

AWARD NUMBER: W81XWH-20-1-0644

TITLE: Driver Gene Networks of Genomic Instability in Prostate Cancer Progression

PRINCIPAL INVESTIGATOR: Dr. Sungyong You, PhD

CONTRACTING ORGANIZATION: Cedars-Sinai Medical Center

REPORT DATE: DECEMBER 2023

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE DECEMBER 2023		2. REPORT TYPE Final		3. DATES COVERED 15 AUG 2020 – 14 AUG 2023	
4. TITLE AND SUBTITLE Driver Gene Networks of Genomic Instability in Prostate Cancer Progression				5a. CONTRACT NUMBER W81XWH-20-1-0644	
				5b. GRANT NUMBER GRANT12937731	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S) Dr. Sungyong You, PhD E-Mail: Sungyong.You@cshs.org				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) CEDARS-SINAI MEDICAL CENTER 8700 BEVERLY BLVD LOS ANGELES CA 90048-1804				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Development Command Fort Detrick, Maryland 21702-5012				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT The less stable the cancer gene is, the faster cancer worsens. This can be used as an important measure of cancer's characteristics in various human cancers. By comparing and analyzing tumors at the genetic or gene expression level that is highly unstable in cancerous cells, we can identify changes in factors that are important to the various genetic manifestations that result in increased genomic instability (GI). This talk will describe the GI driver networks that can distinguish disease subtypes. Progress was made in four key points in the first year of the funding period. I have 1) developed a novel method for GI score computation to identify PCs with highly altered genome, 2) identified the genes (PCGI-E) significantly correlated with GI events, 3) Identified MRs that are relevant to PCGI-E gene signature regulation, and 4) developed PCGI TRN classifier that can identify the PCs with poor survival outcomes.					
15. SUBJECT TERMS Genomic Instability, Driver Network, Prostate Cancer, Transcriptome, Circulating Tumor Cell					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRDC
U	U	U	UU	81	19b. TELEPHONE NUMBER (include area code)

TABLE OF CONTENTS

	<u>Page</u>
1. Introduction	4
2. Keywords	4
3. Accomplishments	4
4. Impact	30
5. Changes/Problems	30
6. Products	31
7. Participants & Other Collaborating Organizations	31
8. Special Reporting Requirements	32
9. Appendices	32

1. INTRODUCTION

Prostate cancer (PC) is a major leading cause of death in men in US. Genomic rearrangement, copy number changes, and increased mutational burden are well characterized features that correlate highly with PC progression. However, drivers of genomic instability (GI) and their relationships in PC progression are elusive. Specific oncogene mutations such as SPOP and TP53 mutations can alter gene expression and promoting other genomic abnormalities. Recent studies suggest that tumors harboring genomic alterations are significantly associated with gene expression phenotypes linked with PC subtypes and loss/gain of function mutations of genetic drivers have been associated with a characteristic transcriptional phenotype.

This project is testing the hypotheses that transcriptional programs governing the disease progression are distinct between PCGI high and low tumors and that the activity of GI driver networks allows us to make clinical predictions about prognosis and likely treatment sensitivity. To this end we identified and assessed the GI driver network concept through systems biology approach to available PC genomic and transcriptomic profiles from multiple cohorts. During the second year of the funding period, we will identify common and distinct transcriptional programs between SPOP and TP53 driven drivers' networks.

Our specific aims have **not** modified from those stated in the original application.

Specific Aim 1. Determine whether the transcriptional programs are distinct between PCGI high and low tumors.

Specific Aim 2. Determine common and distinct transcriptional programs between the two mutually exclusive GI driver networks.

Specific Aim 3. Test whether the GI driver networks evolve in pre/post treatment using PC transcriptome.

2. KEYWORDS

Genomic Instability, Driver Network, Prostate Cancer, Transcriptome, Circulating Tumor Cell

3. ACCOMPLISHMENTS

What were the major goals of the project?

1. 1st Year Tasks

Research Goal 1: Identify transcriptional signature (PCGI-E) describing the specific transcriptional programs in each PCGI high and low tumors.

Milestones:

- 1) PCGI scores for all the tumor samples.
- 2) The most significant genes as a PCGI-E signature in PC genomic alterations.

Target months: 9

Percentage of completion: 100%

Research Goal 2: Develop the PCGI TRN classifier to predict metastatic progression and assess clinical association with PCGI TRN classifier.

Milestones:

- 1) A TRN describing the interaction of the PCGI dependent TFs and their targets in the DEGs.
- 2) A PCGI TRN classifier.
- 3) A PCGI TRN classifier correlated with clinical outcomes.

Target months: 12

Percentage of completion: 100%

2. 2nd Year Tasks

Research Goal 1: Assessment of mutual exclusivity of GI driver networks driven by TP53 and SPOP.

Milestones:

- 1) Testing the mutual exclusivity of SPOP and TP53 aberrations.

Target months: 3

Percentage of completion: 100%

Research Goal 2: Identification of transcriptional signatures of the novel GI drivers and reconstruction of TRNs activated by either TP53-network or SPOP-network.

Milestones:

- 1) Transcriptional signatures of SYNE1 loss or TNFRSF21 gain in PC.
- 2) Two classifiers for TP53-network and SPOP-network.

Target months: 12

Percentage of completion: 100%

3. 3rd Year Tasks

Research Goal 1: Assessment of the association between the GI driver network activities and clinical outcomes.

Milestones:

- 1) Correlation with the GI driver network activities and clinical outcomes in PCGI high tumors.
- 2) Association between GI driver-specific TFs and ARSI response in the CSMC CTC cohort.

Target months: 12

Percentage of completion: 100%

What was accomplished under these goals?

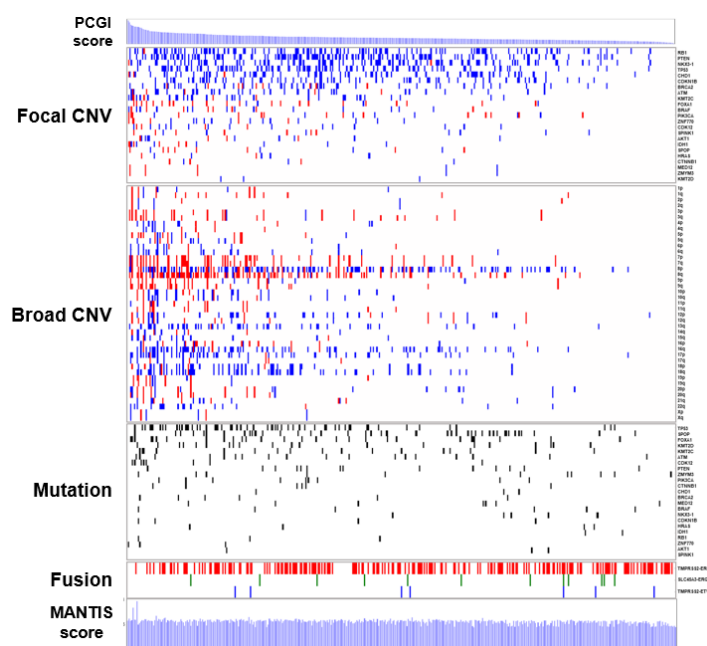
We have developed a novel method to quantify degree of genomic alterations in a single PC sample using five different types of GI events, including broad and focal CNAs, genomic rearrangements, MSI, and somatic mutational burden. Using this, we identified potential links from the genetic alterations to secondary oncogenic driver activations resulting in a large number of transcriptional changes, thereby

Major accomplishments of this funding cycle include:

1. Specific Aim 1: Determine whether the transcriptional programs are distinct between PCGI high and low tumors.

1.1. Major Task 1: Computation of PCGI score using the TCGA, Taylor et al., MET360 cohorts.

1.1.1. Subtask 1: Compute scores of five different types of genomic alterations including focal CNAs, broad CNAs, tumor mutational burden, genomic fusion, and microsatellite instability.



PCGI computations in TCGA PRAD: We have developed a novel scoring method of GI through summation of broad and focal copy number alterations (CNAs), microsatellite instability (MSI), tumor mutational burden (TMB), and genomic rearrangements (fusion). Categories of genomic alteration events were identified in individual tumor samples and summed in order to generate a numerical PCGI score after scaling by Min-Max normalization method. Recurrent genomic alterations were frequent in samples with high PCGI scores (**Figure 1**).

Figure 1. PCGI scores in TCGA PRAD cohort. A heatmap in TCGA ordered by PCGI score; the frequency of broad CNAs and focal CNAs (gain in red, losses in Blue), mutations, and fusion.

1.1.2. Subtask 2: Combine six genomic alteration scores into one score as PCGI score

We identified Identify transcriptional signature (PCGI-E) describing the specific transcriptional programs in each PCGI high and low tumors.

PCGI computations in TCGA PRAD, MET440, and Taylor et al. cohorts: We have obtained genomic alteration profiles of TCGA¹, MET440² and Taylor et al³ cohorts due to the limited availability of the raw data from public databases and applied modified version of PCGI (pseudo-PCGI; here after pPCGI) computation method, which does not include MSI in the score computation because it is weakly correlated to the overall PCGI score with less variability (**Figure 2A**). To assess the use of pPCGI instead of PCGI, we have checked the correlation between PCGI and pPCGI scores in TCGA PRAD cohort. As shown in **Figure 2B**, it has strong correlation (Spearman's rho = 0.85, P<0.001).

Given this, we have computed pPCGI scores for all 3 cohorts including TCGA, MET440, and Taylor et al. Of note, two cohorts MET440 and Taylor et al. including metastases have more samples with higher pPCGI score, while TCGA has a smaller number of samples with high pPCGI due to the samples are from primary prostate tissues (**Figure 2C**). This suggests that the higher PCGI is the more metastases.

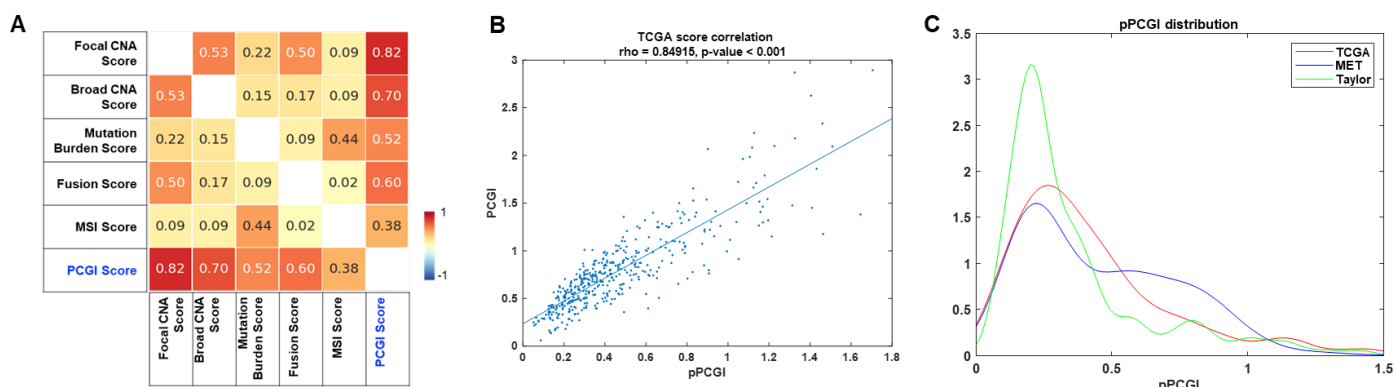


Figure 2. PCGI and pPCGI scores in 3 PC cohorts. A. Correlation between original PCGI score and individual genomic alterations in TCGA PRAD cohort. Correlation was measured by using Spearman's method. B. Comparison of original PCGI and pseudo-PCGI (pPCGI) scores in TCGA PRAD cohort. C. pPCGI distribution of TCGA, MET440, and Taylor et al. cohorts.

1.2. Major Task 2: Identification of differentially expressed genes (DEGs) between PCGI high and low tumors using the TCGA, Taylor et al., and MET360 cohorts.

1.2.1. Subtask 1: Identify DEGs between PCGI high and low tumors and perform functional enrichment analysis of the DEGs

Identification of significantly differentially expressed genes between pPCGI high and low tumors:

In order to identify differentially regulated genes by the genomic alterations. We first divided the patients three groups PCGI high, intermediate, and low groups at upper and lower tiles of the pPCGI score in each cohort. We then compared PCGI high and low groups to select the most significantly differentially expressed genes between the tumors highly genomically altered and less altered. For this, integrated hypothesis testing was employed. Briefly, we computed T, median difference, and Rank-Sum stats. Significance levels of individual stats were obtained from individual null distribution and combined them using Stouffer's method⁴, resulting in combined

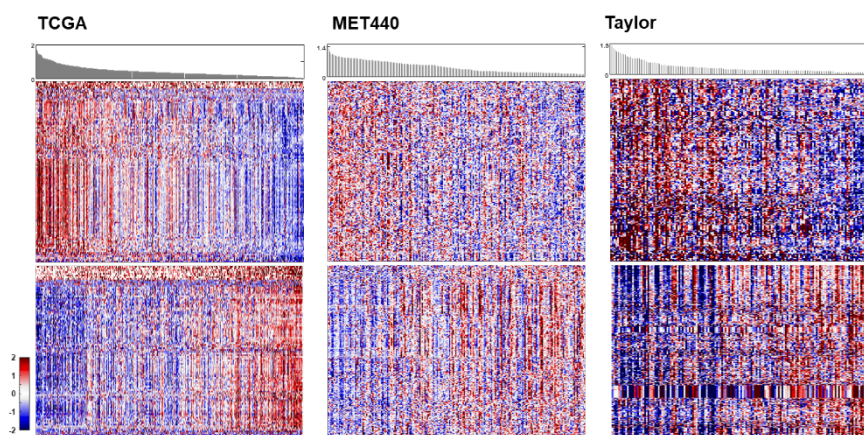


Figure 3. Heatmap of z-score normalized gene expression that are differentially expressed between pPCGI high and low. Top color bars indicate the pPCGI score of each data set. The differentially expressed genes indicate up-regulated genes (up panel) and down-regulated genes (down panel) between pPCGI high and low groups in TCGA (left), MET440 (middle), and Taylor et al., (right). Each column in the heatmap is an individual sample.

p-values. Multiple correction with BH method was performed with the combined p-values. Differentially expressed genes (DEGs) were selected with false discovery rate (FDR) < 0.05 and log2-fold-change 1 (TCGA) or 0.58 (MET440 and Taylor et al.) according to deviation of the log2-fold-change distribution. Finally, we have identified 107, 193, and 201 up-regulated genes in pPCGI high tumors compared to pPCGI low tumors in TCGA, MET440, and Taylor et al., respectively. We also identified 99, 234, and 337 down-regulated genes in the same comparisons in TCGA, MET440, and Taylor et al., respectively. The heatmaps display the differential expression pattern of the genes according to the PCGI scores (**Figure 3**).

1.2.2. Subtask 2: Assess whether the DEGs have consistent expression patterns in Taylor et al., and MET360 data

Comparing GI driven transcriptional changes between TCGA, MET440, and Taylor et al. cohorts: We then compared those DEGs using Venn Diagram to see the common and distinct DEGs between the cohorts. As shown in Figure 3. Notably, there is less overlap of the DEGs between the 3 cohorts. Only 10 up-regulated genes are overlapped between TCGA and Taylor et al. (**Figure 4A**). Down-regulated genes are also a smaller overlapping gene between the cohorts (**Figure 4B**). However, functional enrichment of the DEGs using DAVID software⁵ revealed that these genes are mostly associated with cell cycle, chromatin segregation and mitosis as shown in **Figure 4C**. This indicates that the most correlated genes with GI in PC are associated with cell cycle progression and chromatin regulations. These DEGs is defined as PCGI-E gene signature representing transcriptional changes by GI in PC.

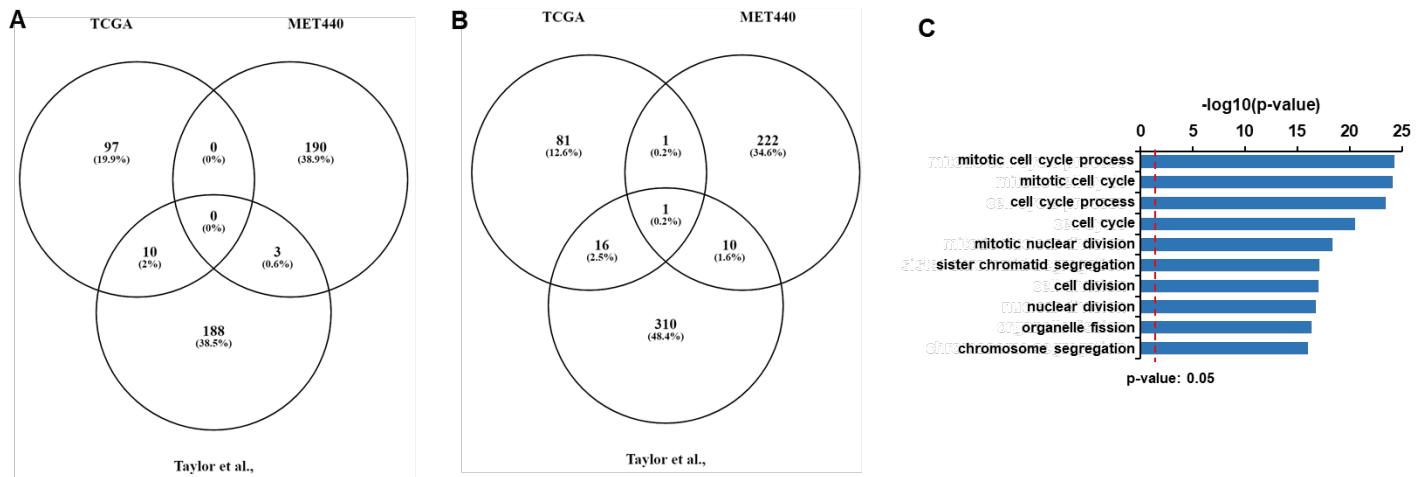


Figure 4. The Venn Diagrams of up- and down-regulated genes between pPCGI high and low tumors. A. Up-regulated genes. B. Down-regulated genes. C. Bar plot displays enrichment of the up-regulated genes.

1.3. Major Task 3: Construction of TRNs of PCGI.

1.3.1. Subtask 1: Identify PCGI dependent TFs

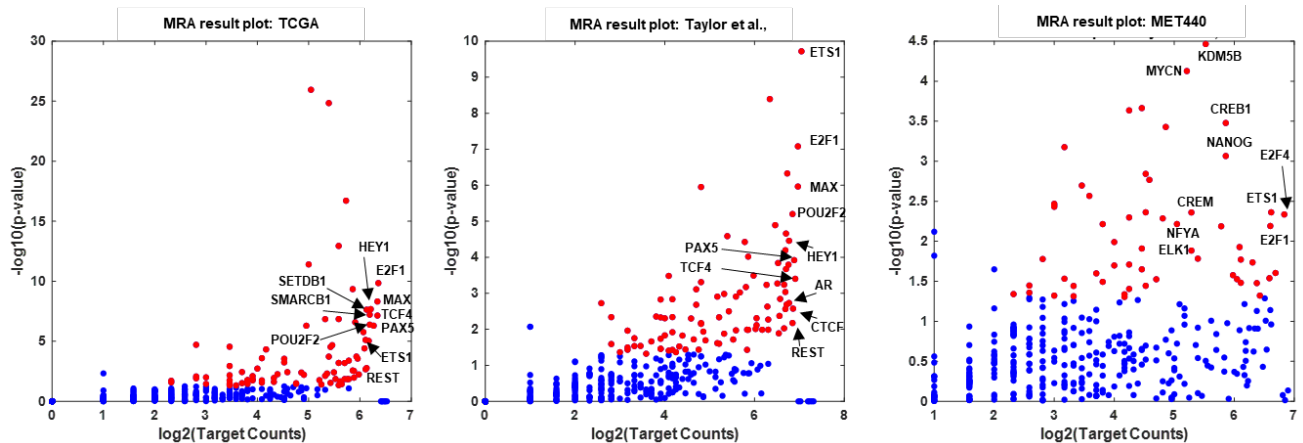


Figure 5. MRA results from TCGA (left), Taylor et al. (middle), and MET440 (right). Red dot indicates significant TFs and blue dot indicates insignificant TFs. The top 10 TFs were labeled with gene symbol.

We developed the PCGI TRN classifier to predict metastatic progression and assess clinical association with PCGI TRN classifier.

Master regulator analysis (MRA) of the DEGs: Given the DEGs, we performed MRA to identify the most relevant transcription factors (TFs) of the differential expression. For this, we have used in-house TF-target relationship database consisting of 799,657 TF-target interactions. Individual TF significance was tested by using null distribution based on 10,000 random sampling of target genes. Finally, we identified significant TFs for each cohort with significance level (p-value) < 0.01 and minimum number of targets ≥ 10 .

TFs were sorted by a ratio of number of targets in the DEGs over total number of targets and top 10 TFs were selected from the sorted list of TFs (**Figure 5**).

Common and distinct master regulators (MRs) and their functional associations: To assess whether transcriptional drivers driven by genomic alterations are different between the cohorts, we compared the top10 TFs from TCGA, MET440, and Taylor et al. cohorts. We found that 8 MRs are overlapped between TCGA and Taylor et al., while MET440 has only 2 MRs overlapped (**Figure 6A**). We further checked whether their target genes are overlapped like MRs overlap, however a small overlap was found (**Figure 6B**). Despite a small overlap of the target genes, their cellular processes are highly overlapped and enriched in cell cycle, chromatin segregation, and mitosis (**Figure 6C**) consistent with the DEG enrichment (**Figure 4C**). Such a high overlap between TCGA and Taylor cohorts could be explained by the sample origin. Those samples are from primary tissue in prostate, pertaining prostate tissue specific transcriptional regulations. However, MET440 samples are from distance tissues of metastases and its transcriptional regulation should be different from prostate lineage gene expression. However, it is still an interesting part of the process involved in cell cycle and chromatin regulations.

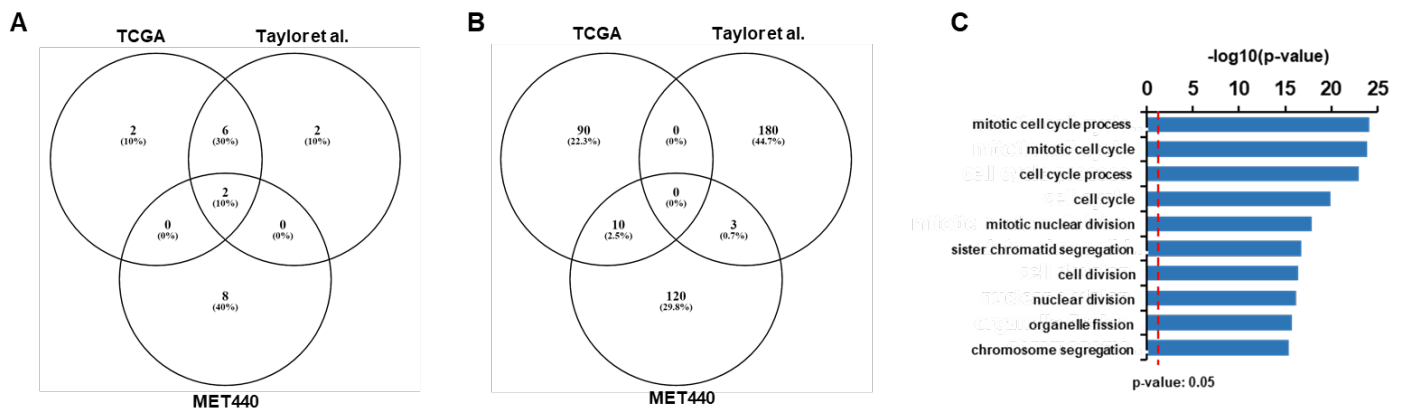


Figure 6. Top10 MRs and their functional association using their target genes. A. Venn Diagram depicts large overlaps of MRs between TCGA and Taylor et al. B. Venn Diagram depicts overlap of MRs' target genes. C. Bar plot displays enriched cellular processes based on the MRs' target genes.

1.3.2. Subtask 2: Test whether the activity of TF is different between PCGI high and low tumors.

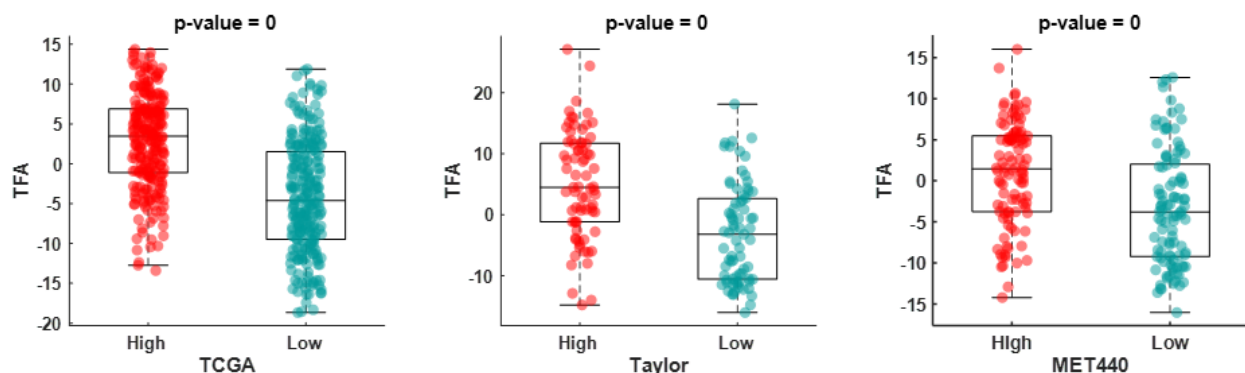


Figure 7. Boxplot of TFA in PCGI high and low tumors from TCGA (left), Taylor et al. (middle), and MET440 (right). p-values were determined by two-sided rank-sum test.

Assessment of the activity of TF (TFA) in PCGI high and low tumors: We determined whether common and distinct MRs have differential activity between PCGI high and low tumors. TFA were calculated the unweighted z-score method⁴ with MR's target gene expression profiles. As shown in **Figure 7**, PCGI high tumors showed high TFA in both primary and metastatic tumors.

1.3.3. Subtask 3: Construct TRNs for PCGI using the large compendium of TF-target interaction information

Reconstruction of the transcriptional regulatory network (TRN) models as a PCGI TRN network: We have identified a total 20 MRs through MRA of the 3 cohorts. We then reconstruct MRs network model describing the interactions between the MRs with distinct number of targets in different sample context such as primary vs. metastasis or prostate vs. non-prostate. Protein-protein interaction information was obtained from STRING database⁶ and construct network model with log2-fold-change between PCGI high and low tumors, as well as number of targets in the DEGs for each cohort. As shown in **Figure 8**, TCGA and Taylor cohorts from primary prostate tumor tissues have similar target enrichment and log2-fold-change pattern, while MET440 has very different target enrichment. These TRN model shows several AR interactors including ETS1 and KDM5B. ETS1 expression is highest in high-grade tumors and increased ETS1 expression and transcriptional activity promotes an aggressive and castrate-resistant phenotype in immortalized PC cells⁷. Histone Demethylase KDM5B (a.k.a. JARID1B) was found to be overexpressed in PC⁸ and it suppresses tumor suppressor genes such as TP53⁹. We also found that NFYA and SETDB1 are active in all 3 cohorts which are associated with mutant TP53¹⁰ or disruption of TP53¹¹. In addition to these, WNT and Notch signaling downstream regulators, TCF4, NANOG, HEY1, CREM and CREB1 are highly active in PCGI high tumors. WNT and Notch signaling interlinked to promote prostate progenitor cell proliferations and differentiation¹². In prostate cancer these MRs might associated with increased cancer cell plasticity, resulting in resistant to chemotherapy and/or radiotherapy¹³. Collectively, MR networks illustrates active transcriptional programs associated largely with 1) increase of castrate-resistant phenotype, 2) inhibition of tumor suppressors, and 3) lineage plasticity.

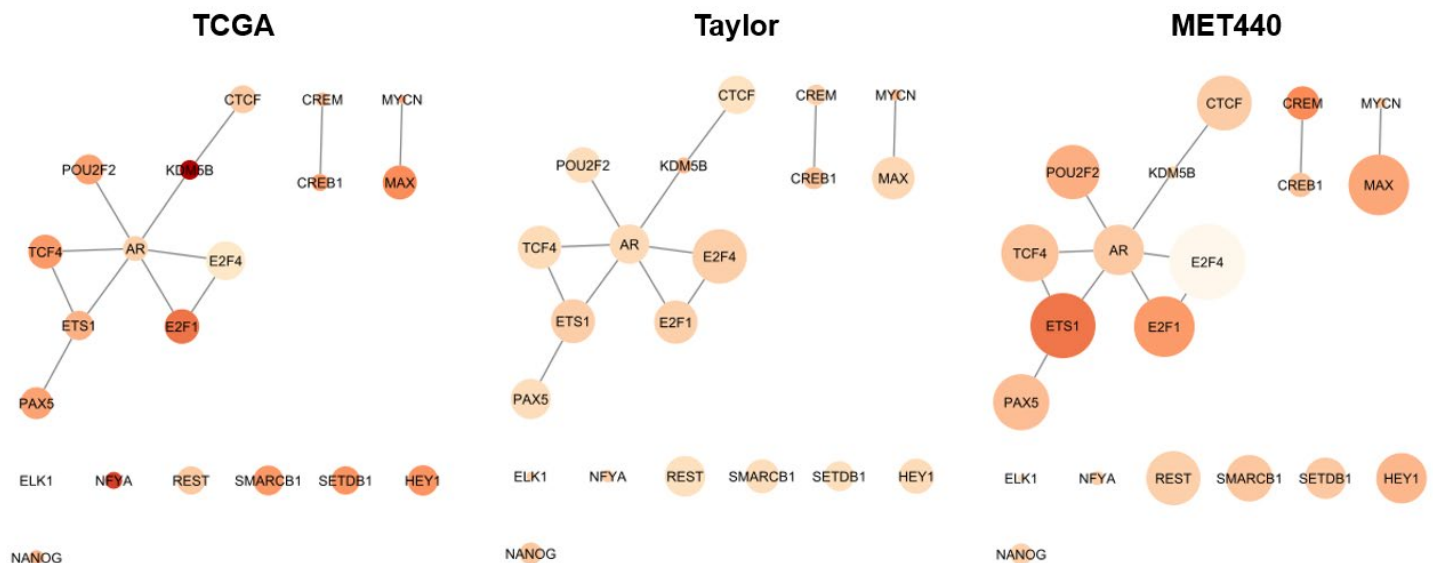


Figure 8. PCGI TRN models describing MRs interactions and their target gene regulation in each cohort. Node color: $-\log_{10}(p\text{-value})$. Node size: number of targets.

1.4. Major Task 4: Development of PCGI TRN based classifier.

1.4.1. Subtask 1: Feature extraction from the PCGI TRN

To perform feature extraction from PCGI TRN, genes exhibiting high correlation with PC GI scores were determined. Spearman's correlation tests were performed to extract the top 10 genes demonstrating the strongest correlations with the PC GI scores and the individual scores constituting the PC GI. These top genes were merged into a signature, which was utilized to compute a gene set Z-score⁶ as shown in **Figure 9**.

1.4.2. Subtask 2: Computation of TRN activation score using a machine learning algorithm.

Using the PCGI TRN, we computed PCGI TNR score and compared with PCGI-E signature score (Figure 9A). PCGI TRN score was calculated using unweighted Z-score method¹⁴ with MRs and their target gene expression profiles. The PCGI TRN score was defined as a PCGI TRN classifier to discriminate PCGI TRN activation high and low tumors.

1.5. Major Task 5: Assessment of the association of activity of TRNs and clinical outcomes

1.5.1. Subtask 1: Assess whether the activity of TRN computed by the classifier of PCGI correlates with clinical outcomes.

Clinical outcome association with PCGI TRN classifier score: For this analysis, Decipher GRID cohort¹⁵ was employed and was grouped by using PCGI TRN score and PCGI-E score at median values (Figure 9A). Of note, PCGI TRN activity (TFA) high & PCGI-E high tumors exhibit poor clinical outcome with fast metastatic progression (Figure 9B). Cox proportional hazard regression analysis showed significant survival difference between TFA-high & PCGI-E high and others (Figure 9C; HR=1.43, CI=1.26-1.63). This demonstrates that our PCGI TNR classifier identifies aggressive PCs.

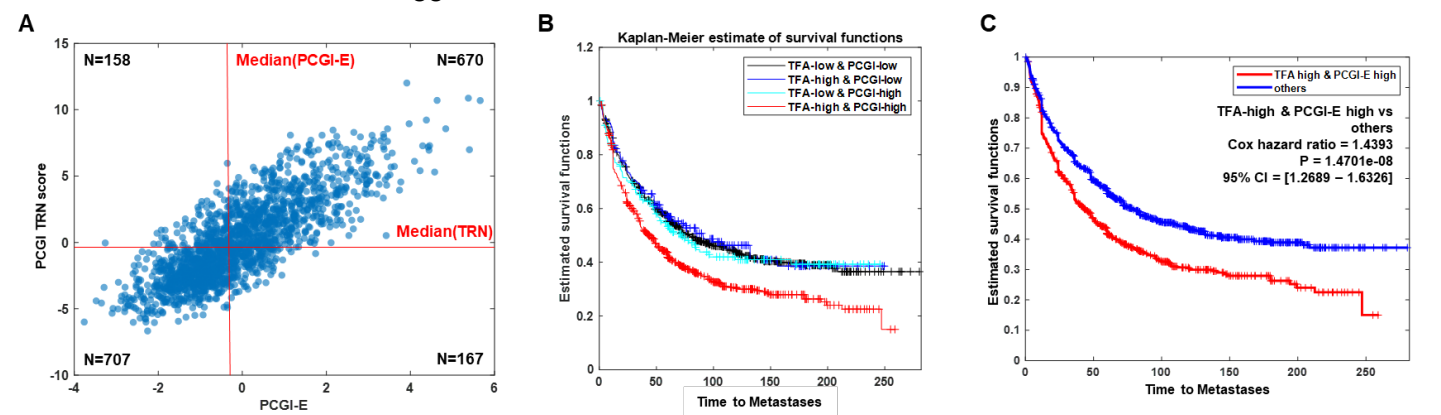


Figure 9. PCGI TRN score distribution and its clinical association. A. Scatter plot shows distribution of PCGI TRN score by PCGI-E score. B. Kaplan-Meier curves depicts distinct survival outcomes between TRN activity (TFA) high & PCGI-high tumors and other tumors. C. Significant survival difference of TFA-high & PCGI-E high versus others was shown.

2. Specific Aim 2: Determine common and distinct transcriptional programs between the two mutually exclusive GI driver networks.

2.1. Major Task 1: Assessment of mutual exclusivity of GI driver networks driven by TP53 and SPOP.

2.1.1. Subtask 1: Reanalyze using our GI workflow described in Aim 1, producing harmonized call sets of genomic alterations

We assessed of mutual exclusivity of GI driver networks driven by TP53 and SPOP.

We identified novel GI driver networks driven by TP53 and SPOP in preliminary research. SPOP and TP53 mutant tumors show mutual exclusivity. In addition, TP53 mutation co-occurs with TNFRSF21, PIK3CA, and BRCA2. SPOP mutation also co-occurs with SYNE1, CDK12, TNFRSF21, BRCA2, PIK3CA. Thus, we tested that whether GI driver networks driven by TP53 and SPOP have also mutual exclusivity (Table 1) and identified interactions between GI drivers and downstream transcriptional programs.

Table 1 Mutual exclusivity of GI driver networks driven TP53 and SPOP.

A	B	Log2 Odds Ratio	q-value	Tendency
TP53	SPOP	-0.569	0.01	Mutual exclusivity

Driver A	Driver B	Log2 Odds Ratio	q-value	Tendency
PIK3CA	TNFRSF21	>3	<0.001	Co-occurrence

TNFRSF21	SYNE1	>3	<0.001	Co-occurrence
PIK3CA	CDK12	2.324	<0.001	Co-occurrence
PIK3CA	SYNE1	2.025	<0.001	Co-occurrence
BRCA2	TNFRSF21	>3	<0.001	Co-occurrence
SPOP	CDK12	1.684	<0.001	Co-occurrence
CDK12	TNFRSF21	2.857	<0.001	Co-occurrence
CDK12	SYNE1	1.859	<0.001	Co-occurrence
SPOP	TNFRSF21	2.174	<0.001	Co-occurrence
BRCA2	PIK3CA	1.586	<0.001	Co-occurrence
BRCA2	SYNE1	1.449	<0.001	Co-occurrence
SPOP	BRCA2	1.151	<0.001	Co-occurrence
TP53	PIK3CA	0.749	<0.001	Co-occurrence
SPOP	SYNE1	0.879	0.004	Co-occurrence
TP53	TNFRSF21	1.031	0.022	Co-occurrence
TP53	BRCA2	0.556	0.024	Co-occurrence
SPOP	PIK3CA	0.705	0.029	Co-occurrence

2.1.2. Assess aberrations of SYNE1, CDK12, BRCA2, PIK3CA, TNFRSF21, TP53, and SPOP in all the tumors from Taylor et al., and MET360 cohort

Reanalysis of harmonized call sets of genomic alterations using Taylor et al. and MET440 cohorts:

We combined Taylor et al.³, and MET440² cohorts (here after harmonized call set) and reanalyzed using our GI workflow developed in the first year. Due to limited availability of the raw data from public databases, we applied a modified version of PCGI (pseudo-PCGI, which we called as pPCGI) computation method which includes copy number alteration (CNA), tumor mutational burden (TMB), and genomic rearrangements (fusion). pPCGI has a strong correlation with PCGI scoring method (Spearman's $\rho = 0.85$, $P < 0.001$) (data not shown). As shown in **Figure 10**, metastatic cancers have higher pPCGI scores than primary cancers.

This suggests that the higher the PCGI, the more metastases, which is consistent with the first year reported results. In the individual analysis of TCGA, Taylor et al, and MET440 datasets, metastatic cancers had higher PCGI scores than primary cancers.

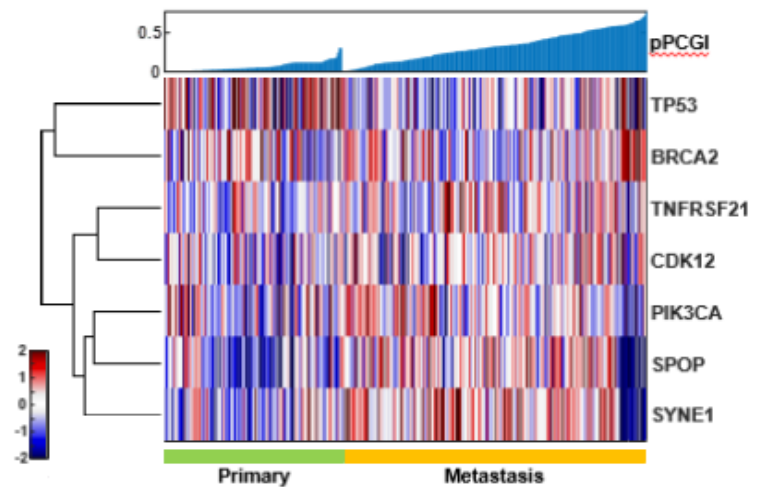


Figure 10. Heatmap of GI driver networks in harmonized call sets.

Top color bars indicate the pPCGI score of harmonized call sets. The differentially expressed genes indicate 7 GI driver gene candidates' expression between primary and metastatic PC samples. Each column in the heatmap is an individual sample (n=358).

Assessment of aberrations of GI drivers: Previously, we developed a driver gene network of GI using PCGI score in PC (n=3,597). We identified 7 driver gene candidates including SPOP, TP53, SYNE1, CDK12, TNFRSF21, BRCA2, and PIK3CA. Among the 7 driver genes, SPOP and TP53 mutant tumors showed mutual exclusivity in PC. SPOP and TP53 mutant have common transcriptional networks including TNFRSF21, BRCA2, PIK3CA. In addition, SPOP mutant has SPOP-specific transcriptional networks including SYNE1 and CDK12. First of all, we analyzed the expression of 7 driver genes in primary and metastatic tumors. Expression of SPOP, SYNE1, TNFRSF21, BRCA2, and CDK12 were statistically significant differences in primary and metastatic tumors (**Figure 11**).

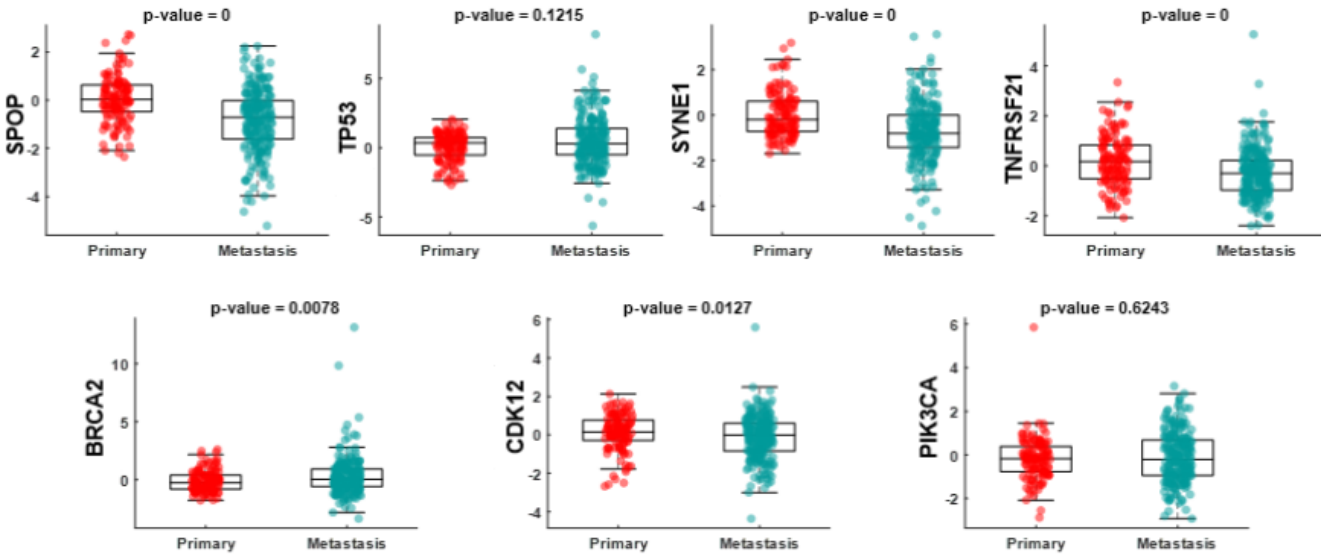


Figure 11. Expression of GI driver networks in primary and metastatic PC. p-values were determined by two-sided rank-sum test.

Expression of GI driver networks in pPCGI high low tumors: We then compared how GI driver networks change with pPCGI high and low in primary and metastatic cancers. Expression of SPOP, TP53, TNFRSF21, and CDK12 is significantly difference between pPCGI low primary tumors and pPCGI high metastatic tumors (**Figure 3**). Especially, SPOP and TNFRSF21 indicated low expression between pPCGI high and low primary tumors and pPCGI high metastatic tumors (**Figure 12**). BRCA2 showed the lowest expression in pPCGI low primary tumors (**Figure 12**). These results suggest that the expression of the GI driver genes is different according to pPCGI high and low tumors.

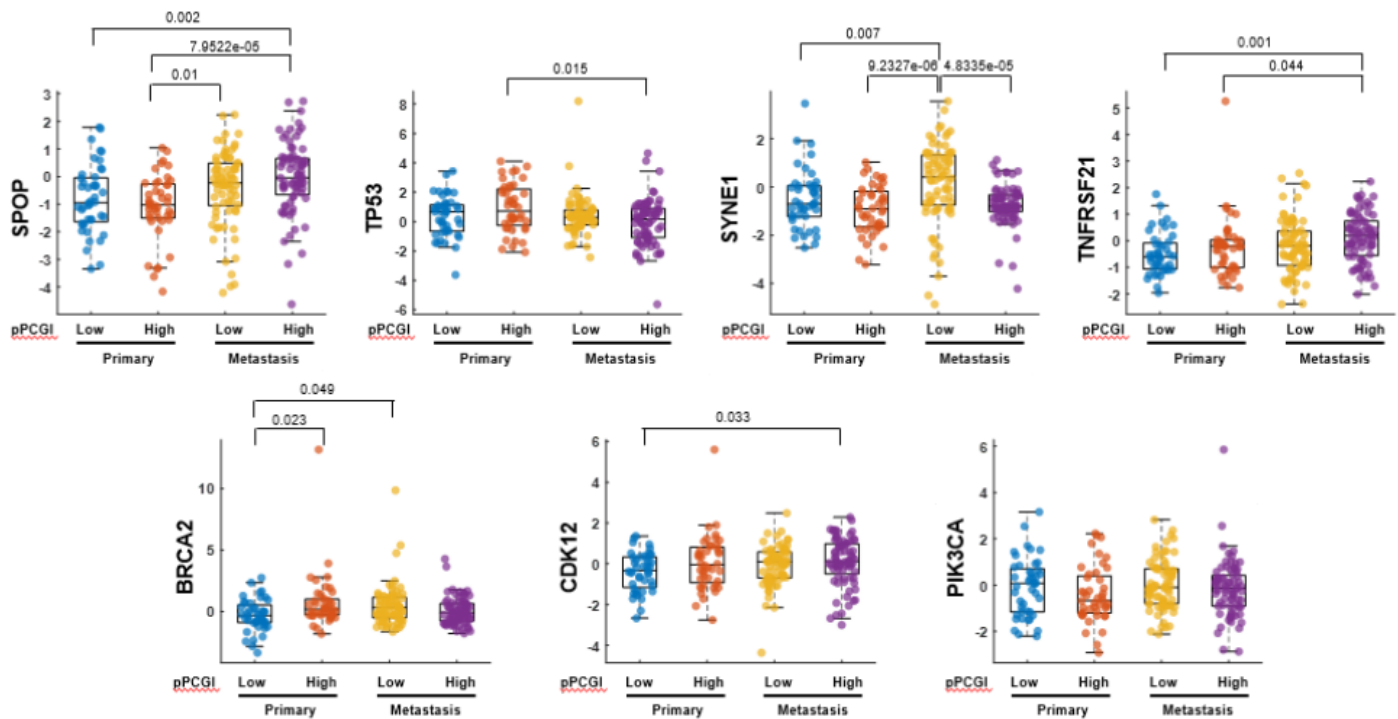


Figure 12. Expression of GI driver network in PCGI high and low tumors. p-values were determined by two-sided rank-sum test.

Expression changes of GI driver networks by SPOP and TP53 aberrations: Then we determined whether these GI driver networks are driven by two mutually exclusive GI drivers, SPOP and TP53 mutant. We also hypothesize that metastatic tumors have SPOP or TP53 mutant more than primary tumors and we tested whether any difference we observe is significant using conducted Fisher's exact tests (**Table 2**). Metastatic tumors had high proportion of SPOP (p-value:0.0028) and TP53 mutant (p-value: 2.8113e-19). In addition, the expression of TNFRSF21 showed a common significant difference according to the SPOP and TP53 mutational status in the harmonized call set (**Figure 13**). CDK12 also presented expression differences in the TP53 mutant state (**Figure 13**). These results suggest that the expression of seven driver genes of the GI can be affected by two mutually exclusive GI drivers, SPOP and TP53 mutant, especially in metastatic cancers.

Table 2. Fisher-exact test of SPOP and TP53 mutation in primary and metastatic prostate cancer.

Total n = 358			Total n = 358		
SPOP			TP53		
	Mutated	Non-mutated		Mutated	Non-mutated
Primary	0	131	Primary	1	130
Metastasis	13	214	Metastasis	86	141

Overexpression of TNFRSF21, Apoptosis regulator, induced apoptosis in a variety of cell types through promoting caspase-3 activation.⁷⁻⁹ mRNA expression was elevated in prostate tumor cells and in advanced stages of PC.¹⁰ Several studies report that death receptor signaling cascades can crosstalk with androgen signaling through inducible knockout of specific death receptors and downstream components or interaction with ARA267-alpha. TNFRSF21 can be a good therapeutic target in prostate involution after androgen ablation. Loss of tumor suppressors such as TP53, RB1, and PTEN losses, have been associated with characteristic gene expression profiles.¹¹⁻¹³ As shown in **Figure 12** and **Figure 13**, the expression of SYNE1 and TNFRSF21 is significantly changed according to the SPOP and TP53 mutant and pPCGI high and low. It is indicated that a large number of transcriptional changes can be associated with SYNE1 or TNFRSF21 genetic aberrations.

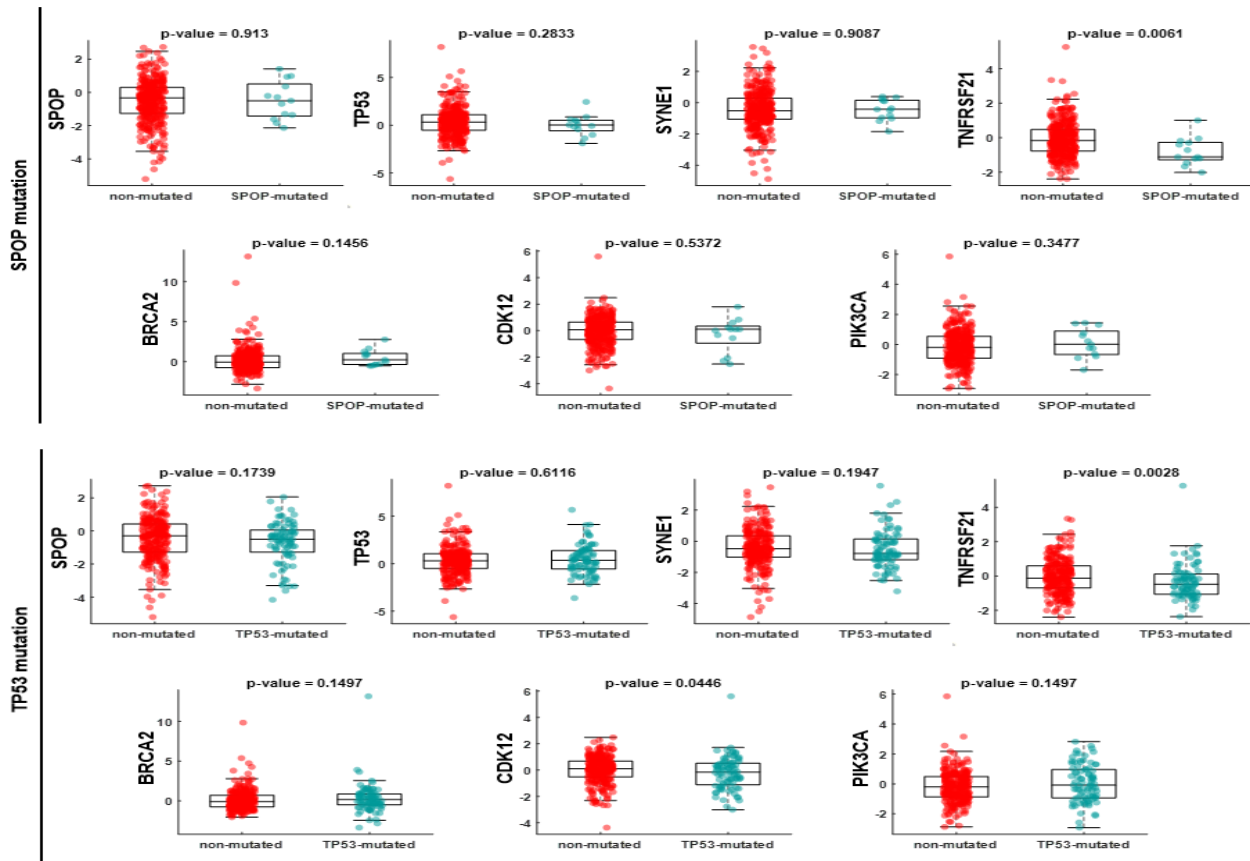


Figure 13. Expression of GI drivers according to SPOP and TP53 mutational status in harmonized call set. p-values were determined by two-sided rank-sum test.

2.2. Major Task 2: Identification of transcriptional signatures of the novel GI drivers, SYNE1, and TNFRSF21.

2.2.1. Subtask 1: Compare the expression profiles of mCRPC cases with aberrations in SYNE1 or TNFRSF21 to a reference set of cases that were wild type for all the GI drivers in the network.

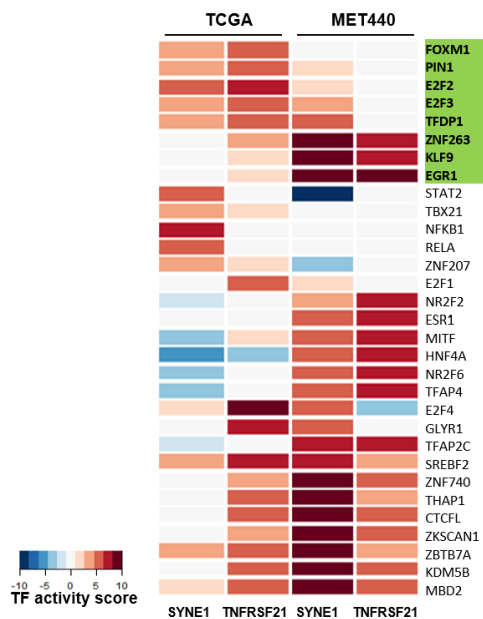


Figure 14. TFA heatmap of patients with TP53, TNFRSF21 mutations in TCGA and MET440.

As cancer progresses from PCa to mCRPC, the activation profiles of TFs change. More TFs show higher activity, and SYNE1 displays high activation in mCRPC (**Figure 14**). SYNE1 loss occurs early in prostate cancer and promotes invasion and metastasis by disrupting chromatin organization and gene regulation¹⁴. Targeting molecular pathways altered by SYNE1 deficiency could potentially block prostate cancer progression¹⁴. Also, TNFRSF21 is overexpressed in prostate cancer progression and associated with recurrence after prostatectomy¹⁵. Blocking TNFRSF21 signaling reduces proliferation and tumor growth in preclinical models¹⁴. By modulating TFs induced by the GI driver network, regulating SYNE1 and TNFRSF21 may be promising targets for novel prostate cancer therapies.

2.2.2. Subtask 2: Test whether depletion of SYNE1 or enforced expression of TNFRSF21 at the RNA and protein levels using LNCaP.

We generated SYNE1 and TNFRSF21 stable cell line systems to identify the association with transcriptional phenotype and genetic aberrations.

Characteristic of genetic aberrations of PC cells: To evaluate whether a large number of transcriptional changes can be associated with SYNE1 or TNFRSF21 aberrations and to determine whether transcriptional changes can be different according to SPOP and TP53 mutant, we generated stable cell lines with depletion of SYNE1 or enforced expression of TNFRSF21 using LNCap and DU145 cells (**Table 3**). Furthermore, we constructed a double knockdown model to assess transcriptional changes in SPOP and SYNE1 deletions. PC3 cells were used to compare the transcription factor differences according to TP53 status (**Table 3**).

Table 3. Known genetic aberrations in cell lines used in this study.

Cell line	TP53	PTEN	Androgen response
LNCap	WT	Null	AD
DU145	MT	Heterozygous	AI
PC3	Del	Null	AI

WT; Wild type, MT; Mutant, Del; Deletion, AI; Androgen-independent growth; AD; Androgen-dependent growth.

Generation of stable cell lines for SYNE1 and TNFRSF21: We hypothesize that aberrations of the novel GI drivers in our GI driver networks such as SYNE1 and TNFRSF21 may similarly constrain a specific transcriptional phenotype. To this, we generated stable cell lines of depletion of SYNE1 (**Figure 15**) or enforced expression of TNFRSF21 (**Figure 16**) according to SPOP and TP53 mutational status. SPOP mRNA expression was reduced by 90% and 20% by shSPOP treatment and was inhibited by 50-10% by DKO in LNCaP and DU145, respectively (**Figure 15A**). SYNE mRNA expression also was reduced by 50% and 10% by RNAi and was inhibited by 90-70% by DKO in LNCaP and DU145, respectively (**Figure 15B**). The SYNE1 protein expression was inhibited by shSYNE1 (**Figure 15C**).

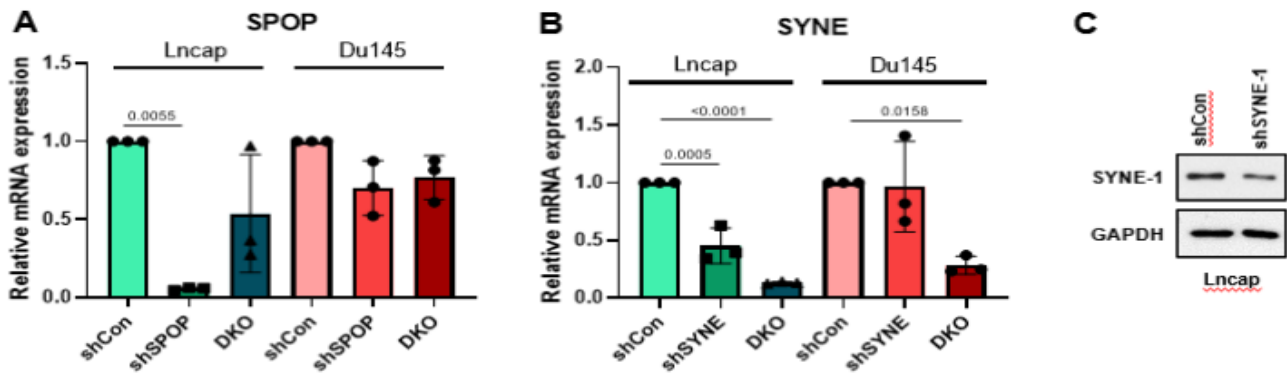


Figure 15. Generation of Stable shRNA-expression cell lines with or with or knockdown of SPOP and SYNE gene in LNCaP and DU145. The knockdown of SPOP, a validated shRNA clone (#TRCN0000139181, target sequence: CACAAGGCTATCTTAGCAGCT) in the vector pLKO1-puro was purchased from Sigma. The knockdown of SYNE-1, a validated shRNA clone (#TRCN0000147281, target sequence: GCGTAGTGATAAGACTGATTT) in the vector pLKO1-Neo was purchased from Sigma. A nonmammalian shRNA control plasmid (Sigma #SHC002) was used as a control. E. coli colonies were selected by 10ug/mL ampicillin. LNCaP and DU145 cells were stably transfected with expression plasmids containing shRNA targeting SPOP, SYNE, and its control scramble sequence. (A and B) qPCR determined the relative SPOP and SYNE mRNA expression levels. (C) Western blotting determined the relative SYNE1 protein

expression levels in LNCaP cells. Gene expression was normalized to GAPDH using the comparative CT method. p-values were determined by one-way ANOVA test. DKO; double knockout, shSPOP/SYNE.

TNFRSF21 mRNA expression was increased over 200-fold and 10-fold in GFP-TNFRSF21-transfected LNCap and DU145 cells (Figure 16).

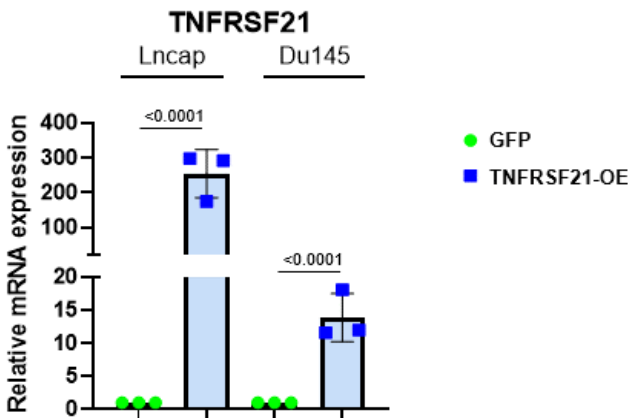


Figure 16. Establishment of stable overexpression of TNFRSF21 in LNCap and DU145 cell lines. The TNFRSF21 overexpression plasmid was obtained by cloning the full-length TNFRSF21 cDNA (NM_014452) into the pLenti-C-mGFP vector (Origene #RC205864L2). E. coli colonies were selected by 34 ug/mL of Chloramphenicol. Gene expression was normalized to GAPDH using the comparative CT method. p-values were determined by unpaired two-tailed Student’s t-test. OE; overexpression.

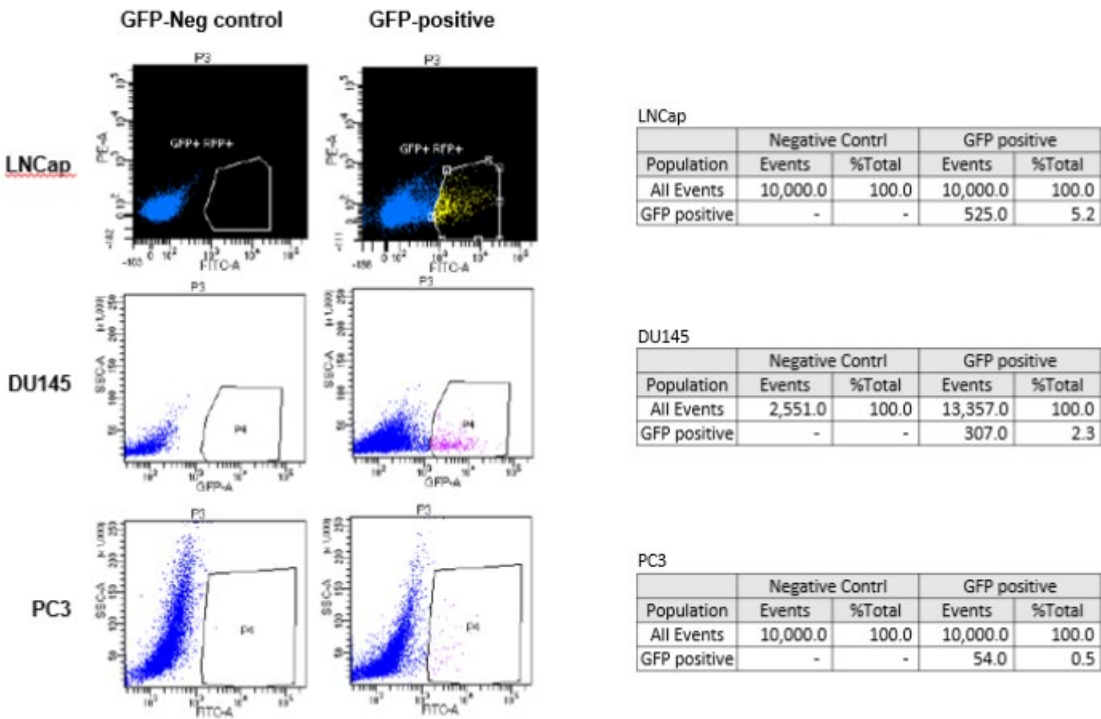


Figure 17. Fluorescence-activated cell sorting of GFP-positive cells. GFP-positive cells were sorted out with GFP signal using flow cytometry (BD Aria III Cell Sorter) three days after the transfection. Only the top 10% of cells were sorted out and placed back into culture. After 4 months of growth, GFP-positive cells were detected to confirm stable cell line generation.

During this project, we generated shSYNE and GFP-TNFRSF21 overexpressed stable cells in PC3. Unlike other cell lines, PC3 cell numbers were good but observed in cells undergoing senescence after transfection. The original transfection rate is 5% after 72 h transfection and we sort GFP cells. After 4 months of growth, GFP-positive cells dropped from 100% to 0.5% (Figure 17). Moreover, PC3 shSPOP cells were isolated as a single colony, but the double knockout of shSPOP/SYNE PC3 cells couldn’t survive. We hypothesized that genetic aberration of PC3, such as TP53 deletion and PTEN-null, can be associated with cell survival and senescence

when a specific gene is defective. Further studies are needed to determine whether the relationship between TP53 and specific defective genes has a significant impact on PC's survival and fate.

Transcriptional changes by the novel GI drivers, SYNE1 and TNFRSF21: We select 5 genes that are highly correlated with PCGI score (**Table 4**). These genes have been reported to be closely related to genomic instability, PC survival, and poor prognosis. MGAT4B show aberrant upregulation in CRPC compared to hormone-sensitive PC and is associated with PC recurrence.¹⁶ PBK, CEP55, and UBE2C are one of the genes affecting genomic instability and are closely related to cell proliferation and survival.¹⁷ EZH2 is associated with tumor progression and poor prognosis in prostate tumorigenesis.¹⁸ In addition, Increased EZH2 expression results in uncontrolled cell division, aberrant mitoses with extra chromosomes, and genomic instability.¹⁹

Table 4. Correlation with 5 genes panel and PCGI score.

EntrezID	Symbol	PCGI_score	
		p-value	Rho
11282	MGAT4B	1.46221E-34	0.5466359
55872	PBK	3.40972E-16	0.3813367
11065	UBE2C	8.9097E-32	0.5266902
55165	CEP55	1.69531E-22	0.4487434
2146	EZH2	1.41952E-31	0.5251885

Evaluation of 5 PCGI-E genes by deletion of SYNE and SPOP: We tested relative mRNA expression of 5 genes by SPOP and SYNE aberrations (**Figure1 8**). PBK, UBE2C, and MGAT4B were elevated by shSYNE1 in LNCap cells. Also, the expression of PBK and MGAT2B was increased in LNCap DKO. Unlike LNCap, deletion of SPOP up-regulated 5 genes in DU145 cells. It seems that the expression difference was due to the difference in TP53 mutant and PTEN-null status between LNCap and DU145 cells. The primers for each gene are listed in **Table 5**.

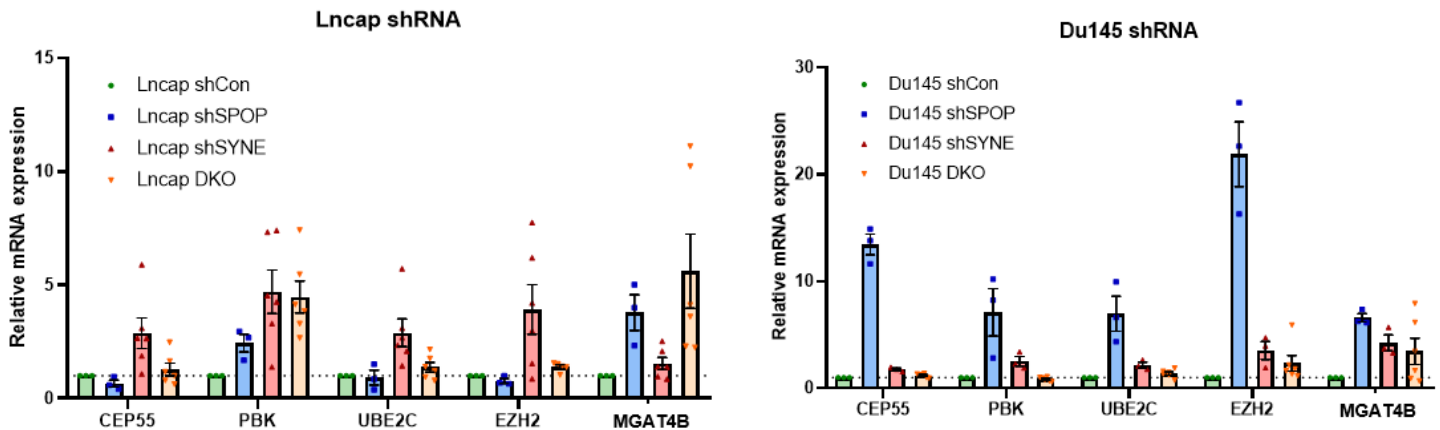


Figure 12. RT-qPCR analysis of gene expression in shSPOP, shSYNE1, and DKO cells. Gene expression was normalized to GAPDH using the comparative CT method. p-values were determined by one-way ANOVA test.

Table 5. Primers were used in this study.

Primers	Sequences	Primers	Sequences
Syne-1-F	AGAGCCAAGTCCTCAACCACCT	PBK-F	GGAGTCTCTCTACCACTGGATG
Syne-1-R	CACCGAAGCATTGTGACAGGTCAC	PBK-R	CACCATCTCTCTCCACAGCTTC
TNFRSF21-F	CCAGTGCCATTGTGGAAAAGGC	UBE2C-F	CTGGCGATAAAGGGATTCTGCC
TNFRSF21-R	CTTCCCACTTGGGCTGCTACAA	UBE2C-R	GCGAGAGCTTATACCTCAGGTC
SPOP-F	GGAAGGCTCCAAACCTCGACAA	MGAT4B-F	TGATGTACGCGCAGTCCAAAGG
SPOP-R	AGCGTTCTCCACGGACAGGTTA	MGAT4B-R	TCATCCAGTCCTCTGAAGGCTG
CEP55-F	TCGACCGTCAACATGTGCAGCA	EZH2-F	GACCTCTGTCTTACTTGTGGAGC

CEP55-R	GGCTCTGTGATGGCAAACATCATG	EZH2-R	CGTCAGATGGTGCCAGCAATAG
---------	--------------------------	--------	------------------------

Evaluation of 5 PCGI-E genes by stable overexpressed TNFRSF21 cells: Overexpressed TNFRSF21 up-regulated 5 genes panel in stable TNFRSF21 overexpressed LNCaP and DU145 cells (**Figure 19**). These results suggest that SYNE1 and TNFRSF21 genetic aberrations can constrain a specific transcriptional phenotype. It also indicates that these transcription factor program changes can be affected by the status of genes such as SPOP, TP53, and PTEN.

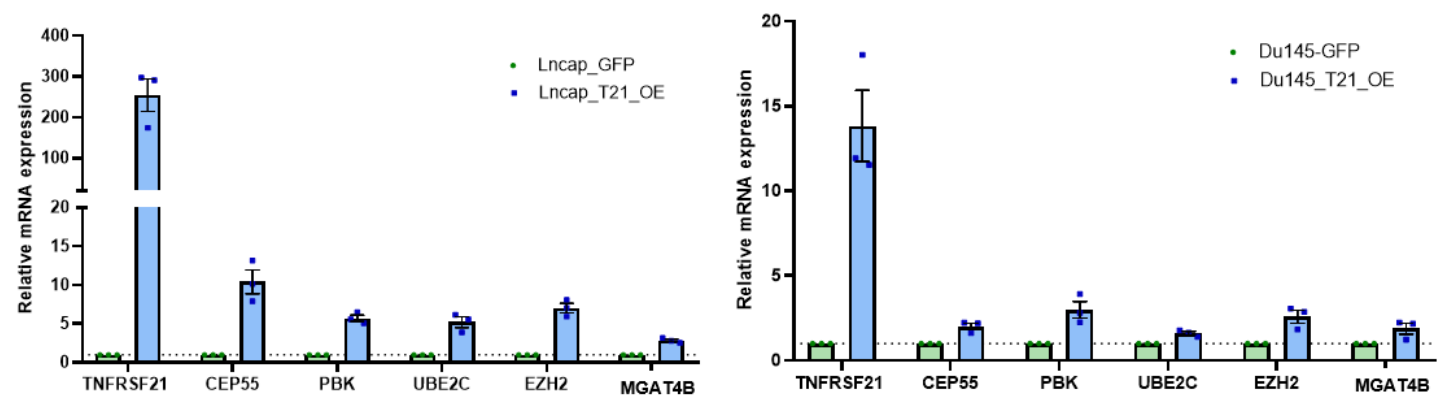


Figure 13. Relative expression of TNFRSF21 and 5 genes in stable TNFRSF21 overexpressed cells. Gene expression was normalized to GAPDH using the comparative CT method. p-values were determined by unpaired two-tailed Student's t-test.

2.3. Major Task 3: Reconstruction of TRNs activated by either TP53-network or SPOP-network.

2.3.1. Subtask 1: Perform MRA to identify key TFs that regulate majority of the genes in the combined signature of 7 GI drivers.

Generation of RNA-seq data using stable cell lines: To determine transcriptional signatures of the novel GI drivers, SYNE1 and TNFRSF21 driven by SPOP and TP53 mutant, we generated high-depth RNA-seq data (60M PE) (**Table 6**).

Table 6. Throughput and quality of RNA-seq data.

	Sample Name	Clean Reads	Clean Base	Read Length	Q20(%)	GC(%)
1	Du145_DKO_1	72286888	2.17E+10	PE150	97.62	51.75
2	Du145_DKO_2	72227083	2.17E+10	PE150	97.51	50.64
3	Du145_DKO_3	72195978	2.17E+10	PE150	97.9	51.06
4	Du145_GFP_1	72235650	2.17E+10	PE150	96.82	52.13
5	Du145_GFP_2	72265906	2.17E+10	PE150	96.79	52
6	Du145_GFP_3	72150447	2.16E+10	PE150	96.84	51.68
7	Du145_T21_2	72153044	2.16E+10	PE150	97.7	51.99
8	Du145_T21_3	72255619	2.17E+10	PE150	97.62	51.53
9	Du145_T21_4	72115021	2.16E+10	PE150	97.18	51.2
10	Du145_shCon_1	65256436	1.96E+10	PE150	97.41	51.65
11	Du145_shCon_3	72143406	2.16E+10	PE150	97.44	51.62
12	Du145_shCon_4	62897784	1.89E+10	PE150	97.29	49.16
13	Du145_shSPOP_1	72257166	2.17E+10	PE150	97.65	51.54
14	Du145_shSPOP_2	72213713	2.17E+10	PE150	97.6	51.39
15	Du145_shSPOP_3	72187330	2.17E+10	PE150	97.47	51.79
16	Du145_shSYNE_1	72207460	2.17E+10	PE150	97.66	51.58
17	Du145_shSYNE_2	72212329	2.17E+10	PE150	97.76	51.39
18	Du145_shSYNE_3	72356194	2.17E+10	PE150	97.47	51.69
19	Lncap_DKO_2	68942420	2.07E+10	PE150	96.83	51.71

20	Lncap_DKO_3	72101668	2.16E+10	PE150	97.03	51.54
21	Lncap_DKO_5	72193831	2.17E+10	PE150	97.4	51.99
22	Lncap_GFP_1	72176344	2.17E+10	PE150	97.61	51.78
23	Lncap_GFP_2	72118471	2.16E+10	PE150	97.49	52.25
24	Lncap_GFP_3	72309172	2.17E+10	PE150	97.45	52.27
25	Lncap_T21_1	72177359	2.17E+10	PE150	97.47	52.6
26	Lncap_T21_2	72328860	2.17E+10	PE150	97.42	52.41
27	Lncap_T21_3	72167028	2.17E+10	PE150	97.39	51.77
28	Lncap_shSPOP_1	72299908	2.17E+10	PE150	97.17	52.26
29	Lncap_shSPOP_2	72133140	2.16E+10	PE150	97.15	51.59
30	Lncap_shSPOP_4	69864458	2.1E+10	PE150	97.5	51.55
31	Lncap_shSYNE_1	72171713	2.17E+10	PE150	96.88	52.01
32	Lncap_shSYNE_2	72324467	2.17E+10	PE150	97.14	52.33
33	Lncap_shSYNE_3	72273542	2.17E+10	PE150	96.75	52.34
34	Lncap_shcon_1	72329040	2.17E+10	PE150	97.5	51.7
35	Lncap_shcon_2	72222574	2.17E+10	PE150	97.18	51.94
36	Lncap_shcon_3	72121842	2.16E+10	PE150	96.96	51.99
37	PC3_WT_1	72205228	2.17E+10	PE150	97.85	51.06
38	PC3_WT_2	72184046	2.17E+10	PE150	97.84	51.27
39	PC3_WT_4	72122673	2.16E+10	PE150	97.44	50.83

Build automated pipeline for RNA-seq short variant discovery: We developed the automated pipeline for RNA-seq short variant discovery (**Table 7**). The protocol is based on DROP, the detection of RNA outliers pipeline, that can integrate aberrant expression, aberrant splicing, and mono-allelic expression from raw sequencing files.

Table 7. Workflow of the automated pipeline for identification of variants and mutations from RNA-seq data.

Steps	Main procedure
Step1	Remove residual adapter sequences and crop low-quality bases with trim_galore
Step2	Align the reads with STAR
Step3	Mark duplicated reads with GATK
Step4	Modify the marked-duplicates bam using GATK's SplitNCigarReads (due to splice junctions)
Step5	Add read groups with GATK
Step6	Perform base recalibration with GATK
Step7	Apply base recalibration with GATK
Step8	Call variants using GATK
Step9	Mark in the all variants with GATK
Step10	Exclude the variants that did not pass the filter with bcftools
Step11	Add gene names to the variants with bcftools

Perform MRA to identify key TFs that regulate majority of the genes in the combined signature of 7 GI drivers:

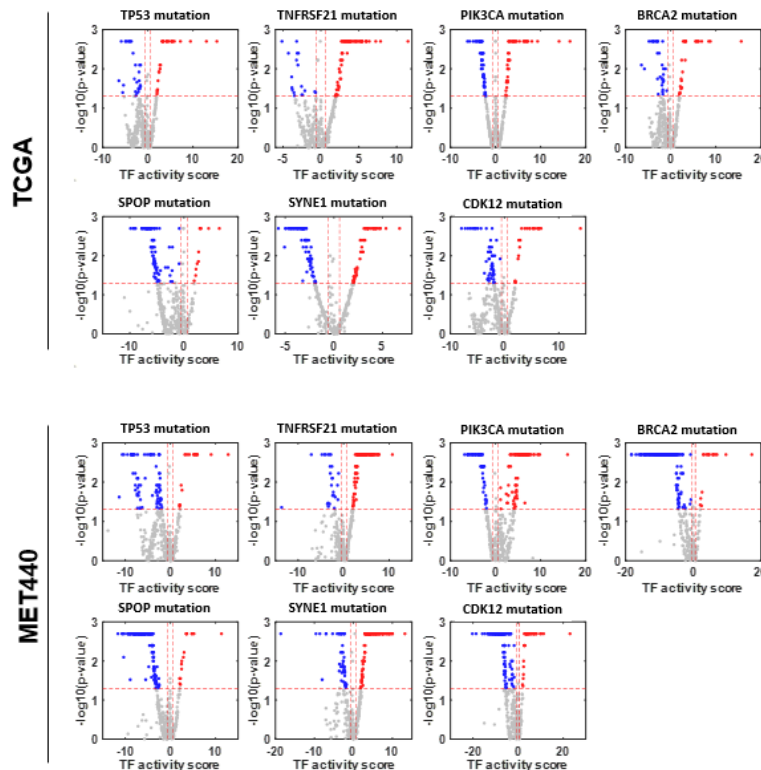


Figure 20. MRA identified significant TF activity based on mutation status in PCa (TCGA) and mCRPC (MET440).

We used decoupleR package to extract biological regulatory activities from omics data using the weighted-mean method. TFA was calculated in TCGA and the MET440 cohort based on the mutation status of GI driver network genes (**Figure 20**).

To identify transcriptional signatures in the GI driver gene network and reconstruct integrated networks describing interactions of GI drivers and downstream transcriptional programs, we selected the top 10 TFs (transcription factors) according to each mutational status from TCGA (localized PCa) and MET440 (mCRPC) cohorts (**Figure 21**). Among the top 10 TFs for each GI driver gene mutation in TCGA and MET440, a total of 14 TFs were commonly identified. In addition, 12 TFs were specifically identified in TCGA only, and 28 TFs in MET440 only. As shown in **Figure 21**, we can observe changes in the activation of downstream transcriptional programs by the GI driver network in localized PCa and mCRPC. As cancer progresses to mCRPC, the activity of transcriptional programs changed with additional TF activations, and specifically, the activity of TFs associated with PIK3CA, SYNE1, and TNFRSF21 mutations was newly acquired.

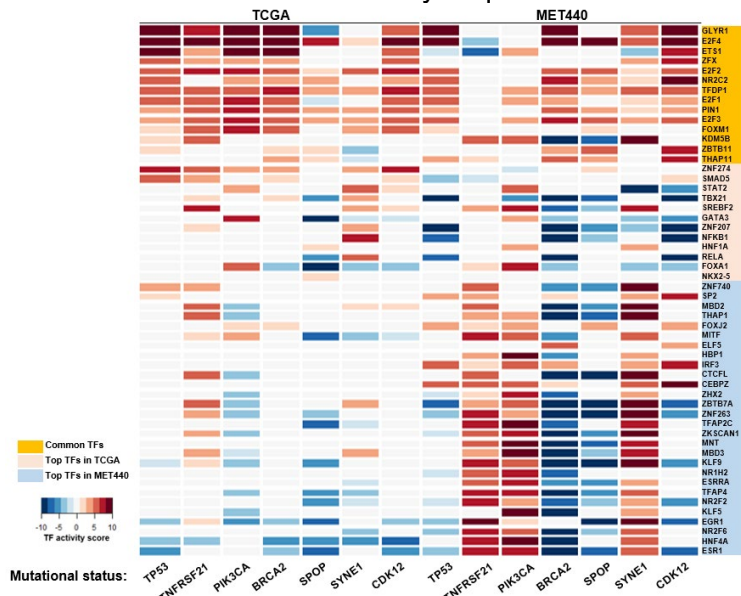


Figure 4. TFA heatmap of the top 10 TFs by mutation status between TCGA and MET440.

2.3.2. Subtask 2: Reconstruct two GI networks governed by TP53 and SPOP

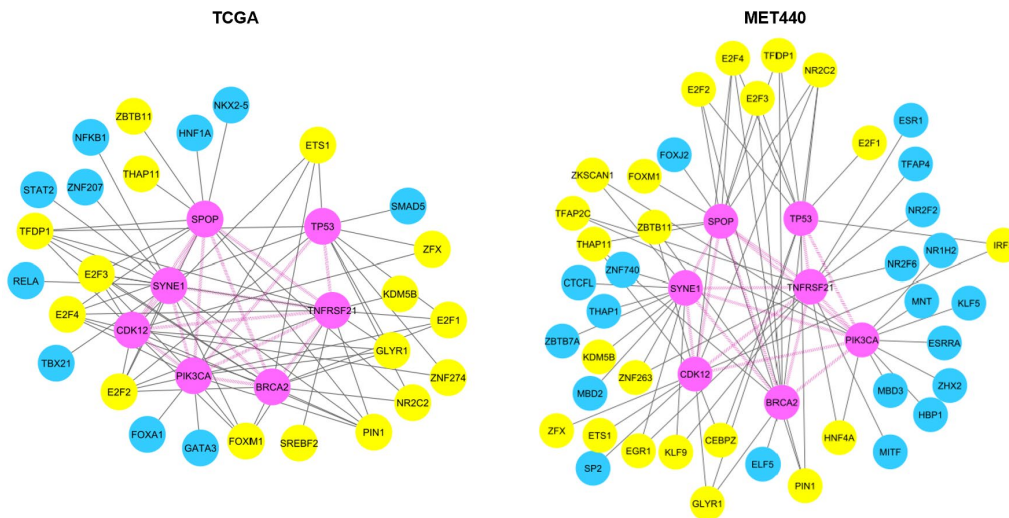


Figure 22. Network models of TCGA and MET440. The network describes interactions between GI driver network genes and TFs. The magenta node color indicates GI driver network genes. Yellow node presents duplicated TF between TCGA and MET440. Magenta arrowed edges presents co-occurrence. Gray arrowed edges represent interaction between GI driver network genes and TFs.

As shown in **Figure 22**, we can observe that in both PCa and mCRPC, downstream transcriptional programs are controlled through shared TFs. Additionally, there is an increase the number of TFs that are activated as the cancer advances from PCa to the more aggressive mCRPC stage.

2.3.3. Subtask 3: Develop and validate two classifiers of TP53-network and SPOP-network.

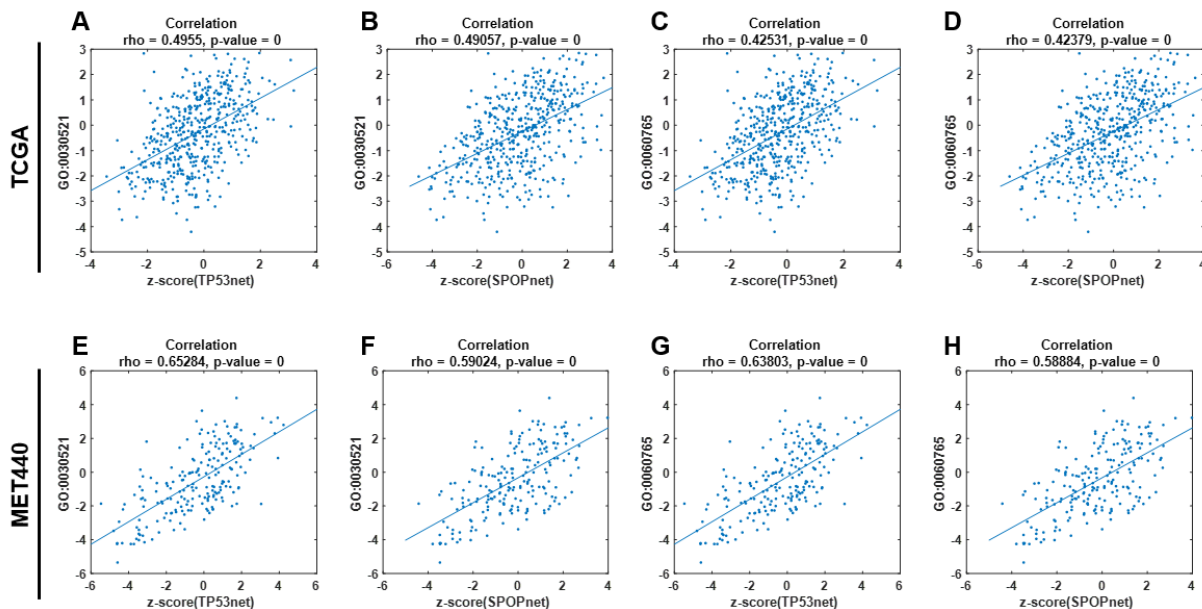


Figure 23. Correlation between AR gene signatures and TP53-net and SPOP-net signature in TCGA PRAD and MET440 cohort. Correlation was measured by using Spearman's method. GO:0030521_GOBP_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY²¹; GO:0060765_GOBP_REGULATION_OF_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY²².

The androgen receptor (AR) drives prostate cancer progression, leading to AR pathway inhibition as standard therapy²³. resistance of aberrant AR activation, requiring more potent AR targeting²⁴. The abnormal activation of AR signaling results in AR dependence in advancing disease. Resistance to therapies targeting the AR pathway

continues to be a significant challenge in the clinic²⁵. In order to evaluate the influence of the TP53-net and SPOP-net on the regulation of these AR aberrations, we computed signature scores for each signature as depicted in **Figure 23** and examined their correlation utilizing z-score approaches⁶. The TP53-net and SPOP-net showed strong correlation with AR signaling pathways, as depicted in **Figure 23**. By modulating the transcription factors regulating TP53-net and SPOP-net, we may be able to assist with overcoming resistance to AR-targeted therapies.

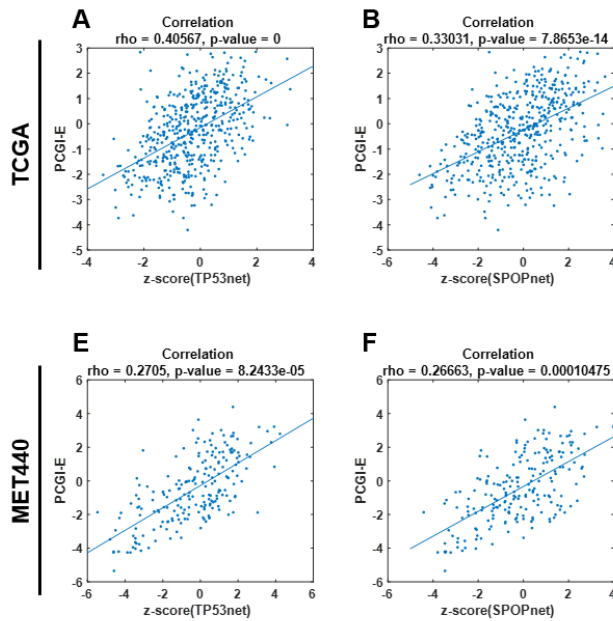


Figure 5. Correlation between PCGI-E s and TP53-net and SPOP-net signature in TCGA PRAD and MET440 cohort. Correlation was measured by using Spearman's method.

During the first year, we developed a scoring approach based on GI and demonstrated an association between the GI signature (PCGI-E) and clinical outcomes. We found that higher PCGI-E and TFA scores strongly correlated with increased mortality rates (**Figure 24**). To determine if TP53-net and SPOP-net are associated with survival and PCGI-E in prostate cancer patients, we performed correlation tests and Kaplan-Meier curve analysis. As shown in **Figure 25**, TP53-net and SPOP-net exhibited strong associations with PCGI-E. However, the Kaplan-Meier curve analysis did not show significant results. While no significant correlations were observed with disease-free survival (DFS) in TCGA or overall survival (OS) in MET440, it appears necessary to stratify with larger cohort samples, calculate TFA, and implement TP53-net and SPOP-net signatures.

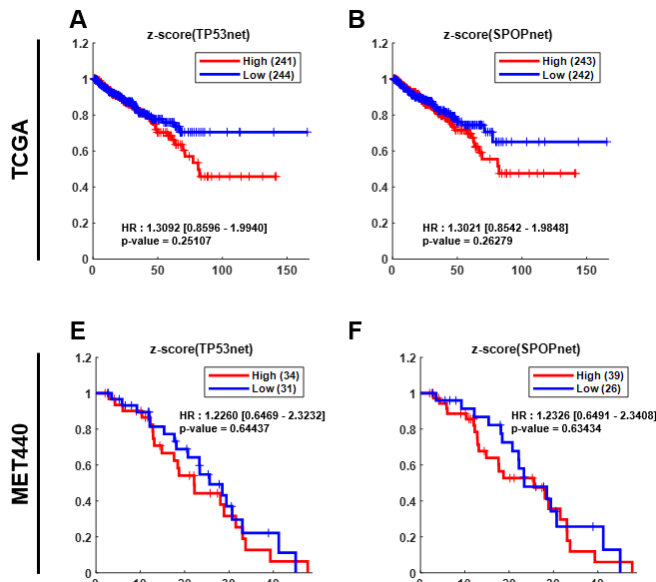


Figure 6. Kaplan-Meier curve analysis. A. DFS in TCGA cohort stratified by z-score of TP53-net and SPOP-net. B. OS in MET440 stratified by z-score of TP53-net and SPOP-net.

Specific Aim 3: Test whether the GI driver networks evolve in pre/post treatment using PC transcriptome.

3.1. Major Task 1: Assessment of the association between the GI driver network activities and clinical outcomes in RP samples and diagnostic needle biopsy sample.

3.1.1. Perform a single sample classification using the GI network classifiers.

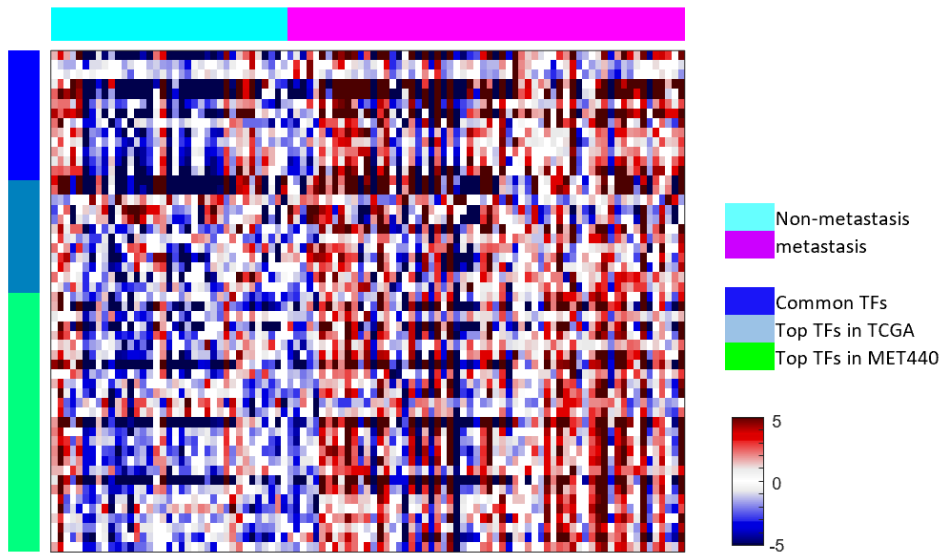


Figure 76. Heatmap of TFA in PNBX cohort.

The TFA of TFs found in PCa and mCRPC were examined in the PNBX data. Diagnostic PNBX samples from a racially and ethnically diverse cohort of 1,927 men who were diagnosed and treated exclusively within a single healthcare system are archived in the VA-GLA Healthcare System. From 1,927 PNBX samples, our collaborator Dr. Isla Garraway selected 99 men with de novo metastatic disease, a metastatic progression that emerged following treatment, and/or high-grade disease without metastatic progression through a review of more than 2,100 cases identified within the VA-GLA cancer registry or procedure logs. Selected cases were categorized into sub-cohorts based on tumor burden at diagnosis or recurrence/progression and were only included if diagnostic PNBX blocks were available with sufficient tissue. These 99 PNBX patient specimens include oligo-metastatic disease (defined by <5 extrapelvic lymph node and/or bone metastases and no visceral metastases), poly-metastatic disease (defined as >5 metastases or any visceral involvement), and high-grade PC cases without metastatic progression (defined as no evidence of metastasis over a median follow-up of 56 month). High-grade PC areas in PNBX were procured for RNA sequencing from formalin-fixed and paraffin-embedded diagnostic PNBX of 99 patients. The strength of these PNBX data lies in the associated well-annotated clinical information and a sufficient number for correlative analyses, but they are classified into subgroup by the degree of metastatic progression. As shown in **Figure 26**, metastatic prostate cancer patients were observed to have higher TFA than non-metastatic patients.

3.1.2. Test the clinical outcomes are correlate with the activity of GI driver network for PCGI high tumors.

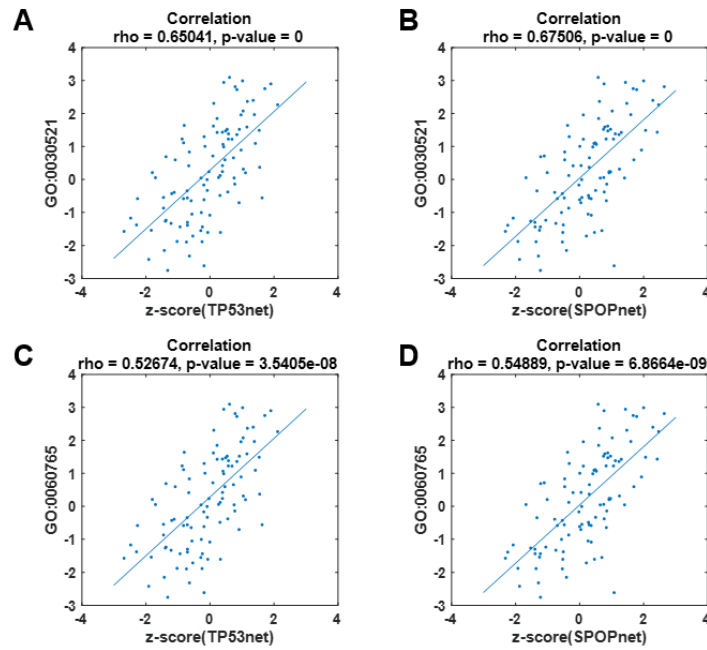


Figure 27. Correlation between AR gene signatures and TP53-net and SPOP-net signature in TCGA PRAD and MET440 cohort. Correlation was measured by using Spearman's method. GO:0030521_GOBP_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY; GO:0060765_GOBP_REGULATION_OF_ANDROGEN_RECEPTOR_SIGNALING_PATHWAY.

The PNBX data also showed significant correlation between the AR signaling signature and TP53-net and SPOP-net signatures (**Figure 27**). And PCGI-E also exhibited strong correlation (**Figure 28**).

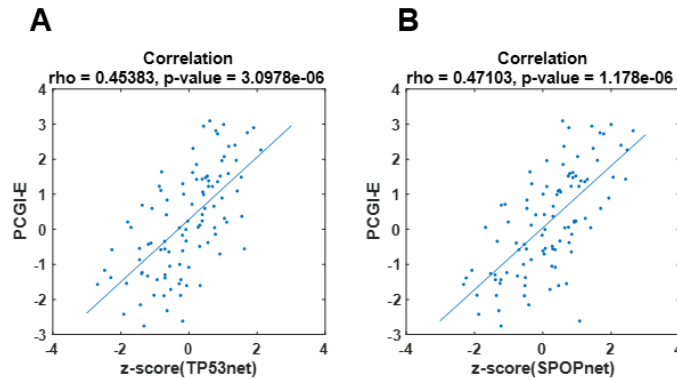


Figure 28. Correlation between PCGI-E s and TP53-net and SPOP-net signature in PNBX cohort. Correlation was measured by using Spearman's method.

3.2. Major Task 2: Assessment of association between GI driver specific TFs and ARSI response in the CSMC CTC cohort.

3.2.1. Subtask 1: Generate NanoString-RNA assay data with the CSMC CTC cohort.

Geneset selection for the PC-CTC detection: The PCGI related genes (54 Top TFs, 5 top PCGI correlated genes, and 27 nuclear instability genes) were adjusted for the CTC detection (**Figure 11**). To minimize background signal from white blood cells, we selected genes have low expression in immune cells using the Differentiation Map²⁶ or high expression in prostate cancer cell line using the Cancer Cell Line Encyclopedia²⁷. Then, we further selected genes have high expression in CTC RNA-seq data²⁸ the VA Greater Los Angeles prostate needle biopsies cohort¹⁷.

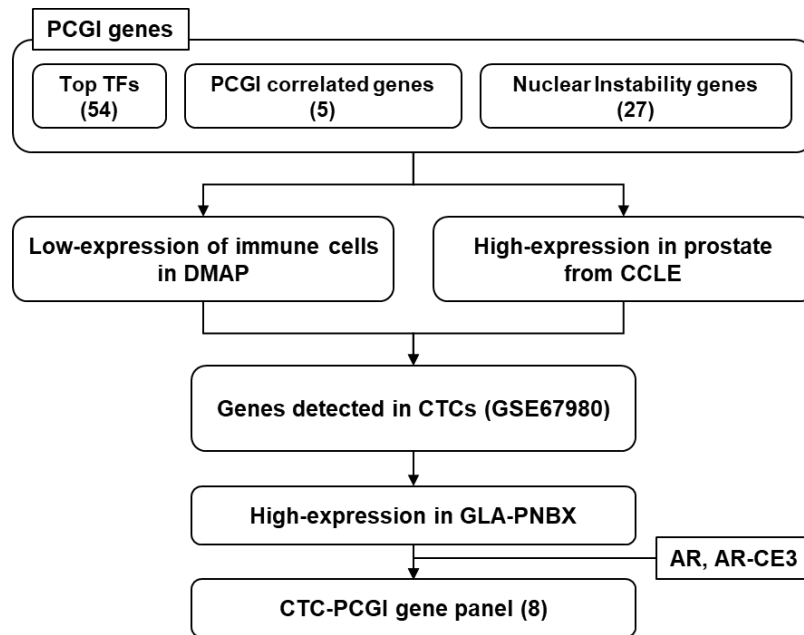


Figure 8. Schematic flow of the CTC-PCGI gene panel selection.

Gene expression pattern of CTC-PCGI gene panel in CSMC PC-CTC cohort: The gene expression of CTC-PCGI gene panel were measured using the NanoString-RNA assay as previously reported²⁹. Briefly, the CTCs from patients' blood were purified by Thermo-responsive (TR)-NanoVelcro CTC purification system^{30, 31}. Then, the NanoString nCounter platform measure the gene expression of purified CTCs³². The expression patterns of 8 CTC-PCGI genes (Top TFs: FOXM1, FOXA1; P CGI correlated genes: EZH2, UBE2C, CEP55; Nuclear Instability gene: DIAPH3; and AR, AR-CE3) for 37 PC patients' samples were presented in **Figure 30A**. We checked the correlation between 8 CTC-PCGI genes to assess the association (**Figure 30B**). Interestingly, the correlation between genes from the same gene set was not very strong, and the correlation with genes from different gene sets was stronger. The correlation patterns are clustered into three groups. EZH2 is strongly associated with DIAPH3, FOXM1 is strongly associated with UBE2C, and AR, CEP55, FOXA1, and AR-CE3 are positively associated with each other.

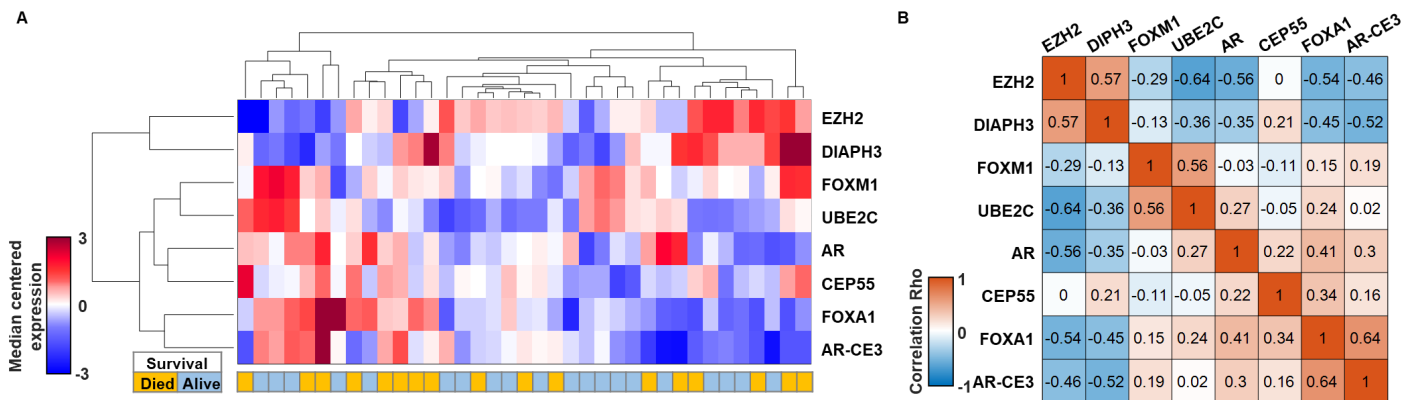


Figure 30. Gene expression patterns of CTC-PCGI gene expression in CSMC-CTC cohort. A. Heatmap of 8 CTC-PCGI genes in CSMC-CTC cohort. B. Correlogram of 8 CTC-PCGI genes.

3.2.2. Subtask 2: P CGI-E scores and/or GI driver specific TF expression will be calculated to predict treatment response/outcomes.

Evaluation of the prognostic association between the CTC-PCGI gene panel and overall survival (OS) in CSMC PC-CTC cohort: We performed to logistic regression analysis using 8 CTC-PCGI gene panel to predict survival and evaluated the performance using area under the receiver operating characteristic (ROC) curve in CSMC PC-CTC cohort (**Table 8 and Figure 31**). The AUC of the logistic regression of 8 CTC-PCGI gene panel was

0.9059 ($p < 0.01$). Among 8 CTC-PCGI genes, DIAPH3, UBE2C, and CEP55 showed high association with OS.

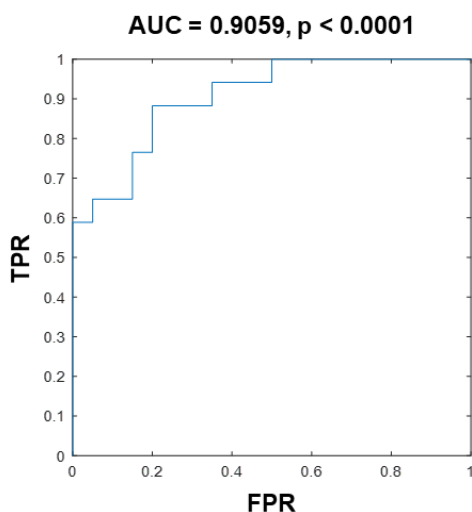


Figure 9 ROC curve of logistic regression of 8 CTC-PCGI gene panel for discriminating OS in CSMC PC-CTC cohort

Table 8. Logistic regression model for predicting survival of CSMC PC-CTC cohort.

Gene	Coefficient
Intercept	0.5993
EZH2	-0.0910
DIAPH3	1.0034
FOXM1	-1.3294
UBE2C	1.4136
AR	0.3348
CEP55	2.7911
FOXA1	-0.7796
AR-CE3	0.6261

Using DIAPH3, UBE2C, and CEP55, we calculated z-score and checked the survival association (**Figure 32**). The Kaplan-Meier survival curve analysis showed that z-score of these three genes accurately discriminated OS (HR = 3.6277, 95% CI = 1.2564-10.4745, $P = 0.0172$).

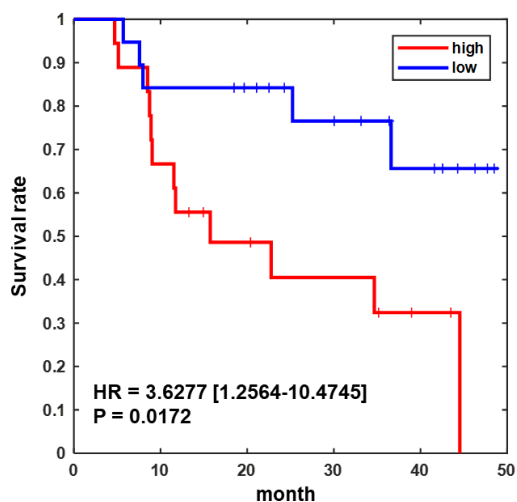


Figure 32. Kaplan-Meier curve analysis of OS in the CSMC PC-CTC cohort stratified by z-score of DIAPH3, UBE2C, and CEP55.

4) Key research accomplishments

- Computed PCGI scores to quantify genomic instability in TCGA, MET360, and Taylor prostate cancer cohorts using various genomic alterations.
- Found higher PCGI scores in metastases vs primary tumors, suggesting PCGI correlates with aggressive disease.
- Identified DEGs between high and low PCGI score tumors in each cohort.
- DEGs were enriched in cell cycle, chromatin regulation, and mitosis, linking PCGI to proliferation.
- Performed master regulator analysis to identify key transcription factors driving PCGI-associated differential expression.
- Reconstructed transcriptional regulatory networks showing interactions between master regulators.
- Developed a PCGI transcriptional regulatory network (TRN) classifier based on master regulators.
- Showed the PCGI TRN classifier can discriminate aggressive prostate cancers with poor clinical outcomes.
- Validated the prognostic potential of the PCGI TRN classifier.
- Assessed mutual exclusivity between TP53 and SPOP mutant tumors as GI drivers.
- Identified co-occurring genomic alterations associated with TP53 and SPOP mutations.
- Generated RNA-seq data from prostate cancer cell lines with knockdown/overexpression of GI drivers.
- Identified transcriptional changes induced by perturbing GI driver genes SYNE1 and TNFRSF21.
- Reconstructed separate transcriptional networks governed by TP53 and SPOP mutations.
- Developed and evaluated classifiers for the TP53 and SPOP mutant networks.
- Evaluated association between GI driver networks and outcomes in biopsy and CTC data.
- Showed prognostic potential of GI network signatures in clinical samples.
- Developed automated pipeline for variant calling from RNA-seq data.
- Performed integrative analyses linking genotypes, phenotypes and clinical outcomes.

5) Conclusion

This study performed an integrative analysis of genomic, transcriptomic and clinical data to elucidate the role of genomic instability (GI) in driving aggressive prostate cancer. By developing a genomic instability score (PCGI), we identified greater GI in metastatic tumors compared to primary tumors across multiple cohorts. Differential expression analysis revealed GI-associated genes were enriched in cell cycle and chromatin regulation. Master regulator analysis identified key transcription factors mediating differential expression, which were used to build prognostic transcriptional regulatory network models of GI. Knockdown and overexpression experiments validated two novel GI drivers, SYNE1 and TNFRSF21, induce specific transcriptional changes. Further, separate models governing TP53 and SPOP mutant networks were constructed, with distinct activities in localized versus metastatic disease. The findings demonstrate GI confers aggressive prostate cancer phenotypes by reprogramming transcriptional networks involved in proliferation, cell cycle progression, lineage plasticity and resistance to standard therapies. By developing network-based classifiers, this study shows clinical potential for GI signatures in prognostic risk stratification. The biological insights uncovered provide rationale for developing new therapeutic approaches targeting GI-driven vulnerabilities. Collectively, this integrative multi-omic analysis establishes GI as a key driver of lethal prostate cancer through coordinated transcriptional reprogramming, and nominates network-based genomic classifiers as promising biomarkers guiding prognosis and precision therapy.

6) Other achievement

A novel GI scoring method (pPCGI) facilitates identification of downstream transcriptional programs driven by highly altered genome and PCGI TNR classifier can separate tumors with poor clinical outcomes. This result also suggests that GI dependent transcriptional network associated with aggressive disease phenotype such as metastatic progression. We further developed a simplified method to compute GI score using precomputed genomic alteration profiles from cBioPortal database with significantly high concordance with original PCGI score (**Figure 1B**). This allows us to access many other data with less computational cost to identify prostate tumors with high genomic alterations.

Stable cell line systems allow assessment of transcriptional programs driven by novel driver genes, SYNE1 and TNFRSF21 according to SPOP and TP53 mutant. Transcriptional signatures and GI driver networks can similarly be driven by novel GI drivers such as SYNE1 and TNFRSF21. This suggests that the regulation of

transcriptional signatures can be altered depending on the mutational state of the genes. In addition, harmonized expression of GI driver networks according to SPOP and TP53 mutant can be a potential domain of targeted therapy for patients who are resistant to specific drugs or conventional treatment.

PTEN deletions in prostate cancer are associated with tumor aggression and poor outcome. Recent studies have implicated that PTEN-null prostate and other cancer cells have been reported to be sensitive to PARP inhibitors (PARPi). PTEN loss also impairs centromere instability, leading to spontaneous replication, mitotic stress, and subsequent stress tolerance.^{33, 34} Our studies showed SYNE1 and TNFRSF21 perturbation can promote synthetic lethal of PTEN null prostate cancer. Given this, SYNE1 and TNFRSF21 can be good therapeutic targets or markers for vulnerable patients to PARPi.³⁵

7) Reference

1. Robinson D, Van Allen EM, Wu YM, et al. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015;161: 1215-1228.
2. Wu YM, Cieslik M, Lonigro RJ, et al. Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer. *Cell*. 2018;173: 1770-1782 e1714.
3. Taylor BS, Schultz N, Hieronymus H, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell*. 2010;18: 11-22.
4. Hwang D, Rust AG, Ramsey S, et al. A data integration methodology for systems biology. *Proc Natl Acad Sci U S A*. 2005;102: 17296-17301.
5. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4: 44-57.
6. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102: 15545-15550.
7. Pan G, Bauer JH, Haridas V, et al. Identification and functional characterization of DR6, a novel death domain-containing TNF receptor. *FEBS Lett*. 1998;431: 351-356.
8. Mi S, Lee X, Hu Y, et al. Death receptor 6 negatively regulates oligodendrocyte survival, maturation and myelination. *Nat Med*. 2011;17: 816-821.
9. Zeng L, Li T, Xu DC, et al. Death receptor 6 induces apoptosis not through type I or type II pathways, but via a unique mitochondria-dependent pathway by interacting with Bax protein. *J Biol Chem*. 2012;287: 29125-29133.
10. Kasof GM, Lu JJ, Liu D, et al. Tumor necrosis factor- α induces the expression of DR6, a member of the TNF receptor family, through activation of NF- κ B. *Oncogene*. 2001;20: 7965-7975.
11. Saal LH, Johansson P, Holm K, et al. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc Natl Acad Sci U S A*. 2007;104: 7564-7569.
12. Ku SY, Rosario S, Wang Y, et al. Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science*. 2017;355: 78-83.
13. Herschkowitz JI, He X, Fan C, Perou CM. The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas. *Breast Cancer Res*. 2008;10: R75.
14. Kratochvill F, Neale G, Haverkamp JM, et al. TNF Counterbalances the Emergence of M2 Tumor Macrophages. *Cell Rep*. 2015;12: 1902-1914.
15. Luense LJ, Veiga-Lopez A, Padmanabhan V, Christenson LK. Developmental programming: gestational testosterone treatment alters fetal ovarian gene expression. *Endocrinology*. 2011;152: 4974-4983.
16. Jiang Y, Mei W, Gu Y, et al. Construction of a set of novel and robust gene expression signatures predicting prostate cancer recurrence. *Mol Oncol*. 2018;12: 1559-1578.
17. Miller ET, You S, Cadaneanu RM, et al. Chromosomal instability in untreated primary prostate cancer as an indicator of metastatic potential. *BMC Cancer*. 2020;20: 398.
18. Deb G, Thakur VS, Gupta S. Multifaceted role of EZH2 in breast and prostate tumorigenesis: epigenetics and beyond. *Epigenetics*. 2013;8: 464-476.
19. Gonzalez ME, DuPrie ML, Krueger H, et al. Histone methyltransferase EZH2 induces Akt-dependent genomic instability and BRCA1 inhibition in breast cancer. *Cancer Res*. 2011;71: 2360-2370.
20. Yepez VA, Mertes C, Muller MF, et al. Detection of aberrant gene expression events in RNA sequencing data. *Nat Protoc*. 2021;16: 1276-1296.
21. Heinlein CA, Chang C. Androgen receptor in prostate cancer. *Endocr Rev*. 2004;25: 276-308.
22. Kim HS, Vassilopoulos A, Wang RH, et al. SIRT2 maintains genome integrity and suppresses tumorigenesis through regulating APC/C activity. *Cancer Cell*. 2011;20: 487-499.

23. Schoendorf A, Rithner CD, Williams RM, Croteau RB. Molecular cloning of a cytochrome P450 taxane 10 beta-hydroxylase cDNA from *Taxus* and functional expression in yeast. *Proc Natl Acad Sci U S A*. 2001;98: 1501-1506.
24. Erfani A. Levels, trends, and determinants of unintended pregnancy in iran: the role of contraceptive failures. *Stud Fam Plann*. 2013;44: 299-317.
25. Ibrahim F, Maragkakis M, Alexiou P, Mourelatos Z. Ribothrypsis, a novel process of canonical mRNA decay, mediates ribosome-phased mRNA endonucleolysis. *Nat Struct Mol Biol*. 2018;25: 302-310.
26. Novershtern N, Subramanian A, Lawton LN, et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell*. 2011;144: 296-309.
27. Barretina J, Caponigro G, Stransky N, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature*. 2012;483: 603-607.
28. Miyamoto DT, Zheng Y, Wittner BS, et al. RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance. *Science*. 2015;349: 1351-1356.
29. Jan YJ, Yoon J, Chen JF, et al. A Circulating Tumor Cell-RNA Assay for Assessment of Androgen Receptor Signaling Inhibitor Sensitivity in Metastatic Castration-Resistant Prostate Cancer. *Theranostics*. 2019;9: 2812-2826.
30. Hou S, Zhao H, Zhao L, et al. Capture and stimulated release of circulating tumor cells on polymer-grafted silicon nanostructures. *Adv Mater*. 2013;25: 1547-1551.
31. Ke Z, Lin M, Chen JF, et al. Programming thermoresponsiveness of NanoVelcro substrates enables effective purification of circulating tumor cells in lung cancer patients. *ACS Nano*. 2015;9: 62-70.
32. Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nat Biotechnol*. 2008;26: 317-325.
33. Fan X, Kraynak J, Knisely JPS, Formenti SC, Shen WH. PTEN as a Guardian of the Genome: Pathways and Targets. *Cold Spring Harb Perspect Med*. 2020;10.
34. Shen WH, Balajee AS, Wang J, et al. Essential role for nuclear PTEN in maintaining chromosomal integrity. *Cell*. 2007;128: 157-170.
35. Fraser M, Zhao H, Luoto KR, et al. PTEN deletion in prostate cancer cells does not associate with loss of RAD51 function: implications for radiotherapy and chemotherapy. *Clin Cancer Res*. 2012;18: 1015-1027.

What opportunities for training and professional development has the project provided?

During the funding period, I promoted an Associate Professor in the Department of Urology and Computational Biomedicine at Cedars-Sinai Medical Center, and recently appointed to an Associate Director of the Center for Artificial Intelligence Research and Education (CAIRE). I have had the benefit of interactions with my collaborators, Drs. Freeman and Posadas, who are the world renown biologist and clinical scientist in the field of prostate cancer. These productive collaborations have stimulated my efforts aimed at enhancing my research capabilities in translational bioinformatics and growing my bioinformatics team within this Department of Urology. For this, I developed these resources: 1) unified data processing and analysis pipelines, 2) computational infrastructure by acquiring workstations and a cloud server used by my team, and 3) an in-house database containing a large collection of cancer data and molecular interaction information. I have actively advertised and shared these resources to my collaborators and the research community within Cedars-Sinai, as well as used for my research project development.

I have been mentoring many laboratory scientists during the funding period by providing both technical working knowledge and analytical strategies that have been crucial to the development of mentees' working research projects. For example, one of my mentees, Dr. Minhyung Kim published 15 papers as coauthor or first author in three years and plays a key role in 7 federally funded grants as a bioinformatician. Dr. Kim was recently promoted to Project Scientist in the Department of Surgery. I typically spend substantial time with mentees and collaborators in one-on-one discussions, in addition to the bioinformatics work. For example, I am a co-mentor of Dr. Ju Dong Yang, a physician scientist, and an awardee of a DoD Career Development Award, and I helped prepare his recent K08 application as a co-mentor. From this collaboration, Dr. Yang and I published 4 papers as co-authors and co-corresponding authors in *Nature Communications* (Na et al., 2020), *Advanced Materials and Technologies* (Na et al., 2021), *Liver Transplant* (Lee et al., 2022), and *Hepatology* (Na et al., 2023). In all, I have trained 3 undergraduate students, 8 graduate students, 4 research associates, 1 physician scientist and 11 post-doctoral fellows through the funding period.

How were the results disseminated to communities of interest?

Research products of this work have been presented at these invited talks: 1) School of Life Science Online Seminar in GIST Korea on June 3rd, 2021; 2) I-HDS (Institute for Healthcare Delivery Science) in collaboration with TCI (Tisch Cancer Institute) Seminar Series. November 1st, 2021; and 3) Department of Pharmacy Seminar Series at Yeungnam University, South Korea. October 26th, 2022.

In the first year of funding period, we published a research article in *Prostate Cancer and Prostatic Disease* as a senior author. This study describes a systematic comparison of PC subtyping methods, PCS and PAM50. We found that PCS is outperform in identification of tumors with fast metastatic progression. In second year of funding period, we published a comprehensive review paper early in *International Journal of Molecular Sciences* as a coauthor. This paper presents a comprehensive description of biomarker development process for prostate cancer detection and identification of aggressive prostate cancer with various omics data. In third year of funding period, we published 1 research paper in *Nano Today* (IF: 18.962) and 1 review paper in *Nature Reviews Urology* (*In press*), and submitted 2 research article in high impact journals. These published papers describe the detection of disease relevant gene expression signature in prostate cancer and bioinformatics approaches in urologic oncology research, respectively. We are preparing for a manuscript describing the results from this project with the goal of submitting it within the year.

What do you plan to do during the next reporting period to accomplish the goals?

After this final report, the transcriptome data produced by this study will be uploaded to the GEO database, and the results of the analysis will be published in a manuscript currently in preparation. The code and results used in this study will be made accessible to other researchers via GitHub.

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

We have made an important conceptual and clinically relevant advance by developing a novel method of characterizing PC using genomic and transcriptomic profiles. Consequently, this project is high impact and high reward, with potentially immediate opportunities to alter clinical practice if the classification scheme can be shown to have clinical utility. This new PC classification scheme we developed might improve prognostication of PC and enable the development of subtype-specific therapies and companion diagnostics. Using computational modeling, we have also identified 44 key TFs which appears to be highly active in CRPC/Met tumors, but which are associated with castrate-resistant phenotype, 2) inhibition of tumor suppressors, and 3) lineage plasticity, and therefore represents a comprehensive catalog of MRs in PC with highly altered genome.

What was the impact on other disciplines?

Nothing to Report.

What was the impact on technology transfer?

Nothing to Report.

What was the impact on society beyond science and technology?

Nothing to Report.

5. CHANGES/PROBLEMS

Changes in approach and reasons for change

Nothing to Report.

Actual or anticipated problems or delays and actions or plans to resolve them

We were able to successfully complete the analysis of cell line RNA-seq data in year 3, which was delayed by the COVID pandemic in year 2.

Changes that had a significant impact on expenditures

Nothing to Report.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents

Nothing to Report.

6. PRODUCTS:

Publications, conference papers, and presentations

Journal publications.

Yoon, J., Kim, M., Posadas, EM., Freedland, SJ., Liu, Y., Davicioni, E., Den, RB., Trock, BJ., Karnes, RJ., Klein, EA., Freeman, MR., **You, S.**, “A comparative study of PCS and PAM50 prostate cancer classification schemes”. *Prostate Cancer and Prostatic Diseases*, 2021 Feb;2;10.1038/s41391-021-00325-4. (2021). PMID: 33531653.

Acknowledgement of federal support (Yes)

Yan Y, Yeon SY, Qian C, **You S**, Yang W. On the Road to Accurate Protein Biomarkers in Prostate Cancer Diagnosis and Prognosis: Current Status and Future Advances. *International Journal of Molecular Sciences*. 2021 Dec 17;22(24):13537. doi: 10.3390/ijms222413537. PMID: 34948334; PMCID: PMC8703658.

Acknowledgement of federal support (Yes)

Wang, JJ., Sun, N., Lee, YT., Kim, M., Vagner, T., Rohena-Rivera, K., Wang, Z., Chen, Z., Zhang, R., Zhang, C., Tang, H., Widjaja, J., Zhang, TX., Qi, D., Teng, PC., Jan, YJ., Hou, KC., Hamann, C., Sandler, H., Daskivich, T., Luthringer, D., Bhowmick, N., Pei, R., **You, S***, Di Vizio, D*, Tseng, HR*, Chen, JF*, Zhu, Y*, Posadas, EM*, “Prostate Cancer Extracellular Vesicle Digital Scoring Assay – a rapid noninvasive approach for quantification of disease-relevant mRNAs”. *Nano Today*. 2023 Feb;48:101746. doi: 10.1016/j.nantod.2022.101746. (2023). PMID: 36711067; PMCID: PMC9879227. * co-corresponding author. Acknowledgement of federal support (Yes)

Adam, RM., Gheinani, AH., Kim, J., **You, S.**, “Bioinformatics in urology - molecular characterization of pathophysiology and response to treatment”. *In Press*. *Nature Reviews Urology*.

Acknowledgement of federal support (Yes)

Teng, PC., Jan, YJ., Chen, JF., Wang, JJ., Cheng, GCWS., Yao, N., Lozano, A., Chu, GCY., Chen, PJ., Ho, H., Yang, Y., Yeo, YH., Lee, YT., Huang, J., Li, KC., Chung, LWK., **You, S.**, Zhu, Y., Freeman, MR., Rogatko, A., Yang, JD., Tseng, HR., Posadas, EM., “Circulating tumor cells with small nuclear size are associated with poor clinical outcomes in advanced prostate cancer”. *Under Review*.

Acknowledgement of federal support (Yes)

Kim, M., Tamukong, P., Galvan, GC., Yang, Q., De Hoedt, A., Freeman, MR., **You, S.**, Freedland, SJ., “Distinct transcriptional programs in prostate cancers from black and white men”. *Under Review*.

Acknowledgement of federal support (Yes)

Conference papers

Wang, JJ., Cavassani, KA., Teng, PC., Chen, JF., Jan, YJ., Chu, CY., Lee, YT., Gao, A., Di Vizio, D., Chung, LWK., **You, S.**, Zhu, Y., Freeman, MR., Rogatko, A., Yang, JD., Tseng, HR., Posadas, EM., “Nuclear size of circulating tumor cells in advanced prostate cancer to reveal a potential biomarker for clinical outcomes and androgen receptor indifference” The 2021 ASCO Genitourinary Cancers Symposium, February 2021.

Wang, JJ., Cavassani, KA., Teng, PC., Chen, JF., Jan, YJ., Chu, CY., Lee, YT., Gao, A., Di Vizio, D., Chung, LWK., **You, S.**, Zhu, Y., Freeman, MR., Rogatko, A., Yang, JD., Tseng, HR., Posadas, EM., “A morphological subset of circulating tumor cells in advanced prostate cancer reveals a potential biomarker for clinical outcomes” The 2021 ASCO Annual Meeting, May 2021.

Kim H., **You, S.**, “An integrative data analysis pipeline for modeling transcriptional regulatory networks in accessible genome using ATAC-seq and RNA-seq data”. The Plant and Animal Genome XXIX Conference, January 2022.

Wang, JJ., Sun, N., Lee, YE., Kim, M., **You, S.**, Tseng, HR., Zhu, Y., Posadas, EM., “Extracellular vesicle-based assay for detecting metastases and dynamic monitoring of prostate cancer”, The 2022 ASCO Genitourinary Cancers Symposium, February 2022.

Wang, JJ., Sun, N., Lee, YE., Kim, M., **You, S.**, Tseng, HR., Zhu, Y., Posadas, EM., “A Nanotechnology-enabled, Extracellular Vesicle-based Assay for Reflecting Progression of Prostate Cancer”, The 2022 USCAP Annual Meeting, March 2022.

Wang, JJ., Kim, M., Teng, PC., Jan, YJ., Chen, JF., Zhu, Y., **You, S.**, Tseng, HR., Posadas, EM., “Luminal and basal subtyping of circulating tumor cells and influence on outcomes in patients with metastatic castration-resistant prostate cancer”, The 2023 ASCO Genitourinary Cancers Symposium, February 2023.

Cavassani, K., Posadas, EM., Teng, PC., Pustchi, SP., Kim, M., Bhowmick, NA., **You, S.**, Freeman, MR., Ellis, L., Di Vizio, D., Tseng, HR., Chung, LWK., Yao, S., Grasso, C., “Nuclear size and nuclear shape instability in prostate cancer epithelial cells during metastatic progression”, AACR Annual Meeting 2023. March 2023.

Qian, C., Yang, Q., Kim, H., Yan, Y., Mostaghel, E., Ellis, L., Knudsen, B., Kaochar, S., Ye, H., Garraway, I., **You, S.**, Freeman, MR., “ONECUT2 activates diverse drug-resistant phenotypes in prostate cancer”, AACR Annual Meeting 2023. March 2023.

Zhang, L., Billet, S., Gonzales, G., Rohena-Rivera, K., Muranaka, H., Lee, Y., **You, S.**, Posadas, EM., Bhowmick NA., “Dopamine receptor antagonists sensitize prostate cancer cells to androgen targeted therapy”, AACR Annual Meeting 2023. March 2023.

Books or other non-periodical, one-time publications.

Nothing to Report.

Other publications, conference papers, and presentations.

Lecture:

You S, Driver gene networks of genomic instability in prostate cancer progression. Gwangju Institute of Science and Technology (GIST), Korea. School of Life Science Online Seminar. June 3rd, 2021.

You S, Driver gene networks of genomic instability in prostate cancer progression. Icahn School of Medicine at Mount Sinai. I-HDS (Institute for Healthcare Delivery Science) in collaboration with TCI (Tisch Cancer Institute) Seminar Series. November 1st, 2021.

You S, Discovery of disease-relevant mRNAs as non-invasive blood biomarkers in liquid biopsy setting through bioinformatic approaches. Department of Pharmacy Seminar Series at Yeungnam University, South Korea. October 26th, 2022.

Website(s) or other Internet site(s)

Nothing to Report.

Technologies or techniques

Nothing to Report.

Inventions, patent applications, and/or licenses

Nothing to Report.

Other Products

Nothing to Report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	Sungyong You
Project Role:	Principal Investigator
Researcher Identifier:	yousung1
Nearest person month worked:	1.8
Contribution to Project:	Dr. You has supervised Drs. Minhyung Kim and Yeonjoo Lee to perform the genomic and transcriptomic data analysis. In addition, Dr. You wrote report and paper and provide lecture to disseminate this study result.
Funding Support:	W81XWH-20-1-0644 PI: You R01 CA246304 PIs: Agopian/Tseng/ You R01 CA253651 PIs: Tseng/Agopian/ You R01 CA252042 PIs: Hitchins/Hailes/ You

Name:	Michael R. Freeman
Project Role:	Co-Investigator
Researcher Identifier:	micfreeman
Nearest person month worked:	0.36
Contribution to Project:	Dr. Freeman has supervised wet-lab experiments Dr. Qian has performed and provides data interpretation for its biological implications in PCa.
Funding Support:	W81XWH-20-1-0644 PI: You R01 CA220327 PI: Freedland/Freeman W81XWH-19-1-0523 PI: Freeman W81XWH-20-1-0695 PI: Freeman

Name:	Edwin M. Posadas
Project Role:	Co-Investigator
Researcher Identifier:	yousung1
Nearest person month worked:	0.36
Contribution to Project:	Dr. Posadas has supervised collection of clinical samples and provides data interpretation for its clinical implications.
Funding Support:	W81XWH-20-1-0644 PI: You W81XWH-19-1-0390 PI: Posadas

Name:	Minhyung Kim
Project Role:	Co-Investigator
Researcher Identifier:	MINHYUNGK
Nearest person month worked:	3.6
Contribution to Project:	Dr. Kim performed differential expression analysis and survival analysis. He also performed CTC data analysis and classifier development and application.
Funding Support:	W81XWH-20-1-0644 PI: You R01 CA246304-01 PIs: Agopian/Tseng/ You

Name:	Yeonjoo Lee
Project Role:	Co-Investigator
Researcher Identifier:	LEEYEO

Nearest person month worked:	8.4
Contribution to Project:	Dr. Lee performed PCGI computation, differential gene expression analysis, TRN network modeling, TRN score computation and classifier development.
Funding Support:	W81XWH-20-1-0644 PI: You

Name:	HyoYoung Kim
Project Role:	Co-Investigator
Researcher Identifier:	
Nearest person month worked:	1.2
Contribution to Project:	Dr. Kim performed mutations and variant calling with RNA-seq data from PCa cell lines. For this, she developed automated pipeline and support mutation analysis of the TCGA and MET440 data.
Funding Support:	W81XWH-20-1-0644 PI: You R01 CA246304 PIs: Agopian/Tseng/ You R01 CA253651 PIs: Tseng/Agopian/ You R01 CA252042 PIs: Hitchins/Hailes/ You

Name:	Chen Qian
Project Role:	Co-Investigator
Researcher Identifier:	CHENQ
Nearest person month worked:	1.2
Contribution to Project:	Dr. Qian performed all wet-lab experiments with PCa cell lines, RNA extraction for RNA-seq data generation, and RT-PCR validation.
Funding Support:	W81XWH-20-1-0644 PI: You R01 CA220327 PI: Freedland/Freeman W81XWH-19-1-0523 PI: Freeman W81XWH-20-1-0695 PI: Freeman

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

Nothing to Report.

What other organizations were involved as partners?

Nothing to Report.

8. SPECIAL REPORTING REQUIREMENTS: Nothing to Report.

9. APPENDICES: Published papers supported by this funding.



Clinical Research

A comparative study of PCS and PAM50 prostate cancer classification schemes

Junhee Yoon¹ · Minhyung Kim¹ · Edwin M. Posadas^{2,3} · Stephen J. Freedland^{1,4} · Yang Liu⁵ · Elai Davicioni⁵ · Robert B. Den⁶ · Bruce J. Trock⁷ · R. Jeffrey Karnes⁸ · Eric A. Klein⁹ · Michael R. Freeman^{1,10,11} · Sungyong You^{1,10}

Received: 25 May 2020 / Revised: 20 December 2020 / Accepted: 15 January 2021
© The Author(s) 2021. This article is published with open access

Abstract

Background Two prostate cancer (PC) classification methods based on transcriptome profiles, a de novo method referred to as the “Prostate Cancer Classification System” (PCS) and a variation of the established PAM50 breast cancer algorithm, were recently proposed. Both studies concluded that most human PC can be assigned to one of three tumor subtypes, two categorized as luminal and one as basal, suggesting the two methods reflect consistency in underlying biology. Despite the similarity, differences and commonalities between the two classification methods have not yet been reported.

Methods Here, we describe a comparison of the PCS and PAM50 classification systems. PCS and PAM50 signatures consisting of 37 (PCS37) and 50 genes, respectively, were used to categorize 9,947 PC patients into PCS and PAM50 classes. Enrichment of hallmark gene sets and luminal and basal marker gene expression were assessed in the same datasets. Finally, survival analysis was performed to compare PCS and PAM50 subtypes in terms of clinical outcomes.

Results PCS and PAM50 subtypes show clear differential expression of PCS37 and PAM50 genes. While only three genes are shared in common between the two systems, there is some consensus between three subtype pairs (PCS1 versus Luminal B, PCS2 versus Luminal A, and PCS3 versus Basal) with respect to gene expression, cellular processes, and clinical outcomes. PCS categories displayed better separation of cellular processes and luminal and basal marker gene expression compared to PAM50. Although both PCS1 and Luminal B tumors exhibited the worst clinical outcomes, outcomes between aggressive and less aggressive subtypes were better defined in the PCS system, based on larger hazard ratios observed.

Conclusion The PCS and PAM50 classification systems are similar in terms of molecular profiles and clinical outcomes. However, the PCS system exhibits greater separation in multiple clinical outcomes and provides better separation of prostate luminal and basal characteristics.

These authors contributed equally: Junhee Yoon, Minhyung Kim

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41391-021-00325-4>.

✉ Sungyong You
Sungyong.You@cshs.org

¹ Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, CA, USA

² Urologic Oncology Program & Uro-Oncology Research Program, Cedars-Sinai Cancer, Cedars-Sinai Medical Center, Los Angeles, CA, USA

³ Division of Oncology, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA

⁴ Division of Urology, Department of Surgery, Veteran Affairs Healthcare System, Durham, NC, USA

⁵ Decipher Biosciences Inc., San Diego, CA, USA

⁶ Department of Radiation Oncology, Jefferson Medical College of Thomas Jefferson University, Philadelphia, PA, USA

⁷ James Buchanan Brady Urological Institute, Johns Hopkins Hospital, Baltimore, MD, USA

⁸ Department of Urology, Mayo Clinic, Rochester, MN, USA

⁹ Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH, USA

¹⁰ Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA

¹¹ Department of Medicine, University of California, Los Angeles, CA, USA

Introduction

There has been much recent progress in the use of prostate cancer (PC) genomics to identify drivers of aggressive disease. These advances have spurred development of commercially available genomic classifiers that can identify aggressive tumors at high risk of progression to metastatic and/or castration resistant PC (CRPC) [1–7]. Despite this technology, there is no universally accepted or widely used molecular subtyping system for PC, unlike other cancers such as breast cancer, in which luminal A, luminal B, and basal subclassification are commonplace and clinically meaningful. Despite these limitations, PC subtypes have been proposed based on genomic criteria, such as various somatic alterations in chromatin sequence, (e.g., TMPRSS-ERG fusions [8, 9]) and androgen receptor amplification [3, 4]. The Cancer Genome Atlas identified several genomic PC subtypes, referred to as ERG, ETV1, ETV4, FLI1, SPOP, FOXA1, IDH1, and Other [7]. Tomlins et al. [10] described four subtypes based on gene expression (ERG+, ETV+, SPINK1+, and Triple Negative (ERG–/ETS–/SPINK1–)). However, the clinical applicability of these PC genomic classifiers has been limited [11, 12].

Many reports have proposed gene expression signatures as a means of tumor classification [13–18]. Two transcriptome-based classification methods were recently reported that categorize PC into three subtypes [19, 20]. Our group described the prostate cancer classification system (PCS), an integrated approach employing activation signatures of 14 pathways associated with PC biology to interrogate a virtual cohort of 1321 clinical samples [19]. Two luminal subtypes (PCS1 and PCS2) and one basal subtype (PCS3) were described in this report. PCS1 tumors exhibited the poorest clinical outcomes, including increased risk of metastatic progression, PC-specific mortality, and overall survival. In contrast, no significant differences in clinical outcomes were observed between PCS2 and PCS3; however, both PCS1 and PCS3 tumors were enriched in bone metastases in comparison to PCS2 [21]. In the initial report, we validated the PCS scheme with ten independent patient cohorts and 19 laboratory models of PC. In line with this, PCS1-specific genes were highly expressed in androgen receptor signaling inhibitor (ARSI) resistant PC [19] and this was independently validated by our recent study using a novel circulating tumor cell RNA assay system [22]. The initial You et al. study [19] was the first to categorize PC into only three subtypes using genomic approaches.

Zhao et al. [20] applied a variation of the widely used PAM50 breast cancer classifier to a large PC transcriptome dataset. PAM50 classifies breast cancer into luminal A (LumA), luminal B (LumB), HER2, Normal-like, and Basal subgroups [23]. The adaptation of PAM50 to PC focused on

luminal and basal PC phenotypes and disregarded the HER2 and Normal-like subtypes described in the standard PAM50 system. Similar to the PCS system, PAM50 revealed that transcriptome data alone can divide PCs into only three subtypes, LumA, LumB, and Basal. These subgroups were associated with varied clinical behaviors. LumB exhibited the worst survival rate, while LumA and Basal subtypes exhibited similar clinical outcomes with better survival rates. In comparison to other classification schemes, which identified as many as seven PC subtypes [1–7], it is intriguing that both PCS and PAM50 concluded that PCs can be categorized into two distinct luminal and one basal subtype using only transcriptome data.

To date, no comparison has been made between these two different three-category classification systems. To address this gap, here we present a comparison of the PCS and PAM50 systems. We applied the PCS and PAM50 methods to two large PC transcriptome datasets: (1) the Prostate Cancer Transcriptome Atlas (PCTA) and (2) the Decipher GRIDTM database (GRID) (Supplementary Table 1). The PCTA is a virtual cohort consisting of 1,321 PC transcriptome profiles. These data were used to develop the PCS [19]. The GRID is a cohort consisting of 8626 PC transcriptome profiles, a subset of which was employed in the Zhao et al. study [20]. We hypothesized that the PCS and PAM50 systems have many similarities as well as differences in terms of molecular profiles and clinical outcomes. We thus performed a comparative analysis assessing three different measures: (1) gene expression patterns, (2) pathway associations, and (3) correlation with clinical outcomes.

Methods

Statistical analysis

To examine the association between clinical outcomes and PCS and PAM50 categories in the tumors from the GRID dataset, we used Kaplan–Meier survival analysis and Cox proportional hazards regression analysis with the following outcomes: biochemical recurrence (BCR), metastasis (Met), and PC-specific mortality (PCSM). To test whether PCS or PAM50 classification was prognostic independent of other clinical variables, multivariable analyses were performed adjusting for pathological grade, which was categorized into three groups with Gleason sum score of less than 7, equal to 7, and more than 7. All analyses were conducted using Python (version 2.7). $p < 0.05$ was considered statistically significant. Detailed information of transcriptome datasets and analysis procedures used in this study are described in Supplementary Methods.

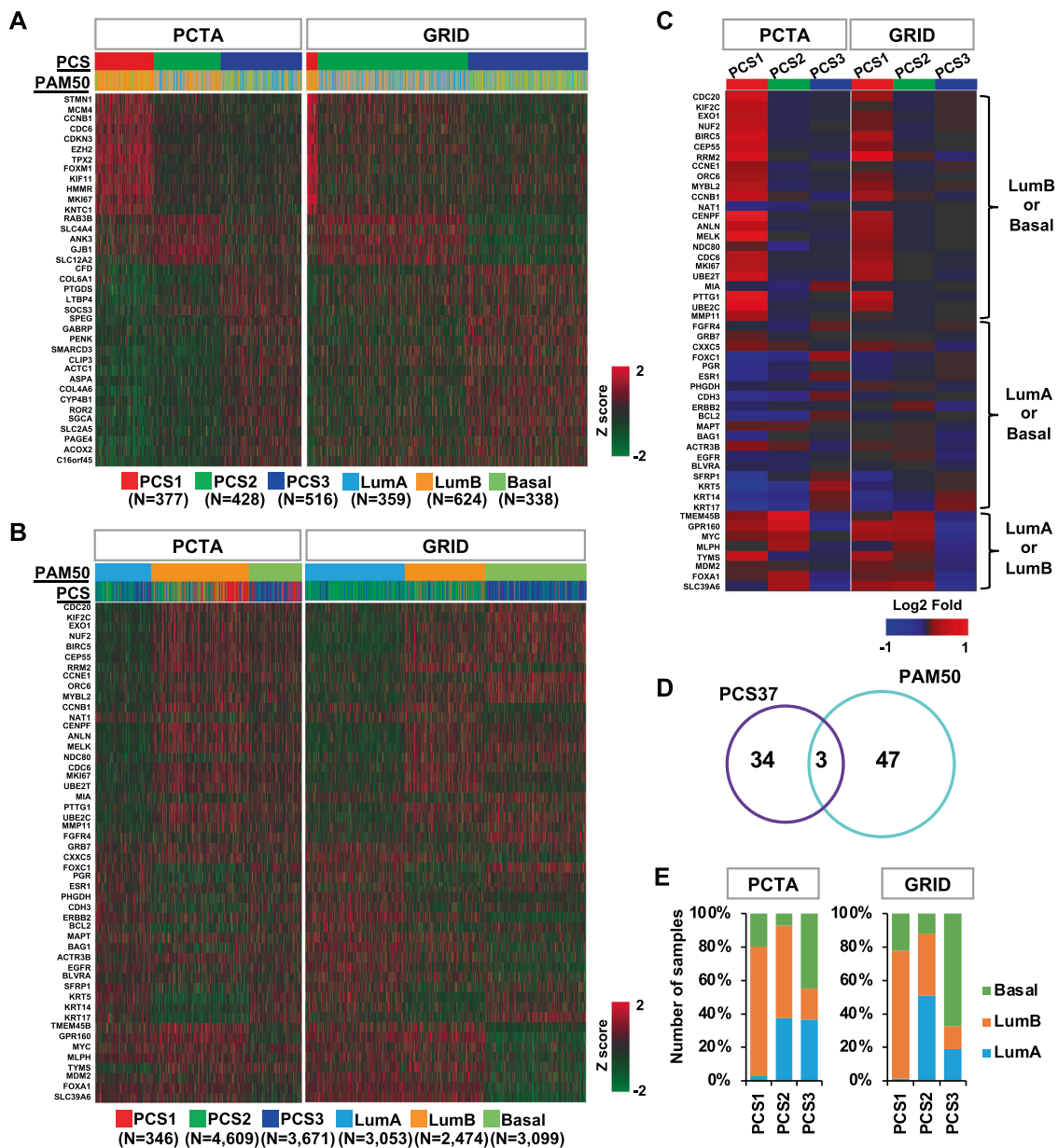


Fig. 1 PCS37 and PAM50 genes in the PCTA and the GRID datasets. The heatmap depicts differential expression of PCS37 and PAM50 in the PCTA and GRID datasets based on **A** PCS and **B** PAM50 grouping. **C** The heatmap shows differential expression of PAM50 genes in the PCTA and GRID datasets based on PCS

grouping. **D** The Venn diagram of PCS37 and PAM50. **E** The distribution of PAM50 subtypes across three PCS subtypes in the PCTA and GRID datasets. PCS prostate cancer subtype, LumA Luminal A subtype, LumB Luminal B subtype, Basal Basal subtype.

Results

Conserved gene expression patterns between PCS and PAM50

We applied the PCS and PAM50 classifiers to assign the tumors from the PCTA cohort ($n = 1,321$) [19] and the GRID cohort ($n = 8,626$) into PCS1-3 and LumA/B and Basal subtypes. Differential expression patterns of PCS37 genes [19] were displayed by PCS classes with the PCTA

and the GRID datasets (Fig. 1A and Supplementary Fig. 1), showing distinct expression patterns of individual PCS class-specific genes. For example, 12 PCS1-specific genes (STMN1, MCM4, CCNB1, CDC6, CDKN3, EZH2, TPX2, FOXM1, KIF11, HMMR, MKI67, and KNTC1) are only highly expressed in PCS1 tumors (Fig. 1A). The GRID contains 4% (346/8,626) PCS1, the most aggressive PCS subtype, as most specimens were primary PCs. In the PCTA cohort, containing 260 metastatic PCs, 26% (377/1,321) of patients were assigned to PCS1. In total, 66% (176/260) of

metastatic PCs belong to this PCS1 subtype. LumB, the most aggressive PAM50 subtype, accounts for 29% (2,474/8,626) of the GRID cohort and 47% (624/1,321) of the PCTA cohort. Of 260 metastatic PCs in the PCTA cohort, 170 (65%) patients were assigned LumB.

We then assessed PAM50 gene expression by PAM50 classes with both datasets (Fig. 1B). These 50 genes also showed distinct expression patterns among the classes. However, their expression was not unique to a specific subtype. For example, some genes that were highly expressed in LumB also showed high expression in Basal. In addition, some genes (e.g., NAT1, GRB7, MAPT) did not show a clear differential expression pattern among the PAM50 classes. Both PCS37 and PAM50 displayed clear separation into three PC classes. We further assessed the expression pattern of PAM50 genes by the PCS classifier (Fig. 1C). Of note, genes with high expression in LumB or Basal were highly expressed in PCS1. These genes exhibit mostly low expression in LumA.

Given that both classification methods resolve all PCs into three subtypes, it is interesting that only three genes (CCNB1, CDC6, and MKI67) overlap between the two systems (Fig. 1D). These three genes show high expression in both PCS1 and LumB. We observed all PCS categories in all classification groups as defined by PAM50 in both PCTA and GRID (Fig. 1E). We found a high frequency of PCS2 in LumA, but not PCS1 and PCS3. LumB was enriched for PCS1, while Basal subtype but not LumA or LumB was enriched for PCS3. This was validated by visualizing the distribution of the tumors from each category and their overlaps using PCS37 and PAM50 gene expression (Supplementary Fig. 2A). We then defined centroids of each categories and estimated the pairwise distances between the PCS and PAM50 categories (Supplementary Fig. 2B and Supplementary Table 2). Consistent with the enrichment of sample numbers shown in Fig. 1E, three subtype pairs (PCS1 versus LumB, PCS2 versus LumA, and PCS3 versus Basal) exhibit the shortest distance between centroids. Taken together, three subtype pairs (PCS1 versus LumB, PCS2 versus LumA, and PCS3 versus Basal) show comparable enrichment in terms of number of tumors commonly included in both subtypes, despite the small number of genes shared by PCS37 and PAM50.

Pairwise overlap of PCS and PAM50 classes in cellular functions

We computed enrichment scores using the GSEA method [24] with hallmark gene sets [25] for all the PCS and PAM50 categories. In the PCTA, PCS1 samples were highly enriched with E2F targets and G2M checkpoint; PCS2 samples were enriched with androgen response, fatty acid metabolism, and cholesterol homeostasis; and PCS3

was enriched with IL6-JAK-STAT3 signaling and KRAS signaling (UP) (Fig. 2A). PAM50-classified samples in the PCTA showed an overall similar enrichment result with PCS classification (Fig. 2A). LumA samples were enriched in fatty acid metabolism, androgen response, cholesterol homeostasis. LumB samples were enriched with the hallmark gene sets of PCS1, such as E2F targets and G2M checkpoint. Basal samples were enriched in the same hallmark gene sets as PCS3 (Fig. 2B). In the GRID, PCS1, and PCS2 were enriched in E2F targets, G2M checkpoint, fatty acid metabolism, androgen response, and cholesterol homeostasis, but PCS3 did not show any significant enrichment of any of these hallmark gene sets. PAM50-classified samples in the GRID showed the same enrichment pattern. LumB exhibited higher ES of E2F targets and G2M checkpoint than LumA, but the highest score was androgen response in LumB-enriched hallmark gene sets (Supplementary Fig. 3A). The highest enrichment score in LumA was androgen response, similar to PCS2 (Supplementary Fig. 3B).

Luminal and basal marker expression in PCS and PAM50 classifications

We assessed expression levels of luminal and basal cell marker genes in the PCTA and the GRID datasets [19, 26]. Expression of these marker genes was highly concordant with PCS classification in both the PCTA (Fig. 3A) and the GRID (Fig. 3B), consistent with our prior results in the initial PCS report [19]. In the PAM50 classes, LumB and Basal subtypes display clear high and low expression of luminal and basal marker genes, respectively. However, LumA tumors exhibit mixed marker gene expression, indicating a composite luminal and basal expression phenotype. This result suggests that the PCS classification scheme is superior in separating prostatic luminal and basal gene expression.

PCS1 and LumB are associated with the worst survival outcomes

We examined relationships between subtype category and the clinical outcomes of PCSM, BCR, and Met using the GRID. PCS1 within the PCS categories (Fig. 4A) and LumB within the PAM50 categories (Fig. 4B) were the most aggressive for all three metrics (PCSM, BCR, and Met). Cox proportional hazard analysis was performed using the two classification systems. In univariable analysis using the PCS scheme, the PCS1 subtype exhibited the highest hazard ratio (HR) (PCSM: 3.91 [2.78–5.52], BCR: 2.48 [1.34–4.08], Met: 3.91 [2.78–5.52]). Similarly, the LumB subtype exhibited the highest HR (PCSM: 1.85 [1.32–2.61], BCR: 1.54 [1.01–2.29], Met: 2.04 [1.58–2.55])

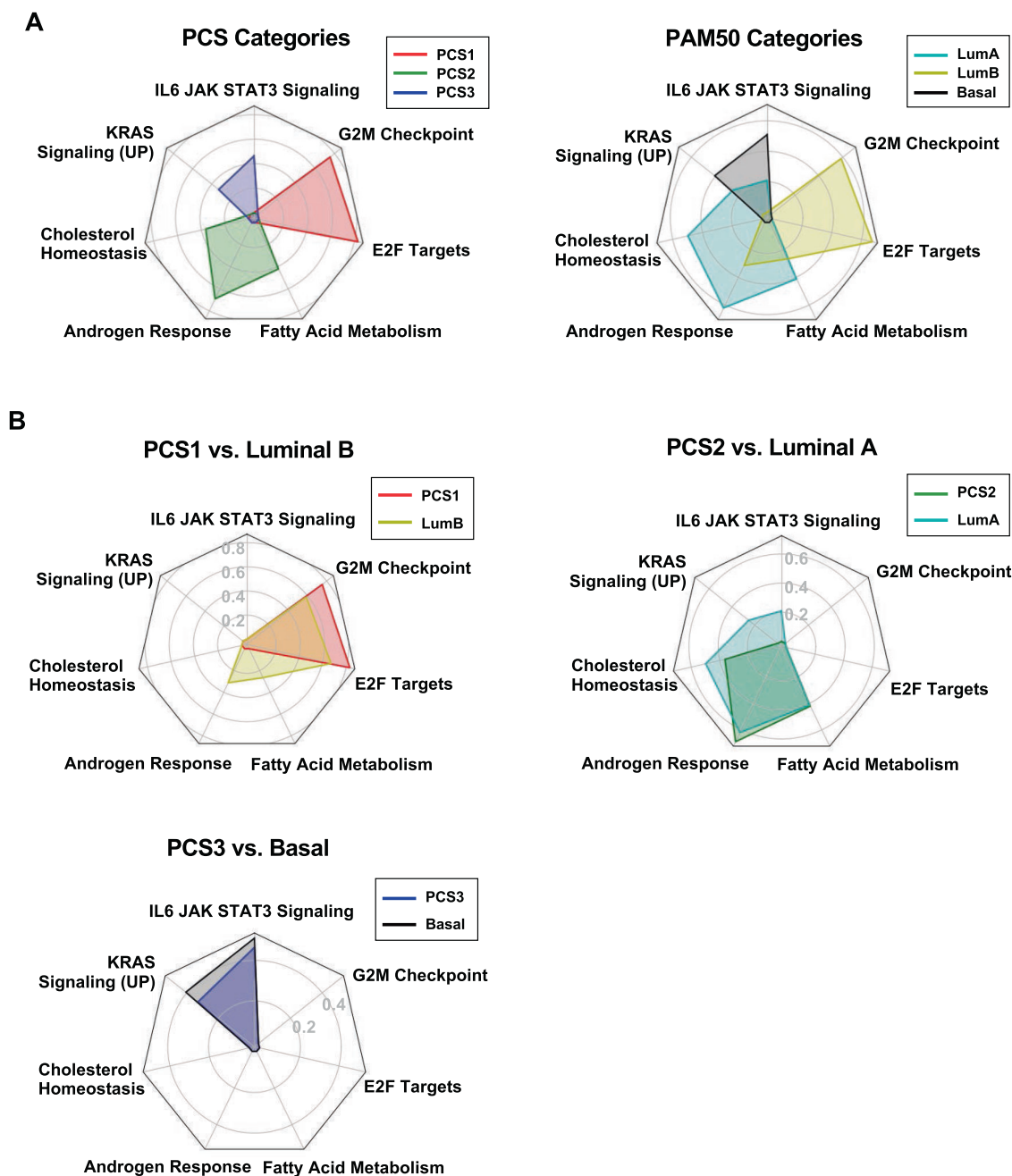


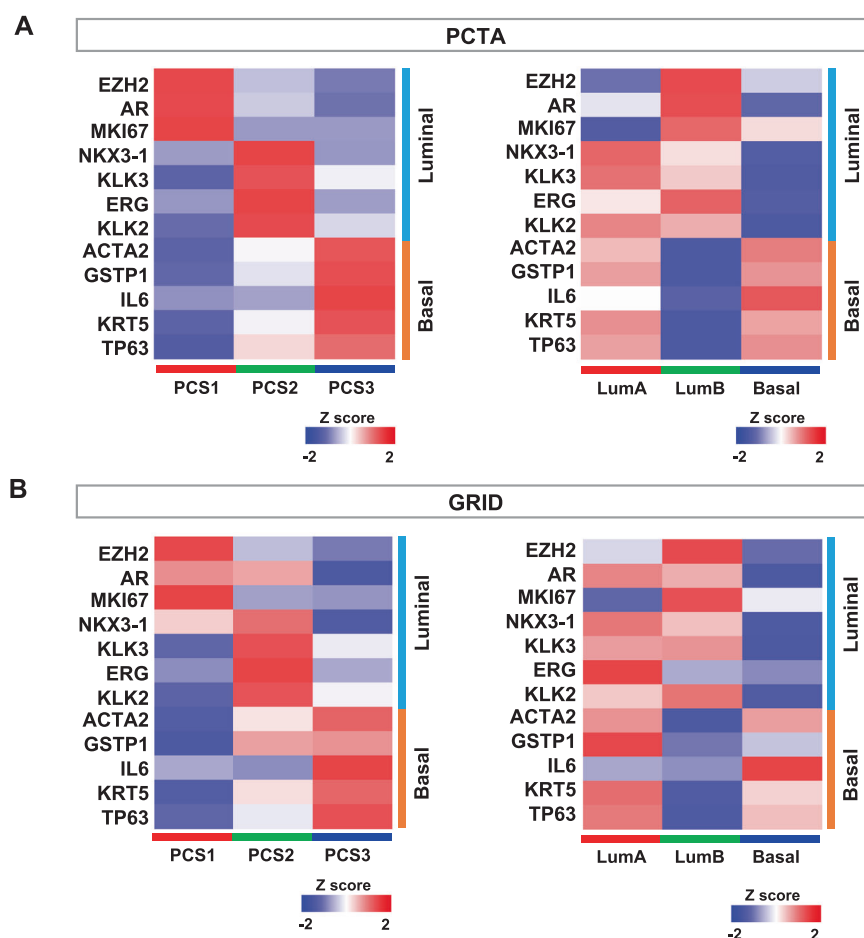
Fig. 2 Enriched cellular processes of PCS and PAM50 subtypes in the PCTA. **A** Enriched seven hallmark gene sets of PCS (left) and PAM50 (right) categories were displayed with radar chart in the PCTA cohort. **B** Radar chart illustrates overlaps of enriched hallmark gene

(Fig. 4C). In all combinations of pairwise comparisons, except for LumB compared to LumA in BCR (HR = 3.41 [2.40–4.87]), PCS1 exhibited higher HRs compared to LumB (Supplementary Table 3). We also performed multivariable analysis of PCS or PAM50 adjusting for Gleason score (Table 1). After the adjustment, PCS still showed larger HR (PCSM: 3.76 [2.17–6.25], BCR: 2.03 [1.39–2.96], Met: 3.38 [2.26–5.06]) compared to PAM50 (PCSM: 1.51 [0.96–2.38], BCR: 1.43 [1.14–1.79], Met:

sets by PCS1 versus LumB, PCS2 versus LumA, and PCS3 versus Basal in the PCTA cohort. LumA Luminal A subtype, LumB Luminal B subtype, Basal Basal subtype.

1.73 [1.30–2.29]). Importantly, we found that controlling for Gleason score did not dramatically affect the HRs in either PCS or PAM50, indicating both are largely independent of Gleason grade for predicting patient outcomes. Of note, both multivariable PCS and PAM50 models with Gleason score exhibited the same concordance index at 0.75. We further checked the association of each subtype and Decipher score [27], which is designed to predict metastasis risk after radical prostatectomy. The Decipher

Fig. 3 Gene expression of Luminal and Basal cell markers in the PCTA and the GRID. The heatmap displays expression of the Luminal and Basal cell marker genes in the PCTA (A) and the GRID (B).



score was significantly higher in PCS1 and LumB compared to other subtypes (Supplementary Fig. 4A, B). Multi-variable analysis of Decipher score after adjustment of Gleason score was performed to compare with PCS and PAM50. Decipher score exhibited the highest HRs and significant *p* values in PCSM, BCR, and Met (Supplementary Fig. 4C).

Discussion

In this study, we performed a comprehensive cross-comparison of subtype assignments obtained from the PCS and PAM50 classification methods. We found that three subtype pairs (PCS1 versus LumB, PCS2 versus LumA, and PCS3 versus Basal) show conserved gene expression patterns in the tumors included in both subtypes. Distance metrics of subtype centroids followed by visualization showed a similarity between these three pairs of subtypes. Despite the fact that only three genes overlap between the PCS37 and PAM50 signatures, these three subtype pairs exhibit very similar enrichment of cellular processes, as well as the distance of the centroids between the subtypes (e.g., PCS1 and LumB). This result suggests high concordance of the two

subtypes in cellular functions and clinical phenotypes. Finally, a further comparison was made in clinical outcomes with PCSM, BCR, and Met, demonstrating that PCS1 and LumB subtypes are the most aggressive tumors with the worst survival outcomes in each categorization scheme.

The two classification methods likely reflect underlying luminal and basal biology as described in basic PC studies [19, 20]. PCS was originally designed as an unbiased classification scheme, with 14 PC relevant pathway activation signatures that revealed two luminal and one basal phenotype, while the PAM50 as applied to PC was biased toward luminal and basal phenotypes at the outset [20]. As such, it is interesting that two distinct PC categorization schemes converged to similar conclusions about the manner in which PCs can be grouped from transcriptome data alone. As shown in Fig. 3, both PCS and PAM50 exhibit differential expression of luminal and basal marker genes. However, PCS classes show consistent luminal and basal marker gene expression with clear separation between PCS1/2 and PCS3, respectively, while LumA tumors have high expression of basal marker genes, which are highly expressed in Basal tumors as well. This suggests that PCS classification more accurately identifies prostate luminal and basal phenotypes expressed within the cancer specimens.

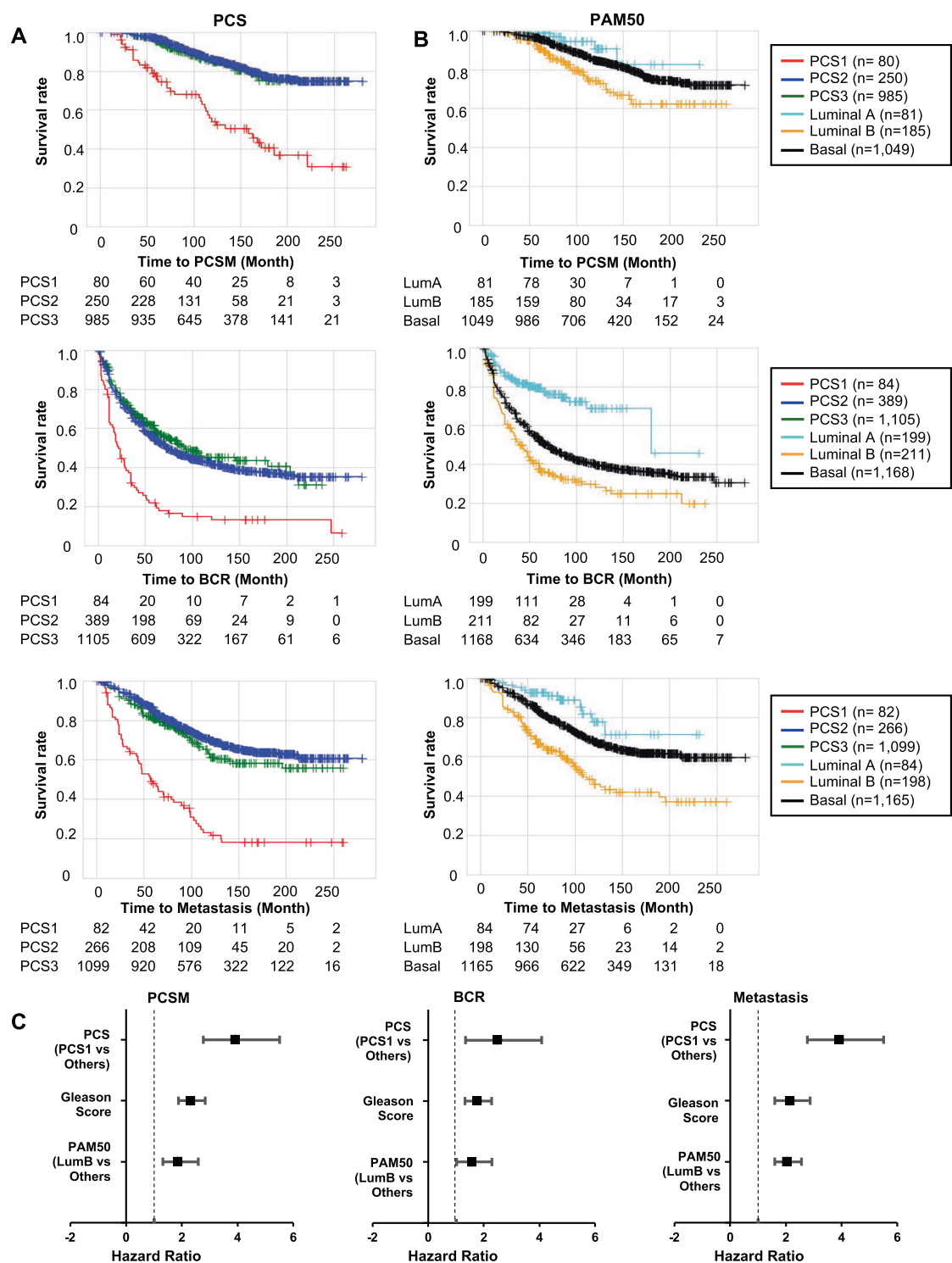


Fig. 4 Clinical outcomes in distinct PCS and PAM50 subtypes in the GRID. Kaplan–Meier survival curves shows differential clinical outcome association of PCS (A) and PAM50 (B) categorization. Tables below the KM plot represent the number at risk. (PCSM

prostate cancer-specific survival, BCR biochemical recurrence, Met metastasis.) C Forest plots display hazard ratios of PCS, PAM50, and Gleason grade against PCSM, BCR, and Met. (GS Gleason score).

The development of classification schema such as PCS and PAM50 in PC represents an important evolution in the field. Rather than considering an isolated biomarker or limited

groups of biomarkers, these approaches focus on patterns of gene expression that ultimately relate to key biological drivers of the disease process. Thus far, these classifiers have shown

Table 1 Multivariable Cox proportional hazard regression analysis for PCS and PAM50 classification with clinical parameters.

Multivariable analysis with PCS				Multivariable analysis with PAM50			
Variable	Hazard ratio	<i>p</i> value	C-index	Variable	Hazard ratio	<i>p</i> value	C-index
PCSM							
PCS	3.76 (2.17–6.52)	<0.001	0.75	PAM50	1.51 (0.96–2.38)	0.073	0.75
Gleason score	2.16 (1.74–2.66)	<0.001		Gleason score	2.25 (1.83–2.78)	<0.001	
BCR							
PCS	2.03 (1.39–2.96)	<0.001	0.66	PAM50	1.43 (1.14–1.79)	0.002	0.67
Gleason score	1.71 (1.55–1.88)	<0.001		Gleason score	1.69 (1.54–1.86)	<0.001	
Met							
PCS	3.38 (2.26–5.06)	<0.001	0.71	PAM50	1.73 (1.30–2.29)	<0.001	0.72
Gleason score	2.03 (1.79–2.31)	<0.001		Gleason score	2.04 (1.80–2.32)	<0.001	

Multivariate analysis of PCS or PAM50 classification system with Gleason score was performed.

the capacity to enhance clinical prediction and prognostication beyond conventional clinical criteria used in practice today. Emerging data with PAM50 and PCS point toward their potential utility in personalizing the approach to PC care, particularly with the use of systemic therapies such as ARSIs (e.g., abiraterone, enzalutamide, apalutamide, and darolutamide) and taxanes (e.g., docetaxel and cabazitaxel). Analysis of samples from the SPARTAN clinical trial revealed clinically important differences in behavior in response to apalutamide, based on PAM50 classification. In SPARTAN, PAM50 Basal patients seemed to have a greater absolute benefit from ADT and apalutamide, suggesting these patients may benefit from intensification with ARSIs [28]. Consistent with this, analysis of PCS1 and LumB specimens in the GRID from patients receiving ADT showed higher rates of PCSM, BCR, and Met compared to PCS2/3 or LumA/Basal (Supplementary Fig. 5). However, absence of information on ADT response in the GRID prohibits analysis of relative performance of the PCS and PAM50 classifications with respect to this clinical variable. While encouraging, prospective validations of these data are needed prior to more widespread deployment of these approaches. Studies are now underway that will provide important insights into this question, including the ongoing NRG-GU-006/BALANCE study (NCT03371719), which used PAM50 as a stratification variable, and has now finished accrual and is expected to yield results within the next 3 years.

While tissue-based classification will remain an important standard in PC, the clinical behavior of this disease makes obtaining tissue samples from patients with metastatic disease challenging. Given the potential importance of these tools in advanced PC/metastatic CRPC, additional means of deploying these technologies is crucial. Fortunately, given recent technical advances, it will become possible to conduct genomic classifications using liquid biopsies [22]. These blood-based tools will also require prospective validation. Liquid biopsies are included in

ongoing studies, including NRG-GU-006/BALANCE. These initial studies will provide a foundational experience for use of PCS and PAM50 in clinical practice.

In order for these classification schemes to be useful in a clinical setting, it is critical that they offer actionable prognostic information. We observed a disparate percentage of subtype assignments within the study cohorts. The GRID consists of a prospective cohort ($n = 7,000$) and a retrospective cohort ($n = 1,626$) as shown in Supplementary Fig. 1. In the GRID, only 3.7% of the prospective cohort and 5% of the retrospective cohort are classified as PCS1, whereas the PCTA cohort contains 19% of primary tumors classified as PCS1. The likely explanation for this difference is that the GRID and the PCTA contain different proportions of high grade tumors or tumors with distinct metastatic potential. However, these three cohorts include a similar percentage of high risk patients (Gleason score > 7), and PCS1 tumors exhibit the worst prognosis in low grade tumors (Gleason score ≤ 7) based on our study [19]. This indicates that the PCS categorization is a variable independent variable of Gleason grade. On the other hand, the proportion of PCS1 (6% in our previous study [19]) and the incidence rate of metastases were similar, with about 3–6% of newly diagnosed PC cases with metastasis [29, 30]. The PCS1 assignment rate can therefore vary according to metastatic potential independent of pathological grade.

It is also well known that dataset composition and choices for reference construction affect subsequent subtype calling. Of note, standard PAM50 classification used in Zhao et al. and the present study is profoundly affected by the composition of the sample cohort used for reference construction. Zhao et al. conducted median centering to the GRID retrospective and prospective cohorts separately prior to applying the PAM50 algorithm. Thus, the percentage of PAM50 categories was assigned different median values from each cohort. In this study, we applied single median

values based on the GRID, combining retrospective and prospective cohorts because the median value approaches the actual population mean as the sample number increases. In future studies, it will be necessary to introduce a method for improving classification robustness by applying appropriate reference construction [31].

PC is one of the malignancies most affected by genetic factors [32], and a range of genomic alterations and structural variations associated with the clinical outcomes have been described [3, 7, 33]. However, the present study is based only on transcriptome data, and did not consider genetic or structural changes associated with the cancers in the individual PCS and PAM50 categories. Thus, our report identified differences and similarities between the two classification schemes but did not make comparisons with other PC subtyping systems. In the future, comparative analysis with other classification methods using multi-omics data [3, 7, 34] will complement the work described here.

In conclusion, PCS and PAM50 present a new lens through which PC may be viewed. There are important similarities in these signatures despite obvious differences in origin and performance (based on existing datasets or samples). Prospective validation of these tools is underway and will help to clarify how they may be most effectively deployed in clinical practice.

Acknowledgements This research was supported by National Cancer Institute (R01-CA220327, R01-CA218356, P01-CA098912, P01-CA233452, U01-CA198900), Department of Defense Prostate Cancer Research Program (PC151088, PC171066, PC180541, PC190482), Steven Spielberg Discovery Fund in Prostate Cancer Research, St. Anthony Prostate Cancer Research Fund, CD McKinnon Memorial Fund for Neuroendocrine Prostate Cancer, Michael & Patricia Berns Family Prostate Cancer Research Fund, Prostate Cancer Foundation Stuart Rahr Young Investigator Award, Prostate Cancer Foundation VALOR Challenge Award.

Compliance with ethical standards

Conflict of interest YL and ED are employees of Decipher Biosciences.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell*. 2010;18:11–22.
2. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. *Cell*. 2013;153:666–77.
3. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature*. 2012;487:239–43.
4. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative clinical genomics of advanced prostate cancer. *Cell*. 2015;161:1215–28.
5. Spratt DE, Dai DLY, Den RB, Troncoso P, Yousefi K, Ross AE, et al. Performance of a prostate cancer genomic classifier in predicting metastasis in men with prostate-specific antigen persistence postprostatectomy. *Eur Urol*. 2018;74:107–14.
6. 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;10:5251.
7. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell*. 2015;163:1011–25.
8. Perner S, Demichelis F, Beroukhi R, Schmidt FH, Mosquera JM, Setlur S, et al. TMPRSS2: ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. *Cancer Res*. 2006;66:8337–41.
9. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature*. 2007;448:595–9.
10. Tomlins SA, Alshalalfa M, Davicioni E, Erho N, Yousefi K, Zhao S, et al. Characterization of 1577 primary prostate cancers reveals novel biological and clinicopathologic insights into molecular subtypes. *Eur Urol*. 2015;68:555–67.
11. Al Hussein Al Awamlh B, Shoag JE. Genomics and risk stratification in high-risk prostate cancer. *Nat Rev Urol*. 2019;16:641–2.
12. Fraser M, Rouette A. Prostate cancer genomic subtypes. *Adv Exp Med Biol*. 2019;1210:87–110.
13. Markert EK, Mizuno H, Vazquez A, Levine AJ. Molecular classification of prostate cancer using curated expression signatures. *Proc Natl Acad Sci USA*. 2011;108:21276–81.
14. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. *Nature*. 2000;406:747–52.
15. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455:1061–8.
16. Howard LE, Zhang J, Fishbane N, Hoedt AM, Klaassen Z, Spratt DE, et al. Validation of a genomic classifier for prediction of metastasis and prostate cancer-specific mortality in African-American men following radical prostatectomy in an equal access healthcare setting. *Prostate Cancer Prostatic Dis*. 2019;23:419–28.
17. Erho N, Crisan A, Vergara IA, Mitra AP, Ghadessi M, Buerki C, et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS ONE*. 2013;8:e66855.
18. Luca BA, Brewer DS, Edwards DR, Edwards S, Whitaker HC, Merson S, et al. DESNT: a poor prognosis category of human prostate cancer. *Eur Urol Focus*. 2018;4:842–50.
19. You S, Knudsen BS, Erho N, Alshalalfa M, Takhar M, Al-Deen Ashab H, et al. Integrated classification of prostate cancer reveals

- a novel luminal subtype with poor outcome. *Cancer Res.* 2016;76:4948–58.
20. Zhao SG, Chang SL, Erho N, Yu M, Lehrer J, Alshalalfa M, et al. Associations of luminal and basal subtyping of prostate cancer with prognosis and response to androgen deprivation therapy. *JAMA Oncol.* 2017;3:1663–72.
 21. You S, Freeman MR. A systems approach to prostate cancer classification-response. *Cancer Res.* 2017;77:7133–5.
 22. Jan YJ, Yoon J, Chen JF, Teng PC, Yao N, Cheng S, et al. A circulating tumor cell-RNA assay for assessment of androgen receptor signaling inhibitor sensitivity in metastatic castration-resistant prostate cancer. *Theranostics.* 2019;9:2812–26.
 23. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol.* 2009;27:1160–7.
 24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA.* 2005;102:15545–50.
 25. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1:417–25.
 26. He Y, Johnson DT, Yang JS, Wu H, You S, Yoon J, et al. Loss of the tumor suppressor, Tp53, enhances the androgen receptor-mediated oncogenic transformation and tumor development in the mouse prostate. *Oncogene.* 2019;38:6507–20.
 27. Ross AE, Feng FY, Ghadessi M, Erho N, Crisan A, Buerki C, et al. A genomic classifier predicting metastatic disease progression in men with biochemical recurrence after prostatectomy. *Prostate Cancer Prostatic Dis.* 2014;17:64–69.
 28. Feng F, Thomas S, Gormley M, Lopez-Gitlitz A, Yu MK, Cheng S, et al. Identifying molecular determinants of response to apalutamide (APA) in patients (pts) with nonmetastatic castration-resistant prostate cancer (nmCRPC) in the SPARTAN study. *Cancer Res* 2019;79:13_Supplement.
 29. Kelly SP, Anderson WF, Rosenberg PS, Cook MB. Past, current, and future incidence rates and burden of metastatic prostate cancer in the United States. *Eur Urol Focus.* 2018;4:121–7.
 30. Weiner AB, Matulewicz RS, Eggener SE, Schaeffer EM. Increasing incidence of metastatic prostate cancer in the United States (2004–2013). *Prostate Cancer Prostatic Dis.* 2016;19:395–7.
 31. Cascianelli S, Molineris I, Isella C, Masseroli M, Medico E. Machine learning for RNA sequencing-based intrinsic subtyping of breast cancer. *Sci Rep.* 2020;10:14071.
 32. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med.* 2000;343:78–85.
 33. Boysen G, Barbieri CE, Prandi D, Blattner M, Chae SS, Dahija A, et al. SPOP mutation leads to genomic instability in prostate cancer. *Elife.* 2015;4:e09207.
 34. Ross-Adams H, Lamb AD, Dunning MJ, Halim S, Lindberg J, Massie CM, et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: a discovery and validation cohort study. *EBioMedicine.* 2015;2:1133–44.



Review

On the Road to Accurate Protein Biomarkers in Prostate Cancer Diagnosis and Prognosis: Current Status and Future Advances

Yiwu Yan ¹, Su Yeon Yeon ¹ , Chen Qian ¹, Sungyong You ^{1,2,3} and Wei Yang ^{1,2,3,4,*}

¹ Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA; yiwu.yan@cshs.org (Y.Y.); suyeon.yeon@cshs.org (S.Y.Y.); chen.qian@cshs.org (C.Q.); sungyong.you@cshs.org (S.Y.)

² Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

³ Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA 90048, USA

⁴ Department of Medicine, University of California Los Angeles, Los Angeles, CA 90095, USA

* Correspondence: wei.yang@cshs.org; Tel.: +1-310-423-7142

Abstract: Prostate cancer (PC) is a leading cause of morbidity and mortality among men worldwide. Molecular biomarkers work in conjunction with existing clinicopathologic tools to help physicians decide who to biopsy, re-biopsy, treat, or re-treat. The past decade has witnessed the commercialization of multiple PC protein biomarkers with improved performance, remarkable progress in proteomic technologies for global discovery and targeted validation of novel protein biomarkers from clinical specimens, and the emergence of novel, promising PC protein biomarkers. In this review, we summarize these advances and discuss the challenges and potential solutions for identifying and validating clinically useful protein biomarkers in PC diagnosis and prognosis. The identification of multi-protein biomarkers with high sensitivity and specificity, as well as their integration with clinicopathologic parameters, imaging, and other molecular biomarkers, bodes well for optimal personalized management of PC patients.

Keywords: prostate cancer; biomarker; proteomics; diagnosis; prognosis



Citation: Yan, Y.; Yeon, S.Y.; Qian, C.; You, S.; Yang, W. On the Road to Accurate Protein Biomarkers in Prostate Cancer Diagnosis and Prognosis: Current Status and Future Advances. *Int. J. Mol. Sci.* **2021**, *22*, 13537. <https://doi.org/10.3390/ijms222413537>

Academic Editor: Pierre Tennstedt

Received: 29 November 2021

Accepted: 14 December 2021

Published: 17 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Prostate cancer (PC) is a leading cause of morbidity and mortality in men, particularly in developed countries, resulting in enormous social and economic costs. PC is the second most commonly diagnosed non-skin cancer and the fifth most lethal cancer in males worldwide. It was estimated that around 1.4 million men will be diagnosed with PC in 2020, with approximately 375,000 PC patients dying from this disease [1]. The incidence and mortality rates of PC are strongly associated with age—the median ages at diagnosis and death are 67 and 80, respectively (<https://seer.cancer.gov/statfacts/html/prost.html> (accessed on 20 November 2021)). Global aging is expected to result in around 16% of the global population being over 65 by 2050, up from 9% in 2019 [2]. As a result, the prevalence of PC and its economic cost are expected to skyrocket in the coming years.

Currently, the cornerstones of PC management include serum prostate-specific antigen (PSA) quantification, digital rectal examination (DRE), and systemic transrectal ultrasound (TRUS)-guided biopsies. PC is frequently discovered before symptoms emerge due to the widespread use of PSA and DRE screening. If the screening results are abnormal, approximately 12 needle core biopsies (small pieces of tissue) are collected from various locations of the prostate for histological analysis. To diagnose PC, a pathologist examines the collected biopsy tissue under a microscope for aberrant histological alterations.

PC is a remarkably heterogeneous disease that can range from indolent to very aggressive [3]. It can be divided into a number of intermediate clinical states, each of which may benefit from a different therapeutic modality. As a result, if a pathologist detects cancer, the next step is to determine the PC's aggressiveness so that the patient can receive optimal care. Gleason grading, developed by Donald F. Gleason in 1966 [4], is widely

regarded as the best predictor of prognosis in localized PC. The Gleason grading system categorizes PC tumors into five Gleason patterns, ranging from well differentiated (grade 1) to poorly differentiated (grade 5) [3]. Because PC tumors often contain cancer cells of varying grades, each PC tumor is assigned two Gleason grades (the most prominent and the second most prominent), the sum of which is reported as a Gleason score (GS) ranging from 2 to 10. However, in contemporary PC diagnosis, GSs of 2–5 are rare, so they are almost exclusively 6–10. To reflect the current diagnostic and therapeutic approaches, the GS system was modified to categorize PCs into five distinct Grade Groups (GGs) [5]. These are GG1 through GG5, which correspond to GSs of ≤ 6 , 7 (3 + 4), 7 (4 + 3), 8, and 9–10, respectively [3].

Despite its utility, the Gleason grading does not accurately predict disease outcomes for individual patients and is subject to inter-observer variability. To further improve risk stratification, various tools were developed by combining GS with other clinicopathologic parameters, such as PSA, tumor-node-metastasis (TNM) classification, age, and percentage of positive biopsies [3]. Among these risk assessment tools, the most widely used ones include the Partin table, the D'Amico risk group, the Cancer of the Prostate Risk Assessment (CAPRA) score, and the Kattan nomogram [3]. However, the performance of these tools in predicting PC aggressiveness is suboptimal. For example, some PCs classified as intermediate-risk tumors are actually high risk. Thus, additional information is required to better assess the risk of a given patient.

2. Clinical Needs of Molecular Biomarkers in PC Diagnosis and Prognosis

Molecular biomarkers supplement existing clinicopathologic tools for PC diagnosis and prognosis by providing additional and valuable information about the biological behavior of PC tumors. To improve the management of PC patients, a number of molecular biomarkers have been developed to address the following questions (Figure 1).

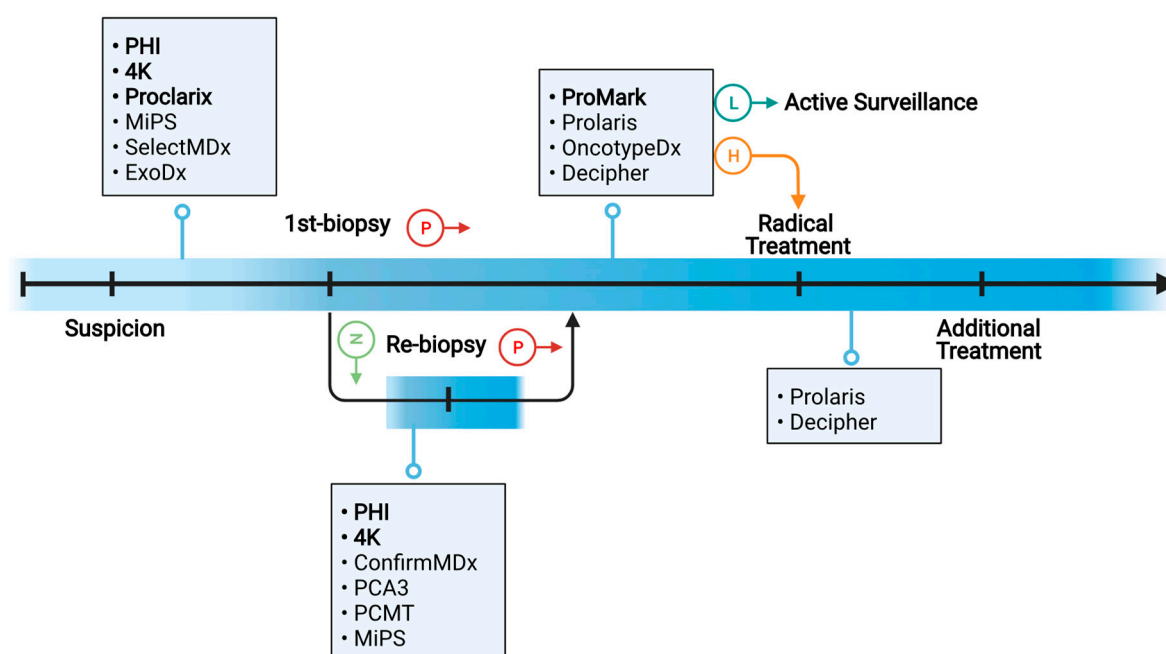


Figure 1. Schematic overview of the clinical needs for molecular biomarkers in various settings. Boxes present commercially available PC biomarkers, among which protein biomarkers are in bold font.

First, who should be biopsied? When PC tumors are diagnosed early, they can be treated with surgery and radiotherapy, which can be curative. However, only about 30% of prostate biopsy procedures result in PC diagnosis [6–8]. Furthermore, approximately 40% of positive biopsies detect clinically insignificant (GS ≤ 6 , i.e., GG = 1) PC [7], which is rarely fatal if untreated. Prostate biopsies are not without risks; they can cause patient

anxiety and discomfort, as well as potential side effects such as pain, fever, blood in the urine, and infections [9]. Therefore, biomarkers are needed to detect clinically significant ($GS \geq 7$, i.e., $GG \geq 2$) PC while minimizing unnecessary and invasive prostate biopsies. Commercially available molecular biomarkers for this purpose include Prostate Health Index (PHI), four-kallikrein (4 K) score, Proclarix, Mi Prostate Score (MiPS), SelectMDX, and ExoDx [10,11].

Second, after a negative initial biopsy, who should be re-biopsied? Because PC is a multifocal disease, and only a small proportion (~1%) of the prostate is sampled, the standard biopsy strategy is prone to sampling error, missing 25–35% of all PC and 10–20% of clinically significant PC [12]. Therefore, if the initial systemic biopsy result is negative but clinical suspicion of PC persists (e.g., continued elevation of serum PSA levels), patients are recommended to be re-biopsied. Total serum PSA levels and serum PSA kinetics, as currently used, are ineffective indicators of PC and frequently result in unnecessary repeat biopsies [13]. To improve prediction accuracy, newer molecular biomarkers have been developed. Among these, commercially available biomarkers include PHI, 4 Kscore, ConfirmMDx, PCA3, Prostate Core Mitomix Test (PCMT), and MiPS [10,11].

Third, after being diagnosed with PC, should a patient receive definitive treatment or be monitored by active surveillance (AS)? In the hope of a cure, approximately 87% of patients with newly diagnosed PC elect for definitive treatments such as radical prostatectomy (RP) and radiotherapy (RT) [8]. However, such treatments can cause significant complications such as urinary, bowel, and sexual dysfunction, lowering the quality of life of PC patients. In comparison, AS allows PC patients to avoid the significant side effects of PC treatment until their disease progresses to the point where treatment is required (if at all). Thus, AS is increasingly being used on patients with low-risk or favorable-intermediate-risk PC [14]. To distinguish PC patients who can be safely monitored by AS from those who require definitive treatment, accurate risk stratification is the key. Because prostate biopsies only sample a small portion of the prostate gland, clinicopathological risk assessment on biopsy specimens is inherently flawed. According to a large-scale study involving 10,273 patients, 44% of low-risk cases were upgraded and 9.7% were up-staged at RP [15]. On the other hand, 18–62% downgrading could occur between needle biopsy and RP [16]. For higher-resolution risk stratification, a number of molecular biomarkers including ProMark, Prolaris, Oncotype DX, and Decipher have been developed and commercialized [10,11,17]. These tests assess underlying biology from biopsy tissue and thus partially address the issues of tumor heterogeneity and biopsy under-sampling.

Fourth, who needs additional treatment following radical PC treatment? A large-scale study found that at 10 years after RP, the cumulative incidence of biochemical recurrence (BCR) and metastasis was 13% and 6%, respectively, among 3089 men with intermediate- or high-risk PC defined by the National Comprehensive Cancer Network (NCCN) [18]. Nonetheless, risk stratification based on clinicopathologic risk factors is inadequate and should be improved. The aforementioned biomarkers such as Decipher and Prolaris have been used in the post-RP setting to improve the prediction of PC recurrence risk [19].

This review will focus on protein biomarkers in PC diagnosis and prognosis because of the following reasons. First, molecular biomarkers (particularly genomic biomarkers) in PC diagnosis and prognosis have already been extensively reviewed [10,11,20–22], yet few review articles specifically focus on PC protein biomarkers. Second, proteins are major functional molecules in cells and the primary determinants of most phenotypic traits. Hence, protein biomarkers have a high clinical potential, particularly for routine monitoring, because their expression can reflect disease activity in real time. Last, the past few years have witnessed remarkable progress in global or targeted protein quantification, allowing for the discovery and validation of novel protein biomarkers with clinical relevance. Because the Special Issue focuses on biomarkers for diagnosis and prognosis of urological tumors, we will not discuss protein biomarkers for treatment response or resistance. Here, we will first review Food and Drug Administration (FDA)-approved and commercially available protein biomarkers for PC diagnosis and prognosis. We will then

discuss the biological sources for biomarker discovery and their pros and cons. Next, we will describe proteomic approaches for the global discovery and targeted validation of novel PC protein biomarkers. Last, we will discuss the challenges and potential solutions for identifying clinically useful PC protein biomarkers.

3. FDA-Approved and Commercially Available Protein Biomarkers for PC Diagnosis and Prognosis

PSA, a kallikrein-like serine protease glycoprotein encoded by the *KLK3* gene, was approved by the FDA in 1986 as a prognostic biomarker in PC and then in 1994 as a diagnostic tool for PC detection among asymptomatic men. It is the best validated and most widely used molecular PC biomarker employed by clinicians. However, PSA is prostate-specific but not PC-specific. It could be elevated not only in PC but also in non-cancerous conditions such as benign prostatic hyperplasia (BPH) and prostatitis. As a result, the positive predictive value (PPV) for a PSA level of >4.0 ng/mL is only about 30% [23]. In other words, slightly less than one in three men with an elevated PSA will have PC detected on biopsy. Furthermore, PSA has a low specificity for clinically significant ($GG \geq 2$) PC, resulting in the over-diagnosis of clinically insignificant ($GG = 1$) PC. Consequently, a large number of men undergo unnecessary prostate biopsies. Additionally, PSA screening can produce false-negative results. A cutoff of 4.0 ng/mL is estimated to miss about 15% of PC cases, including 2.3% of clinically significant PC cases [24]. Due to the limitations of PSA, novel biomarkers for detecting clinically significant PC have been developed. Among these are several commercially available protein biomarkers, including serum-based PHI, 4 KScore, and Proclarix, as well as biopsy tissue-based ProMark.

PHI (Beckman Coulter) combines total PSA (tPSA), free PSA (fPSA), and $[-2]$ proPSA (p2PSA) into a single score to predict the likelihood of PC on biopsy. tPSA comprises free (unbound) PSA and bound (predominantly to α -1-antichymotrypsin) PSA. fPSA is PSA unbound to any carrier molecules or proteins. p2PSA is a truncated form of proPSA (the inactive precursor of PSA) that contains two pro-leader amino acids. In a 2011 multicenter study of 892 patients, PHI demonstrated an AUC of 0.70 for PC in men with 2–10 ng/mL PSA and normal DRE, outperforming p2PSA (0.56), fPSA (0.62), and tPSA (0.53) [25]. PHI was premarket approved by the FDA in 2012 (No. P090026) for PC diagnosis in men over 50, with 4–10 ng/mL PSA and a negative DRE.

The 4 Kscore (OPKO Health) test measures the protein levels of four kallikreins—namely, tPSA, fPSA, intact PSA, and human kallikrein 2—and then combines these parameters with clinical information on age, DRE outcome, and prior negative biopsy status into an algorithm. It is reported as a percentage likelihood of harboring $GG \geq 2$ PC (0–100%). In a large prospective validation study of 1012 men, 4 Kscore had an AUC of 0.82 for predicting $GG \geq 2$ PC [26].

Proclarix (Proteomedix) combines tPSA and fPSA with cancer-related glycoproteins thrombospondin-1 (THBS1) and cathepsin D (CTSD) as well as age to calculate a patient's probability of having clinically significant PC on biopsy [27]. It is intended for use in men with a prostate volume ≥ 35 mL, no history of PC, a normal DRE result, and elevated serum PSA levels (2–10 ng/mL). At a sensitivity of 90%, the Proclarix test has a specificity of 42%, a negative predictive value (NPV) of 95%, and a PPV of 25% [27].

ProMark (MetaMark Genetics) measures the expression levels of eight proteins that are involved in cell signaling, stress response, and cell proliferation [28]. These proteins are SMAD4, PDSS2, HSPA9, FIS, YBOX1, DERL1, and CUL2. ProMark is a personalized prognostic test to distinguish patients with early-stage PC ($GS = 3 + 3$ and $3 + 4$) for AS from those who require RP. Using a quantitative multiplex immunofluorescence platform, the ProMark test measures the expression levels of the eight proteins in formalin-fixed biopsy tissue specimens. Subsequently, ProMark reports a score ranging from 0 to 1 that reflects the probability of detecting adverse prostate pathology in the same patient's RP specimen [29]. For ProMark scores of > 0.8 or 0.9 , the predictive value for unfavorable pathological characteristics after RP can be as high as 76.9% and 100%, respectively [29]. According to the latest NCCN guidelines (version 1.2022) [17], ProMark is recommended

for men with very-low-risk or low-risk PC on biopsy, a life expectancy of at least 10 years, and no prior PC treatment.

4. Clinical Biospecimens for the Discovery and Validation of PC Protein Biomarkers

Despite their clinical utility, the aforementioned FDA-approved and commercially available PC protein biomarkers lack the specificity and sensitivity needed to confidently adjust the course of PC biopsy and treatment. Thus, novel PC protein biomarkers with improved diagnostic and prognostic values are urgently needed. Common clinical biospecimens for the discovery and validation of such biomarkers include patient tissue, expressed prostatic secretions (EPS), EPS-urine, serum/plasma, and extracellular vesicles (EVs) isolated from biofluids. The benefits and drawbacks of these clinical specimens will be discussed further below. We will not discuss preclinical models such as PC cell lines, patient-derived xenografts (PDXs), organoids, and transgenic or genetically engineered mouse models. Using these preclinical models can be a good starting point for directing biomarker discovery towards PC-relevant pathways and increasing confidence in existing biomarker candidates. Nonetheless, they fall short of fully capturing the vast heterogeneity of human PC.

Tissue is frequently used in the discovery of PC protein biomarkers because it closely reflects tumor biology. Compared with biofluid specimens, tissue specimens allow for more direct sampling of proteomic changes in tumor cells and the microenvironment. Nevertheless, prostate tissue biopsy is invasive and only provides a limited snapshot of the tumor. Furthermore, prostate biopsy under-samples multi-focal prostate tumors, making a comprehensive view of PC tumors in individual patients difficult.

Prostate-proximal fluids include direct-EPS and EPS-urine. Direct-EPS is a prostatic fluid collected from patients undergoing RP by massaging the organ and expelling 0.5–1 mL of fluid immediately prior to surgical removal. It is anticipated to contain a relatively high concentration and purity of prostate-secreted proteins. Despite this, EPS has limited clinical utility and cannot be obtained longitudinally because it is collected just prior to RP. EPS-urine, also called post-DRE urine, is first-catch urine collected after a DRE. EPS-urine contains a small fraction of EPS that is expelled during the DRE and collected by the urine. It is easy to obtain and can be collected longitudinally.

Serum/plasma-based biomarkers are particularly appealing in the context of PC because they can be routinely measured pre-, post-, or on-treatment and assayed alongside PSA. Serum/plasma-based liquid biopsies are more likely to fully capture diverse information that reflects intratumoral, microenvironmental, and systemic conditions than tissue biopsies, which only sample ~1% of the prostate. Furthermore, serum/plasma biomarker analysis can be performed in a fast and high-throughput manner, which is critical in a clinical setting to determine a more appropriate PC management plan in a timely fashion. Nonetheless, serum/plasma is currently not very suitable for biomarker discovery using mass spectrometry (MS). This is because tumor-derived proteins are present at low concentrations in the circulation, making them difficult to detect using MS-based shotgun proteomics. Furthermore, serum/plasma proteins have a dynamic range of 10–12 orders of magnitude (from <5 pg/mL to ~50 mg/mL), whereas MS only covers a dynamic range of 4–5 orders of magnitude [30]. As a result, most low-abundance serum/plasma proteins cannot be reliably detected and quantified by MS. Immunodepletion can be used before MS analysis to remove high-abundance proteins, making low-abundance proteins easier to detect and quantify. However, immunodepletion has some limitations such as variable depletion efficiencies for high-abundance proteins, concomitant loss of non-targeted proteins, and decreased sample preparation throughput [31].

EVs are phospholipid bilayer membrane-coated vesicles released by most cell types in physiological and pathological conditions [32]. Of note, EVs are more abundant in biofluid samples from PC patients than from control subjects [33]. EVs are highly heterogeneous in size, biogenesis, function, content, and membrane markers [34]. Because tumor-derived EVs carry tumor-specific cargos and are released into various human body

fluids, EVs represent an attractive source of cancer biomarkers. Because EV contents are well protected within a lipid membrane, they are stable in circulation. To isolate EVs from biological fluids, a variety of techniques and commercial products have been developed. Common isolation methods include (1) ultracentrifugation, (2) precipitation, (3) ultrafiltration, (4) size-exclusion chromatography, (5) affinity interactions, and (6) microfluidic devices and microchips [35]. However, standardized methods for EV isolation remain to be established so that they can be transferred across research or clinical labs.

5. Proteomic Approaches for Global Discovery of Novel Protein Biomarkers from Clinical PC Specimens

5.1. MS-Based Protein Biomarker Discovery

MS is a biophysical technique that allows for the structural analysis of various biomolecules in the form of gas-phase ions, resulting in their detection and quantification. MS can be broadly classified into bottom-up and top-down MS. In bottom-up MS, proteins are digested by an endoproteinase (e.g., trypsin) into peptides prior to MS analysis. In comparison, top-down MS analyzes intact proteins. Compared with top-down MS, bottom-up MS is more well established and approximately two orders of magnitude more sensitive [36]. At present, bottom-up MS enables the identification and quantification of thousands of different proteins in cultured cells and tissue specimens. Thus, it has served as the workhorse tool for the global discovery of novel protein biomarkers.

In bottom-up (also known as shotgun) proteomics, extracted proteins are digested with MS-grade endoproteinases to obtain peptides prior to liquid chromatography (LC)-tandem mass spectrometry (MS/MS) analysis (Figure 2A). Trypsin and Lys-C are the most commonly used endoproteinases because they yield peptides with positively charged C-termini, which are amenable to ionization and thus detectable by MS. The sample complexity can be reduced further if peptides are fractionated using LC methods such as high-pH reversed-phase LC or strong cation exchange LC. Peptide fractionation improves proteomic coverage (i.e., number of detected proteins), but it requires substantially longer instrument time and more input material (typically >40 ug of protein).

In shotgun proteomics, two independent MS scan modes are commonly used: data-dependent acquisition (DDA) and data-independent acquisition (DIA) (Figure 2B,C). DDA-MS was first reported in the 1990s [37], and it has since become the standard for global proteomics [38]. DIA-MS was first described in the early 2000s [39,40]. However, it did not gain widespread acceptance until the introduction of sequential windowed acquisition of all theoretical fragment ion spectra (SWATH) in 2012 [41]. Later, other DIA-MS methods were developed, including multiplexed MS/MS [42] and data-independent acquisition parallel accumulation-serial fragmentation (diaPASEF) [43]. In DDA-MS, MS/MS scans are acquired with narrow isolation windows, e.g., two units of mass/charge (m/z) ratio, centered on peptide precursor ions with the highest intensities (typically top 10–20) in an MS scan. In DIA-MS, MS/MS scans are acquired with wide isolation windows (e.g., 25 units of m/z ratio) that do not target any particular peptide precursor ions [44]. Because DDA-MS stochastically selects precursor ions, whereas DIA-MS consistently collects fragment ion spectra for all detected precursor ions, the latter provides better reproducibility than the former [41]. Nonetheless, owing to continuous improvements in the speed and sensitivity of mass spectrometers, as well as the development of novel algorithms that enable global targeting of thousands of peptides (e.g., MaxQuant.Live [45]), the stochastic nature of DDA-MS is becoming less problematic. Moreover, DDA-MS can be coupled with stable isotope labeling to achieve highly multiplexed protein quantification [46]. For instance, TMTpro enables simultaneous quantification of proteins in up to 16 different samples in a single LC-MS/MS run [47,48]. In comparison, DIA-MS is predominantly label-free and samples are analyzed individually. Although several multiplexed DIA-MS methods, such as NeuCoDIA [49] and MdFDIA [50], have been developed, their utility in cancer biomarker discovery remains to be proven.

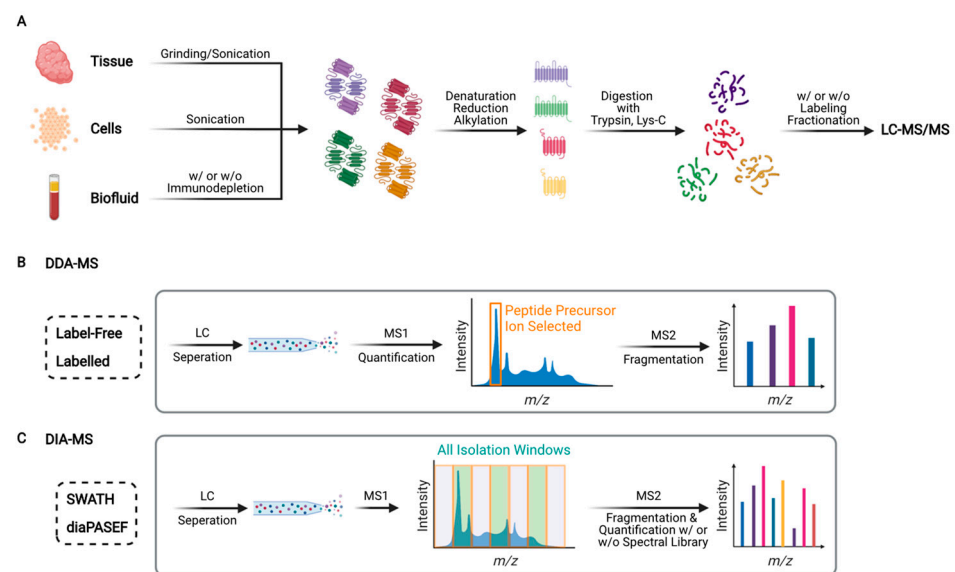


Figure 2. Schematic overview of bottom-up proteomics. **(A)** Typical workflow for global proteomic analysis. Proteins are extracted from tissues, cells, or biofluids, subsequently digested into peptides by an endoproteinase (e.g., trypsin or Lys-C), followed by label-free or isotope labeling-based quantitative proteomic analysis using liquid chromatography-tandem mass spectrometry (LC-MS/MS). **(B)** Schematic of data-dependent acquisition (DDA)-mass spectrometry (MS). Peptides are separated by reversed-phase LC and converted into positively charged gas-phase precursor ions, whose mass/charge (m/z) ratios are measured by MS. Peptide precursor ions with the highest intensities are isolated and broken into product (fragment) ions by MS/MS (also called MS2). **(C)** Schematic of data-independent acquisition (DIA)-MS. Representative methods include sequential window acquisition of all theoretic mass spectra (SWATH) and data-independent acquisition parallel accumulation-serial fragmentation (diaPASEF). Unlike DDA-MS, DIA-MS acquires MS/MS scans with wide isolation windows (e.g., 25 m/z) that do not target any particular peptide precursor ions.

Notably, a major advantage of MS-based proteomics is that it can analyze not only protein abundance but also various post-translational modifications (PTMs) and protein-protein interactions (PPIs) on the proteome scale [38]. Widespread PTMs include protein phosphorylation, glycosylation, and palmitoylation. Drake et al. applied DDA-MS-based label-free tyrosine phosphoproteomics to compare 35 prostate tissue specimens, including 18 metastatic castration-resistant PC (mCRPC) and 12 primary PC samples [51]. The study identified 297 phosphotyrosine peptides and revealed intra-patient similarity and inter-patient heterogeneity of activated tyrosine kinases such as SRC, EGFR, RET, and ALK. Later, the same group performed global phosphoproteomics analysis of 34 tissue specimens including 16 mCRPC samples and 11 treatment-naïve localized PC samples [52]. A total of 8051 phosphopeptides were identified, and six major signaling pathways were found to be enriched in mCRPC specimens when compared with other samples. Liu et al. conducted N-glycoproteomic analysis of PC tissue specimens, including 10 normal prostate, 24 non-aggressive PC, 16 aggressive PC, and 25 metastatic PC, by SWATH-MS [53]. They discovered and validated that two glycoproteins encoded by *NAAA* and *PTK7* were significantly associated with PC aggressiveness. Dong et al. performed a global N-glycoproteomic analysis of 74 aggressive PC and 68 non-aggressive PC urine samples [54]. The study showed that a three-protein signature including urinary ACP, urinary CLU, and serum PSA provides an AUC of 0.86 in distinguishing aggressive PC from non-aggressive PC. Protein palmitoylation (more accurately known as S-acylation) is a widespread PTM that plays critical roles in reversibly regulating protein localization, activity, stability, and complex formation [55,56]. For global analysis of palmitoylated proteins, we developed Palmitoyl-Protein Identification and Site Characterization (PalmPISC) [57] and low-background Acyl-biotinyl Exchange (LB-ABE) [58]. Using these palmitoyl-proteomics approaches, we

profiled the palmitoyl-proteomes in platelets [59] and EVs shed by PC cells [60], identifying the substrate proteins of DHHC3 in PC and breast cancer [61]. We are currently applying palmitoyl-proteomics to analyze total plasma samples and plasma EVs from PC patients, which holds great potential for uncovering novel PC protein biomarkers. Nonetheless, global analysis of protein PTMs to identify novel protein biomarkers remains limited, mainly due to the requirement for a relatively large amount of proteins (typically 500–1000 µg per sample) and a longer time for sample processing.

PPIs are frequently rewired in diseases including cancer [62,63]. Thus, aberrant PPIs are promising and likely more specific disease biomarkers than abnormal protein expression. We are the first to identify protein complexes associated with PC progression in clinical tissue specimens on the proteome scale [64]. In this proof-of-concept study, we coupled tandem mass tagging (TMT)-synchronous precursor selection (SPS)-MS/MS/MS (MS3) with differential expression and co-regulation analyses to compare the differences between protein complexes in PC-adjacent normal prostate, low-grade PC, and high-grade PC tissue specimens ($n = 9$ in each group). Our study identified 28 differentially assembled protein complexes in low-grade PC versus normal prostate, 22 differentially assembled protein complexes in high-grade PC versus normal prostate, and 22 differentially assembled protein complexes in high-grade PC versus low-grade PC. Further exploitation of these deregulated protein complexes is anticipated to reveal novel PPI biomarkers for aggressive PC.

Table 1 summarizes candidate protein biomarkers that have been identified by MS-based discovery proteomics. Well-known biomarkers such as KLK3/PSA, FOLH1/PSMA, and TMPRSS2 were repeatedly identified. Interestingly, TGM4 was the second most frequently identified after KLK3/PSA. TGM4 is upregulated in PC patients with higher Gleason scores and higher PSA levels, and its protein levels correlate with tumor recurrence [65].

Table 1. List of candidate PC protein markers identified by MS-based discovery proteomics.

Potential Biomarker	Sample Cohort	Source	Method	Ref.
NAAA, PTK7	10 normal, 24 non-aggressive PC, 16 aggressive PC, 25 metastatic PC	Tissue (OCT)	DIA-MS (label-free N-glycoproteomics)	[53]
MSK2, CPT2, COPA, NPY	28 PC, 8 PC-adjacent normal	Tissue (FFPE)	DDA-MS (super-SILAC)	[66]
PGM3, PYCR1, GAA, HNRNPM, TALDO1, HNRNPL, GGCT, CTSH, NPEPP5, USP5, SUCLG2, HEXB, NDRG1, STEAP4, DDAH2, CTSD, COPA, TSTA3, PSMB5, TUFM, HSP90B1	14 PC, 9 matched non-malignant	Tissue (fresh frozen)	DDA-MS (label-free)	[67]
RPL28, RBM4, RPL5, NCL, ATP5H, THRAP3, H1FX, SNRPA1, RPL23, PPIB, TPD52, HNRNPL, HNRNPUL1, RALY, RPL10A, APEH, GOT1, USP14, RAB3D, DCXR, DPT, PPL, QDPR, SOD3, OLFML3, EPHX2, EMILIN1, FMOD, GDF15	4 GS3 + 3 PC, 4 GS4 + 4 PC	Tissue (frozen)	DDA-MS (label-free)	[68]
ATR, MRE11, RAD21, RAD23A, RAD23B, RAD50, RAD9A, CHEK1, XRCC5, XRCC6	12 BPH, 18 PC	Tissue (OCT)	DIA-MS (label-free)	[69]
ACO2, CS, FH, IDH3A, MDH2, OGDH, SUCLA2, SUCLG1	10 BPH, 17 untreated PC, 11 CRPC	Tissue (fresh frozen)	DIA-MS (label-free)	[70]
NDRG3, PARP1, ABHD11, SSH3	5 PC w/o metastasis, 5 PC w/lymph node metastases, 5 lymph node metastases	Tissue (FFPE)	DDA-MS (label-free)	[71]
IGKV3D-20, RNASET2, TACC2, ANXA7, LMOD1, PRCP, GYG1, NDUFV1, H1FX, APOBEC3C, CTSZ	5 BPH, 50 PC	Tissue (fresh)	DDA-MS (label-free)	[72]

Table 1. Cont.

Potential Biomarker	Sample Cohort	Source	Method	Ref.
TGM2, NDRG3, KLK3/PSA, AKT1, PTEN, NKX3-1, KRAS, ATM	76 PC	Tissue (OCT ¹)	DDA-MS (label-free)	[73]
CARS2, NFKB2, ENPP4, PDSS2 (high-grade vs. low-grade); YBX1, SETSIP, FASN, PYCR1, PDSS2, FOLH1, SPON2 (high grade vs. normal); NSUN2, HEXB, HEXA, EPCAM, PYCR1 (low grade vs. normal)	9 adjacent normal, 9 low-grade PC, 9 high-grade PC	Tissue (OCT)	DDA-MS (TMT)	[64]
SRM, NOLC1, PTGIS	10 non-malignant, 8 PC, 2 metastatic	Tissue (frozen)	DDA-MS (TMT)	[74]
FASN, TPP1, SPON2	9 BPH, 8PC	Tissue	DIA-MS (label-free)	[75]
ALB, ACTG2, FLNA, MYH11, DES, TAGLN, COL6A3, HBB, ACTB, HIST1H2AH	5 BPH, 17 PC	Tissue (fresh frozen)	DDA-MS (label-free)	[76]
SFN, MME, PARK7, TIMP1, TGM4	8 extracapsular, 8 organ-confined	Direct-EPS	DDA-MS (label-free)	[77]
KLK3/PSA, PAP, MSMB, FOLH1/PSMA, TMPRSS2	6 BPH, 5PC	EPS-urine	DDA-MS (label-free)	[78]
ACPP, ATRN, GP2, KLK11, PTPRN2, NPTN, CPE, RNASE2 (low in aggressive PC). CD97, ORM1, AFM, UMOD, PTGDS, GRN, SERPINA1, CLU, LRG1, LOX, DSC2 (high in aggressive PC)	74 aggressive PC, 68 non-aggressive PC	EPS-urine	DIA-MS (label-free N-glycoproteomics)	[54]
KLK3/PSA, ACPP, TGM4, FOLH1/PSMA	12 noncancer, 12 PC	EPS urinary EV	DDA-MS (label-free)	[79]
SCIN, AMBP, FABP5, CHMP4C, CHMP2B, BAIAP2, GRN, SYTL2, CALR, CHMP4A, DNPH1	11 negative biopsy, 18 PC including 5 GS6, 7 GS 7, and 6 GS 8–9	EPS urinary EV	DDA-MS (iTRAQ ²)	[80]
KLK2, KLK3/PSA, FOLH1/PSMA, MSMB, ACPP, TGM4, NDRG1, NKX3-1, FKBP5, FAM129A, RAB27A, FASN, NEFH	12 BPH, 12 PC	EPS urinary EV	DDA-MS (label-free)	[81]
B2M, PGA3, MUC3	83 BPH, 90 PC	Urine	DDA-MS (iTRAQ)	[82]
TM256/C17orf61, LAMPTOR1, VATL, ADIRF, KLK3/PSA, FOLH1/PSMA, TGM4, TMPRSS2, GOLPH3	15 noncancer, 17 PC	Urinary EV	DDA-MS (label-free)	[83]
C1QB, APOA4, CO9, ANT3, VTDB, PLMN, GPX3, ITIH4, CFAI, APOH, VTNC, IBP3, CLUS, APOA2, PEDE, TETN, CD14, LG3BP, CFAH, FCN3, HPT, CO3, APOA1, APOC3, SAMP, HEMO, CO6, KLK3/PSA, A2MG, A1At, APOE, A2GI, TTHY, C1S, ZAG, AMBP, KNG1, CO4A, AACT, CAV1, TRFE	3 PC with BCR, 3 control	Immunodepleted serum	DDA-MS (label-free)	[84]

¹ Optimal cutting temperature. ² Isobaric tag for relative and absolute quantification.

Despite its many strengths, MS-based discovery proteomics has some drawbacks. First, it is biased toward high-abundance proteins and provides less robust detection and quantification of low-abundance proteins. Second, analyzing complex biological matrices such as serum and plasma remains a formidable challenge. Despite progress [85,86], it remains difficult to detect proteins below low $\mu\text{g/mL}$ or high ng/mL levels without extensive sample fractionation or protein enrichment. Of note, serum/plasma protein biomarkers with clinical applicability are often present in the pg/mL to sub ng/mL range [30]. Third, MS instruments are expensive, and their operation requires specialized skills, limiting widespread adoption in clinical labs. These limitations can be addressed, at least in part, by immunoassay- or aptamer-based discovery proteomics as described below.

5.2. Immunoassay-Based Protein Biomarker Discovery

5.2.1. Antibody Array

In an antibody array, a large collection of distinct capture antibodies is immobilized onto a solid support surface (Figure 3A). For an assay, each antibody array is incubated with one test sample, where tens to hundreds of different proteins and phosphoproteins

are measured simultaneously. For protein quantification, proteins in samples are labeled by one or more fluorescent dyes. The dye labeling can be direct via chemical conjugation. Alternatively, proteins can be biotinylated so that they can be probed with fluorescently labeled streptavidin [87]. Miller et al. used antibody arrays containing 184 different antibodies to analyze serum samples from 33 PC patients and 20 controls, leading to the identification of five differentially expressed proteins: von Willebrand factor, immunoglobulin M, α 1-antichymotrypsin, villin, and immunoglobulin G [88]. Shafer et al. used antibody arrays containing 102 antibodies to analyze serum samples from healthy controls, organ-confined PC, non-organ-confined PC, and benign prostatic conditions ($n = 92$ in each group), resulting in the discovery of some differentially expressed proteins including thrombospondin 1 [89]. However, antibody arrays may suffer from the limitations of batch-to-batch variability, antibody stability, and high cost. As a result, antibody arrays have not yet been frequently used to identify novel PC biomarkers.

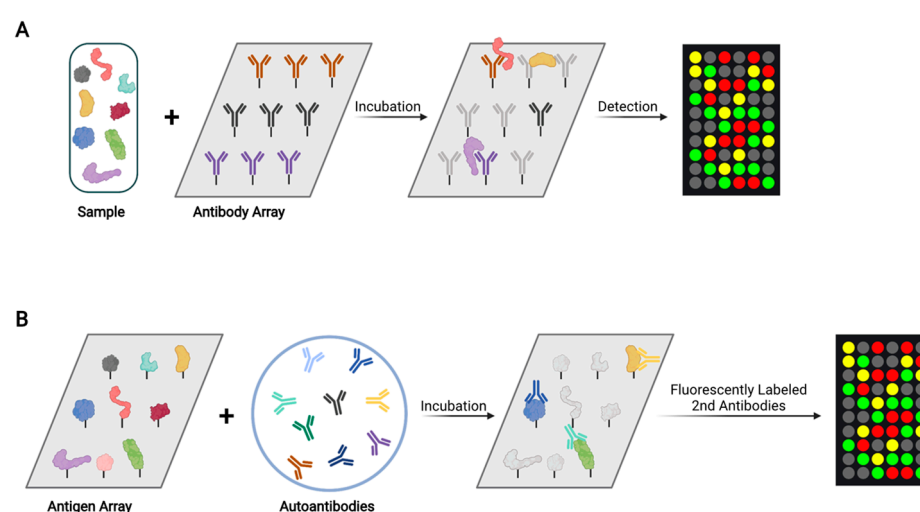


Figure 3. Schematic overview of antibody array and antigen array analyses. **(A)** In an antibody array, each spot contains one type of antibody and each array is incubated with one test sample. For protein quantification, proteins are fluorescently labeled (either directly or indirectly) and incubated with an antibody array. **(B)** In an antigen array, each spot contains one purified protein and each array is incubated with one test sample. For protein quantification, fluorescently labeled secondary antibodies are incubated with an antigen array.

5.2.2. Antigen Array

Tumor-associated autoantibodies are promising diagnostic and prognostic biomarkers [90]. They are easily accessible in blood specimens, have a long half-life, and may be significantly more abundant than tumor antigens due to antibody amplification response [90]. Each array contains purified proteins spotted onto nitrocellulose filters that are adhered to glass slides. For analysis, arrays are incubated with patient samples, and tumor-associated autoantibodies in the samples bind to their cognate antigens on the array (Figure 3B). After washing, the arrays are incubated with fluorescently labeled secondary antibodies. The fluorescence intensities of bound autoantibodies are used to quantify them. Using an array containing 123 tumor-associated antigens, Adeola et al. measured autoantibodies in serum samples from 20 PC patients, 32 BPH patients, and 15 controls [91]. They identified 41 candidate PC biomarkers, including GAGE1, ROPN1, SPANXA1, and PRKCZ.

5.2.3. Proximity Extension Assay (PEA)

The PEA technology (Olink Proteomics) was developed based on proximity-dependent DNA ligation (Figure 4) [92]. In PEA, either two matched monoclonal antibodies, or one batch of polyclonal antibody split into two fractions, are covalently linked with two different 40-mer oligonucleotides at the 3'- and 5'-end, respectively. To the 3'-linked probe,

a 56-mer DNA oligo that consists of 40 nt complementary to that probe, 7 nt spacer, and 9 nt complementary to the corresponding 5'-linked probe is hybridized. The hybridized proximity probe pair is then incubated with a sample that contains the antigens of interest, resulting in antigen-proximity probe pair binding. As a result, the oligonucleotides come into contact with one another and hybridize. The hybridizing oligo is then extended over the other probe arm using a DNA polymerase. The resulting DNA template can be detected and quantified by quantitative polymerase chain reaction (qPCR) or next-generation sequencing (NGS).

Using the Olink Immuno-oncology and Oncology II panel, Liu et al. measured the abundance of 92 target proteins in serum samples from men without PC, patients with low-risk primary PC, patients with high-risk primary PC, and patients with metastatic PC ($n = 20$ per group) [93]. Nine proteins (i.e., PTN, MK, PVRL4, EPHA2, TFPI-2, hK11, SYND1, ANGPT2, and hK14) were found to be elevated in metastatic PC patients, compared with other groups. In another study, the same group applied PEA to measure the protein levels of 184 target proteins in pre- and post-operative serum samples from ten patients with high-grade and high-volume PC [94]. Six proteins were found to be significantly reduced after RP: CASP8, MSLN, FGFBP1, ICOSLG, TIE2, and S100A4.

Currently, the Olink Explore 1536/384 allows for the quantification of 1463 distinct proteins (<https://www.olink.com/products-services/explore/> (accessed on 25 November 2021)). An expanded version of the platform, Olink Explore 3072, allows for the quantification of ~3000 unique proteins and is available for pre-order. It is expected that Olink will be increasingly applied to identify protein biomarkers for PC diagnosis and prognosis.

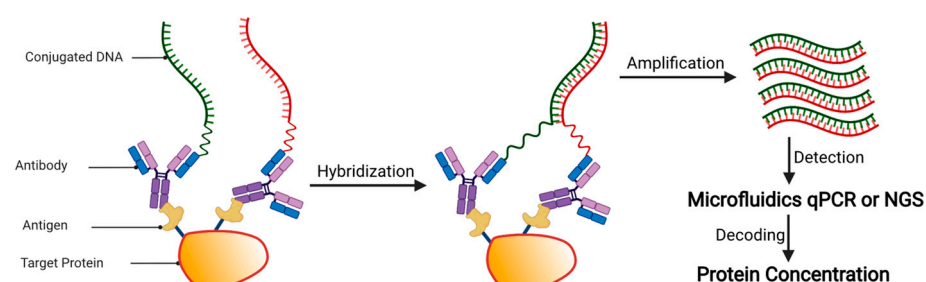


Figure 4. Schematic overview of proximity extension assay (PEA) analysis. Upon sample incubation, the antibody-based proximity probe pair binds to its specific antigens on the same protein. As a result, the pair of probes come in close proximity and hybridize. The addition of a DNA polymerase causes the hybridizing oligo to be extended, resulting in a DNA template that can be detected and quantified by quantitative PCR (qPCR) or next-generation sequencing (NGS).

5.3. Aptamer-Based Protein Biomarker Discovery

Aptamers, also known as chemical antibodies, are short oligonucleotide sequences that have a high affinity for binding to a protein target. Aptamers have several advantages as follows [95]. First, they are not immunogenic, making them suitable for clinical use. Second, due to their chemical synthesis and modification, aptamers can be more easily, quickly, and economically produced than antibodies, resulting in little inter-batch variability. Third, aptamers are highly stable in harsh environments, ensuring longer shelf life and easier storage and transport. Fourth, because of their nanometric size, aptamers can interact with a wide range of targets, ranging from inorganic molecules to whole cells, as well as being able to penetrate tissues and internalize into cells. Fifth, aptamers can be screened without prior knowledge of the target molecules, allowing the discovery of previously unknown biomarkers.

The slow off-rate modified aptamer (SOMAmer) assay (SOMAscan) is a multiplex proteomic platform (Figure 5). In SOMAscan, each SOMAmer is modified with a biotin group, a photocleavable group, and a fluorescent tag. After incubating SOMAMers with the sample, any formed SOMAmer–protein complexes are captured by streptavidin beads via streptavidin–biotin interaction. The captured proteins are then biotinylated, and the

complexes are released from the beads via photocleavage and washing. Another set of streptavidin beads is then added to the mixture to recapture the SOMAmer–protein complexes via the biotinylated proteins. SOMAmers are eluted using specific pH conditions and hybridized to complementary DNA sequences on a proprietary microarray chip. The concentrations of SOMAmers, which are proportional to target protein concentrations, are quantified by fluorescence intensity.

To identify biomarkers indicative of treatment failure, Welton et al. applied SOMAscan to analyze EVs isolated from 11 plasma samples and 5 urine samples [96]. However, no proteins were found to be significantly different between the treatment-naïve and treatment-resistant groups. In another study, Dudani et al. used SOMAscan to compare five PC samples and five matched normal adjacent tissue samples [97]. Proteases such as uPA and PRSS3 were found to be more abundant in the PC samples than in the normal adjacent tissue samples.

The SOMAscan workflow is highly automated, allowing for high-throughput analysis of 7000 proteins in 680 samples in a single day (<https://somalologic.com/life-sciences/> (accessed on 25 November 2021)). It was reported that the median inter-assay coefficient of variation (CV) was ~5% [98]. SOMAscan is expected to be used more frequently in discovering novel PC biomarkers, especially those in body fluids. However, it should be noted that about 7% of SOMAmers have cross-reactivity to another protein [99].

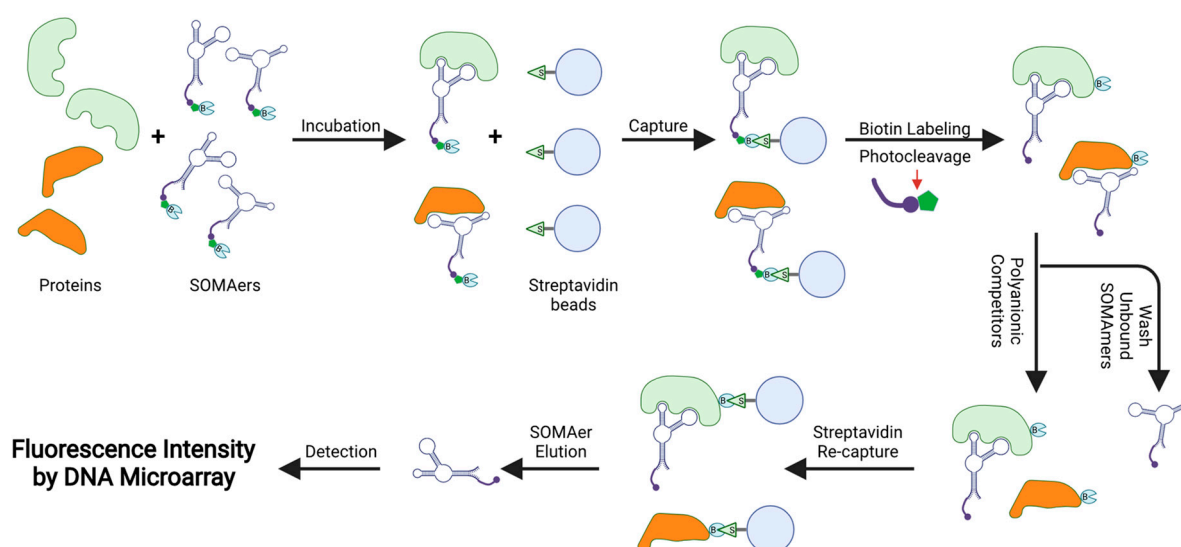


Figure 5. Schematic overview of the SOMAscan analysis. Each SOMAmer contains a biotin (B) group, a photo-cleavable link, and a fluorescent tag at the 5' end. SOMAmers are mixed with the test sample, forming SOMAmer–protein complexes. The complexes are captured on streptavidin beads via strong biotin–streptavidin interaction. The captured proteins are then biotinylated and the SOMAmer–protein complexes are released from beads using ultraviolet light. Polyanionic competitors are added to promote the dissociation between proteins and non-specific SOMAmers. The SOMAmer–protein complexes are recaptured on new streptavidin beads. Protein-bound SOMAmers are eluted, hybridized to custom arrays of SOMAmer-complementary oligonucleotides, and quantified by fluorescence intensities, which are proportional to the concentrations of their cognate target proteins.

6. Computational Approaches for Prioritizing Protein Biomarker Candidates

A single biomarker can only provide a limited diagnostic or prognostic value. It is widely accepted that a panel of multiple biomarkers is more clinically useful than a single molecular biomarker. One key challenge is determining the best combination of individual biomarkers from massive omics data sets. For this, three feature selection methods are available: filter-based, wrapper-based, and embedded (Figure 6) [100]. Filter-based feature selection is computationally fast and simple, yet interaction with the classifier is ignored. Commonly used methods include fold change, ANOVA, Student's *t*-test, and

Mann–Whitney–Wilcoxon test (Figure 6A). Wrapper-based feature selection looks for the best subset of features based on their predictive power, but it is compute-intensive. Methods in this category include sequential forward selection, sequential backward elimination, and genetic algorithm (Figure 6B). Embedded feature selection may require a dataset to be randomly divided into training and testing sets to avoid model overfitting or underfitting. Methods for this include decision tree, support vector machine, and random forest (Figure 6C), all of which have been applied to identify optimal combinations of PC protein biomarkers [101–103].

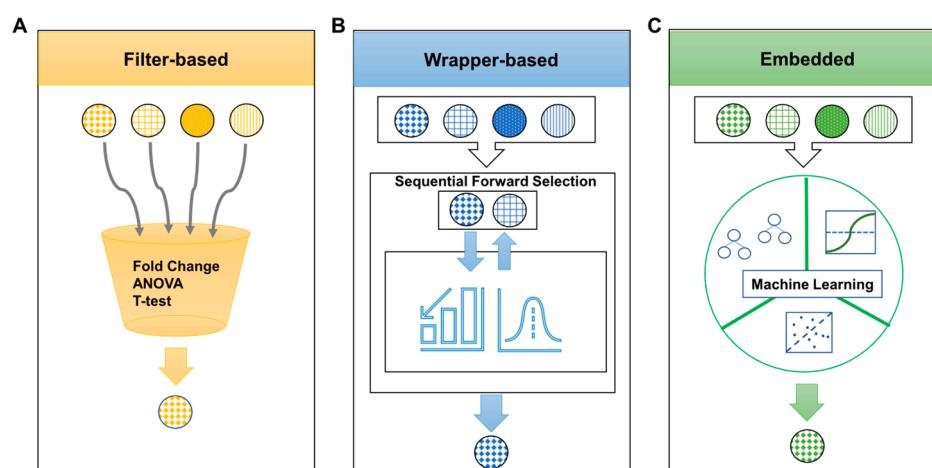


Figure 6. Schematic of three different feature selection methods for determining optimal protein biomarkers. (A) Filter-based methods are based on choosing the differential feature according to discriminating metrics such as p -value. Metrics are calculated from a statistical method such as fold change, ANOVA, and Student’s t -test. This method ranks proteins according to the selected criteria that put highly redundant or differentially expressed proteins on the top rank. (B) Wrapper-based methods look for the best subset of features based on their predictive power. Generation of a feature subset and assessment function is repeated until the optimal subset is returned through the learning algorithm. The feature subset with the highest performance is returned as a result. Sequential forward selection is one of the examples of this method and uses a bottom-up search technique to find the best subset. (C) Embedded methods use various machine learning techniques to select the optimal subset of features. Random forest, support vector machine and artificial neural network are examples of embedded feature selection methods.

7. Proteomic Approaches for Multiplexed Targeted Validation of Candidate PC Biomarker Proteins

Multiplex protein measurements reduce time, cost, and sample volume. Currently, commonly used multiplexed targeted proteomics methods include antibody-independent targeted MS as well as antibody-dependent reverse phase protein array (RPPA) and Luminex. These methods are complementary and should be integrated to achieve optimal results.

7.1. MS-Based Multiplexed Targeted Proteomics

Currently, the most widely used MS-based multiplexed targeted proteomics approaches are selected reaction monitoring (SRM) [104], also called multiple reaction monitoring (MRM), and parallel reaction monitoring (PRM) [105,106] (Figure 7A,B). SRM and PRM assays are typically carried out in triple quadrupole (QqQ) (e.g., QTRAP series) and quadrupole-orbitrap (Q-OT) (e.g., Q Exactive series) mass spectrometers, respectively. In SRM, a predefined series of transitions (i.e., precursor–product ion pairs) is monitored over time for precise quantification of each target peptide. In PRM, targeted MS/MS is applied to simultaneously monitor all product ions of a targeted peptide with high resolution and mass accuracy. Unlike SRM, the selection of the best transitions in PRM can be defined in a post-acquisition step. SRM and PRM have similar linearity, dynamic range, and precision,

with the latter requiring less method development than the former [107,108]. Notably, once validated on an instrument, SRM or PRM assays can be transferred across sites and clinical laboratories. Table 2 summarizes SRM- or PRM-based targeted proteomics validation studies of candidate PC protein biomarkers.

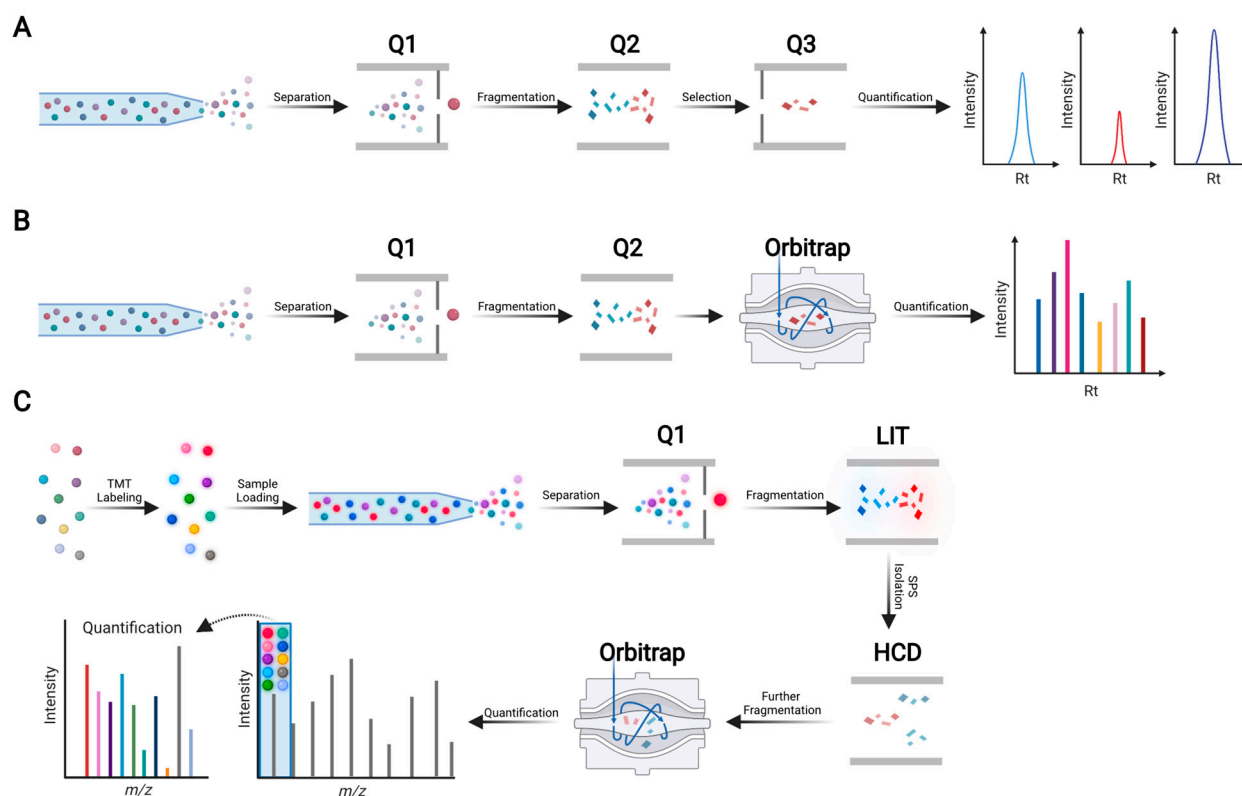


Figure 7. Schematic overview of MS-based targeted proteomics methods. **(A)** Schematic of selected reaction monitoring (SRM), also known as multiple reaction monitoring (MRM). For peptide quantification, three to five selected fragment ions from a single peptide precursor ion are measured sequentially. SRM is typically performed on a triple quadrupole (QqQ) mass spectrometer. The first quadrupole (Q1) isolates a predefined peptide precursor ion, the second quadrupole (Q2) is a collision cell where isolated precursor ions are broken into product ions (also called fragment ions), and the third quadrupole (Q3) isolates predefined product ions. Such predefined pairs of precursor and product ions are called transitions, which provide high specificity and sensitivity to quantify peptides that are surrogates of proteins of interest. **(B)** Schematic of parallel monitoring reaction (PRM). PRM employs a high-resolution Orbitrap mass analyzer to simultaneously monitor many product ions. Because transitions do not need to be defined in advance, PRM is easier to set up than SRM. **(C)** Schematic of TOMAHAQ (triggers by offset, multiplexed, accurate mass, high-resolution, absolute quantification). Peptides derived from 10 (or 16) samples are labeled with 10-plex (or 16-plex) tandem mass tag (TMT) reagents, which consist of 10 (or 16) different isobaric compounds with the same mass and chemical structure. Subsequently, an equal amount of differentially TMT-labeled peptides is pooled into one tube, followed by LC separation and targeted MS analysis. Rt: retention time; LIT: linear ion trap; HCD: higher-energy collisional dissociation.

Table 2. List of MS-based targeted proteomics validation studies of candidate PC protein markers.

Potential Biomarker	Sample Cohort	Source	Method	Ref.
FASN, TPP1, SPON2	16 BPH, 57 PC	Tissue	PRM-MS analysis of 6 peptides corresponding to 3 target proteins	[75]
ADSV, TGM4, CD63, GLPK5, SPHMPSA, PAPP	54 noncancer, 22 low-grade PC, 31 high-grade PC	EPS urinary EV	SRM-MS analysis of 64 peptides corresponding to 64 target proteins	[109]
C1QB, APOA4, CO9, ANT3, VTDB, PLMN, GPX3, ITIH4, CFAI, APOH, VTNC, IBP3, CLUS, APOA2, PEDE, TETN, CD14, LG3BP, CFAH, FCN3, HPT, CO3, APOA1, APOC3, SAMP, HEMO, CO6, KLK3/PSA, A2MG, A1At, APOE, A2GI, TTHY, C1S, ZAG, AMBP, KNG1, CO4A, AACT, CAV1, TRFE	86 time-point samples from 3 PC patients with BCR and 3 controls	Immunodepleted serum	SRM-MS analysis of 59 peptides corresponding to 41 target proteins	[84]
ITIH2, CD44, IGHG2, CDH13	25 aggressive PC, 25 non-aggressive PC	Serum	PRM-MS analysis of 41 N-glycosite-containing peptides corresponding to 37 target proteins	[110]

SRM and PRM are both label-free targeted proteomics methods, and samples must be analyzed separately (1–2 h per sample). In comparison, TOMAHAQ (triggered by offset, multiplexed, accurate mass, high-resolution, absolute quantification) combines sample multiplexing with targeted proteomics to significantly increase throughput (Figure 7C) [111]. For instance, using TOMAHAQ, 131 different peptides were quantified across 180 cell lysate samples in only 48 h [111]. Nevertheless, TOMAHAQ can only be implemented on expensive tribrid mass spectrometers such as Orbitrap Eclipse and Orbitrap Fusion Lumos, limiting its widespread use in research and clinical applications.

7.2. Antibody-Based Multiplexed Targeted Proteomics

Compared with MS, antibody-based immunoassays provide higher throughput and sensitivity. For example, the enzyme-linked immunosorbent assay (ELISA) is one of the most widely used tools for protein quantification in research settings, as well as the gold standard in clinical laboratories for detecting single analytes. Notably, ELISAs can achieve 1–10 pg/mL detection limits without sample pretreatment [112]. Nevertheless, ELISA is not well suitable for multiplexed targeted proteomics because it requires a relatively high sample volume (as high as 50 µL sample per analyte) for analysis as well as high cost (USD 100,000—USD 1,000,000 per assay) and long lead time (1–2 years) for assay development [113]. To enable multiplexed protein detection and quantification, several immunoassay technologies, such as RPPA and microsphere bead capture, have been developed and commercialized [114].

7.2.1. RPPA

RPPAs are widely used for high-throughput, multiplexed, and quantitative analysis of target proteins and phosphoproteins in tissue lysates, cultured cell lines, and, to a lesser extent, biological fluids (Figure 8A) [115]. Each array can contain hundreds of different test samples, one for each spot. Of note, owing to the high sensitivity of RPPA, only a very small amount of protein (equivalent to ~200 cells) is required for each test sample. Each array is also printed with control samples containing varying amounts of protein, allowing for the generation of a calibration curve for protein quantification. In RPPA analysis, each array is probed with one single antibody that can be detected using fluorescent, colorimetric, or

chemiluminescent assays. The analytical sensitivity of RPPA has been reported to range from picogram to femtogram levels, with a CV of <15% [116].

The first time RPPA was used was in the study of PC [117]. In this study, PC progression was found to be significantly associated with increased AKT1 phosphorylation, decreased ERK phosphorylation, and suppression of apoptosis pathways. In another study, RPPA was used to examine signaling pathways in normal, tumor, and stromal cells isolated from PC tissue specimens using laser capture microdissection (LCM) [118]. AKT1 and GSK3 β phosphorylation levels were higher in tumor cells than in normal cells, whereas ERK, p38, and PKC α phosphorylation levels were lower. RPPA has also been applied to analyze tumor cells isolated by LCM from treatment-naïve localized PC, hormone-refractory localized PC, and metastatic PC tissue specimens [119]. The study found that ERBB2 and BCL-2 phosphorylation levels were higher in metastatic PC than in primary PC, whereas ERK, p38, and JNK phosphorylation levels were lower. In yet another study, RPPAs were probed with 190 validated antibodies to analyze 152 primary PC samples [120]. The study identified three clusters with (1) high apoptosis and DNA damage response pathway activities, (2) a high EMT pathway score, and (3) increased PI3K–AKT, MAPK, and RTK activities, respectively. To explore the molecular architecture of the tumor microenvironment in human PC, RPPAs were probed with 124 antibodies to analyze epithelial and stromal cells isolated by LCM from 18 PC patients [121]. The study identified a protein network activated in the malignant PC tumor microenvironment. In the so-far largest RPPA analysis of PC specimens, RPPAs were probed with 225 validated antibodies to analyze 351 primary PC specimens and 7312 patient samples from 30 other cancer types [122]. The study showed that the level of PI3K–AKT–mTOR pathway activity in primary PC was average among the 31 evaluated cancer types. More recently, RPPA was used to measure key antigens and activated signaling in EVs isolated from PC patients' sera [123]. The study showed that PD-L1, ERG, Integrin- β 5, Survivin, TGF- β , phosphorylated TSC2, and partners of MAPK and mTOR pathways are differentially expressed in tumor-derived EVs.

Taken together, these studies demonstrate that RPPA is a valuable tool for targeted proteomics analysis of PC specimens, owing to its high throughput, sensitivity, cost-effectiveness, and quick turnaround time. Therefore, RPPAs are particularly useful in the clinical setting. It should be noted, however, that RPPA is antibody-dependent, necessitating extensive antibody and assay validation.

7.2.2. Microsphere Bead Capture (Luminex)

Figure 8B shows a schematic of microsphere bead capture analysis. The analyte-specific capture antibodies are immobilized on 6.5 μ m superparamagnetic microsphere beads that are color-coded. For multiplexed targeted proteomics, a mixture of antibody-coated beads is used to capture target proteins. Subsequently, biotinylated detection antibodies specific to the target proteins are added to form an antibody–antigen sandwich. Phycoerythrin (PE)-conjugated streptavidin is added to bind to the biotinylated detection antibodies. Beads are read on a dual-laser flow-based detection instrument: one laser classifies the beads, and the other laser determines the magnitude of the PE-derived signal that is proportional to the target protein bound to the beads. Experiments can be performed in 96- or 384-well microtiter plates, allowing for high throughput. The limit of detection is about 1–10 pg/mL, and the dynamic range is about 3–4 orders of magnitude [124]. In practice, up to about 30 target proteins can be analyzed in each assay [124].

Tsaur et al. used Luminex to measure the concentrations of six cytokines in serum samples from 39 PC patients and 15 healthy donors [125]. They found that CCL2 was significantly more abundant in the serum samples of PC patients compared with controls, suggesting that CCL2 is a potential diagnostic biomarker for PC. Al-Mazidi et al. used Luminex to analyze 27 cytokines in plasma samples from 19 healthy controls, 29 untreated patients with nonmetastatic PC, 20 patients with metastatic PC who received chemotherapy and reported pain, and 10 chemotherapy-treated patients with no pain [126]. They found that the concentrations of IL-6, IL-8, Eotaxin, VEGF, and IP-10 are significantly higher

in the plasma of chemotherapy-treated patients with pain than the other groups. These cytokines are potential targets for pain control in PC patients receiving chemotherapy. Shore et al. applied Luminex to measure the concentrations of 32 cellular growth factors in serum samples from 64 patients with non-aggressive PC and 120 patients with aggressive PC [127]. The concentrations of PSA, IL-7, and VEGF were found to be significantly higher in aggressive PC than non-aggressive PC.

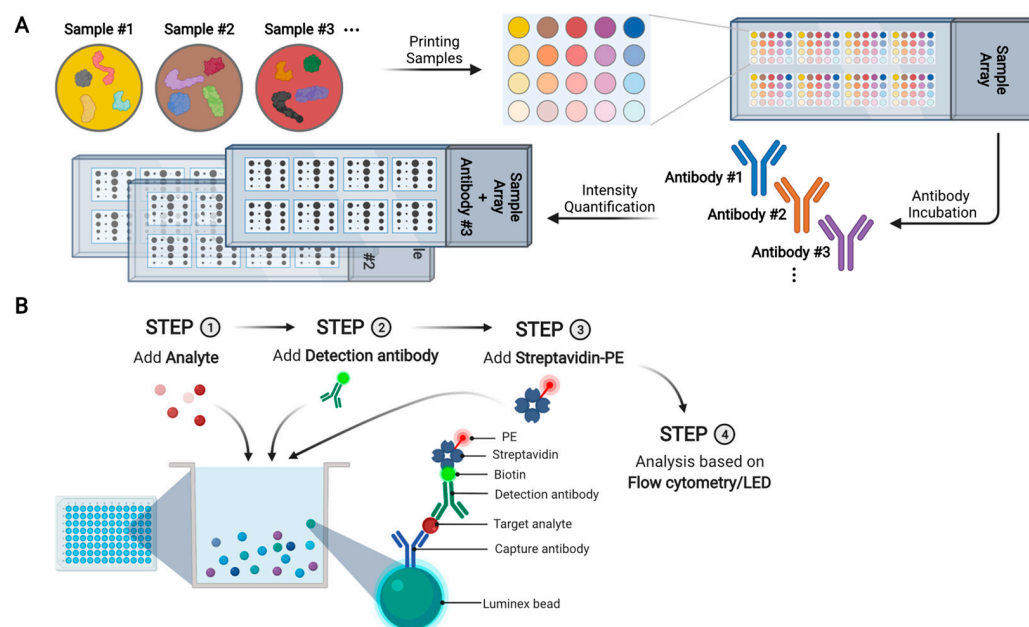


Figure 8. Schematic overview of reverse phase protein array (RPPA) and Luminex microspheres bead capture. (A) Schematic of RPPA. Each spot contains a single sample and each array is probed with one specific antibody. The bound antibody can be quantified (either directly or indirectly) by fluorescent, colorimetric, or chemiluminescent assays. (B) Schematic of Luminex microspheres bead capture assay. Analyte-specific capture antibodies are immobilized on superparamagnetic microsphere beads that are color-coded. After incubating a test sample with antibody-coated microsphere beads, target proteins are captured. Biotinylated detection antibodies specific to the target proteins are added, leading to the formation of an antibody–antigen sandwich. Phycoerythrin (PE)-conjugated streptavidin is added, so that the protein amount can be quantified based on the intensities of PE-derived signal.

8. Challenges and Potential Solutions in Identifying and Validating Clinically Valuable PC Protein Biomarkers

In the past decade, a large number of candidate protein biomarkers for PC diagnosis and prognosis have been identified. Nonetheless, very few have been approved for clinical use, and their spread in the clinical routine is very slow. The reasons are multi-faceted, and we offer potential solutions as follows. First, most studies stop at the discovery phase and fail to proceed to biomarker verification and validation, largely due to a lack of financial support, patient specimens, or both. A small number of studies proceeded to biomarker verification, yet few have used a completely independent cohort of samples to rule out false positives caused by sample collection and processing. To address this issue, collaborative efforts involving multiple institutions and, more optimally, multiple nations are required for large-scale validation studies of prioritized biomarker candidates. Second, most discovery cohorts are small in size, and few studies share the same biomarker candidates due to a lack of standardized methods for sample collection and processing, data acquisition, and bioinformatics analysis. Given the wealth of available data on biomarker candidates, a meta-analysis would help prioritize biomarker candidates for large-scale validation. Third, rather than identifying protein biomarkers that can provide

additional information to the established clinicopathological parameters (e.g., Gleason grading), many studies identified proteins that are proxies for these parameters. As a result, the clinical values of such candidate protein biomarkers may be insignificant. A potential solution is to identify candidate protein biomarkers associated with clinically relevant endpoints, such as the occurrence of metastases, disease-specific mortality, and overall survival. Fourth, most PC protein biomarker candidates were identified using specimens from white men, so their clinical values for patients of other races (e.g., African American) remain unknown. To address this issue, biomarker discovery studies should be conducted in more diverse cohorts of patient specimens. Fifth, PC is a highly heterogeneous disease, so the discovery of protein biomarkers with both high sensitivity and high specificity is inherently difficult. To overcome this issue, a large sample size is required for biomarker discovery, and a relatively large panel of protein biomarkers is very likely required to recapitulate the heterogeneity. Sixth, it remains unclear how the short-term and long-term temporal dynamics of biomarkers affect their validity and clinical utility. Longitudinal studies are needed to address this question. Seventh, despite marked advances in MS instruments and techniques, the proteomics workflows need to be further simplified and standardized, and the costs need to be substantially reduced. Ultimately, identified and validated protein biomarkers should be cost-effective, provide additional information to what PSA already provides, and be easily incorporated into routine workflows in clinical laboratories.

9. Conclusions

The past decade has witnessed dramatic progress in MS-, antibody-, and aptamer-based global or targeted proteomics. Many promising PC protein biomarker candidates have been identified, among which some are undergoing biomarker validation in large cohorts of PC specimens. Although several hurdles must be overcome, we are optimistic that clinically valuable protein biomarkers will be identified, validated, and commercialized in the near future. In combination with other parameters including histology, clinical data, imaging, and other molecular biomarkers, protein biomarkers will guide physicians in deciding optimal personalized management for their patients.

Author Contributions: Writing—original draft preparation, Y.Y., S.Y.Y. and C.Q.; writing—review and editing, S.Y. and W.Y.; supervision, W.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Cancer Institute R01CA218526 (W.Y.) and Department of Defense W81XWH-20-1-0644 (S.Y.).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: BioRender was used to construct Figures 1–5, 7 and 8 after obtaining a paid license for publication authorization. Due to space constraints, it was not possible to include an exhaustive list of references for all of the work discussed. We apologize to the authors whose important contributions could not be adequately described or cited.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Sung, H.; Ferlay, J.; Siegel, R.L.; Laversanne, M.; Soerjomataram, I.; Jemal, A.; Bray, F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J. Clin.* **2021**, *71*, 209–249. [[CrossRef](#)]
2. Department of Economic and Social Affairs, U.N. World Population Prospects 2019: Highlights. *Dept. Econ. Soc. Aff. World Popul. Prospect.* **2019**, 46.
3. Rebello, R.J.; Oing, C.; Knudsen, K.E.; Loeb, S.; Johnson, D.C.; Reiter, R.E.; Gillissen, S.; Van der Kwast, T.; Bristow, R.G. Prostate Cancer. *Nat. Rev. Dis. Prim.* **2021**, *7*, 9. [[CrossRef](#)]

4. Gleason, D.F. Classification of Prostatic Carcinomas. *Cancer Chemother. Rep.* **1966**, *50*, 125–128.
5. Epstein, J.I. Prostate Cancer Grading: A Decade After the 2005 Modified System. *Mod. Pathol.* **2018**, *31*, 47–63. [\[CrossRef\]](#)
6. Welch, H.G.; Fisher, E.S.; Gottlieb, D.J.; Barry, M.J. Detection of Prostate Cancer via Biopsy in the Medicare-SEER Population During the PSA Era. *J. Natl. Cancer Inst.* **2007**, *99*, 1395–1400. [\[CrossRef\]](#)
7. Shah, N.; Ioffe, V.; Huebner, T.; Hristova, I. Prostate Biopsy Characteristics: A Comparison Between the Pre- and Post-2012 United States Preventive Services Task Force (USPSTF) Prostate Cancer Screening Guidelines. *Rev. Urol.* **2018**, *20*, 77–83. [\[CrossRef\]](#)
8. Schröder, F.H.; Hugosson, J.; Roobol, M.J.; Tammela, T.L.J.; Zappa, M.; Nelen, V.; Kwiatkowski, M.; Lujan, M.; Mänttinen, L.; Lilja, H.; et al. Screening and Prostate Cancer Mortality: Results of the European Randomised Study of Screening for Prostate Cancer (ERSPC) at 13 years of Follow-up. *Lancet* **2014**, *384*, 2027–2035. [\[CrossRef\]](#)
9. Loeb, S.; Vellekoop, A.; Ahmed, H.U.; Catto, J.; Emberton, M.; Nam, R.; Rosario, D.J.; Scattoni, V.; Lotan, Y. Systematic Review of Complications of Prostate Biopsy. *Eur. Urol.* **2013**, *64*, 876–892. [\[CrossRef\]](#) [\[PubMed\]](#)
10. Visser, W.; de Jong, H.; Melchers, W.; Mulders, P.; Schalken, J. Commercialized Blood-, Urinary- and Tissue-Based Biomarker Tests for Prostate Cancer Diagnosis and Prognosis. *Cancers* **2020**, *12*, 3790. [\[CrossRef\]](#)
11. Meehan, J.; Gray, M.; Martínez-Pérez, C.; Kay, C.; McLaren, D.; Turnbull, A.K. Tissue- and Liquid-based Biomarkers in Prostate Cancer Precision Medicine. *J. Pers. Med.* **2021**, *11*, 664. [\[CrossRef\]](#)
12. Capitanio, U.; Pfister, D.; Emberton, M. Repeat Prostate Biopsy: Rationale, Indications, and Strategies. *Eur. Urol. Focus* **2015**, *1*, 127–136. [\[CrossRef\]](#)
13. Blute, M.L.; Abel, E.J.; Downs, T.M.; Kelcz, F.; Jarrard, D.F. Addressing the Need for Repeat Prostate Biopsy: New Technology and Approaches. *Nat. Rev. Urol.* **2015**, *12*, 435–444. [\[CrossRef\]](#)
14. Liu, J.L.; Patel, H.D.; Haney, N.M.; Epstein, J.I.; Partin, A.W. Advances in the Selection of Patients with Prostate Cancer for Active Surveillance. *Nat. Rev. Urol.* **2021**, *18*, 197–208. [\[CrossRef\]](#)
15. Dinh, K.T.; Mahal, B.A.; Ziehr, D.R.; Muralidhar, V.; Chen, Y.W.; Viswanathan, V.B.; Nezoslosky, M.D.; Beard, C.J.; Choueiri, T.K.; Martin, N.E.; et al. Incidence and Predictors of Upgrading and Up Staging among 10,000 Contemporary Patients with Low Risk Prostate Cancer. *J. Urol.* **2015**, *194*, 343–349. [\[CrossRef\]](#)
16. Wenzel, M.; Würnschimmel, C.; Chierigo, F.; Flammia, R.S.; Tian, Z.; Shariat, S.F.; Gallucci, M.; Terrone, C.; Saad, F.; Tilki, D.; et al. Nomogram Predicting Downgrading in National Comprehensive Cancer Network High-risk Prostate Cancer Patients Treated with Radical Prostatectomy. *Eur. Urol. Focus* **2021**, *49*, 1–8. [\[CrossRef\]](#)
17. Schaeffer, E.; Srinivas, S.; Antonarakis, E.S.; Armstrong, A.J.; Cheng, H.H.; D’Amico, A.V.; Davis, B.J.; Desai, N.; Dorff, T.; Eastham, J.A.; et al. Prostate Cancer, Version 1.2022, NCCN Clinical Practice Guidelines in Oncology. *Natl. Compr. Cancer Netw.* **2021**.
18. Ross, A.E.; Yousefi, K.; Davicioni, E.; Ghadessi, M.; Johnson, M.H.; Sundi, D.; Tosoian, J.J.; Han, M.; Humphreys, E.B.; Partin, A.W.; et al. Utility of Risk Models in Decision Making after Radical Prostatectomy: Lessons from a Natural History Cohort of Intermediate- and High-Risk Men. *Eur. Urol.* **2016**, *69*, 496–504. [\[CrossRef\]](#)
19. Olleik, G.; Kassouf, W.; Aprikian, A.; Hu, J.; Vanhuyse, M.; Cury, F.; Peacock, S.; Bonnevier, E.; Palenius, E.; Dragomir, A. Evaluation of New Tests and Interventions for Prostate Cancer Management: A Systematic Review. *JNCCN J. Natl. Compr. Cancer Netw.* **2018**, *16*, 1340–1351. [\[CrossRef\]](#)
20. Couñago, F.; López-Campos, F.; Díaz-Gavela, A.A.; Almagro, E.; Fenández-Pascual, E.; Henríquez, I.; Lozano, R.; Espinós, E.L.; Gómez-Iturriaga, A.; de Velasco, G.; et al. Clinical Applications of Molecular Biomarkers in Prostate Cancer. *Cancers* **2020**, *12*, 1550. [\[CrossRef\]](#)
21. Balázs, K.; Antal, L.; Sáfrány, G.; Lumniczky, K. Blood-derived Biomarkers of Diagnosis, Prognosis and Therapy Response in Prostate Cancer Patients. *J. Pers. Med.* **2021**, *11*, 296. [\[CrossRef\]](#)
22. Eyrych, N.W.; Morgan, T.M.; Tosoian, J.J. Biomarkers for Detection of Clinically Significant Prostate Cancer: Contemporary Clinical Data and Future Directions. *Transl. Androl. Urol.* **2021**, *10*, 3091–3103. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Wolf, A.M.D.; Wender, R.C.; Etzioni, R.B.; Thompson, I.M.; D’Amico, A.V.; Volk, R.J.; Brooks, D.D.; Dash, C.; Guessous, I.; Andrews, K.; et al. American Cancer Society Guideline for the Early Detection of Prostate Cancer: Update 2010. *CA Cancer J. Clin.* **2010**, *60*, 70–98. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Thompson, I.M.; Pauler, D.K.; Goodman, P.J.; Tangen, C.M.; Lucia, M.S.; Parnes, H.L.; Minasian, L.M.; Ford, L.G.; Lippman, S.M.; Crawford, E.D.; et al. Prevalence of Prostate Cancer among Men with a Prostate-Specific Antigen Level ≤ 4.0 ng per Milliliter. *N. Engl. J. Med.* **2004**, *350*, 2239–2246. [\[CrossRef\]](#)
25. Catalona, W.J.; Partin, A.W.; Sanda, M.G.; Wei, J.T.; Klee, G.G.; Bangma, C.H.; Slawin, K.M.; Marks, L.S.; Loeb, S.; Broyles, D.L.; et al. A Multicenter Study of [-2]pro-prostate Specific Antigen Combined With Prostate Specific Antigen and Free Prostate Specific Antigen for Prostate Cancer Detection in the 2.0 to 10.0 ng/ml Prostate Specific Antigen Range. *J. Urol.* **2011**, *185*, 1650–1655. [\[CrossRef\]](#)
26. Parekh, D.J.; Punnen, S.; Sjöberg, D.D.; Asroff, S.W.; Bailen, J.L.; Cochran, J.S.; Concepcion, R.; David, R.D.; Deck, K.B.; Dumbadze, I.; et al. A Multi-institutional Prospective Trial in the USA Confirms that the 4Kscore Accurately Identifies Men with High-grade Prostate Cancer. *Eur. Urol.* **2015**, *68*, 464–470. [\[CrossRef\]](#)
27. Klocker, H.; Golding, B.; Weber, S.; Steiner, E.; Tennstedt, P.; Keller, T.; Schiess, R.; Gillissen, S.; Horninger, W.; Steuber, T. Development and Validation of a Novel Multivariate Risk Score to Guide Biopsy Decision for the Diagnosis of Clinically Significant Prostate Cancer. *BJUI Compass* **2020**, *1*, 15–20. [\[CrossRef\]](#)

28. Shipitsin, M.; Small, C.; Choudhury, S.; Giladi, E.; Friedlander, S.; Nardone, J.; Hussain, S.; Hurley, A.D.; Ernst, C.; Huang, Y.E.; et al. Identification of Proteomic Biomarkers Predicting Prostate Cancer Aggressiveness and Lethality Despite Biopsy-sampling Error. *Br. J. Cancer* **2014**, *111*, 1201–1212. [\[CrossRef\]](#)
29. Blume-Jensen, P.; Berman, D.M.; Rimm, D.L.; Shipitsin, M.; Putzi, M.; Nifong, T.P.; Small, C.; Choudhury, S.; Capela, T.; Coupal, L.; et al. Biology of Human Tumors Development and Clinical Validation of an In Situ Biopsy-based Multimarker Assay for Risk Stratification in Prostate Cancer. *Clin. Cancer Res.* **2015**, *21*, 2591–2600. [\[CrossRef\]](#)
30. Anderson, N.L.; Anderson, N.G. The Human Plasma Proteome: History, Character, and Diagnostic Prospects. *Mol. Cell. Proteomics* **2002**, *1*, 845–867. [\[CrossRef\]](#)
31. Patel, B.B.; Barrero, C.A.; Braverman, A.; Kim, P.D.; Jones, K.A.; Chen, D.E.; Bowler, R.P.; Merali, S.; Kelsen, S.G.; Yeung, A.T. Assessment of Two Immunodepletion Methods: Off-Target Effects and Variations in Immunodepletion Efficiency may Confound Plasma Proteomics. *J. Proteome Res.* **2012**, *11*, 5947–5958. [\[CrossRef\]](#)
32. Théry, C.; Witwer, K.W.; Aikawa, E.; Alcaraz, M.J.; Anderson, J.D.; Andriantsitohaina, R.; Antoniou, A.; Arab, T.; Archer, F.; Atkin-Smith, G.K.; et al. Minimal Information for Studies of Extracellular Vesicles 2018 (MISEV2018): A Position Statement of the International Society for Extracellular Vesicles and update of the MISEV2014 Guidelines. *J. Extracell. Vesicles* **2018**, *7*. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Tavoosidana, G.; Ronquist, G.; Darmanis, S.; Yan, J.; Carlsson, L.; Wu, D.; Conze, T.; Ek, P.; Semjonow, A.; Eltze, E.; et al. Multiple Recognition Assay Reveals Prostatomes as Promising Plasma Biomarkers for Prostate Cancer. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 8809–8814. [\[CrossRef\]](#)
34. Van Niel, G.; D'Angelo, G.; Raposo, G. Shedding Light on the Cell Biology of Extracellular Vesicles. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 213–228. [\[CrossRef\]](#)
35. Pang, B.; Zhu, Y.; Ni, J.; Thompson, J.; Malouf, D.; Bucci, J.; Graham, P.; Li, Y. Extracellular Vesicles: The Next Generation of Biomarkers for Liquid Biopsy-based Prostate Cancer Diagnosis. *Theranostics* **2020**, *10*, 2309–2326. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Timp, W.; Timp, G. Beyond Mass Spectrometry, the Next Step in Proteomics. *Sci. Adv.* **2020**, *6*, eaax8978. [\[CrossRef\]](#)
37. Yates, J.R.; Eng, J.K.; McCormack, A.L.; Schieltz, D. Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Database. *Anal. Chem.* **1995**, *67*, 1426–1436. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Aebersold, R.; Mann, M. Mass-spectrometric Exploration of Proteome Structure and Function. *Nature* **2016**, *537*, 347–355. [\[CrossRef\]](#) [\[PubMed\]](#)
39. Purvine, S.; Eppel, J.T.; Yi, E.C.; Goodlett, D.R. Shotgun Collision-induced Dissociation of Peptides Using a Time of Flight Mass Analyzer. *Proteomics* **2003**, *3*, 847–850. [\[CrossRef\]](#) [\[PubMed\]](#)
40. Venable, J.D.; Dong, M.Q.; Wohlschlegel, J.; Dillin, A.; Yates, J.R. Automated Approach for Quantitative Analysis of Complex Peptide Mixtures from Tandem Mass Spectra. *Nat. Methods* **2004**, *1*, 39–45. [\[CrossRef\]](#)
41. Gillet, L.C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteom.* **2012**, *11*, O111.016717. [\[CrossRef\]](#)
42. Egertson, J.D.; Kuehn, A.; Merrihew, G.E.; Bateman, N.W.; MacLean, B.X.; Ting, Y.S.; Canterbury, J.D.; Marsh, D.M.; Kellmann, M.; Zabrouskov, V.; et al. Multiplexed MS/MS for Improved Data-independent Acquisition. *Nat. Methods* **2013**, *10*, 744–746. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Meier, F.; Brunner, A.D.; Frank, M.; Ha, A.; Bludau, I.; Voytik, E.; Kaspar-Schoenefeld, S.; Lubeck, M.; Raether, O.; Bache, N.; et al. diaPASEF: Parallel Accumulation–serial Fragmentation Combined with Data-independent Acquisition. *Nat. Methods* **2020**, *17*, 1229–1236. [\[CrossRef\]](#)
44. Gillet, L.C.; Leitner, A.; Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu. Rev. Anal. Chem.* **2016**, *9*, 1–24. [\[CrossRef\]](#)
45. Wichmann, C.; Meier, F.; Winter, S.V.; Brunner, A.D.; Cox, J.; Mann, M. MaxQuant.live Enables Global Targeting of More Than 25,000 Peptides. *Mol. Cell. Proteom.* **2019**, *18*, 982–994. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Tian, X.; Permentier, H.P.; Bischoff, R. Chemical Isotope Labeling for Quantitative Proteomics. *Mass Spectrom. Rev.* **2021**, 1–31. [\[CrossRef\]](#)
47. Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C.; et al. TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-based Isobaric 16-plex Tandem Mass Tag Reagent Set. *Anal. Chem.* **2019**, *91*, 15941–15950. [\[CrossRef\]](#)
48. Li, J.; Van Vranken, J.G.; Pontano Vaite, L.; Schweppe, D.K.; Huttlin, E.L.; Etienne, C.; Nandhikonda, P.; Viner, R.; Robitaille, A.M.; Thompson, A.H.; et al. TMTpro Reagents: A Set of Isobaric Labeling Mass Tags Enables Simultaneous Proteome-wide Measurements Across 16 Samples. *Nat. Methods* **2020**, *17*, 399–404. [\[CrossRef\]](#)
49. Minogue, C.E.; Hebert, A.S.; Rensvold, J.W.; Westphall, M.S.; Pagliarini, D.J.; Coon, J.J. Multiplexed Quantification for Data-Independent Acquisition. *Anal. Chem.* **2015**, *87*, 2570–2575. [\[CrossRef\]](#) [\[PubMed\]](#)
50. Di, Y.; Zhang, Y.; Zhang, L.; Tao, T.; Lu, H. MdFDIA: A Mass Defect Based Four-Plex Data-Independent Acquisition Strategy for Proteome Quantification. *Anal. Chem.* **2017**, *89*, 10248–10255. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Drake, J.M.; Graham, N.A.; Lee, J.K.; Stoyanova, T.; Faltermeier, C.M.; Sud, S.; Titz, B.; Huang, J.; Pienta, K.J.; Graeber, T.G.; et al. Metastatic Castration-resistant Prostate Cancer Reveals Inpatient Similarity and Interpatient Heterogeneity of Therapeutic Kinase Targets. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, E4762–E4769. [\[CrossRef\]](#)

52. Drake, J.M.; Paull, E.O.; Graham, N.A.; Lee, J.K.; Smith, B.A.; Titz, B.; Stoyanova, T.; Faltermeier, C.M.; Uzunangelov, V.; Carlin, D.E.; et al. Phosphoproteome Integration Reveals Patient-Specific Networks in Prostate Cancer. *Cell* **2016**, *166*, 1041–1054. [\[CrossRef\]](#)
53. Liu, Y.; Chen, J.; Sethi, A.; Li, Q.K.; Chen, L.; Collins, B.; Gillet, L.C.J.; Wollscheid, B.; Zhang, H.; Aebersold, R. Glycoproteomic Analysis of Prostate Cancer Tissues by SWATH Mass Spectrometry Discovers N-acyl ethanolamine Acid Amidase and Protein Tyrosine Kinase 7 as Signatures for Tumor Aggressiveness. *Mol. Cell. Proteom.* **2014**, *13*, 1753–1768. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Dong, M.; Lih, T.M.; Chen, S.-Y.; Cho, K.-C.; Eguev, R.V.; Höti, N.; Zhou, Y.; Yang, W.; Mangold, L.; Chan, D.W.; et al. Urinary Glycoproteins Associated with Aggressive Prostate Cancer. *Theranostics* **2020**, *10*, 11892–11907. [\[CrossRef\]](#)
55. Wang, Y.; Yang, W. Proteome-scale Analysis of Protein S-Acylation Comes of Age. *J. Proteome Res.* **2021**, *20*, 14–26. [\[CrossRef\]](#)
56. Zhou, B.; An, M.; Freeman, M.R.; Yang, W. Technologies and Challenges in Proteomic Analysis of Protein S-acylation. *J. Proteom. Bioinform.* **2014**, *7*, 256–263. [\[CrossRef\]](#)
57. Yang, W.; Di Vizio, D.; Kirchner, M.; Steen, H.; Freeman, M.R. Proteome-scale Characterization of Human S-acylated Proteins in Lipid Raft-enriched and Non-raft Membranes. *Mol. Cell. Proteom.* **2010**, *9*, 54–70. [\[CrossRef\]](#)
58. Zhou, B.; Wang, Y.; Yan, Y.; Mariscal, J.; Di Vizio, D.; Freeman, M.R.; Yang, W. Low-background Acyl-biotinyl Exchange Largely Eliminates the Coisolation of Non-S-acylated Proteins and Enables Deep S-acylproteomic Analysis. *Anal. Chem.* **2019**, *91*, 9858–9866. [\[CrossRef\]](#)
59. Dowal, L.; Yang, W.; Freeman, M.R.; Steen, H.; Flaumenhaft, R. Proteomic Analysis of Palmitoylated Platelet Proteins. *Blood* **2011**, *118*, e62–e73. [\[CrossRef\]](#) [\[PubMed\]](#)
60. Mariscal, J.; Vagner, T.; Kim, M.; Zhou, B.; Chin, A.; Zandian, M.; Freeman, M.R.; You, S.; Zijlstra, A.; Yang, W.; et al. Comprehensive Palmitoyl-Proteomic Analysis Identifies Distinct Protein Signatures for Large and Small Cancer-derived Extracellular Vesicles. *J. Extracell. Vesicles* **2020**, *9*, 1764192. [\[CrossRef\]](#)
61. Sharma, C.; Yang, W.; Steen, H.; Freeman, M.R.; Hemler, M.E. Antioxidant Functions of DHHC3 Suppress Anti-cancer Drug Activities. *Cell. Mol. Life Sci.* **2021**, *78*, 2341–2353. [\[CrossRef\]](#)
62. Cheng, S.S.; Yang, G.J.; Wang, W.; Leung, C.H.; Ma, D.L. The Design and Development of Covalent Protein-protein Interaction Inhibitors for Cancer Treatment. *J. Hematol. Oncol.* **2020**, *13*, 1–14. [\[CrossRef\]](#)
63. Cheng, R.; Jackson, P.K. Identifying Cancer Drivers. *Science* **2021**, *374*, 38–39. [\[CrossRef\]](#)
64. Zhou, B.; Yan, Y.; Wang, Y.; You, S.; Freeman, M.R.; Yang, W. Quantitative Proteomic Analysis of Prostate Tissue Specimens Identifies Deregulated Protein Complexes in Primary Prostate Cancer. *Clin. Proteomics* **2019**, *16*, 1–18. [\[CrossRef\]](#)
65. Ablin, R.J.; Owen, S.; Jiang, W.G. Prostate Transglutaminase (TGase-4) Induces Epithelial-to-Mesenchymal Transition in Prostate Cancer Cells. *Anticancer Res.* **2017**, *37*, 481–487. [\[CrossRef\]](#)
66. Iglesias-Gato, D.; Wikström, P.; Tyanova, S.; Lavalley, C.; Thysell, E.; Carlsson, J.; Hägglöf, C.; Cox, J.; Andrén, O.; Stattin, P.; et al. The Proteome of Primary Prostate Cancer. *Eur. Urol.* **2016**, *69*, 942–952. [\[CrossRef\]](#)
67. Myers, J.S.; von Lersner, A.K.; Sang, Q.X.A. Proteomic Upregulation of Fatty Acid Synthase and Fatty Acid Binding Protein 5 and Identification of Cancer- and Race-specific Pathway Associations in Human Prostate Cancer Tissues. *J. Cancer* **2016**, *7*, 1452–1464. [\[CrossRef\]](#)
68. Staunton, L.; Tonry, C.; Lis, R.; Espina, V.; Liotta, L.; Inzitari, R.; Bowden, M.; Fabre, A.; O’Leary, J.; Finn, S.P.; et al. Pathology-driven Comprehensive Proteomic Profiling of the Prostate Cancer Tumor Microenvironment. *Mol. Cancer Res.* **2017**, *15*, 281–293. [\[CrossRef\]](#) [\[PubMed\]](#)
69. Guo, T.; Li, L.; Zhong, Q.; Rupp, N.J.; Charmpi, K.; Wong, C.E.; Wagner, U.; Rueschoff, J.H.; Jochum, W.; Fankhauser, C.D.; et al. Multi-region Proteome Analysis Quantifies Spatial Heterogeneity of Prostate Tissue Biomarkers. *Life Sci. Alliance* **2018**, *1*, e201800042. [\[CrossRef\]](#)
70. Latonen, L.; Afyounian, E.; Jylhä, A.; Nättinen, J.; Aapola, U.; Annala, M.; Kivinummi, K.K.; Tammela, T.T.L.; Beuerman, R.W.; Uusitalo, H.; et al. Integrative Proteomics in Prostate Cancer Uncovers Robustness Against Genomic and Transcriptomic Aberrations During Disease Progression. *Nat. Commun.* **2018**, *9*, 1176. [\[CrossRef\]](#) [\[PubMed\]](#)
71. Müller, A.-K.; Föll, M.; Heckelmann, B.; Kiefer, S.; Werner, M.; Schilling, O.; Biniossek, M.L.; Jilg, C.A.; Drendel, V. Proteomic Characterization of Prostate Cancer to Distinguish Nonmetastasizing and Metastasizing Primary Tumors and Lymph Node Metastases. *Neoplasia* **2018**, *20*, 140–151. [\[CrossRef\]](#)
72. Kawahara, R.; Recuero, S.; Nogueira, F.C.S.; Domont, G.B.; Leite, K.R.M.; Srougi, M.; Thaysen-Andersen, M.; Palmisano, G. Tissue Proteome Signatures Associated with Five Grades of Prostate Cancer and Benign Prostatic Hyperplasia. *Proteomics* **2019**, *19*. [\[CrossRef\]](#)
73. Sinha, A.; Huang, V.; Livingstone, J.; Wang, J.; Fox, N.S.; Kurganovs, N.; Ignatchenko, V.; Fritsch, K.; Donmez, N.; Heisler, L.E.; et al. The Proteogenomic Landscape of Curable Prostate Cancer. *Cancer Cell* **2019**, *35*, 414–427.e6. [\[CrossRef\]](#) [\[PubMed\]](#)
74. Kwon, O.K.; Ha, Y.S.; Na, A.Y.; Chun, S.Y.; Kwon, T.G.; Lee, J.N.; Lee, S. Identification of Novel Prognosis and Prediction Markers in Advanced Prostate Cancer Tissues Based on Quantitative Proteomics. *Cancer Genom. Proteom.* **2020**, *17*, 195–208. [\[CrossRef\]](#)
75. Zhu, T.; Zhu, Y.; Xuan, Y.; Gao, H.; Cai, X.; Piersma, S.R.; Pham, T.V.; Schelfhorst, T.; Haas, R.R.G.D.; Bijnsdorp, I.V.; et al. DPHL: A DIA Pan-human Protein Mass Spectrometry Library for Robust Biomarker Discovery. *Genom. Proteom. Bioinform.* **2020**, *18*, 104–119. [\[CrossRef\]](#) [\[PubMed\]](#)

76. Latosinska, A.; Davaliev, K.; Makridakis, M.; Mullen, W.; Schanstra, J.P.; Vlahou, A.; Mischak, H.; Frantzi, M. Molecular Changes in Tissue Proteome during Prostate Cancer Development: Proof-of-Principle Investigation. *Diagnostics* **2020**, *10*, 655. [\[CrossRef\]](#) [\[PubMed\]](#)
77. Kim, Y.; Ignatchenko, V.; Yao, C.Q.; Kalatskaya, I.; Nyalwidhe, J.O.; Lance, R.S.; Gramolini, A.O.; Troyer, D.A.; Stein, L.D.; Boutros, P.C.; et al. Identification of Differentially Expressed Proteins in Direct Expressed Prostatic Secretions of Men with Organ-confined Versus Extracapsular Prostate Cancer. *Mol. Cell. Proteom.* **2012**, *11*, 1870–1884. [\[CrossRef\]](#)
78. Principe, S.; Kim, Y.; Fontana, S.; Ignatchenko, V.; Nyalwidhe, J.O.; Lance, R.S.; Troyer, D.A.; Alessandro, R.; Semmes, O.J.; Kislinger, T.; et al. Identification of Prostate-enriched Proteins by In-depth Proteomic Analyses of Expressed Prostatic Secretions in Urine. *J. Proteome Res.* **2012**, *11*, 2386–2396. [\[CrossRef\]](#) [\[PubMed\]](#)
79. Principe, S.; Jones, E.E.; Kim, Y.; Sinha, A.; Nyalwidhe, J.O.; Brooks, J.; Semmes, O.J.; Troyer, D.A.; Lance, R.S.; Kislinger, T.; et al. In-depth Proteomic Analyses of Exosomes Isolated from Expressed Prostatic Secretions in Urine. *Proteomics* **2013**, *13*, 1667–1671. [\[CrossRef\]](#) [\[PubMed\]](#)
80. Fujita, K.; Kume, H.; Matsuzaki, K.; Kawashima, A.; Ujike, T.; Nagahara, A.; Uemura, M.; Miyagawa, Y.; Tomonaga, T.; Nonomura, N. Proteomic Analysis of Urinary Extracellular Vesicles from High Gleason Score Prostate Cancer. *Sci. Rep.* **2017**, *7*, 42961. [\[CrossRef\]](#) [\[PubMed\]](#)
81. Dhondt, B.; Geurickx, E.; Tulkens, J.; Van Deun, J.; Vergauwen, G.; Lippens, L.; Miinalainen, I.; Rappu, P.; Heino, J.; Ost, P.; et al. Unravelling the Proteomic Landscape of Extracellular Vesicles in Prostate Cancer by Density-based Fractionation of Urine. *J. Extracell. Vesicles* **2020**, *9*, 1736935. [\[CrossRef\]](#) [\[PubMed\]](#)
82. Jedinak, A.; Curatolo, A.; Zurakowski, D.; Dillon, S.; Bhasin, M.K.; Libermann, T.A.; Roy, R.; Sachdev, M.; Loughlin, K.R.; Moses, M.A. Novel Non-invasive Biomarkers that Distinguish Between Benign Prostate Hyperplasia and Prostate Cancer. *BMC Cancer* **2015**, *15*, 1–9. [\[CrossRef\]](#)
83. Overbye, A.; Skotland, T.; Koehler, C.J.; Thiede, B.; Seierstad, T.; Berge, V.; Sandvig, K.; Llorente, A. Identification of Prostate Cancer Biomarkers in Urinary Exosomes. *Oncotarget* **2015**, *6*, 30357–30376. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Tonry, C.L.; Doherty, D.; O'Shea, C.; Morrissey, B.; Staunton, L.; Flatley, B.; Shannon, A.; Armstrong, J.; Pennington, S.R. Discovery and Longitudinal Evaluation of Candidate Protein Biomarkers for Disease Recurrence in Prostate Cancer. *J. Proteome Res.* **2015**, *14*, 2769–2783. [\[CrossRef\]](#)
85. Geyer, P.E.; Voytik, E.; Treit, P.V.; Doll, S.; Kleinhempel, A.; Niu, L.; Müller, J.B.; Buchholtz, M.; Bader, J.M.; Teupser, D.; et al. Plasma Proteome Profiling to Detect and Avoid Sample-related Biases in Biomarker Studies. *EMBO Mol. Med.* **2019**, *11*, e10427. [\[CrossRef\]](#)
86. Deutsch, E.W.; Omenn, G.S.; Sun, Z.; Maes, M.; Pernemalm, M.; Palaniappan, K.K.; Letunica, N.; Vandenbrouck, Y.; Brun, V.; Tao, S.; et al. Advances and Utility of the Human Plasma Proteome. *J. Proteome Res.* **2021**, *20*, 5241–5263. [\[CrossRef\]](#)
87. Alhamdani, M.S.; Schröder, C.; Hoheisel, J.D. Oncoproteomic Profiling with Antibody Microarrays. *Genome Med.* **2009**, *1*, 68. [\[CrossRef\]](#)
88. Miller, J.C.; Zhou, H.; Kwekel, J.; Cavallo, R.; Burke, J.; Butler, E.B.; Teh, B.S.; Haab, B.B. Antibody Microarray Profiling of Human Prostate Cancer Sera: Antibody Screening and Identification of Potential Biomarkers. *Proteomics* **2003**, *3*, 56–63. [\[CrossRef\]](#) [\[PubMed\]](#)
89. Shafer, M.W.; Mangold, L.; Partin, A.W.; Haab, B.B. Antibody Array Profiling Reveals Serum TSP-1 as a Marker to Distinguish Benign from Malignant Prostatic Disease. *Prostate* **2007**, *67*, 255–267. [\[CrossRef\]](#)
90. Heo, C.K.; Bahk, Y.Y.; Cho, E.W. Tumor-associated Autoantibodies as Diagnostic and Prognostic Biomarkers. *BMB Rep.* **2012**, *45*, 677–685. [\[CrossRef\]](#)
91. Adeola, H.A.; Smith, M.; Kaestner, L.; Blackburn, J.M.; Zerbini, L.F. Novel Potential Serological Prostate Cancer Biomarkers Using CT100+ Cancer Antigen Microarray Platform in a Multi-cultural South African cohort. *Oncotarget* **2016**, *7*, 13945–13964. [\[CrossRef\]](#)
92. Lundberg, M.; Eriksson, A.; Tran, B.; Assarsson, E.; Fredriksson, S. Homogeneous Antibody-based Proximity Extension Assays Provide Sensitive and Specific Detection of Low-abundant Proteins in Human Blood. *Nucleic Acids Res.* **2011**, *39*, e102. [\[CrossRef\]](#) [\[PubMed\]](#)
93. Liu, S.; Shen, M.; Hsu, E.-C.; Zhang, C.A.; Garcia-Marques, F.; Nolley, R.; Koul, K.; Rice, M.A.; Aslan, M.; Pitteri, S.J.; et al. Discovery of PTN as a Serum-based Biomarker of Pro-metastatic Prostate Cancer. *Br. J. Cancer* **2021**, *124*, 896–900. [\[CrossRef\]](#)
94. Liu, S.; Garcia-Marques, F.; Zhang, C.A.; Lee, J.J.; Nolley, R.; Shen, M.; Hsu, E.C.; Aslan, M.; Koul, K.; Pitteri, S.J.; et al. Discovery of CASP8 as a Potential Biomarker for High-risk Prostate Cancer Through a High-multiplex Immunoassay. *Sci. Rep.* **2021**, *11*, 7612. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Campos-Fernández, E.; Oliveira Alqualo, N.; Moura Garcia, L.C.; Coutinho Horácio Alves, C.; Ferreira Arantes Vieira, T.D.; Caixeta Moreira, D.; Alonso-Goulart, V. The Use of Aptamers in Prostate Cancer: A Systematic Review of Theranostic Applications. *Clin. Biochem.* **2021**, *93*, 9–25. [\[CrossRef\]](#)
96. Welton, J.L.; Brennan, P.; Gurney, M.; Webber, J.P.; Spary, L.K.; Carton, D.G.; Falcón-Pérez, J.M.; Walton, S.P.; Mason, M.D.; Tabi, Z.; et al. Proteomics Analysis of Vesicles Isolated From Plasma and Urine of Prostate Cancer Patients Using a Multiplex, Aptamer-based Protein Array. *J. Extracell. Vesicles* **2016**, *5*, 31209. [\[CrossRef\]](#)
97. Dudani, J.S.; Ibrahim, M.; Kirkpatrick, J.; Warren, A.D.; Bhatia, S.N. Classification of Prostate Cancer Using a Protease Activity Nanosensor Library. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 8954–8959. [\[CrossRef\]](#)

98. Walker, K.A.; Chen, J.; Zhang, J.; Fornage, M.; Yang, Y.; Zhou, L.; Grams, M.E.; Tin, A.; Daya, N.; Hoogeveen, R.C.; et al. Large-scale Plasma Proteomic Analysis Identifies Proteins and Pathways Associated with Dementia Risk. *Nat. Aging* **2021**, *1*, 473–489. [\[CrossRef\]](#)
99. Joshi, A.; Mayr, M. In Aptamers They Trust: Caveats of the SOMAscan Biomarker Discovery Platform from SomaLogic. *Circulation* **2018**, *138*, 2482–2485. [\[CrossRef\]](#)
100. Tang, J.; Wang, Y.; Luo, Y.; Fu, J.; Zhang, Y.; Li, Y.; Xiao, Z.; Lou, Y.; Qiu, Y.; Zhu, F. Computational Advances of Tumor Marker Selection and Sample Classification in Cancer Proteomics. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 2012–2025. [\[CrossRef\]](#) [\[PubMed\]](#)
101. Clairefond, S.; Ouellet, V.; Péant, B.; Barrès, V.; Karakiewicz, P.I.; Mes-masson, A.M.; Saad, F. Expression of ERBB Family Members as Predictive Markers of Prostate Cancer Progression and Mortality. *Cancers* **2021**, *13*, 1688. [\[CrossRef\]](#) [\[PubMed\]](#)
102. Wagner, M.; Naik, D.N.; Pothén, A.; Kasukurti, S.; Devineni, R.R.; Adam, B.L.; Semmes, O.J.; Wright, G.L. Computational Protein Biomarker Prediction: A Case Study for Prostate Cancer. *BMC Bioinform.* **2004**, *5*, 1–9. [\[CrossRef\]](#)
103. Toth, R.; Schiffmann, H.; Hübner-Magg, C.; Büschek, F.; Höflmayer, D.; Weidemann, S.; Lebok, P.; Fraune, C.; Minner, S.; Schlomm, T.; et al. Random Forest-based Modelling to Detect Biomarkers for Prostate Cancer Progression. *Clin. Epigenetics* **2019**, *11*, 1–15. [\[CrossRef\]](#) [\[PubMed\]](#)
104. Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected Reaction Monitoring for Quantitative Proteomics: A Tutorial. *Mol. Syst. Biol.* **2008**, *4*, 222. [\[CrossRef\]](#)
105. Peterson, A.C.; Russell, J.D.; Bailey, D.J.; Westphall, M.S.; Coon, J.J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteom.* **2012**, *11*, 1475–1488. [\[CrossRef\]](#) [\[PubMed\]](#)
106. Gallien, S.; Duriez, E.; Crone, C.; Kellmann, M.; Moehring, T.; Domon, B. Targeted Proteomic Quantification on Quadrupole-Orbitrap Mass Spectrometer. *Mol. Cell. Proteom.* **2012**, *11*, 1709–1723. [\[CrossRef\]](#)
107. Domon, B.; Gallien, S. Recent Advances in Targeted Proteomics for Clinical Applications. *Proteom. Clin. Appl.* **2015**, *9*, 423–431. [\[CrossRef\]](#)
108. Faria, S.S.; Morris, C.F.M.; Silva, A.R.; Fonseca, M.P.; Forget, P.; Castro, M.S.; Fontes, W. A Timely Shift from Shotgun to Targeted Proteomics and How It Can Be Groundbreaking for Cancer Research. *Front. Oncol.* **2017**, *7*, 13. [\[CrossRef\]](#)
109. Sequeiros, T.; Rigau, M.; Chiva, C.; Montes, M.; Garcia-Grau, I.; Garcia, M.; Diaz, S.; Celma, A.; Bijnsdorp, I.; Campos, A.; et al. Targeted Proteomics in Urinary Extracellular Vesicles Identifies Biomarkers for Diagnosis and Prognosis of Prostate Cancer. *Oncotarget* **2017**, *8*, 4960–4976. [\[CrossRef\]](#)
110. Thomas, S.N.; Harlan, R.; Chen, J.; Aiyetan, P.; Liu, Y.; Sokoll, L.J.; Aebersold, R.; Chan, D.W.; Zhang, H. Multiplexed Targeted Mass Spectrometry-Based Assays for the Quantification of N-Linked Glycosite-Containing Peptides in Serum. *Anal. Chem.* **2015**, *87*, 10830–10838. [\[CrossRef\]](#)
111. Erickson, B.K.; Rose, C.M.; Braun, C.R.; Erickson, A.R.; Knott, J.; McAlister, G.C.; Wühr, M.; Paulo, J.A.; Everley, R.A.; Gygi, S.P. A strategy to Combine Sample Multiplexing with Targeted Proteomics Assays for High-throughput Protein Signature Characterization. *Mol. Cell* **2017**, *65*, 361–370. [\[CrossRef\]](#)
112. Zhang, S.; Garcia-D'Angeli, A.; Brennan, J.P.; Huo, Q. Predicting Detection Limits of Enzyme-linked Immunosorbent Assay (ELISA) and Bioanalytical Techniques in General. *Analyst* **2013**, *139*, 439–445. [\[CrossRef\]](#) [\[PubMed\]](#)
113. Wang, P.; Whiteaker, J.R.; Paulovich, A.G. The Evolving Role of Mass Spectrometry in Cancer Biomarker Discovery. *Cancer Biol. Ther.* **2009**, *8*, 1083–1094. [\[CrossRef\]](#) [\[PubMed\]](#)
114. Ren, A.H.; Diamandis, E.P.; Kulasingam, V. Uncovering the Depths of the Human Proteome: Antibody-based Technologies for Ultrasensitive Multiplexed Protein Detection and Quantification. *Mol. Cell. Proteom.* **2021**, 100155. [\[CrossRef\]](#) [\[PubMed\]](#)
115. Masuda, M.; Yamada, T. Signaling Pathway Profiling Using Reverse-phase Protein Array and Its Clinical Applications. *Expert Rev. Proteom.* **2017**, *14*, 607–615. [\[CrossRef\]](#)
116. Akbani, R.; Becker, K.-F.; Carragher, N.; Goldstein, T.; de Koning, L.; Korf, U.; Liotta, L.; Mills, G.B.; Nishizuka, S.S.; Pawlak, M.; et al. Realizing the Promise of Reverse Phase Protein Arrays for Clinical, Translational, and Basic Research: A Workshop Report. *Mol. Cell. Proteom.* **2014**, *13*, 1625–1643. [\[CrossRef\]](#)
117. Paweletz, C.P.; Charboneau, L.; Bichsel, V.E.; Simone, N.L.; Chen, T.; Gillespie, J.W.; Emmert-Buck, M.R.; Roth, M.J.; Petricoin III, E.F.; Liotta, L.A. Reverse Phase Protein Microarrays Which Capture Disease Progression Show Activation of Pro-survival Pathways at the Cancer Invasion Front. *Oncogene* **2001**, *20*, 1981–1989. [\[CrossRef\]](#)
118. Grubb, R.L.; Calvert, V.S.; Wulkuhle, J.D.; Paweletz, C.P.; Linehan, W.M.; Phillips, J.L.; Chuaqui, R.; Valasco, A.; Gillespie, J.; Emmert-Buck, M.; et al. Signal Pathway Profiling of Prostate Cancer Using Reverse Phase Protein Arrays. *Proteomics* **2003**, *3*, 2142–2146. [\[CrossRef\]](#)
119. Grubb, R.L.; Deng, J.; Pinto, P.A.; Mohler, J.L.; Chinnaiyan, A.; Rubin, M.; Linehan, W.M.; Liotta, L.A.; Petricoin, E.F.; Wulkuhle, J.D. Pathway Biomarker Profiling of Localized and Metastatic Human Prostate Cancer Reveal Metastatic and Prognostic Signatures. *J. Proteome Res.* **2009**, *8*, 3044–3054. [\[CrossRef\]](#)
120. Abeshouse, A.; Ahn, J.; Akbani, R.; Ally, A.; Amin, S.; Andry, C.D.; Annala, M.; Aprikian, A.; Armenia, J.; Arora, A.; et al. Cancer Genome Atlas Research Network The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **2015**, *163*, 1011–1025. [\[CrossRef\]](#)
121. Pin, E.; Stratton, S.; Belluco, C.; Liotta, L.; Nagle, R.; Hodge, K.A.; Deng, J.; Dong, T.; Baldelli, E.; Petricoin, E.; et al. A Pilot Study Exploring the Molecular Architecture of the Tumor Microenvironment in Human Prostate Cancer Using Laser Capture Microdissection and Reverse Phase Protein Microarray. *Mol. Oncol.* **2016**, *10*, 1585–1594. [\[CrossRef\]](#) [\[PubMed\]](#)

-
122. Zhang, Y.; Kwok-Shing Ng, P.; Kucherlapati, M.; Chen, F.; Liu, Y.; Tsang, Y.H.; de Velasco, G.; Jeong, K.J.; Akbani, R.; Hadjipanayis, A.; et al. A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations. *Cancer Cell* **2017**, 1–13. [[CrossRef](#)]
 123. Signore, M.; Alfonsi, R.; Federici, G.; Nanni, S.; Addario, A.; Bertuccini, L.; Aiello, A.; Di Pace, A.L.; Sperduti, I.; Muto, G.; et al. Diagnostic and Prognostic Potential of the Proteomic Profiling of Serum-Derived Extracellular Vesicles in Prostate Cancer. *Cell Death Dis.* **2021**, *12*, 1–14. [[CrossRef](#)] [[PubMed](#)]
 124. Kupcova Skalninkova, H.; Cizkova, J.; Cervenka, J.; Vodicka, P. Advances in Proteomic Techniques for Cytokine Analysis: Focus on Melanoma Research. *Int. J. Mol. Sci.* **2017**, *18*, 2697. [[CrossRef](#)] [[PubMed](#)]
 125. Tsauro, I.; Noack, A.; Makarevic, J.; Oppermann, E.; Waaga-Gasser, A.M.; Gasser, M.; Borgmann, H.; Huesch, T.; Gust, K.M.; Reiter, M.; et al. CCL2 Chemokine as a Potential Biomarker for Prostate Cancer: A Pilot Study. *Cancer Res. Treat.* **2015**, *47*, 306–312. [[CrossRef](#)]
 126. Al-Mazidi, S.; Farhat, K.; Nedjadi, T.; Chaudhary, A.; Zin Al-Abdin, O.; Rabah, D.; Al-Zoghaibi, M.; Djouhri, L. Association of Interleukin-6 and Other Cytokines With Self-reported Pain in Prostate Cancer Patients Receiving Chemotherapy. *Pain Med.* **2018**, *19*, 1058–1066. [[CrossRef](#)] [[PubMed](#)]
 127. Shore, N.D.; Pieczonka, C.M.; Henderson, R.J.; Bailen, J.L.; Saltzstein, D.R.; Concepcion, R.S.; Beebe-Dimmer, J.L.; Ruterbusch, J.J.; Levin, R.A.; Wissmueller, S.; et al. Development and Evaluation of the MiCheck Test for Aggressive Prostate Cancer. *Urol. Oncol. Semin. Orig. Investig.* **2020**, *38*, 683.e11–683.e18. [[CrossRef](#)]



Prostate cancer extracellular vesicle digital scoring assay – a rapid noninvasive approach for quantification of disease-relevant mRNAs

Jasmine J. Wang^{a,b,c,1}, Na Sun^{c,d,1}, Yi-Te Lee^c, Minhyung Kim^e, Tatyana Vagner^f, Krizia Rohena-Rivera^g, Zhili Wang^d, Zijiang Chen^a, Ryan Y. Zhang^c, Junseok Lee^c, Ceng Zhang^c, Hubert Tang^c, Josephine Widjaja^c, Tiffany X. Zhang^c, Dongping Qi^c, Pai-Chi Teng^b, Yu Jen Jan^b, Kuan-Chu Hou^c, Candace Hamann^a, Howard M. Sandler^{b,h}, Timothy J. Daskivich^{b,i}, Daniel J. Luthringer^{b,j}, Neil A. Bhowmick^{b,e,g}, Renjun Pei^d, Sungyong You^{b,e,f,*}, Dolores Di Vizio^{b,e,f,j,**}, Hsian-Rong Tseng^{c,k,***}, Jie-Fu Chen^{l,****}, Yazhen Zhu^{c,k,***}, Edwin M. Posadas^{a,b,*****}

^a Division of Medical Oncology, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA

^b Cedars-Sinai Cancer, Cedars-Sinai Medical Center, Los Angeles, CA, USA

^c California NanoSystems Institute, Crump Institute for Molecular Imaging, Department of Molecular and Medical Pharmacology, University of California, Los Angeles, Los Angeles, CA, USA

^d Key Laboratory for Nano-Bio Interface, Suzhou Institute of Nano-Tech and Nano-Bionics, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Suzhou, PR China

^e Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA

^f Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, CA, USA

^g Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA

^h Department of Radiation Oncology, Cedars-Sinai Medical Center, Los Angeles, CA, USA

ⁱ Division of Urology, Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, CA, USA

^j Department of Pathology and Laboratory Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA

^k Jonsson Comprehensive Cancer Center, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA

^l Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

ARTICLE INFO

Article history:

Received 17 June 2022

Received in revised form 30 November 2022

Accepted 25 December 2022

Available online 3 January 2023

ABSTRACT

Optimizing outcomes in prostate cancer (PCa) requires precision in characterization of disease status. This effort was directed at developing a PCa extracellular vesicle (EV) Digital Scoring Assay (DSA) for detecting metastasis and monitoring progression of PCa. PCa EV DSA is comprised of an EV purification device (i.e., EV Click Chip) and reverse-transcription droplet digital PCR that quantifies 11 PCa-relevant mRNA in purified PCa-derived EVs. A Met score was computed for each plasma sample based on the expression of the 11-gene

Abbreviations: DSA, Digital Scoring Assay; EV, extracellular vesicle; PCa, prostate cancer; RT-ddPCR, reverse transcription-droplet digital polymerase chain reaction

* Correspondence to: Surgery and Biomedical Sciences, Cedars-Sinai Medical Center, 8700 Beverly Blvd., Los Angeles, CA 90048, USA.

** Correspondence to: Department of Biomedical Sciences, Cedars-Sinai Medical Center, 8700 Beverly Blvd., Los Angeles, CA 90048, USA.

*** Corresponding authors at: California NanoSystems Institute, Crump Institute for Molecular Imaging, Department of Molecular and Medical Pharmacology, University of California, Los Angeles, Los Angeles, CA, USA.

**** Correspondence to: Department of Pathology, Memorial Sloan Kettering Cancer Center, 1275 York Ave., New York, NY, USA.

***** Correspondence to: Division of Medical Oncology, Department of Medicine, Cedars-Sinai Medical Center, 8700 Beverly Boulevard, Los Angeles, CA 90048, USA.

E-mail addresses: jasmine.wang@cshs.org (J.J. Wang), nsun2013@sinano.ac.cn (N. Sun), Yi-TeLee@mednet.ucla.edu (Y.-T. Lee), Minhyung.Kim@cshs.org (M. Kim), Tatyana.Vagner@cshs.org (T. Vagner), Krizia.Rohena-Rivera@cshs.org (K. Rohena-Rivera), zlwang2013@sinano.ac.cn (Z. Wang), Zijiang.Chen@cshs.org (Z. Chen), RYZhang@mednet.ucla.edu (R.Y. Zhang), JunseokLee@mednet.ucla.edu (J. Lee), cengzhang1208@outlook.com (C. Zhang), HubertTang@mednet.ucla.edu (H. Tang), josephine.r.widjaja@gmail.com (J. Widjaja), tiffanyxyz1516@ucla.edu (T.X. Zhang), dongpingqi@gmail.com (D. Qi), paichi.teng@gmail.com (P.-C. Teng), alex1322kimo@gmail.com (Y.J. Jan), houlfrank2003@gmail.com (K.-C. Hou), Candace.Hamann@cshs.org (C. Hamann), Howard.Sandler@cshs.org (H.M. Sandler), Timothy.Daskivich@cshs.org (T.J. Daskivich), Daniel.Luthringer@cshs.org (D.J. Luthringer), Neil.Bhowmick@cshs.org (N.A. Bhowmick), rpei2011@sinano.ac.cn (R. Pei), Sungyong.You@cshs.org (S. You), Dolores.Divizio@cshs.org (D. Di Vizio), HRTseng@mednet.ucla.edu, hrteng@mednet.ucla.edu (H.-R. Tseng), chenj14@mskcc.org, ussjsf@gmail.com (J.-F. Chen), YazhenZhu@mednet.ucla.edu (Y. Zhu), Edwin.Posadas@csmc.edu, Edwin.Posadas@csmc.edu (E.M. Posadas).

¹ These authors contributed equally to this work.

Keywords:

Prostate cancer
Liquid biopsy
Extracellular vesicle
Metastatic biomarker
Nanotechnology

panel using the weighted Z score method. Under optimized conditions, the EV Click Chips outperformed the ultracentrifugation or precipitation method of purifying PCa-derived EVs from artificial plasma samples. Using PCa EV DSA, the Met score distinguished metastatic ($n = 20$) from localized PCa ($n = 20$) with an area under the receiver operating characteristic curve of 0.88 (95% CI:0.78–0.98). Furthermore, longitudinal analysis of three PCa patients showed the dynamics of the Met scores reflected clinical behavior even when disease was undetectable by imaging. Overall, a sensitive PCa EV DSA was developed to identify metastatic PCa and reveal dynamic disease states noninvasively. This assay may complement current imaging tools and blood-based tests for timely detection of metastatic progression that can improve care for PCa patients.

© 2022 Published by Elsevier Ltd.

1. Introduction

In the United States (US), prostate cancer (PCa) is the most common, non-cutaneous, solid tumor malignancy affecting men. More than 268,490 cases will be diagnosed in 2022. [1] PCa is also the second leading cause of cancer-related death in American men and will claim more than 34,500 lives this year. [1] Advances in PCa care and diagnosis have brought forward several new tests and interventions including molecular imaging, antibody-targeted therapeutics, and immune therapies. [2] Additionally, the role of focal therapies has emerged as a means of controlling limited recurrences in the hope of avoiding protracted toxicity from systemic therapy. Given the cost and specificity of these treatments, [3] there is a growing need for simple, rapid, and affordable tests which can help to improve the triage of diagnostics and therapeutics in the growing pool of patients who need treatment.

This demand has fueled an interest in exploring liquid biopsies that noninvasively provide molecular insights into the underlying disease. [4] While many have turned to circulating tumor DNA, [5,6] these approaches are limited to characterization of fragmented DNA [7] and lack the ability to trace transcriptomic changes which are known to be highly relevant to the evolution of PCa. [8] Other groups have used various technologies to isolate circulating tumor cells (CTCs), [9,10] which provide rich genetic information. However, the time and cost of isolating and preserving the integrity of CTCs are challenging. In addition, clinical applications are limited due to the rarity of CTCs in relatively early-stage disease.

There has been growing interest in exploring the potential of extracellular vesicles (EVs) to address this unmet need. [4] EVs are phospholipid bilayer-enclosed particles released from both normal and cancer cells into body fluids, e.g., urine and blood. [4] These EVs carry membrane-bound markers and genetic materials including miRNA, mRNA, and fragments of DNA inherited from their parental cells. [11] By transferring these biomolecules between cells, EVs can function as a vehicle for intercellular communication. [11] This process has also been shown to prime the pre-metastatic niche at distant organs and facilitate the invasion and dissemination of cancer cells. [4,11] Compared to CTCs, EVs have advantages in their higher frequency in circulation [12] and the simplicity of sample preservation. [13] In addition, considering the relatively low rate of DNA mutations and the frequently altered transcriptomic patterns in PCa [8,14], EV-derived mRNA may provide more biological information relevant to PCa progression, which may not be obtained from ctDNA. As such, profiling mRNA in PCa-derived EVs holds great potential for rapid and accurate characterization of PCa. This approach may offer the opportunity to dynamically monitor the evolving biology of PCa over its natural history. One major challenge of this approach, however, is that PCa-derived EVs typically co-exist with those from non-tumor sources, resulting in complex background signals. [15] Thus, selective purification of PCa-derived EVs, which constitute only a minor portion of total EVs in plasma samples, is a major technical barrier.

To overcome this, many efforts have been dedicated to enriching PCa-derived EVs and analyzing their molecular contents. [16–18] For

example, Han et al. introduced an ultrasensitive DNA tetrahedron-based thermophoretic assay to detect mRNA in PCa-derived EVs and demonstrated its excellent performance in distinguishing PCa from benign prostate hyperplasia. [16] Others proposed nanomaterial-based strategies for multiplexed biodetection of potential biomarkers in cancer-derived EVs. [19–21] Recently, our team developed a novel microfluidic device coupled with a unique nanosurface (i.e., EV Click Chip) for purification of tumor-derived EVs. [18] By using reverse transcription-droplet digital polymerase chain reaction (RT-ddPCR) for quantification of the disease relevant mRNAs from purified EVs, we developed an EV Digital Scoring Assay (DSA) (i.e., EV Click Chip plus RT-ddPCR). This technology was initially applied to the identification of early-stage hepatocellular carcinoma within a group of cirrhotic patients. [18] With modifications of this approach, we aimed to apply the EV DSA to PCa and benchmark the performance of this assay in reflecting the clinical states of patients with PCa.

In this study, we optimized the EV DSA for rapid and sensitive analysis of the mRNA contents from PCa-derived EVs. The optimized EV DSA outperformed ultracentrifugation and a commercial EV precipitation assay in purifying PCa-derived EVs. A panel of 11 PCa-relevant mRNA markers was developed through a rigorous screening framework for specific detection of the underlying PCa in the blood background. Finally, a pilot study with 40 PCa patients showed the ability of this EV DSA to distinguish metastatic from localized PCa. In an analysis of serial blood samples from 3 patients, changes in the mRNA signatures detected by the EV DSA associated with clinical behavior observed over the course of treatment. These findings demonstrated the potential utility of this assay in PCa which may provide a novel means of interrogating EVs suitable for deployment into the clinical setting that may be used to improve personalized care.

2. Results

2.1. Optimization of PCa EV Digital Scoring Assay

As illustrated in Fig. 1, the PCa EV DSA was comprised of two major components: EV Click Chip for purification of PCa-derived EVs and RT-ddPCR for quantification of the mRNAs from the purified EVs. Two antibodies were selected for EV Click Chip to facilitate capture of PCa-derived EVs with high sensitivity and specificity: anti-epithelial cell adhesion molecule (EpCAM) and anti-prostate-specific membrane antigen (PSMA). These markers were selected on the basis of their well-characterized and strong expression on the surface of PCa epithelial cells. [22–25] Artificial plasma samples were used (Fig. 2a) to evaluate the performance of EV Click Chip throughout the optimization process. 22Rv1-derived EVs spiked into healthy donor plasma were used for assay optimization focusing on measurements of the AR-V7 transcript, a known androgen receptor variant present in 22Rv1 derived-EVs and absent in healthy donors' plasma. The counts of AR-V7 transcripts were used to calculate the EV recovery yield (Fig. 2a). The counts of AR-V7 transcripts in the artificial plasma samples and the EV Click Chip-purified EVs were

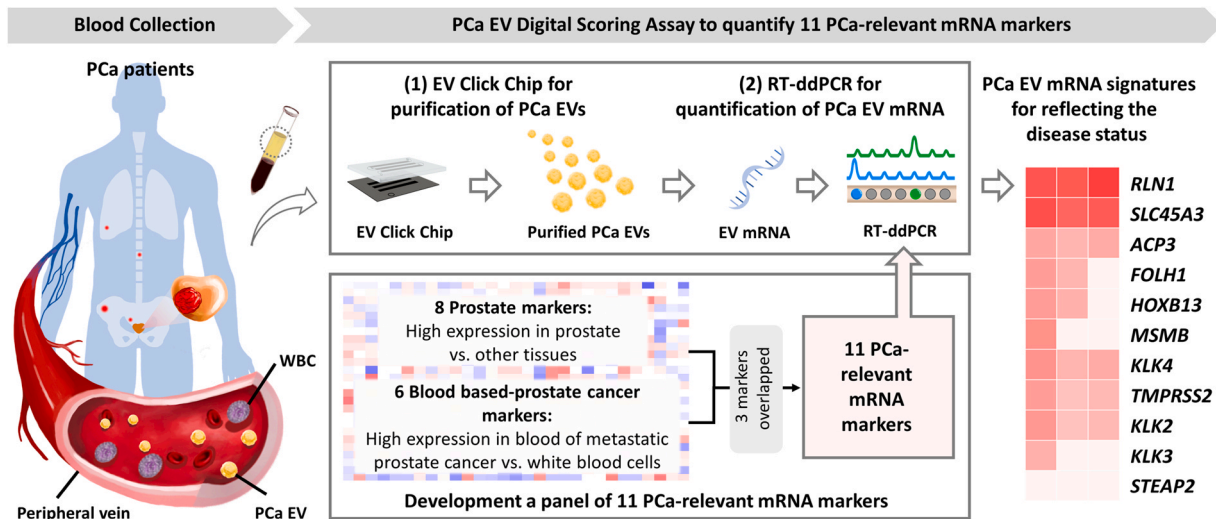


Fig. 1. PCa EV Digital Scoring Assay is comprised of two major components: (1) EV Click Chip for purifying PCa-derived EVs from plasma samples, and (2) RT-ddPCR for quantifying mRNA markers from the purified PCa-derived EVs. In parallel, a panel of 11 PCa-relevant genes, comprised of tissue-based prostate markers and blood-based PCa markers, was adopted to reflect disease state. EV, extracellular vesicle; PCa, prostate cancer; RT-ddPCR, reverse transcription-droplet digital polymerase chain reaction.

labeled as AR-V7 transcripts_{pre-purified} and AR-V7 transcripts_{post-purified}, respectively. The EV recovery yield obtained by EV Click Chips was calculated using the following Eq. (1): [18].

$$\text{EV recovery yield} = \frac{\text{AR} - \text{V7 transcripts}_{\text{post-purified}}}{\text{AR} - \text{V7 transcripts}_{\text{pre-purified}}} \quad (1)$$

The optimal concentrations of trans-cyclooctene (TCO)-anti-EpCAM and TCO-anti-PSMA for PCa DSA were determined by testing the EV recovery yields with a series of different concentrations of these two antibodies individually. As shown in Fig. 2b & 2c, for a 100-μL artificial plasma sample, optimal EV recovery yields were achieved using 25 ng for TCO-anti-EpCAM and TCO-anti-PSMA, respectively. EV recovery yield was substantially higher with the combination of TCO-anti-EpCAM and TCO-anti-PSMA at their optimal concentrations as compared to individual antibody (TCO-anti-EpCAM, $P = 0.006$; TCO-anti-PSMA, $P = 0.002$, Fig. 2d).

Using this dual capture antibody combination and the optimized assay parameters from our previous study, [18] the dynamic range of capture performance of EV Click Chip was investigated using artificial plasma samples spiked with different amount of 22Rv1-derived EVs containing 0–5000 AR-V7 transcripts in a 500-μL volume. Our results demonstrated consistent capture (Fig. 2e) across a wide range of concentration of artificial plasma samples. In addition, the results confirmed the linearity of the EV recovery yields by EV Click Chip ($y = 0.74x$, $r^2 = 0.99$, Fig. 2f). Consistency in the triplicate measurements of the capture performance revealed the high precision of the optimized assay (Supplementary Table 1). We also demonstrated high purity (90.38%) of the captured EVs by EV Click Chip (see Supplementary Methods for measurement and calculation of purity).

Lastly, we compared the EV recovery yields between the optimized EV Click Chip, ultracentrifugation, and ExoQuick® ULTRA EV Isolation Kit for Serum and Plasma (a commercially available assay). All artificial plasma samples were analyzed in triplicate. The EV recovery yield of the optimized EV Click Chip ($92.5 \pm 5.4\%$; Fig. 2g) was significantly higher than those from ultracentrifugation ($41.0 \pm 3.2\%$; $P < 0.001$) and the ExoQuick® EV isolation methods ($34.9 \pm 2.2\%$; $P < 0.001$).

Characterization of EVs was carried following guidelines from the International Society for Extracellular Vesicles (ISEV). [26] The morphologies of the 22Rv1-derived EVs before capture, the EVs captured on EV Click Chip, and the EVs released from EV Click Chip

were characterized by transmission electron microscopy (TEM; Supplementary Figure 1a, 1b, and 1c). The size of 22Rv1-derived EVs used for optimization ranged from 30 to 450 nm in diameter with a mean of 186.3 nm by nanoparticle tracking analysis (NTA; Supplementary Figure 1d). Size distribution and intactness of EVs were not altered during purification. This was consistent with the findings from our previous study. [18] (Supplementary Figure 1d). Western blot analysis showed the presence of epithelial tissue specific transmembrane protein EpCAM (MISEV2018 [26] Category 1) and cytosolic protein Annexin V (Category 2), as well as the absence of non-EV co-isolated structures apolipoprotein APOA1 (Category 3) and intracellular mitochondria cytochrome C (Category 4) in LNCaP-derived EVs and artificial PCa plasma samples purified by EV Click Chips (Supplementary Figure 1e). Both immunocapture markers i.e., EpCAM and PSMA were identified on the LNCaP-derived EVs before and after purification. None of the above markers were detected in male healthy donor plasma purified by EV Click Chip, confirming the specificity of EV Click Chip for capturing EpCAM and PSMA positive EVs. LNCaP cell lysate and healthy donor plasma were analyzed in parallel as positive controls for these markers.

2.2. Development and validation of a panel of 11 PCa-relevant genes

After the optimization of PCa-derived EV purification, a panel of mRNA markers reflecting the underlying malignancy was developed. The selection framework for these genes is shown in the Fig. 3a. First, to ensure the markers were highly specific to prostate rather than other tissues, 120 candidate genes were nominated on the basis of elevated expression in prostatic epithelium in contrast to 36 other tissues in The Tissue Atlas of The Human Protein Atlas. [27,28] Eight genes specific to prostate tissue were selected by applying a threshold of a greater than 5-fold change when comparing prostate and non-prostate tissues. In parallel, to incorporate markers associated with metastatic PCa, the gene set from Miyamoto et al. [29] was considered given its use in a PCa liquid biopsy setting as blood-based biomarkers for detection of disseminated PCa. The six top expressed genes in their dataset (GEO GSE67980) were incorporated into the final panel. As 3 genes overlapped between the above gene sets, this process yielded 11 PCa-relevant genes, including ACP3, FOLH1, HOXB13, KLK2, KLK3, KLK4, MSMB, RLN1, SLC45A3, STEAP2, and TMPRSS2 (Supplementary Table 2). Since the majority of the background EVs circulating in cancer patient's blood originated from

a Optimization of PCa EV Digital Scoring Assay

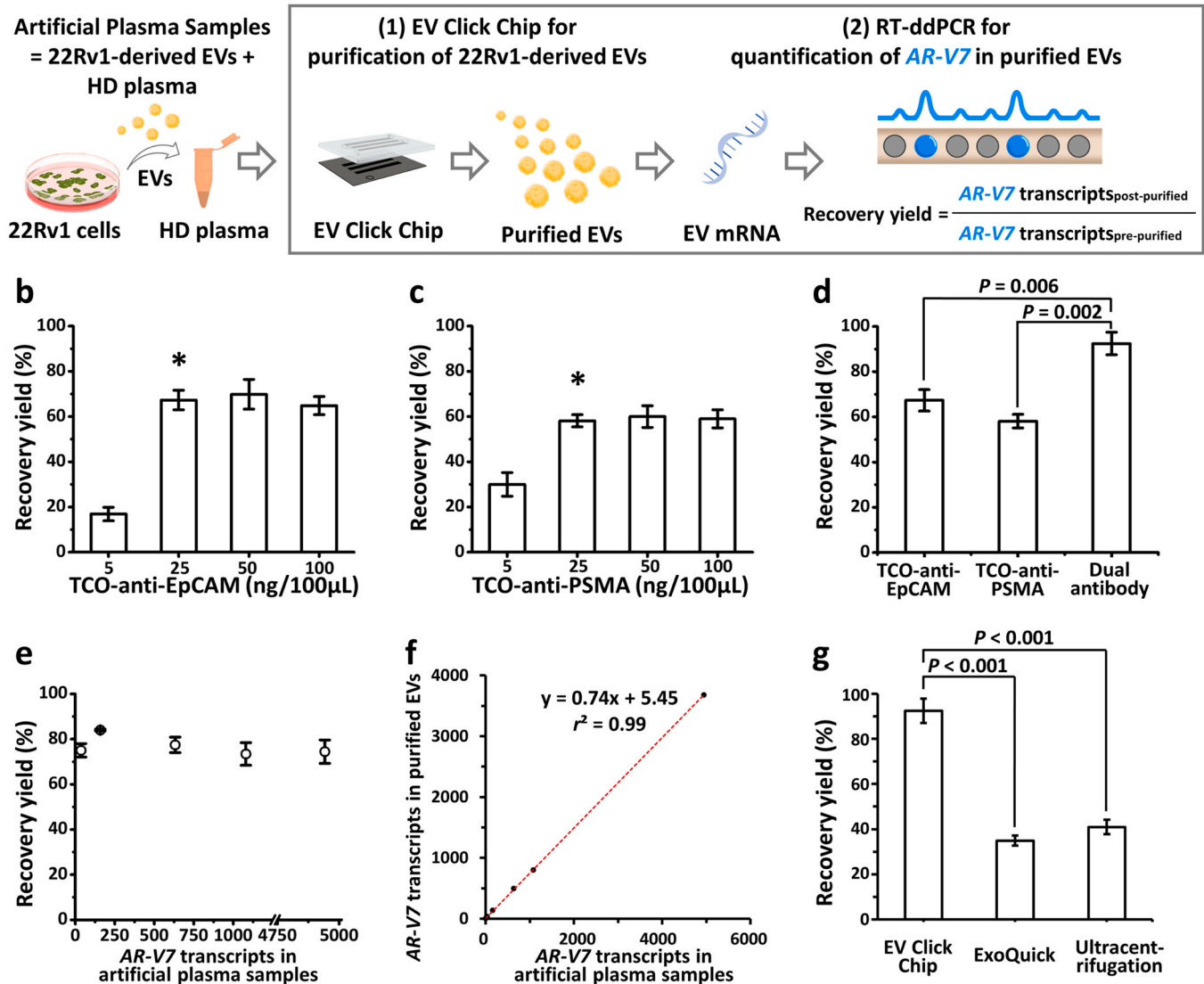


Fig. 2. Optimization of capture antibody and EV Click Chip using artificial PCa EV samples. (a) Schematic workflow of a quantitative method to evaluate the performance of EV Click Chip. Artificial plasma samples were prepared by spiking 22Rv1-derived EVs into the plasma from a male healthy donor. EV Click Chips were utilized to purify 22Rv1-derived EVs and RT-ddPCR was applied to quantify the *AR-V7* transcripts in the purified EVs to obtain the recovery yield. (b-c) Recovery yields observed for EV Click Chip at different concentrations of TCO-anti-EpCAM and TCO-anti-PSMA. Data are presented as means $\pm$ SD of three independent assays. *Indicates the concentration selected for further optimization studies. (d) The recovery yields observed in the presence of individual and dual antibodies, i.e., TCO-anti-EpCAM, TCO-anti-PSMA, and combination of the two antibodies. Data are presented as means $\pm$ SD of three independent assays. (e) Capture performance of EV Click Chip using artificial sample containing 0–5000 *AR-V7* transcripts per 500-μL volume. (f) Dynamic linearity range of the recovery yields observed for EV Click Chip across artificial sample containing 0–5000 *AR-V7* transcripts. (g) The recovery yields observed for optimized EV Click Chip, ExoQuick® ULTRA EV isolation kit and ultracentrifugation. Data are presented as means $\pm$ SD of three independent assays. EpCAM, epithelial cell adhesion molecule; EV, extracellular vesicle; PCa, prostate cancer; PSMA, prostate-specific membrane antigen; RT-ddPCR, reverse transcription-droplet digital polymerase chain reaction; SD, standard deviation; TCO, trans-cyclooctene.

immune cells, [30] it was crucial to ensure the selected genes are expressed at low levels in immune cells to minimize background noise. The expression level of these 11 genes was further evaluated in PCa tissues (from The Cancer Genome Atlas [TCGA]), [8] PCa cell lines (from Cancer Cell Line Encyclopedia [CCLE]), [31] and immune cells (from Differentiation MAP [DMAP]), [32] As shown in Fig. 3b, these 11 genes were highly expressed in PCa tissues and PCa cell lines, while lowly expressed in immune cells ($P < 0.001$, respectively), making them potential PCa markers for a blood-based assay.

2.3. Performance of the 11-gene panel in PCa cell line-derived EVs

Primers and probes for these 11 genes were validated using RT-ddPCR to confirm specificity (Fig. 4a). Strong signals were detected

in PCa cells while no signal was detected in white blood cells (WBCs) from a male healthy donor. To ensure that these genes were also expressed in PCa-derived EVs with limited expression in background EVs on EV Click Chip, the following were tested: (1) PCa cell line-derived EVs without purification of EV Click Chip, (2) PCa cell line-derived EVs purified from artificial plasma samples by EV Click Chip, and (3) EVs isolated from a male healthy donor's plasma by EV Click Chip (Fig. 4b). Compared to the strong signals in PCa cell line-derived EVs, no signals were detected in healthy donor's EVs with the exception of low *RLN1*. In addition, similar expression profiles were observed in PCa cell line-derived EVs before and after purification, which indicated that the mRNA components were not affected by the purification process. Finally, we compared the expression of the 11 genes in PCa cells and their derived EVs. There was a high

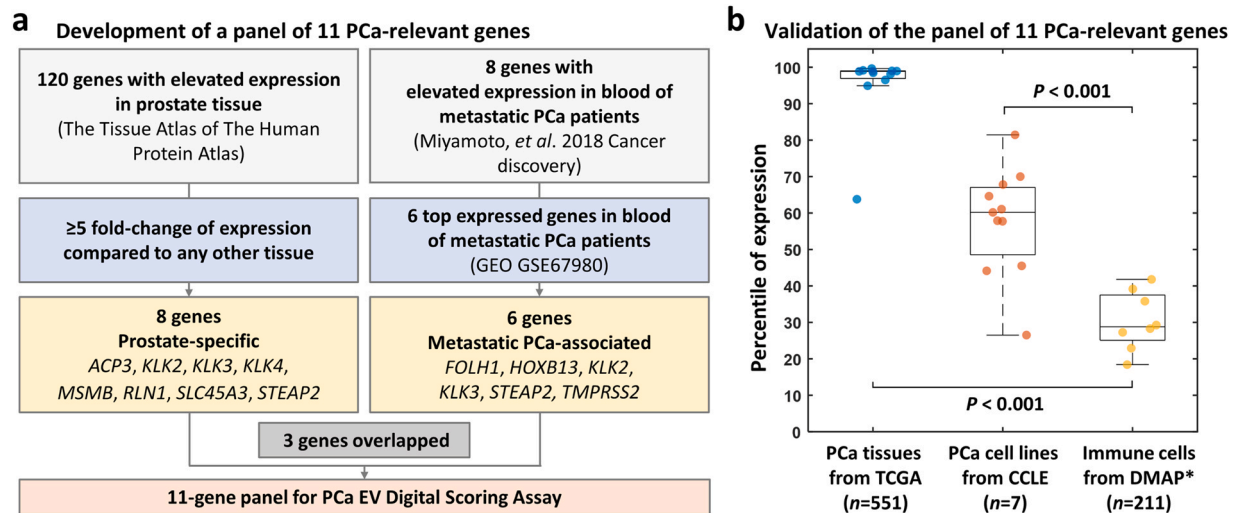


Fig. 3. Development and validation of a panel of 11 PCa-relevant genes. (a) Schematic flow of the selection of 11-gene panel for PCa EV Digital Scoring Assay, including *ACP3*, *FOLH1*, *HOXB13*, *KLK2*, *KLK3*, *KLK4*, *MSMB*, *RLN1*, *SLC45A3*, *STEAP2*, and *TMPRSS2*. (b) Box plot of the percentile of expression level of the 11-gene panel in primary PCa patients' tissues from TCGA, PCa cell lines from CCLE, and immune cells from DMAP. Each dot represents the mean of percentile of each maker. *Expression of *STEAP2*, *SLC45A3*, and *KLK4* are not available in DMAP. CCLE, Cancer Cell Line Encyclopedia; DMAP, Differentiation MAP dataset; EV, extracellular vesicle; PCa, prostate cancer; TCGA, The Cancer Genome Atlas.

correlation in gene expression patterns between EVs and their parental cells (LNCaP, $r = 0.95$, $P < 0.001$; C4-2B, $r = 0.74$, $P = 0.01$; 22Rv1, $r = 0.89$, $P < 0.001$). These findings point to the utility of these 11 genes in the context of the EV DSA.

2.4. Met scores for differentiating metastatic from localized PCa

Having validated the panel and PCa EV DSA on artificial samples, we designed a pilot study to benchmark the performance of this assay on clinical samples (Fig. 5a). For this pilot study, plasma samples from two groups of patients were utilized: (1) localized PCa ($n = 20$); (2) metastatic PCa ($n = 20$). Demographic and clinical information of the patients are presented in Table 1.

Expression of the 11 genes from the PCa EV DSA is shown as a heatmap in Fig. 5b with $\log_2$ -transformed transcript counts. A Met score for each patient sample was calculated using the weighted Z score method based on the 11-gene signature. [33] As shown in Fig. 5c, the Met scores of metastatic PCa samples were significantly higher ($P < 0.001$) than those from patients with localized PCa. Receiver operating characteristic (ROC) analysis further confirmed the ability of the Met score to distinguish metastatic PCa from localized PCa with an area under the ROC curve (AUROC) of 0.88 (95% CI=0.78–0.98; sensitivity=85%, specificity=75%; Fig. 5d) at the optimal cutoff of -0.30, which significantly surpassed the predictive power of serum prostate-specific antigen (PSA) concentration (AUROC=0.64, 95% CI=0.45–0.84, $P = 0.03$; sensitivity=65%, specificity=75%) at the cutoff of 10.25 ng/mL. The addition of PSA to Met score did not increase the predictive power (AUROC=0.89, 95% CI=0.79–0.99, $P = 0.46$; sensitivity=85%, specificity=75%). To investigate the performance of each gene in discriminating metastatic from localized PCa, a Met score was calculated for each individual gene and ROC analyses were performed. Among the 11 genes, the ROC analyses for *RLN1*, *SLC45A3*, and *ACP3* were statistically significant when being considered as sole predictors for metastatic PCa (Supplementary Figure 2). While *FOLH1* (PSMA) and *STEAP2* were exclusively seen in EVs from metastatic PCa, their discriminatory power did not meet statistical significance in our limited samples.

To consider the impact of multiple tumor features (macroscopic and microscopic) on the Met scores, a subgroup analysis was conducted. The Met scores calculated from the 11-gene panel did not show a strong correlation with T stage or Gleason score

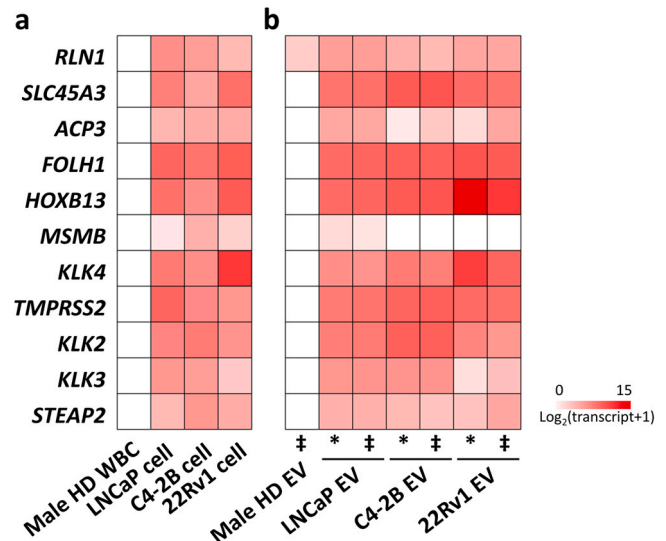


Fig. 4. Performance of the 11-gene panel in PCa cell line-derived EVs. (a) Expression heatmap of the 11 gene-panel with LNCaP, C4-2B, and 22Rv1 cells (positive controls) and male healthy donor (HD) WBCs (negative control). (b) Expression heatmap of the 11 gene-panel in the PCa cell line-derived EVs before purification (labeled as *), purified PCa cell line-derived EVs from artificial plasma samples (i.e., a male HD's plasma spiked with LNCaP, C4-2B, and 22Rv1-derived EVs, respectively) by EV Click Chip (labeled as ‡), and purified EVs from a male HD's plasma by EV Click Chip. Transcript counts are $\log_2$ -transformed as shown in the heatmaps. ‡ after purification by EV Click Chip. * before purification by EV Click Chip. EV, extracellular vesicle; HD, healthy donor; PCa, prostate cancer; RT-ddPCR, reverse transcription-droplet digital polymerase chain reaction; WBC, white blood cell.

(Supplementary Figure 3a and 3b). No association was identified between Met scores and tumor volume/mass determined at the final pathologic evaluation of radical prostatectomy (Supplementary Figure 3c and 3d). On the other hand, among patients with metastatic PCa, the Met scores were significantly higher in metastatic castration-resistant prostate cancer (mCRPC) than those in metastatic castration-sensitive prostate cancer (mCSPC) ($P = 0.04$, Supplementary Figure 3e). Lastly, we found the Met scores were not associated with serum PSA concentration in both localized and metastatic PCa (Supplementary Figure 3f and 3g).

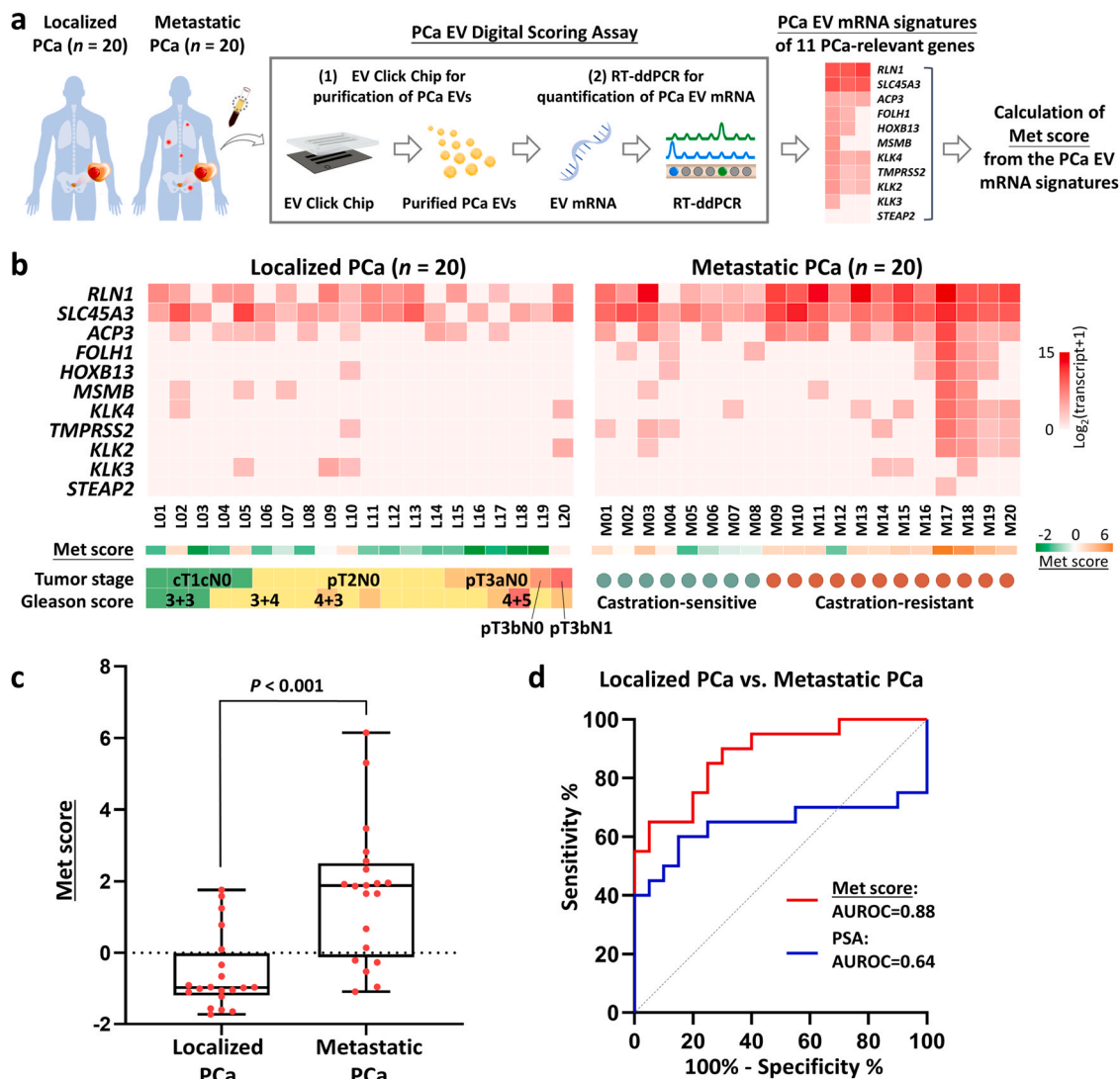


Fig. 5. Performance of PCa EV Digital Scoring Assay in distinguishing metastatic PCa patients from localized PCa patients (a) The workflow of PCa EV Digital Scoring Assay for quantifying the 11-gene panel and calculating Met score in purified EVs from patients with localized PCa (n = 20) and metastatic PCa (n = 20). (b) Heatmap of the 11 genes in patients with localized PCa (n = 20) and metastatic PCa (n = 20). Transcript counts were log₂-transformed. Corresponding Met scores were calculated from the expression of the 11 genes. (c) Box plot of the distribution of Met scores among localized PCa (n = 20) and metastatic PCa patients (n = 20). (d) ROC curves for Met score and serum PSA concentration to distinguish metastatic from localized PCa (Met score: AUROC=0.88, $P < 0.001$, 95% CI=0.78–0.98; serum PSA concentration: AUROC=0.64, $P = 0.12$, 95% CI=0.45–0.84). AUROC, area under the receiver operating characteristic curve; CI, confidence interval; EV, extracellular vesicle; PCa, prostate cancer; PSA, prostate-specific antigen; ROC, receiver operating characteristic.

2.5. Dynamic performance of the PCa EV Digital Scoring Assay

As noted above, an advantage of blood-based biomarker assays is the capacity to examine the changes in information in the face of clinical events of importance over time. To illustrate the dynamic performance of this assay, we analyzed selected samples from three patients in our cohort over the course of their treatment.

Case A. (Fig. 6a): This was a patient with a pT3bN0 Gleason 3 + 4 PCa who experienced a persistent PSA elevation following radical prostatectomy. Conventional imaging identified an isolated metastasis to the right iliac bone that was treated with metastasis directed radiotherapy without treating the prostatic fossa. After the treatment, his disease became quiescent by biochemical and radiographic surveillance for four years. At this timepoint of low and stable disease burden, his Met score was also the lowest compared to subsequent blood draws. The patient was noted to have a rising serum PSA concentration four years post-prostatectomy and finally had a ⁶⁸Ga-PSMA positron emission

tomography and computed tomography (PET/CT) when his serum PSA concentration rose to 1.4 ng/mL. The PET/CT identified an L4 lesion and periaortic lymph nodes suited for metastasis directed radiotherapy. Two blood draws during this progression phase revealed a consecutive increase of Met score, corresponding to a continuous increase of disease burden and/or activity clinically.

This patient declined castration therapy and opted for radiation to the metastatic lesion. He did not experience a decline in his serum PSA concentration following radiation but did have a ⁶⁸Ga-PSMA PET/CT that no longer showed the L4 lesion or the nodes. The Met score after the treatment remained stable, which was commensurate with the clinical picture. The lack of benefit from focal treatment was further confirmed by the emergence of a new rib metastasis emerging on CT and bone scan one year later.

In this clinical case, the Met score reflected the increase of disease burden and their lack of response to focal treatment. Of note, the Met score detected the progression of disease when the disease burden was below the limits of detection even with PSMA PET/CT.

Table 1
Clinical characteristics of PCa patients.

Characteristic	Localized PCa (n = 20)	Characteristic	Metastatic PCa (n = 20)
Median Age, y (IQR)	64 (60.3–69.8)	Median Age, y (IQR)	69.5 (60.8–76.5)
Age < 60, n (%)	4 (20.0)	Age < 60, n (%)	5 (25.0)
Age ≥ 60, n (%)	16 (80.0)	Age ≥ 60, n (%)	15 (75.0)
Race/ethnicity, n (%)		Race/ethnicity, n (%)	
White	10 (50.0)	White	13 (65.0)
Black	6 (30.0)	Black	3 (15.0)
Asian	3 (15.0)	Asian	2 (10.0)
Hispanic	1 (5.0)	Hispanic	2 (10.0)
Median PSA, ng/mL (IQR)	7.4 (5.5–11.4)	Median PSA, ng/mL (IQR)	22.1 (1.6–222.5)
The extent of the primary tumor, n (%)		Castration sensitivity, n (%)	
T1c	6 (30.0)	Castration-sensitive	7 (35.0)
T2	8 (40.0)	Castration-resistant	13 (65.0)
T3	6 (30.0)		
Nodal involvement, n (%)			
N0	19 (95.0)		
N1	1 (5.0)		
Gleason Score, n (%)			
3 + 3	3 (15.0)		
3 + 4	12 (60.0)		
4 + 3	4 (20.0)		
4 + 5	1 (5.0)		
Treatment choice, n (%)			
Active surveillance	6 (30.0)		
Radical prostatectomy	14 (70.0)		

IQR, interquartile range; PCa, prostate cancer; PSA, prostate-specific antigen.

Case B. (Fig. 6b): This patient presented with biochemical recurrence after radical prostatectomy followed by salvage radiotherapy. His serum PSA concentration rose to 6.4 ng/mL. While conventional CT and bone scans failed to identify metastatic deposits, a ^{68}Ga -PSMA-PET/CT identified multiple retroperitoneal lymph nodes and an osseous metastasis at L4. His pattern of spread was deemed as not appropriate for focal therapy. As such, intensified androgen deprivation therapy (ADT) was initiated. His serum PSA declined and has now been undetectable.

In this case, conventional imaging failed to detect the metastatic disease and was not informative in monitoring the patient's response, whereas the Met score decreased after the initiation of medical castration which mirrored his clinical status.

Case C. (Fig. 6c): This patient presented with advanced, metastatic prostate cancer with a serum PSA concentration in excess of 2000 ng/mL. Imaging showed diffuse bone metastases. As such, the patient was started on intensified ADT with abiraterone acetate and prednisone. He had a favorable biochemical and clinical response. Two years after starting the therapy, he presented with acute back pain and upper extremity weakness. Though his PSA remained stable, imaging showed a new lesion at T7 and in the right tibia noted as concerning for metastatic progression. Due to the neurologic symptoms, a magnetic resonance imaging (MRI) of the spine was obtained and was read as suspicious for a metastatic lesion. The clinical picture raised concerns for neuroendocrine transformation and progression. Upon referral to neurosurgery, the diagnosis was changed to compression fracture resulting from drug-induced osteoporosis. Radiographs of the tibia showed no correlation with the bone scan—i.e., no fracture or sclerosis to indicate metastatic disease. With conservative measures, the patient's symptoms abated. He has remained clinically well for an additional one year and continues therapy.

Four serial blood samples were evaluated by PCa EV DSA. The first two bloods were drawn while his disease was responding to intensified ADT very well. The last two bloods were drawn after the new suspicious lesions emerged. Met scores of across these four blood samples remained stable regardless of the emergence of new lesions. This case demonstrated the assay was capable of clarifying

progression from non-progression events detected by radiographs that could have resulted in dramatically different courses of clinical action – in this case, chemotherapy for progression versus bone modifying agents for osteoporosis.

3. Discussion

In the current study, we have presented the development of a PCa EV DSA, which combines Click Chip technology for purification of PCa-derived EVs with RT-ddPCR for quantification of the disease-relevant mRNAs. In parallel, an 11-gene mRNA panel associated with the underlying PCa was generated through a rigorous selection process (Fig. 1). Using this approach, the Met scores calculated from the expression of the 11 genes had the capacity to distinguish metastatic from localized PCa. Finally, in the presented cases, the changes in the Met scores reflected the changes in disease burden in the face of treatment over its course. This points toward the potential for using this assay to monitor PCa over its disease course.

Blood-based testing has been an important part of PCa clinical care. Since its introduction in the 1980s, serum PSA measurements have become widely used and have played an important role in advancing the field. [34] It is clear, however, to all practitioners that there are important limitations to the value of this approach. PSA measurements have not been highly accurate as a means of identifying metastatic disease. [35] As a particular example, in the setting of biochemical recurrence following radical prostatectomy, absolute PSA concentration can help risk stratify patients for benefit from salvage radiotherapy, but it cannot be used to identify patients with disease beyond a standard radiation port who are likely to progress through standard salvage radiotherapy. [36] Moreover, while changes in PSA in an individual patient are useful, there is a significant variation in the rate of PSA secretion from prostatic tissues that can obscure the clinical picture. For example, it is recognized that irritation of the prostate gland can lead to elevated serum PSA concentrations not related to cancer. [37] Conversely, anaplastic variants of PCa often produce lower amounts of PSA than more typical acinar adenocarcinomas. [38] As such, there remains a need for a non-PSA based means of assessing disease behavior that can be used to address these clinical scenarios.

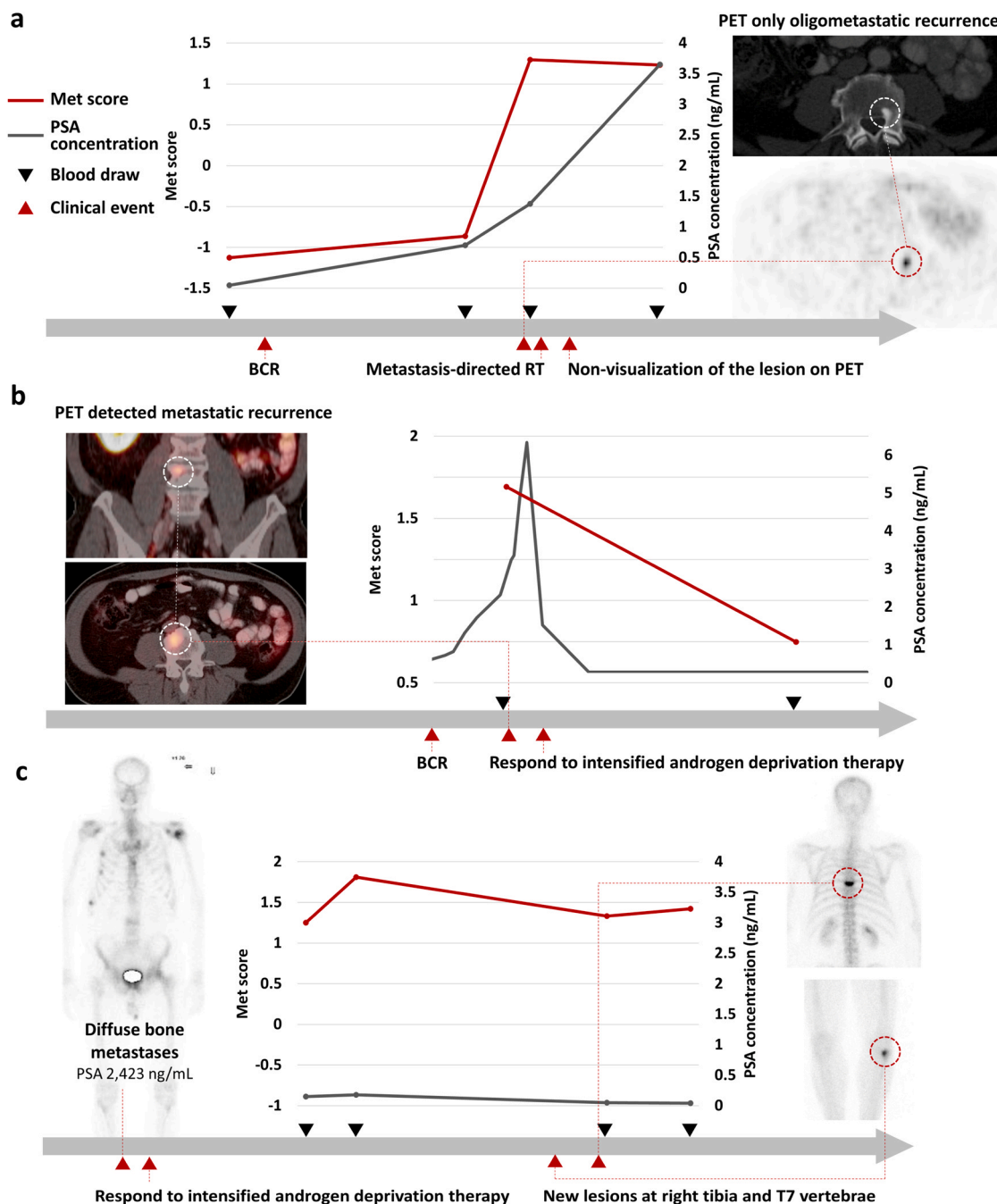


Fig. 6. Dynamic performance of the PCa EV Digital Scoring Assay. Met scores from the PCa EV DSA were generated for three PCa patients over the course of treatment. (a) Patient undergoing metastasis-directed therapy via stereotactic radiotherapy without subsequent benefit; (b) Patient with clinical response to intensified androgen deprivation therapy; and (c) Patient with suspected progression by radiograph and symptoms later found to be erroneous. BCR, biochemical recurrence; EV, extracellular vesicle; PCa, prostate cancer; PET, positron emission tomography; PSA, prostate specific antigen; RT, radiotherapy.

The utility of EV transcripts enriched from urine as a noninvasive diagnostic biomarker for detecting high-risk PCa has been investigated over the past decade. [39,40] By measuring *PCA3* and *ERG* transcripts, the ExoDx Prostate (IntelliScore) (EPI) test can non-invasively detect high-grade (Gleason score ≥ 7) PCa and has been used to help minimize biopsies among patients who likely only have low-grade disease. [39,40] In contrast to numerous urine-based EV-mRNA studies in this field, to our knowledge, few blood-based studies investigated the association between EV-mRNA and the disease behavior of PCa. Reported studies placed emphasis on the prognostic role of *AR-V7* and identified castration-resistant PCa patients with *AR-V7* in EVs as at risk for disease progression and poor survival.

[41–43] While these existing EV assays focus on only one or two transcripts within a single stage of PCa, the clinical relevance and significance of mRNA in EVs have not been widely elucidated across the broad spectrum of PCa despite the growing understanding of aberrant transcription in PCa evolution. [8,44] As such, there has remained a need for an EV assay profiling a panel of multiple PCa-relevant mRNAs with the potential for reflecting the disease state of PCa patients and their evolving biology.

As demonstrated in our previous study, EV Click Chips allow for rapid purification of tumor-derived EVs. [18] Integration of the nanosubstrate-surface, dual-antibody capture, and click chemistry-mediated EV capture/release process not only enhanced purification

efficiency of PCa-derived EVs, but also minimized nonspecific binding of particles from the background. [18] It is worthy to note that after optimization, EV Click Chip outperformed ultracentrifugation and commercial ExoQuick Assay in specific purification of PCa-derived EVs. This underscores its advantage in extraction of PCa-specific information from small volumes of patient plasma. In addition, the use of RT-ddPCR further allows for absolute quantification of mRNA transcripts, reducing bias when measuring signals from rare vesicles. [45] Lastly, we successfully applied the established EV DSA [18] to PCa, indicating this assay can be rapidly adapted to different diseases and clinical settings.

In our analysis of clinical samples, the mRNA contents of PCa-derived EV were profiled using PCa EV DSA for a cohort of patients with differing states across the disease spectrum of PCa. With 40 patients being profiled, we noted that Met scores were significantly higher in metastatic patients than in localized patients. This may be attributed to increased EV shedding [46] and alteration in mRNA contents during metastatic progression. Interestingly, we observed that *FOLH1* (*PSMA*) and *STEAP2* were exclusively found in EVs from metastatic PCa. Others have reported that *FOLH1* (*PSMA*) and *STEAP2* are overexpressed in aggressive PCa cells and tissues. Several studies have also pointed to their roles in promoting angiogenesis, migration, and invasion. [47–49] Our observation highlights the performance of Met scores as means of describing the expression pattern of the 11-gene panel. This approach outperformed individual genes in distinguishing the disease states. In addition, it is important to note that the genes in this panel are involved in multiple pathways associated with PCa progression in addition to androgen receptor signaling. This is also supported by the fact that the Met score was not correlated with serum PSA concentrations in both localized and metastatic PCa patients. This underscores that our approach provides information reflecting PCa disease status that is independent of serum PSA measurements. Overall, our study demonstrated significant changes in gene expression identified from tumor-derived EVs across different stages of PCa. Further investigations are warranted to explore the biological roles of this expression signature in EVs of metastatic PCa.

To further explore the dynamic change of EV-based signatures over the course of disease evolution, we performed analyses of specimens collected over time and across therapies for given patients. The presented cases illustrate the dynamic performance of the PCa EV DSA in the clinical setting.

The first case (Fig. 6a) showed the Met score accords with the persistent increase of micro-metastatic disease below the PET-detection threshold that blossomed into metastatic lesions one year later. This points toward the ability of this assay to detect molecular signals from microscopic deposits of disease which even the PSMA-PET imaging could not identify. Detection of micro-metastatic disease is still a critical unmet need in PCa. Patients with micro-metastatic disease would not benefit from the focal therapy including salvage radiotherapy and metastasis-directed therapy due to the presence of disease outside of treated areas, as shown in this case. The development of a test with sensitivity beyond the current imaging modalities will greatly improve shared-decision making around these focal treatments to optimize outcomes and quality of life.

In the second case (Fig. 6b), the initial high Met score reflected his metastatic disease detected by PSMA-PET/CT, which is in line with our findings that metastatic PCa patients have higher Met scores (Fig. 5c). In addition, the Met score decreased in response to intensified ADT which was also reflected by a decrease in serum PSA concentration and improvement in imaging. This was consistent with our observation that the Met scores of mCSPC patients responding to intensified ADT were lower than those of mCRPC patients progressing under ADT (Supplementary Figure 3e). As many genes in our panel are regulated by androgen receptor activation,

[50] we posited that the Met score may reflect the effect of the overall clinical effect of androgen suppression. Further studies that explore the Met score as a biomarker for hormonal responsiveness of PCa are ongoing.

Pathologic fracture due to bone metastasis of PCa can be challenging to distinguish from compression fracture as a result of osteoporosis due to long-term use of ADT. In the third case (Fig. 6c), the PCa EV DSA accurately indicated the ^{99}Tc -bone scan finding was not caused by PCa metastasis. This identification of the cause of the lesion can result in dramatically different clinical action on whether to initiate the next line of PCa treatment. In conclusion, this vignette illustrates the utility of this assay to provide PCa-specific molecular information as an adjunct to imaging.

We do recognize that there are limitations to the study as presented. First, due to the lack of comprehensive PCa EV-derived mRNA data in the current scientific literature or publicly available databases, the genes selected for the initial assay were selected from tissue-based datasets. These may not provide the most suitable candidates for an EV-based study. Thus, we conducted a PCa cell line-derived EV interrogation to confirm the applicability of this gene panel in EVs. As the knowledge of EV cargo continues to evolve, this assay can be easily adapted using alternative genes to identify and reflect specific changes of disease biology—particularly those that may aid in optimizing selection of therapies. Second, the pilot study was a small, single-center experience. This limits the correlation of Met scores with more clinical parameters. However, the results confirm the value and clinical utility of this novel approach—profiling PCa relevant EV-mRNA to depict metastatic status of PCa patients and dynamic change of their underlying PCa. Recent studies [51–53] have demonstrated that machine learning approaches combined with CTC- or EV-based liquid biopsies significantly enhanced sensitivity and specificity for disease detection. Larger cohorts combined with advanced machine learning approaches are warranted for further validating the precision, accuracy, and clinical performance of this assay. Given the stability of EVs in frozen plasma and the rapid processing afforded by the PCa EV DSA, this system is ready for deployment in both current and historic PCa clinical studies to validate the EV-mRNA signatures as biomarkers for disease monitoring.

4. Conclusions

In summary, we developed a novel PCa EV DSA to characterize variations in gene expression within EVs across a spectrum of PCa patients. To our knowledge, this is the first study demonstrating that EV-derived mRNAs can reflect the dynamics of the biological processes correlating to the clinical behavior of PCa. Our results warrant additional validation studies of the PCa EV DSA in a larger, multi-center cohort to facilitate its application in clinical settings. Finally, this assay may offer the opportunity to complement current tools including imaging and PSA to provide greater insights into the active biology of an underlying PCa for better personalized care.

5. Materials and methods

5.1. Cell culture

22Rv1 (ATCC CRL-2505) and LNCaP (ATCC CRL-1740) cells were purchased from American Type Culture Collection. C4-2B cells were generously provided by Dr. Leland W. K. Chung. Cells were cultured in RPMI 1640 media (catalog# 10040CV, Thermo Fisher Scientific, USA) with 10% fetal bovine serum and 100 U mL⁻¹ penicillin-streptomycin in a humidified incubator with 5% CO₂. To harvest the PCa cell line-derived EVs, cells were grown to 70% confluency in 18 Nunc™ EasYDish™ Dishes (145 cm², catalog# 150468, Thermo Fisher Scientific, USA) and incubated in serum-free media for

24–48 h. The serum-free media was then collected for harvesting cell line-derived EVs.

5.2. Harvest of PCa cell line-derived EVs by ultracentrifugation

The collected serum-free media incubated with 22Rv1, C4–2B, and LNCaP cells was first centrifuged at 300 g (4 °C) for 5 min three times and then centrifuged at 2800 g (4 °C) for 10 min to remove cells and cell debris. Afterwards, the supernatant was transferred to Open-Top Thinwall Ultra-Clear Tubes (catalog# 344058, Beckman Coulter, Inc., USA), and ultracentrifuged at 100,000 g (4 °C) for 70 min using SW 32 Ti rotor and Optima L-100 XP Ultracentrifuge (Beckman Coulter, Inc., USA). Finally, the pelleted PCa cell line-derived EVs were resuspended with 200-μL ice-cold PBS and collected.

5.3. Preparation of artificial PCa plasma samples

To prepare the artificial plasma samples mimicking plasma samples collected from PCa patients, a 50-μL aliquot of EVs derived from LNCaP cells was spiked into 200-μL male healthy donors' plasma. This was used for western blot characterization of PCa EVs throughout the purification process by EV Click Chip. Ten μL aliquots of EV pellets derived from 22Rv1, C4–2B, and LNCaP cells, respectively, were spiked into 90-μL of male healthy donors' plasma. These were used for optimization of the purification of PCa-derived EVs and validation of the 11-gene panel.

5.4. Purification of EVs from artificial plasma sample

Artificial plasma was centrifuged at 10,000g for 10 min to remove cell debris before EV purification. EV purification using an ExoQuick® ULTRA EV Isolation Kit for Serum and Plasma (category# EQUltra-20A-1, System Biosciences, USA) was carried out per the manufacturer's protocol. Ultracentrifugation was performed at 100,000 g (4 °C) for 70 min using an SW 32 Ti rotor and Optima L-100 XP Ultracentrifuge (Beckman Coulter, Inc., USA).

5.5. PCa EV Digital Scoring Assay

PCa EV Digital Scoring Assay is comprised of two major components: (1) EV Click Chip for purifying PCa-derived EVs from plasma samples, and (2) RT-ddPCR for quantifying mRNA markers from the purified PCa-derived EVs.

EV Click chip is a nanostructured microfluidic chip utilizing both immunoaffinity and click chemistry to purify the tumor-derived EVs. [18] The following three major approaches were integrated to purify EVs: (1) nanostructured substrates, (2) microfluidic chaotic mixers, and (3) covalent chemistry-mediated EV capture/release.

First, the densely packed silicon nanowires substrates (SiNWS) of EV Click Chip substantially increase the chip surface area interacting with EVs. [18] In addition, EV Click Chip's microfluidic chaotic mixer enhances repeated physical contact between the flow-through tumor-derived EVs and SiNWS. [18] These two structures work synergically to increase the capture efficiency of EVs onto the chip. Second, a pair of click chemistry motifs, i.e., tetrazine (Tz) and trans-cyclooctene (TCO), are conjugated onto SiNWS and capture antibodies (anti-EpCAM and anti-PSMA), respectively. Plasma samples are incubated with the TCO-grafted antibodies to label prostate cancer (PCa)-derived EVs. Once plasma samples are introduced to EV Click Chips, the click chemistry reaction between Tz-conjugated SiNWS and TCO-conjugated PCa-derived EVs is extremely fast, selective, irreversible, and bioorthogonal, [54] resulting in immobilization of the PCa-derived EVs onto the chip surface. Subsequent exposure to a disulfide bonds reducing agent, 1,4-dithiothreitol (DTT) [55] leads to the specific release of the PCa-derived EVs from the SiNWS by cleaving the embedded disulfide bond

between captured PCa-derived EV and EV Click Chip, which further increases the purity of the harvested EVs. Further detailed mechanisms and complete protocol of EV Digital Scoring Assay (DSA) have been described in our previous study. [18].

5.5.1. Preparation of TCO-antibody conjugates

For capturing PCa-derived EVs, TCO-antibody conjugates (i.e., TCO-anti-EpCAM and TCO-anti-PSMA) were prepared as described below. Goat anti-human EpCAM (catalog#AF960, R&D Systems, Inc., USA) and sheep anti-human PSMA (catalog#AF4234, R&D Systems, Inc., USA) antibodies were incubated with TCO-PEG4-NHS ester (catalog# A137–2, Click Chemistry Tools, USA) in phosphate-buffered saline (PBS) at room temperature for 30 min. All the TCO-antibody conjugates were freshly prepared just prior to use.

5.5.2. EV Click Chip for purification of PCa-derived EVs

First, the artificial plasma samples were incubated with individual capture TCO-antibody (5, 25, 50, 100 ng TCO-anti-EpCAM and TCO-anti-PSMA, respectively, per 100 μL artificial plasma samples, Fig. 2b–c) or dual capture TCO-antibody combination (25 ng TCO-anti-EpCAM and 25 ng TCO-anti-PSMA per 100 μL artificial plasma samples, Fig. 2d) for 30 min at room temperature to determine and validate the optimal concentrations of capture TCO-antibodies for EV Click Chips. The optimal dual capture TCO-antibody combination (25 ng TCO-anti-EpCAM and 25 ng TCO-anti-PSMA per 100 μL plasma samples) was used for clinical plasma samples in the subsequent clinical study. Second, after incubation with individual capture TCO-antibody or dual capture TCO-antibody combination, the artificial plasma samples or clinical plasma samples were then introduced into the tetrazine (Tz)-grafted EV Click Chip microfluidic devices, where the PCa-derived EVs in the plasma samples were rapidly and irreversibly captured through the click reaction between TCO and Tz. Third, to release these PCa-derived EVs from the chips, 100 μL 1,4-dithiothreitol (DTT) (50 mM) was injected into EV Click Chips at the flow rate of 1.0 mL/h to cleave the embedded disulfide bonds linking the PCa-derived EVs and EV Click Chip. Lastly, the released EVs were collected and DTT was removed before the subsequent RNA extraction process.

5.5.3. RT-ddPCR for quantification of mRNA markers from the purified PCa-derived EVs

After the removal of DTT, these purified PCa-derived EVs were then lysed with 700 μL QIAzol Lysis Reagent (catalog#79306, Qiagen, USA). RNA extraction and subsequent cDNA conversion were conducted using miRNeasy Micro Kits (catalog#217084, Qiagen, USA) and Maxima H Minus First Strand cDNA Synthesis Kits (catalog#K1652, Thermo Scientific, USA), respectively, according to the manufacturer's instructions. The resultant cDNA was aliquoted into 6 portions for the quantification of the 11 genes by duplex ddPCR on a QX200 Droplet Digital PCR System (catalog#1864001, Bio-Rad Laboratories, Inc.) according to the manufacturer's instructions. Information on primers and probes (catalog#4331182, #4448489, Thermo Scientific, USA) used for each tested gene in the RT-ddPCR assay is provided in [Supplementary Table 3](#). The transcript copies of each gene were calculated based on the counts of positive droplets using the QuantaSoft™ software.

5.6. Characterization of EV before and after purification

Characterization of EVs before and after purification was carried out following the recommendations of the International Society for Extracellular Vesicles (ISEV). [26] In brief, nanoparticle tracking analysis (NTA), western blotting analysis, and transmission electron microscopy (TEM) were conducted. Detailed descriptions of the steps for NTA, western blotting, and TEM are provided in the [Supplementary Methods](#).

5.7. Enrollment of PCa patients

All the participants in this study were enrolled in Cedars-Sinai Medical Center (CSMC) Institutional Review Board (IRB)-approved protocols (IRB #Pro00042197, #Pro00033050, #Pro00051931). Samples were collected only after the participants provided written informed consent. The samples utilized in this study were collected between January 2016 - December 2020. We included two groups of PCa patients (localized PCa [$n = 20$] and metastatic PCa [$n = 20$]) with a pathologically proven diagnosis of prostate adenocarcinoma for clinical validation of PCa EV DSA: (1) Patients with clinically localized PCa were defined as who are on active surveillance or prior to radical prostatectomy (prostate intact) without any evidence of cancer metastasis either on next-generation imaging (PSMA PET/CT, ^{68}Ga or ^{18}F -based) or conventional radiographs (CT/MRI, ^{99}Tc -single-photon emission computed tomography [SPECT] bone scan). (2) Patients with metastatic PCa were defined as who had metastatic PCa detected by conventional radiographs. Additionally, three PCa patients with serial blood collections over the course of their natural history were included to evaluate the dynamic performance of PCa EV DSA. For patients classified by PSMA PET/CT, a metastatic lesion was defined by the presence of a PET-detectable lesion with an uptake greater than background and an SUV minimum of 10. This was consistent with previous studies utilizing PSMA PET as a screening modality for detection of metastatic disease.

5.8. Clinical blood sample processing

Peripheral venous blood samples were collected from PCa patients or healthy donors in a BD Vacutainer glass tube (BD, USA) with acid citrate dextrose at CSMC. Blood samples were sequentially centrifuged at 530 g for 10 min and then 4600 g for 10 min to collect plasma within 4 h of collection. Plasma samples were stored at -80°C and thawed immediately in a 37°C water bath before use. After thawed, plasma samples were centrifuged at 10,000g for 10 min before running through the PCa EV DSA described above.

5.9. Statistical analysis

To assess the linearity of the EV recovery yields in artificial plasma samples by the PCa DSA, linear regression analysis was performed to calculate the slope and coefficient of determination (r^2). During the optimization process of PCa EV DSA, EV recovery yields were expressed as Mean $\pm$ standard deviation. Significant differences between different purification methods were evaluated using two sample t test.

Percentiles of the expression of the 11 selected genes in primary PCa tissues (from The Cancer Genome Atlas [TCGA]), [8] PCa cell lines (from Cancer Cell Line Encyclopedia [CCLE]), [31] and immune cells (from Differentiation MAP dataset [DMAP]) [32] were compared using the Wilcoxon rank-sum test. The correlation of the expression of the selected genes between PCa cell lines and their derived EVs was computed as Pearson's correlation coefficient.

The Met score of the selected 11 genes, which represents the likelihood estimate of 11-gene upregulation, was computed from the mRNA expression of these genes using a weighted Z score method. [33] After the median centering of expression data across the samples, Met scores were computed by the error-weighted mean of the expression values of the 11 genes in a sample. Briefly, for each sample, if Mean(sig) and Mean(all) are the mean expression of the genes in the signature and all the genes in the genome, respectively, and σ is the standard deviation of all the genes, then the Met score used to measure the relative expression level of the signature in the patient sample is calculated as below:

$$\text{Met score} = \frac{(\text{Mean}(\text{sig}) - \text{Mean}(\text{all})) \times \sqrt{n}}{\sigma}$$

where n is the number of genes in signature. The value of Met score is represented in units of standard deviation, which measures the differences between the gene expression of the signatures and those from the genome. We also assume that the mean of all the genes in the genome follows standard normal distribution with median-centered data, which is zero. A positive value of Met score from a patient sample indicates collective upregulation of the genes in the signature, while a negative value suggests collective downregulation. The magnitude of the Met score represents the degree of upregulation or downregulation.

ROC curve analysis was used to evaluate the performance of Met score and PSA concentration, respectively, focusing on the capability of each to distinguish localized PCa from metastatic PCa. Youden's index was used to identify the optimal cutoff of Met score and PSA concentration. The AUROC between Met score and PSA concentration was compared using the paired DeLong's test. Pearson's correlation coefficients were computed for correlation between Met score and continuous variables. A Mann-Whitney U test was used to compare the differences of Met score between categorical variables.

To assess statistical power of 40 samples in this pilot study, we performed a power analysis. Briefly, the chance of having a type I error (i.e., false negative) is less than 0.001 in this pilot study consisting of 40 patients. In addition, the sample size of 20 for each group provides 0.998 power for discriminating localized from metastatic patients using a one-sided two-sample t -test, type I error (α) of 0.05. This high power suggests the probability of having a type II error (i.e., false positive) is only 0.002.

All statistical analyses were performed using R statistical software (version 4.1.0; R Foundation, Vienna, Austria), GraphPad Prism (version 9.2.0; GraphPad Software, Inc., CA, USA), and MATLAB (Mathworks, Natick, MA, USA) with two-sided tests and a significance level of 0.05.

Financial support and sponsorship

This work is supported by United States Department of Defense (PC180192, PC171066, PC190482); and National Institute of Health (R01 CA218356, U01 CA198900, P01 CA233452, R01 CA255727, R01 CA253651, R01 CA246304, U01 EB026421, R21 CA240887, and R21 CA235340). This work has also received support from developmental funds from Cedars-Sinai Cancer. The funders had no role in the collection of data; the design and conduct of the study; management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

CRediT authorship contribution statement

Jasmine J. Wang: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Visualization, Writing – Original Draft, Writing – Review & Editing. **Na Sun:** Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing – Review & Editing. **Yi-Te Lee:** Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft, Writing – Review & Editing. **Minhyung Kim:** Data Curation, Formal Analysis, Methodology, Writing – Review & Editing. **Tatyana Vagner:** Methodology, Writing – Review & Editing. **Krizia Rohena-Rivera:** Methodology, Writing – Review & Editing. **Zhili Wang:** Investigation, Writing – Review & Editing. **Zijing Chen:** Data Curation, Writing – Review & Editing. **Ryan Y. Zhang:** Investigation, Writing – Review & Editing. **Tiffany X. Zhang:** Visualization, Writing – Review & Editing. **Howard M. Sandler:** Resources, Writing – Review & Editing. **Timothy J.**

Daskivich: Resources, Writing – Review & Editing. **Neil A. Bhowmick:** Resources, Writing – Review & Editing. **Sungyong You:** Conceptualization, Formal Analysis, Funding Acquisition, Methodology, Supervision, Writing – Review & Editing. **Dolores Di Vizio:** Resources, Supervision, Writing – Review & Editing. **Hsian-Rong Tseng:** Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Writing – Original Draft, Writing – Review & Editing. **Jie-Fu Chen:** Supervision, Writing – Review & Editing. **Yazhen Zhu:** Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Writing – Review & Editing. **Edwin M. Posadas:** Conceptualization, Funding Acquisition, Methodology, Resources, Supervision, Writing – Original Draft, Writing – Review & Editing. **All the authors:** Writing – Review & Editing.

Data availability

Data will be made available on request.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests, Hsian-Rong Tseng reports a relationship with CytoLumina Technologies Corp. that includes: equity or stocks. Hsian-Rong Tseng reports a relationship with Pulsar Therapeutics Corp. that includes: equity or stocks. Hsian-Rong Tseng, Yazhen Zhu, Na Sun has patent Covalent chemistry enables extracellular vesicle purification on nanosubstrates - toward early detection of hepatocellular carcinoma pending to Hsian-Rong Tseng, Yazhen ZHU, Na Sun, Vatche G. AGOPIAN. Corresponding author Dr. Edwin Posadas is an uncompensated advisor to CytoLumina Technologies Corp.

Acknowledgments

The authors would like to thank Dr. Leland Chung for providing C4–2B cells. The results shown in this study are part based upon data from the Human Protein Atlas (<http://www.proteinatlas.org>), The Cancer Genome Atlas (<https://www.cancer.gov/tcga>), Cancer Cell Line Encyclopedia (<https://sites.broadinstitute.org/ccle/>), and Differentiation MAP dataset (<https://rdrr.io/github/mchikina/CellCODE/man/DMAP.html>). We acknowledge the efforts of all the authors and staffs in the creation of the Human Protein Atlas, The Cancer Genome Atlas, Cancer Cell Line Encyclopedia, and Differentiation MAP dataset.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.nantod.2022.101746](https://doi.org/10.1016/j.nantod.2022.101746).

References

- R.L. Siegel, K.D. Miller, H.E. Fuchs, A. Jemal, Cancer statistics, 2022, *CA Cancer J. Clin.* 72 (1) (2022) 7–33, <https://doi.org/10.3322/caac.21708>
- R.J. Rebello, C. Oing, K.E. Knudsen, et al., Prostate cancer, *Nat. Rev. Dis. Prim.* 7 (1) (2021) 9, <https://doi.org/10.1038/s41572-020-00243-0>
- J.G. Trogdon, A.D. Falchook, R. Basak, W.R. Carpenter, R.C. Chen, Total medicare costs associated with diagnosis and treatment of prostate cancer in elderly men, *JAMA Oncol.* 5 (1) (2019) 60–66, <https://doi.org/10.1001/jamaoncol.2018.3701>
- Z. Gao, B. Pang, J. Li, N. Gao, T. Fan, Y. Li, Emerging Role of Exosomes in Liquid Biopsy for Monitoring Prostate Cancer Invasion and Metastasis, *Front Cell Dev. Biol.* 9 (2021) 679527, <https://doi.org/10.3389/fcell.2021.679527>
- H. Tukachinsky, R.W. Madison, J.H. Chung, et al., Genomic analysis of circulating tumor DNA in 3,334 patients with advanced prostate cancer identifies targetable BRCA alterations and AR resistance mechanisms, *Clin. Cancer Res* 27 (11) (2021) 3094–3105, <https://doi.org/10.1158/1078-0432.CCR-20-4805>
- E. Chen, C.L. Cario, L. Leong, et al., Cell-free DNA concentration and fragment size as a biomarker for prostate cancer, *Sci. Rep.* 11 (1) (2021) 5040, <https://doi.org/10.1038/s41598-021-84507-z>
- S. Volik, M. Alcaide, R.D. Morin, C. Collins, Cell-free DNA (cfDNA): clinical significance and utility in cancer shaped by emerging technologies, *Mol. Cancer Res* 14 (10) (2016) 898–908, <https://doi.org/10.1158/1541-7786.MCR-16-0044>
- Cancer Genome Atlas Research N, The molecular taxonomy of primary prostate cancer, *Cell* 163 (4) (2015) 1011–1025, <https://doi.org/10.1016/j.cell.2015.10.025>
- J. Dong, J.F. Chen, M. Smalley, et al., Nanostructured substrates for detection and characterization of circulating rare cells: from materials research to clinical applications, *Adv. Mater.* 32 (1) (2020) e1903663, <https://doi.org/10.1002/adma.201903663>
- D.T. Miyamoto, Y. Zheng, B.S. Wittner, et al., RNA-Seq of single prostate CTCs implicates noncanonical Wnt signaling in antiandrogen resistance, *Science* 349 (6254) (2015) 1351–1356, <https://doi.org/10.1126/science.aab0917>
- G. van Niel, G. D'Angelo, G. Raposo, Shedding light on the cell biology of extracellular vesicles, *Nat. Rev. Mol. Cell Biol.* 19 (4) (2018) 213–228, <https://doi.org/10.1038/nrm.2017.125>
- A. Nanou, M.C. Miller, L.L. Zeune, et al., Tumour-derived extracellular vesicles in blood of metastatic cancer patients associate with overall survival, *Br. J. Cancer* 122 (6) (2020) 801–811, <https://doi.org/10.1038/s41416-019-0726-9>
- Q. Ge, Y. Zhou, J. Lu, Y. Bai, X. Xie, Z. Lu, miRNA in plasma exosome is stable under different storage conditions, *Molecules* 19 (2) (2014) 1568–1575, <https://doi.org/10.3390/molecules19021568>
- E. Shtivelman, T.M. Beer, C.P. Evans, Molecular pathways and targets in prostate cancer, *Oncotarget* 5 (17) (2014) 7217–7259, <https://doi.org/10.18632/oncotarget.2406>
- S. Ferguson, R. Weissleder, Modeling EV Kinetics for Use in Early Cancer Detection, *Adv. Biosyst.* 4 (12) (2020) e1900305, <https://doi.org/10.1002/adbi.201900305>
- Z. Han, F. Wan, J. Deng, et al., Ultrasensitive detection of mRNA in extracellular vesicles using DNA tetrahedron-based thermophoretic assay, *Nano Today* (2021) 38, <https://doi.org/10.1016/j.nantod.2021.101203>
- J. Deng, S. Zhao, J. Li, et al., One-Step Thermophoretic AND Gate Operation on Extracellular Vesicles Improves Diagnosis of Prostate Cancer, *Angew. Chem.* (2022), <https://doi.org/10.1002/ange.202207037>
- N. Sun, Y.T. Lee, R.Y. Zhang, et al., Purification of HCC-specific extracellular vesicles on nanosubstrates for early HCC detection by digital scoring, *Nat. Commun.* 11 (1) (2020) 4489, <https://doi.org/10.1038/s41467-020-18311-0>
- X. Liu, W. Wu, D. Cui, X. Chen, W. Li, Functional Micro-/Nanomaterials for Multiplexed Biodetection, *Adv. Mater.* 33 (30) (2021) e2004734, <https://doi.org/10.1002/adma.202004734>
- D. Zou, W. Wu, J. Zhang, et al., Multiplex detection of miRNAs based on aggregation-induced emission luminogen encoded microspheres, *RSC Adv.* 9 (68) (2019) 39976–39985, <https://doi.org/10.1039/c9ra07680h>
- Z. Yu, S. Lin, F. Xia, et al., ExoSD chips for high-purity immunomagnetic separation and high-sensitivity detection of gastric cancer cell-derived exosomes, *Biosens. Bioelectron.* 194 (2021) 113594, <https://doi.org/10.1016/j.bios.2021.113594>
- P.T. Went, A. Lugli, S. Meier, et al., Frequent EpCam protein expression in human carcinomas, *Hum. Pathol.* 35 (1) (2004) 122–128, <https://doi.org/10.1016/j.humpath.2003.08.026>
- G. Spizzo, D. Fong, M. Wurm, et al., EpCAM expression in primary tumour tissues and metastases: an immunohistochemical analysis, *J. Clin. Pathol.* 64 (5) (2011) 415–420, <https://doi.org/10.1136/jcp.2011.090274>
- J.S. Ross, C.E. Sheehan, H.A. Fisher, et al., Correlation of primary tumor prostate-specific membrane antigen expression with disease recurrence in prostate cancer, *Clin. Cancer Res* 9 (17) (2003) 6357–6362.
- D.G. Bostwick, A. Pacelli, M. Blute, P. Roche, G.P. Murphy, Prostate specific membrane antigen expression in prostatic intraepithelial neoplasia and adenocarcinoma, *Cancer* 82 (11) (1998) 2256–2261, [https://doi.org/10.1002/\(sici\)1097-0142\(19980601\)82:11<2256::Aid-cnrc22>3.0.Co;2-s](https://doi.org/10.1002/(sici)1097-0142(19980601)82:11<2256::Aid-cnrc22>3.0.Co;2-s)
- C. Thery, K.W. Witwer, E. Aikawa, et al., Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines, *J. Extra Vesicles* 7 (1) (2018) 1535750, <https://doi.org/10.1080/20013078.2018.1535750>
- M. Uhlen, C. Zhang, S. Lee, et al., A pathology atlas of the human cancer transcriptome, *Science* 357 (6352) (2017), <https://doi.org/10.1126/science.aan2507>
- Human Protein Atlas available from <http://www.proteinatlas.org>. Accessed Aug 11, 2021.
- D.T. Miyamoto, R.J. Lee, M. Kalinich, et al., An RNA-based digital circulating tumor cell signature is predictive of drug response and early dissemination in prostate cancer, *Cancer Disco* 8 (3) (2018) 288–303, <https://doi.org/10.1158/2159-8290.CD-16-1406>
- A. Nanou, L.L. Zeune, L. Terstappen, Leukocyte-derived extracellular vesicles in blood WITH AND Without EpCAM Enrichment, *Cells* 8 (8) (2019), <https://doi.org/10.3390/cells8080937>
- J. Barretina, G. Caponigro, N. Stransky, et al., The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity, *Nature* 483 (7391) (2012) 603–607, <https://doi.org/10.1038/nature11003>
- N. Novershtern, A. Subramanian, L.N. Lawton, et al., Densely interconnected transcriptional circuits control cell states in human hematopoiesis, *Cell* 144 (2) (2011) 296–309, <https://doi.org/10.1016/j.cell.2011.01.004>
- D.M. Levine, D.R. Haynor, J.C. Castle, et al., Pathway and gene-set activation measurement from mRNA expression data: the tissue distribution of human pathways, *Genome Biol.* 7 (10) (2006) R93, <https://doi.org/10.1186/gb-2006-7-10-r93>

- [34] M.J. Barry, Clinical practice. Prostate-specific-antigen testing for early diagnosis of prostate cancer, *N. Engl. J. Med* 344 (18) (2001) 1373–1377, <https://doi.org/10.1056/NEJM200105033441806>
- [35] F.B. Thomsen, M. Westerberg, H. Garmo, et al., Prediction of metastatic prostate cancer by prostate-specific antigen in combination with T stage and Gleason Grade: Nationwide, population-based register study, *PLoS One* 15 (1) (2020) e0228447, <https://doi.org/10.1371/journal.pone.0228447>
- [36] A.K. Lee, A.V. D'Amico, Utility of prostate-specific antigen kinetics in addition to clinical factors in the selection of patients for salvage local therapy, *J. Clin. Oncol.* 23 (32) (2005) 8192–8197, <https://doi.org/10.1200/JCO.2005.03.0007>
- [37] R.B. Nadler, P.A. Humphrey, D.S. Smith, W.J. Catalona, T.L. Ratliff, Effect of inflammation and benign prostatic hyperplasia on elevated serum prostate specific antigen levels, *J. Urol.* 154 (2 Pt 1) (1995) 407–413, <https://doi.org/10.1097/00005392-199508000-00023>
- [38] L.H. Schwartz, L.R. LaTrenta, E. Bonaccio, W.K. Kelly, H.I. Scher, D.M. Panicek, Small cell and anaplastic prostate cancer: correlation between CT findings and prostate-specific antigen level, *Radiology* 208 (3) (1998) 735–738, <https://doi.org/10.1148/radiology.208.3.9722854>
- [39] M.J. Donovan, M. Noerholm, S. Bentink, et al., A molecular signature of PCA3 and ERG exosomal RNA from non-DRE urine is predictive of initial prostate biopsy result, *Prostate Cancer Prostatic Dis.* 18 (4) (2015) 370–375, <https://doi.org/10.1038/pcan.2015.40>
- [40] J. McKiernan, M.J. Donovan, V. O'Neill, et al., A novel urine exosome gene expression assay to predict high-grade prostate cancer at initial biopsy, *JAMA Oncol.* 2 (7) (2016) 882–889, <https://doi.org/10.1001/jamaoncol.2016.0097>
- [41] M. Del Re, E. Biasco, S. Crucitta, et al., The detection of androgen receptor splice variant 7 in plasma-derived exosomal RNA strongly predicts resistance to hormonal therapy in metastatic prostate cancer patients, *Eur. Urol.* 71 (4) (2017) 680–687, <https://doi.org/10.1016/j.eururo.2016.08.012>
- [42] M. Del Re, S. Crucitta, A. Sbrana, et al., AR-V7 and AR-FL expression is associated with clinical outcome: a translational study in patients with castrate resistant prostate cancer, *BJU Int.* (2019), <https://doi.org/10.1111/bju.14792>
- [43] F.H. Joncas, F. Lucien, M. Rouleau, et al., Plasma extracellular vesicles as phenotypic biomarkers in prostate cancer patients, *Prostate* 79 (15) (2019) 1767–1776, <https://doi.org/10.1002/pros.23901>
- [44] S. Chen, V. Huang, X. Xu, et al., Widespread and functional RNA circularization in localized prostate cancer, *e22, Cell* 176 (4) (2019) 831–843, <https://doi.org/10.1016/j.cell.2019.01.025>
- [45] L.B. Pinheiro, V.A. Coleman, C.M. Hindson, et al., Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification, *Anal. Chem.* 84 (2) (2012) 1003–1011, <https://doi.org/10.1021/ac202578x>
- [46] C.N. Biggs, K.M. Siddiqui, A.A. Al-Zahrani, et al., Prostate extracellular vesicles in patient plasma as a liquid biopsy platform for prostate cancer using nanoscale flow cytometry, *Oncotarget* 7 (8) (2016) 8839–8849, <https://doi.org/10.18632/oncotarget.6983>
- [47] H. Whiteland, S. Spencer-Harty, C. Morgan, et al., A role for STEAP2 in prostate cancer progression, *Clin. Exp. Metastasis* 31 (8) (2014) 909–920, <https://doi.org/10.1007/s10585-014-9679-9>
- [48] R.E. Conway, C. Rojas, J. Alt, et al., Prostate-specific membrane antigen (PSMA)-mediated laminin proteolysis generates a pro-angiogenic peptide, *Angiogenesis* 19 (4) (2016) 487–500, <https://doi.org/10.1007/s10456-016-9521-x>
- [49] S.E.A. Burnell, S. Spencer-Harty, S. Howarth, et al., STEAP2 knockdown reduces the invasive potential of prostate cancer cells, *Sci. Rep.* 8 (1) (2018) 6252, <https://doi.org/10.1038/s41598-018-24655-x>
- [50] M. Stanbrough, G.J. Bubley, K. Ross, et al., Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer, *Cancer Res* 66 (5) (2006) 2815–2825, <https://doi.org/10.1158/0008-5473.CAN-05-4000>
- [51] B. Wang, S. Zhang, J. Meng, et al., Evaporation-Induced rGO coatings for highly sensitive and non-invasive diagnosis of prostate cancer in the PSA gray zone, *Adv. Mater.* 33 (40) (2021) e2103999, <https://doi.org/10.1002/adma.202103999>
- [52] R. Zarnowski, A. Noll, M.G. Chevrette, et al., Coordination of fungal biofilm development by extracellular vesicle cargo, *Nat. Commun.* 12 (1) (2021) 6235, <https://doi.org/10.1038/s41467-021-26525-z>
- [53] L. Min, B. Wang, H. Bao, et al., Advanced nanotechnologies for extracellular vesicle-based liquid biopsy, *Adv. Sci. (Weinh.)* 8 (20) (2021) e2102789, <https://doi.org/10.1002/advs.202102789>
- [54] M.L. Blackman, M. Royzen, J.M. Fox, Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity, *J. Am. Chem. Soc.* 130 (41) (2008) 13518–13519, <https://doi.org/10.1021/ja8053805>
- [55] E.A. Smith, M.J. Wanat, Y. Cheng, S.V.P. Barreira, A.G. Frutos, R.M. Corn, Formation, spectroscopic characterization, and application of sulfhydryl-terminated alkanethiol monolayers for the chemical attachment of DNA onto gold surfaces, *Langmuir* 17 (8) (2001) 2502–2507, <https://doi.org/10.1021/la001064q>