

Tumor subtype defines distinct pathways of molecular and clinical progression in primary prostate cancer

Deli Liu,^{1,2,3,4} Michael A. Augello,^{1,2} Ivana Grbesa,^{1,2} Davide Prandi,⁵ Yang Liu,⁶ Jonathan E. Shoag,^{2,7,8} R. Jeffrey Karnes,⁹ Bruce J. Trock,¹⁰ Eric A. Klein,¹¹ Robert B. Den,¹² Francesca Demichelis,¹³ Elai Davicioni,⁶ Andrea Sboner,^{1,3,4,14} and Christopher E. Barbieri^{1,2,4}

¹Sandra and Edward Meyer Cancer Center and ²Department of Urology, Weill Cornell Medicine, New York, New York, USA. ³HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York, USA. ⁴Englander Institute for Precision Medicine of Weill Cornell Medicine and NewYork-Presbyterian Hospital, New York, New York, USA. ⁵Fondazione Bruno Kessler (FBK), Trento, Italy. ⁶GenomeDx Bioscience Inc, Vancouver, British Columbia, Canada. ⁷University Hospitals Cleveland Medical Center, Cleveland, Ohio, USA. ⁸Case Western Reserve University School of Medicine, Cleveland, Ohio, USA. ⁹Department of Urology, Mayo Clinic, Rochester, Minnesota, USA. ¹⁰Department of Urology, Johns Hopkins University, Baltimore, Maryland, USA. ¹¹Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, Ohio, USA. ¹²Department of Radiation Oncology, Sidney Kimmel Medical College at Thomas Jefferson University, Philadelphia, Pennsylvania, USA. ¹³Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento, Italy. ¹⁴Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, New York, USA.

BACKGROUND. Molecular characterization of prostate cancer (PCa) has revealed distinct subclasses based on underlying genomic alterations occurring early in the natural history of the disease. However, how these early alterations influence subsequent molecular events and the course of the disease over its long natural history remains unclear.

METHODS. We explored the molecular and clinical progression of different genomic subtypes of PCa using distinct tumor lineage models based on human genomic and transcriptomic data. We developed transcriptional classifiers, and defined “early” and “late” categories of molecular subclasses from 8,158 PCa patients. Molecular subclasses were correlated with clinical outcomes and pathologic characteristics using Kaplan-Meier and logistic regression analyses.

RESULTS. We identified *PTEN* and *CHD1* alterations as subtype-specific late progression events specifically in *ERG*-overexpressing (*ERG*+) and *SPOP*-mutant tumors, respectively, and 2 distinct progression models consisting of *ERG/PTEN* (normal to *ERG*+ to *PTEN*-deleted) and *SPOP/CHD1* (normal to *SPOP*-mutated to *CHD1*-deleted) with shared early tumorigenesis but distinct pathways toward progression. We found that within *ERG*+ and *SPOP*-mutant subtypes, late events were associated with worse prognosis. Importantly, the clinical and pathologic features associated with distinct late events at radical prostatectomy were strikingly different; *PTEN* deletions were associated with increased locoregional stage, while *CHD1* deletions were only associated with increased grade, despite equivalent metastatic potential.

CONCLUSION. These findings suggest a paradigm in which specific subtypes of PCa follow distinct pathways of progression, at both the molecular and clinical levels. Therefore, the interpretation of common clinical parameters such as locoregional tumor stage may be influenced by the underlying tumor lineage, and potentially influence management decisions.

FUNDING. Prostate Cancer Foundation, National Cancer Institute, Urology Care Foundation, Damon Runyon Cancer Research Foundation, US Department of Defense, and the AIRC Foundation.

Authorship note: AS and CEB jointly supervised this work.

Conflict of interest: YL reports employment with Decipher Biosciences. RJK reports patents, royalties, and other intellectual property with Decipher Biosciences. BJT reports consulting or advisory roles with Decipher Biosciences and Myriad Genetics, and research funding from Myriad Genetics and MDxHealth. EAK reports consulting or advisory roles with Decipher Biosciences and Genomic Health, and is a member of the Speakers' Bureau for Genomic Health. RBD reports a consulting or advisory role for Decipher Biosciences, is a member of the Speakers' Bureau for Bayer, and reports research funding from Medivation, Astellas Pharma, and Decipher Biosciences, and travel, accommodations, and expense reimbursement from Decipher Biosciences. ED reports employment, holding a leadership position, stock and other ownership interests, and travel, accommodations, and expense reimbursement from Decipher Biosciences, as well as a patent: Cancer diagnostics using biomarkers (patent 20140066323). CEB is a coinventor on a patent application filed regarding *SPOP* mutations in prostate cancer by Weill Cornell Medicine.

Copyright: © 2021, American Society for Clinical Investigation.

Submitted: March 30, 2021; **Accepted:** April 7, 2021; **Published:** May 17, 2021.

Reference information: *J Clin Invest.* 2021;131(10):e147878.

<https://doi.org/10.1172/JCI147878>. Submitted: January 19, 2021; Accepted: April 7, 2021; Published: May 17, 2021.

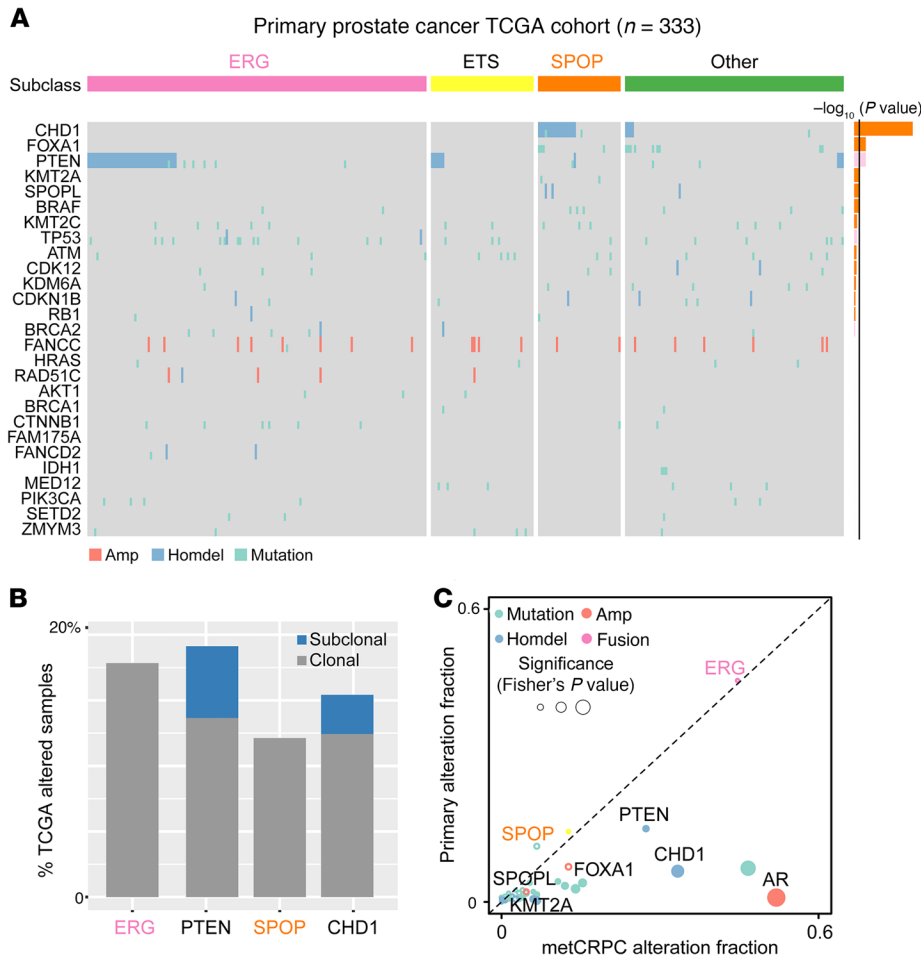


Figure 1. Identification of subclass-specific late progression events in localized prostate cancer. (A) Enrichment of recurrent genomic alterations in *ERG*-fusion and *SPOP*-mutant subclasses from TCGA localized prostate cancer (PCa) cohort ($n = 333$). The alteration enrichment between 2 subclasses was calculated by 2-sided Fisher's exact test. Orange denotes enrichment in *SPOP* mutant, pink denotes enrichment in *ERG* subclass. Amp, amplification; homdel, homozygous deletion. (B) Clonality results of *ERG* fusion, *PTEN* deletion, *SPOP* mutation, and *CHD1* deletion from TCGA cohort. Alteration frequency from each event is shown on the y axis, and different colors represent clonal and subclonal changes. (C) Enrichment of genomic alterations from localized PCa to metCRPC nominate progression events. Alteration percentages in metCRPC samples ($n = 150$) are shown on the x axis, and those in localized TCGA samples ($n = 333$) are shown on the y axis. The significance of enrichment (2-sided Fisher's exact test P value) is shown by the size of the dots: small, $P < 0.05$; medium, $P < 0.01$; and large, $P < 0.001$. Genes in bold have significant enrichments of genomic alterations by using Fisher's exact test for alteration burden.

Introduction

Prostate cancer (PCa) is a clinically and molecularly heterogeneous disease (1-7). Emerging next-generation DNA and RNA sequencing (RNA-seq) data point toward different molecular subclasses of PCa (3, 4, 8, 9), defined by underlying genomic alterations. Approximately 50% of primary PCAs harbor gene fusions involving members of the *ETS* family of transcription factors, including *ERG*, *ETV1*, *ETV4*, and *FLI1* (1, 2, 4-6), while another key subclass representing approximately 10% of PCAs is defined by recurrent mutations in *SPOP* (3, 4, 10-13). PCa has a natural history spanning decades and our understanding of the evolution of PCa over time is only beginning to emerge (4-6). Subtype-defining events like *ERG* fusion (which leads to *ERG* overexpression) and *SPOP* mutation appear to occur in early tumor development (4, 12, 14, 15); events occurring later in the natural history of these cancers may drive progression to more aggressive local disease, transition to metastatic phenotypes, or resistance to therapy (4, 12, 16, 17). However, understanding how specific subtype-defining events influence alterations that occur later and their impact on the clinical course of the disease remain unclear (4-7, 12, 18).

We previously established a framework using an RNA-based model to classify tumor subtype from transcriptional data (7), allowing the interrogation of cohorts with the long follow-up necessary to define clinical outcomes (18-29). Here, we established distinct tumor lineage models of PCa progression, by defining

early and late progression events within specific subtypes, and investigating their unique and shared transcriptional alterations and signaling pathways. We developed transcriptional classifiers to categorize subtype-specific early and late states, and applied these to a retrospective cohort including 1,626 patient samples and a prospective cohort including 6,532 samples using microarray-based gene expression data from a clinically available prognostic assay (the Decipher Prostate Cancer Test).

Results

Defining genomic late progression events in ERG-fusion and SPOP-mutant subclasses. To understand molecular progression in specific subtypes of PCa, we initially took an unbiased approach to define genomic alterations (including point mutations, amplifications, and homozygous deletions) associated with specific subclasses (Figure 1A, Supplemental Figure 1, and Supplemental Table 1; supplemental material available online with this article; <https://doi.org/10.1172/JCI147878DS1>). In the *ERG*-fusion subclass, *PTEN* deletion was the most enriched alteration, while *CHD1* deletion was the most enriched alteration in the *SPOP*-mutant subclass (Figure 1A), consistent with prior results (4, 12, 30).

We next attempted to distinguish between early alterations and those more likely to represent late progression events. By investigating the clonal architectures of those genomic events in The Cancer Genome Atlas (TCGA) primary PCa cohort (4),

we found that all *ERG* fusions and *SPOP* mutations were clonal changes, consistent with early alterations. In contrast, a significant fraction of *PTEN* and *CHD1* deletions were subclonal changes (Figure 1B and Supplemental Table 2), more suggestive of late progression events and consistent with previous findings (15). Furthermore, we compared the fraction of samples with these alterations in advanced metastatic castration-resistant prostate cancer (metCRPC) (16) and primary PCa (4); *PTEN* and *CHD1* deletions were enriched in metCRPC (Figure 1C and Supplemental Table 3), again consistent with late progression events (16, 31). Overall, these results confirmed that specific subtypes of PCa are associated with subsequent molecular changes; tumors with *ERG* fusions later may acquire *PTEN* deletions, while *SPOP*-mutant tumors may progress with *CHD1* deletion.

Identification of 2 tumor lineage models: *ERG/PTEN* and *SPOP/CHD1*. To understand the transcriptional landscape of molecular progression within subtypes, we established a tumor lineage model with 3 PCa states: (a) normal (benign prostate samples), (b) “early” (*ERG* overexpressing or *SPOP* mutant), or (c) “late” (*PTEN* or *CHD1* deleted) cancer from TCGA cohort (4). We investigated transcriptional alterations via unbiased differential expression analyses across these states within each genomically defined subtype (32). We hypothesized that transcriptional changes associated with disease progression follow a specific pattern: increasing or decreasing steadily from the normal to early to late states (Figure 2A). Using the 2 models, (a) normal to *ERG*+ to *PTEN*-deleted (*ERG/PTEN*) and (b) normal to *SPOP*-mutated to *CHD1*-deleted (*SPOP/CHD1*), we found 3,160 *ERG/PTEN* and 1,654 *SPOP/CHD1* progressively upregulated and downregulated genes (Figure 2A and Supplemental Tables 4 and 5). In contrast, testing the reverse order of events (normal to *PTEN*-deleted to *ERG*+, or normal to *CHD1*-deleted to *SPOP*-mutated) returned very few altered genes (Supplemental Figure 2 and Supplemental Tables 6 and 7), supporting the temporal sequence of our original models. To define convergent signaling between the 2 lineage models, we compared affected genes and nominated pathways, and found that upregulated genes shared among both subtype progression models were enriched in cell cycle function, while shared downregulated genes were enriched in focal adhesion function (Figure 2B, Supplemental Figure 3, and Supplemental Tables 8 and 9), consistent with broad common processes of tumorigenesis within the 2 tumor lineages (1, 4). In contrast, uniquely altered genes displayed different functional annotation (Supplemental Figure 3).

By comparing the transcriptional pathways between these 2 tumor lineages, we identified similar enriched functions from the normal to early states (Supplemental Figure 4 and Supplemental Tables 10 and 14), but divergent signatures from the early to late states, in multiple localized PCa cohorts: TCGA (4), Taylor (1), and International Cancer Genome Consortium Prostate Adenocarcinoma - Canada (ICGC PRAD-CA) (33) (Figure 2C, Supplemental Figure 5, and Supplemental Tables 11–13 and 15–17). To further validate these transcriptional differences and determine if underlying genomic alterations were causative, we examined the transcriptomes of prostate organoids and tissue samples from genetically engineered mouse models with conditional deletion of *Pten* or *Chd1*, corresponding to the late state with each subtype (34, 35). Gene set enrichment analysis (GSEA) demonstrat-

ed distinct enriched functions between these 2 tumor lineages, in both human samples and genetically engineered mouse models (Figure 2D, Supplemental Figure 6, and Supplemental Table 18). For instance, tumor lineage of *SPOP/CHD1* was positively enriched in androgen response signature (Figure 2D), consistent with higher androgen receptor transcriptional activity in *SPOP*-mutant samples (4, 11) and higher prostate-specific antigen (PSA) in *SPOP*-mutant cases (7).

To further explore the transcriptional divergence between these tumor lineages, we examined putative upstream transcriptional regulators for gene expression changes (36). We detected similar predicted upstream regulators from the normal to early states (Supplemental Figure 7 and Supplemental Tables 19 and 20). However, distinct upstream regulators were identified from the early to late states in multiple cohorts: TCGA (4), Taylor (1), and ICGC PRAD-CA (33) (Figure 2E, Supplemental Figures 8 and 9, and Supplemental Tables 21–26). Specifically, growth/survival-related kinases such as MEK, PI3K, and Erb-B2 receptor tyrosine kinase 2 (*ERBB2*) were predicted to be active in the *ERG/PTEN* lineage but inhibited in the *SPOP/CHD1* lineage, while kinase inhibitors showed the opposite trend, suggesting distinct activities between the tumor lineages (Figure 2E). Consistent with its status as a known oncogenic *SPOP* substrate, *TRIM24* was predicted to be activated only in the *SPOP/CHD1* lineage, whereas putative *GATA2* activity was restricted to the *ERG/PTEN* lineage (37–39). Broadly speaking, these analyses credential 2 distinct transcription-based tumor lineage progression models consisting of *ERG/PTEN* and *SPOP/CHD1*, with shared early tumorigenesis but distinct pathways toward progression.

Development of SCaPT models to classify *PTEN* and *CHD1* deletions from transcriptional data. We next sought to understand the impact of subtype-specific progression on clinical outcomes (7, 18) using RNA-based machine-learning classifiers, similarly to that which we have previously reported (7). We developed subclass predictor based on transcriptional data (SCaPT) models to categorize prostate tumors according to subtype-specific molecular events (*ERG/PTEN* and *SPOP/CHD1*). To define signatures of *PTEN* and *CHD1* deletions, we selected transcriptional features specific for these genomic events using TCGA cohort (ref. 4 and Figure 3, A and B). We next utilized support vector machine (SVM) (40–42) models and performed 10-fold cross validation to define the best features and models with highest sensitivity and specificity (Figure 3A and Supplemental Figure 10), and thereby established 2 RNA-based classifiers for *PTEN* and *CHD1* deletions (Supplemental Tables 27 and 28). With unsupervised hierarchical clustering using the *PTEN*- and *CHD1*-deleted signatures on TCGA training data, we found expected enrichment of cases with *PTEN* and *CHD1* genomic deletions (Figure 3C and Supplemental Figures 11 and 12). To further validate these models, we applied our *PTEN* and *CHD1* transcriptional classifiers to an independent cohort (1), and found approximately 80% sensitivity and 90% specificity compared with genomic annotations (Figure 3D). These results demonstrated that our SCaPT models classify *PTEN*- and *CHD1*-deleted subclasses on the basis of transcriptional data with high accuracy and confidence.

Tumor lineage in 8,158 patients using the SCaPT models and decision tree. We applied RNA-based classifiers (SCaPT) and

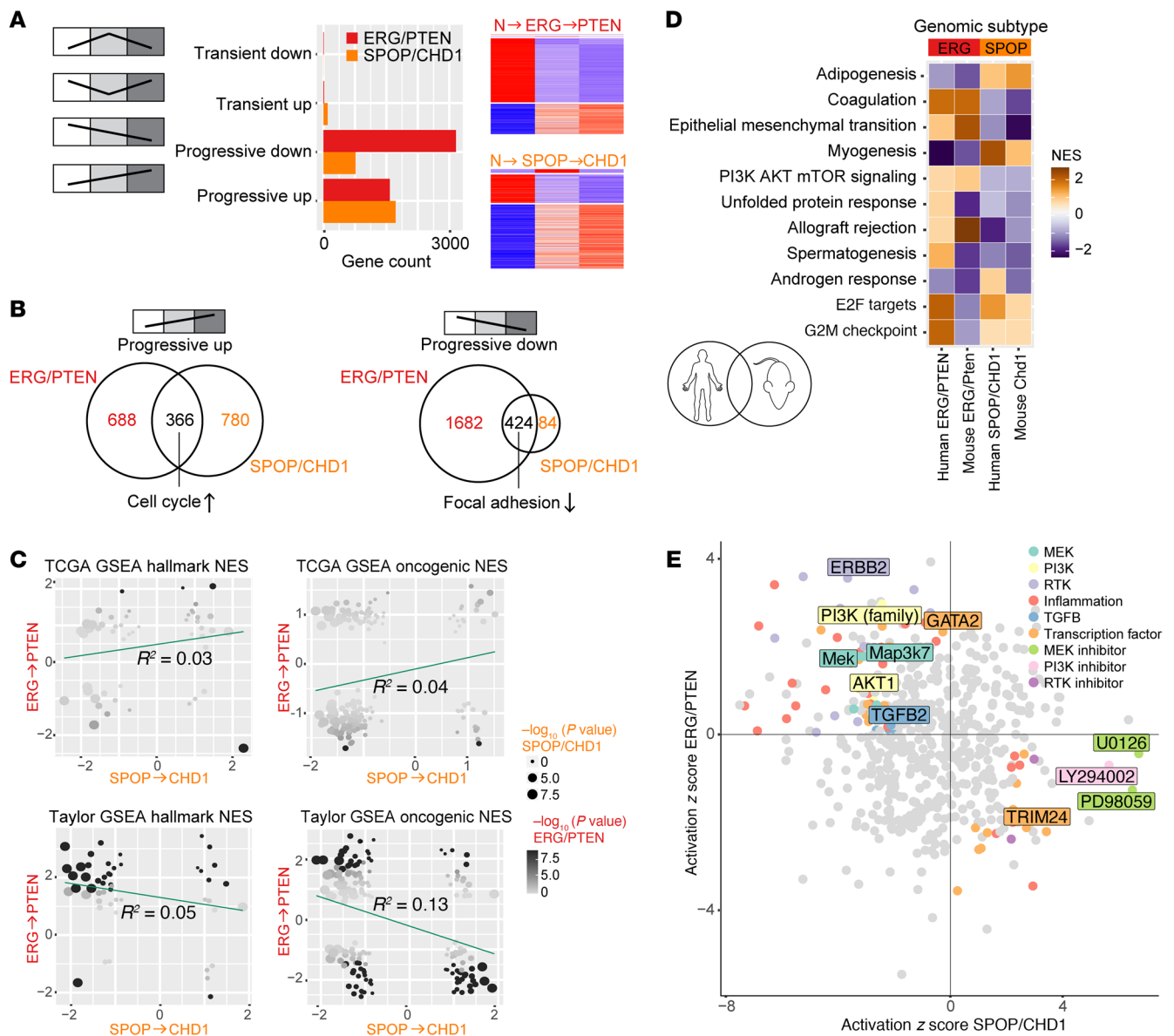


Figure 2. Transcriptional alterations of 2 distinct tumor lineage models: *ERG/PTEN* and *SPOP/CHD1*. (A) Two distinct tumor lineage models of PCa progression: *ERG/PTEN* and *SPOP/CHD1* via ImpulseDE2 identified from TCGA cohort. Total genes in each category (transiently and progressively upregulated and downregulated) are represented in the bar plot with corresponding heatmaps. (B) Venn diagrams of shared and uniquely upregulated and downregulated genes between the 2 tumor lineage models. Numbers of shared and unique altered genes are indicated. (C) Normalized enrichment score (NES) from “early” to “late” states between the 2 tumor lineage models in TCGA and Taylor cohorts. R^2 values of the linear regression model are shown. (D) Distinct pathways with NES from early to late events in TCGA cohort, *ERG/PTEN* mouse tissue, and *Chd1* mouse organoid samples. (E) Divergent predicted upstream regulators from early to late events between the 2 tumor lineage models in TCGA cohort. Different colors represent upstream regulator groups.

decision tree to define tumor lineage in 8,158 patients from retrospective and prospective Genomics Resource Information Database (GRID) cohorts (refs. 7, 18; Figure 4A, and Supplemental Figure 13). Among the retrospective cohort with 1,626 radical prostatectomy specimens, we classified 8% (range, 4% to 10%) of samples to be *CHD1*^{del} (*CHD1* deleted), 8% (2% to 13%) of samples to be *SPOP*^{mut} (*SPOP* mutant), and 2% (1% to 4%) of samples to be *SPOP*^{mut}+*CHD1*^{del} (*SPOP* mutant with *CHD1* deletion) (Figure 4B and Supplemental Figure 14). Previously defined expression thresholds (7, 18) classified 42% (35% to 68%) as *ERG* fusion (overexpression, *ERG*+), and 14% (8% to 29%) to be *ERG*+*PTEN*^{del} (*ERG* fusion with *PTEN* deletion),

and 28% (21% to 39%) to be *ERG*+*PTEN*^{wt} (*ERG* fusion without *PTEN* deletion) (Figure 4B and Supplemental Figure 14). Expression thresholds defined 9% (7% to 12%) as non-*ERG* *ETS* fusion, and 35% (12% to 38%) without outlier expression, which we defined as an “other” subclass (Figure 4B). Among the prospective cohort with 6,532 radical prostatectomy specimens, we classified 7% of cases to be *CHD1*^{del}, 4% of cases to be *SPOP*^{mut}, 15% as *ERG*+*PTEN*^{del}, 24% as *ERG*+*PTEN*^{wt}, 9% as *ETS*, and 36% as other subclass (Supplemental Figure 15). Overall, the percentage of each molecular subclass is consistent with previous PCa studies (1, 4–6), supporting the validity of our SCaPT models and decision tree.

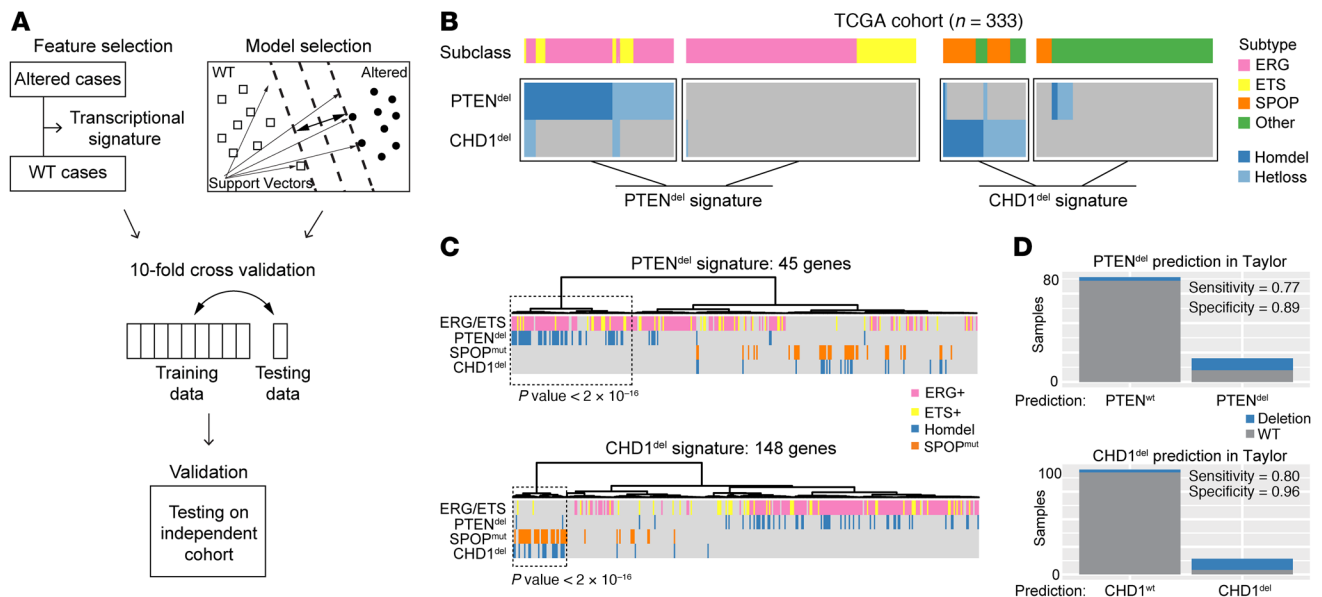


Figure 3. Development of SCaPT models to classify *PTEN* and *CHD1* deletions from transcriptional data. (A) Overview of SCaPT models to predict *PTEN* and *CHD1* deletions from transcriptional data, including steps of feature selection, model selection, 10-fold cross validation, and validation testing on independent cohort. (B) *PTEN* signature of differentially expressed genes between *PTEN*-deleted and WT samples from TCGA *ETS*-fusion samples, and *CHD1* signature of differentially expressed genes between *CHD1*-deleted and WT samples from TCGA non-*ETS*-fusion samples. Different colors represent molecular subclasses. Homdel, homozygous deletion; hetloss, heterozygous loss. (C) Significant enrichment of *PTEN*- and *CHD1*-deleted samples with *PTEN* and *CHD1* features based on unsupervised hierarchical clustering of TCGA samples. Different colors represent genomic alterations. (D) Accuracy and confidence of *PTEN*- and *CHD1*-deleted subtype classifications by SCaPT model determined by testing on an independent data set (n = 106).

Late progression events are associated with worse clinical prognosis. To investigate the association of molecular progression with clinical progression and patient prognosis, we examined the clinical outcomes associated with early and late progression events within each molecular subclass (Supplemental Figure 16). We found worse metastasis-free survival in both *CHD1*^{del} and *PTEN*^{del} tumors compared with the early state within each subtype (*SPOP*^{mut} and *PTEN*^{wt}) (Figure 4, C and D). Of note, early states of each subtype had similar favorable prognosis, while both late states showed similar unfavorable prognosis (Supplemental Figure 16). Endpoints of biochemical recurrence-free survival and PCa-specific mortality-free survival rates followed similar patterns (Supplemental Figure 17), consistent with previous findings (7, 43). These results show that genomic alterations defined as late progression events at the molecular level also show clear evidence of more aggressive disease, consistent with clinical progression. Furthermore, these data suggest that the degree of progression within each subtype, rather than the initial lineage, is more associated with clinical prognosis.

Distinct clinical and pathologic characteristics among late progression events. Finally, having established that molecular progression within each subtype was associated with similar prognosis regarding detection of metastatic disease, we examined the association of clinical and pathologic characteristics in the 2 late-progressed states, using retrospective and prospective cohorts of 8,158 radical prostatectomy specimens, compared to various references. Consistent with known association with aggressive disease features, we found that tumors with predicted *PTEN* deletion were more likely to harbor adverse pathologic features at radical prostatectomy: lymph node invasion,

extracapsular extension, seminal vesicle invasion, and higher Gleason score in both retrospective and prospective cohorts (Figure 5, A and C, Supplemental Figures 18 and 19, and Supplemental Tables 29 and 30), consistent with pathologic features of late progression events. Strikingly, however, tumors with predicted *CHD1* deletion were only associated with higher Gleason score but no other adverse clinical features (Figure 5, A and C, and Supplemental Tables 29 and 30). When compared with the early event of *SPOP* mutation, *CHD1* deletion was associated with higher Gleason score in the retrospective cohort only (Supplemental Figures 18 and 19). Similarly, higher tumor stage (T3/T4) was associated with predicted *PTEN* deletion, but not *CHD1* deletion (Figure 5, B and D). We further validated these associations in TCGA cohort (4), using genomic events to annotate subclass, rather than transcriptional signatures (Supplemental Figure 20).

By comparing enriched signaling pathways between lymph node invasion, and from early to late states of 2 tumor lineages — *ERG/PTEN* and *SPOP/CHD1* — we identified similar enriched functions from lymph node invasion and *ERG/PTEN* lineage. Strikingly, we found divergent signatures between lymph node invasion and *SPOP/CHD1* lineage, supporting the clinical findings that *PTEN*-deleted tumors were more likely to harbor adverse pathological features including lymph node invasion, whereas *CHD1* deletion was not associated with locoregional adverse features (Supplemental Figure 21). Interestingly, when compared with lymph node invasion and *ERG/PTEN* lineage, *SPOP/CHD1* lineage showed dysregulation in metabolism-related pathways (Supplemental Figure 22), which has been shown to represent a hallmark of cancer progression and metastasis (44–46).

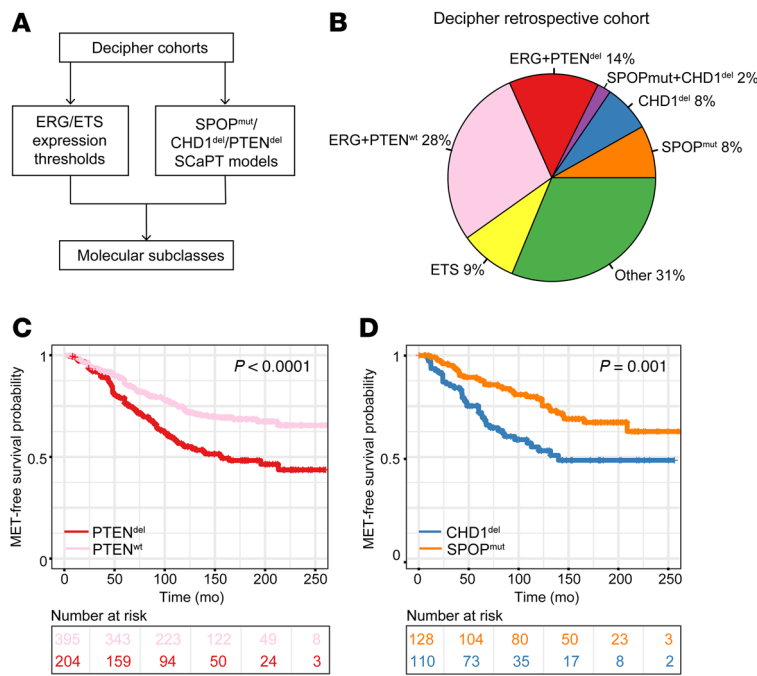


Figure 4. The molecular subclass prediction via SCAPT models and its prognostic outcomes from the Decipher retrospective cohort. (A) Overview of molecular subclass classification in Decipher cohorts via SCAPT models and gene expression thresholds. (B) Subclass classifications from the Decipher retrospective cohort with 1,626 samples, on the basis of SCAPT models and decision tree. Different colors represent molecular subclasses. (C) Significant difference in prognostic outcome between *PTEN*-deleted and WT subclasses via Kaplan-Meier analysis for metastasis-free (MET-free) survival rates. (D) Significant difference in prognostic outcome between *CHD1*-deleted and *SPOP*-mutant subclasses via Kaplan-Meier analysis for MET-free survival rates.

Broadly, these results demonstrate that despite similar metastatic potential, *PTEN*-deleted tumors show evidence of locoregional progression at radical prostatectomy, while *CHD1* deletion is only associated with higher Gleason score, suggesting distinct pathways to metastatic disease.

Discussion

PCa is a molecularly heterogeneous disease, with a specific temporal sequence of early and late genomic events (4–6). Previous studies demonstrated that *ERG* rearrangements and *SPOP* mutations designate mutually exclusive tumor subclasses (4, 12, 13), which may represent distinct tumor lineages. In this study, we examined co-occurring genomic alterations, defined their temporal sequence, and established 2 tumor lineage models of PCa progression: *ERG* overexpression followed by *PTEN* deletion, and *SPOP* mutation followed by *CHD1* deletion. By investigating these tumor lineage models, we found progressive changes in transcriptional alterations, with shared altered genes from the 2 tumor lineages enriched for broadly common tumorigenic functions (cell cycle and focal adhesion), while uniquely altered genes were enriched in distinct signaling pathways that were subsequently validated using mouse model systems. These data provide a framework for exploring the unique biology of these tumor lineages, allowing future functional studies to define the relationship between the early and late progression events using in vitro and in vivo models,

and investigate why specific early events favor specific subsequent genomic alterations.

In addition to its molecular heterogeneity, PCa also displays striking clinical variability. Here, we provide a clear link between molecular and clinical progression, showing that genomic alterations defined as late progression events are associated with worse prognosis. Furthermore, our data suggest that in terms of impact on clinical prognosis, the initial tumor lineage is much less important than the degree of progression within the lineage, nominating biomarker-driven strategies for risk stratification and surveillance of patients.

Clinical and pathological features after prostatectomy are the critical components of risk assessment in PCa. Current risk stratification guidelines, such as those from the NCCN (47), AUA/ASTRO (48), and EAU-ESTRO-SIOG (49) rely heavily on surgical pathology, including tumor stage and lymph node invasion status, to guide management decisions for clinically localized disease. In this study, we found that different tumor lineages showed similar prognostic outcomes with respect to distant metastasis, but divergent clinical and pathological features at radical prostatectomy: *PTEN*-deleted tumors were more likely to harbor stage-associated adverse pathological features, such as lymph node invasion, extracapsular extension, and seminal vesicle invasion, whereas *CHD1* deletion was only associated with higher Gleason score but not stage-associated adverse features. These data potentially nominate distinct clinical pathways toward distant metastasis in specific molecular subtypes of PCa, with the potential to provide guidance for therapies and imaging focused on specific patterns of disease progression. Whether tumor lineages and molecular subclasses will add clinical value to current risk stratification tools remains unclear, and need to be prospectively tested in future clinical studies. However, these data do provide compelling rationale to consider molecular subclass in future clinical trial designs.

The relationship between specific molecular events in cancers can offer insight into functional interactions; co-occurrence can suggest cooperativity or predisposition, while mutual exclusivity can suggest epistasis or synthetic lethality. There is evidence for several of these interactions regarding PCa molecular features. *SPOP* and *ERG* have been reported as both functionally redundant (50, 51) and synthetically unfavorable (13, 52). Similarly, *CHD1* and *PTEN* have been reported to be synthetically essential (53–55) through effects on the immune microenvironment (56). Here, we present the concept that these relationships are related to tumor lineage, established by early events (*SPOP* mutation and *ERG* fusion), with specific late events (*CHD1* and *PTEN* deletion) restricted to each lineage. This paradigm has implications for the distinct biology, progression pathways, and clinical features of these disease subtypes, which need to be further explored in both preclinical and clinical studies. Moreover, current genomic and clinical data are derived from bulk tumor samples and limited by intratumor heterogeneity. Molecular and clinical progression for distinct subtypes needs to be further investigated at the single-cell level.

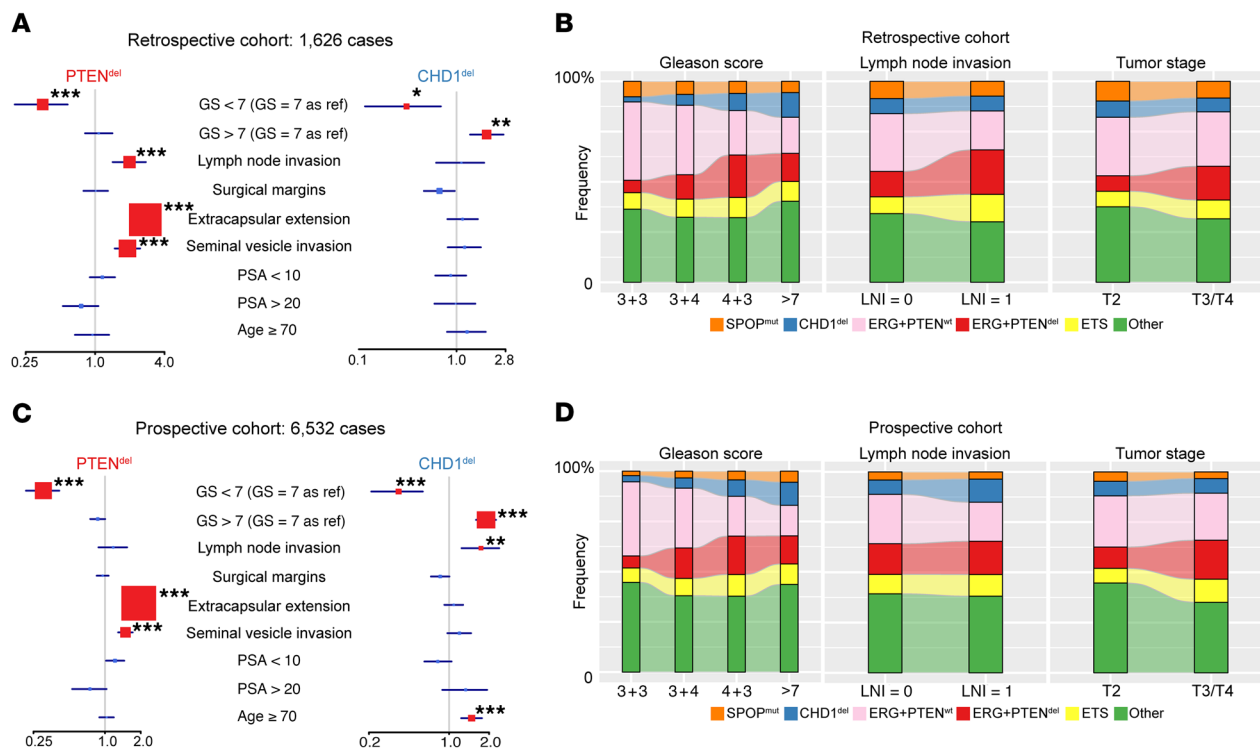


Figure 5. Distinct pathological characteristics in *CHD1*- and *PTEN*-deleted subclasses from the Decipher retrospective and prospective cohorts. (A) Clinical and pathological difference between *PTEN*- and *CHD1*-deleted status in the Decipher retrospective cohort ($n = 1,626$) via univariate analyses, with other samples as reference. GS, Gleason score. **(B)** Alluvial diagrams of Gleason scores, lymph node invasion status (LNI), and tumor stages from molecular subclasses in retrospective cohort. Different colors represent molecular subclasses. **(C)** Clinical and pathological difference between *PTEN*- and *CHD1*-deleted status in the Decipher prospective cohort ($n = 6,532$) via univariate analyses, with other samples as reference. In **A** and **C**, box size indicates significance and red color indicates $P < 0.05$. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. **(D)** Alluvial diagrams of Gleason scores, lymph node invasion status, and tumor stages from defined molecular subclasses in prospective cohort.

In conclusion, we established mutually exclusive tumor lineage models of PCa progression: *ERG/PTEN* and *SPO/CHD1*. Using transcriptional classifiers to categorize progressive events, we predict lineage and progression status from a large population of human patients, and find that molecularly defined late progression events are associated with worse clinical outcome, but may be associated with distinct clinical pathways toward metastasis. More broadly, these data suggest a paradigm in which specific subtypes of PCa follow distinct molecular pathways of tumor progression, and the interpretation of common risk stratification parameters such as locoregional tumor staging may be influenced by the underlying tumor lineage and degree of molecular progression.

Methods

RNA-seq and microarray data of PCa samples. A total of 8,622 radical prostatectomy tumor expression profiles were used for training, testing, and validation in SCaPT models. For training and testing, we utilized RNA-seq expression and DNA genomic alteration data from TCGA prostate cancer project ($n = 333$) (4), and Human Exon 1.0 ST microarray data from a Memorial Sloan Kettering Cancer Center (MSKCC) primary PCa ($n = 131$) cohort (1). For validation, the expression profiles of retrospective ($n = 1,626$) and prospective ($n = 6,532$) cohorts were derived from the Decipher GRID registry (NCT02609269). The retrospective GRID cohort was pooled from 7 published microarray studies: Cleveland Clinic (CCF, ref. 57), Eras-

mus MC (EMC, ref. 19), Johns Hopkins (JHMI, ref. 25), MSKCC (1), Mayo Clinic (Mayo I and Mayo II, refs. 20, 21), and Thomas Jefferson University (TJU, ref. 22). The prospective GRID cohort was from clinical use of the Decipher test (GenomeDx Biosciences Laboratory). DNA and RNA from the TCGA and MSKCC cohorts were extracted from fresh frozen radical prostatectomy tumor tissue, as previously described (1, 4). RNA from the GRID cohorts was extracted from routine formalin-fixed, paraffin-embedded radical prostatectomy tumor tissues, amplified and hybridized to Human Exon 1.0 ST microarrays (Thermo Fisher Scientific) (18). Gene expression data of the ICGC PRAD-CA cohort ($n = 144$) were downloaded from the ICGC Data Portal (<https://dcc.icgc.org/>), and genomic alterations were annotated based on a previous study (33).

Clonality analysis. Primary PCa and metCRPC samples from TCGA (4) and the SU2C-PCF (16) cohorts were uniformly reprocessed through an in-house pipeline. Briefly, the output includes somatic copy-number aberrations (SCNAs), computed on paired whole-exome samples by CNVkit (58), and tumor ploidy and purity assessments together with SCNA clonality analysis by CLONETv2 (59, 60). CLONETv2 uses the sequence coverage of the set of patient’s specific germline heterozygous SNPs at sites of somatic hemizygous and homozygous deletions to assess the fraction of tumor cells harboring the aberration, i.e., the clonality of the deletion. Similarly, SNV clonality estimation is performed, where the variant allelic fraction (VAF) is adjusted for tumor purity; observed and theoretical VAFs were com-

pared to estimate the percentage of aberrant tumor cells. Finally, for each aberration, minimum and maximum estimated clonality values are provided to account for sequencing coverage instability; lower and upper bounds allow the discretization of continuous clonality values into clonal or subclonal statuses. The clonality analysis was performed on top-quality data samples, including 297 primary and 355 metastatic PCa samples, respectively.

Differential expression analysis from normal to early and late progression events. Based on the strategy from time-course RNA-seq analysis, we performed differential expression analysis to identify early and late progression events based on multiple categories: from normal prostate to *ERG* fusion without *PTEN* deletion and then *ERG* fusion with *PTEN* deletion samples ($N \rightarrow ERG \rightarrow PTEN$), from normal prostate to *ERG* fusion with *PTEN* deletion and then *ERG* fusion without *PTEN* deletion samples ($N \rightarrow PTEN \rightarrow ERG$), from normal prostate to *SPOP* mutation without *CHD1* deletion and then *SPOP* mutation with *CHD1* deletion samples ($N \rightarrow SPOP \rightarrow CHD1$), and from normal prostate to *SPOP* mutation with *CHD1* deletion and then *SPOP* mutation without *CHD1* deletion samples ($N \rightarrow CHD1 \rightarrow SPOP$), via ImpulseDE2 (32), based on TCGA-PRAD reads count data (<https://portal.gdc.cancer.gov/>) (4). The transiently and progressively upregulated and downregulated genes were derived from ImpulseDE2 output at $FDR < 10^{-10}$, and compared between 2 tumor lineages. The altered signaling pathways from *Pten*-deleted mouse tissue and *Chd1*-deleted mouse organoid were downloaded from previous studies (34, 35). The enriched pathway analyses were performed via DAVID (61) and GSEA (62–64) at $P < 0.05$. The enriched signaling pathways associated with lymph node invasion were identified by comparing the samples with and without lymph node invasion via DESeq2 (65) and GSEA from TCGA-PRAD RNA-seq data. The upstream regulators were predicted by Ingenuity Pathway Analysis (IPA) software (36) at $P < 0.05$ based on differentially expressed genes.

Development of SCaPT models to classify *PTEN* and *CHD1* deletions purely from transcriptional data. The expression of *PTEN* and *CHD1* was not correlated with genomic alterations well, and other mechanisms such as methylation and signaling alterations may contribute to transcriptional alterations. Therefore, we developed 2 SCaPT models to classify *PTEN* and *CHD1* deletions purely from transcriptional data, based on a similar strategy from a previous *SPOP* prediction model (7). The RNA-based classifier via SCaPT models includes feature and model selections, on the basis of an SVM model (40–42). First, transcriptional features were selected for *PTEN* and *CHD1* deletions, by comparing *PTEN*-deleted and WT samples with *ETS* family gene fusions, and comparing *CHD1*-deleted and WT samples that lack *ETS* family gene fusions, via Wilcoxon's rank-sum test and controlled for false discovery. Second, we included SVM in SCaPT model selection with different cost parameters, because the prediction for *PTEN* deletions from samples with *ETS* family gene fusions and prediction for *CHD1* deletions from samples lacking *ETS* family gene fusions are 2-class classification problems, and SVM is intrinsically suited for 2-class problems. Then, we performed 10-fold cross validation to find the best feature and cost with highest sensitivities and specificities on TCGA training data set. Finally, we established 2 RNA-based classifiers of SCaPT models including *PTEN* transcription signature with 45 differentially expressed genes, and *CHD1* transcription signature with 148 differentially expressed genes. In our SCaPT models, training data were defined as the transcriptional z scores of *PTEN* and *CHD1*

features from TCGA cohort. Testing data were the transcriptional z scores from RNA-seq or microarray expression data of *PTEN* and *CHD1* signatures.

Classification of molecular subclasses in PCa by decision tree. In each individual study of retrospective and prospective GRID cohorts, *ERG*+ (*ERG* overexpressing) and *ETS*+ (*ETV1*, *ETV4*, *ETV5*, or *FLII* overexpressing) subclasses were firstly classified based on previously developed microarray-based expression thresholds (7, 18). Next, we predicted *SPOP*-mutant and *PTEN*- and *CHD1*-deleted subclasses using SCaPT models. Then we classified samples with both *ERG* overexpression and predicted *PTEN*-deleted status as *ERG*+*PTEN*^{del} subclass, and *ERG*-overexpressing and predicted *PTEN* WT status as *ERG*+*PTEN*^{wt} subclass. For *ERG* and *ETS* WT samples, we classified *CHD1*^{del} subclass based on predicted *CHD1*-deleted status without predicted *SPOP*-mutant status, *SPOP*^{mut} subclass without predicted *CHD1*-deleted status, and *SPOP*^{mut}+*CHD1*^{del} with both predicted *CHD1*-deleted and *SPOP*-mutant status. The remaining samples without *PTEN*, *CHD1*, and *SPOP* calling and outlier expression were considered as “other” subclass (Supplemental Figure 13).

Statistics. Statistical analyses were performed in R v3.5.1 (R Foundation). All statistical tests were 2-sided, and a P value of less than 0.05 was considered significant. We evaluated the associations between molecular subclasses and patient outcomes including biochemical recurrence, metastasis, and PCa-specific mortality, based on Kaplan-Meier analysis. Univariate logistic regression analyses were performed on the combined cohort to test the statistical association between molecular subclasses and clinical variables, including age, race, preoperative PSA, Gleason score, lymph node invasion, surgical margin status, extracapsular extension, and seminal vesicle invasion.

Author contributions

CEB, AS, and DL designed the research studies. DL, MAA, IG, DP, YL, JES, RJK, BJT, EAK, RBD, and ED conducted experiments and acquired data. DL, MAA, and DP analyzed the data. DL and CEB wrote the manuscript. MAA, IG, DP, BJT, FD, and AS helped to revise the manuscript.

Acknowledgments

We are grateful to the prostate cancer patients and families who contributed to this research. We thank The Cancer Genome Atlas Research Network (TCGA) for providing the prostate cancer genomic and transcriptomic data. We thank the Weill Cornell Medicine Genomics and Epigenomics Core Facility and the MSKCC cBioPortal. This work was supported by the Prostate Cancer Foundation (to CEB, DL, and MAA), NCI (P50CA211024 and R01CA233650 to CEB), a Urology Care Foundation Rising Star in Urology Research Award (to CEB), Damon Runyon Cancer Research Foundation MetLife Foundation Family Clinical Investigator Award (to CEB), US Department of Defense Postdoctoral Fellowship (W81XWH-17-1-0137 to MAA), and Foundation AIRC (IG 19221 to FD).

Address correspondence to: Christopher E. Barbieri, Belfer Research Building, BRB 1452, 413 East 69th Street, New York, New York 10021, USA. Phone: 646.962.6295; Email: chb9074@med.cornell.edu. Or to: Andrea Sboner, Weill Greenberg Center, 1305 York Avenue, New York, New York 10021, USA. Phone: 646.962.2241; Email: ans2077@med.cornell.edu.

1. Taylor BS, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell*. 2010;18(1):11–22.
2. Berger MF, et al. The genomic complexity of primary human prostate cancer. *Nature*. 2011;470(7333):214–220.
3. Barbieri CE, et al. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nat Genet*. 2012;44(6):685–689.
4. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell*. 2015;163(4):1011–1025.
5. Espiritu SMG, et al. The evolutionary landscape of localized prostate cancers drives clinical aggression. *Cell*. 2018;173(4):1003–1013.
6. Gerhauser C, et al. Molecular evolution of early-onset prostate cancer identifies molecular risk markers and clinical trajectories. *Cancer Cell*. 2018;34(6):996–1011.e8.
7. Liu D, et al. Impact of the SPOP mutant subtype on the interpretation of clinical parameters in prostate cancer. *JCO Precis Oncol*. 2018;2:1–13.
8. Lapointe J, et al. Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. *Cancer Res*. 2007;67(18):8504–8510.
9. Tomlins SA, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature*. 2007;448(7153):595–599.
10. Blattner M, et al. SPOP mutations in prostate cancer across demographically diverse patient cohorts. *Neoplasia*. 2014;16(1):14–20.
11. Blattner M, et al. SPOP mutation drives prostate tumorigenesis *in vivo* through coordinate regulation of PI3K/mTOR and AR signaling. *Cancer Cell*. 2017;31(3):436–451.
12. Arora K, Barbieri CE. Molecular subtypes of prostate cancer. *Curr Oncol Rep*. 2018;20(8):58.
13. Shoag J, et al. SPOP mutation drives prostate neoplasia without stabilizing oncogenic transcription factor ERG. *J Clin Invest*. 2018;128(1):381–386.
14. Boysen G, et al. SPOP mutation leads to genomic instability in prostate cancer. *Elife*. 2015;4:e09207.
15. Baca SC, et al. Punctuated evolution of prostate cancer genomes. *Cell*. 2013;153(3):666–677.
16. Abida W, et al. Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci U S A*. 2019;116(23):11428–11436.
17. Teo MY, et al. Treatment of advanced prostate cancer. *Annu Rev Med*. 2019;70:479–499.
18. Tomlins SA, et al. Characterization of 1577 primary prostate cancers reveals novel biological and clinicopathologic insights into molecular subtypes. *Eur Urol*. 2015;68(4):555–567.
19. Boormans JL, et al. Identification of TDRD1 as a direct target gene of ERG in primary prostate cancer. *Int J Cancer*. 2013;133(2):335–345.
20. Erho N, et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS One*. 2013;8(6):e66855.
21. Karnes RJ, et al. Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at risk patient population. *J Urol*. 2013;190(6):2047–2053.
22. Den RB, et al. Genomic prostate cancer classifier predicts biochemical failure and metastases in patients after postoperative radiation therapy. *Int J Radiat Oncol Biol Phys*. 2014;89(5):1038–1046.
23. Klein EA, et al. A genomic classifier improves prediction of metastatic disease within 5 years after surgery in node-negative high-risk prostate cancer patients managed by radical prostatectomy without adjuvant therapy. *Eur Urol*. 2015;67(4):778–786.
24. Johnson MH, et al. SPINK1 defines a molecular subtype of prostate cancer in men with more rapid progression in an at risk, natural history radical prostatectomy cohort. *J Urol*. 2016;196(5):1436–1444.
25. Ross AE, et al. Tissue-based genomics augments post-prostatectomy risk stratification in a natural history cohort of intermediate- and high-risk men. *Eur Urol*. 2016;69(1):157–165.
26. Zhao SG, et al. The landscape of prognostic outlier genes in high-risk prostate cancer. *Clin Cancer Res*. 2016;22(7):1777–1786.
27. Nguyen PL, et al. Ability of a genomic classifier to predict metastasis and prostate cancer-specific mortality after radiation or surgery based on needle biopsy specimens. *Eur Urol*. 2017;72(5):845–852.
28. Spratt DE, et al. Individual patient-level meta-analysis of the performance of the decipher genomic classifier in high-risk men after prostatectomy to predict development of metastatic disease. *J Clin Oncol*. 2017;35(18):1991–1998.
29. Zhao SG, et al. Associations of luminal and basal subtyping of prostate cancer with prognosis and response to androgen deprivation therapy. *JAMA Oncol*. 2017;3(12):1663–1672.
30. Demichelis F, et al. Distinct genomic aberrations associated with ERG rearranged prostate cancer. *Genes Chromosomes Cancer*. 2009;48(4):366–380.
31. Armenia J, et al. The long tail of oncogenic drivers in prostate cancer. *Nat Genet*. 2018;50(5):645–651.
32. Fischer DS, et al. Impulse model-based differential expression analysis of time course sequencing data. *Nucleic Acids Res*. 2018;46(20):e119.
33. Fraser M, et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature*. 2017;541(7637):359–364.
34. Chen Y, et al. ETS factors reprogram the androgen receptor cistrome and prime prostate tumorigenesis in response to PTEN loss. *Nat Med*. 2013;19(8):1023–1029.
35. Augello MA, et al. CHD1 loss alters ar binding at lineage-specific enhancers and modulates distinct transcriptional programs to drive prostate tumorigenesis. *Cancer Cell*. 2019;35(4):603–617.
36. Krämer A, et al. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014;30(4):523–530.
37. Rodriguez-Bravo V, et al. The role of GATA2 in lethal prostate cancer aggressiveness. *Nat Rev Urol*. 2017;14(1):38–48.
38. Wang Y, et al. GATA2 negatively regulates PTEN by preventing nuclear translocation of androgen receptor and by androgen-independent suppression of PTEN transcription in breast cancer. *Hum Mol Genet*. 2012;21(3):569–576.
39. Groner AC, et al. TRIM24 is an oncogenic transcriptional activator in prostate cancer. *Cancer Cell*. 2016;29(6):846–858.
40. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273–297.
41. Bennett KP, Campbell C. Support vector machines: hype or hallelujah? *SIGKDD Explor Newsl*. 2000;2(2):1–13.
42. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Trans Intell Syst Technol*. 2011;2(3):1–27.
43. Oh-Hohenhorst SJ, et al. CHD1 loss negatively influences metastasis-free survival in RO-resected prostate cancer patients and promotes spontaneous metastasis *in vivo* [published online January 7, 2021]. *Cancer Gene Ther*. <https://doi.org/10.1038/s41417-020-00288-z>.
44. Faubert B, et al. Metabolic reprogramming and cancer progression. *Science*. 2020;368(6487):eaaw5473.
45. Murray PJ. Cancer metastasis linked to macrophage size, shape, and metabolism. *J Exp Med*. 2020;217(11):e20201259.
46. Wei Q, et al. Metabolic rewiring in the promotion of cancer metastasis: mechanisms and therapeutic implications. *Oncogene*. 2020;39(39):6139–6156.
47. Mohler J, et al. Prostate Cancer, version 2.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw*. 2019;17(5):479–505.
48. Sanda MG, et al. Clinically localized prostate cancer: AUA/ASTRO/SUO guideline. Part I: risk stratification, shared decision making, and care options. *J Urol*. 2018;199(3):683–690.
49. Mottet N, et al. EAU-ESTRO-SIOG guidelines on prostate cancer. Part I: screening, diagnosis, and local treatment with curative intent. *Eur Urol*. 2017;71(4):618–629.
50. An J, et al. Truncated ERG oncoproteins from TMPRSS2-ERG fusions are resistant to SPOP-mediated proteasome degradation. *Mol Cell*. 2015;59(6):904–916.
51. Gan W, et al. SPOP promotes ubiquitination and degradation of the ERG oncoprotein to suppress prostate cancer progression. *Mol Cell*. 2015;59(6):917–930.
52. Bernasocchi T, et al. Dual functions of SPOP and ERG dictate androgen therapy responses in prostate cancer. *Nat Commun*. 2021;12(1):734.
53. Zhao D, et al. Synthetic essentiality of chromatin remodelling factor CHD1 in PTEN-deficient cancer. *Nature*. 2017;542(7642):484–488.
54. Trock BJ, et al. PTEN loss and chromosome 8 alterations in Gleason grade 3 prostate cancer cores predicts the presence of un-sampled grade 4 tumor: implications for active surveillance. *Mod Pathol*. 2016;29(7):764–771.
55. Fontugne J, et al. Recurrent prostate cancer genomic alterations predict response to brachytherapy treatment. *Cancer Epidemiol Biomarkers Prev*. 2014;23(4):594–600.
56. Zhao D, et al. Chromatin regulator CHD1 remodels the immunosuppressive tumor microenvironment in PTEN-deficient prostate cancer. *Cancer Discov*. 2020;10(9):1374–1387.
57. Klein EA, et al. A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. *Eur Urol*. 2014;66(3):550–60.
58. Talevich E, et al. CNVkit: genome-wide copy number detection and visualization from targeted DNA sequencing. *PLoS Comput Biol*.

- 2016;12(4):e1004873.
59. Prandi D, Demichelis F. Ploidy- and purity-adjusted allele-specific DNA analysis using CLONETv2. *Curr Protoc Bioinformatics*. 2019;67(1):e81.
60. Prandi D, et al. Unraveling the clonal hierarchy of somatic genomic aberrations. *Genome Biol*. 2014;15(8):439.
61. Huang da W, et al. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc*. 2009;4(1):44–57.
62. Subramanian A, 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(43):15545–15550.
63. Liberzon A, et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739–1740.
64. Liberzon A, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417–425.
65. Love MI, et al. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.