D)*** *Distribution of tumor fraction (TF%) across five mutation-defined groups: BRCA1/2 or PALB2, other HRR genes, CDK12, TP53/PTEN/RB1, and samples without pathogenic mutations.* ***E)*** *shallowHRD scores (sHRD) in the same groups, limited to samples with TF > 20%. Dashed lines indicate sHRD thresholds of 15 and 20.* ***F)*** *Correlation between TF% and sHRD score within each genomic subgroup. Dashed lines mark TF = 20% and sHRD = 20. Spearman correlation statistics are reported per panel. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

As expected, GIS, quantified by sHRD score, was most pronounced in the *BRCA1/2/PALB2* group (median sHRD = 25.5), with 75% of samples exceeding the HRD-positive threshold of 20 **(Figure 16E)**. In contrast, tumors harboring mutations in other HRR genes exhibited intermediate sHRD scores, comparable to those observed in samples without detectable pathogenic mutations or in *TP53/PTEN/RB1*-mutated tumors, indicating an absence of widespread chromosomal instability in these groups. However, the *TP53/PTEN/RB1* group was heterogeneous, with some samples exhibiting elevated sHRD, likely due to secondary genomic events. *CDK12*-mutant samples showed intermediate sHRD values, reflecting the localized nature of tandem duplications rather than widespread chromosomal loss.

Correlation analysis between TF and sHRD within each subgroup further highlighted these differences **(Figure 16F)**. Strong positive correlations were observed in *BRCA*-mutant(Spearman $\rho > 0.82$, $p < 0.0001$), suggesting that increasing ctDNA shedding parallels escalating GIS in these contexts. In contrast, *CDK12*-mutant and HRR-wild-type tumors showed weak or absent correlations, consistent with CNA burden being uncoupled from overall tumor fraction in these subtypes.

At the individual gene level, *BRCA2* alterations were significantly associated with elevated sHRD scores ($p < 0.05$, two-sided Mann–Whitney U test; **Figure 17A**), reinforcing their central role in driving genome-wide instability. Genome-wide correlation of focal copy number events with sHRD score further localized key contributors to chromosomal instability **(Figure 17B)**, with strong associations at 13q (including *RB1*), 10q (*PTEN*), and 8q (*MYC*). These loci overlapped with the most recurrent alterations in subgroup-specific CNA profiles, underscoring their biological relevance and diagnostic value as markers of high-HRD states.

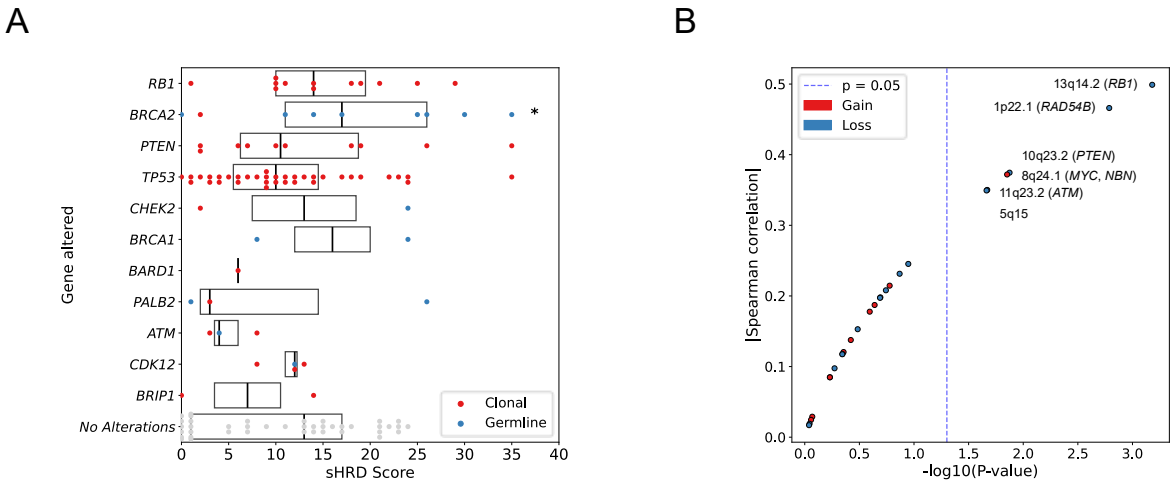


Figure 17: Association between sHRD score and genomic alterations.

A) *Distribution of sHRD scores across samples harboring pathogenic alterations in HRR-related genes. Each point represents one sample, colored by mutation type: clonal (red) or germline (blue). Boxplots summarize the distribution of sHRD scores for each gene, and samples without any pathogenic alterations are shown as a reference group ("No Alterations"). Asterisks indicate statistically significant differences (Mann–Whitney U test $p < 0.05$).* ***B)*** *Volcano-style plot showing the association between focal copy number events and sHRD score across the genome. Each point represents a genomic locus, with the x-axis indicating significance ($-\log_{10}p$) and the y-axis showing the absolute Spearman correlation between copy number and sHRD score. Red dots indicate gains, blue dots indicate losses. A dashed vertical line marks the $p = 0.05$ significance threshold. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

To explore chromosomal distribution of SCNAs in greater detail, we generated genome-wide SCNA frequency maps across 5 Mb bins for each major genomic subgroup **(Figure 18)**. *BRCA1/2/PALB2*-mutated tumors exhibited a high prevalence of arm-level deletions, particularly on 13q and 10q, which harbor *RB1* and *PTEN*, respectively. In contrast, *TP53*-mutated samples displayed more focal and heterogeneous SCNA patterns, while *CDK12*-mutant and HRR-wild-type samples showed fewer high-frequency alterations. Stratification by heterozygous deletions (HETD) further highlighted regions of consistent copy number loss shared across *BRCA*- and *TP53*-altered subtypes **(Figure 18C)**. A consensus SCNA track visualized at the bottom of each panel revealed recurrently altered loci, supporting subgroup-specific patterns of chromosomal instability.

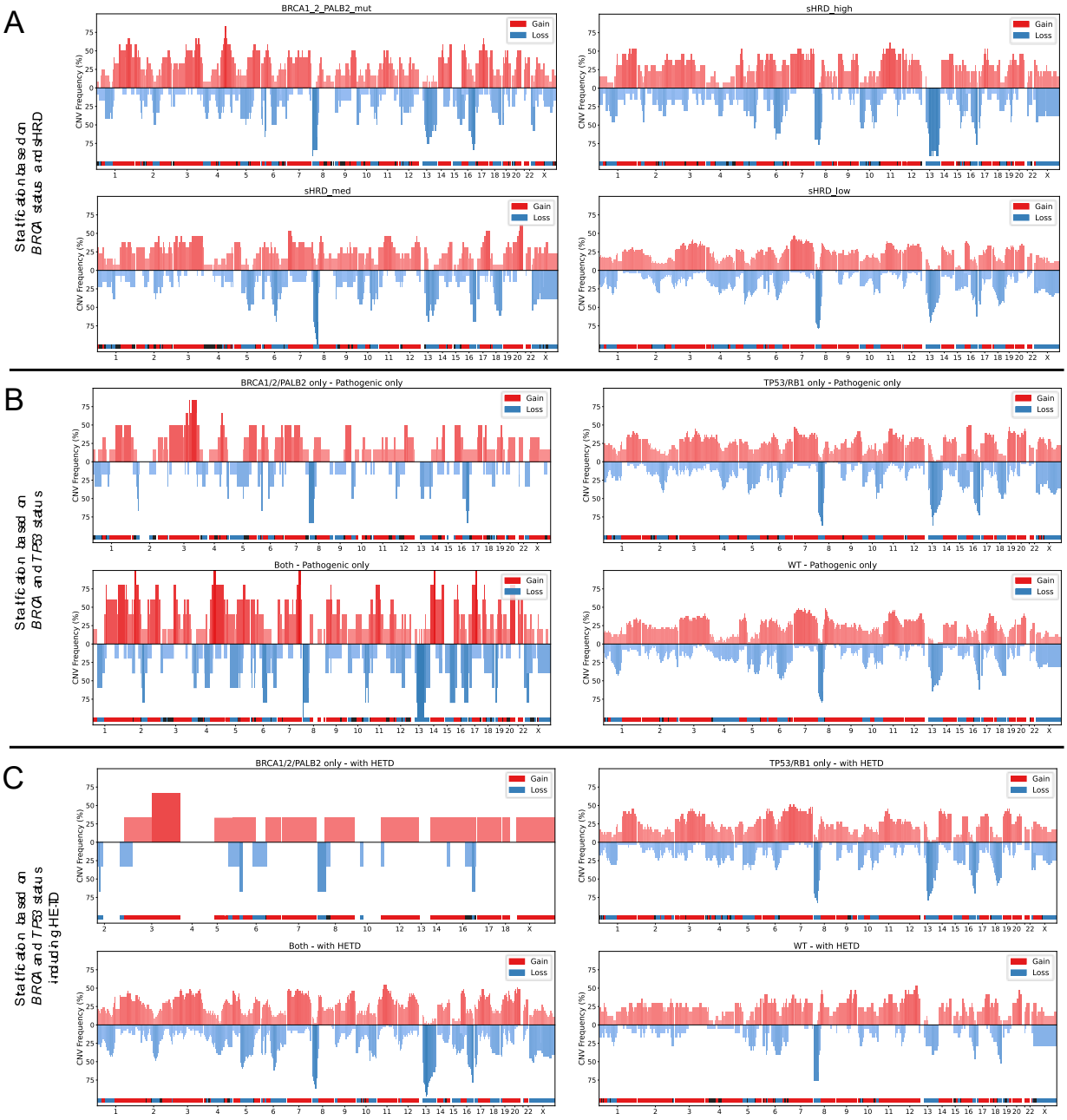


Figure 18: Genome-wide copy number alteration frequencies across cfDNA PCa subgroups.

Each panel shows genome-wide copy number alteration (CNA) frequencies across 5 Mb bins along chromosomes 1–22 and X. Red bars represent the percentage of samples with copy number gains; blue bars represent copy number losses. A horizontal consensus track at the bottom of each panel marks regions with predominant gains (red), losses (blue), or neutral (gray) calls based on ≥10% frequency difference. Chromosome boundaries are indicated along the x-axis. Included samples exhibit tumor fraction (TF) > 10% ***A)*** *CNA frequency plots for all high-tumor-fraction samples grouped by sHRD status (high, medium, low) and BRCA1/2/PALB2 mutation status.* ***B)*** *CNA profiles for subgroups defined by pathogenic mutations only, comparing BRCA1/2/PALB2-only, TP53/RB1-only, both, and wild-type (WT) cases.* ***C)*** *CNA profiles for the same subgroups as B, but including cases with heterozygous deletions (HETD) in the mutation group definitions. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

Complementing these findings, SV analysis of the high-coverage WGS dataset identified 7,171 events affecting coding regions, with large deletions comprising the dominant class. Notably, no recurrent gene fusions involving HRR genes were detected, suggesting that gene disruption in this context is not primarily mediated by balanced rearrangements. Instead, we observed frequent multi-gene deletions spanning key components of the HRR pathway, including *BRCA1/2, PALB2, RAD51B/C/D, FANCD2, FANCM, ATR*, and *CDK12*. These extensive deletions often encompassed multiple adjacent genes, consistent with the arm-level and focal copy number losses identified in SCNA analyses. Together, these results indicate that large-scale genomic loss represents a predominant structural mechanism underlying HRD in PCa, reinforcing the notion that functional HRR deficiency can arise through cumulative chromosomal attrition rather than discrete, gene-specific rearrangements **(Figure 19).**

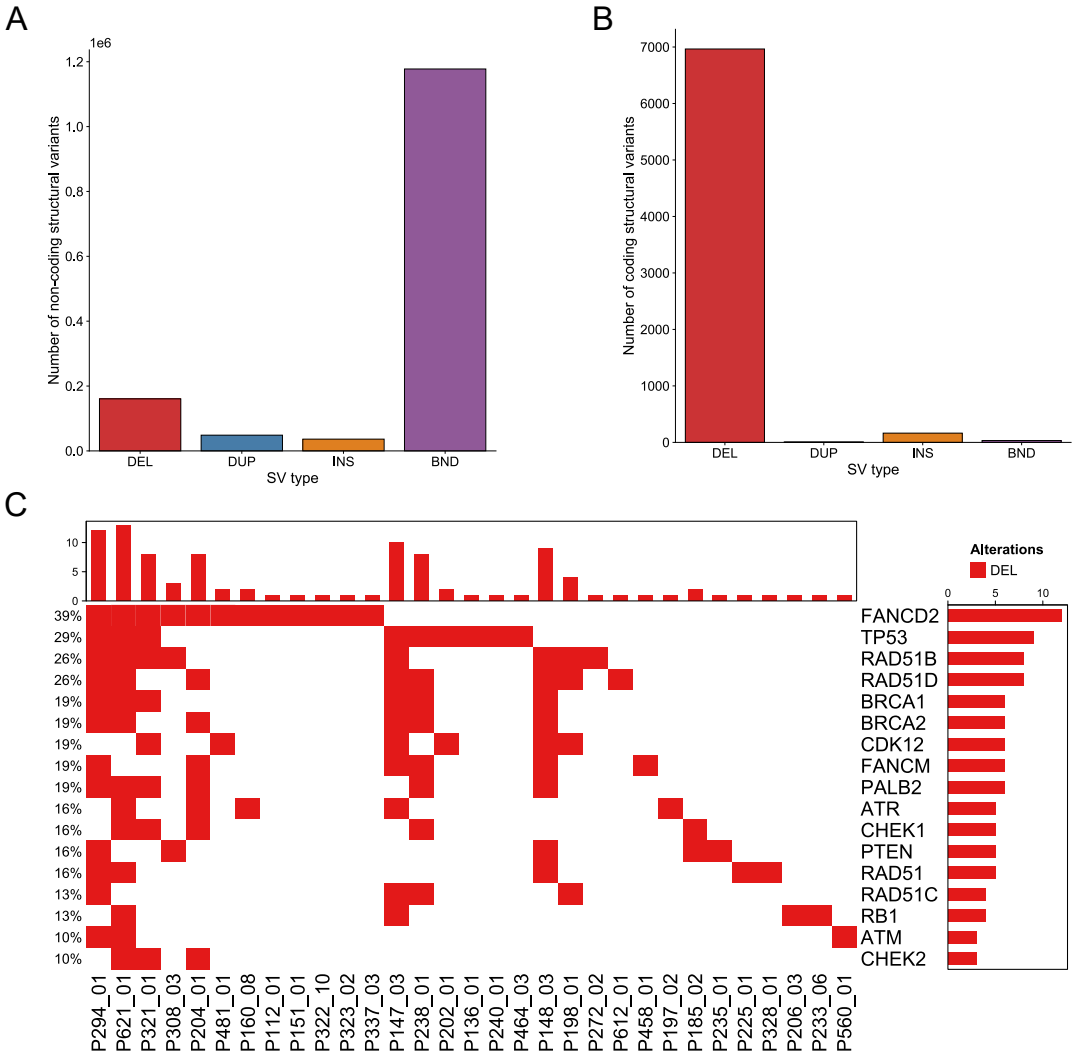


Figure 19: Structural variant landscape in cfDNA high-coverage WGS samples.

A) *Distribution of structural variants (SVs) detected across cfDNA samples, classified by type; deletions, duplications, insertions, and breakends.* ***B)*** *Coding-region subset of SVs, illustrating the predominance of large deletions relative to other variant types.* ***C)*** *Oncoprint summarizing homologous recombination repair (HRR) genes affected by structural variants. Each cell denotes the presence of a deletion (red), duplication (blue), insertion (orange), or translocation/breakend (purple). Multi-gene deletions frequently involved BRCA1/2, PALB2, RAD51 paralogs, and other HRR regulators, whereas no recurrent gene fusions were identified. Duplications, insertions, and other rearrangements within HRR genes were comparatively rare. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

Altogether, these data demonstrated that cfDNA-based SCNA profiling can effectively differentiate molecular subtypes of PCa and capture the functional consequences of HRD. *BRCA*-driven tumors were characterized by broad arm-level SCNAs and high sHRD scores, *CDK12* tumors displayed focal tandem duplications, and *TP53*-altered cases fell along a spectrum of instability. This supported a multi-dimensional classification framework that incorporated not just mutation status, but the structural genomic architecture inferred from plasma.

4.4 Clinical implications of GIS and HRD-related alterations on survival

To evaluate the prognostic relevance of GIS and HRD-related alterations in cfDNA, we performed survival analyses incorporating TF, sHRD scores, and mutation-based subgrouping. Kaplan-Meier analysis revealed that patients with TF ≥ 5% had significantly worse OS compared to those with lower TF values (median OS: 7 vs. not reached; HR = 7.71, $p < 0.0001$) **(Figure 20A)**. This association was independent of clinical subgroup and remained significant in multivariable models, supporting the value of TF as a dynamic, non-invasive biomarker of disease burden and prognosis. PFS was also reduced in the high-TF group (median PFS: 2 vs. 5 months; HR = 2.33, $p = 0.009$) **(Figure 20E)**, consistent with faster clinical progression in patients with elevated ctDNA levels.

We then assessed survival outcomes based on the presence of key genomic alterations. Alterations in TP53 and/or RB1, whether through mutation or deletion, were strongly associated with poor OS (median: 6 vs. 33 months; HR = 3.25; $p < 0.0001$) **(Figure 20B)**, reflecting their central role in aggressive tumor phenotypes and therapeutic resistance. Similarly, patients harboring BRCA1, BRCA2, or PALB2 mutations exhibited shorter OS (median: 5 vs. 11 months; HR = 1.87; $p = 0.006$) **(Figure 20C)**, although this subgroup did not show significant differences in PFS, potentially due to treatment heterogeneity or limited sample size.

To further dissect the interaction between these pathways, we stratified patients into four combined mutation groups: (i) no pathogenic mutations, (ii) *TP53*/RB1-only, (iii) *BRCA1/2/PALB2*-only, and (iv) dual mutation cases **(Figure 20D)**. Median OS was longest in patients without detectable pathogenic mutations (33 months), followed by the BRCA-only group (20 months), *TP53/RB1*-only group (9 months), and combined alteration group, which had the poorest survival (3 months; HR = 3.59; $p < 0.0001$). Although the comparison between *TP53*/*RB1*-only and *BRCA*-only groups did not reach statistical significance ($p = 0.082$), the overall trend suggests additive effects of these pathways on disease progression.

These findings highlighted the clinical relevance of ctDNA-derived biomarkers for prognostication in advanced PCa. TF, sHRD score, and HRR/*TP53* pathway alterations each provided independent and complementary information on disease aggressiveness and outcome, supporting their integration into future risk stratification frameworks. We next examined whether GIS, quantified by the sHRD metric, had similar clinical implications. Patients with sHRD scores below 20 had significantly longer OS than those with high sHRD (median OS: 10 vs. 3 months; HR = 2.40, $p = 0.001$) **(Figure 21A)**. Interestingly, sHRD score did not correlate with PFS (HR = 1.14, $p = 0.67$) **(Figure 21B)**, suggesting that GIS is more reflective of long-term disease biology than short-term treatment response. These findings reinforced the concept that genomic scarring reflects cumulative tumor evolution and vulnerability, particularly in HRD-positive disease.

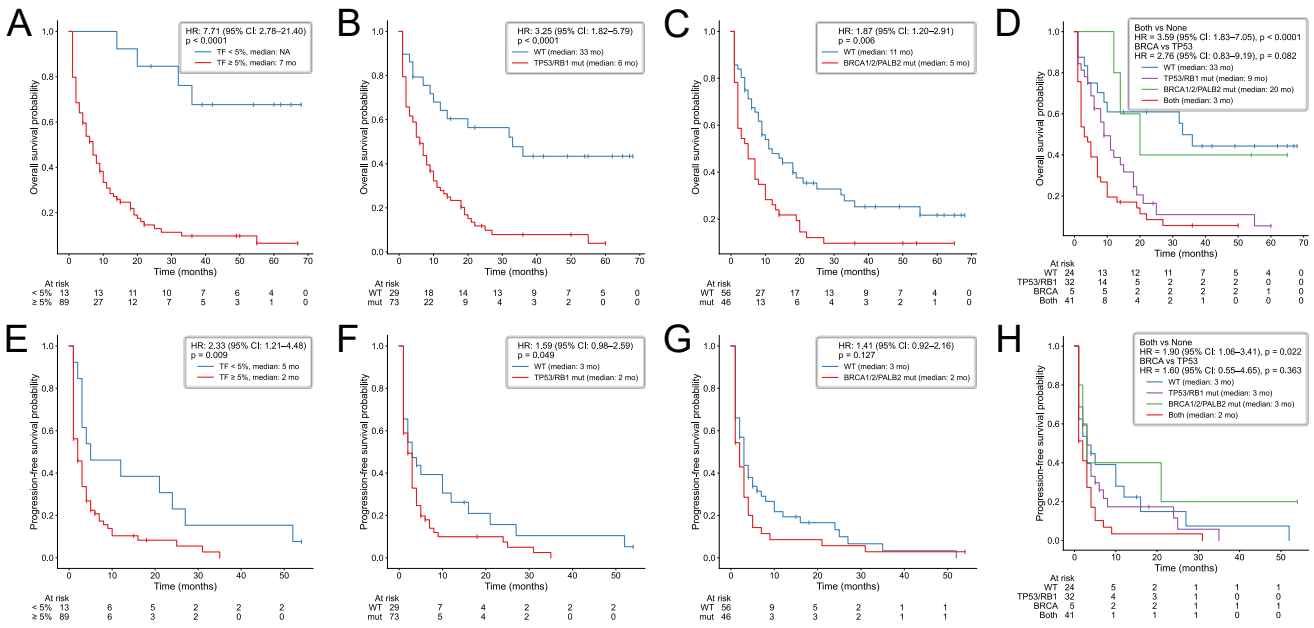


Figure 20: Survival analysis by ctDNA burden and HRD-related mutations in PCa.

*Overall survival (OS) of patients stratified by **A)** tumor fraction (TF), **B)** presence of TP53 or RB1 mutations **C)** presence of BRCA/PALB2 mutations, and **D)** no pathogenic mutations detected, detected TP53/RB1 mutations, detected BRCA/PALB2 only, and detected TP53/RB1/BRCA/PALB2 mutations. Progression-free survival (PFS) of patients stratified by **E)** tumor fraction (TF), **F)** presence of TP53 or RB1 mutations **G)** presence of BRCA/PALB2 mutations, and **H)** no pathogenic mutations detected, detected TP53/RB1 mutations, detected BRCA/PALB2 only, and detected TP53/RB1/BRCA/PALB2 mutations. Median survival times are indicated in the legend. Hazard ratios (HR), 95% confidence intervals (CI), and log-rank p-values were derived from Cox proportional hazards models and pairwise log-rank tests. Censoring is shown with vertical tick marks, and at-risk counts are displayed below each curve. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

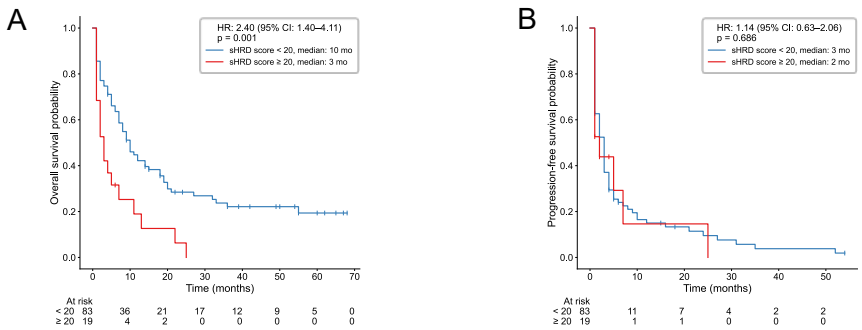


Figure 21: Kaplan–Meier survival analysis stratified by GIS (sHRD score).

A) *Overall survival (OS) and* ***B)*** *progression-free survival (PFS) for patients with available clinical follow-up, stratified by sHRD score: low (<20, blue) vs. high (≥20, red). Median survival times are indicated in the legend for each group. Hazard ratios (HR), 95% confidence intervals (CI), and log-rank p-values were computed using Cox proportional hazards models. Censoring is marked with vertical tick marks, and at-risk counts are shown below the time axis. Survival curves were estimated using the Kaplan–Meier method. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

4.5 Homologous recombination specific signatures can be detected in ctDNA to potentially guide therapy

To further characterize the mutational processes shaping the cfDNA genomes of advanced PCa, we performed comprehensive mutational signature analysis across 21 tumor–germline matched WES data sets. Somatic single-nucleotide variants (SNVs), insertions and deletions (indels), and allele-specific copy number alterations were analyzed and decomposed using COSMIC v3.4 reference signatures, applying strict refitting via the MutationalPatterns and SigProfiler frameworks.

Signature contributions were aggregated and visualized for three distinct signature classes: single-base substitutions (SBS), indels (ID), and copy number (CN) features **(Figure 22A-C)**. Among SBS signatures, SBS3-the hallmark of BRCA-related HRD-was enriched in samples harboring *BRCA2* or *PALB2* mutations **(Figure 22A)**. Other recurrent SBS signatures included SBS1 (age-related) and SBS5 (clock-like), observed broadly across all samples. Indel signatures ID6 and ID8, both linked to defective DNA repair pathways, were also elevated in HRD-positive tumors, particularly those with biallelic BRCA2 loss.

CN signature analysis highlighted the frequent contribution of CN19, CN17, and CN25, which have been associated with HRD-like profiles in previous prostate and ovarian cancer studies (191). In contrast, CDK12-mutant tumors exhibited a distinct CN signature composition, lacking CN19 and CN25

but showing modest contributions from CN4 and CN14, consistent with focal duplications and tandem repeat expansions characteristic of this subtype.

To visualize the diversity and structure of mutational processes, we performed t-distributed stochastic neighbor embedding (t-SNE) on the SNV and ID mutations. This unsupervised dimensionality reduction approach clustered samples into distinct genomic groups, reflecting differences in underlying DNA repair defects **(Figure 22D-E)**. HRD-positive tumors with *BRCA1/2* or *PALB2* mutations and elevated sHRD scores formed a cohesive cluster marked by high SBS3 and ID6 activity. *CDK12*-mutant samples occupied a separate region of the embedding, consistent with their unique mutational architecture, whereas *TP53*-only and HRR-wild-type tumors displayed more heterogeneous distributions with minimal HRD-associated signature activity. These patterns further supported the association between specific mutational signatures and defined genomic subtypes in PCa.

Taken together, these findings demonstrated that mutational signatures associated with defective DNA repair, particularly SBS3 and ID6 could be reliably inferred from cfDNA using WES. These signatures distinguish *BRCA*-deficient tumors from other PCa subtypes and offer a molecular basis for classifying HRD in liquid biopsy material. Moreover, the distinct signature profiles observed in *CDK12*-mutant tumors highlighted the potential of mutational signatures to capture biologically and clinically relevant subtypes, informing future strategies for genomic stratification and therapeutic targeting.

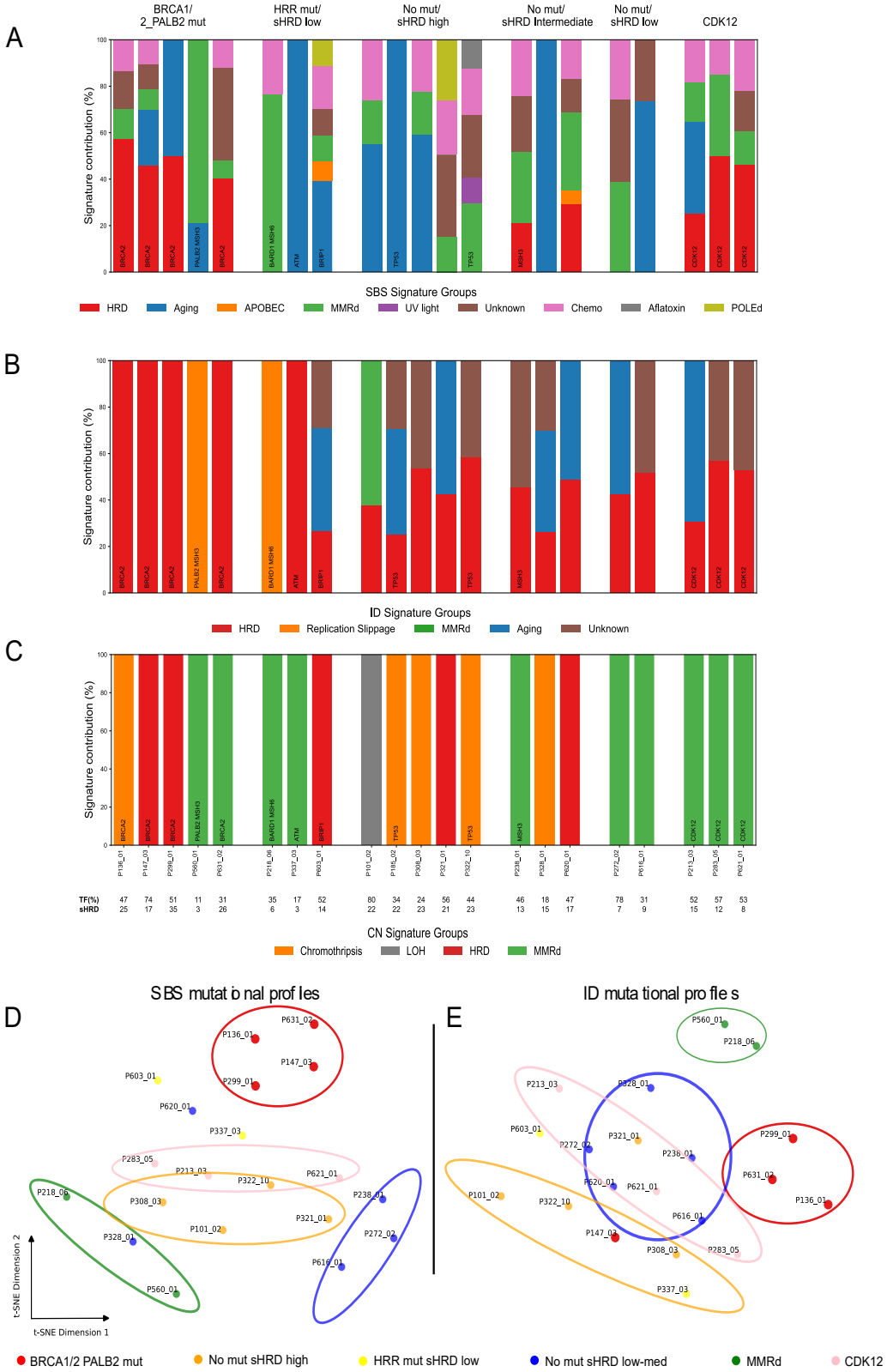


Figure 22: Mutational signatures reveal distinct genomic footprints of HRD, MMR deficiency, and CDK12 inactivation.

A) *Single base substitution (SBS) signatures across mutation-defined and sHRD-stratified groups.* ***B)*** *Indel (ID) signature composition per sample.* ***C)*** *Copy number (CN) signature contributions reveal enrichment of CN25. t-SNE clustering of* ***D)*** *SBS and* ***E)*** *ID signatures separate HRD, CDK12, and MMRd cases into partially distinct groups. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

4.6 HRD detection from cfDNA using fragmentomic features

Beyond mutation and copy number features, cfDNA fragmentomic features - particularly fragment length and fragment end characteristics - offer a complementary window into tumor-intrinsic chromatin organization and DNA repair status. To evaluate whether such fragmentomic signals were correlated with HRD, we extracted fragment length profiles and 5′ end sequence motifs from high-quality cfDNA WGS data. Analyses focused on fragments between 80 and 400 base pairs in length, with lengths normalized and binned into nine 32-base windows to quantify variability within each region. These bin-wise Z-score standard deviations served as measures of fragment length dispersion, a property previously linked to nucleosomal organization and cell-of-origin (121,143).

Fragment length variability was increased in samples harboring *BRCA2* mutations and elevated GIS, whereas samples lacking HRD-associated alterations showed more constrained fragment size distributions **(Figure 23A)**. Among the high sHRD group, many samples showed increased variability across multiple bins, suggesting disrupted chromatin organization. By contrast, *TP53*- and *CDK12*-mutated samples exhibited intermediate or heterogeneous patterns, possibly reflecting a mix of fragmentomic phenotypes unrelated to canonical HRD.

We also assessed 5′ end motif usage by extracting the four-base sequences flanking the start of each cfDNA fragment. A total of 256 possible 4-mers were collapsed into 25 clusters using spectral clustering. After filtering low-abundance motifs, 19 clusters remained and were used as input for downstream modeling.

To evaluate the translational relevance of our findings, we constructed a cfDNA fragmentomics-based classifier to identify HRD cases using a penalized logistic regression framework. Given the limited number of HRD-positive samples (n = 7), a conservative modeling strategy was employed to minimize overfitting. Feature extraction was restricted to healthy control cfDNA profiles (n = 14) to prevent data leakage, resulting in a final set of nine fragment-length and nineteen end-motif variables. Model performance was assessed through seven-fold cross-validation across 41 PCa samples. The classifier based solely on fragment-length features achieved an AUC of 0.79 and an average precision (AP) of 0.69, whereas the end-motif-only model yielded more modest results (AUC = 0.61, AP = 0.53). When both feature categories were combined, the AUC remained at 0.79, and the AP improved to 0.70, demonstrating that integrating fragment-length and end-motif information enhances predictive accuracy **(Figure 23B)**.

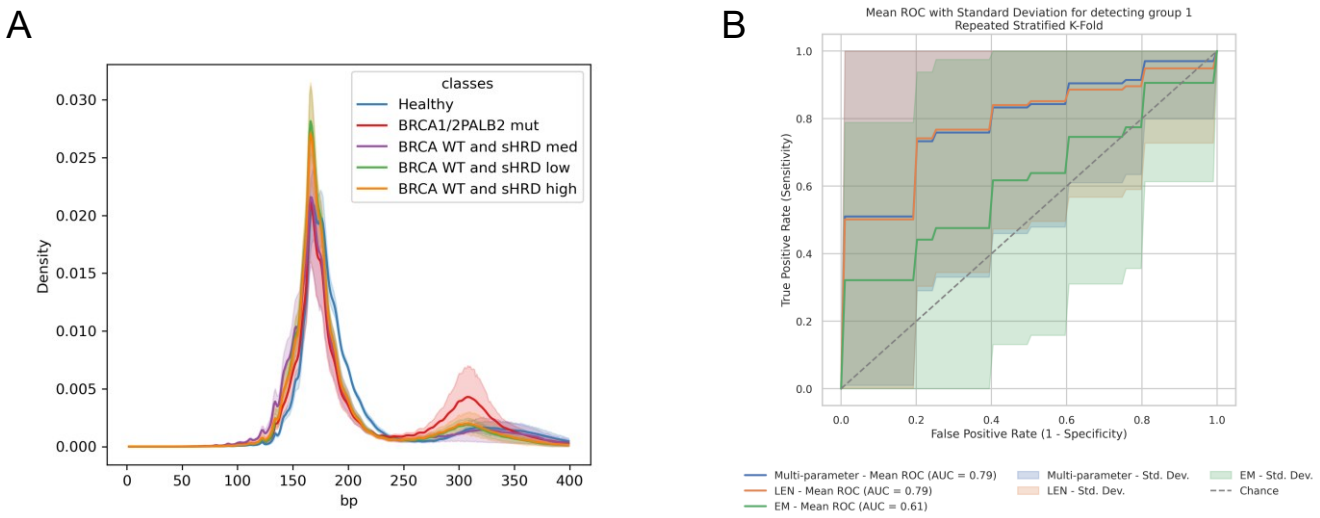


Figure 23: Fragmentomics reveal non-mutational footprints of HRD in cfDNA.

A) Fragment length density plots comparing cfDNA fragmentation patterns across groups. B) Receiver operating characteristic (ROC) curves evaluating the performance of a logistic regression classifier trained on fragment length (LEN), end motif (EM), or combined features to detect HRD-positive cases. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.

To account for the potential influence of TF on cfDNA fragment size profiles, samples were stratified by mutation status and compared with *BRCA*/*PALB2*-wildtype tumors and healthy controls. Fragment counts were subsampled to achieve balanced coverage across groups. TF did not differ significantly between HRD and HR-proficient, indicating that neither sequencing depth nor TF introduced measurable bias in fragment size distributions **(Figure 24A-B)**.

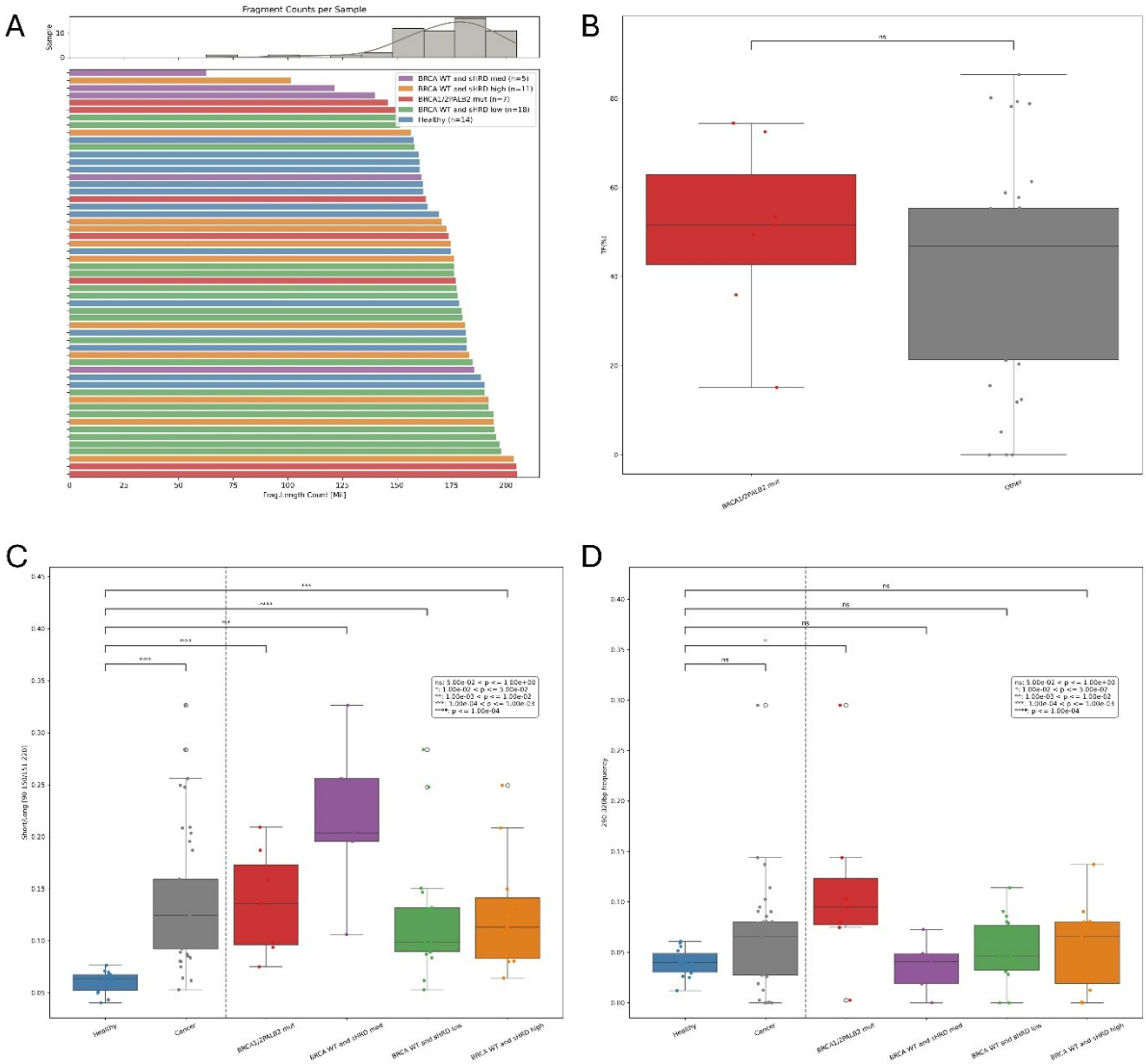


Figure 24: Fragment length distributions stratified by HRR mutation status.

(A) *Total fragment counts across all cfDNA samples, demonstrating comparable sequencing depth among groups.* ***(B)*** *Tumor fraction comparison between HRR-mutated (BRCA1/2 or PALB2) and HRR-wildtype tumors, indicating no significant difference.* ***(C)*** *Global enrichment of short cfDNA fragments in all PCa groups relative to healthy controls.* ***(D)*** *Selective enrichment of 290–320 bp fragments observed exclusively in BRCA1/2- or PALB2-mutated cases, corresponding to dinucleosome-associated fragment lengths. Adjusted p-values were calculated using two-sided Mann-Whitney U tests with Benjamini–Hochberg false-discovery-rate correction, as detailed in the Methods. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

Together, these results demonstrated that cfDNA fragment length dynamics and end motif usage encode information about tumor chromatin structure and repair competence. These features could complement traditional mutation-based analyses and may offer an orthogonal, low-input approach to identifying HRD in cfDNA, particularly when tumor-derived DNA is sufficiently abundant.

4.7 A subset of zinc finger transcription factor binding sites is differentially accessible in HRD cfDNA.

Chromatin accessibility patterns inferred from cfDNA coverage could serve as an indirect readout of transcriptional regulation and epigenomic dysregulation. To explore whether chromatin structure differs between HRD and HR-proficient tumors, we quantified cfDNA fragment coverage at predefined TFBS using a windowed approach centered on known motifs. Coverage signals were normalized and compared across genomic subgroups to identify regions of differential accessibility.

Co-mutated tumors harboring concurrent alterations in *BRCA2* and *TP53*/*RB1* were categorized within the *TP53*/*RB1*/NE group, as combined loss of *TP53* and *RB1* is known to promote lineage plasticity and NE transdifferentiation. These epigenetic and transcriptional changes could alter nucleosome organization and potentially obscure HRD-specific chromatin accessibility signals.

Global analysis revealed that HRD-positive samples, defined by the presence of *BRCA1*, *BRCA2*, or *PALB2* mutations and/or high GIS, exhibited significantly reduced cfDNA coverage around specific zinc finger (ZNF) transcription factor binding sites compared to HRD-negative cases (Benjamini-Hochberg adjusted $p < 0.05$, two-tailed Mann–Whitney U test) **(Figure 25)**. These differences suggested that *BRCA*-deficient tumors may undergo chromatin remodeling or transcriptional silencing at ZNF loci, a finding consistent with reports linking ZNF proteins to DNA repair (192–198).

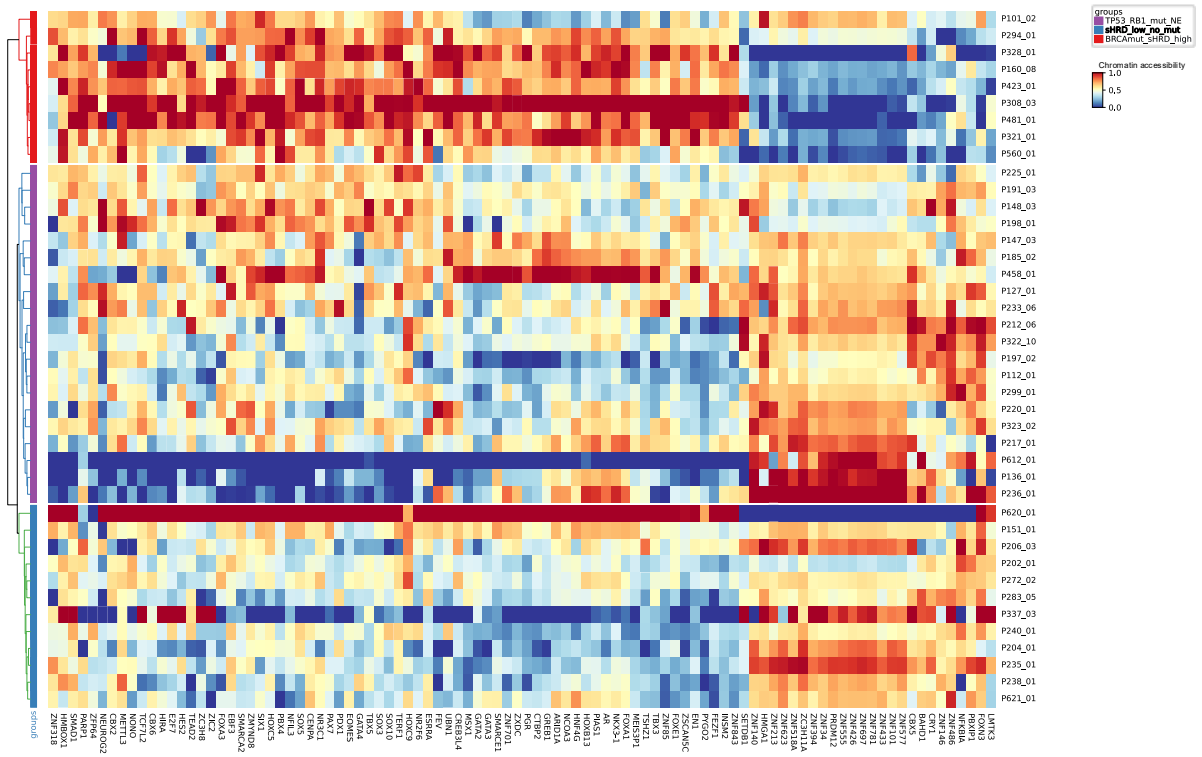


Figure 25: Chromatin accessibility reveal potentially mechanistical insight into HRD.

Heatmap showing differential chromatin accessibility at transcription factor binding sites (TFBS) across PCa subgroups. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.

This reduction in coverage around ZNF binding sites was consistently observed when HRD cases were compared both to HR-proficient and to *TP53/RB1*/NE-altered tumors, with the strongest signal attenuation evident in the latter group **(Figure 26)**. These findings suggested that ZNF-associated regions represented a distinct fragmentomic and chromatin signature of HRD, potentially reflecting altered nucleosomal architecture or repressive chromatin remodeling in *BRCA2*-deficient PCa.

These accessibility shifts may reflect broader alterations in chromatin architecture that accompany *BRCA* loss, such as changes in replication timing, DNA methylation, or the formation of repressive domains. From a diagnostic perspective, TFBS accessibility features derived from cfDNA could offer an additional layer of HRD detection, especially in combination with mutational and fragmentomic signals.

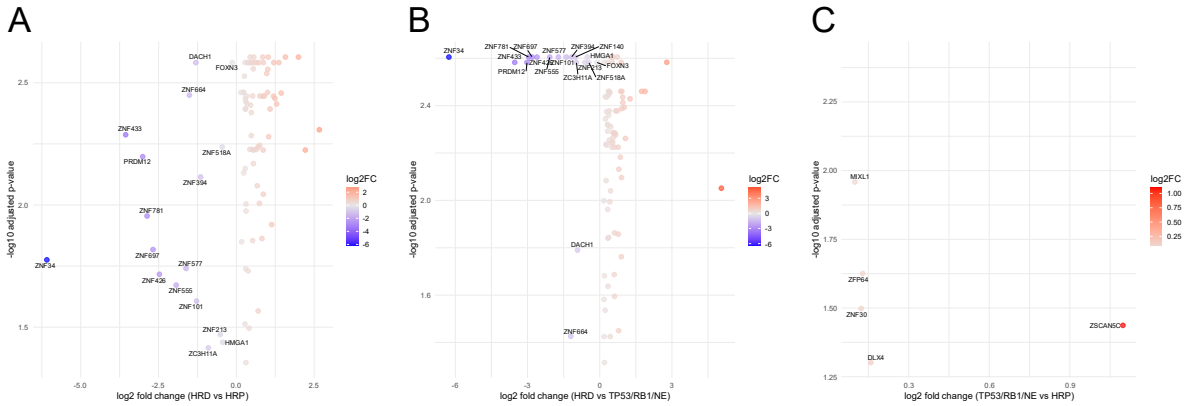


Figure 26: Differential chromatin accessibility at transcription factor binding sites (TFBS).

(A) *Comparison of BRCA2/sHRD-high versus HRR-proficient tumors.* ***(B)*** *Comparison of BRCA2/sHRD-high versus TP53/RB1/NE tumors.* ***(C)*** *Comparison of TP53/RB1/NE versus HRR-proficient tumors. Each panel displays the log_2 fold change (x-axis) against the $-log_{10}$ adjusted p-value (y-axis) for chromatin accessibility at zinc finger (ZNF) TFBS inferred from cfDNA. Adjusted p-values were calculated using two-sided Mann–Whitney U tests with Benjamini–Hochberg false-discovery-rate correction, as described in the Methods. Each point represents a single TFBS, with color indicating both the direction and magnitude of the log_2 fold change. Gene symbols are annotated for significantly altered loci. Collectively, these comparisons reveal consistent reductions in accessibility at specific ZNF TFBS in HRD cfDNA, with the most pronounced effects observed in BRCA2/sHRD-high cases relative to TP53/RB1/NE tumors. Figure and legend adapted from Vlachos et al. (180). Reproduced with permission from the original authors, under the terms of the Creative Commons CC BY license.*

5 Discussion

Metastatic PCa remains a lethal and molecularly heterogeneous disease, with treatment resistance frequently driven by alterations in DNA repair and other core oncogenic pathways (1). HRD has emerged as a clinically actionable vulnerability, particularly in tumors harboring *BRCA1/2* or *PALB2* alterations, but its detection is challenged by limited access to high-quality tumor tissue in advanced disease (199). Liquid biopsy approaches using ctDNA offer a minimally invasive means to profile HRD-related genomic and structural features longitudinally, providing a powerful framework for patient stratification and therapeutic monitoring in precision oncology (200).

The present study demonstrates that a multimodal analysis of ctDNA, including mutational and copy number profiling, mutational signatures, fragment length variability, and chromatin accessibility, can be used to identify HRD-positive PCa with high resolution. Beyond single-gene mutation testing, our approach captures diverse genomic consequences of HRD and reveals molecular subtypes with distinct structural and epigenetic features. By integrating these layers, we propose a framework for HRD

detection that can complement or even substitute for tissue-based assays in patients with sufficient ctDNA content. This has potential implications for biomarker-driven therapy selection, clinical trial stratification, and the development of real-time, blood-based companion diagnostics (201).

5.1 Interpretation of key findings

A central observation of this work is that BRCA-driven HRD in PCa manifests most consistently as a genome-wide structural phenotype rather than solely as a mutational event. Tumors harboring *BRCA1*, *BRCA2*, or *PALB2* alterations exhibited high sHRD scores, extensive arm-level copy number alterations, and frequent multi-gene deletions affecting core HRR components. These features collectively reflect cumulative chromosomal attrition arising from defective DNA repair and provide a functional readout of HRD that extends beyond individual gene mutations. Importantly, tumors with alterations in other HRR genes such as *ATM*, *CHEK2*, or *RAD51* paralogs generally lacked this widespread GIS, despite harboring pathogenic or likely pathogenic variants. This dissociation highlights a key limitation of mutation-based HRD classification and suggests that not all HRR gene alterations confer equivalent functional consequences. Our data therefore support a model in which HRD should be defined as a structural and evolutionary state of the genome, rather than inferred solely from gene-level alterations.

We observed that HRD-associated alterations could be broadly stratified into three genomic subtypes: (i) classical *BRCA1/2* or *PALB2*-mutated tumors, (ii) non-BRCA HRR-altered cases with variable GIS, and (iii) *CDK12*-mutated tumors with distinct copy number and mutational profiles. *BRCA2* emerged as the most frequently mutated HRR gene **(Figure 14A)**, often co-occurring with *PTEN* or *RB1* loss, a pattern consistent with prior tissue-based studies (202). This suggests that *BRCA*-driven HRD often emerges within a broader context of tumor suppressor disruption, potentially accelerating GIS and disease aggressiveness. In contrast, *TP53* mutations were inversely correlated with *BRCA2* alterations, suggesting distinct evolutionary trajectories and mutually exclusive genomic programs **(Figure 13E)** (203).

While mutations in canonical HRR genes (e.g. *BRCA1*/*2* and *PALB2*) remain primary biomarkers of HRD, increasing attention has turned to the impact of non-HRR tumor suppressors including *TP53*, *PTEN*, and *RB1* on shaping the HRD phenotype and treatment responses. *TP53* is the most frequently mutated gene in human cancers and is critical for maintaining genomic integrity. Although TP53 does not directly participate in HRR, loss of p53 function permits accumulation of DNA damage and exacerbates GIS. Indeed, *TP53* mutations often co-occur with HRR gene alterations in advanced tumors (including PCa), suggesting cooperative genomic programs (204,205). Some studies further indicate that intact p53 activity is associated with better responsiveness to PARPi, whereas *TP53*

inactivation can attenuate the efficacy of these drugs (206). However, the precise mechanistic role of *TP53* in modulating HRD remains under investigation, and its value as an independent predictor of PARPi response is still inconclusive.

PTEN, a well-known tumor suppressor and negative regulator of PI3K–AKT signaling, has also been implicated in HRD biology (207). Loss of PTEN function is associated with defects in the repair of DNA double-strand breaks by homologous recombination (208). Mechanistically, PTEN deficiency leads to reduced formation of RAD51 nuclear foci in response to DNA damage, a hallmark of impaired HR repair (208). This RAD51 dysfunction induced by PTEN loss creates a state of "functional HRD" even in the absence of *BRCA1/2* mutations. Consistent with this, preclinical studies have demonstrated that *PTEN*-null or -mutant cells are hypersensitive to PARPi (208). Nevertheless, clinical data on the PTEN-PARPi relationship remain discrepant. The impact of *PTEN* loss on therapeutic response appears to be context-dependent, varying by tumor type and the presence of co-mutations in other pathways (e.g. cell cycle or AR signaling). Thus, while *PTEN*-inactivated tumors often exhibit GIS, *PTEN* status alone has not yet been adopted as a standalone biomarker for PARPi benefit in the clinic.

RB1, another key tumor suppressor frequently inactivated in prostate and other cancers, exhibits a nuanced, context-specific role in HRD. In certain settings, loss of *RB1* can counterintuitively restore HR repair capacity. A recent mechanistic study in PCa models showed that *RB1* inactivation upregulates E2F1, which in turn drives re-expression of *BRCA2* and other HR genes (209). Through this E2F1-mediated pathway, *RB1* loss effectively reactivates HR, allowing tumor cells to efficiently repair DNA damage despite HRR gene defects. Consequently, RB1-deficient tumors can become resistant to PARPi even in the presence of genomic "scarring" indicative of HRD. This phenomenon was demonstrated by Miao *et al.*, who showed that *RB1* loss leads to upregulation of *BRCA2*, thereby restoring homologous recombination capacity. Importantly, knockdown of *BRCA2* resensitized *RB1*-deficient PCa cells to PARPi, confirming that elevated *BRCA2* expression following *RB1* loss is a key driver of resistance (209).

Adding further complexity, other evidence suggests that *RB1* loss can synergize with BRCA dysfunction in different biological contexts. Notably, a recent multi-cohort analysis of tubo-ovarian high-grade serous carcinoma found that patients with concurrent *RB1* loss and germline *BRCA1/2* mutations had significantly improved overall survival compared to those with BRCA mutations alone (210). In this study, co-occurrence of *RB1* loss and *BRCA* deficiency was associated with exceptionally long survival (median OS 9.3 years vs. 3.1 years in *BRCA*-only cases) and enhanced responsiveness to DNA-damaging chemotherapies. Functional experiments indicated that *BRCA*-mutant ovarian cancer cells became more sensitive to cisplatin and paclitaxel upon *RB1* knockout. Moreover, tumors harboring *BRCA*/*RB1* co-loss displayed a distinct immunogenic profile with increased CD8+ T-cell infiltration and

upregulation of interferon signaling pathways. These co-deficient tumors showed transcriptional evidence of cGAS–STING pathway activation and elevated antigen presentation, suggesting that *RB1* loss in a BRCA-deficient background may heighten tumor immunogenicity alongside chemosensitivity (210). Such findings hint that in certain contexts, *RB1* inactivation does not simply negate HRD, but rather shifts the tumor into a state that is more vulnerable to immune-mediated and DNA-damaging therapies, even if it is less responsive to PARP inhibition.

These observations highlight the complex and sometimes opposing influences of *TP53*, *PTEN*, and *RB1* on the HRD phenotype and treatment outcomes. Although these genes are not classical components of the HRR pathway, their inactivation can indirectly promote GIS, alter DNA repair capacity, or modulate sensitivity to therapy. Notably, *RB1* loss emerges as a double-edged sword: on one hand enabling HR repair (through E2F1-driven *BRCA2* upregulation) and thereby conferring PARPi resistance, but on the other hand potentially cooperating with BRCA deficiencies to create an immunostimulatory, therapeutically exploitable state. These context-dependent effects underscore the importance of integrated genomic and transcriptomic profiling when evaluating HRD status. In a precision oncology setting, assessing the status of *TP53*, *PTEN*, and *RB1* alongside canonical HRR genes may improve patient stratification for PARPi versus alternative or combination therapies. However, further clinical validation is required before non-HRR gene alterations can be reliably used as standalone biomarkers for HRD or for predicting PARPi response.

Besides canonical *BRCA1/2* and *PALB2* alterations, our WES analysis identified pathogenic variants in several FANC pathway genes, including *FANCA*, *FANCL*, *FANCD2*, and *FANCM* **(Figure 15)**. The Fanconi complex plays a central role in resolving DNA interstrand crosslinks and coordinating homologous recombination repair through its interaction with BRCA-associated proteins. Disruption of this pathway has been linked to GIS, replication stress, and increased susceptibility to DNA-damaging agents in multiple tumor types (211). The presence of FANC gene mutations in our cohort, occasionally co-occurring with *BRCA2* or *PALB2* loss, suggests that defects in the Fanconi network may represent an additional route to HRD in PCa. These findings support the inclusion of FANC genes in cfDNA diagnostic panels, which could improve the sensitivity of HRD detection and broaden patient eligibility for PARP- or platinum-based therapies.

Our data also highlight the central role of SCNAs in driving GIS in HRD-positive cases. Samples with *BRCA2* or *PALB2* mutations displayed highly fragmented copy number profiles and elevated sHRD scores, even when only monoallelic loss was observed **(Figure 16)**. This suggests that partial loss of HRR function can still have profound effects on genome stability, in line with emerging evidence questioning the necessity of strict biallelic inactivation in PCa (212). However, it is possible that in some cases the second hit was missed due to epigenetic regulation such as promoter hypermethylation.

Moreover, high-sHRD cases were enriched for deletions at canonical instability loci (e.g., *RB1*, *PTEN*, *ATM*) and showed novel recurrent events, including deletions at 5q15 **(Figure 17B)**, implicating these regions as additional contributors to the HRD phenotype (213).

Complementing the mutational data, SV analysis of hcWGS revealed extensive large-scale deletions affecting key HRR loci, including *BRCA1/2*, *PALB2*, *RAD51B/C/D*, *FANCD2*, *FANCM*, *ATR*, and *CDK12* **(Figure 19)**. These multi-gene deletions indicate that structural gene loss is a major driver of HRD in PCa, consistent with prior reports that deletions contribute more frequently to HRD than gene fusions or small indels. The absence of recurrent fusion events in our dataset further supports this notion. In addition, the presence of rare, complex rearrangements suggests that HRD phenotypes can emerge from cumulative chromosomal instability rather than single-gene lesions. Comprehensive tumor-normal WGS or RNA-seq analyses, potentially with long-read sequencing technology, could help clarify these structural mechanisms, resolve complex rearrangements, and identify cryptic events that may be missed by tumor-only cfDNA assays. Together, these findings highlight the importance of integrating SV profiling with sequence-level mutation analysis to achieve a complete understanding of HRD architecture in PCa.

Beyond BRCA-related HRD, we identified *CDK12*-mutant tumors as a distinct molecular class with unique genomic hallmarks. These samples lacked classical HRD-associated mutational signatures but were enriched for FTDs, a known consequence of *CDK12* inactivation **(Figure 16C)** (214). These findings support prior observations that *CDK12* loss drives oncogenesis through fusion gene generation and enhanced neoantigen burden, which may confer sensitivity to immune checkpoint blockade rather than PARP inhibition (72). Importantly, *CDK12* mutations were mutually exclusive with *BRCA1/2* and *TP53* alterations, underscoring their role in a separate evolutionary pathway **(Figure 13E)**.

Mutational signature profiling revealed that tumors with *BRCA1/2/PALB2* mutations were enriched for single-base substitution signature SBS3 and indel signature ID6 **(Figure 22)**. While SBS3 is well established as a hallmark of HRD, our data show that it can also appear in *CDK12*-mutated samples, raising questions about the specificity of this signature in for HRD testing (40). Conversely, samples lacking HRR mutations but exhibiting high sHRD often showed low SBS3 signal, reinforcing the need for multimodal classification strategies. The t-SNE clustering of SBS and ID signatures confirmed that *CDK12*, MMR-deficient, and HRD tumors could be resolved into distinct clusters, validating the use of signature-based stratification in liquid biopsy material **(Figure 22D-E)**.

Finally, our analysis uncovered unexpected overlaps between HRD and mismatch repair deficiency (MMRd), with ~38% of WES-profiled samples showing features of both. This co-enrichment of HRD- and MMRd-associated signatures, such as SBS6 and CN25, suggests possible mechanistic interactions between these repair pathways, potentially mediated by altered mutational landscapes or

compensatory repair activation **(Figure 22A-C)** (215,216). These findings add to a growing body of evidence indicating that DNA repair deficiencies often converge, and that integrated profiling is essential for accurate tumor classification. This finding contradicts the classical notion of co-exclusivity between these two repair deficiencies (217).

5.2 Clinical implications of GIS and HRD-associated features

From a clinical standpoint, our results highlight the prognostic relevance of plasma-derived biomarkers, particularly TF and HRD-related genomic alterations. Patients with TF ≥5% had significantly worse OS and PFS **(Figure 20A,E)**, confirming that increased tumor burden results in poorer survival (218). Similarly, elevated sHRD scores (≥20) were associated with poorer OS, reinforcing the value of GIS as an independent prognostic marker **(Figure 21A-B)**.

Among genomic subtypes, mutations in *TP53* and *RB1* conferred the worst outcomes, particularly when present in combination with *BRCA1*, *BRCA2*, or *PALB2* alterations. These co-altered cases exhibited extremely poor survival, suggesting additive or synergistic disruption of DNA repair, checkpoint control, and apoptotic signaling. Conversely, patients with *BRCA*-pathway alterations alone had intermediate outcomes, while those without HRD-related mutations had the most favorable survival profiles **(Figure 20B,F)**. Notably, despite the prevalence of *BRCA* and *PALB2* mutations in our cohort, only two patients received PARPi therapy. This limited exposure likely underestimates the potential clinical benefit of PARPi in HRD-positive cases, especially given the genomic features observed. These findings underscore the clinical importance of accurate HRD detection to guide eligibility for DNA repair-targeted therapies.

Our analysis also recapitulated known clinical associations. Neuroendocrine differentiation, present in a small subset of patients, was associated with poor prognosis and reflects a transition to a more aggressive lineage (219). In contrast, castration-sensitive tumors, often harboring fewer alterations and lower TF, were linked to longer survival, highlighting the prognostic value of integrating molecular and clinical phenotyping **(Figure 12)**.

These results support a framework in which TF, sHRD, mutation profile, and molecular subtype are jointly evaluated to stratify risk and inform therapeutic decision-making. cfDNA-based analysis offers a minimally invasive route to capture these dimensions in real time, and may serve as a critical tool in personalizing treatment for advanced PCa. In future clinical workflows, plasma-based HRD classifiers could be applied at multiple points along the disease course to optimize therapy selection, monitor resistance evolution, and identify candidates for emerging HR-directed or immunotherapeutic strategies.

5.3 Fragmentomics and epigenomic markers of HRD

Recent studies employing integrated fragmentomics models have demonstrated high sensitivity for both cancer detection and molecular subtyping (144,220). Building on this framework, our cfDNA classifier incorporated fragment length and end-motif features to identify HRD-associated fragmentation patterns in PCa. We hypothesized that the altered chromatin organization characteristic of HRD would generate distinctive fragmentation profiles detectable in plasma cfDNA. Our results support this hypothesis and align with emerging mechanistic evidence that histone modifications, nucleosome positioning, and chromatin accessibility collectively shape cfDNA fragmentation signatures (221). These observations suggest that HRD-associated chromatin states can be inferred from cfDNA fragmentomic features, providing a functional, noninvasive window into tumor epigenetic architecture. Indeed, we observed that tumors with *BRCA1/2* or *PALB2* mutations exhibited an enrichment of dinucleosomal fragments (290–320 bp), a pattern absent in HRR-proficient and *TP53/RB1*-altered cases **(Figure 23A)**. This suggests aberrant nucleosomal spacing and reduced chromatin compaction, both of which are consistent with impaired DSB repair and disrupted nucleosome maintenance in HRD-positive contexts (222).

To translate these features into a predictive model, we trained a penalized logistic regression classifier using both fragment length variability and motif cluster frequencies as input features. The combined model demonstrated strong discrimination of HRD-positive versus HRR-proficient samples (AUC = 0.79, AP = 0.69; Figure 17B). Fragment length metrics alone outperformed end motif features (AUC = 0.61, AP = 0.53), but combining both modestly improved performance (AP = 0.70), indicating that these fragmentomic layers capture complementary aspects of the HRD phenotype **(Figure 23B)**.

Notably, prior to this work fragmentomics had not been commonly applied or reported in the context of HRD. Although early studies have shown that cfDNA fragmentation patterns can capture tumor-specific features, no published evidence has explicitly compared fragmentomic profiles between HR-deficient and HR-proficient tumors. This represents a critical knowledge gap, given the central role of HRD in the repair of DSB. It is biologically plausible that HRD tumors would shed cfDNA with distinctive fragmentation patterns, such as shorter average fragment lengths, greater fragment end variability, or specific mono- and dinucleotide motifs at cleavage sites, arising from the dysregulated nuclease activity and chromatin disorganization associated with defective DNA repair pathways (121). By demonstrating differences like an excess of dinucleosomal fragments in HRD cases, our study begins to fill this gap. Future investigations should further evaluate whether these fragmentomic differences can serve as reliable and specific biomarkers for HRD status in the liquid biopsy setting.

To further investigate epigenomic correlates of HRD, we analyzed TFBS accessibility using cfDNA coverage data **(Figure 24)**. We observed that HRD tumors, particularly those harboring *BRCA2*

mutations and high GIS, showed a consistent reduction in accessibility at ZNF transcription factor binding sites, including ZNF34 and ZNF781. This reduction was not observed in *TP53*/*RB1*-altered or neuroendocrine subtypes, suggesting that diminished ZNF accessibility reflects an HRD-specific chromatin signature rather than general tumor aggressiveness.

ZNF transcription factors are among the most abundant DNA-binding proteins in the human genome and play diverse roles in gene regulation, chromatin remodeling, and increasingly, DNA repair. Several ZNFs have been mechanistically linked to DSB repair. For example, ZNF281 and ZBTB24 facilitate NHEJ through interactions with XRCC4 and PARP1, while PHF6 promotes 53BP1 recruitment and checkpoint recovery at damage sites (192–194). Additionally, ZNF-domain-containing proteins like KDM2A and ZMYND8 regulate homologous recombination by promoting RAD51 and BRCA1 foci formation via histone methylation pathways (195,196).

Although ZNF34, ZNF433, and ZNF781 have not yet been directly linked to DNA repair, their consistent inaccessibility in HRD samples suggests a role in either parallel or regulatory components of the DNA damage response. It is plausible that epigenetic silencing or nucleosomal occlusion at ZNF TFBS interferes with the transcription of key repair mediators or chromatin modifiers. PRDM12, for instance, is a histone methyltransferase that facilitates H3K9 dimethylation, promoting transcriptional repression and potentially compounding repair inefficiency in these tumors (197). A recent systematic analysis further supports the role of ZNF proteins as chromatin-embedded sensors or effectors in the DSB response, often recruited in a PARP-dependent manner (198).

Collectively, these findings point to a broader model in which altered TFBS accessibility particularly at ZNF-associated loci acts as a measurable epigenetic footprint of HRD. This has two important implications. First, it highlights a mechanistic link between chromatin remodeling and defective homologous recombination. Second, it raises the possibility that such fragmentomic and accessibility-derived signals can be incorporated into cfDNA-based classifiers, especially in cases where mutation data alone are insufficient for classification. Whether this chromatin phenotype is reversible or therapeutically targetable through epigenetic reprogramming remains an open question but represents a promising direction for future HRD-targeted interventions.

Consistent with this model, emerging evidence indicates that HR-deficient tumors undergo distinctive chromatin and TF binding reprogramming. For instance, loss of *BRCA2* not only triggers GIS but also causes aberrant chromatin relaxation; BRCA2-deficient cells exhibit diminished heterochromatin marks (e.g. H3K9me3) and an opening of normally silenced genomic regions (223) This de-repression is accompanied by transcriptional shifts, including unscheduled activation of oncogenic pathways (notably, upregulation of MYC-driven gene targets upon biallelic *BRCA2* loss). Notably, such HRD-linked chromatin alterations are beginning to be detectable via liquid biopsy: high-resolution cfDNA

fragment analyses can capture tumor-specific nucleosome positioning and TF “footprints”, and recent nucleosome mapping studies in plasma have revealed accessible TF-binding sites reflective of tumor regulatory states (224,225). These findings strengthen the mechanistic link between defective homologous repair and epigenomic remodeling in cancer, suggesting that HRD-induced changes in TFBS accessibility, in both tumor tissue and circulating DNA, are tangible hallmarks of the HRD phenotype (226).

The convergence of these fragmentomic and epigenomic signals supports the notion that HRD not only alters the genetic landscape but also imposes a distinct physical and structural signature on tumor-derived cfDNA. Integration of fragmentomic features with mutation and copy number information improved classification performance and enabled more nuanced stratification of borderline or ambiguous cases.

5.4 Limitations

While this study presents a comprehensive multimodal cfDNA-based framework for characterizing HRD in PCa, several important limitations should be acknowledged. First, the approach is inherently biased toward samples with high TF, which are more likely to yield reliable mutation, copy-number, and fragmentomic data. Consequently, the method may perform suboptimally in early-stage or low-shedding tumors, limiting its generalizability across the full clinical spectrum. Future adaptations may require ultra-sensitive techniques, integration of complementary biomarkers, or probabilistic modeling approaches to extend applicability to minimal residual disease and early detection contexts.

Second, the number of confirmed HRD-positive cases, particularly those harboring clonal *BRCA1*, *BRCA2*, or *PALB2* mutations, was relatively small. This restricts statistical power for subgroup analyses and increases the risk of overfitting in the machine learning components of the fragmentomics classifier. To minimize this risk, we used a simple predefined logistic regression model without hyperparameter optimization and implemented repeated cross-validation to ensure that model performance was not driven by individual samples. Validation in independent, larger, and clinically diverse cohorts will nevertheless be essential to confirm reproducibility and generalizability.

Third, the targeted HRR panel employed in this study covered a restricted set of DNA repair genes and did not encompass the full spectrum of homologous recombination– and Fanconi anemia–related loci now known to contribute to HRD. As a result, rare or noncanonical drivers of homologous recombination deficiency may have been missed. Expanding future assays to include a more comprehensive HRR and DNA damage repair gene set could improve detection sensitivity and capture a broader range of clinically relevant alterations.

Fourth, the ground truth for HRD classification was defined primarily by cfDNA-detected *BRCA1/2* and *PALB2* mutations, as high-quality matched tumor tissue was available for only a subset of patients. This reliance may underestimate the prevalence of functional HRD, yet it highlights the clinical need for liquid biopsy based functional assays capable of complementing or replacing tissue-based testing when archival material is limited.

Fifth, while our analysis revealed characteristic fragment length variability, end-motif composition, and TFBS accessibility patterns associated with HRD, these signals are indirect and observational. We cannot infer causality from the chromatin-level features captured in cfDNA, nor exclude potential confounding factors such as cell-of-origin differences or chromatin disorganization secondary to necrosis. Future validation using orthogonal epigenomic approaches, such as cfChIP-seq, or ATAC-seq, will be crucial to confirm their mechanistic basis.

Sixth, the detection of complex structural variants and gene fusions affecting HRR genes was limited by the sequencing depth and design of the current cfDNA assay. While high-coverage WGS enabled partial identification of large-scale deletions, the resolution of balanced rearrangements or transcriptionally active fusions would require deeper whole-genome sequencing or complementary RNA-based approaches. Integrating such analyses in future studies could provide a more complete view of HRD-associated genomic architecture.

Seventh, the absence of treatment-linked outcome data limited our ability to evaluate the predictive value of the identified HRD markers. Although molecular profiles strongly suggested HRD in a subset of tumors, the lack of direct correlation with response to PARPi or platinum-based therapies precludes assessment of clinical benefit. Longitudinal sampling and treatment-matched cohorts will be needed to establish the translational utility of these biomarkers in therapeutic decision-making.

Finally, the study design did not include spatial tumor profiling or multi-analyte integration, precluding dissection of cfDNA signals by tumor origin or subclonal composition. Moreover, while SV analysis identified extensive deletions affecting HRR and Fanconi genes, and fragmentomic profiling revealed distinct chromatin signatures, additional integration with methylation, immune, or imaging-derived data could further enhance biological resolution and predictive accuracy.

Despite these limitations, the present findings underscore the feasibility and promise of multimodal cfDNA analysis for non-invasive HRD assessment. The combined evaluation of genetic, structural, and epigenetic features establishes a foundation for future tiered HRD testing frameworks are complemented by fragmentomic and chromatin accessibility metrics. Ongoing validation in independent cohorts and prospective clinical trials will be essential to confirm reproducibility, define performance thresholds, and support the integration of this approach into precision oncology workflows.

5.5 Future directions

This study demonstrates the feasibility of using multimodal cfDNA analysis to non-invasively characterize HRD in PCa. The next critical step will be the prospective validation of these findings in larger, clinically annotated cohorts, ideally incorporating serial plasma sampling during treatment to correlate cfDNA-derived biomarkers with therapeutic response and resistance trajectories, particularly in patients receiving PARPi or other DNA repair targeted therapies.

To extend analytical sensitivity and broaden clinical applicability, future work should explore repeated plasma sampling and the integration of orthogonal cfDNA-derived layers, including methylation, copy-number variation, and nucleosome occupancy, combined with advanced normalization and probabilistic modeling frameworks. A key methodological priority will be the systematic determination of the LOD for fragmentomic and HRD-associated features, especially in low-ctDNA contexts, to ensure their reliable application in diverse clinical settings.

Another promising avenue lies in extending these methods to earlier disease stages and minimal residual disease (MRD) contexts. At present, analyses rely on sufficient TF to capture mutational, structural, and fragmentomic signals. Advances in assay sensitivity, noise modeling, and background correction could enable detection of HRD-related cfDNA signatures in low-shedding tumors, providing opportunities for early therapeutic intervention or adaptive treatment strategies.

Further development of cfDNA-based classifiers should emphasize integration of multiple biomarker layers into a single unified model. Combining mutational status, GIS, fragment length variability, end-motif usage, and TFBS accessibility may enhance classification accuracy and robustness across variable cfDNA quality and tumor content. These integrated classifiers would support mutation-independent HRD detection, with potential applications in treatment eligibility assessment, resistance monitoring, and stratification for biomarker-driven clinical trials.

From a translational standpoint, we envision a tiered HRD assessment framework in clinical practice. Patients could first undergo targeted sequencing for HRR mutations and GIS evaluation from WGS, followed by fragmentomic and chromatin accessibility profiling as a complementary layer to refine HRD classification, particularly in ambiguous or discordant cases. Independent validation in prospective multicenter studies will be required to establish diagnostic performance, define standardized workflows, and confirm feasibility in real-world oncology settings.

Finally, expanding the molecular scope of cfDNA analysis beyond HR by incorporating MMR, replication stress, or chromothripsis-related signatures could enable comprehensive cfDNA-based profiling of GIS. As analytical pipelines mature, cfDNA has the potential to evolve into a standalone diagnostic substrate

for real-time genomic stratification in PCa, reducing reliance on tissue biopsies and facilitating dynamic, longitudinal monitoring throughout the disease course.

6 Conclusion

To the best of our knowledge, this is the first study to demonstrate that fragmentomic features and changes at transcription factor binding site accessibility are functionally linked to HRD in a cancer setting. By integrating mutational profiling, GIS, fragment length distributions, end-motif patterns, and chromatin accessibility from cfDNA, we establish a multimodal framework for liquid biopsy based HRD classification.

Whereas most current assays rely exclusively on gene mutation status or loss-of-heterozygosity metrics, our findings reveal that HRD-positive tumors exhibit distinct structural and epigenetic signatures that are detectable in circulation. Notably, we observed dinucleosome enrichment and characteristic end-motif shifts in *BRCA2*-deficient samples, together with a consistent loss of accessibility at ZNF binding sites, a previously undescribed feature in this context. These signals complement traditional genomic measures and likely represent the functional consequences of impaired DNA repair and altered chromatin organization.

Although this retrospective study was limited by sample size and the small number of patients treated with PARPi, it demonstrates the feasibility of classifier models that extend beyond sequence-based information to detect HRD with higher resolution. Fragmentomic and chromatin accessibility features offer a mutation-independent avenue for HRD assessment, particularly valuable in cases with ambiguous genotypes or low–variant allele fraction mutations.

In summary, our work underscores the potential of cfDNA as a comprehensive substrate for multimodal HRD detection in PCa. By capturing both structural and chromatin-level features in plasma, this approach bridges the gap between molecular complexity and clinical implementation-moving the field closer to personalized, minimally invasive precision oncology for patients with DNA repair–deficient PCa.

7 References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024 Feb;74(1):12–49. doi:10.3322/caac.21820 PubMed PMID: 38230766.

2. Global Cancer Facts & Figures 5th Edition [Internet]. American Cancer Society; 2024 [cited 2024 Jul 12]. Available from: https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/global-cancer-facts-and-figures/global-cancer-facts-and-figures-2024.pdf

3. Sawada N. Risk and preventive factors for prostate cancer in Japan: The Japan Public Health Center-based prospective (JPHC) study. Journal of Epidemiology. 2017 Jan;27(1):2–7. doi:10.1016/j.je.2016.09.001

4. Gleason DF, Mellinger GT. Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. J Urol. 1974 Jan;111(1):58–64. doi:10.1016/s0022-5347(17)59889-4 PubMed PMID: 4813554.

5. Epstein JI, Zelefsky MJ, Sjoberg DD, Nelson JB, Egevad L, Magi-Galluzzi C, et al. A Contemporary Prostate Cancer Grading System: A Validated Alternative to the Gleason Score. European Urology. 2016 Mar 1;69(3):428–35. doi:10.1016/j.eururo.2015.06.046

6. Gordetsky J, Epstein J. Grading of prostatic adenocarcinoma: current state and prognostic implications. Diagnostic Pathology. 2016 Mar 9;11(1):25. doi:10.1186/s13000-016-0478-2

7. Wyatt AW. Rare Genetic Drivers of Lethal Prostate Cancer. JAMA Oncology. 2023 Sep 21. doi:10.1001/jamaoncol.2023.3352

8. Jaratlerdsiri W, Jiang J, Gong T, Patrick SM, Willet C, Chew T, et al. African-specific molecular taxonomy of prostate cancer. Nature. 2022;609(7927):552–9. doi:10.1038/s41586-022-05154-6

9. Mucci LA, Hjelmborg JB, Harris JR, Czene K, Havelick DJ, Scheike T, et al. Familial Risk and Heritability of Cancer Among Twins in Nordic Countries. JAMA. 2016 Jan 5;315(1):68–76. doi:10.1001/jama.2015.17703

10. Bratt O, Drevin L, Akre O, Garmo H, Stattin P. Family History and Probability of Prostate Cancer, Differentiated by Risk Category: A Nationwide Population-Based Study. JNCI: Journal of the National Cancer Institute. 2016 Oct 1;108(10):djw110. doi:10.1093/jnci/djw110

11. Zhu Y, Wei Y, Zeng H, Li Y, Ng CF, Zhou F, et al. Inherited mutations in Chinese men with prostate cancer. Journal of the National Comprehensive Cancer Network. 2022;20(1):54–62.

12. Nyberg T, Tischkowitz M, Antoniou AC. BRCA1 and BRCA2 pathogenic variants and prostate cancer risk: systematic review and meta-analysis. British Journal of Cancer. 2022;126(7):1067–81. doi:10.1038/s41416-021-01675-5

13. Taylor RA, Fraser M, Rebello RJ, Boutros PC, Murphy DG, Bristow RG, et al. The influence of BRCA2 mutation on localized prostate cancer. Nature Reviews Urology. 2019;16(5):281–90. doi:10.1038/s41585-019-0164-8

14. Khan HM, Cheng HH. Germline genetics of prostate cancer. The Prostate. 2022 Aug 1;82(S1):S3–12. doi:https://doi.org/10.1002/pros.24340

15. Paluch-Shimon S, Cardoso F. PARP inhibitors coming of age. Nature Reviews Clinical Oncology. 2021;18(2):69–70. doi:10.1038/s41571-020-00452-2

16. Bhamidipati D, Haro-Silerio JI, Yap TA, Ngoi N. PARP inhibitors: enhancing efficacy through rational combinations. British Journal of Cancer. 2023;129(6):904–16. doi:10.1038/s41416-023-02326-7

17. Tolkach Y, Kristiansen G. The Heterogeneity of Prostate Cancer: A Practical Approach. Pathobiology. 2018;85(1–2):108–16. doi:10.1159/000477852

18. Rebello RJ, Oing C, Knudsen KE, Loeb S, Johnson DC, Reiter RE, et al. Prostate cancer. Nature Reviews Disease Primers. 2021;7(1):9. doi:10.1038/s41572-020-00243-0

19. Francini E, Taplin ME. Prostate cancer: Developing novel approaches to castration-sensitive disease. Cancer. 2017 Jan 1;123(1):29–42. doi:https://doi.org/10.1002/cncr.30329

20. Khalaf DJ, Annala M, Taavitsainen S, Finch DL, Oja C, Vergidis J, et al. Optimal sequencing of enzalutamide and abiraterone acetate plus prednisone in metastatic castration-resistant prostate cancer: a multicentre, randomised, open-label, phase 2, crossover trial. The Lancet Oncology. 2019;20(12):1730–9.

21. Humphrey PA. Histological variants of prostatic carcinoma and their significance. Histopathology. 2012 Jan 1;60(1):59–74. doi:https://doi.org/10.1111/j.1365-2559.2011.04039.x

22. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet. 2018 May;50(5):645–51. doi:10.1038/s41588-018-0078-z PubMed PMID: 29610475; PubMed Central PMCID: PMC6107367.

23. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature. 2017 Jan 19;541(7637):359–64. doi:10.1038/nature20788 PubMed PMID: 28068672.

24. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015 Nov 5;163(4):1011–25. doi:10.1016/j.cell.2015.10.025 PubMed PMID: 26544944; PubMed Central PMCID: PMC4695400.

25. Hamid AA, Gray KP, Shaw G, MacConaill LE, Evan C, Bernard B, et al. Compound Genomic Alterations of TP53, PTEN, and RB1 Tumor Suppressors in Localized and Metastatic Prostate Cancer. European Urology. 2019 Jul;76(1):89–97. doi:10.1016/j.eururo.2018.11.045

26. Maekawa S, Takata R, Obara W. Molecular Mechanisms of Prostate Cancer Development in the Precision Medicine Era: A Comprehensive Review. Cancers. 2024 Jan 25;16(3):523. doi:10.3390/cancers16030523

27. Zhang Y, Shi M, Zhang Y, Wang J, Zhang H, Ren G. RB1 and p53 are diagnostic markers for treatment-related neuroendocrine prostate cancer: a clinical and pathological analysis of 23 cases. Am J Clin Exp Urol. 2025;13(2):118–31. doi:10.62347/GRQJ8158 PubMed PMID: 40401001; PubMed Central PMCID: PMC12089225.

28. Ofner H, Kramer G, Shariat SF, Hassler MR. TP53 Deficiency in the Natural History of Prostate Cancer. Cancers. 2025 Feb 14;17(4):645. doi:10.3390/cancers17040645

29. Antar A, Xia Y, Al-Mistarehi AH, Papali P, Alfonzo Horowitz M, Sriram S, et al. PTEN Mutations Associated with Increased Recurrence and Decreased Survival in Patients with Prostate Cancer Spinal

Metastasis. Curr Oncol. 2025 Jun 4;32(6):331. doi:10.3390/curroncol32060331 PubMed PMID: 40558274; PubMed Central PMCID: PMC12192451.

30. Thompson LH, Schild D. Homologous recombinational repair of DNA ensures mammalian chromosome stability. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2001;477(1):131–53. doi:https://doi.org/10.1016/S0027-5107(01)00115-4

31. Scully R, Panday A, Elango R, Willis NA. DNA double-strand break repair-pathway choice in somatic mammalian cells. Nature Reviews Molecular Cell Biology. 2019;20(11):698–714. doi:10.1038/s41580-019-0152-0

32. Huang R, Zhou PK. DNA damage repair: historical perspectives, mechanistic pathways and clinical translation for targeted cancer therapy. Signal Transduction and Targeted Therapy. 2021;6(1):254. doi:10.1038/s41392-021-00648-7

33. Helleday T. The underlying mechanism for the PARP and BRCA synthetic lethality: clearing up the misunderstandings. Mol Oncol. 2011 Aug;5(4):387–93. doi:10.1016/j.molonc.2011.07.001 PubMed PMID: 21821475; PubMed Central PMCID: PMC5528309.

34. Shi Z, Chen B, Han X, Gu W, Liang S, Wu L. Genomic and molecular landscape of homologous recombination deficiency across multiple cancer types. Scientific Reports. 2023;13(1):8899. doi:10.1038/s41598-023-35092-w

35. Onji H, Murai J. Reconsidering the mechanisms of action of PARP inhibitors based on clinical outcomes. Cancer Science. 2022 Jun 29;n/a(n/a). doi:https://doi.org/10.1111/cas.15477

36. Cong K, Peng M, Kousholt AN, Lee WTC, Lee S, Nayak S, et al. Replication gaps are a key determinant of PARP inhibitor synthetic lethality with BRCA deficiency. Molecular Cell. 2021;81(15):3128-3144.e7. doi:https://doi.org/10.1016/j.molcel.2021.06.011

37. Popova T, Manié E, Rieunier G, Caux-Moncoutier V, Tirapo C, Dubois T, et al. Ploidy and Large-Scale Genomic Instability Consistently Identify Basal-like Breast Carcinomas with BRCA1/2 Inactivation. Cancer Research. 2012 Oct;72(21):5454–62. doi:10.1158/0008-5472.CAN-12-1470

38. Birkbak NJ, Wang ZC, Kim JY, Eklund AC, Li Q, Tian R, et al. Telomeric allelic imbalance indicates defective DNA repair and sensitivity to DNA-damaging agents. Cancer discovery. 2012;2(4):366–75.

39. 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(10):1776–82.

40. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature. 2020;578(7793):94–101. doi:10.1038/s41586-020-1943-3

41. Goodall J, Mateo J, Yuan W, Mossop H, Porta N, Miranda S, et al. Circulating Cell-Free DNA to Guide Prostate Cancer Treatment with PARP Inhibition. Cancer Discovery. 2017 Aug 31;7(9):1006–17. doi:10.1158/2159-8290.CD-17-0261

42. Shore N, Oliver L, Shui I, Gayle A, Wong OY, Kim J, et al. Systematic Literature Review of the Epidemiology of Advanced Prostate Cancer and Associated Homologous Recombination Repair Gene Alterations. J Urol. 2021 Apr;205(4):977–86. doi:10.1097/JU.0000000000001570 PubMed PMID: 33332152.

43. Shui IM, Burcu M, Shao C, Chen C, Liao CY, Jiang S, et al. Real-world prevalence of homologous recombination repair mutations in advanced prostate cancer: an analysis of two clinico-genomic databases. Prostate Cancer Prostatic Dis. 2024 Dec;27(4):728–35. doi:10.1038/s41391-023-00764-1 PubMed PMID: 38057611; PubMed Central PMCID: PMC11543596.

44. van der Wiel AMA, Schuitmaker L, Cong Y, Theys J, Van Hoeck A, Vens C, et al. Homologous Recombination Deficiency Scar: Mutations and Beyond-Implications for Precision Oncology. Cancers (Basel). 2022 Aug 27;14(17):4157. doi:10.3390/cancers14174157 PubMed PMID: 36077694; PubMed Central PMCID: PMC9454578.

45. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, et al. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. Clin Cancer Res. 2016 Aug 1;22(15):3764–73. doi:10.1158/1078-0432.CCR-15-2477 PubMed PMID: 26957554; PubMed Central PMCID: PMC6773427.

46. Armstrong AJ, Taylor A, Haffner MC, Abida W, Bryce AH, Karsh LI, et al. Germline and somatic testing for homologous repair deficiency in patients with prostate cancer (part 1 of 2). Prostate Cancer Prostatic Dis. 2025 Sep;28(3):652–61. doi:10.1038/s41391-024-00901-4 PubMed PMID: 39354185; PubMed Central PMCID: PMC12399426.

47. Serritella AV, Taylor A, Haffner MC, Abida W, Bryce A, Karsh LI, et al. Therapeutic implications of homologous repair deficiency testing in patients with prostate cancer (Part 2 of 2). Prostate Cancer Prostatic Dis. 2025 Sep;28(3):601–9. doi:10.1038/s41391-024-00887-z PubMed PMID: 39333696; PubMed Central PMCID: PMC12399429.

48. Hussain M, Mateo J, Fizazi K, Saad F, Shore N, Sandhu S, et al. Survival with Olaparib in Metastatic Castration-Resistant Prostate Cancer. N Engl J Med. 2020 Dec 10;383(24):2345–57. doi:10.1056/NEJMoa2022485 PubMed PMID: 32955174.

49. Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. Nature Genetics. 2001 Mar 1;27(3):247–54. doi:10.1038/85798

50. Piazza A, Heyer WD. Homologous Recombination and the Formation of Complex Genomic Rearrangements. Trends in Cell Biology. 2019 Feb 1;29(2):135–49. doi:10.1016/j.tcb.2018.10.006

51. Ratnaparkhe M, Wong JKL, Wei PC, Hlevnjak M, Kolb T, Simovic M, et al. Defective DNA damage repair leads to frequent catastrophic genomic events in murine and human tumors. Nature Communications. 2018 Nov 12;9(1):4760. doi:10.1038/s41467-018-06925-4

52. Grenon M, Gilbert C, Lowndes NF. Checkpoint activation in response to double-strand breaks requires the Mre11/Rad50/Xrs2 complex. Nature Cell Biology. 2001 Sep 1;3(9):844–7. doi:10.1038/ncb0901-844

53. Scott SP, Pandita TK. The cellular control of DNA double-strand breaks. Journal of cellular biochemistry. 2006;99(6):1463–75.

54. Chapman JR, Taylor MRG, Boulton SJ. Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. Molecular Cell. 2012 Aug 24;47(4):497–510. doi:10.1016/j.molcel.2012.07.029

55. Biehs R, Steinlage M, Barton O, Juhász S, Künzel J, Spies J, et al. DNA Double-Strand Break Resection Occurs during Non-homologous End Joining in G1 but Is Distinct from Resection during Homologous Recombination. Molecular Cell. 2017 Feb 16;65(4):671-684.e5. doi:10.1016/j.molcel.2016.12.016

56. Symington Lorraine S. Role of RAD52 Epistasis Group Genes in Homologous Recombination and Double-Strand Break Repair. Microbiology and Molecular Biology Reviews. 2002 Dec 1;66(4):630–70. doi:10.1128/mmbr.66.4.630-670.2002

57. Prakash R, Zhang Y, Feng W, Jasin M. Homologous recombination and human health: the roles of BRCA1, BRCA2, and associated proteins. Cold Spring Harbor perspectives in biology. 2015;7(4):a016600.

58. Tarsounas M, Sung P. The antitumorigenic roles of BRCA1–BARD1 in DNA repair and replication. Nature Reviews Molecular Cell Biology. 2020 May 1;21(5):284–99. doi:10.1038/s41580-020-0218-z

59. Liu J, Doty T, Gibson B, Heyer WD. Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. Nature Structural & Molecular Biology. 2010 Oct 1;17(10):1260–2. doi:10.1038/nsmb.1904

60. Thorslund T, McIlwraith MJ, Compton SA, Lekomtsev S, Petronczki M, Griffith JD, et al. The breast cancer tumor suppressor BRCA2 promotes the specific targeting of RAD51 to single-stranded DNA. Nature Structural & Molecular Biology. 2010 Oct 1;17(10):1263–5. doi:10.1038/nsmb.1905

61. Ranjha L, Howard SM, Cejka P. Main steps in DNA double-strand break repair: an introduction to homologous recombination and related processes. Chromosoma. 2018 Jun;127(2):187–214. doi:10.1007/s00412-017-0658-1 PubMed PMID: 29327130.

62. San Filippo J, Sung P, Klein H. Mechanism of Eukaryotic Homologous Recombination. Annual Review of Biochemistry. Annual Reviews; 2008. p. 229–57. doi:https://doi.org/10.1146/annurev.biochem.77.061306.125255

63. Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, et al. Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21. Science. 1990 Dec 21;250(4988):1684–9. doi:10.1126/science.2270482

64. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, et al. A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1. Science. 1994 Oct 7;266(5182):66–71. doi:10.1126/science.7545954

65. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, et al. Identification of the breast cancer susceptibility gene BRCA2. Nature. 1995 Dec 1;378(6559):789–92. doi:10.1038/378789a0

66. Roy R, Chun J, Powell SN. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nature Reviews Cancer. 2012 Jan 1;12(1):68–78. doi:10.1038/nrc3181

67. Lord CJ, Ashworth A. The DNA damage response and cancer therapy. Nature. 2012 Jan 1;481(7381):287–94. doi:10.1038/nature10760

68. Kuchenbaecker KB, Hopper JL, Barnes DR, Phillips KA, Mooij TM, Roos-Blom MJ, et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA. 2017 Jun 20;317(23):2402–16. doi:10.1001/jama.2017.7112

69. Gruber JJ, Afghahi A, Timms K, DeWees A, Gross W, Aushev VN, et al. A phase II study of talazoparib monotherapy in patients with wild-type BRCA1 and BRCA2 with a mutation in other homologous recombination genes. Nature Cancer. 2022 Oct 1;3(10):1181–91. doi:10.1038/s43018-022-00439-1

70. de Bono Johann, Mateo Joaquin, Fizazi Karim, Saad Fred, Shore Neal, Sandhu Shahneen, et al. Olaparib for Metastatic Castration-Resistant Prostate Cancer. New England Journal of Medicine. 2020 May 28;382(22):2091–102. doi:10.1056/NEJMoa1911440

71. Marshall CH, Sokolova AO, McNatty AL, Cheng HH, Eisenberger MA, Bryce AH, et al. Differential Response to Olaparib Treatment Among Men with Metastatic Castration-resistant Prostate Cancer Harboring BRCA1 or BRCA2 Versus ATM Mutations. European Urology. 2019 Oct 1;76(4):452–8. doi:10.1016/j.eururo.2019.02.002

72. Wu YM, Cieślik M, Lonigro RJ, Vats P, Reimers MA, Cao X, et al. Inactivation of CDK12 delineates a distinct immunogenic class of advanced prostate cancer. Cell. 2018;173(7):1770–82.

73. Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, et al. Promoter Hypermethylation and BRCA1 Inactivation in Sporadic Breast and Ovarian Tumors. JNCI: Journal of the National Cancer Institute. 2000 Apr 5;92(7):564–9. doi:10.1093/jnci/92.7.564

74. Jacot W, Lopez-Crapez E, Mollevi C, Boissière-Michot F, Simony-Lafontaine J, Ho-Pun-Cheung A, et al. BRCA1 Promoter Hypermethylation is Associated with Good Prognosis and Chemosensitivity in Triple-Negative Breast Cancer. Cancers. 2020;12(4). doi:10.3390/cancers12040828

75. Turner N, Tutt A, Ashworth A. Hallmarks of "BRCAness" in sporadic cancers. Nat Rev Cancer. 2004 Oct;4(10):814–9. doi:10.1038/nrc1457 PubMed PMID: 15510162.

76. Telli ML, Timms KM, Reid J, Hennessy B, Mills GB, Jensen KC, et al. Homologous Recombination Deficiency (HRD) Score Predicts Response to Platinum-Containing Neoadjuvant Chemotherapy in Patients with Triple-Negative Breast Cancer. Clinical Cancer Research. 2016 Jul 31;22(15):3764–73. doi:10.1158/1078-0432.CCR-15-2477

77. de Caceres II, Battagli C, Esteller M, Herman JG, Dulaimi E, Edelson MI, et al. Tumor Cell-Specific BRCA1 and RASSF1A Hypermethylation in Serum, Plasma, and Peritoneal Fluid from Ovarian Cancer Patients. Cancer Research. 2004 Sep 16;64(18):6476–81. doi:10.1158/0008-5472.CAN-04-1529

78. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013 Aug 1;500(7463):415–21. doi:10.1038/nature12477

79. Davies H, Glodzik D, Morganella S, Yates LR, Staaf J, Zou X, et al. HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. Nature Medicine. 2017 Apr 1;23(4):517–25. doi:10.1038/nm.4292

80. Lee CY, Cheng WF, Lin PH, Chen YL, Huang SH, Lei KH, et al. An activity-based functional test for identifying homologous recombination deficiencies across cancer types in real time. Cell Reports Medicine. 2023 Nov 21;4(11):101247. doi:10.1016/j.xcrm.2023.101247

81. Compadre AJ, van Biljon LN, Valentine MC, Llop-Guevara A, Graham E, Fashemi B, et al. RAD51 Foci as a Biomarker Predictive of Platinum Chemotherapy Response in Ovarian Cancer. Clinical Cancer Research. 2023 Jul 5;29(13):2466–79. doi:10.1158/1078-0432.CCR-22-3335

82. Meijer TG, Nguyen L, Van Hoeck A, Sieuwerts AM, Verkaik NS, Ladan MM, et al. Functional RECAP (REpair CAPacity) assay identifies homologous recombination deficiency undetected by DNA-based BRCAness tests. Oncogene. 2022 Jun 1;41(26):3498–506. doi:10.1038/s41388-022-02363-1

83. MyChoice® CDx Technical Information [Internet]. [cited 2025 Mar 22]. Available from: https://s3.amazonaws.com/myriad-web/myChoiceCDx/downloads/myChoiceCDxTech.pdf

84. FoundationOne®CDx Technical Information [Internet]. [cited 2025 Oct 22]. Available from: https://info.foundationmedicine.com/hubfs/FMI%20Labels/FoundationOne_CDx_Label_Technical_Info.pdf

85. FoundationOne® Liquid CDx Technical Information [Internet]. [cited 2025 Mar 22]. Available from: https://info.foundationmedicine.com/hubfs/FMI%20Labels/FoundationOne_Liquid_CDx_Label_Technical_Info.pdf

86. BRACAnalysis CDx® Technical Information [Internet]. [cited 2025 Mar 22]. Available from: https://s3.amazonaws.com/myriad-web/BRACAnalysisCDxTS.pdf

87. The Tempus HRD Test [Internet]. [cited 2025 Mar 22]. Available from: https://www.tempus.com/wp-content/uploads/2020/05/Tempus-Tech-Spotlight-The-Tempus-HRD-Test.pdf

88. Homologous Recombination Deficiency. Caris Life Sciences. [Internet]. [cited 2025 Mar 22]. Available from: https://www.carislifesciences.com/physicians/physician-tests/whole-exome-sequencing/homologous-recombination-deficiency-hrd/

89. Zheng G, Lin MT, Lokhandwala PM, Beierl K, Netto GJ, Gocke CD, et al. Clinical mutational profiling of bone metastases of lung and colon carcinoma and malignant melanoma using next-generation sequencing. Cancer Cytopathology. 2016 Oct 1;124(10):744–53. doi:10.1002/cncy.21743

90. Gulhan DC, Lee JJK, Melloni GEM, Cortés-Ciriano I, Park PJ. Detecting the mutational signature of homologous recombination deficiency in clinical samples. Nat Genet. 2019 May;51(5):912–9. doi:10.1038/s41588-019-0390-2 PubMed PMID: 30988514.

91. Eeckhoutte A, Houy A, Manié E, Reverdy M, Bieche I, Marangoni E, et al. ShallowHRD: detection of homologous recombination deficiency from shallow whole genome sequencing. Bioinformatics. 2020;36(12):3888–9.

92. Callens C, Rodrigues M, Briaux A, Frouin E, Eeckhoutte A, Pujade-Lauraine E, et al. Shallow whole genome sequencing approach to detect Homologous Recombination Deficiency in the PAOLA-1/ENGOT-OV25 phase-III trial. Oncogene. 2023 Nov 1;42(48):3556–63. doi:10.1038/s41388-023-02839-8

93. 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 Jul 2;4(1):16. doi:10.1038/s41523-018-0066-6

94. Pozzorini C, Andre G, Coletta T, Buisson A, Bieler J, Ferrer L, et al. GIInger predicts homologous recombination deficiency and patient response to PARPi treatment from shallow genomic profiles. Cell Rep Med. 2023 Dec 19;4(12):101344. doi:10.1016/j.xcrm.2023.101344 PubMed PMID: 38118421; PubMed Central PMCID: PMC10772634.

95. Nguyen L, W. M. Martens J, Van Hoeck A, Cuppen E. Pan-cancer landscape of homologous recombination deficiency. Nat Commun. 2020 Nov 4;11(1):5584. doi:10.1038/s41467-020-19406-4

96. Wang X, Xu Y, Zhang Y, Wang S, Zhang X, Yi X, et al. HRD-MILN: Accurately estimate tumor homologous recombination deficiency status from targeted panel sequencing data. Front Genet. 2022 Sep 28;13:990244. doi:10.3389/fgene.2022.990244

97. Liu R, Roberts E, Parsons HA, Stover EH, Choudhury AD, Rhoades J, et al. DirectHRD enables sensitive scar-based classification of homologous recombination deficiency. Nucleic Acids Research. 2025 Apr 22;53(8):gkaf313. doi:10.1093/nar/gkaf313

98. Ritch EJ, Herberts C, Warner EW, Ng SWS, Kwan EM, Bacon JVW, et al. A generalizable machine learning framework for classifying DNA repair defects using ctDNA exomes. NPJ Precis Oncol. 2023 Mar 13;7(1):27. doi:10.1038/s41698-023-00366-z PubMed PMID: 36914848; PubMed Central PMCID: PMC10011564.

99. Bergstrom EN, Abbasi A, Díaz-Gay M, Galland L, Ladoire S, Lippman SM, et al. Deep Learning Artificial Intelligence Predicts Homologous Recombination Deficiency and Platinum Response From Histologic Slides. JCO. 2024 Oct 20;42(30):3550–60. doi:10.1200/JCO.23.02641

100. Schirris Y, Gavves E, Nederlof I, Horlings HM, Teuwen J. DeepSMILE: Contrastive self-supervised pre-training benefits MSI and HRD classification directly from H&E whole-slide images in colorectal and breast cancer. Medical Image Analysis. 2022 Jul;79:102464. doi:10.1016/j.media.2022.102464

101. Loeffler CML, El Nahhas OSM, Muti HS, Seibel T, Cifci D, van Treeck M, et al. Direct prediction of Homologous Recombination Deficiency from routine histology in ten different tumor types with attention-based Multiple Instance Learning: a development and validation study. medRxiv. 2023 Mar 10;2023.03.08.23286975. doi:10.1101/2023.03.08.23286975 PubMed PMID: 36945540; PubMed Central PMCID: PMC10029072.

102. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language Models are Few-Shot Learners [Internet]. arXiv; 2020 [cited 2025 Jun 24]. Available from: https://arxiv.org/abs/2005.14165 doi:10.48550/ARXIV.2005.14165

103. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Internet]. arXiv; 2018 [cited 2025 Jun 24]. Available from: https://arxiv.org/abs/1810.04805 doi:10.48550/ARXIV.1810.04805

104. Liu X, Zhang F, Hou Z, Mian L, Wang Z, Zhang J, et al. Self-supervised Learning: Generative or Contrastive. IEEE Trans Knowl Data Eng. 2021;1–1. doi:10.1109/TKDE.2021.3090866

105. Sanabria M, Hirsch J, Joubert PM, Poetsch AR. DNA language model GROVER learns sequence context in the human genome. Nat Mach Intell. 2024 Jul 23;6(8):911–23. doi:10.1038/s42256-024-00872-0

106. Dalla-Torre H, Gonzalez L, Mendoza-Revilla J, Lopez Carranza N, Grzywaczewski AH, Oteri F, et al. Nucleotide Transformer: building and evaluating robust foundation models for human genomics. Nat Methods. 2025 Feb;22(2):287–97. doi:10.1038/s41592-024-02523-z

107. Benegas G, Ye C, Albors C, Li JC, Song YS. Genomic language models: opportunities and challenges. Trends in Genetics. 2025 Apr;41(4):286–302. doi:10.1016/j.tig.2024.11.013

108. Liu J, Yang M, Yu Y, Xu H, Wang T, Li K, et al. Advancing bioinformatics with large language models: components, applications and perspectives. United States; 2025.

109. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention Is All You Need [Internet]. arXiv; 2017 [cited 2025 Jun 24]. Available from: https://arxiv.org/abs/1706.03762 doi:10.48550/ARXIV.1706.03762

110. Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, et al. Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods. 2021 Oct;18(10):1196–203. doi:10.1038/s41592-021-01252-x

111. Ji Y, Zhou Z, Liu H, Davuluri RV. DNABERT: pre-trained Bidirectional Encoder Representations from Transformers model for DNA-language in genome. Kelso J, editor. Bioinformatics. 2021 Aug 9;37(15):2112–20. doi:10.1093/bioinformatics/btab083

112. Su J, Lu Y, Pan S, Murtadha A, Wen B, Liu Y. RoFormer: Enhanced Transformer with Rotary Position Embedding [Internet]. arXiv; 2021 [cited 2025 Jun 24]. Available from: https://arxiv.org/abs/2104.09864 doi:10.48550/ARXIV.2104.09864

113. Lord CJ, Ashworth A. BRCAness revisited. Nat Rev Cancer. 2016 Feb;16(2):110–20. doi:10.1038/nrc.2015.21

114. Avsec Ž, Agarwal V, Visentin D, Ledsam JR, Grabska-Barwinska A, Taylor KR, et al. Effective gene expression prediction from sequence by integrating long-range interactions. Nat Methods. 2021 Oct;18(10):1196–203. doi:10.1038/s41592-021-01252-x

115. Gao H, Hamp T, Ede J, Schraiber JG, McRae J, Singer-Berk M, et al. The landscape of tolerated genetic variation in humans and primates. Science. 2023 Jun 2;380(6648):eabn8153. doi:10.1126/science.abn8197

116. Cheng J, Novati G, Pan J, Bycroft C, Žemgulytė A, Applebaum T, et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science. 2023 Sep 22;381(6664):eadg7492. doi:10.1126/science.adg7492

117. Zhu G, Guo YA, Ho D, Poon P, Poh ZW, Wong PM, et al. Tissue-specific cell-free DNA degradation quantifies circulating tumor DNA burden. Nature Communications. 2021 Apr 13;12(1):2229. doi:10.1038/s41467-021-22463-y

118. Tsui WHA, Jiang P, Lo YMD. Cell-free DNA fragmentomics in cancer. Cancer Cell. 2025 Oct;43(10):1792–814. doi:10.1016/j.ccell.2025.09.006

119. Liu D, Yehia L, Dhawan A, Ni Y, Eng C. Cell-free DNA fragmentomics and second malignant neoplasm risk in patients with PTEN hamartoma tumor syndrome. Cell Rep Med. 2024 Feb 20;5(2):101384. doi:10.1016/j.xcrm.2023.101384 PubMed PMID: 38242121; PubMed Central PMCID: PMC10897513.

120. Bettegowda C, Sausen M, Leary RJ, Kinde I, Wang Y, Agrawal N, et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. Sci Transl Med. 2014 Feb 19;6(224). doi:10.1126/scitranslmed.3007094

121. Cristiano S, Leal A, Phallen J, Fiksel J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. Nature. 2019 Jun 1;570(7761):385–9. doi:10.1038/s41586-019-1272-6

122. Chabon JJ, Hamilton EG, Kurtz DM, Esfahani MS, Moding EJ, Stehr H, et al. Integrating genomic features for non-invasive early lung cancer detection. Nature. 2020;580(7802):245–51. doi:10.1038/s41586-020-2140-0

123. Moser T, Kühberger S, Lazzeri I, Vlachos G, Heitzer E. Bridging biological cfDNA features and machine learning approaches. Trends in Genetics. 2023 Apr 1;39(4):285–307. doi:10.1016/j.tig.2023.01.004

124. Wan JCM, Sasieni P, Rosenfeld N. Promises and pitfalls of multi-cancer early detection using liquid biopsy tests. Nat Rev Clin Oncol. 2025 Aug;22(8):566–80. doi:10.1038/s41571-025-01033-x

125. Bando H, Kagawa Y, Kato T, Akagi K, Denda T, Nishina T, et al. A multicentre, prospective study of plasma circulating tumour DNA test for detecting RAS mutation in patients with metastatic colorectal cancer. British Journal of Cancer. 2019 May 1;120(10):982–6. doi:10.1038/s41416-019-0457-y

126. Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, et al. Direct detection of early-stage cancers using circulating tumor DNA. Sci Transl Med. 2017 Aug 16;9(403). doi:10.1126/scitranslmed.aan2415 PubMed PMID: 28814544; PubMed Central PMCID: PMC6714979.

127. Bae JH, Liu R, Roberts E, Nguyen E, Tabrizi S, Rhoades J, et al. Single duplex DNA sequencing with CODEC detects mutations with high sensitivity. Nature Genetics. 2023 May 1;55(5):871–9. doi:10.1038/s41588-023-01376-0

128. Zarrei M, MacDonald JR, Merico D, Scherer SW. A copy number variation map of the human genome. Nature Reviews Genetics. 2015 Mar 1;16(3):172–83. doi:10.1038/nrg3871

129. Klega K, Imamovic-Tuco A, Ha G, Clapp AN, Meyer S, Ward A, et al. Detection of Somatic Structural Variants Enables Quantification and Characterization of Circulating Tumor DNA in Children With Solid Tumors. JCO Precis Oncol. 2018;2018. doi:10.1200/PO.17.00285 PubMed PMID: 30027144; PubMed Central PMCID: PMC6049092.

130. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. Nature Communications. 2017;8(1):1324. doi:10.1038/s41467-017-00965-y

131. Heitzer E, Ulz P, Belic J, Gutschi S, Quehenberger F, Fischereder K, et al. Tumor-associated copy number changes in the circulation of patients with prostate cancer identified through whole-genome sequencing. Genome Medicine. 2013;5(4):30. doi:10.1186/gm434

132. Leary RJ, Kinde I, Diehl F, Schmidt K, Clouser C, Duncan C, et al. Development of Personalized Tumor Biomarkers Using Massively Parallel Sequencing. Science Translational Medicine. 2010 Feb 24;2(20):20ra14-20ra14. doi:10.1126/scitranslmed.3000702

133. Logsdon GA, Vollger MR, Eichler EE. Long-read human genome sequencing and its applications. Nature Reviews Genetics. 2020 Oct 1;21(10):597–614. doi:10.1038/s41576-020-0236-x

134. Jain M, Koren S, Miga KH, Quick J, Rand AC, Sasani TA, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. Nature Biotechnology. 2018 Apr 1;36(4):338–45. doi:10.1038/nbt.4060

135. Chen Y, Wang AY, Barkley CA, Zhang Y, Zhao X, Gao M, et al. Deciphering the exact breakpoints of structural variations using long sequencing reads with DeBreak. Nat Commun. 2023 Jan 17;14(1):283. doi:10.1038/s41467-023-35996-1

136. Valle-Inclan JE, Stangl C, De Jong AC, Van Dessel LF, Van Roosmalen MJ, Helmijr JCA, et al. Optimizing Nanopore sequencing-based detection of structural variants enables individualized circulating tumor DNA-based disease monitoring in cancer patients. Genome Med. 2021 Dec;13(1):86. doi:10.1186/s13073-021-00899-7

137. Sakamoto Y, Zaha S, Suzuki Y, Seki M, Suzuki A. Application of long-read sequencing to the detection of structural variants in human cancer genomes. Computational and Structural Biotechnology Journal. 2021;19:4207–16. doi:10.1016/j.csbj.2021.07.030

138. Degasperi A, Amarante TD, Czarnecki J, Shooter S, Zou X, Glodzik D, et al. A practical framework and online tool for mutational signature analyses show intertissue variation and driver dependencies. Nature Cancer. 2020 Feb 1;1(2):249–63. doi:10.1038/s43018-020-0027-5

139. Jin H, Gulhan DC, Geiger B, Ben-Isvy D, Geng D, Ljungström V, et al. Accurate and sensitive mutational signature analysis with MuSiCal. Nature Genetics. 2024 Mar 1;56(3):541–52. doi:10.1038/s41588-024-01659-0

140. Jiang N, Wu Y, Rozen SG. Benchmarking 13 tools for mutational signature attribution, including a new and improved algorithm. Briefings in Bioinformatics. 2025 Jan 1;26(1):bbaf042. doi:10.1093/bib/bbaf042

141. Hollizeck S, Wang N, Wong SQ, Litchfield C, Guinto J, Ftouni S, et al. Unravelling mutational signatures with plasma circulating tumour DNA. Nature Communications. 2024 Nov 14;15(1):9876. doi:10.1038/s41467-024-54193-2

142. Wan JCM, Stephens D, Luo L, White JR, Stewart CM, Rousseau B, et al. Genome-wide mutational signatures in low-coverage whole genome sequencing of cell-free DNA. Nature Communications. 2022 Aug 23;13(1):4953. doi:10.1038/s41467-022-32598-1

143. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA Comprises an In Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. Cell. 2016 Jan 14;164(1):57–68. doi:10.1016/j.cell.2015.11.050

144. Doebley AL, Ko M, Liao H, Cruikshank AE, Santos K, Kikawa C, et al. A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. Nature Communications. 2022 Dec 3;13(1):7475. doi:10.1038/s41467-022-35076-w

145. Esfahani MS, Hamilton EG, Mehrmohamadi M, Nabet BY, Alig SK, King DA, et al. Inferring gene expression from cell-free DNA fragmentation profiles. Nature Biotechnology. 2022 Apr 1;40(4):585–97. doi:10.1038/s41587-022-01222-4

146. Sun K, Jiang P, Cheng SH, Cheng THT, Wong J, Wong VWS, et al. Orientation-aware plasma cell-free DNA fragmentation analysis in open chromatin regions informs tissue of origin. Genome Res. 2019 Mar;29(3):418–27. doi:10.1101/gr.242719.118 PubMed PMID: 30808726; PubMed Central PMCID: PMC6396422.

147. Jiang P, Xie T, Ding SC, Zhou Z, Cheng SH, Chan RWY, et al. Detection and characterization of jagged ends of double-stranded DNA in plasma. Genome Res. 2020 Aug;30(8):1144–53. doi:10.1101/gr.261396.120 PubMed PMID: 32801148; PubMed Central PMCID: PMC7462074.

148. Zhu G, Jiang P, Li X, Peng W, Choy LYL, Yu SCY, et al. Methylation-Associated Nucleosomal Patterns of Cell-Free DNA in Cancer Patients and Pregnant Women. Clinical Chemistry. 2024 Nov 1;70(11):1355–65. doi:10.1093/clinchem/hvae118

149. Zhu G, Rahman CR, Getty V, Odinokov D, Baruah P, Carrié H, et al. Fragle: Universal ctDNA quantification using deep learning of fragmentomic profiles. bioRxiv. 2024 Jan 1;2023.07.28.550922. doi:10.1101/2023.07.28.550922

150. Bai J, Jiang P, Ji L, Lam WKJ, Zhou Q, Ma MJL, et al. Histone modifications of circulating nucleosomes are associated with changes in cell-free DNA fragmentation patterns. Proc Natl Acad Sci USA. 2024 Oct 15;121(42):e2404058121. doi:10.1073/pnas.2404058121

151. Budhraja KK, McDonald BR, Stephens MD, Contente-Cuomo T, Markus H, Farooq M, et al. Genome-wide analysis of aberrant position and sequence of plasma DNA fragment ends in patients with cancer. Sci Transl Med. 2023 Jan 11;15(678):eabm6863. doi:10.1126/scitranslmed.abm6863 PubMed PMID: 36630480; PubMed Central PMCID: PMC10080578.

152. Liu J, Shen H, Chen K, Li X. Large language model produces high accurate diagnosis of cancer from end-motif profiles of cell-free DNA. Briefings in Bioinformatics. 2024 Jul 25;25(5):bbae430. doi:10.1093/bib/bbae430

153. Shen H, Liu J, Chen K, Li X. Language model enables end-to-end accurate detection of cancer from cell-free DNA. Brief Bioinform. 2024 Jan 22;25(2):bbae053. doi:10.1093/bib/bbae053 PubMed PMID: 38385880; PubMed Central PMCID: PMC10883418.

154. Ju J, Zhao X, An Y, Yang M, Zhang Z, Liu X, et al. Cell-free DNA end characteristics enable accurate and sensitive cancer diagnosis. Cell Reports Methods. 2024 Oct;4(10):100877. doi:10.1016/j.crmeth.2024.100877

155. Helzer KT, Sharifi MN, Sperger JM, Shi Y, Annala M, Bootsma ML, et al. Fragmentomic analysis of circulating tumor DNA-targeted cancer panels. Annals of Oncology. 2023 Sep;34(9):813–25. doi:10.1016/j.annonc.2023.06.001

156. Hou Y, Meng X, Zhou X. Systematically Evaluating Cell-Free DNA Fragmentation Patterns for Cancer Diagnosis and Enhanced Cancer Detection via Integrating Multiple Fragmentation Patterns. Advanced Science. 2024 Aug;11(30):2308243. doi:10.1002/advs.202308243

157. Liu J, Dai L, Wang Q, Li C, Liu Z, Gong T, et al. Multimodal analysis of cfDNA methylomes for early detecting esophageal squamous cell carcinoma and precancerous lesions. Nature Communications. 2024 May 2;15(1):3700. doi:10.1038/s41467-024-47886-1

158. Mathios D, Niknafs N, Annapragada AV, Bobeff EJ, Chiao EJ, Boyapati K, et al. Detection of brain cancer using genome-wide cell-free DNA fragmentomes. Cancer Discovery. 2025 Apr 29. doi:10.1158/2159-8290.CD-25-0074

159. Nguyen VTC, Nguyen TH, Doan NNT, Pham TMQ, Nguyen GTH, Nguyen TD, et al. Multimodal analysis of methylomics and fragmentomics in plasma cell-free DNA for multi-cancer early detection and localization. Choi M, editor. eLife. 2023 Oct 11;12:RP89083. doi:10.7554/eLife.89083

160. Kim SY, Jeong S, Lee W, Jeon Y, Kim YJ, Park S, et al. Cancer signature ensemble integrating cfDNA methylation, copy number, and fragmentation facilitates multi-cancer early detection. Experimental & Molecular Medicine. 2023 Nov 1;55(11):2445–60. doi:10.1038/s12276-023-01119-5

161. Carter HB, Albertsen PC, Barry MJ, Etzioni R, Freedland SJ, Greene KL, et al. Early detection of prostate cancer: AUA Guideline. The Journal of urology. 2013;190(2):419–26.

162. Ahdoot M, Wilbur AR, Reese SE, Lebastchi AH, Mehralivand S, Gomella PT, et al. MRI-targeted, systematic, and combined biopsy for prostate cancer diagnosis. New England Journal of Medicine. 2020;382(10):917–28.

163. Litchfield K, Stanislaw S, Spain L, Gallegos LL, Rowan A, Schnidrig D, et al. Representative Sequencing: Unbiased Sampling of Solid Tumor Tissue. Cell Reports. 2020 May 5;31(5):107550. doi:10.1016/j.celrep.2020.107550

164. Bass EJ, Pantovic A, Connor MJ, Loeb S, Rastinehad AR, Winkler M, et al. Diagnostic accuracy of magnetic resonance imaging targeted biopsy techniques compared to transrectal ultrasound guided biopsy

of the prostate: a systematic review and meta-analysis. Prostate Cancer and Prostatic Diseases. 2022;25(2):174–9. doi:10.1038/s41391-021-00449-7

165. Ahdoot M, Wilbur AR, Reese SE, Lebastchi AH, Mehralivand S, Gomella PT, et al. MRI-Targeted, Systematic, and Combined Biopsy for Prostate Cancer Diagnosis. The New England journal of medicine. 2020 Mar 5;382(10):917–28. doi:10.1056/NEJMoa1910038

166. Halabi S, Kelly WK, Ma H, Zhou H, Solomon NC, Fizazi K, et al. Meta-Analysis Evaluating the Impact of Site of Metastasis on Overall Survival in Men With Castration-Resistant Prostate Cancer. Journal of Clinical Oncology. 2016 Mar 7;34(14):1652–9. doi:10.1200/JCO.2015.65.7270

167. Zheng G, Lin MT, Lokhandwala PM, Beierl K, Netto GJ, Gocke CD, et al. Clinical mutational profiling of bone metastases of lung and colon carcinoma and malignant melanoma using next-generation sequencing. Cancer Cytopathology. 2016 Oct 1;124(10):744–53. doi:https://doi.org/10.1002/cncy.21743

168. Herberts C, Annala M, Sipola J, Ng SWS, Chen XE, Nurminen A, et al. Deep whole-genome ctDNA chronology of treatment-resistant prostate cancer. Nature. 2022;608(7921):199–208. doi:10.1038/s41586-022-04975-9

169. De Mattos-Arruda L, Weigelt B, Cortes J, Won HH, Ng CKY, Nuciforo P, et al. Capturing intra-tumor genetic heterogeneity by de novo mutation profiling of circulating cell-free tumor DNA: a proof-of-principle. Annals of Oncology. 2014 Sep 1;25(9):1729–35. doi:10.1093/annonc/mdu239

170. Warner EW, Van der Eecken K, Murtha AJ, Kwan EM, Herberts C, Sipola J, et al. Multiregion sampling of de novo metastatic prostate cancer reveals complex polyclonality and augments clinical genotyping. Nature Cancer. 2024. doi:10.1038/s43018-023-00692-y

171. Quigley D, Alumkal JJ, Wyatt AW, Kothari V, Foye A, Lloyd P, et al. Analysis of circulating cell-free DNA identifies multiclonal heterogeneity of BRCA2 reversion mutations associated with resistance to PARP inhibitors. Cancer discovery. 2017;7(9):999–1005.

172. Arends CM, Galan-Sousa J, Hoyer K, Chan W, Jäger M, Yoshida K, et al. Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. Leukemia. 2018;32(9):1908–19. doi:10.1038/s41375-018-0047-7

173. Duffy MJ, Diamandis EP, Crown J. Circulating tumor DNA (ctDNA) as a pan-cancer screening test: is it finally on the horizon? Vol. 59. 2021;Clinical Chemistry and Laboratory Medicine (CCLM)59(8):1353–61. doi:doi:10.1515/cclm-2021-0171

174. Heitzer E, Haque IS, Roberts CES, Speicher MR. Current and future perspectives of liquid biopsies in genomics-driven oncology. Nature Reviews Genetics. 2019;20(2):71–88. doi:10.1038/s41576-018-0071-5

175. Heitzer E, van den Broek D, Denis MG, Hofman P, Hubank M, Mouliere F, et al. Recommendations for a practical implementation of circulating tumor DNA mutation testing in metastatic non-small-cell lung cancer. ESMO Open. 2022;7(2):100399. doi:https://doi.org/10.1016/j.esmoop.2022.100399

176. Tsui WHA, Ding SC, Jiang P, Lo YMD. Artificial intelligence and machine learning in cell-free-DNA-based diagnostics. Genome Res. 2025 Jan 22;35(1):1–19. doi:10.1101/gr.278413.123 PubMed PMID: 39843210; PubMed Central PMCID: PMC11789496.

177. Pellegrino B, Musolino A, Llop-Guevara A, Serra V, De Silva P, Hlavata Z, et al. Homologous Recombination Repair Deficiency and the Immune Response in Breast Cancer: A Literature Review. Translational Oncology. 2020 Feb 1;13(2):410–22. doi:10.1016/j.tranon.2019.10.010

178. Kwan EM, Wyatt AW, Chi KN. Towards clinical implementation of circulating tumor DNA in metastatic prostate cancer: Opportunities for integration and pitfalls to interpretation. Frontiers in Oncology [Internet]. 2022;12. Available from: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2022.1054497

179. Cimadamore A, Cheng L, Massari F, Santoni M, Pepi L, Franzese C, et al. Circulating Tumor DNA Testing for Homology Recombination Repair Genes in Prostate Cancer: From the Lab to the Clinic. Int J Mol Sci. 2021 May 24;22(11). doi:10.3390/ijms22115522 PubMed PMID: 34073818; PubMed Central PMCID: PMC8197269.

180. Vlachos G, Moser T, Lazzeri I, Moser MJ, Glawitch L, Bauernhofer ET, et al. Functional footprints of homologous recombination deficiency in prostate cancer revealed by ctDNA fragmentation and transcription factor accessibility. Br J Cancer. 2026 Jan 9. doi:10.1038/s41416-025-03301-0

181. Belic J, Koch M, Ulz P, Auer M, Gerhalter T, Mohan S, et al. Rapid Identification of Plasma DNA Samples with Increased ctDNA Levels by a Modified FAST-SeqS Approach. Clinical Chemistry. 2015 Jun 1;61(6):838–49. doi:10.1373/clinchem.2014.234286

182. Vasimuddin Md, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS) [Internet]. Rio de Janeiro, Brazil: IEEE; 2019 [cited 2025 Oct 19]. p. 314–24. Available from: https://ieeexplore.ieee.org/document/8820962/ doi:10.1109/IPDPS.2019.00041

183. Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics. 2016 Apr 15;32(8):1220–2. doi:10.1093/bioinformatics/btv710 PubMed PMID: 26647377.

184. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Dec;17(1):122. doi:10.1186/s13059-016-0974-4

185. Broad Institute. Picard Toolkit [Internet]. Broad Institute, GitHub repository; 2018. Available from: http://broadinstitute.github.io/picard/

186. Md V, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems [Internet]. arXiv; 2019 [cited 2025 Oct 24]. Available from: https://arxiv.org/abs/1907.12931 doi:10.48550/ARXIV.1907.12931

187. Lazzeri I, Spiegl BG, Hasenleithner SO, Speicher MR, Kircher M. LBFextract: Unveiling transcription factor dynamics from liquid biopsy data. Computational and Structural Biotechnology Journal. 2024 Dec 1;23:3163–74. doi:10.1016/j.csbj.2024.08.007

188. Spiegl B, Kapidzic F, Röner S, Kircher M, Speicher MR. GCparagon: evaluating and correcting GC biases in cell-free DNA at the fragment level. NAR Genomics and Bioinformatics. 2023 Oct 11;5(4):lqad102. doi:10.1093/nargab/lqad102

189. Van Dessel LF, Van Riet J, Smits M, Zhu Y, Hamberg P, Van Der Heijden MS, et al. The genomic landscape of metastatic castration-resistant prostate cancers reveals multiple distinct genotypes with potential clinical impact. Nat Commun. 2019 Nov 20;10(1):5251. doi:10.1038/s41467-019-13084-7

190. Viswanathan SR, Ha G, Hoff AM, Wala JA, Carrot-Zhang J, Whelan CW, et al. Structural Alterations Driving Castration-Resistant Prostate Cancer Revealed by Linked-Read Genome Sequencing. Cell. 2018 Jul;174(2):433-447.e19. doi:10.1016/j.cell.2018.05.036

191. Steele CD, Abbasi A, Islam SMA, Bowes AL, Khandekar A, Haase K, et al. Signatures of copy number alterations in human cancer. Nature. 2022 Jun 30;606(7916):984–91. doi:10.1038/s41586-022-04738-6

192. Nicolai S, Mahen R, Raschellà G, Marini A, Pieraccioli M, Malewicz M, et al. ZNF281 is recruited on DNA breaks to facilitate DNA repair by non-homologous end joining. Oncogene. 2020 Jan 1;39(4):754–66. doi:10.1038/s41388-019-1028-7

193. Singh JK, van Attikum H. DNA double-strand break repair: Putting zinc fingers on the sore spot. Seminars in Cell & Developmental Biology. 2021 May 1;113:65–74. doi:10.1016/j.semcdb.2020.09.003

194. Warmerdam DO, Alonso-de Vega I, Wiegant WW, van den Broek B, Rother MB, Wolthuis RM, et al. PHF6 promotes non-homologous end joining and G2 checkpoint recovery. EMBO reports. 2020 Jan 7;21(1):e48460. doi:10.15252/embr.201948460

195. Cao LL, Wei F, Du Y, Song B, Wang D, Shen C, et al. ATM-mediated KDM2A phosphorylation is required for the DNA damage repair. Oncogene. 2016 Jan 1;35(3):301–13. doi:10.1038/onc.2015.81

196. Gong F, and Miller KM. Double duty: ZMYND8 in the DNA damage response and cancer. Cell Cycle. 2018 Feb 16;17(4):414–20. doi:10.1080/15384101.2017.1376150

197. Di Tullio F, Schwarz M, Zorgati H, Mzoughi S, Guccione E. The duality of PRDM proteins: epigenetic and structural perspectives. The FEBS Journal. 2022 Mar 1;289(5):1256–75. doi:10.1111/febs.15844

198. Izhar L, Adamson B, Ciccia A, Lewis J, Pontano-Vaites L, Leng Y, et al. A Systematic Analysis of Factors Localized to Damaged Chromatin Reveals PARP-Dependent Recruitment of Transcription Factors. Cell Reports. 2015 Jun 9;11(9):1486–500. doi:10.1016/j.celrep.2015.04.053

199. Pritchard CC, Mateo J, Walsh MF, De Sarkar N, Abida W, Beltran H, et al. Inherited DNA-Repair Gene Mutations in Men with Metastatic Prostate Cancer. N Engl J Med. 2016 Aug 4;375(5):443–53. doi:10.1056/NEJMoa1603144

200. Diaz LA, Bardelli A. Liquid Biopsies: Genotyping Circulating Tumor DNA. JCO. 2014 Feb 20;32(6):579–86. doi:10.1200/JCO.2012.45.2011

201. Boerrigter E, Groen LN, Van Erp NP, Verhaegh GW, Schalken JA. Clinical utility of emerging biomarkers in prostate cancer liquid biopsies. Expert Review of Molecular Diagnostics. 2020 Feb 1;20(2):219–30. doi:10.1080/14737159.2019.1675515

202. Mateo J, Carreira S, Sandhu S, Miranda S, Mossop H, Perez-Lopez R, et al. DNA-Repair Defects and Olaparib in Metastatic Prostate Cancer. N Engl J Med. 2015 Oct 29;373(18):1697–708. doi:10.1056/NEJMoa1506859

203. Sztupinszki Z, Diossy M, Krzystanek M, Borcsok J, Pomerantz MM, Tisza V, et al. Detection of Molecular Signatures of Homologous Recombination Deficiency in Prostate Cancer with or without BRCA1/2 Mutations. Clin Cancer Res. 2020 Jun 1;26(11):2673–80. doi:10.1158/1078-0432.CCR-19-2135 PubMed PMID: 32071115; PubMed Central PMCID: PMC8387086.

204. Westphalen CB, Fine AD, André F, Ganesan S, Heinemann V, Rouleau E, et al. Pan-cancer Analysis of Homologous Recombination Repair–associated Gene Alterations and Genome-wide Loss-of-Heterozygosity Score. Clinical Cancer Research. 2022 Apr 1;28(7):1412–21. doi:10.1158/1078-0432.ccr-21-2096

205. Knijnenburg TA, Wang L, Zimmermann MT, Chambwe N, Gao GF, Cherniack AD, et al. Genomic and Molecular Landscape of DNA Damage Repair Deficiency across The Cancer Genome Atlas. Cell Reports. 2018 Apr 3;23(1):239-254.e6. doi:10.1016/j.celrep.2018.03.076

206. Smeby J, Kryeziu K, Berg KCG, Eilertsen IA, Eide PW, Johannessen B, et al. Molecular correlates of sensitivity to PARP inhibition beyond homologous recombination deficiency in pre-clinical models of colorectal cancer point to wild-type TP53 activity. eBioMedicine. 2020 Sep;59:102923. doi:10.1016/j.ebiom.2020.102923

207. Mendes-Pereira AM, Martin SA, Brough R, McCarthy A, Taylor JR, Kim J, et al. Synthetic lethal targeting of PTEN mutant cells with PARP inhibitors. EMBO Molecular Medicine. 2009 Sep 1;1(6–7):315–22. doi:10.1002/emmm.200900041

208. Dedes KJ, Wetterskog D, Mendes-Pereira AM, Natrajan R, Lambros MB, Geyer FC, et al. PTEN Deficiency in Endometrioid Endometrial Adenocarcinomas Predicts Sensitivity to PARP Inhibitors. Sci Transl Med. 2010 Oct 13;2(53). doi:10.1126/scitranslmed.3001538

209. Miao C, Tsujino T, Takai T, Gui F, Tsutsumi T, Sztupinszki Z, et al. RB1 loss overrides PARP inhibitor sensitivity driven by RNASEH2B loss in prostate cancer. Sci Adv. 2022 Feb 18;8(7):eabl9794. doi:10.1126/sciadv.abl9794 PubMed PMID: 35179959; PubMed Central PMCID: PMC8856618.

210. Saner FAM, Takahashi K, Budden T, Pandey A, Ariyaratne D, Zwimpfer TA, et al. Concurrent RB1 Loss and BRCA Deficiency Predicts Enhanced Immunologic Response and Long-term Survival in Tubo-ovarian High-grade Serous Carcinoma. Clinical Cancer Research. 2024 Aug 15;30(16):3481–98. doi:10.1158/1078-0432.CCR-23-3552

211. Vogel A, Haupts A, Kloth M, Roth W, Hartmann N. A novel targeted NGS panel identifies numerous homologous recombination deficiency (HRD)-associated gene mutations in addition to known BRCA mutations. Diagn Pathol. 2024 Jan 6;19(1):9. doi:10.1186/s13000-023-01431-8

212. 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 Dec 8;6(23):e152789. doi:10.1172/jci.insight.152789

213. Fettke H, Dai C, Kwan EM, Zheng T, Du P, Ng N, et al. BRCA-deficient metastatic prostate cancer has an adverse prognosis and distinct genomic phenotype. EBioMedicine. 2023;95.

214. Sokol ES, Pavlick D, Frampton GM, Ross JS, Miller VA, Ali SM, et al. Pan-Cancer Analysis of CDK12 Loss-of-Function Alterations and Their Association with the Focal Tandem-Duplicator Phenotype. Oncologist. 2019 Dec;24(12):1526–33. doi:10.1634/theoncologist.2019-0214 PubMed PMID: 31292271; PubMed Central PMCID: PMC6975947.

215. Incorvaia L, Bazan Russo TD, Gristina V, Perez A, Brando C, Mujacic C, et al. The intersection of homologous recombination (HR) and mismatch repair (MMR) pathways in DNA repair-defective tumors. npj Precision Oncology. 2024 Sep 5;8(1):190. doi:10.1038/s41698-024-00672-0

216. Farmanbar A, Firouzi S, Kneller R, Khiabanian H. Mutational signatures reveal ternary relationships between homologous recombination repair, APOBEC, and mismatch repair in gynecological cancers. Journal of Translational Medicine. 2022 Feb 2;20(1):65. doi:10.1186/s12967-022-03259-0

217. Farmanbar A, Kneller R, Firouzi S. Mutational signatures reveal mutual exclusivity of homologous recombination and mismatch repair deficiencies in colorectal and stomach tumors. Scientific Data. 2023 Jul 1;10(1):423. doi:10.1038/s41597-023-02331-8

218. Mehra N, Dolling D, Sumanasuriya S, Christova R, Pope L, Carreira S, et al. Plasma Cell-free DNA Concentration and Outcomes from Taxane Therapy in Metastatic Castration-resistant Prostate Cancer from Two Phase III Trials (FIRSTANA and PROSELICA). European Urology. 2018 Sep;74(3):283–91. doi:10.1016/j.eururo.2018.02.013

219. Aggarwal R, Huang J, Alumkal JJ, Zhang L, Feng FY, Thomas GV, et al. Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. JCO. 2018 Aug 20;36(24):2492–503. doi:10.1200/JCO.2017.77.6880

220. Chabon JJ, Hamilton EG, Kurtz DM, Esfahani MS, Moding EJ, Stehr H, et al. Integrating genomic features for non-invasive early lung cancer detection. Nature. 2020;580(7802):245–51.

221. Bai J, Jiang P, Ji L, Lam WKJ, Zhou Q, Ma MJL, et al. Histone modifications of circulating nucleosomes are associated with changes in cell-free DNA fragmentation patterns. Proc Natl Acad Sci USA. 2024 Oct 15;121(42):e2404058121. doi:10.1073/pnas.2404058121

222. Liu Y. At the dawn: cell-free DNA fragmentomics and gene regulation. Br J Cancer. 2022 Feb 1;126(3):379–90. doi:10.1038/s41416-021-01635-z

223. Gupta K. Transcriptional and Chromatin Disorientation Associated with BRCA2 Inactivation [Internet]. Apollo - University of Cambridge Repository; 2023 [cited 2025 Jul 11]. Available from: https://www.repository.cam.ac.uk/handle/1810/353610 doi:10.17863/CAM.99709

224. Rao S, Han AL, Zukowski A, Kopin E, Sartorius CA, Kabos P, et al. Transcription factor-nucleosome dynamics from plasma cfDNA identifies ER-driven states in breast cancer. Sci Adv. 2022 Aug 26;8(34):eabm4358. doi:10.1126/sciadv.abm4358 PubMed PMID: 36001652; PubMed Central PMCID: PMC9401618.

225. De Sarkar N, Patton RD, Doebley AL, Hanratty B, Adil M, Kreitzman AJ, et al. Nucleosome Patterns in Circulating Tumor DNA Reveal Transcriptional Regulation of Advanced Prostate Cancer Phenotypes. Cancer Discovery. 2023 Mar 1;13(3):632–53. doi:10.1158/2159-8290.cd-22-0692

226. Bayona-Feliu A, Herrera-Moyano E, Badra-Fajardo N, Galván-Femenía I, Soler-Oliva ME, Aguilera A. The chromatin network helps prevent cancer-associated mutagenesis at transcription-replication conflicts. Nat Commun. 2023 Oct 28;14(1). doi:10.1038/s41467-023-42653-0