Lee *et al. Clinical Proteomics* (2025) 22:13
https://doi.org/10.1186/s12014-025-09536-6

Clinical Proteomics

RESEARCH

Open Access

Circulatory prostate cancer proteome landscapes and prognostic biomarkers in metastatic castrate resistant prostate cancer

Hyejung Lee[1], Jincheng Shen[1], Muhammad Zaki Fadlullah[2], Anna Neibling[3], Claire Hanson[4], Enos Ampaw[4], Tengda Lin[1], Matt Larsen[4], Jennifer Lloyd[4], Benjamin L. Maughan[4], Umang Swami[4], Sumati Gupta[4], Jonathan Tward[5], Skyler B. Johnson[5], Brock O'Neil[6], Bogdana Schmidt[6], Christopher B. Dechet[6], Benjamin Haaland[1], Liang Wang[7], Aik-Choon Tan[2] and Manish Kohli[4*]

Abstract

Background Plasma-based high-plex proteomic profiling were performed in prostate cancer (PC) patients using the Olink® Explore Proximity Extension Assay to identify plasma proteins associated in different PC states and to explore potential prognostic biomarkers. The progressive PC states include local, organ-confined PC (local PC), metastatic hormone-sensitive PC (mHSPC) and metastatic castrate-resistant PC (mCRPC).

Methods Plasma samples were uniformly processed from 84 PC patients (10 patients with local PC; 29 patients with mHSPC; 45 patients with mCRPC). A proteome-wide association study was performed to identify proteins differentially overexpressed in progressive cancer states. Specifically, a sequential screening approach was employed where proteins overexpressed from one disease state were assessed for overexpression in the progressive disease state. Linear regression, analysis of variance, and t-tests were used for this approach. Differentially expressed proteins (DEPs) in mCRPC were then used to construct a prognostic model for overall survival (OS) in mCRPC patients using the Cox Proportional Hazard Model. The predictive performance of this model was assessed using time-dependent area under the receiver operating characteristic curves (tAUC) in an independent sample of mCRPC patients. The tAUC of the prognostic model was then compared to that of a model excluding DEPs to evaluate the added value of circulatory proteins in predicting survival.

Results Of 736 tumor-associated proteins, 26 were differentially expressed across local PC, mHSPC, and mCRPC states. Among these, 20 were overexpressed in metastatic states compared to local, and in mCRPC compared to mHSPC states. Of these 20 proteins, Ribonucleoside-diphosphate reductase subunit M2 (RRM2) was identified as a prognostic biomarker for OS in mCRPC, with a hazard ratio of 2.30 (95% confidence interval (CI) 1.17–4.51) per normalized expression unit increase. The tAUC of the model including previously identified clinical prognostic factors was 0.62 (95% CI 0.29–0.91), whereas the model that includes RRM2 with clinical prognostic factors was 0.87 (95% CI 0.51–0.98).

Conclusions Plasma proteome profiling can identify novel circulatory DEPs associated with mCRPC state survivals. Overexpression of RRM2 is linked to poor mCRPC survival and its inclusion alongside conventional prognostic factors enhances the predictive performance of the prognostic model.

Keywords Prostate cancer, Proteomes, Landscape, Prognostic model

*Correspondence:
Manish Kohli
Manish.kohli@hci.utah.edu
Full list of author information is available at the end of the article

Introduction

Prostate cancer (PC) is a significant cause of mortality, accounting for over 35,000 deaths among males in the United States and over 325,000 deaths globally in 2024 [1]. The predominant clinical management of metastatic prostate cancer (mPC) is based on disruption of the androgen receptor-signaling axis, which, prior to 2015, was achieved with single-agent androgen deprivation therapy (ADT). Since then, therapeutic management for metastatic hormone-sensitive prostate cancer (mHSPC) state has evolved rapidly, and ADT-based combinations with multiple drugs are the standard of care [2]. While the therapeutic landscape has quickly evolved, monitoring clinical response or identifying recurrence after treatment is conducted largely by measuring serum prostate-specific antigen (PSA) protein levels in the blood. PSA monitoring has been used extensively, but the prognostic and predictive power of single or serial PSA levels in metastatic states has yet to be validated for clinical use. Multivariable clinical factors like performance status, presence of visceral sites of metastasis, combined with non-specific proteins including serum alkaline phosphatase and serum lactate dehydrogenase (LDH) levels, have demonstrated prognostic accuracy [3] in metastatic castrate-resistant prostate cancer (mCRPC), but these are not tumor-biology specific markers. Integrative proteomic profiling using high-throughput mass spectrometry in prostate cancer tissue specimens across different states of progressive disease has revealed distinctive state-specific protein profiles [4]. Progressive state-specific genomic alterations using plasma cell-free DNA in prostate cancer from a blood sample has been performed [5], however sensitive proteome-based technologies, such as liquid chromatography-mass spectrometry (LC–MS)-based global quantitative proteomics has been challenging to perform for identification of novel tumor biology blood-based protein candidates with prognostic or predictive utility in different states of tumor progression [6]. Candidate proteins specifically overexpressed in mCRPC may potentially capture a diverse tumor-host integrated landscape compared to serum PSA alone [7]. We used high-throughput proteomic profiling in plasma based on next-generation sequencing (NGS) approaches to identify blood-based proteomic candidate biomarkers. The differential abundance of secreted cancer-related proteins in progressive cancer states was evaluated in prostate cancer patients with local organ-confined cancer, metastatic hormone-sensitive prostate cancer (mHSPC), and mCRPC states.

Methods

Patient cohort: A real-world clinically annotated plasma biobank obtained from PC patients visiting a single tertiary cancer center (Huntsman Cancer Institute, University of Utah) was used while receiving standard-of-care cancer management. This biobank includes research blood samples collected under an IRB-approved protocol (IRB# 00089989; IRB# 00139755) and clinical information collection. Blood biospecimens were collected for research in the biobank after written informed consent was obtained, and all samples collected were prospectively collected and uniformly processed. Clinical outcomes including patient demographics and survival outcomes were obtained from electronic medical records.

An initial "discovery" cohort of cross-sectional PC patients enrolled in four different states of clinical progression between August 2020 and February 2022 representing a continuum of clinical progression constituted of local PC (Cohort A), mHSPC (Cohort B), mCRPC experiencing only biochemical relapse (defined as serially rising PSA on continuous mHSPC treatments without the appearance of new image-based metastasis) (Cohort C) and mCRPC with patients progressing with clinical/imaging-based criteria in addition to biochemical relapse (Cohort D). The goal was to determine the differential protein expression over disease state progression. To minimize treatment effects on protein expression, only cohort-specific pre-treatment samples were analyzed. Details of specimen processing of the blood samples collected from all patients in the registry are provided under "Patient Methods" section of the "Supplementary Methods." An independent secondary cohort of 23 pre-treatment mCRPC patients, enrolled between August 2020 and September 2023, was collected to assess the predictive performance of prognostic models identified from the discovery data.

Proteomic methods: The tumor-associated proteins were preselected by Olink® Explore from widely used public-access bioinformatic databases, including UniProt, the Human Protein Atlas (HPA), Gene Ontology (GO), and DisGeNET [8]. The list of proteins included from these databases on the Olink® Explore platform was based on biomarkers that passed Olink's validation tests and requirements, as previously reported [8]. These proteins were measured using Olink® Explore 3072 (Olink Proteomics AB, Uppsala, Sweden) according to the manufacturer's instructions by Psomagen Inc. (Rockville, MD) for the discovery dataset and by Olink® for the secondary dataset. The secondary dataset was

processed using the intensity normalization method for batch correction and was bridged to the discovery dataset [9] to ensure the NPX values of each protein from the discovery and secondary datasets were comparable. The Olink® protocol is based on a Proximity Extension Assay (PEA) [10] coupled with readout via NGS for measuring protein profiles from 80 μL of plasma. Technical details of the NGS-based PEA are provided in the "Proteomic Methods" section of "Supplementary Methods." Readouts of the PEA measurements are reported as "Normalized Protein Expression" (NPX) units. The NPX is a unit of protein expression to quantify the relative amount of the protein expressed among the samples being analyzed. One-unit increase in NPX equates to a twofold increase in protein concentration. Further details for calculating protein abundance based on the NPX values are provided in the Proteomic Methods section of "Supplementary Methods."

Study approach and statistical methods

We hypothesized that differentially expressed proteins (DEPs) are associated with progressive PC states. To explore state-specific expression patterns, we utilized 736 tumor-associated plasma proteins measured on the Olink® Explore 3072 platform across four patient cohorts. Proteins overexpressed in specific cancer progressive states were considered for further evaluation as potential prognostic biomarkers. Consequently, the aims of this study were to (1) identify proteins that are gradually overexpressed over the course of disease progression, and (2) assess the prognostic significance of these proteins in predicting overall survival (OS) in mCRPC.

A schematic overview of our approach is presented in Fig. 1, outlining the study's stepwise methodology through four interconnected objectives. In Objective 1, we identified DEPs across three PC states: localized PC (Cohort A), mHSPC (Cohort B), and mCRPC (Cohorts C + D). Each of the 736 proteins (measured in NPX units) were analyzed using ordinary least squares linear regression, with the protein expression regressed on a trivariate grouping variable representing the three PC states. Analysis of Variance (ANOVA) was conducted to test for differences in protein expression among these states. The false discovery rate (FDR) was controlled at 10% (q-values < 0.1) [11], and DEPs with the q-value < 0.1 were classified as significantly overexpressed in a particular state of progression.

In Objective 2, we focused on the identified DEPs in Objective 1 that were overexpressed in metastatic cohorts (mHSPC + mCRPC) relative to local PC. The aim was to pinpoint DEPs associated with the transition metastatic disease states compared to localized PC. Welch's one-sided t-tests [12] were conducted to compare protein expression levels between local and metastatic cohorts, with the alternative hypothesis positing higher mean expression in the metastatic group. DEPs with p-values ≤ 0.1 were considered overexpressed in metastatic PC. Adjustment for multiple testing was not performed for this objective, as the proteins had already been identified to be differentially expressed with FDR adjustment in Objective 1. This objective can be thought as an extension of Objective 1, where the location and direction of the differential expression were identified.

In Objective 3, we analyzed identified DEPs associated with metastatic prostate cancer states which were overexpressed in mCRPC compared to mHSPC state. This allowed for the identification of DEPs exhibiting progressively higher expression in the metastatic state. Welch's one-sided t-tests were conducted, with the alternative hypothesis suggesting higher expression in mCRPC state. DEPs meeting the p-value ≤ 0.1 were classified as overexpressed in the mCRPC state. The FDR correction was not applied for the same reason as stated for objective 2.

In the final Objective 4, we determined the prognostic value of DEPs overexpressed in mCRPC for predicting OS in this mCRPC cohort. Thus, analyses were restricted to the mCRPC patients, with OS defined from the date of blood sample collection to either death or administrative censoring (August 31, 2022, for the discovery dataset; April 4, 2024, for the secondary dataset). Each protein was analyzed individually using Cox proportional hazards (PH) models while adjusting for key clinical prognostic factors [3, 13, 14]. The clinical prognostic factors in mCRPC state included as binary variables were: albumin levels below the first sample quartile (Q1), hemoglobin levels below Q1, alkaline phosphatase (ALP) levels above the third sample quartile (Q3), prostate-specific antigen (PSA) levels above Q3, and LDH levels > 222 U/L, the upper limit of normal [15]. If any of the clinical prognostic factors were missing, they were assumed to be missing completely at random (MCAR) [16]; therefore, no imputation was performed, and only complete cases were included in the analyses.

mCRPC DEPs with positive coefficients in the Cox models were interpreted as having an association with an elevated risk of death. Since the primary focus of this study was to identify hazardous overexpressed proteins, only those with positive coefficients and q-values < 0.1 (to account for multiple testing) were classified as marginally prognostic and hazardous for survival in mCRPC.

To account for the simultaneous expression of multiple proteins a multivariable Cox PH model was fitted to estimate their effects in the presence of one another. Protein identified as hazardous based on the individual

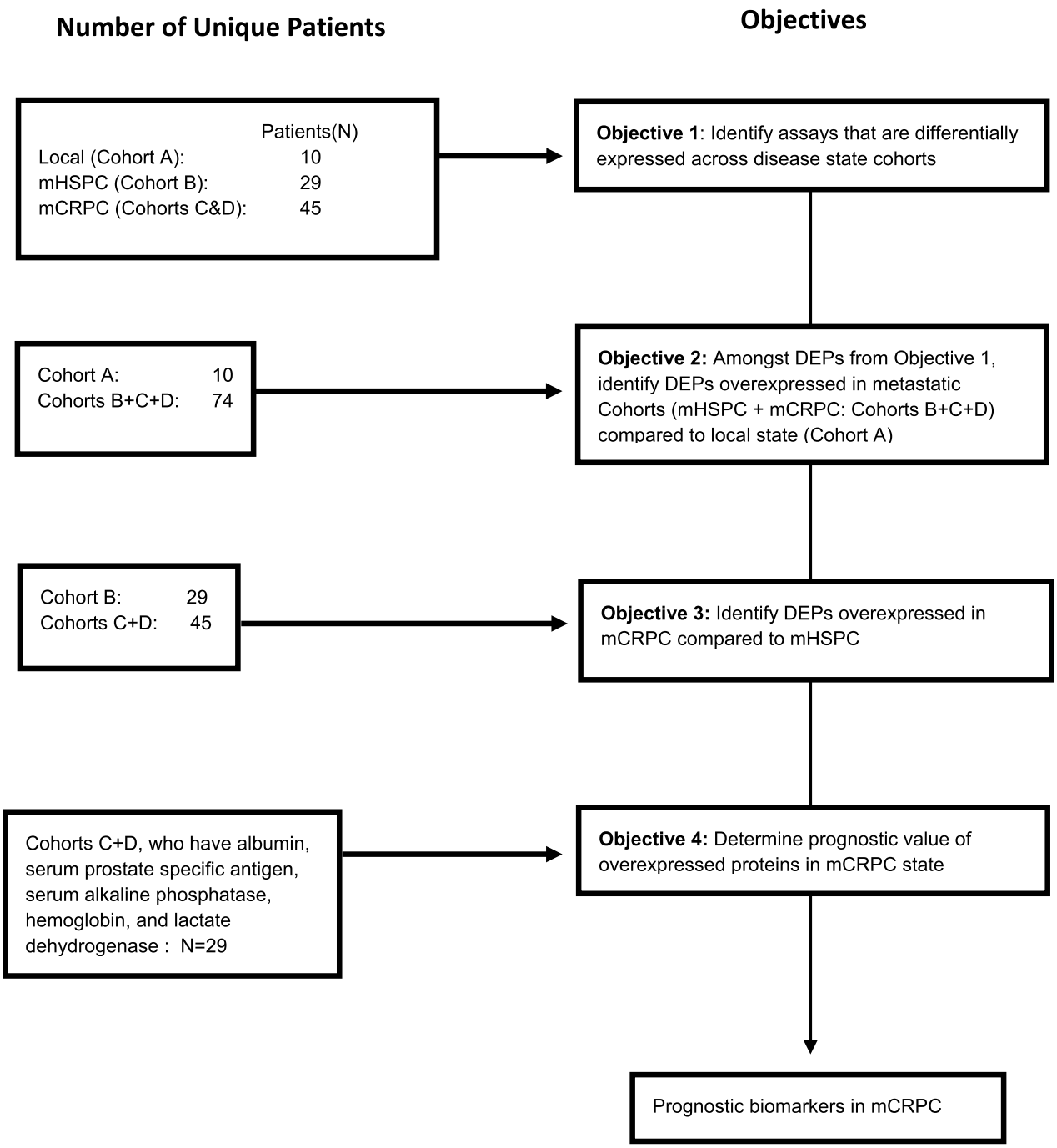

Fig. 1 Study approach and objectives: Cohort A consists of plasma samples from patients with local, organ-confined stage prostate cancer. Cohort B included plasma from metastatic hormone-sensitive prostate cancer patients, and Cohort C included plasma from patients with metastatic castrate-resistant prostate cancer who experienced biochemical relapse (defined as serially rising Serum Prostate Specific Antigen (PSA) on continuous metastatic Hormone-Sensitive Prostate Cancer (mHSPC) treatments) without new clinical evidence of progression. Cohort D included patients with metastatic castration-resistant prostate cancer who progressed with clinical/imaging-based criteria in addition to biochemical relapses. (DEPs: Differentially Expressed Proteins)

Cox models were included in this multivariable Cox PH model, along with the clinical risk factors (i.e., albumin levels below the first quartile for the range, hemoglobin levels below the first quartile for the range, ALP levels above the third quartile for the range, PSA levels above the third quartile for the range, and LDH levels > 222 U/L). To address potential collinearity, pairs of proteins with high correlation (Pearson correlation > 0.8) were managed by removing the protein with the weaker association with survival, determined by a higher p-value for the estimated coefficient in the multivariable Cox PH model. This model is hereafter referred to as the "integrated" model.

A comparative Cox PH model, referred to as the "clinical" model, was also fitted, which included only the currently known non-specific clinical prognostic factors. The predictive performance of the integrated and clinical models was assessed using time-dependent area under the curve (tAUC) [17] on the secondary dataset. The tAUC values, ranging from 0 to 1, measures the concordance between predicted and observed survival times, with higher values indicating greater predictive accuracy. Bias-adjusted bootstrap confidence intervals, derived from 10,000 replicates, were calculated to account for asymmetry in the bootstrap estimates [18]. All statistical analyses were performed using R software (version 4.4.1).

Results

The discovery dataset included 84 unique PC patients in various states of disease progression with plasma PEA-NGS sequencing. Table 1 provides a summary of the demographic characteristics of patients in this discovery cohort. For metastatic states, high and low volumes were based on the ChemoHormonal Therapy Versus Androgen Ablation Randomized Trial for Extensive Disease in Prostate Cancer (CHAARTED) clinical trial definitions [19].

Overall distribution of targeted plasma proteins

The landscape of the expression-based heatmap (NPX value-based) for all 736 oncology proteins grouped by cohort is shown in Fig. 2. The NPX values were individually standardized and truncated for comparison and graphical purposes. We note that 72/736 proteins on the preselected Olink® platform that have been previously reported in the published literature to be associated with prostate tumor biology. The names of these 72 proteins are listed in Supplementary Table SI, and they appear on the heatmap in the top 72 rows.

DEPs in PC progressive states

Among the 736 proteins, 26 proteins were identified as DEPs across the three cancer progressive states (local PC, mHSPC, and mCRPC) (Objective 1). Heatmap of these 26 proteins grouped by the three cancer progressive states are shown in Fig. 3. Glial cell line-derived neurotrophic factor family receptor alpha-1 (GFRA1) was among the 26 DEPs that was also in the 72 proteins that have been previously reported in the published literature to be associated with prostate tumor biology [20, 21]. Supplementary Table SII shows p-values from ANOVA tests of all 736 proteins and their corresponding q-values, in the order of smallest q-value at the top. Violin plots of NPX expression grouped by the three states for all 26 DEPs are shown in Supplementary Figure S1.

Twenty of the 26 DEPs were overexpressed in the metastatic states compared to local (Objective 2) with all of these 20 proteins overexpressed in mCRPC compared to mHSPC state (Objective 3). Supplementary Figure S2 shows the 20 proteins in the order from the greatest mean difference in NPX between the local and metastatic states. Supplementary Table SIII lists the mean and standard deviation of the proteins, in both the local and metastatic states, and their respective p-values from the Welch's t-test. Violin plots of the 20 DEPs are shown in Supplementary Figure S3 in the order of greatest to the smallest difference in the mean NPX value between the mHSPC and mCRPC. Supplementary Table SIV lists the mean and standard deviation of the proteins, in both the mHSPC and mCRPC states, and their respective p-values from the Welch's t-test.

Association of DEPs with mCRPC survival

Among the 45 mCRPC patients in the discovery dataset, sixteen did not have at least one clinical prognostic factor listed in the electronic medical records. Thus, the remaining 29 patients with no missing data were used to build a prognostic survival model. The median mCRPC survival for the cohort was 22.90 months (range 1.94–23.46 months). Among the 20 overexpressed DEPs in the mCRPC, ribonucleoside-diphosphate reductase subunit M2 (RRM2), iron-sulfur cluster scaffold homolog, mitochondrial (NFU1), and Leucine zipper protein 2 (LUZP2) and Jupiter microtubule associated homolog 2(JPT2) were associated with survival based on the individual Cox PH models (HR 2.30 (95% CI 1.17–4.51) per 1 NPX increase for RRM2, p-value = 0.02; HR 0.18 (95% CI 0.03–0.97) per 1 NPX increase for LUZP2, p-value = 0.05; HR 1.96 (95% CI 1.08–3.54) per 1 NPX increase for NFU1, p-value = 0.03). Of these three proteins, RRM2 and NFU1 had increased hazard ratios for poor survival and were included in multivariable Cox PH model. However, these two proteins were highly correlated (Pearson

Table 1 Patient cohort demographics

	Localized Disease (Cohort A) (N = 10)	Tx Naive mHSPC (Cohort B) (N = 29)	Biochemical mCRPC (Cohort C) (N = 8)	Clinical mCRPC (Cohort D) (N = 37)
Age				
Median [Min, Max]	68.0 [59.0, 79.0]	73.0 [58.0, 85.0]	67.5 [53.0, 82.0]	71.0 [56.0, 85.0]
Albumin (grams/deciliter)				
Median [Min, Max]	NA [NA, NA]	NA [NA, NA]	4.0 [3.8, 4.6]	3.9 [3.5, 4.5]
Missing	10 (100%)	29 (100%)	1 (12.5%)	8 (21.6%)
PSA (ng/mL)				
Median [Min, Max]	9.4 [2.9, 38.1]	12.2 [0.3, 1401.8]	4.8 [0.2, 54.9]	8.8 [0.1, 1863.6]
Hemoglobin (grams/deciliter)				
Median [Min, Max]	NA [NA, NA]	NA [NA, NA]	14.5 [13.0, 15.7]	12.5 [8.9, 15.0]
Missing	10 (100%)	29 (100%)	1 (12.5%)	5 (13.5%)
ALP (U/L)				
Median [Min, Max]	86.0 [68.0, 104.0]	77.0 [47.0, 987.0]	65.0 [46.0, 151.0]	117.0 [41.0, 639.0]
Missing	8 (80.0%)	0 (0%)	1 (12.5%)	0 (0%)
Gleason score at initial diagnosis				
< 8	7 (70.0%)	10 (34.5%)	2 (25.0%)	13 (35.1%)
> = 8	3 (30.0%)	18 (62.1%)	6 (75.0%)	23 (62.2%)
Missing	0 (0%)	1 (3.4%)	0 (0%)	1 (2.7%)
De novo metastatic stage				
Metastatic	0 (0%)	8 (27.6%)	3 (37.5%)	18 (48.6%)
Nonmetastatic	0 (0%)	19 (65.5%)	5 (62.5%)	19 (51.4%)
Missing	10 (100%)	2 (6.9%)	0 (0%)	0 (0%)
CHAARTED volume at the time of first metastasis				
Low	0 (0%)	19 (65.5%)	6 (75.0%)	18 (48.6%)
High	0 (0%)	9 (31.0%)	2 (25.0%)	17 (45.9%)
Unknown	0 (0%)	0 (0%)	0 (0%)	1 (2.7%)
Missing	10 (100%)	1 (3.4%)	0 (0%)	1 (2.7%)
Systemic treatments for mCRPC patients				
Androgen receptor pathway inhibitors (ARPIs)	None	None	8	20
Chemotherapy	None	None	0	17

PSA: Serum Prostate Specific Antigen; ALP: Serum Alkaline Phosphatase; CHAARTED: Chemo-Hormonal Therapy Versus Androgen Ablation-Randomized Trial for Extensive Disease in Prostate Cancer; mHSPC: metastatic Hormone-Sensitive Prostate Cancer; mCRPC: metastatic Castrate Resistant Prostate Cancer

correlation = 0.83). Thus, we initially fitted the final model with both RRM2 and NFU1 with known clinical prognostic factors but removed NFU1 because of high correlation for RRM2 which had stronger significance (p-value of NFU1 = 0.39; p-value of RRM2 = 0.16). The final Cox PH model was built on RRM2 and the five clinical prognostic factors of PSA, serum albumin, hemoglobin, alkaline phosphatase and serum LDH.

Marginal association between RRM2 and survival is shown in Supplementary Figure S4A using Kaplan–Meier (KM) curve in the 29 patients. RRM2 was categorized into tertiles based on distribution of the NPX levels. As can be observed in Figure S4A, patients with lower RRM2 expression were observed to have longer survival, while those with higher expression had shorter survival (log-rank test with $p < 0.0001$) (Objective 4). Interestingly, we explored NPX-values for RRM2 in five metastatic prostate cancer patients with neuroendocrine pathology and found that median NPX-values were two-fold higher when compared to the 29 mCRPC patients, although no formal statistical analyses was performed because of the small number of patients with confirmed neuroendocrine pathology.

In the second mCRPC cohort, one patient was removed due to missing LDH data, leaving 22/23 patients for assessing the predictive performance of the integrated and clinical prognostic models. The median survival for the secondary cohort was 30.88 months (range 5.26–31.93 months). The distribution of all clinical prognostic factors and RRM2 values for the

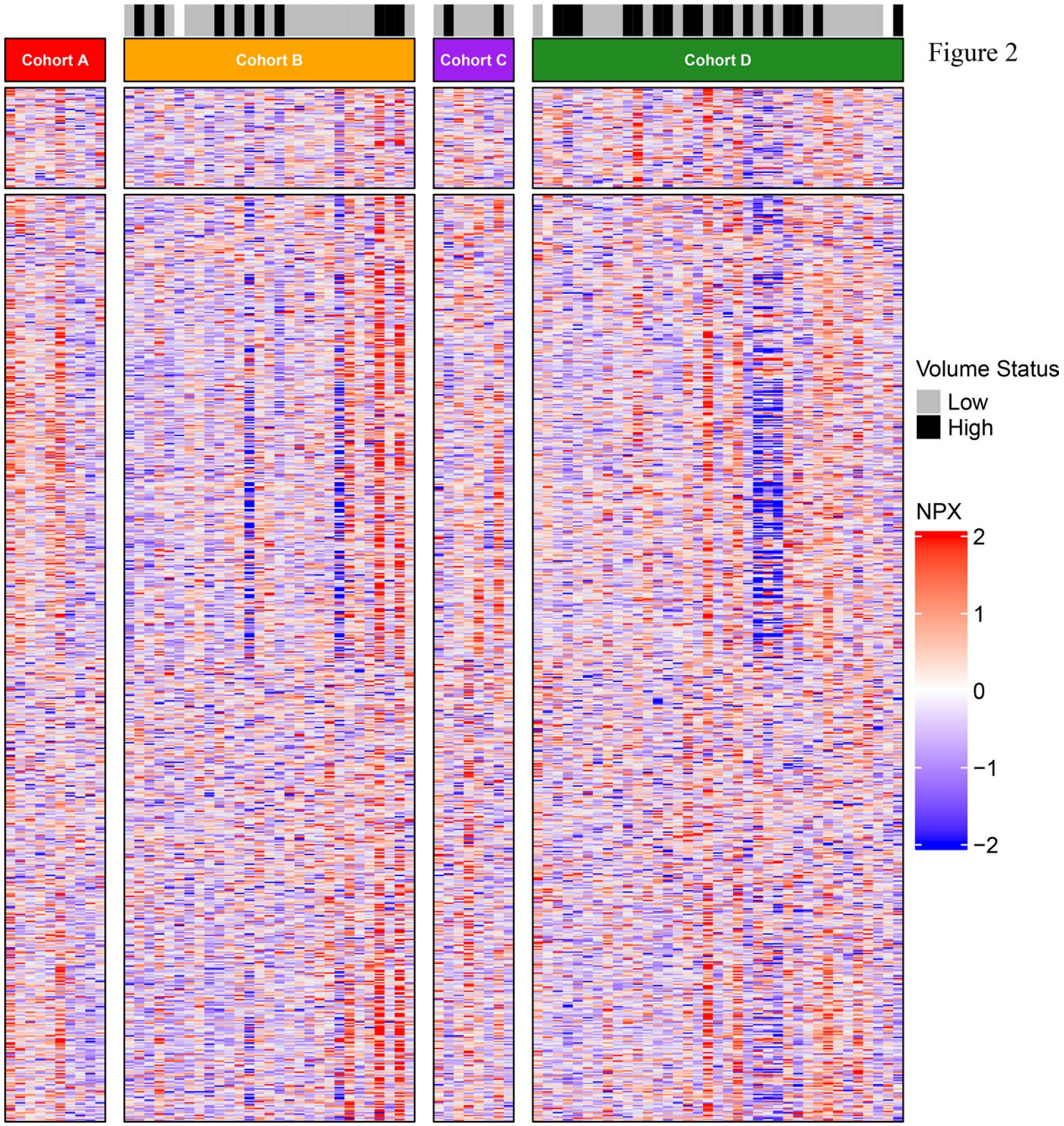

Fig. 2 Heatmap of the normalized protein expression (NPX) of 736 oncology proteins grouped by cohorts: The top 72 rows represent proteins that have been previously reported to be associated with prostate cancer biology and are also listed in the Olink® panel. The names of these 72 proteins can be found in the Supplementary Table SI. This is followed by a blank row, below which the NPX expression patterns of the remaining proteins are listed in the panel. Each column represents a patient with protein expression in red representing overexpression and low expression in blue using NPX units. The NPX values were individually standardized by subtracting the sample mean and dividing by the sample standard deviation to facilitate comparability across proteins. To address the issue of extreme values dominating the color scale in heatmap visualizations—thereby obscuring patterns in the data—the standardized values were truncated at -2 and 2. This truncation ensured that the coloring scale remained interpretable and effectively highlighted meaningful variations

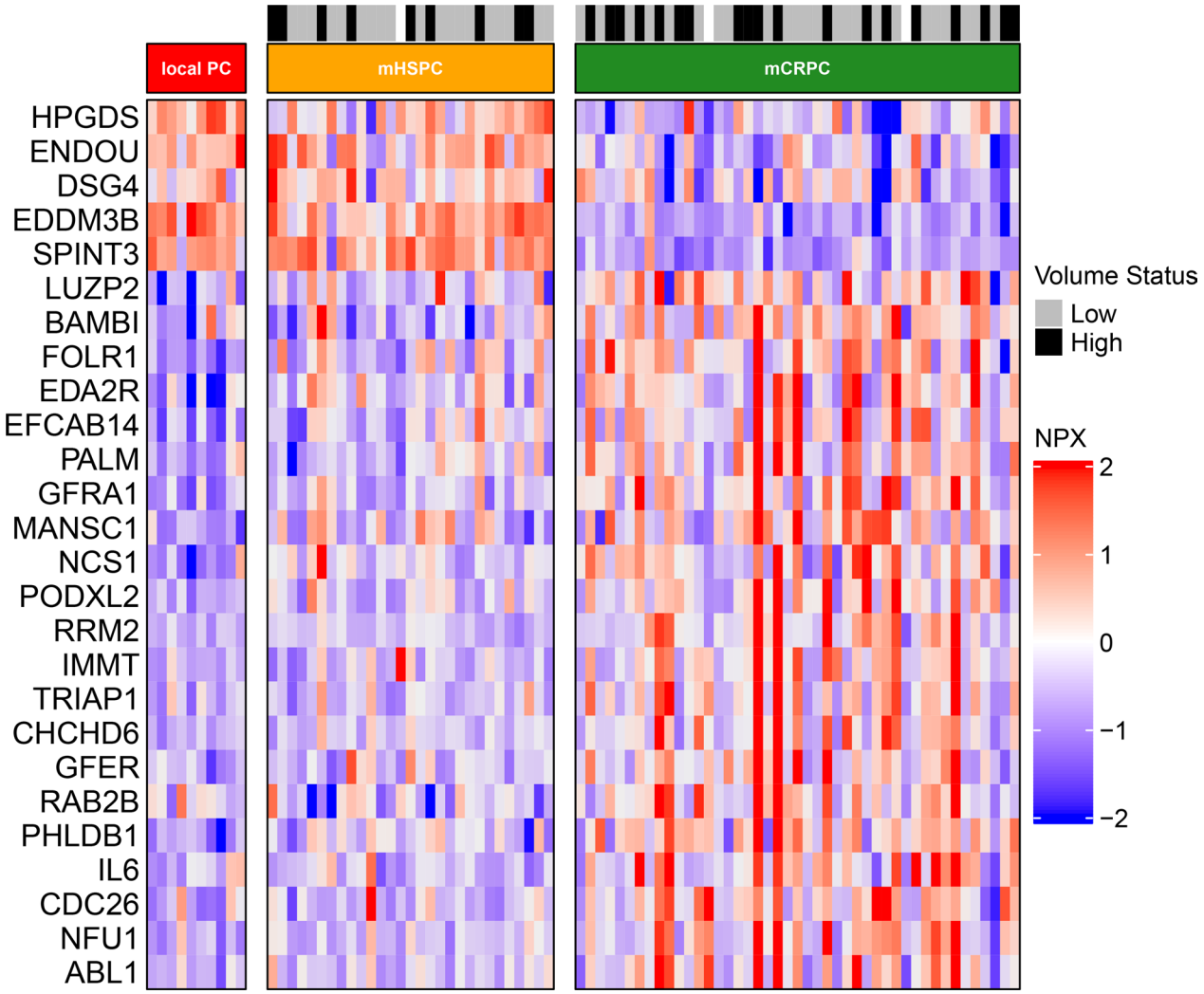

Fig. 3 Heatmap of the normalized protein expression (NPX) of 26 differentially expressed proteins grouped by three cancer progressive states: Heatmap of NPX values for 26 differentially expressed proteins identified in Objective 1. The rows represent proteins and columns represent patient sample. Proteins are clustered using Euclidean distance. Patient samples are grouped by the three cancer progressive states as done in Objective 1. The NPX values are individually standardized by subtracting the sample mean and dividing by the sample standard deviation to facilitate comparability across proteins

discovery and secondary mCRPC cohorts is shown in Supplementary Table V. Supplementary Figure S4B illustrates the survival of the secondary mCRPC cohort based on RRM2, as observed in the discovery dataset (Supplementary Figure S4A). The tAUC (95% CI) of the integrated and clinical models were 0.87 (95% CI 0.51–0.98) and 0.62 (95% CI 0.29–0.91), respectively with the integrated model observed to have a higher point estimate of tAUC with greater precision.

Discussion

Our goal of this study was to initially define the prostate cancer circulatory proteome landscape using a large number of tumor associated proteins and then to identify mCRPC-state associated proteins that have potential prognostic relevance to improve on current non-specific protein biomarkers. Specific tumor-biology associated prognostic proteins in mCRPC state may not only serve to improve current prognostic models but also serve as targets for drug delivery. We were able to observe 26/736 DEPs, of which 20 were increasingly overexpressed with metastatic states. Of these 20, overexpression of RRM2 as

a biomarker for overall survival of mCRPC was observed to be significant and integrating RRM2 expression in the two independent mCRPC cohorts with non-specific clinical prognostic factors was observed to increase predictive accuracy for survival as observed by tAUCs (tAUC = 0.87, 95% CI 0.51–0.98) over current clinical model (tAUC = 0.62, 95% CI 0.29–0.91). Gene Ontology annotations for RRM2 include oxidoreductase activity, ribonucleoside-diphosphate reductase activity, and thioredoxin disulfide as acceptor [22, 23]. These functions make RRM2 a critical component in DNA synthesis, which is essential for both cell replication and repair [24]. Previously the overexpression of RRM2 as a rate-limiting enzyme involved in DNA repair and synthesis has been evaluated as a prognostic biomarker in oral squamous cell carcinoma using tissue microarrays and it was observed that compared with normal and dysplastic tissues, the expression of RRM2 in human primary oral squamous cell carcinomas was significantly increased, and its overexpression was correlated with advanced pathological grade, recurrence and poor survival [25]. Our study is the first to identify a link between circulatory RRM2 and prostate cancer progression, as well as its potential as a prognostic biomarker in mCRPC patients.

The strength of this study is that we determined a circulatory proteome landscape using a large set of proteins across prostate cancer states to identify candidate proteins differentially overexpressed in each state of progression. We used advancements in sequencing technology enabling quantification of circulatory proteins in plasma and combined it with statistical rigor to demonstrate an approach that identifies significant proteins in PC progression and biomarkers in survival. Using novel proteome-profiling technology to relatively non-invasive biospecimen collections has not been previously performed to quantify proteins in plasma. The DEP identified RRM2, appears to enhance the predictive accuracy of the prognostic model for mCRPC patients when included in addition to the currently used non-tumor biology associated clinical prognostic factors. We also observed that the RRM2 integrated prognostic model was able to predict survival in a separate, independent mCRPC cohort. The use of prospectively collected uniformly processed plasma samples and not the use of "samples of convenience" is another strength of the study as it mitigates pre-analytic bias.

However, there are limitations of the study for generalizability as the findings are based on a small sample size (84 prostate cancer patients in the discovery cohort and 22 in the secondary mCRPC cohort), which highlights the need to reproduce the findings in larger prospectively collected randomized study cohorts. Our approach to find proteins overexpressed over the course of disease progression employ a sequential screening approach. Using a one-step modeling strategy to identify overexpressed proteins across the three disease states may be an alternate approach but may not yield differentially expressed proteins relevant to a state of progression and we were driven by dual objectives: first, to identify proteins that are increasingly more expressed from early to late stages, and second, to determine key proteins involved in transition to metastasis, and then are significant when hormonal therapy is no longer effective. Finally, the analysis focused solely on proteins linked to tumor pathways, excluding those related to inflammation and metabolism due to the project's early stage. However, cancer-related metabolic and inflammatory pathways could also provide proteomic biomarkers that influence outcomes.

In conclusion, we determined tumor biology and state-specific protein biomarkers in mCRPC state to enhance prognosis. Defining mCRPC lethality based on molecular protein biomarkers linked to cancer biology which currently relies on non-specific proteins lacking tumor-pathway associations is likely to provide prognostic value and identify potential novel targets for drug therapies, if these can be validated in independent mCRPC cohorts.

Abbreviations

PC	Prostate cancer
PEA	Proximity Extension Assay
local PC	Local, organ-confined prostate cancer
mHSPC	Metastatic hormone-sensitive prostate cancer
mCRPC	Metastatic castrate-resistant prostate cancer
DEPs	Differentially expressed proteins
OS	Overall survival
tAUC	Time-dependent area under the receiver operating characteristic curves
CI	Confidence interval
mPC	Metastatic prostate cancer
ADT	Androgen deprivation therapy
LDH	Lactate dehydrogenase
LC–MS	Liquid chromatography–mass spectrometry
NGS	Next-generation sequencing
HPA	Human Protein Atlas
GO	Gene Ontology
ANOVA	Analysis of variance
FDR	False discovery rate
PH	Proportional hazards
ALP	Serum alkaline phosphatase
MCAR	Missing completely at random
RRM2	Ribonucleoside-diphosphate reductase subunit M2
GFRA1	Glial cell line-derived neurotrophic factor family receptor alpha-1
Q1	First sample quartile
Q3	Third sample quantile
U/L	Units per liter
NFU1	NFU1 iron-sulfur cluster scaffold homolog, mitochondrial
JPT2	Jupiter microtubule associated homolog 2

KM Kaplan–Meier
LUZP2 Leucine zipper protein 2

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12014-025-09536-6.

Supplementary material 1. Supplementary Figure S1: "Protein expression levels of all assays in local, mHSPC, and mCRPC states": Violin plots of NPX expression of 26 assays across different states of progression found to be differently expressed in at least one of the three states (local, mHSPC, mCRPC).

Supplementary material 2. Supplementary Figure S2: "Protein expression levels in local and metastatic states": Violin plots of 20 assays that are overexpressed in metastatic state compared to the local state.

Supplementary material 3. Supplementary Figure S3: "Protein expression levels in mHSPC and mCRPC states": Violin plots of 20 assays that are more expressed in mCRPC compared to the mHSPC state.

Supplementary material 4. Supplementary Figure S4: "Kaplan-Meier survival plot of mCRPC patients, grouped by RRM2 NPX expression": RRM2 is categorized into three different groups, consisting of values lower than 1st quartile (green), interquartile range (violet), and higher than 3rd quartile (orange). Figure A) is for the discovery dataset, and the expression level of RRM2 has visibly negative association with survival probability. Figure B) is for the secondary dataset, and RRM2 expression exceeding 3rd quantile has consistently smaller probabilities over the other groups, while the survival probabilities of IQR and > Q3 expression groups are crossing over time. In both subplots, the log rank test for median survival shows significantly difference in median survival (p-value < 0.0001 for discovery data set and p-value = 0.01 for secondary data set).

Supplementary material 5. Supplementary Table SI: "PC Proteins (N = 72)": List of 72 proteins that have been reported in published literature to be associated with prostate cancer biology which were also integrated on the pre-selected Olink® platform. Protein name column shows the full name show the protein that is synthesized from the corresponding gene shown in the Gene name column. Supplementary Table SII: "DEP p-values (N = 736)": P-values from ANOVA testing differential expression of the protein expression (measured in NPX) across local, mHSPC, and mCRPC states. Their corresponding q-values are also provided. The table is ordered from the most significant q-value at the top to the least at the bottom. The top 26 assays are significant with q-value cut off at 0.1. The interpretation of the q-value is as follows: The p-value of PALM is 0.0007, with a corresponding q-value of 0.0384. This indicates that the minimum false positive discovery rate is 38.4% when rejecting all p-values ≤ 0.0007. This interpretation applies similarly to all other q-values. Supplementary Table SIII: "Metastatic V. Local (N = 20)": Comparison of protein expression levels between local and metastatic states. Only the 20 proteins that are identified to be overexpressed in the metastatic state are shown (Welch's one sided t-test with p-value < 0.1). For each protein (Assay), the sample mean and standard deviation in the metastatic state (Metastatic mean (sd)) and local state (Local mean (sd)) are shown, along with the p-value of Welch's t-test (p-value). The table is ordered from the greatest mean difference in NPX between the local and metastatic states. Supplementary Table SIV: "mHSPC Vs. mCRPC (N = 20)": Comparison of protein expression levels between mHSPC and mCRPC states. All 20 proteins that are found to be overexpressed in metastatic states (as shown in Supplementary Table SIII) were also found to be overexpressed in mCRPC compared to mHSPC states (Welch's one-sided t-test with p-value < 0.1). The sample mean and standard deviations of each state are provided, along with the p-value from Welch's t-test. The table is ordered from the greatest mean difference in NPX between the mCRPC and mHSPC states. Supplementary Table SV: "Data comparison": Descriptive summary of RRM2 (the protein identified to hold prognostic value in mCRPC patients) and clinical risk factors of the discovery and secondary datasets. Clinical risk factors are dichotomized using cut-off values derived from the discovery dataset sample quantiles. These cut-off values were subsequently applied to the secondary dataset, rather than recalculating quantiles within the secondary dataset itself. Binary classifications of clinical risk factors include albumin < Q1, prostate-specific antigen > Q3, hemoglobin < Q1, alkaline phosphatase > Q3, and lactate dehydrogenase > 222 U/L, all of which are associated with increased hazard of death in prostate cancer patients. A higher proportion of albumin, prostate-specific antigen, and lactate dehydrogenase risk factors were observed in the discovery dataset, whereas a higher proportion of hemoglobin and alkaline phosphatase risk factors were observed in the secondary dataset. Additionally, the mean expression of RRM2 was higher in patients from the secondary dataset

Supplementary material 6. Supplementary Methods: Supplementary methods on patient and Olink® data processing.

Acknowledgements
None.

Author contributions
JL, BLM, US, SG, JT, SBJ, BON, BS, CBD, LW, MK acquired patients and AN, CH, ML, EA processed samples or clinical data. HL, BH MK designed the study. TL, EA provided a descriptive summary of the acquired data and database access. HL performed the statistical analysis and provided interpretation. JS and BH provided guidance for statistical analysis. MZF and ACT contributed to discussions of study design and interpretation of results, reviewed the manuscript, and provided critical feedback to enhance its clarity and scientific rigor. MK, JS, and HL drafted the manuscript or substantively revised it. Study oversight was performed by MK. All authors commented on and provided feedback and final approval on the manuscript.

Funding
From institutional start-up funds made available by Department of Internal Medicine and Huntsman Cancer Institute to MK.

Availability of data and materials
Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate
This biobank includes research blood samples collected under an IRB-approved protocol with clinical outcomes after obtaining a written informed consent (IRB# 00089989; IRB# 00139755) at the University of Utah.

Consent for publication
All patients involved in this study provided written consent for research publication.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Population Health Sciences, Division of Biostatistics, University of Utah, Salt Lake City, UT, USA. [2]Departments of Oncological Sciences and Biomedical Informatics, University of Utah-Huntsman Cancer Institute, Salt Lake City, UT, USA. [3]Population Sciences Study Support, University of Utah-Huntsman Cancer Institute, Salt Lake City, UT, USA. [4]Division of Oncology, Department of Internal Medicine, University of Utah-Huntsman Cancer Institute, Salt Lake City, UT, USA. [5]Department of Radiation Oncology, University of Utah-Huntsman Cancer Institute, Salt Lake City, UT, USA. [6]Department of Surgery, Division of Urology, University of Utah-Huntsman Cancer Institute, Salt Lake City, UT, USA. [7]Department of Tumor Microenvironment and Metastasis, Moffitt Cancer Center, Tampa, FL, USA.

Received: 28 February 2025 Accepted: 7 April 2025
Published online: 18 April 2025

References

1. Sung H, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2021;71:209–49.
2. Omrcen T. Systemic triple therapy in metastatic hormone sensitive prostate cancer (Mhspc). Acta Clin Croat. 2022;61(Suppl 3):81–5.
3. Halabi S, et al. External validation of a prognostic model of overall survival in men with chemotherapy-naive metastatic castration-resistant prostate cancer. J Clin Oncol. 2023;41(15):2736–46.
4. Latonen L, et al. Integrative proteomics in prostate cancer uncovers robustness against genomic and transcriptomic aberrations during disease progression. Nat Commun. 2018;9(1):1176.
5. Kohli M, et al. Clinical and genomic insights into circulating tumor DNA-based alterations across the spectrum of metastatic hormone-sensitive and castrate-resistant prostate cancer. EBioMedicine. 2020;54: 102728.
6. Bhawal R, et al. Challenges and opportunities in clinical applications of blood-based proteomics in cancer. Cancers. 2020;12(9):2428.
7. Kohli M, et al. Serum proteomics on the basis of discovery of predictive biomarkers of response to androgen deprivation therapy in advanced prostate cancer. Clin Genitourin Cancer. 2019;17(4):248–53.
8. Proteomics O. How was the protein library content for Olink Explore selected? - Olink. 2023. https://olink.com/faq/how-was-the-protein-library-for-explore-content-for-olink-expl-selected/. Accessed 3 July 2023.
9. Introduction to bridging Olink® NPX datasets. 2024. https://cran.r-project.org/web/packages/OlinkAnalyze/vignettes/bridging_introduction.html. Accessed 26 April 2024.
10. Assarsson E, et al. Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. PLoS ONE. 2014;9(4):e95192–e95192.
11. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci USA. 2003;100(16):9440–5.
12. Welch BL. The generalization of `Student's' problem when several different population variances are involved. Biometrika. 1947;34(1/2):28–35.
13. Halabi S, et al. Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. J Clin Oncol. 2014;32(7):671–7.
14. Halabi S, et al. Prognostic model predicting metastatic castration-resistant prostate cancer survival in men treated with second-line chemotherapy. J Natl Cancer Inst. 2013;105(22):1729–37.
15. Farhana A, Lappin SL. Biochemistry, lactate dehydrogenase. 2023.
16. Little RJA, Rubin DB. Statistical analysis with missing data. 2014.
17. Uno H, et al. Evaluating prediction rules for t-year survivors with censored regression models. J Am Stat Assoc. 2007;102(478):527–37.
18. Davison AC, Hinkley DV. Bootstrap Methods and their Application. 1997. Series Number 1: p. x-x.
19. Sweeney CJ, et al. Chemohormonal therapy in metastatic hormone-sensitive prostate cancer. N Engl J Med. 2015;373(8):737–46.
20. Meng M, Wu YC. LMX1B activated circular RNA GFRA1 modulates the tumorigenic properties and immune escape of prostate cancer. J Immunol Res. 2022;2022:7375879.
21. Huber RM, et al. DNA damage induces GDNF secretion in the tumor microenvironment with paracrine effects promoting prostate cancer treatment resistance. Oncotarget. 2015;6(4):2134–47.
22. Botstein D, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25(1):25–9.
23. Aleksander SA, et al. The Gene Ontology knowledgebase in 2023. Genetics. 2023;224(1): iyad031.
24. Stelzer G, et al. The GeneCards suite: from gene data mining to disease genome sequence analyses. Curr Protoc Bioinform. 2016;2016(1):1.30.1-1.30.33.
25. Wang S, et al. Overexpression of RRM2 is related to poor prognosis in oral squamous cell carcinoma. Oral Dis. 2021;27(2):204–14.

Publisher's Note