The Prostate

WILEY

The Prostate

ORIGINAL ARTICLE OPEN ACCESS

A Gene Expression Tumor Signature Optimizing Partial Area-Under-the-Curve (pAUC) to Improve Specificity for Indolent Prostate Cancer

Travis A. Gerke[1,2] | Svitlana Tyekucheva[3,4] | Konrad H. Stopsack[2,5] | Christopher J. Sweeney[6] | Rosina T. Lis[7] | Elizabeth N. Westling[2,8] | Massimo Loda[9,10,11] | Meir J. Stampfer[2] | Kathryn L. Penney[2] | Giovanni Parmigiani[3,4] | Lorelei A. Mucci[2]

[1]Prostate Cancer Clinical Trials Consortium, Memorial Sloan Kettering Cancer Center, New York, New York, USA | [2]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA | [3]Department of Data Science, Dana-Farber Cancer Institute, Boston, Massachusetts, USA | [4]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, Massachusetts, USA | [5]Department of Epidemiologic Methods and Etiologic Research, Leibniz Institute for Prevention Research and Epidemiology – BIPS, Bremen, Germany | [6]South Australian Immunogenomics Cancer Center, University of Adelaide, Adelaide, Australia | [7]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA | [8]Elson S. Floyd College of Medicine, Washington State University, Spokane, Washington, USA | [9]Department of Pathology, Weill Cornell Medical Center, New York, New York, USA | [10]Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK | [11]Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts, USA

Correspondence: Lorelei A. Mucci (lmucci@hsph.harvard.edu)

Received: 12 September 2025 | **Revised:** 21 March 2026 | **Accepted:** 16 April 2026

Funding: DF/HCC Specialized Programs in Research Excellence (SPORE) in Prostate Cancer, Grant/Award Number: 2P50CA090381; National Cancer Institute, Grant/Award Number: T32 CA009001

Keywords: gene expression | indolent prostate cancer | partial area-under-the-curve | prognostic biomarkers

ABSTRACT

Purpose: A key clinical challenge in prostate cancer is the identification and validation of biomarkers with high specificity for indolent long-term outcomes. We applied a novel statistical method to identify tumor transcriptomic biomarkers that optimally predicted patients with low metastatic potential.

Methods: Using tumor whole-transcriptome data from the Health Professionals Follow-up Study (HPFS, discovery set) and Physicians' Health Study (PHS, validation set), we compared patients who died of prostate cancer or developed metastases ("lethal," $n = 113$) and patients with > 8 years of metastasis-free survival ("indolent," $n = 291$). Whole transcriptome tumor gene expression data were generated using an Affymetrix array. We applied a novel method for optimizing a partial area under the curve (pAUC) that up-weighted indolent cases with a predefined 80%–100% specificity. This method leverages weighted logistic lasso regression, with weights chosen via cross-validation to reduce overfitting.

Results: Median age at cancer diagnosis was 66 years; median follow-up for outcomes was 14 years. We identified a 40-gene transcriptome signature of indolent prostate cancer, which, compared to Gleason grade groups, improved the pAUC over the predefined 80%–100% specificity range by 1.72-fold ($p < 0.001$) and improved overall AUC from 0.85 to 0.93 ($p < 0.001$). The signature improved positive predictive value for indolent tumors > 2-fold with minimal decrease in negative predictive value. Importantly, the 40-gene signature showed high discrimination among intermediate Gleason 7 tumors (Grade groups 2 and 3, AUC 0.88, 95% CI: 0.79–0.95).

Travis A. Gerke and Svitlana Tyekucheva contributed equally to first authorship.

Kathryn L. Penney, Giovanni Parmigiani, and Lorelei A. Mucci contributed equally to senior authorship.

Conclusion: Incorporating pAUC into prognostic signature development improved identification of prostate tumors with low risk of metastatic potential. Its clinical application may help reduce overtreatment and overdiagnosis of indolent prostate cancers, and the pAUC may be relevant beyond prostate cancer.

1 | Introduction

Prostate cancer is a malignancy of two extremes. On the one hand, it is the second leading cause of cancer mortality among men in the United States, leading to 35,770 estimated deaths in 2025. On the other, there is a large pool of prostate cancers detected by screening with prostate-specific antigen (PSA) which would likely not have presented symptomatically over a person's lifetime. It is estimated that 45 million American men have undiagnosed prostate cancer [1], and that two-thirds or more of screen-detected prostate cancers are overdiagnosed and overtreated [2]. If a low-risk prostate cancer is diagnosed, active surveillance may be the most appropriate course of action [3]. However, many such patients are treated with radical prostatectomy or radiation [4], resulting in unnecessary side effects, detriments to quality of life, and increased health care costs.

Prognostic information at diagnosis has the potential to help guide treatment decisions for men with prostate cancer. Standard clinical tools for risk stratification include PSA level, Gleason grade group (tumor grade), and clinical stage, which have appreciable sensitivity for detecting aggressive prostate cancers; however, their specificity for indolent disease is low [5]. An expanding collection of tissue biomarkers has been proposed [6] to supplement clinical factors in identifying aggressive prostate cancer.

Most prior work has focused on identification of markers for aggressive or lethal prostate cancer. In contrast, we sought to prioritize the identification of a tumor molecular signature that accurately predicts indolent prostate cancer, and to do so above and beyond the information provided by Gleason grade group.

2 | Materials and Methods

2.1 | Participant Selection

This study is nested within two nationwide studies, the Health Professionals Follow-up Study (HPFS) and Physicians' Health Study (PHS). HPFS is a prospective cohort of 51,529 US male health professionals initiated in 1986. The PHS I and II were randomized trials of aspirin and supplements among 29,069 US male physicians in the primary prevention of cardiovascular disease and cancer. The methods for follow-up, prostate cancer confirmation, and mortality outcomes are similar in the cohorts and have previously been described [7–9].

We assembled a prostate tumor biobank of archival formalin-fixed, paraffin-embedded (FFPE) prostate tumor tissue sections obtained clinically from radical prostatectomy (91% in the present study) or transurethral resection of the prostate (9%) from patients with incident prostate cancer in the cohorts [10–11]. Thus, the majority of patients in this study underwent curative-intent surgery, and the specificity of a biomarker for indolent tumors will tend to be underestimated (see Discussion).

For this gene expression profiling study, we undertook a case-control design with cumulative incidence sampling among eligible participants with available tumor tissue [12, 13]. Lethal cases were prostate cancer patients who were diagnosed with or developed distant metastases from prostate cancer or died of their cancer at any time during follow-up through 2012. Indolent controls were men who survived at least 8 years after their cancer diagnosis and did not have evidence of metastases during follow-up. Given the long natural history of prostate cancer and to avoid misclassification of the outcome, we selected a minimum of 8 years to give adequate follow-up for a metastatic event to occur. Lethal cases and controls were not matched on any clinical or demographic factors to avoid overmatching. Ultimately, 404 prostate cancer patients were sampled for the transcriptomic analysis. The case-control study within HPFS, our training set, included 254 men who were diagnosed with prostate cancer between 1986 and 2004, 83 of whom experienced a lethal event over a median follow-up of 13.7 years (IQR 10.1–16.1, Table 1). The case-control study within PHS, our validation dataset, included 150 men with prostate cancer occurring between 1982 and 2005, among whom 30 had a metastatic event or died from prostate cancer over a median follow-up of 14.4 years (IQR 11.5–18.0). Study pathologists completed a standardized re-review of histopathology, including Gleason grade groups.

2.2 | RNA Expression Profiling and Data Pre-Processing

Gene expression profiling of RNA extracted from tumor specimens paired whole transcriptome amplification with microarray technologies that performed well on FFPE tissue [14]. Our previous publication [12] provides details on RNA extraction with the Agencourt FormaPure FFPE kit (Beckman Coulter, Brea, CA), amplification using the WT-Ovation FFPE System V2 (NuGEN, San Carlos, CA), and microarray hybridization using a GeneChip Human Gene 1.0 ST microarray (Affymetrix, Santa Clara, CA), using the RMA method [15, 16]. Gene symbol mapping was done using Bioconductor annotation package pd.hugene.1.0.st.v1 [17]. Gene expression data are available through Gene Expression Omnibus (accession number GSE79021). For model building, feature filtering criteria were chosen through stratified 10-fold cross-validation in the training data. This strategy retained markers expressed at a level of 4.37 or higher in at least 80% of the training samples, reducing 33,297 measured features to 4377 non-control features for analysis.

2.3 | Statistical Analysis

Because of its larger sample size, the case-control set within HPFS was used as training dataset for signature discovery and prognostic model building. Our principal aim was to identify

TABLE 1 | Characteristics of prostate cancer patients with tumor gene expression profiling data in the training set from the Health Professionals Follow-up Study (HPFS, 1986–2004) and in the validation set from the Physicians' Health Study (PHS, 1982–2005).

	HPFS	PHS
N	254	150
Years of cancer diagnosis	1986–2004	1982–2005
Age at diagnosis – median (Q1–Q3), years	66.0 (62.0–70.0)	65.9 (61.7–69.5)
Lethal events – *n* (%)	83 (33)	30 (20)
Years of follow-up – median (Q1–Q3)	13.7 (10.1–16.1)	14.4 (11.5–18.0)
Gleason grade groups – *n* (%)		
Grade group 1 (6)	24 (9)	33 (22)
Grade group 2 (3 + 4)	91 (36)	48 (32)
Grade group 3 (4 + 3)	74 (29)	28 (19)
Grade groups 4, 5 (8–10)	65 (26)	41 (27)
Clinical TNM stage – *n* (%)		
T1/T2 N0/Nx M0/Mx	214 (84)	136 (91)
T3 N0/Nx M0/Mx	21 (8)	6 (4)
T4/N1/M1	14 (6)	6 (4)
Prostate-specific antigen, ng/mL – *n* (%)		
0–4	24 (9)	17 (11)
4–10	117 (46)	79 (53)
10–20	43 (17)	20 (13)
> 20	28 (11)	13 (9)
Tissue from prostatectomy – *n* (%)	236 (93)	133 (89)

patients, with reasonable confidence, who could be managed with active surveillance because of a high probability of indolent disease. As such, our goal in marker selection was to discover a set of markers that, when combined with Gleason grade groups, would prioritize a high specificity of at least 80% while maintaining high sensitivity for lethal disease. By prioritizing accurate classification of indolent prostate cancers, this approach stands in contrast to more common strategies that aim to optimize the overall area under the curve (AUC), a measure that can be high even when many indolent cancers are misclassified [18]. We therefore maximized a partial area under the curve (pAUC) [18, 19] using a novel statistical method which is based on weighted lasso logistic regression [20]. We chose the pAUC using a specificity for indolent disease of 80%; this value was prespecified to meet the important clinical needs to reduce the number of false negatives of missing a lethal prostate cancer while maintaining sensitivity. Briefly, our method selects and combines markers to optimize pAUC by up-weighting patients with indolent disease in the training data, with the number of markers and weights chosen through cross-validation.

In the primary validation analysis, regression coefficients estimated from the HPFS were applied to the PHS marker data to obtain predicted risks of lethal disease. Two regression models were of interest: the first was a logistic regression that included Gleason grade groups alone (categorical, 6 (grade group 1), 3 + 4 (grade group 2), 4 + 3 (grade group 3), 8–10 (grade groups 4 and 5)), while the second was the logistic lasso that included both Gleason grade groups and 40 selected markers. Confidence intervals for the pAUC of each model and their pAUC difference were calculated with 5000 bootstrap resamples. The same approach was used to test for statistical significance of the difference. As a secondary analysis, the full AUC was tested for improvement and confidence intervals were calculated [21].

We also created a simplified signature score using a sign-averaging method [22]. This technique considers only signs, not coefficients per marker, which allows comparing signatures when coefficients are unavailable or may not transport to a different study population. This method adds or subtracts standardized marker values in the validation data based on whether their marginal association with lethal prostate cancer in the training data is positive or negative, respectively. Thus, large values of the score imply many poor prognostic features.

Marker values were averaged when gene symbols mapped to more than one feature. Logistic regression models that included Gleason grade groups and the signature score were fit in the HPFS for each signature, including our own, and external performance was assessed in the PHS. All analyses were conducted using R version 3.1.1. Partial AUC statistics were calculated using the pROC package [23] and lasso logistic regressions were estimated using the glmnet [24] package for R.

The research was approved by institutional review boards at Harvard T.H. Chan School of Public Health and Brigham and Women's Hospital, and those of participating registries as required. Participants provided written informed consent.

3 | Results

There were 254 prostate cancer patients in the HPFS training set and 150 in the PHS validation set (Table 1). Median age was

66 years at diagnosis. Although there was a slightly higher proportion of low grade, low stage tumors in the PHS case-control set, the two samples were generally similar across clinical characteristics.

Optimization of pAUC with weighted lasso logistic regression in the HPFS training set selected 40 genes (Table 2). When the coefficients estimated from the training set were applied to the validation data of the PHS set, pAUC was significantly improved by 1.72-fold, from 0.082 (95% CI: 0.053–0.118) with Gleason grade groups alone to 0.141 (95% CI: 0.109–0.171; $p < 0.001$ for difference) when Gleason grade group was combined with the 40-gene set (Figure 1). The signature showed accuracy with high specificity over the highlighted pAUC range. The largest possible value for the chosen pAUC would be 0.20 given our specificity threshold of 80% (0%–20% false negative).

The full AUC was improved from 0.847 (95% CI: 0.786–0.908) with Gleason alone to 0.932 (95% CI: 0.891–0.973) with the addition of the signature ($p < 0.001$ for difference; Figure 1). Sensitivity remained at 80% with a specificity of 91%, and dropped to 70% with a specificity of 94%. When considering the signature using more transportable signs (Table 2) instead of coefficients, the gene set score was able to discriminate indolent from lethal cancer with an AUC of 0.835 (95% CI: 0.754–0.916) and, when combined with Gleason, produced a pAUC of 0.116 (95% CI: 0.081–0.150) in the validation set.

The sign-averaged 40-gene signature score discriminated between indolent and lethal prostate cancer across patients with different Gleason grade groups (Figure 2). When the PHS set was restricted to men with the intermediate-risk Gleason score 7 (grade groups 3 and 4), our signature had substantial discriminatory value with an AUC of 0.879 (95% CI: 0.786–0.954).

4 | Discussion

Using a novel analytic strategy and a unique prostate cancer population with long-term and complete follow-up, we identified a 40-gene signature that improved the specificity of Gleason scoring to distinguish indolent from lethal prostate cancers. Of note, the signature performed well among Gleason grade groups 3 and 4 (Gleason score 7) cancers, for which risk prediction is particularly challenging. Because most diagnosed prostate cancers in current practice are slow growing or indolent, even slight improvements in prognostic test specificity can lead to substantial gains in positive predictive value (PPV) and enhance clinical utility.

Developing candidate signatures that improve prognostication beyond what is known from clinicopathological factors has been challenging [6, 25–27]. Inadequate long-term follow-up may drive subpar model robustness or necessitate the use of surrogate outcomes such as biochemical recurrence. We aimed to overcome these barriers by directly considering the sensitivity-specificity balance in our training approach and by testing and validating the signature with respect to lethal disease. The transcriptomic data were assayed in HPFS and PHS together, reducing differences in biomarker measurement error between the test and validation sets. To enhance the potential that this signature would perform well in other transcriptomics datasets, we also tested the signature using more transportable signs showing good prognostic measures.

TABLE 2 | The top 40 genes identified in the training set from the Health Professionals Follow-up Study (HPFS), ranked by order of coefficient magnitude. Sign indicates the direction of association between gene expression and lethal prostate cancer.

Rank	Gene	Location	Sign	Rank	Gene	Location	Sign
1	*MT3*	16q13	—	21	*ENSG00000207091*	15q14	—
2	*POLR2E*	19p13	—	22	*FAM156A*	Xp11	+
3	*KRTAP5-7*	11q13	+	23	*ASB16*	17q21	—
4	*CHD3*	17p13	—	24	*MTDH*	8q22	+
5	*VAMP2*	17p13	—	25	*GAB1*	4q31	—
6	*RNY4P8*	17q21	—	26	*ENSG00000244469*	22q13	—
7	*SMARCE1*	17q21	+	27	*PTMS*	12p13	—
8	*NUDT16L1*	16p13	—	28	*AZGP1*	7q22	—
9	*ENSG00000201868*	1p34	—	29	*DDX19B*	16q22	—
10	*WAC*	10p12	—	30	*ENSG00000240384*	1q21	—
11	*LSM14B*	20q13	+	31	*UBAP2L*	1q21	+
12	*TPT1*	13q14	—	32	*RELA*	11q13	+
13	*IGFBP6*	12q13	+	33	*SLC44A2*	19p13	—
14	*CYB5D1*	17p13	—	34	*GATAD2B*	1q21	+
15	*WTAP*	6q25–q27	—	35	*HMOX2*	16p13	—
16	*FOXK2*	17q25	+	36	*ZNF154*	19q13	+
17	*FOS*	14q24	—	37	*FAM36A*	1q44	+
18	*SNX5*	20p11	+	38	*LASP1*	17q11–q21	+
19	*SFRP1*	8p11	—	39	*SRSF11*	1p31	—
20	*LOC388428*	17q25	+	40	*SPECC1L*	22q11	—

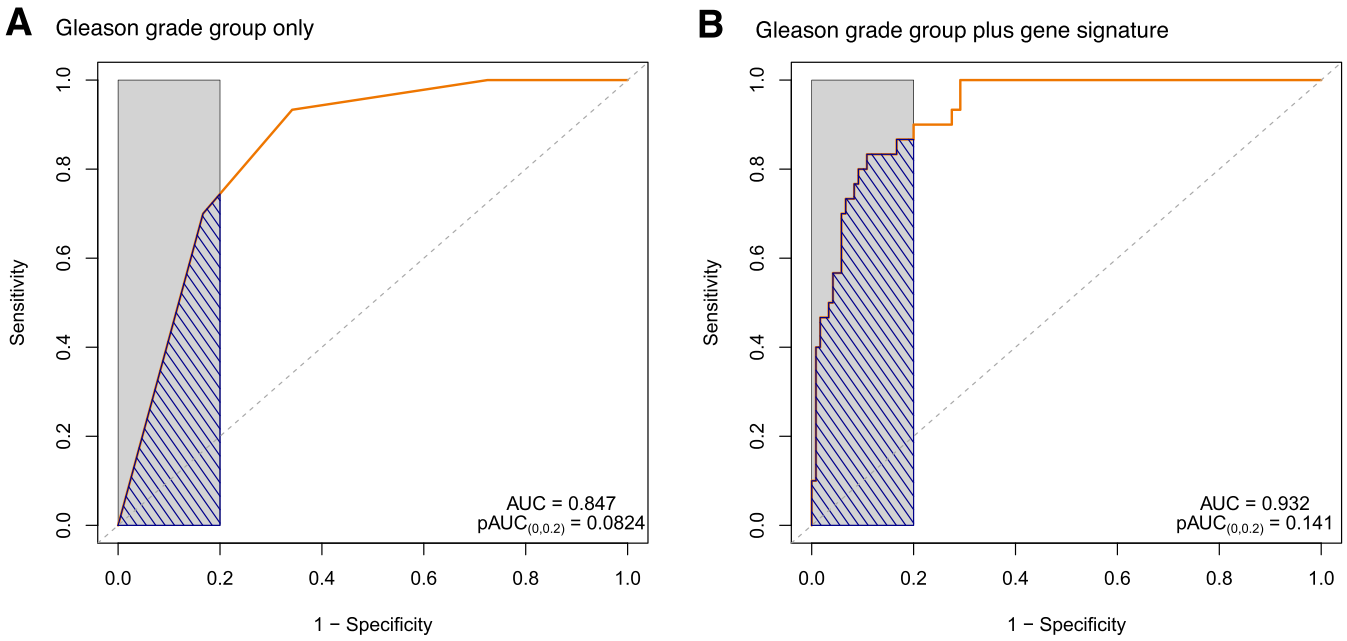

FIGURE 1 | Receiver operating curves (ROC) for models with Gleason grade groups alone (A) vs. the model that combines Gleason grade groups and the 40 selected genes (B), in the Physicians' Health Study (PHS) validation data. Partial area under the curve (AUC) over specificity range [0.80, 1.00] is highlighted. [Color figure can be viewed at wileyonlinelibrary.com]

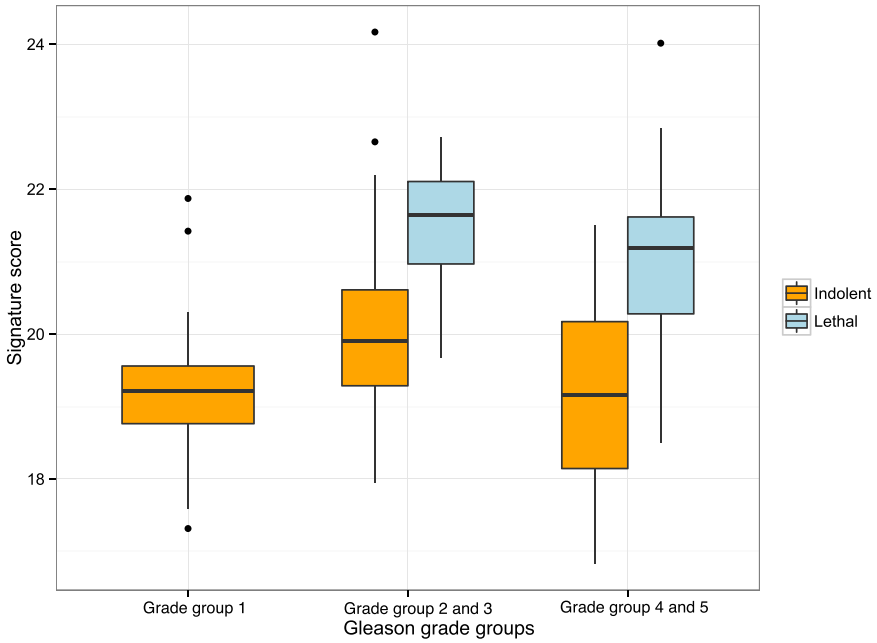

FIGURE 2 | Distribution of the sign-averaged signature score in lethal and indolent cancers in the validation set from the Physician' Health Study (PHS), stratified by Gleason grade groups. No lethal events were observed among Gleason grade group 1 (Gleason 6) tumors. [Color figure can be viewed at wileyonlinelibrary.com]

Interestingly, although changes in full AUC were not our primary analytic focus, our signature was also able to improve the overall discriminatory performance of Gleason grade group in both the training and validation sets. This observation suggests that discovery efforts focusing on the relatively stable biomarkers in the indolent subtype compared to potentially multiple aggressive subtypes [28] may enjoy broader discriminatory power. Another consideration with our approach is the interpretation of the pAUC results as a function of setting the specificity of the models. For example, there would be an increase in specificity from the signature if one fixed the specificity to be that of Gleason grade groups 4 and 5 (score > 7), and likewise the impact on the signature if one fixed the sensitivity to be the same as Gleason> 7.

As an illustrative example of the signature's impact, suppose we chose to stratify patients with Gleason score 7 or higher (grade groups 2–5) as likely lethal, and those with 6 (grade group 1) as

likely indolent. In the PHS validation data, this approach would have a sensitivity of 1.00 (as all lethal events occurred in the Gleason grade groups 2 or higher group) and a specificity of 0.28 (as there were many false positives). Thus, this binarization of Gleason can be viewed as a highly sensitive marker. By contrast, the addition of our signature maintained a specificity of 0.80 with a sensitivity of 0.90. Now, we can apply our results and examine the potential impact, assuming two estimates for the cumulative incidence of lethal cancer between 10 and 15 years of follow-up from randomized trials of radical prostatectomy vs. watchful waiting [3, 29]. In a prostatectomy arm, the risk of lethal cancer may be as low as 5.8%, while in a watchful waiting cohort, it may reach 20.7%. With these values, Gleason grade group alone would have PPVs of 7.9% (prostatectomy) and 26.6% (watchful waiting) and NPVs of 100% (both settings). Our signature classifier would have PPVs of 21.7% (prostatectomy) and 54.0% (watchful waiting) and NPVs of 99.2% (prostatectomy) and 96.8% (watchful waiting). The increase in correct classifications with the gene signature is substantial, although the tradeoff with regards to the added false negatives for lethal tumors compared to Gleason grade group alone is apparent. Further studies that elucidate clinically appropriate scoring cutoffs to best balance sensitivity and specificity, ideally incorporating patient preferences, would be necessary prior to widespread clinical adoption.

The duration and completeness of clinical follow-up from the two independent cohorts in this study provide a considerable strength. The long-term study period permits the use of lethal disease as an endpoint, rather than a surrogate such as biochemical recurrence. However, as a result, the included patients may have a somewhat higher-risk disease than that diagnosed today, especially among those in our underlying cohorts diagnosed prior to the introduction of PSA screening. This limitation is less of a concern in that our signature is designed to target accuracy for non-lethal patients, which represents the majority of diagnosed prostate cancer patients in the United States. Further strengths of our study include the use of a validated mRNA expression profiling platform for FFPE tissues [14] and a standardized histopathological re-review of all tumor samples. The central re-review helped set a high benchmark with Gleason scores that outperformed locally assigned scores [30]. Put differently, the sensitivity of routine Gleason scores may be lower than in our data. For example, we did not observe lethal events among Gleason score 6 cancers, and this case distribution would likely differ with biopsy specimens prior to pathologic upgrading [31].

One potential limitation of the study is the use of data derived from prostatectomy tumor specimens to derive the indolent signature. Some men in this study who were treated with radical prostatectomy were cured from lethal tumors that would have metastasized if left untreated: “category B” patients as per our “ABC model” for prostate cancer biomarkers [32] or the “exposure preventive” response type of causal effects in individuals [33]. The result is a likely decrease in the measurable specificity of our classifier and an underestimate of the prognostic value of our signature. Ideally, the signature would be validated on prostate biopsy specimens in cohorts with decades long follow-up. While the prevalence of such individuals in our study population is unknowable [32, 33], trials of radical prostatectomy vs. watchful waiting/active surveillance indicate that on the absolute scale, the proportion of men who benefit from treatment is small for intermediate and low-risk disease [3, 29]. Our gene expression analyses were conducted in a research setting, although we undertook best laboratory practices, and found high reproducibility of transcriptomics in technical replicates [14]. Whether the validity of the signature holds in the context of platforms outside of the Affymetrix arrays will need to be examined. Moreover, to facilitate clinical translation, biomarker panels conducted in CLIA certified settings will be needed to ensure accurate test results to patients.

Collectively, we show the application of methodology that creates a prognostic model with high specificity for indolent disease. Applied in a post-prostatectomy setting, as in ours, such a model may be useful in helping address fears of cancer recurrence despite therapy. Further work is needed to validate the specific markers identified here, and other models developed in recent years, across additional prospective trials. In particular, if evaluated in prostate biopsy cohorts of men undergoing active surveillance with long-term follow-up for lethal disease, such models may help identify men who can avoid immediate treatment and safely and confidently be managed with a low intensity active surveillance program, such as MRI alone, with a very low risk of progression to needing treatment. More generally, our work showcases the potential value of an explicit focus on indolent disease during development of prognostic biomarkers in prostate cancer and perhaps in other cancer entities. Our methodology, or other developments focusing on pAUC optimization [34], may be useful in that regard.

Acknowledgments

We are grateful to the participants and staff of the Physicians' Health Study and Health Professionals Follow-up Study for their valuable contributions. We thank Michael Pitt for valuable assistance with gene expression profiling. We would like to acknowledge the key contributions of Alex Forrest Hay at Affymetrix and Dr. Edward Fox (in memoriam) at the Dana-Farber Cancer Institute Microarray Core. This project was supported by the DF/HCC Specialized Programs in Research Excellence (SPORE) in Prostate Cancer (2P50CA090381), funding from the National Cancer Institute (P01 CA055075, U01 CA167552, and R01 CA141298), the A. David Mazzone Career Development Award, and the Prostate Cancer Foundation. T.A.G. was supported by the National Cancer Institute grant T32 CA009001. K.L.P., K.H.S., and L.A.M. were Young Investigators of the Prostate Cancer Foundation.

Conflicts of Interest

Travis A. Gerke, Svitlana Tyekucheva, Elizabeth N. Westling, Massimo Loda, and Meir J. Stampfer declare they have no competing interests. Christopher J. Sweeney receives research funding from Janssen, Astellas/Pfizer, Sanofi, Bayer, and Dendreon, was a consultant for BMS, Novartis, Sanofi, Janssen, Astellas, Bayer, Genentech/Roche, Pfizer, Lilly, Hengrui, CellCentric, PointBiopharma, and Astra Zeneca, holds stock options for Leuchemix, and owns royalties and other intellectual properties: Parthenolide (Indiana University), dimethylamino parthenolide (Leuchemix), Exelixis: Abiraterone plus cabozantinib combination, and FRAS1 SNP and tristetraprolin as biomarkers of lethal prostate cancer. Rosina T. Lis received research funding from MSKCC and Harvard T.H. Chan School of Public Health, and is a consultant for Janssen/J&J. Giovanni Parmigiani holds equity as a co-founder in Phaeno Biotechnologies, is on the SAB of Realm Idx (which owns Ambry Genetics), and currently consults for Delphi Diagnostics and (pro bono) for Martingale Labs. He also co-leads the BayesMendel laboratory, whose BayesMendel package which includes several ML

tools for the computation of carrier probability of cancer susceptibility genes and future cancer risk is licensed by DFCI. He does not derive any personal income from these licenses. All revenues are assigned to the lab for software maintenance and upgrades. None of these influenced this study. Kathryn Penney is currently employed at ConcertAI. Lorelei A. Mucci receives research funding from Astra Zeneca to Harvard University, and holds equity interest in Convergent Therapeutics. None of these influenced this study.

Data Availability Statement

The gene expression data used in this project are available on Gene Expression Omnibus GSE79021. Additional data from the HPFS are available through a project request: https://hsph.harvard.edu/research/health-professionals/resources/for-external-collaborators/.

References

1. J. L. Jahn, E. L. Giovannucci, and M. J. Stampfer, "The High Prevalence of Undiagnosed Prostate Cancer at Autopsy: Implications for Epidemiology and Treatment of Prostate Cancer in the Prostate-Specific Antigen-Era," *International Journal of Cancer* 137, no. 12 (2015): 2795–2802.

2. S. Loeb, M. A. Bjurlin, J. Nicholson, et al., "Overdiagnosis and Overtreatment of Prostate Cancer," *European Urology* 65, no. 6 (2014): 1046–1055.

3. T. J. Wilt, M. K. Brawer, K. M. Jones, et al., "Radical Prostatectomy Versus Observation for Localized Prostate Cancer," *New England Journal of Medicine* 367, no. 3 (2012): 203–213.

4. B. A. Mahal, S. Butler, I. Franco, et al., "Use of Active Surveillance or Watchful Waiting for Low-Risk Prostate Cancer and Management Trends Across Risk Groups in the United States, 2010–2015," *JAMA* 321, no. 7 (2019): 704–706.

5. O. Andrén, K. Fall, L. Franzén, S. O. Andersson, J. E. Johansson, and M. A. Rubin, "How Well Does the Gleason Score Predict Prostate Cancer Death? A 20-Year Follow-Up of a Population Based Cohort in Sweden," *Journal of Urology* 175, no. 4 (2006): 1337–1340.

6. S. E. Eggener, R. B. Rumble, A. J. Armstrong, et al., "Molecular Biomarkers in Localized Prostate Cancer: ASCO Guideline," *Journal of Clinical Oncology* 38 (2020): 1474–1494.

7. E. Giovannucci, Y. Liu, E. A. Platz, M. J. Stampfer, and W. C. Willett, "Risk Factors for Prostate Cancer Incidence and Progression in the Health Professionals Follow-Up Study," *International Journal of Cancer* 121, no. 7 (2007): 1571–1578.

8. J. M. Chan, M. J. Stampfer, J. Ma, P. H. Gann, J. M. Gaziano, and E. L. Giovannucci, "Dairy Products, Calcium, and Prostate Cancer Risk in the Physicians' Health Study," *American Journal of Clinical Nutrition* 74, no. 4 (2001): 549–554.

9. J. M. Gaziano, R. J. Glynn, W. G. Christen, et al., "Vitamins E and C in the Prevention of Prostate and Total Cancer in Men: The Physicians' Health Study II Randomized Controlled Trial," *JAMA* 301, no. 1 (2009): 52–62.

10. A. Pettersson, R. T. Lis, A. Meisner, et al., "Modification of the Association Between Obesity and Lethal Prostate Cancer by TMPRSS2:ERG," *JNCI: Journal of the National Cancer Institute* 105, no. 24 (2013): 1881–1890.

11. T. A. Gerke, N. E. Martin, Z. Ding, et al., "Evaluating a 4-Marker Signature of Aggressive Prostate Cancer Using Time Dependent AUC," *Prostate* 75, no. 16 (2015): 1926–1933.

12. K. L. Penney, J. A. Sinnott, S. Tyekucheva, et al., "Association of Prostate Cancer Risk Variants With Gene Expression in Normal and Tumor Tissue," *Cancer Epidemiology, Biomarkers & Prevention* 24 (2015): 255–260.

13. J. A. Sinnott, S. F. Peisch, S. Tyekucheva, et al., "Prognostic Utility of a New mRNA Expression Signature of Gleason Score," *Clinical Cancer Research* 23, no. 1 (2017): 81–87.

14. S. Tyekucheva, N. E. Martin, E. C. Stack, et al., "Comparing Platforms for Messenger RNA Expression Profiling of Archival Formalin-Fixed, Paraffin-Embedded Tissues," *Journal of Molecular Diagnostics* 17, no. 4 (2015): 374–381.

15. R. A. Irizarry, "Summaries of Affymetrix GeneChip Probe Level Data," *Nucleic Acids Research* 31, no. 4 (2003): e15.

16. R. A. Irizarry, "Exploration, Normalization, and Summaries of High Density Oligonucleotide Array Probe Level Data," *Biostatistics* 4, no. 2 (2003): 249–264.

17. B. Carvalho, pd.hugene.1.0.st.v1: Platform Design Info for Affymetrix HuGene-1 0-st-v1. R Package Version 3.8.0.

18. D. K. McClish, "Analyzing a Portion of the ROC Curve," *Medical Decision Making* 9, no. 3 (1989): 190–195.

19. M. L. Thompson and W. Zucchini, "On the Statistical Analysis of ROC Curves," *Statistics in Medicine* 8, no. 10 (1989): 1277–1290.

20. T. Gerke, S. Tyekucheva, L. Mucci, et al., "Logistic Push: A Regression Framework for Partial AUC Optimization," *arXiv* (2016; arXiv:1606.06562.

21. E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the Areas Under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach," *Biometrics* 44, no. 3 (1988): 837–845.

22. S. D. Zhao, G. Parmigiani, C. Huttenhower, and L. Waldron, "Más-o-Menos: A Simple Sign Averaging Method for Discrimination in Genomic Data Analysis," *Bioinformatics* 30, no. 21 (2014): 3062–3069.

23. X. Robin, N. Turck, A. Hainard, et al., "pROC: An Open-Source Package for R and S+ to Analyze and Compare ROC Curves," *BMC Bioinformatics* 12 (2011): 77.

24. J. Friedman, T. Hastie, and R. Tibshirani, "Regularization Paths for Generalized Linear Models Via Coordinate Descent," *Journal of Statistical Software* 33, no. 1 (2010): 1–22.

25. S. F. Shariat, M. W. Kattan, A. J. Vickers, P. I. Karakiewicz, and P. T. Scardino, "Critical Review of Prostate Cancer Predictive Tools," *Future Oncology* 5, no. 10 (2009): 1555–1584.

26. L. A. Mucci, Y. Pawitan, F. Demichelis, et al., "Nine-Gene Molecular Signature Is Not Associated With Prostate Cancer Death in a Watchful Waiting Cohort," *Cancer Epidemiology, Biomarkers & Prevention* 17, no. 1 (2008): 249–251.

27. A. Sboner, F. Demichelis, S. Calza, et al., "Molecular Sampling of Prostate Cancer: A Dilemma for Predicting Disease Progression," *BMC Medical Genomics* 3 (2010): 8.

28. K. L. Penney, J. A. Sinnott, K. Fall, et al., "mRNA Expression Signature of Gleason Grade Predicts Lethal Prostate Cancer," *Journal of Clinical Oncology* 29, no. 17 (2011): 2391–2396.

29. A. Bill-Axelson, L. Holmberg, M. Ruutu, et al., "Radical Prostatectomy Versus Watchful Waiting in Early Prostate Cancer," *New England Journal of Medicine* 364, no. 18 (2011): 1708–1717.

30. J. R. Stark, S. Perner, M. J. Stampfer, et al., "Gleason Score and Lethal Prostate Cancer: Does 3 + 4 = 4 + 3?," *Journal of Clinical Oncology* 27 (2009): 3459–3464.

31. J. I. Epstein, Z. Feng, B. J. Trock, and P. M. Pierorazio, "Upgrading and Downgrading of Prostate Cancer From Biopsy to Radical Prostatectomy: Incidence and Predictive Factors Using the Modified Gleason Grading System and Factoring in Tertiary Grades," *European Urology* 61, no. 5 (2012): 1019–1024.

32. A. Pettersson, T. Gerke, K. Fall, et al., "The ABC Model of Prostate Cancer: A Conceptual Framework for the Design and Interpretation of Prognostic Studies," *Cancer* 123, no. 9 (2017): 1490–1496.

33. S. Greenland and J. M. Robins, "Identifiability, Exchangeability, and Epidemiological Confounding," *International Journal of Epidemiology* 15, no. 3 (1986): 413–419.

34. Q. Yan, L. E. Bantis, J. L. Stanford, and Z. Feng, "Combining Multiple Biomarkers Linearly to Maximize the Partial Area Under the ROC Curve," *Statistics in Medicine* 37 (2018): 627–642.