



Combining Spatial Transcriptomics, Pseudotime, and Machine Learning Enables Discovery of Biomarkers for Prostate Cancer

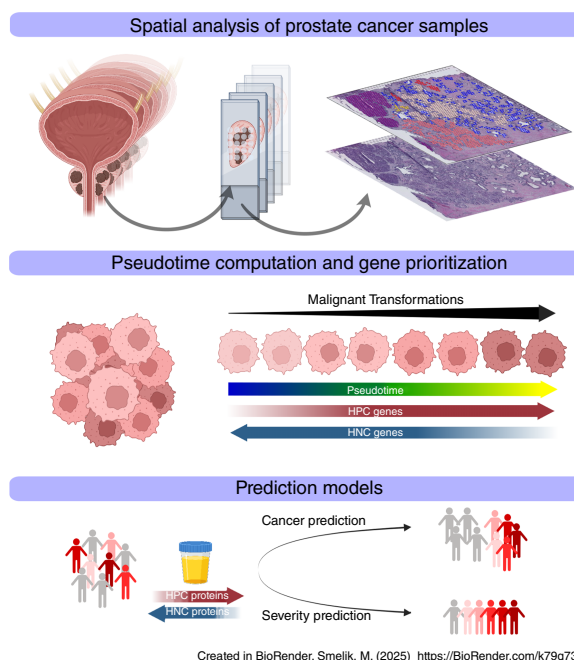
Martin Smelik¹, Daniel Diaz-Roncero Gonzalez¹, Xiaojing An², Rakesh Heer³, Lars Henningsohn¹, Xinxiu Li¹, Hui Wang⁴, Yelin Zhao¹, and Mikael Benson¹

ABSTRACT

Early cancer diagnosis is crucial but challenging owing to the lack of reliable biomarkers that can be measured using routine clinical methods. The identification of biomarkers for early detection is complicated by each tumor involving changes in the interactions between thousands of genes. In addition to this staggering complexity, these interactions can vary among patients with the same diagnosis as well as within the same tumor. We hypothesized that reliable biomarkers that can be measured with routine methods could be identified by exploiting three facts: (i) the same tumor can have multiple grades of malignant transformation; (ii) these grades and their molecular changes can be characterized using spatial transcriptomics; and (iii) these changes can be integrated into models of malignant transformation using pseudotime. Pseudotime models were constructed based on spatial transcriptomic data from three independent prostate cancer studies to prioritize the genes that were most correlated with malignant transformation. The identified genes were associated with cancer grade, copy-number aberrations, hallmark pathways, and drug targets, and they encoded candidate biomarkers for prostate cancer in mRNA, IHC, and proteomics data from the sera, prostate tissue, and urine of more than 2,000 patients with prostate cancer and controls. Machine learning-based prediction models revealed that the biomarkers in urine had an AUC of 0.92 for prostate cancer and were associated with cancer grade. Overall, this study demonstrates the diagnostic potential of combining spatial transcriptomics, pseudotime, and machine learning for prostate cancer, which should be further tested in prospective studies.

Significance: Integrating spatial transcriptomics, pseudotime, and machine learning analyses is effective for identifying prostate

cancer biomarkers that are reliable in different settings and measurable with routine methods, providing potential early diagnosis strategies.



Created in BioRender. Smelik, M. (2025) <https://BioRender.com/k79g739>

This article is part of a special series: Driving Cancer Discoveries with Computational Research, Data Science, and Machine Learning/AI.

Introduction

Early cancer diagnosis may significantly reduce mortality but is complicated by the lack of accurate biomarkers that are easy to measure in the clinic. For example, PSA is a widely used biomarker for screening

prostate cancer. Although the use of PSA is associated with decreased mortality, its lack of sensitivity and specificity has resulted in many unnecessary biopsies or missed prostate cancer diagnoses (1, 2). This has led to significant research efforts to identify accurate biomarkers that

¹Division of ENT Diseases, Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institute, Stockholm, Sweden.

²Department of Pathology, Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing, China. ³Department of Surgery, Imperial College London, London, United Kingdom. ⁴Jiangsu Key Laboratory of Immunity and Metabolism, Department of Pathogen Biology and Immunology, School of Basic Medical Science, Xuzhou Medical University, Xuzhou, China.

M. Smelik, D. Diaz-Roncero Gonzalez, Y. Zhao, and M. Benson contributed equally to this article.

Corresponding Authors: Martin Smelik, Department of Clinical Science, Intervention and Technology (CLINTEC), Karolinska Institute, Stockholm 14186, Sweden. E-mail: martin.smelik@ki.se; Mikael Benson, mikael.benson@ki.se; and Yelin Zhao, yelin.zhao@ki.se

Cancer Res 2025;85:2514–26

doi: 10.1158/0008-5472.CAN-25-0269

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2025 The Authors; Published by the American Association for Cancer Research

can be measured in peripheral blood using routine methods for prostate cancer and other cancers (3–7). Examples of conventional biomarkers to diagnose prostate cancer include the Prostate Health Index. This measures three types of PSA and has AUCs ranging from 0.7 to 0.77, which are superior to PSA (0.56–0.6; refs. 8, 9). AUCs may be further improved by combining biomarkers with imaging. Multiomics analyses of local tumor tissues have led to the identification of candidate biomarkers with very high AUCs (10–13). However, when measured in peripheral blood, these biomarkers had variable AUCs (14).

A potential solution could be to measure biomarkers in local body fluids, such as urine (10–13). By combining urine biomarkers with clinical factors, or imaging, accuracy has been further improved (AUCs of 0.78 and 0.89; ref. 15). Recent proteomic analyses of urine have resulted in the identification of more than 200 proteins that collectively reached AUCs ranging from 0.71 to 0.81 for diagnosing prostate cancer (16). Another study (14) based on transcriptomics showed that only 18 urine mRNAs resulted in an AUC of 0.82 for prostate cancer.

Despite the high AUCs, the authors recognized important limitations. The analyses were based on bulk data, in which important signals in the data may be buried. This limitation is important because prostate cancer and other cancers involve altered interactions between multiple cell types in the same tumor and not only between tumor cells (14).

Another limitation was that the biomarkers were prioritized using data-driven methods. This did not explain the selection of biomarkers nor their relative mechanistic importance compared with many others. The clinical relevance of such explainability is increasingly recognized (17). First, these limitations may be important if biomarkers cannot be replicated in different clinical settings, which can vary greatly (18). If so, analyses of variations in the most important mechanisms may help to trace patient characteristics underlying such variations. Second, prioritization of the most important mechanisms, such as pathways, may also have implications for precision medicine, in that such mechanisms may include drug targets and biomarkers for those drugs. Third, clinical implementation of novel diagnostic and therapeutic methods will require explainable solutions for informed decisions by clinicians and patients.

However, emerging technologies may provide solutions to systematically prioritize biomarkers based on an understanding of their relative importance. Single-cell RNA sequencing (scRNA-seq) and spatial transcriptomics (ST) can measure genome-wide changes in mRNA expression in thousands of individual cells (scRNA-seq), as well as their location in a tumor (ST; ref. 19).

Such analyses have shown changes in the expression of thousands of genes across tumor cells and many other cell types. These changes can vary between different locations in the same tumor, time points in the same patient, as well as between patients with the same cancer diagnosis (20–25). A computational method called pseudotime (PT) can systematically organize cells in a linear trajectory based on the similarity of their expression profiles. PTs derived from bulk, scRNA-seq, and ST analyses of prostate cancer and other tumors, as well as cancer cell lines, support the idea that PTs model malignant transformations (20, 26–29).

This led us to hypothesize that the genes most correlated with PT in most tumors are reliable biomarkers. In support of our hypothesis, we found that PT was associated with the histologic grade of malignancy, copy-number aberrations (CNA), cancer-related pathways, and potential drug targets.

We identified 45 genes that were highly positively or negatively correlated with PT trajectories. The diagnostic relevance of the

mRNA and protein products of these genes was supported by analyses of sera, prostate tissue, and urine samples from more than 2,000 patients with prostate cancer and controls. As protein analyses are more readily performed with routine clinical methods, we used machine learning to construct prediction models based on proteins for prostate cancer diagnosis. Those models reached an AUC of 0.69 in sera and 0.92 in urine. This considerably outperformed the models based only on serum PSA levels. Furthermore, we found that these biomarkers outperformed serum PSA levels in predicting cancer grades. Hence, we propose that prospective studies of candidate biomarkers are warranted and that analytic principles are tested to identify biomarkers in other cancers. To the best of our knowledge, combining ST, PT, and machine learning to identify biomarkers is a novel approach. Therefore, we made the data and codes freely available for clinical studies on the diagnostic potential of our strategy for prostate cancer and other cancers.

Materials and Methods

Spatial data retrieval and sample annotation

We retrieved ST data from the HEST-1k dataset (30), which contains more than 1,000 ST profiles, including 320 cancer samples from 25 cancer types. Human prostate cancer samples with at least 1,000 spots were filtered. To avoid overrepresentation of a single slide, we excluded samples from a study (31), which analyzed the technical replicates of sample MEND156 from the author's previous study (32). The resulting samples were annotated by an expert pathologist.

To annotate the retrieved data, we used Semla software in R (33). Semla facilitates the annotation of Seurat objects containing spatial data, allowing the selection and annotation of specific spots for later analysis. Histologic annotations were performed on all hematoxylin and eosin–stained whole-slide images, categorizing spots into normal glands, prostatic intraepithelial neoplasia (PIN), and cancer. The cancer spots were further divided into five grades according to the Gleason score in the fifth edition of the World Health Organization classification of male and genitourinary tumors from 2022. Only spots identified by pathologist as normal, PIN, or cancer were subsequently analyzed. Once the labeling process was completed, the annotated Seurat object was saved for subsequent analyses. The workflow for the spatial data processing is shown in Supplementary Fig. S1.

Data preprocessing and outlier removal

To focus on malignant transformation, we retained samples containing both benign and cancerous cells, excluding spots that were predominantly composed of non-epithelial cells. Noisy spots and genes were filtered by removing spots with fewer than 100 genes and genes expressed in fewer than three spots. Additionally, spots with more than 20% mitochondrial genes were excluded. The data were log-normalized and scaled, and highly variable genes were identified and annotated for further analyses.

Principal component analysis (PCA) was used to identify outlier spots and reduce data dimensionality by calculating the orthogonal components. First, PCA was performed for each sample considering the distribution of highly variable genes. Outlier spots were removed by retaining only those between the 0.0025 and 0.9975 quantiles across the first 10 components. After outlier removal, reduction by PCA was recomputed to reduce dimensionality, retaining the first 10 components for downstream analyses.

All steps were performed using the Scanpy (1.9.8) Python (RRID: SCR_008394, v 3.12.8) package with `normalize_total`, `log1p`, `pca`,

and scale. All functions were applied with default parameters, unless stated otherwise.

PT

We used PT to track prostate cancer progression in ST data. PT reflects cellular progression by calculating the distances among the diffusion vectors of each spot. The analysis followed the workflow recommended in the Scanpy package. First, Louvain clustering (34) with a resolution of 0.5 was used to cluster the spots. Small clusters with fewer than 20 spots were excluded from further analyses. The rationale behind this is that smaller clusters can be faulty, which could be because of lower cell quality or quantity in that region or a higher ratio of non-epithelial cells. To construct PTs, we first identified a root spot, which served as a reference point, representing a benign state. To provide a more accurate identification of the root spot, the cluster with normal spots was used as an input to the `set_root` function from the `stlearn` package (35). This method uses a semiautomated approach to identify the root spot based on the input cluster. The transitions between the spots were assessed using diffusion-like random walks (36). Shortly, the probability of transitioning between two spots depends on the similarity of their gene expression profiles. The diffusion map was computed using a `diffmap` function. The motivation for using diffusion map was mainly its ability to handle highly complex, non-linear data as well as being more robust to handle noise, compared with other alternatives, such as PCA. Each spot was assigned a PT value ranging from 0 to 1. PT (37) was implemented using the `dpt` function in the Scanpy package.

CNA analysis

To analyze the CNA of the spots in the spatial samples, we used a CopyKAT (38) method. The benign cells were used as reference, and we used Pearson correlation as a distance function. We used default parameters to run the analysis. The final heatmaps were reordered using the hierarchical clustering with Ward agglomeration.

Enrichment analysis

Gene set enrichment analysis (39) is a method used to identify sets of genes that are overrepresented within a known list of biologically relevant genes. GSEapy (40) was used to identify pathways and drugs for which the associated genes were overrepresented in our list of relevant genes.

The associated pathways were identified using the MSigDB_Hallmark_2020 database, which stores the gene sets associated with specific pathways. Only significantly represented pathways were retained for further analysis (adj. *P* value < 0.05).

Potential drugs were identified using DGIdb_Drug_Targets_2024, which stores gene sets associated with specific drugs. Only significantly enriched drugs were retained (adj. *P* value < 0.05).

scRNA-seq analysis

To analyze scRNA expression in patients with prostate cancer, we analyzed a publicly available dataset from a study on the heterogeneity of tumor-associated epithelial cells in prostate cancer (41). The dataset included scRNA expression data from 11 patients, comprising six biopsies from three patients (two biopsies per patient) and 12 radical prostatectomy (RP) samples from eight patients. Of these, four patients provided paired RP samples (one from a tumor-only section and one from a normal region), whereas the remaining four patients provided a single tumor-only RP sample. Briefly, the samples were sequenced using the SeqWell S³ protocol (42) and pre-processed using the DropSeq (43) workflow.

Differential expression analysis between malignant and nonmalignant cells was performed using the FindMarkers function of the Seurat library in R. The log₂ fold change was computed for all genes with an adjusted *P* value below 0.05. Each sample was analyzed separately. The results were saved and plotted in Python (RRID: SCR_008394, v. 3.12.8) using the Seaborn and Matplotlib (RRID: SCR_008624) packages.

IHC of SPON2 in healthy prostate tissues, benign prostate hyperplasia, and prostate cancer

Eleven prostate biopsy specimens, including cancer cells of three different Gleason grades, matched adjacent normal glands, and hyperplastic glands, were collected at the Xiyuan Hospital of China Academy of Chinese Medical Sciences from 2019 to 2024. All pathologic slides were independently reviewed and verified by two trained pathologists according to the Gleason score in the fifth edition of the World Health Organization classification of male and genitourinary tumors from 2022.

Paraffin-embedded tissues were sectioned (4 μm) and prepared. The tissue sections were placed in an incubator at 72°C for 40 minutes for deparaffinization, followed by dewaxing in xylene three times for 10 minutes. Rehydration was performed using a graded ethanol series (100%, 100%, 95%, and 75%) for 2 minutes at each concentration. The sections were washed five times with PBS (pH 7.3) for 2 minutes, and antigen retrieval was performed using EDTA (pH 9.0) at a high pressure and 120°C for 2.5 minutes. The sections were immediately cooled to room temperature and washed with PBS. Endogenous peroxidase activity was eliminated by incubating sections in 3% H₂O₂ for 10 minutes at room temperature. The sections were then incubated with SPON2 antibody (diluted 1:600; Proteintech, cat. #20513-1-AP, RRID: AB_2878696) for 1 hour at 37°C. The sections were then incubated with anti-mouse/rabbit IgG for 30 minutes at room temperature. After three washes, immunoreactivity was visualized using diaminobenzidine (DAKO) for 3 minutes.

Three random digital images at ×200 magnification per specimen were photographed using an Olympus BX41 fluorescence microscope (Olympus) to calculate the average optical density (AOD). The digital images were obtained using a digital pathology system (NanoZoomer 2.0-RS slide scanner, Hamamatsu) at ×400 magnification. AOD [AOD = integrated optical density (IOD)/area] was calculated using Image-Pro Plus v6.0 software (Media Cybernetics; RRID: SCR_007369) and was used in this study for statistical analysis.

Bulk RNA analysis

Two bulk RNA-seq datasets were used in this study: one from local tissue and one from urine.

We retrieved the bulk RNA expression of 1,223 prostate samples from a comprehensive prostate cancer transcriptome atlas (44). This dataset, constructed by batch correction of 11 studies, was filtered to retain 714 primary cancer samples and 174 normal prostate tissue samples for further analysis. The workflow for bulk RNA expression processing is shown in Supplementary Fig. S2.

The differentially expressed genes between primary cancer and normal samples were identified using a Student *t* test, *P* values were adjusted for multiple testing using the Benjamini-Hochberg method, and the log₂ fold change was computed. Final plots were constructed using the Matplotlib package (RRID: SCR_008624) in Python (RRID: SCR_008394, v. 3.12.8).

Urine bulk RNA expression was retrieved from a recent prostate cancer study (14). Briefly, a first-catch urine sample following digital rectal examination (DRE) was obtained and analyzed using

OpenArray Technology (Thermo Fisher Scientific) in the original study. For our analysis, data were retrieved, focusing on the columns “Differential Expression (Cancer vs. Normal) – logFC” and “Differential Expression (Cancer vs. Normal) – adj. P.Val.” Only candidate biomarkers with significant differential expression (adj. *P* value <0.05) were included. The log₂-fold changes in these biomarkers were extracted and visualized using Matplotlib (RRID: SCR_008624) in Python (RRID: SCR_008394, v3.12.8).

Tissue proteomics analysis

We analyzed the prostate proteomic data of 306 patients from a previous study (45). Briefly, three surgical samples were collected from each patient: one corresponding to benign tissue and the other two corresponding to primary and secondary Gleason score grading areas, respectively. The samples were annotated independently by two pathologists, and the protein abundance was determined using data-independent acquisition proteomics technology. Analogous to bulk RNA analysis, Student *t* test and the Benjamini–Hochberg method were applied to identify differentially expressed proteins. The log₂ fold change was computed to estimate the level of change in expression between benign and malignant tissues for the candidate biomarkers.

UK Biobank proteomics

We analyzed data from the UK Biobank (UKBB), a prospective cohort study comprising more than 500,000 individuals recruited from the United Kingdom (46). In this analysis, proteomics data as well as clinical information, including the date of recruitment, date of diagnosis, and diagnostic codes based on the International Classification of Diseases, 10th Edition (ICD-10), were utilized. Patients with prostate cancer were identified using ICD-10 code C61, with inclusion restricted to individuals diagnosed at the time of sampling. Healthy controls were defined as those without ICD-10 codes during the follow-up period. The MatchIt (47) package was used to match patients with prostate cancer with healthy controls, with age as the matching criterion.

Blood samples for proteomic analysis were collected during the baseline assessment conducted between 2006 and 2010. The follow-up period was October 31, 2022. Proteomic profiling of blood plasma samples was performed using the Olink Explore 1536 platform, which captures 2,923 unique proteins.

Urine proteomics

Urine proteomics data were analyzed using the results of a recent study (16). Briefly, urine samples from patients with cancer and BPH controls were collected before and after DRE. After centrifugation, the extracellular vesicles of urine were isolated, and proteomic analysis was performed using mass spectrometry to determine protein abundance. The post-DRE urine centrifuged at 20,000 × *g* for 30 minutes was reported to be the most accurately reflected prostate tissue proteome and was therefore used in our analysis. Serum PSA levels were measured in all the patients. Datasets from both the discovery and validation datasets were included in our analysis, which included 123 patients with cancer and 36 BPH controls. The workflow for urine proteomic processing is presented in Supplementary Fig. S3.

Construction of prediction models

Machine learning approaches were employed to build predictive models for the proteomic datasets described above. The same pipeline was used in both studies.

Proteins with more than 20% missing values were excluded from analysis. Data were divided into training (70%) and test (30%) sets, and

missing data (2% for urine proteomics and 8% for UKBB data, respectively) were imputed using the MissForest model. Next, we used ridge regression to determine the optimal parameter. The β coefficients from the model were used to assess the importance of the candidate biomarkers. For cancer prediction, a binomial model was used with the candidate biomarkers as features, and cancer status was used as the response variable. For grade prediction, a Gaussian model was used with the candidate biomarkers as features and the transformed [healthy = 0, grade group 1 (GG1) = 1, GG2 = 2, GG3 = 3, and GG4 = 4] grade as a response variable. All models were constructed using training data, and the test data were then used to evaluate the predictions of cancer status and progression. The baseline model was computed using linear regression with PSA values as the feature variable. To evaluate the results, the AUCs were used for cancer prediction. For grade prediction, correlation, root mean squared error (RMSE), and R-squared errors were used. We used bootstrapping to estimate the uncertainty of the AUCs and ROC curves. To do that, we randomly sampled the patients for the training of the models in a stratified manner for 1,000 times. The results were evaluated using all the patients that were not used in the training of the models. The mean (IQR) values for both AUC and ROC curves are presented in the results. In order to analyze whether the candidate biomarkers overperform the other genes or proteins, we used the same prediction pipeline for randomly selected genes/proteins. We selected the same number of genes/proteins as the number of our candidate biomarkers captured in the datasets. We repeated this analysis 1,000 times to get a distribution of results with randomly selected proteins.

The missForest, missForestPredict, glm, and cv.glmnet functions of the missForestPredict and glmnet R packages were used.

Survival analysis

From UKBB, 340 patients with prostate cancer were followed from the assessment until the end of the follow-up period (October 31, 2022). We used the Xena tool (48) to download the information about the progression-free interval (PFI) and the follow-up period of the patients from The Cancer Genome Atlas and linked them using the patient ID with the tissue bulk RNA data used in the article. The missing proteomics data from UKBB were excluded (>20% of missing values) and imputed using the same approach as in the prediction models. We used the Cox proportional hazards model from the survival (RRID: SCR_021137) package in R with the proteins/genes as covariates and the overall survival (OS)/PFI as the response variable.

Ethics statement

The study was conducted in accordance with ethical guidelines of the Declaration of Helsinki. Analyses of samples from UKBB were based on ethical approval from the National Information Governance Board for Health and Social Care and the National Health Service Northwest Multi-Center Research Ethics Committee. All participants gave written informed consent through electronic signatures before enrolling in the study. This research has been conducted under approved UKBB project ID 102162. The analyses of prostate tissues with IHC were approved by the Ethics Committee of Xiyuan Hospital, China Academy of Chinese Medical Sciences, Beijing, after obtaining written informed consent from the patients.

Data availability

The scRNA data used in this study are publicly available in the Gene Expression Omnibus under the identifier GSE176031, the bulk

RNA data are publicly available in the BioStudies database under the identifier E-MTAB-9930, the tissue proteomics data are publicly available in iProX under the identifier IPX0005748000, the urine proteomics data are publicly available in MassIVE under the identifier MSV000092061, and the UKBB data are available at <https://www.ukbiobank.ac.uk/> for approved researchers through the UKBB data access protocol. IHC figures are available at Zenodo: 10.5281/zenodo.14699624. All other raw data are available upon request from the corresponding author.

This article includes code hosted on Code Ocean, a computational reproducibility platform that allows users to view,

modify, run, and download the code: <https://codeocean.com/capsule/3933609/tree/v1>.

Results

Patient summary and histologic annotation of the spots

To study prostate cancer progression, we used publicly available ST samples from a repository of ST data (30). Histologic annotations were performed by a pathologist who independently examined all the hematoxylin and eosin–stained whole-slide images, categorizing spots into normal glands, PIN, and cancer. The cancer spots

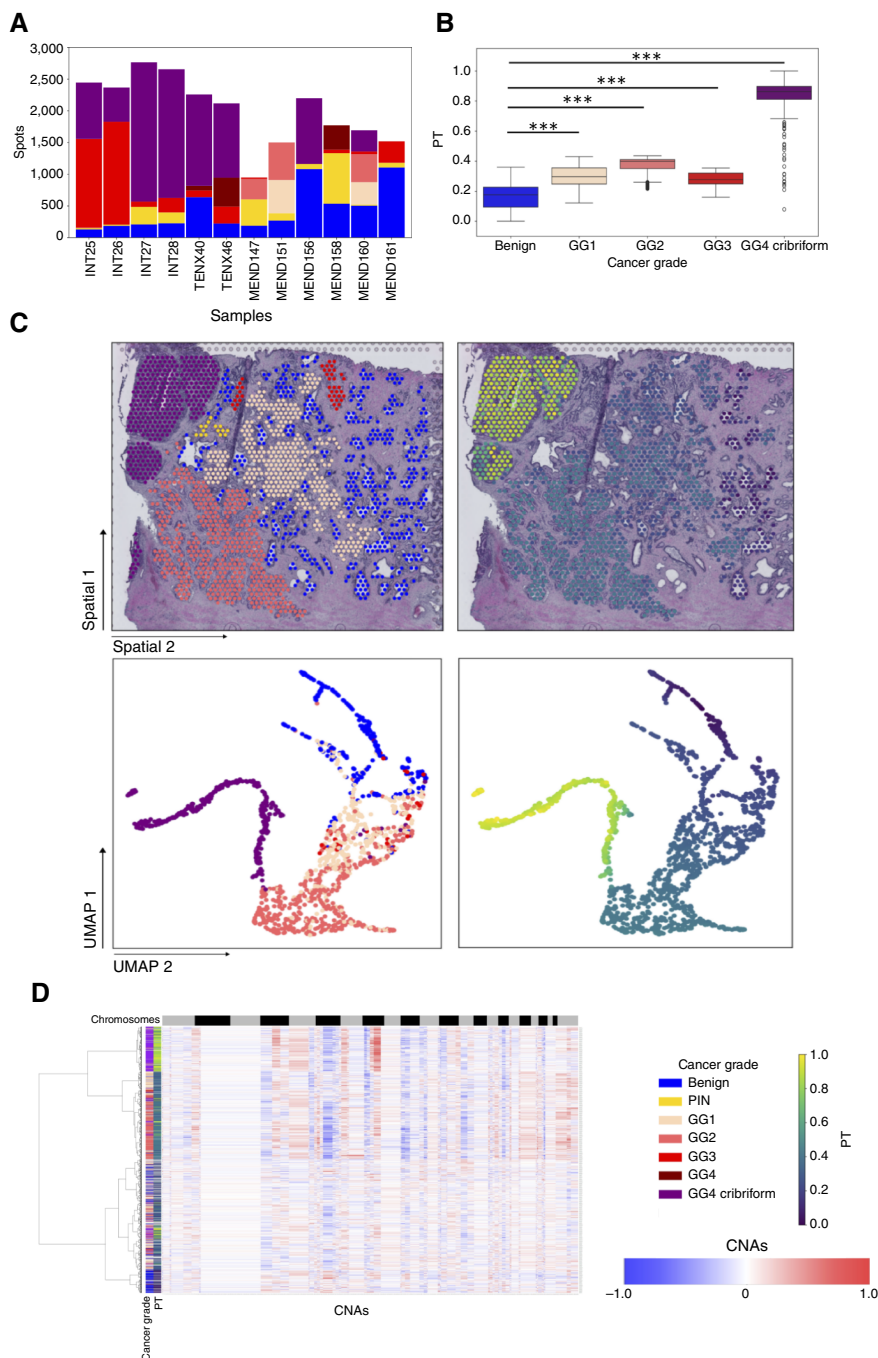
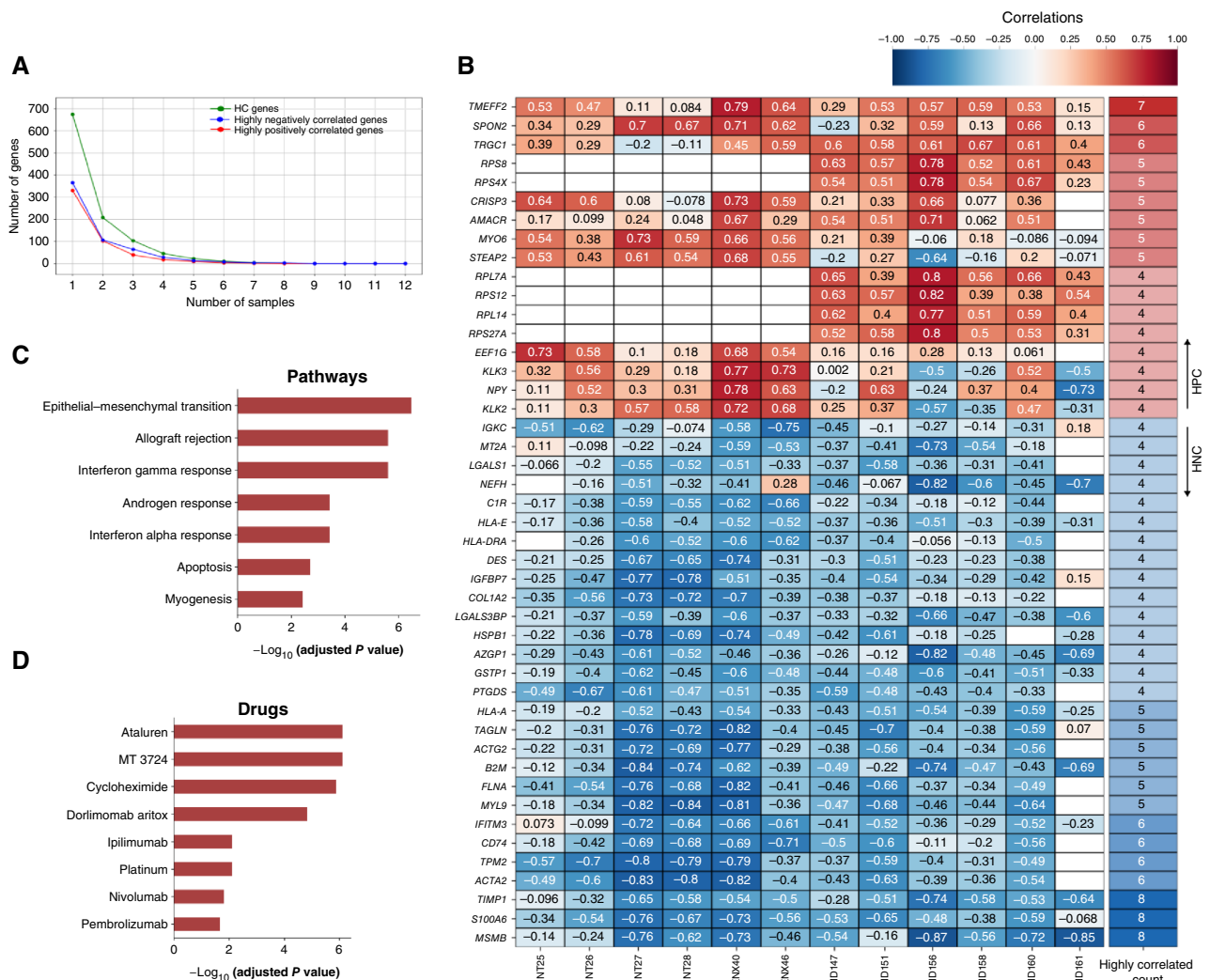


Figure 1.

PT is associated with histologic and molecular development of prostate cancer. **A**, A histogram showing the distribution of cancer grade per sample. **B**, Boxplots showing the PT distribution in different cancer grades in one representative sample (MEND160). The centerline corresponds to the median, the box limits to the upper and lower quartiles, the whiskers to the 1.5× IQR, and the points to outliers. ***, $P < 0.001$ as assessed by the Mann-Whitney U test with Holm method-adjusted P values. **C**, The spatial plots and Uniform Manifold Approximation and Projection (UMAP) plots in MEND160. Each dot corresponds to one spot. The color of the spots in the left figure corresponds to cancer grade and to PT in the figure on the right. **D**, CNA analysis in MEND160. Chromosomes are given on x-axis and spots on the y-axis are ordered using unsupervised hierarchical clustering. The colors on the bars next to the heatmap represent histologic annotations and PT, respectively. Plots **B–D** were constructed using the spots from sample MEND160. Color labels for plots **A–D** are presented on the bottom right area of the figure.

**Figure 2.**

Prioritization and pathway analysis of genes that were highly correlated with PT. **A**, Number of unique genes that were highly correlated with PT. The x-axis corresponds to the number of samples. **B**, The correlation coefficients of the genes with the PT. Only genes that were consistently either positively (red) or negatively (blue) correlated in at least four samples are shown. The HC count corresponds to the number of samples with the correlation coefficient greater than 0.5 or less than -0.5 in red or blue, respectively. HNC, highly negatively correlated; HPC, highly positively correlated. **C**, Enrichment of hallmark pathways among the 45 HC genes. **D**, Drug enrichment analysis of the 45 HC genes.

were further subcategorized into five grades based on the Gleason score criteria. After quality control and removal of outliers for each sample, we retained ST samples from six patients (see Materials and Methods; Supplementary Fig. S1). The number of samples from each patient ranged from one to four, which resulted in the analysis of 12 ST samples. The patient summary is provided in Supplementary Data S1, and the sample characteristics and inclusion criteria are provided in Supplementary Table S1. A total of 19,695 spots were included in the final analysis, representing an average of 1,641 spots per sample and 4,800 unique genes per spot. The distribution of different cancer grades varied greatly, ranging from predominantly benign samples to those with a high number of malignant sites (Fig. 1A). This led us to address the main question of this study, i.e., finding reliable biomarkers given the complexity and variability of cellular and molecular changes.

PT is associated with histologic and molecular changes in prostate cancer

PT is a quantitative metric that organizes cells into trajectories based on expression similarity. Here, we used PT to construct cellular trajectories starting with the expression profiles of spots in histologically normal tissue from prostate cancer. Each spot included 1 to 10 cells. Each spot was assigned a value of 0 to 1. A low value, close to 0, indicates a profile close to the normal tissue, whereas a value close to 1 indicates spots for which expression profiles differed most from those in the normal tissue. We hypothesized that increased PT would be associated with the progression of malignant transformations as measured by histologic changes and CNAs.

We used a semiautomated approach (35) to identify benign cellular spots that represented the initial state of the developmental trajectory

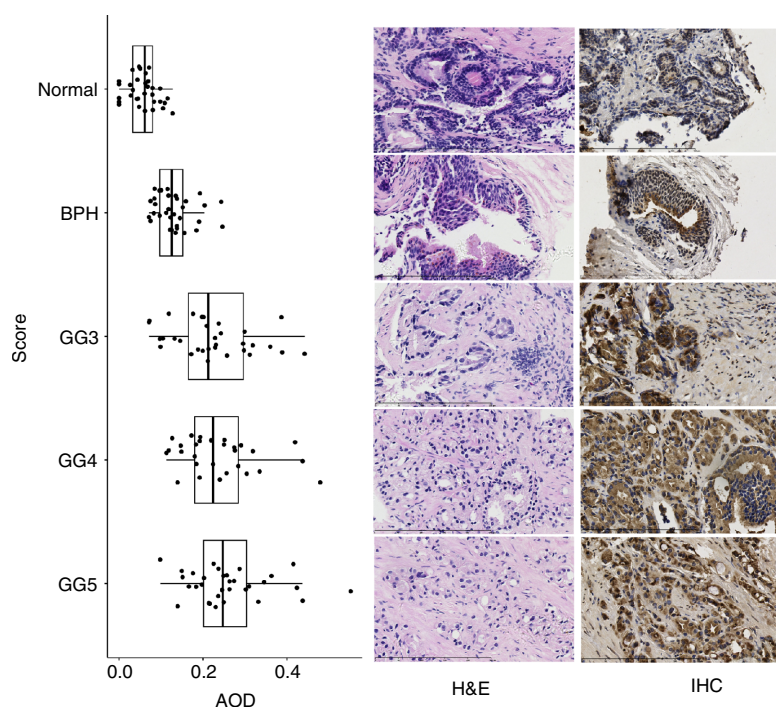


Figure 3.

SPON2 IHC in healthy prostate, BPH, and prostate cancer. SPON2 AOD and representative samples stained with hematoxylin and eosin (H&E) and SPON2 (IHC) for normal, BPH, and prostate cancer samples. The centerline corresponds to the median, the box limits to the upper and lower quartiles, the whiskers to the $1.5 \times$ IQR, and the points correspond to samples.

and computed the PT for each sample separately. A representative example of the correlation between PT and malignant transformation is shown in **Fig. 1B–D**. These results show the correlations between the resulting PT and histologic annotations for sample MEND160, both spatially and using the Uniform Manifold Approximation and Projection. Supplementary Figs. S4–S14 provide visualizations for all other samples. We found that in each sample, PT was significantly correlated with cancer progression based on histologic annotations (P value < 0.05 , Supplementary Data S2, Supplementary Fig. S15). We also found that groups of spots predicted as aneuploid by CopyKAT (38) had significantly (P value < 0.05 , Mann–Whitney U test) higher PT than spots predicted as diploid in all samples, except MEND147 (Supplementary Data S3).

Taken together, the associations between PT and histologic changes, as well as the CNA analysis, supported that PT provided a higher-order representation of malignant transformations.

The genes most correlated with PT were associated with cancer hallmark pathways and enriched for drug targets

As PT was associated with malignant transformations, as measured by histologic changes and CNAs, we hypothesized that the genes most consistently correlated with PT would be promising biomarkers. To identify such genes, we computed the Spearman correlation coefficient between each gene and PT for each sample. We found that the number of highly correlated (HC; absolute value of correlation higher than 0.5) genes varied greatly between the samples, ranging from 12 to 436, with a median and IQR value of 48.50 (45.00–70.25; Supplementary Fig. S16). Despite the great variations, we found that many HC genes were present in multiple samples. Of the 674 HC genes, more than 30% were shared by at least two samples (**Fig. 2A**). Therefore, we applied the elbow method (**Fig. 2A**) to identify genes that were consistently highly expressed in at least four samples. The elbow method is a practical approach used to estimate the optimal number of samples needed

for an analysis. Once this optimal number is reached, including additional samples is not likely to significantly change the number of genes identified or the overall results of the study. This approach yielded 45 genes, of which, 17 were highly positively correlated and 28 were highly negatively correlated (**Fig. 2B**).

Considering that the ST analyses did not have single-cell resolution, we next analyzed whether the HC genes were expressed in epithelial cells using the scRNA-seq data. We analyzed 21,743 single cells from 18 samples, six of which were biopsies from three different patients and 12 from RPs from eight different patients. We computed the differentially expressed genes between malignant and nonmalignant epithelial cells in each patient separately, as well as all patients together. We found that the direction of log fold change aligned with our PT analysis in 29 of 31 differentially expressed candidate biomarkers (Supplementary Fig. S17). This independent analysis supported the relevance of the HC genes. Moreover, a literature review showed that 30 of the 45 HC genes had been associated with malignant transformation, and 11 of the HC genes were specifically associated with prostate cancer (Supplementary Note S1). To obtain a functional overview of all the 45 potential biomarkers, we analyzed their enrichment in the molecular signatures database of hallmark pathways using gene set enrichment analysis. We found that the most significant pathways included epithelial–mesenchymal transition and other prostate cancer–related pathways, such as androgen response and apoptosis (**Fig. 2C**).

Next, we evaluated whether HC genes were significantly enriched for drug targets based on The Drug Gene Interaction Database (49). This resulted in eight significant drugs, of which, six were relevant or potentially relevant for prostate cancer: one of the drugs, pembrolizumab had been already approved for treating prostate cancer (50). Three drugs had been or are currently being tested in clinical trials for prostate cancer, namely ipilimumab (51), platinum (52), and nivolumab (53). Two others have demonstrated *in vitro*

inhibitory effects on cancer cells, namely ataluren (54) and cycloheximide (55) (Fig. 2D). These findings further supported the relevance of the prioritized genes. Therefore, these 45 HC genes will be referred to as candidate biomarkers.

IHC shows correlation between mRNA and protein changes in prostate tissues

The local protein expression of several candidate biomarkers had been previously analyzed by IHC or tissue microarrays in prostate cancer, including transmembrane protein with EGF-like and two follistatin-like domains 2, Spondin-2 (SPON2), β -microseminoprotein, S100 calcium-binding protein A6, and TIMP metalloproteinase inhibitor 1 (a literature-based summary of the candidate biomarkers in relation to malignant transformation is given in the Supplementary Note S1; refs. 56–61).

These analyses revealed protein changes in the same direction as the mRNA changes described above. However, a striking finding in all of these studies was the great inter-patient variability in mRNA expression. This had been observed in spatial and single-cell data and was also found at the protein level in IHC and tissue microarray analyses. Such variations between patients with prostate cancer could confound analyses of differences between patients and controls. However, the variations could also have a clinically relevant explanation. The analyses above showed that the HC genes were associated with malignant transformation. Thus, the observed variability could imply that the candidate biomarkers had diagnostic potential for prostate cancer grade. To analyze this potential, we focused on SPON2 because it is thought to have pivotal roles in cancer progression, by promoting growth, development, and cell proliferation. Also, several studies support its potential as a biomarker (62). In our analyses, it was one of the genes most positively correlated with PT. This led us to perform IHC analyses of SPON2 in multiple biopsies from 11 patients (Supplementary Data S1). This encompassed 32 normal, 33 BPH, 32 GG3, 29 GG4, and 32 GG5 samples (Supplementary Table S2). These analyses showed no or weak SPON2 expression in healthy controls and hyperplastic biopsies. Although there were considerable intra- and inter-individual variations, high SPON2 expression was found in prostate cancer, which increased in high-grade tumors (Fig. 3).

Taken together, our findings supported the diagnostic potential of the candidate biomarkers. However, they also exhibited considerable intra- and inter-individual variations, which were associated with cancer grades. This indicated that altered proportions between multiple candidate biomarkers would be more diagnostically useful than changes in the absolute levels of individual biomarkers. This led us to evaluate whether predictive models based on multiple biomarkers could improve the diagnostic accuracy.

Analyses of tumor tissue, urine, and sera supported the diagnostic relevance of the candidate biomarkers

To test the diagnostic relevance of the candidate biomarkers, we first analyzed the bulk RNA expression data from 714 prostate cancer tumor tissue samples and 174 normal samples (Supplementary Data S1; see Materials and Methods; Supplementary Fig. S2; ref. 44). We found that 33 of 36 differentially expressed (adj. P value < 0.05) candidate biomarkers had the same directionality as predicted from the ST data, i.e., up- or downregulated (Fig. 4A).

As a number of studies point to discrepancies between RNA and protein expression, we further analyzed the extent to which our candidate biomarkers were translated into proteins. We analyzed prostate proteomics data from 306 patients (Supplementary Data

S1). Each patient provided three samples corresponding to the primary and secondary grading areas as well as the adjacent benign area. We found that 28 of 31 candidate biomarkers that were differentially expressed (adj. P value < 0.05) had the same directionality as their transcripts at the protein level (Fig. 4A).

As biomarkers are routinely analyzed in the blood, we constructed prediction models based on the protein levels of the candidate biomarkers in plasma from 340 patients with prostate cancer and age-matched healthy controls from the UKBB (Supplementary Data S1). We applied the bootstrap approach with 1,000 repetitions to estimate the uncertainty of the predictions. The feature importance of the prediction model is presented in Supplementary Fig. S18. This analysis resulted in a mean (IQR) AUC of 0.69 (0.68–0.71; Fig. 4B). By comparison, the mean (IQR) AUC for PSA alone was 0.66 (0.66–0.66). Such AUC levels are of limited clinical value compared with PSA, which is routinely measured in the clinic. To further compare the relevance of the candidate biomarkers, we repeated the prediction pipeline using the same number of randomly chosen proteins. This was repeated 1,000 times. This approach yielded a mean (IQR) AUC of 0.58 (0.54–0.61).

We performed the Cox proportional hazards analysis to test the association of the candidate biomarkers with the OS in 340 patients with prostate cancer from the UKBB, as well as with the PFI in 505 patients with prostate cancer from The Cancer Genome Atlas data. We found that four of 20 proteins expressed in the UKBB were significantly (P value < 0.05) associated with the OS (Supplementary Fig. S19) and five of 45 genes were significantly (P value < 0.05) associated with PFI (Supplementary Fig. S20).

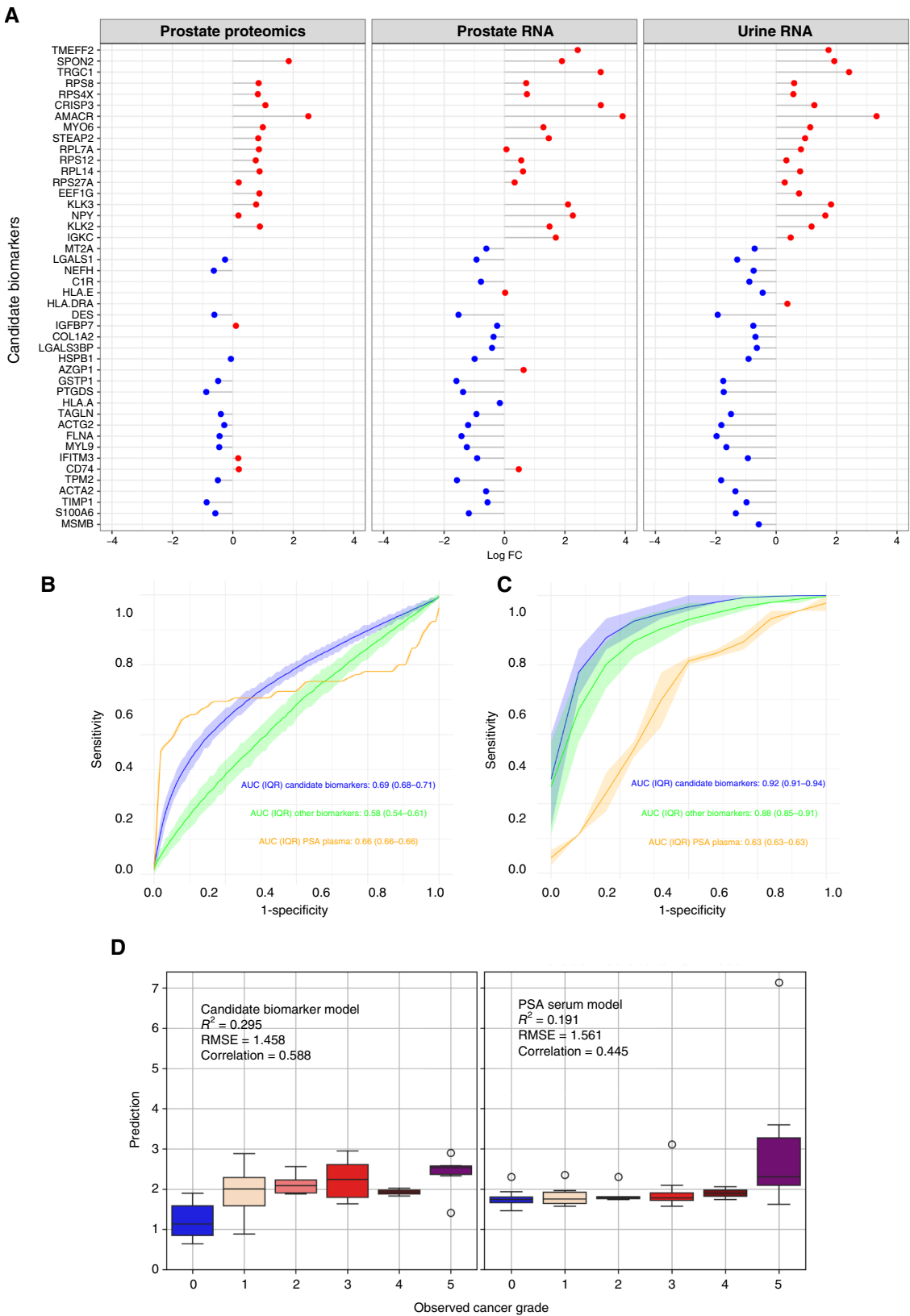
Prediction models based on the candidate biomarkers showed high AUCs in urine

As suggested in introduction, a potential solution to improve the prediction performance of our models could be to analyze the expression of candidate biomarkers in urine. We first analyzed the mRNA expression of candidate biomarkers using data from a previous study of mRNAs in the urine of men diagnosed with localized prostate cancer (14). The study included 761 men with a median (IQR) age of 63 years (58–68), of whom 293 had cancer of GG2 or greater, 106 had cancer of GG1, and 362 had no cancer (Supplementary Data S1). We found that 38 of 39 candidate biomarkers were differentially expressed (adj. P value < 0.05) in the same direction as in our spatial analysis (Fig. 4A).

As we found that the candidate biomarkers were consistently differentially expressed as mRNAs in multiple datasets in both tissues and urine, we hypothesized that they would have diagnostic potential as proteins, which are more stable and easier to measure using routine methods.

We constructed prediction models for prostate cancer diagnosis based on recently published urinary proteomes (16). In that publication, a prediction model based on more than 200 proteins resulted in high AUC values. The material consisted of 123 patients with prostate cancer and 36 patients with BPH (Supplementary Data S1; see Materials and Methods; Supplementary Fig. S3). Urine was collected after DRE, and urinary extracellular vesicles of size 20,000 $\times$ g were isolated.

We divided the dataset into training (70%) and test (30%) samples and constructed models using ridge regression. As the number of patients was limited, we performed the bootstrapping with 1,000 repetitions to estimate the uncertainty of the results. For comparison with the widely used PSA test, predictions based on the individual's serum PSA levels were used as a baseline. Furthermore, to assess the added prediction value of candidate biomarkers, we performed the analysis with the same number of randomly chosen



proteins and repeated it 1,000 times. Although serum PSA differentiated the patients with cancer in test samples with a mean (IQR) AUC of 0.63 (0.63–0.63) and the randomly chosen genes with a mean (IQR) AUC of 0.88 (0.85–0.91), the candidate biomarkers yielded a mean (IQR) AUC of 0.92 (0.91–0.94; **Fig. 4C**). The relative importance of the features in the prediction model is presented in Supplementary Fig. S21.

Given that the candidate biomarkers were identified using PT, which is a measure of malignant transformation, we tested whether the candidate biomarkers could be beneficial in predicting the grade of cancer progression. We used the training data to construct prediction models for cancer grade and tested them using the test data. The relative importance of the features in the prediction models is presented in Supplementary Fig. S22. We found that our model ($R^2 = 0.295$, RMSE = 1.458, correlation = 0.588) outperformed the baseline model based on serum PSA ($R^2 = 0.191$, RMSE = 1.561, correlation = 0.445), thus providing potential clinical utility in the detection of clinically significant prostate cancer (**Fig. 4D**).

Discussion

In this study, we aimed to address a general problem in cancer diagnosis, namely that there is a lack of accurate biomarkers that can be measured using routine clinical methods, focusing on prostate cancer. Data-driven analyses of multiomics data, such as machine learning, have identified several promising combinations of biomarkers for prostate and other cancers. However, it has proven difficult to demonstrate the relevance of such biomarkers across different clinical settings. This problem may depend on the daunting complexity and heterogeneity of prostate and other cancers. Each tumor may involve thousands of genes that can vary between different tumors with the same diagnosis, as well as within the same tumor at different time points, locations, and cell types. Therefore, the prioritization of biomarkers that are both reliable in different settings and measurable with routine methods involves formidable challenges. Here, we hypothesized that these challenges could be systematically addressed using ST and PT to characterize the genes most correlated with malignant transformations in most prostate tumors. This hypothesis was based on a previous study on prostate cancer which showed that PT could organize bulk transcriptomic data from different stages of prostate cancer into a trajectory of malignant transformations. Using PT, this study identified and experimentally validated a potential drug target, *EZH2*, which is an activator of androgen receptor signaling (44).

However, to the best of our knowledge, PT and ST have not been systematically used to identify biomarkers. In support of our analysis, we found that PT was associated with malignant transformations as measured by histologic changes and CNAs. The relevance of CNAs lies in that these consist of gains and losses of DNA. They have important carcinogenic roles by contributing to cancer initiation and progression by increasing chromosomal instability and aneuploidy (63). We also found that groups of spots predicted as aneuploid had significantly higher PT than spots predicted as diploid in all samples. To test whether PT also reflected mechanisms underlying malignant transformation, we performed pathway analysis of the genes most correlated

with PT. This showed that they were enriched with cancer hallmark pathways, like transepithelial migration and androgen response. Those pathways included several genes of known importance for prostate cancer. A summary of the potential relevance of the candidate biomarkers for cancer progression is given in Supplementary Note S1. Some examples can be given here. As mentioned in the results, SPON2 has pleiotropic roles in tumor progression and has been proposed as a biomarker. AMACR is frequently overexpressed in prostate cancer and has been proposed as a biomarker (64). Moreover, a large body of literature implicates the androgen-regulated protein transmembrane protein with EGF-like and two follistatin-like domains 2 in prostate cancer (65). In further support of clinical relevance, the HC genes were also enriched for drug targets in prostate cancer. The candidate biomarkers were also differentially expressed between patients with prostate cancer and controls in independent samples from different cohorts that encompassed peripheral blood, prostate tissues, and urine. The limitations of our study include the fact that PT was based on ST, which is hampered by variable accuracy in measuring mRNA expression, particularly for low-abundance transcripts. Another limitation is that PT may not fully capture the complexity and heterogeneity of branching events in different tumors. Moreover, we measured only the ST data from the 12 samples. This could be an important limitation because of the considerable intra- and inter-individual variation in mRNA expression. However, the relevance of the candidate biomarkers was supported by analyses of mRNA and protein expression levels in more than 2,000 cancer samples from prostate tissue and urine. Furthermore, the mRNA variations in the ST data were associated with a clinical phenotype of potential clinical relevance, namely, prostate cancer grade. This was confirmed at the protein level. IHC of independent prostate tissues showed that a candidate biomarker, SPON2, had higher expression in prostate cancer than in healthy and hyperplastic tissues, as well as in high prostate cancer grade.

This led us to test the diagnostic potential of the candidate biomarker proteins in plasma from UKBB as blood is generally used for the detection of prostate cancer biomarkers. Using machine learning, we constructed a prediction model in a training set, which reached an AUC of 0.69 in a test set. We found that this was better than randomly chosen proteins but a marginal improvement compared with PSA. One explanation could be variable accuracy of blood biomarkers compared with biomarkers in local fluids (66). Because of this and previous studies of prostate cancer biomarkers in urine (14, 44), we hypothesized that a prediction model using urine could be more effective for predicting prostate cancer. Using a similar model on urine data, we obtained an AUC of 0.92 in the test set. By comparison, the corresponding AUC of a model using PSA only reached an AUC of 0.63, which was similar to the AUC in plasma from UKBB. Interestingly, in urine, the randomly chosen candidate biomarkers reached an AUC of 0.88. This supports that proteins in local fluids, like urine in the case of prostate cancer, are more informative for prediction models than those from blood and thus have better diagnostic potential. Although our candidate biomarkers further improved prediction performance, they may also have another important advantage, namely explainability. As mentioned in the introduction, this is increasingly recognized as crucial for clinical implementation of precision diagnostics and

Figure 4.

Validation of candidate biomarkers in prostate tissue and urine from prostate cancer and controls. **A**, Log-fold change (FC) of differentially expressed candidate biomarkers using mRNAs in prostate tissue, mRNAs in urine, and proteins in urine from prostate cancer compared with controls. **B**, ROC curves of prediction models based on plasma proteomics from UKBB. **C**, ROC curves of prediction models based on urine proteins and serum PSA. **D**, Prediction of prostate cancer grade based on the candidate biomarker levels in urine (left) and the PSA levels in serum (right).

therapeutics. In both cases, clinical decisions are more likely to be more successful if they are based on informed and participatory decisions by patients and clinicians. The candidate biomarkers were associated with malignant transformation in ST data, enrichment for CNAs, cancer hallmark pathways, drug targets, and grade. Although these findings reflect the complexity of malignant transformation, increasing efforts are devoted to find solutions for presentation of such information in helping clinicians evaluate and discuss clinical decisions with patients. By comparison, biomarkers derived only from data-driven methods often function as black boxes, making it difficult for clinicians and patients to understand how predictions are made (17). Another problem that can result from a lack of mechanistic understanding of biomarkers is that the accuracy of diagnostic predictions can vary greatly between clinical settings. If so, understanding of the reasons for such variations may help context-dependent clinical implementation and interpretation of biomarkers. A limitation of the prediction models using urine protein levels of the candidate biomarkers is that they were analyzed in less than 200 samples, which may affect the robustness of the results. To approximate the uncertainty of the predictions, we employed bootstrapping and demonstrated that the IQR of AUCs for our models indicated high predictive accuracy. In this context, it is of note that our findings showed agreement in analyses of blood, tissue and urine from almost 2,000 patients.

In further support of the relevance of the candidate biomarkers, we found that they correlated with prostate cancer grade. Notably, the candidate biomarkers included known drug targets for prostate cancer. This raises the possibility that ST and PT can be used to identify drug targets and that the protein levels of these targets may be used as a companion diagnostic tool to personalize treatment with such drugs. As the analysis of proteins in urine is a tractable diagnostic tool, we propose prospective studies to test the diagnostic potential of the candidate biomarkers in prostate cancer. Given the increasing availability of ST data in the public domain, we propose that this analytic approach can be applied to prioritize biomarkers in other cancers. We have made all data and methods freely available for such studies.

References

- Vigneswaran HT, Eklund M, Discacciati A, Nordström T, Hubbard RA, Perlis N, et al. Stockholm3 in a multiethnic cohort for prostate cancer detection (SEPTA): a prospective multicentered trial. *J Clin Oncol* 2024;42:3806–16.
- Tonry C, Finn S, Armstrong J, Pennington SR. Clinical proteomics for prostate cancer: understanding prostate cancer pathology and protein biomarkers for improved disease management. *Clin Proteomics* 2020;17:41.
- Björnebo L, Discacciati A, Falagario U, Vigneswaran HT, Jäderling F, Grönberg H, et al. Biomarker vs MRI-enhanced strategies for prostate cancer screening: the STHLM3-MRI randomized clinical trial. *JAMA Netw Open* 2024;7:e247131.
- Fonseca NM, Maurice-Dror C, Herberts C, Tu W, Fan W, Murtha AJ, et al. Prediction of plasma ctDNA fraction and prognostic implications of liquid biopsy in advanced prostate cancer. *Nat Commun* 2024;15:1828.
- Hachem S, Yehya A, El Masri J, Mavingire N, Johnson JR, Dwead AM, et al. Contemporary update on clinical and experimental prostate cancer biomarkers: a multi-omics-focused approach to detection and risk stratification. *Biology (Basel)* 2024;13:762.
- Mi H, Sivagnanam S, Betts CB, Liudahl SM, Jaffee EM, Coussens LM, et al. Quantitative spatial profiling of immune populations in pancreatic ductal adenocarcinoma reveals tumor microenvironment heterogeneity and prognostic biomarkers. *Cancer Res* 2022;82:4359–72.
- Stojanovic L, Abbotts R, Tripathi K, Coon CM, Rajendran S, Abbasi Farid E, et al. ZNF1 functions as a master regulator of epigenetically induced pathogen mimicry and inflammasome signaling in cancer. *Cancer Res* 2025;85:1183–98.
- Scattoni V, Lazzeri M, Lughezzani G, De Luca S, Passera R, Bollito E, et al. Head-to-head comparison of prostate health index and urinary PCA3 for predicting cancer at initial or repeat biopsy. *J Urol* 2013;190:496–501.
- Agbetuyi-Tayo P, Gbadebo M, Rotimi OA, Rotimi SO. Advancements in biomarkers of prostate cancer: a review. *Technol Cancer Res Treat* 2024;23:15330338241290029.
- Wood SL, Knowles MA, Thompson D, Selby PJ, Banks RE. Proteomic studies of urinary biomarkers for prostate, bladder and kidney cancers. *Nat Rev Urol* 2013;10:206–18.
- Zhang XL, Wu ZZ, Xu Y, Wang JG, Wang YQ, Cao MQ, et al. Saliva proteomic analysis reveals possible biomarkers of renal cell carcinoma. *Open Chem* 2020;18:918–26.
- Bosch LJW, de Wit M, Pham TV, Coupé VMH, Hiemstra AC, Piersma SR, et al. Novel stool-based protein biomarkers for improved colorectal cancer screening: a case-control study. *Ann Intern Med* 2017;167:855–66.
- Martinez-Garcia E, Coll-de la Rubia E, Lesur A, Dittmar G, Gil-Moreno A, Cabrera S, et al. Cervical fluids are a source of protein biomarkers for early, non-invasive endometrial cancer diagnosis. *Cancers (Basel)* 2023;15:911.
- Tosoian JJ, Zhang Y, Xiao L, Xie C, Samora NL, Niknafs YS, et al. Development and validation of an 18-gene urine test for high-grade prostate cancer. *JAMA Oncol* 2024;10:726–36.
- Pavlovic B, Bräutigam K, Dartiguenave F, Martel P, Rakauskas A, Cesson V, et al. Urine biomarkers can predict prostate cancer and PI-RADS score prior to biopsy. *Sci Rep* 2024;14:18148.

Authors' Disclosures

M. Smelik reports grants and other support from Swedish Cancer Society, Radiumhemmet, and Swedish Research Council during the conduct of the study. D. Diaz-Roncero Gonzalez reports grants and other support from Swedish Cancer Society, Radiumhemmet, and Swedish Research Council during the conduct of the study. X. Li reports grants and other support from Swedish Cancer Society, Radiumhemmet, and Swedish Research Council during the conduct of the study. H. Wang reports grants from National Natural Science Foundation of China during the conduct of the study. No disclosures were reported by the other authors.

Authors' Contributions

M. Smelik: Conceptualization, data curation, investigation, visualization, methodology, writing—original draft, writing—review and editing. **D. Diaz-Roncero Gonzalez:** Conceptualization, data curation, investigation, visualization, methodology, writing—original draft, writing—review and editing. **X. An:** Validation, writing—original draft, writing—review and editing. **R. Heer:** Methodology, writing—original draft, writing—review and editing. **L. Henningsohn:** Writing—original draft, writing—review and editing. **X. Li:** Methodology, writing—original draft, writing—review and editing. **H. Wang:** Investigation, methodology, writing—original draft, writing—review and editing. **Y. Zhao:** Conceptualization, supervision, methodology, writing—original draft, writing—review and editing. **M. Benson:** Conceptualization, resources, supervision, funding acquisition, writing—original draft, project administration, writing—review and editing.

Acknowledgments

This work was supported by funding from the Swedish Cancer Society CAN 2017/411, Radiumhemmet, Swedish Research Council, and the National Natural Science Foundation of China (82171791 to H. Wang). Generative AI has been used in the writing process of this manuscript. The purpose of using it was to sharpen the ideas and thus improve the readability. We take full responsibility for the content of the presented work.

Note

Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

Received January 21, 2025; revised March 20, 2025; accepted April 23, 2025; posted first April 28, 2025.

16. Khoo A, Govindarajan M, Qiu Z, Liu LY, Ignatchenko V, Waas M, et al. Prostate cancer reshapes the secreted and extracellular vesicle urinary proteomes. *Nat Commun* 2024;15:5069.
17. Li X, Loscalzo J, Mahmud AKMF, Aly DM, Rzhetsky A, Zitnik M, et al. Digital twins as global learning health and disease models for preventive and personalized medicine. *Genome Med* 2025;17:11.
18. Wadhwa A, Roscoe C, Duran EA, Kwan L, Haraldsen CL, Shelton JB, et al. Neighborhood deprivation, race and ethnicity, and prostate cancer outcomes across California health care systems. *JAMA Netw Open* 2024;7:e242852.
19. Shi J, Wei X, Xun Z, Ding X, Liu Y, Liu L, et al. The web-based portal SpatialTME integrates histological images with single-cell and spatial transcriptomics to explore the tumor microenvironment. *Cancer Res* 2024;84:1210–20.
20. Hou W, Ji Z, Chen Z, Wherry EJ, Hicks SC, Ji H. A statistical framework for differential pseudotime analysis with multiple single-cell RNA-seq samples. *Nat Commun* 2023;14:7286.
21. Jainarayanan A, Mouroug-Anand N, Arbe-Barnes EH, Bush AJ, Bashford-Rogers R, Frampton A, et al. Pseudotime dynamics of T cells in pancreatic ductal adenocarcinoma inform distinct functional states within the regulatory and cytotoxic T cells. *iScience* 2023;26:106324.
22. Kang K, Wu Y, Han C, Wang L, Wang Z, Zhao A. Homologous recombination deficiency in triple-negative breast cancer: multi-scale transcriptomics reveals distinct tumor microenvironments and limitations in predicting immunotherapy response. *Comput Biol Med* 2023;158:106836.
23. Moncada R, Barkley D, Wagner F, Chiodin M, Devlin JC, Baron M, et al. Author Correction: Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat Biotechnol* 2020;38:1476.
24. Cui Zhou D, Jayasinghe RG, Chen S, Herndon JM, Iglesia MD, Navale P, et al. Spatially restricted drivers and transitional cell populations cooperate with the microenvironment in untreated and chemo-resistant pancreatic cancer. *Nat Genet* 2022;54:1390–405.
25. Andrieux G, Das T, Griffin M, Straehle J, Paine SML, Beck J, et al. Spatially resolved transcriptomic profiles reveal unique defining molecular features of infiltrative 5ALA-metabolizing cells associated with glioblastoma recurrence. *Genome Med* 2023;15:48.
26. Ren H, Walker BL, Cang Z, Nie Q. Identifying multicellular spatiotemporal organization of cells with SpaceFlow. *Nat Commun* 2022;13:4076.
27. Wang W, Li T, Xie Z, Zhao J, Zhang Y, Ruan Y, et al. Integrating single-cell and bulk RNA sequencing data unveils antigen presentation and process-related CAFs and establishes a predictive signature in prostate cancer. *J Transl Med* 2024;22:57.
28. Wang M, Deng C, Yang C, Yan M, Lu H, Zhang Y, et al. Unraveling temporal and spatial biomarkers of epithelial-mesenchymal transition in colorectal cancer: insights into the crucial role of immunosuppressive cells. *J Transl Med* 2023;21:794.
29. Iida K, Okada M. Identifying key regulatory genes in drug resistance acquisition: modeling pseudotime trajectories of breast cancer single-cell transcriptome. *Cancers (Basel)* 2024;16:1884.
30. Jaume G, Doucet P, Song A, Lu MY, Almagro Pérez C, Wagner S, et al. HEST-1k: a dataset for spatial transcriptomics and histology image analysis. *Adv Neural Inf Process Syst* 2024;37:53798–833.
31. Mirzazadeh R, Andrusivova Z, Larsson L, Newton PT, Galicia LA, Abalo XM, et al. Spatially resolved transcriptomic profiling of degraded and challenging fresh frozen samples. *Nat Commun* 2023;14:509.
32. Erickson A, He M, Berglund E, Marklund M, Mirzazadeh R, Schultz N, et al. Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature* 2022;608:360–7.
33. Larsson L, Franzén L, Ståhl PL, Lundeberg J. Seqla: a versatile toolkit for spatially resolved transcriptomics analysis and visualization. *Bioinformatics* 2023;39:btad626.
34. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. *J Stat Mech Theor Exp* 2008;2008:P10008.
35. Pham D, Tan X, Balderson B, Xu J, Grice LF, Yoon S, et al. Robust mapping of spatiotemporal trajectories and cell-cell interactions in healthy and diseased tissues. *Nat Commun* 2023;14:7739.
36. Haghverdi L, Büttner M, Wolf FA, Büttner F, Theis FJ. Diffusion pseudotime robustly reconstructs lineage branching. *Nat Methods* 2016;13:845–8.
37. Wolf FA, Hamer FK, Plass M, Solana J, Dahlin JS, Göttgens B, et al. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biol* 2019;20:59.
38. Gao R, Bai S, Henderson YC, Lin Y, Schalck A, Yan Y, et al. Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. *Nat Biotechnol* 2021;39:599–608.
39. 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 U S A* 2005;102:15545–50.
40. Fang Z, Liu X, Peltz G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 2022;39:btac757.
41. Song H, Weinstein HNW, Allegakoen P, Wadsworth MH 2nd, Xie J, Yang H, et al. Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nat Commun* 2022;13:141.
42. Hughes TK, Wadsworth MH 2nd, Gierahn TM, Do T, Weiss D, Andrade PR, et al. Second-strand synthesis-based massively parallel scRNA-seq reveals cellular states and molecular features of human inflammatory skin pathologies. *Immunity* 2020;53:878–94.e877.
43. Petukhov V, Guo J, Baryawno N, Severe N, Scadden DT, Samsonova MG, et al. dropEst: pipeline for accurate estimation of molecular counts in droplet-based single-cell RNA-seq experiments. *Genome Biol* 2018;19:78.
44. Bolis M, Bossi D, Valleria A, Ceserani V, Cavalli M, Impellizzeri D, et al. Dynamic prostate cancer transcriptome analysis delineates the trajectory to disease progression. *Nat Commun* 2021;12:7033.
45. Sun R, A J, Yu H, Wang Y, He M, Tan L, et al. Proteomic landscape profiling of primary prostate cancer reveals a 16-protein panel for prognosis prediction. *Cell Rep Med* 2024;5:101679.
46. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779.
47. Ho DE, Imai K, King G, Stuart EA. MatchIt: nonparametric preprocessing for parametric causal inference. *J Stat Softw* 2011;42:1–28.
48. Goldman MJ, Craft B, Hastie M, Repcheck K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* 2020;38:675–8.
49. Cannon M, Stevenson J, Stahl K, Basu R, Coffman A, Kiwala S, et al. DGIdb 5.0: rebuilding the drug-gene interaction database for precision medicine and drug discovery platforms. *Nucleic Acids Res* 2024;52:D1227–35.
50. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 2014;32:381–6.
51. Kim JE, Lee K, Kim IY. Current update on prostate cancer immunotherapy. *J Urol Oncol* 2023;21:14–22.
52. Schmid S, Omlin A, Higano C, Sweeney C, Martinez Chanza N, Mehra N, et al. Activity of platinum-based chemotherapy in patients with advanced prostate cancer with and without DNA repair gene aberrations. *JAMA Netw Open* 2020;3:e2021692.
53. Markowski MC, Taplin M-E, Aggarwal R, Sena LA, Wang H, Qi H, et al. Bipolar androgen therapy plus nivolumab for patients with metastatic castration-resistant prostate cancer: the COMBAT phase II trial. *Nat Commun* 2024;15:14.
54. Bommer RK, Kethireddy S, Govindapur RR, Eppakayala L. Synthesis, biological evaluation and docking studies of 1,2,4-oxadiazole linked 5-fluorouracil derivatives as anticancer agents. *BMC Chem* 2021;15:30.
55. Zhao X, Cao M, Liu JJ, Zhu H, Nelson DR, Liu C. Reactive oxygen species is essential for cycloheximide to sensitize lexitumab-induced apoptosis in hepatocellular carcinoma cells. *PLoS One* 2011;6:e16966.
56. Georgescu C, Corbin JM, Thibivilliers S, Webb ZD, Zhao YD, Koster J, et al. A TMEFF2-regulated cell cycle derived gene signature is prognostic of recurrence risk in prostate cancer. *BMC Cancer* 2019;19:423.
57. Qian X, Li C, Pang B, Xue M, Wang J, Zhou J. Spondin-2 (SPON2), a more prostate-cancer-specific diagnostic biomarker. *PLoS One* 2012;7:e37225.
58. Bergström SH, Järemo H, Nilsson M, Adamo HH, Bergh A. Prostate tumors downregulate microseminoprotein-beta (MSMB) in the surrounding benign prostate epithelium and this response is associated with tumor aggressiveness. *Prostate* 2018;78:257–65.
59. Rehman I, Cross SS, Azzouzi A-R, Catto JWF, Deloume JC, Larre S, et al. S100A6 (Calcylin) is a prostate basal cell marker absent in prostate cancer and its precursors. *Br J Cancer* 2004;91:739–44.
60. dos Reis ST, Viana NI, Iscaife A, Pontes-Junior J, Dip N, Antunes AA, et al. Loss of TIMP-1 immune expression and tumor recurrence in localized prostate cancer. *Int Braz J Urol* 2015;41:1088–95.

61. Brehmer B, Biesterfeld S, Jakse G. Expression of matrix metalloproteinases (MMP-2 and -9) and their inhibitors (TIMP-1 and -2) in prostate cancer tissue. *Prostate Cancer Prostatic Dis* 2003;6:217–22.
62. Zhang J, Liu G, Liu Y, Yang P, Xie J, Wei X. The biological functions and related signaling pathways of SPON2. *Front Oncol* 2024;13:1323744.
63. Steele CD, Abbasi A, Islam SMA, Bowes AL, Khandekar A, Haase K, et al. Signatures of copy number alterations in human cancer. *Nature* 2022;606:984–91.
64. Jin X, Ji J, Niu D, Yang Y, Tao S, Wan L, et al. Urine exosomal AMACR is a novel biomarker for prostate cancer detection at initial biopsy. *Front Oncol* 2022;12:904315.
65. Masood M, Grimm S, El-Bahrawy M, Yagüe E. TMEFF2: a transmembrane proteoglycan with multifaceted actions in cancer and disease. *Cancers (Basel)* 2020;12:3862.
66. Smelik M, Zhao Y, Mansour Aly D, Mahmud AF, Sysoev O, Li X, et al. Multiomics biomarkers were not superior to clinical variables for pan-cancer screening. *Commun Med (Lond)* 2024;4:234.