The Oncologist®

Genitourinary Cancer

Identification of Somatic Gene Signatures in Circulating Cell-Free DNA Associated with Disease Progression in Metastatic Prostate Cancer by a Novel Machine Learning Platform

Edwin Lin,[a,b,†] Andrew W. Hahn,[c,†] Roberto H. Nussenzveig,[a,†] Sergiusz Wesolowski,[b] Nicolas Sayegh,[a] Benjamin L. Maughan,[a] Taylor McFarland,[a] Nityam Rathi,[a] Deepika Sirohi,[a] Guru Sonpavde,[d] Umang Swami,[a] Manish Kohli,[a] Thereasa Rich,[e] Oliver Sartor,[f] Mark Yandell,[b] Neeraj Agarwal[a]

Divisions of [a]Oncology, Departments of Internal Medicine, Huntsman Cancer Institute and [a]Human Genetics, University of Utah, Salt Lake City, Utah, USA; [c]Division of Cancer Medicine, MD Anderson Cancer Center, Houston, Texas, USA; [d]Department of Hematology/Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA; [e]Guardant Health, Inc, Redwood City, California, USA; [f]Department of Oncology, Tulane University, New Orleans, Louisiana, USA

[†]Contributed equally

Disclosures of potential conflicts of interest may be found at the end of this article.

Key Words. Metastatic prostate cancer • Castration-resistant • Castration-sensitive • Cell-free DNA • Next-generation sequencing • Machine learning • Genomics

Abstract

Purpose. Progression from metastatic castration-sensitive prostate cancer (mCSPC) to a castration-resistant (mCRPC) state heralds the lethal phenotype of prostate cancer. Identifying genomic alterations associated with mCRPC may help find new targets for drug development. In the majority of patients, obtaining a tumor biopsy is challenging because of the predominance of bone-only metastasis. In this study, we hypothesize that machine learning (ML) algorithms can identify clinically relevant patterns of genomic alterations (GAs) that distinguish mCRPC from mCSPC, as assessed by next-generation sequencing (NGS) of circulating cell-free DNA (cfDNA).

Experimental Design. Retrospective clinical data from men with metastatic prostate cancer were collected. Men with NGS of cfDNA performed at a Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory at time of diagnosis of mCSPC or mCRPC were included. A combination of supervised and unsupervised ML algorithms was used to obtain biologically interpretable, potentially actionable insights into genomic signatures that distinguish mCRPC from mCSPC.

Results. GAs that distinguish patients with mCRPC ($n = 187$) from patients with mCSPC ($n = 154$) (positive predictive value = 94%, specificity = 91%) were identified using supervised ML algorithms. These GAs, primarily amplifications, corresponded to androgen receptor, Mitogen-activated protein kinase (MAPK) signaling, Phosphoinositide 3-kinase (PI3K) signaling, G1/S cell cycle, and receptor tyrosine kinases. We also identified recurrent patterns of gene- and pathway-level alterations associated with mCRPC by using Bayesian networks, an unsupervised machine learning algorithm.

Conclusion. These results provide clinical evidence that progression from mCSPC to mCRPC is associated with stereotyped concomitant gain-of-function aberrations in these pathways. Furthermore, detection of these aberrations in cfDNA may overcome the challenges associated with obtaining tumor bone biopsies and allow contemporary investigation of combinatorial therapies that target these aberrations. ***The Oncologist*** 2021;26:1–10

Implications for Practice: The progression from castration-sensitive to castration-resistant prostate cancer is characterized by worse prognosis and there is a pressing need for targeted drugs to prevent or delay this transition. This study used machine learning algorithms to examine the cell-free DNA of patients to identify alterations to specific pathways and genes associated with progression. Detection of these alterations in cell-free DNA may overcome the challenges associated with obtaining tumor bone biopsies and allow contemporary investigation of combinatorial therapies that target these aberrations.

Correspondence: Neeraj Agarwal, M.D., Huntsman Cancer Institute, University of Utah, 2000 Circle of Hope Drive, Suite 5726, Salt Lake City, Utah 84112, USA. Telephone: 801-213-5658; e-mail: neeraj.agarwal@hci.utah.edu; or Mark Yandell, Ph.D., Eccles Institute of Human Genetics, University of Utah, 15 N. 2030 E., Salt Lake City, Utah 84112, Telephone: 801-587-7707; e-mail: myandell@genetics.utah.edu Received March 4, 2021; accepted for publication June 4, 2021. http://dx.doi.org/10.1002/onco.13869

Introduction

Prostate cancer is the second most common cause of cancer-related deaths in American men [1]. Metastatic castration-resistant prostate cancer (mCRPC) is the lethal form of disease. When newly diagnosed, metastatic prostate cancer (mPC) responds rapidly to androgen deprivation therapy (ADT) by medical or surgical castration, and therefore, this state is called metastatic castration-sensitive prostate cancer (mCSPC). However, almost all men with mCSPC eventually experience disease progression on ADT after a median of 16 months to mCRPC. In current studies, the median overall survival after onset of mCRPC is ~3 years [2].

In 2004, chemotherapy with docetaxel, a microtubule inhibitor, was approved for the treatment of mCRPC [3]. This was followed by approval of novel androgen axis inhibitors abiraterone and enzalutamide; a novel taxane, cabazitaxel; a vaccine, sipuleucel-T; and a radiopharmaceutical, radium-223, in 2013 [4]. However, all of these agents are associated with a modest improvement in median overall survival in the mCRPC setting in the range of 3–4 months each. More recently, multiple phase III trials showed improved survival outcomes by using chemotherapy with docetaxel, or novel androgen axis inhibitors in combination with ADT for the treatment of castration-sensitive disease. These results have led to these agents now increasingly being used for treatment of men with mCSPC [3, 5]. Movement of these agents, hitherto available for treating mCRPC, to treatment of castration-sensitive disease poses further challenges to treatment of mCRPC by limiting the number of therapeutic options for these men.

Single-agent targeted therapies for mCRPC, such as Poly [ADP-Ribose] Polymerase (PARP) inhibitors, just received regulatory approval, and many others are under active investigation [6, 7]. However, mCRPC can rapidly become resistant against single-agent therapies through heterogeneous patterns of genomic alteration [8]. In order to improve patient outcomes, combinatorial treatments that target critical signaling pathways or anticipate mechanisms of resistance in mCRPC are needed.

Next-generation sequencing (NGS) of cell-free DNA (cfDNA) is used as a relatively noninvasive clinical assay to detect somatic genomic alterations (GAs) in metastatic solid tumors [9]. Currently, PARP inhibitors are the only molecularly targeted drugs approved for mPC. One of the major reasons for the lack of multiple approved targeted options is the absence of readily biopsiable sites of metastasis in these patients. Eighty percent of patients with mPC have bone-only metastases, which are not feasible to biopsy in the vast majority of cases, a critical factor limiting development of molecularly targeted therapy in mPC. Genomic profiling of tumor somatic alterations in circulating cfDNA has the potential to circumvent this limitation. Machine learning (ML) algorithms extract generalizable patterns and associations from large, complex data sets (e.g., cancer genomics data) [10]. In principle, ML algorithms can be applied to large genomic data sets in order to reveal novel patterns of GAs associated with the progression from mCSPC to mCRPC. These patterns of genomic evolution can, in turn, correspond to clinically actionable mechanistic insights that are applicable to treatment development. Although machine learning has been applied to early detection of cancer in cfDNA, to our knowledge there have been no reports describing its use for identification of molecular targets of disease progression in cfDNA. In this article, we hypothesize that ML algorithms can identify clinically relevant patterns of genomic alterations that distinguish mCRPC from mCSPC, as assessed by NGS of circulating cfDNA.

Materials and Methods

Study Design and Cohort Characteristics

This study was performed with approval from the University of Utah Institutional Review Board (IRB #67518). Clinical data were retrospectively collected on men with metastatic prostate cancer. mCRPC was diagnosed per Prostate Cancer Working Group. 2.0 criteria [11] by either prostate-specific antigen (PSA) or radiographic progression, or by clinical progression as determined by the investigator, whichever occurred earlier. In this cohort, 341 men had undergone Guardant360 (Guardant Health, Inc., Redwood City, CA) testing. At the time of testing, 154 men had mCSPC and 187 men had mCRPC. Clinical characteristics of the cohort are summarized in Table 1.

Table 1. Clinical characteristics of patients with mCSPC and patients with mCRPC

Variable	mCSPC ($n = 154$)	mCRPC ($n = 187$)	Mann-Whitney p value
Age, yr, median (range)	65 (40–97)	63 (40–89)	.14
Gleason score, median (IQR)	8 (7–9)	8 (7–9)	.24
PSA, ng/mL, median (IQR)	20.1 (7.5–81.4)	20.9 (7–90)	.34
Number of systemic lines prior to NGS[a]			
0	NA	59	NA
1	NA	54	NA
2	NA	37	NA
3	NA	37	NA

IQR represents the range between the 25th and 75th percentiles.
[a]Lines of systemic therapy included, but were not limited to, the following agents: abiraterone, enzalutamide, docetaxel, and radium-223.
Abbreviations: IQR, interquartile range; mCRPC, metastatic castration-resistant prostate cancer; mCSPC, metastatic castration-sensitive prostate cancer; NA, not applicable; NGS, next-generation sequencing; PSA, Prostate-specific antigen.

All patient reports were included in this study, even those with no detectable alterations. Our rationale for this decision was to provide a more accurate and generalizable representation of ML performance on new data. Additionally, the diagnostic sensitivity and specificity of individual GAs could be assessed. Furthermore, the detection of a GA is dependent on technical factors and underlying biological differences that can lead to a difference in cfDNA shedding in mCSPC versus mCRPC. The inclusion of patient samples without detectable alterations would empirically account for the combination of these latent biological and technical variables.

Next-Generation Sequencing, Variant Calling, and Functional Annotation

NGS of cfDNA was performed during routine standard-of-care treatment using a commercially available panel, Guardant360. Briefly, cfDNA isolation, library construction and sequencing, and quality-control methods were performed as previously described [12, 13]. Variant calling was done by a validated custom bioinformatics pipeline that uses molecular barcoding and double-stranded consensus sequence representation to achieve >99.99% per-position analytic specificity [12, 13]. Single nucleotide variants (SNVs), small insertions/deletions (InDels), gene rearrangements/fusions, and copy number amplifications (CNAs) were called in a panel of cancer-relevant genes (supplemental online Table 1). CNAs were determined by evaluating whether gene-level probe counts were overrepresented compared with the baseline diploid count, established within each patient's own sample, using prespecified statistical decision thresholds [12, 13]. The reportable ranges for SNVs, InDels, fusions, and CNAs are >0.04%, >0.02%, >0.04%, and > 2.12 copies, respectively [12, 13]. Germline alterations were filtered out as previously described [12, 13]. Variants of unknown significance and synonymous alterations were excluded from further analyses. Only characterized alterations predicted to be functionally significant by COSMIC [14], cBioPortal for Cancer Genomics [15], UniProt [16], and literature catalogued by PubMed and the International agency for research on cancer (IARC) database [17] for TP53 were used in the analysis.

Principal Components Analysis

We used principal components analysis, an unsupervised ML algorithm, for data exploration and visualization. The sci-kit learn package version 0.21.3 in a Python version 3.6.8 (Python Software Foundation, Delaware, USA) was used for this analysis. Gene alteration counts were scaled to a standard normal distribution with mean $=$ 0 and SD $=$ 1. Correlations between GAs and phenotype were visualized by plotting each sample's variance along the first two principal components. Correlations between genes, and between genes and phenotype, were identified by a loadings plot.

Supervised Machine Learning

The sci-kit learn package v0.21.3 in Python v3.6.8 was used for supervised ML classification. All available supervised machine learning algorithms in this package were trained to classify patient as mCRPC or mCSPC using somatic genomic alterations in cfDNA data, and included Bayesian algorithms (Bernoulli, Gaussian, and multinomial naïve Bayes), kernel-based algorithms (Gaussian process, support vector machine, quadratic discriminant analysis), tree-based algorithms (random forest, gradient boosting, adaptive boosting), logistic regression, deep neural networks, and k-nearest neighbors. A dummy classifier using stratified estimates was used as a negative control.

Evaluation of Supervised Machine Learning Algorithm Performance

In order to avoid optimistic bias in model evaluation [18], we used a nested cross-validation strategy consisting of an inner cross-validation for hyperparameter tuning and parameter selection, followed by an outer cross-validation for model evaluation. Hyperparameter optimization for each machine learning algorithm was performed using a random search approach. Empirically, 1,000 iterations of fivefold repeated, stratified cross-validation was sufficient for consistent model selection. Under- and overfitting was assessed by comparing inner and outer cross-validation metrics (supplemental online Fig. 1B). Algorithm performance was assessed by overall accuracy, sensitivity, specificity, positive predictive value (PPV), negative predictive value, and area under the curve (AUC) for the receiver operating characteristic. Performance metrics were determined using 60 iterations of repeated stratified fivefold cross-validation (supplemental online Fig. 1C, 1D). Mean performance metrics and variance did not change with 100, 200, and 500 iterations of repeat stratified cross-validation.

Identification of mCRPC-Associated Genes

In order to identify genes associated with mCRPC, we used a combination of feature selection and supervised machine learning. First, joint mutual information was calculated between each gene and phenotype. Then, subsets of genes were generated using both forward and backward feature selection, in which genes were iteratively added based on rank order of joint mutual information. Then, we trained supervised machine learning algorithms to use these sets of genes to classify samples as either mCRPC or mCSPC. The association between a gene set and mCRPC was quantified by the performance metrics of supervised machine learning algorithms fitted to it. The rank order of feature importance by joint mutual information was highly similar between feature importance in Bernoulli naïve Bayes, support vector machine with a linear kernel, logistic regression, and random forest.

Identification of mCRPC-Associated Pathways

In order to identify pathways enriched for GAs associated with mCRPC, we performed Gene Ontology enrichment analysis [19, 20]. The optimal subset of genes identified by recursive feature selection to be informative for classifying mCRPC were used for this analysis. The analysis type used was the PANTHER Overrepresentation Test (released July 11, 2019). The Reactome pathway database (version 65, released March 12, 2019) was used for annotation.

Homo sapiens was chosen for the reference gene list. Statistical significance was assessed by Fisher's exact test with false discovery rate correction.

Unsupervised Machine Learning

Bayesian networks were used to model conditional dependencies by edges in a directed graph. The pomegranate package in Python was used for this analysis. The exact structure was learned for a Bayesian network that consisted of phenotype and gene alterations associated with mCRPC. Gene alterations were modeled as binary variables, as this assumption was valid for the vast majority of samples. At the pathway level, Bayes nets were fitted using the presence or absence of GAs as a discrete variable, and then using GA counts in pathways as discrete variables. Relative risk was calculated by averaging the bidirectional conditional likelihood between two variables. Bayes nets were refitted to 1,000 permutations of nonparametric bootstrapping in order to obtain robust estimates and confidence intervals for network structure and conditional likelihoods.

Statistical Analysis

The two-sided Student's *t* test was used to compare the total number of GAs per patient between mCRPC and mCSPC. The χ^2 test was used to assess statistically significant differences in the frequency of GAs in each gene between mCRPC and mCSPC. False discovery rate was controlled to 0.05 using Benjamini-Hochberg (supplemental online Fig. 1A). Statistical significance was defined as $p < .05$.

Results

We first sought to determine whether GAs could be used to distinguish mCRPC from mCSPC. In order to answer this question, we retrospectively identified a cohort of patients with metastatic prostate cancer who underwent NGS of cfDNA (Materials and Methods; supplemental online Table 1) as a part of routine care ($n = 341$, mCSPC $= 154$, mCRPC $= 187$). Clinical cohort characteristics are presented in Table 1. Data exploration by principal components analysis revealed distinct clusters corresponding to patients with mCRPC and patients with mCSPC (Fig. 1A) and GAs that contributed to cluster formation (Fig. 1B). In mCSPC, no significant differences in the number of GAs were observed between patients with mCSPC on ADT versus patients who had not yet started ADT. Men with mCRPC had more GAs than men with mCSPC (mean 4.5 vs. 1.9, $p < .0001$), and more unique GAs (supplemental online Fig. 1A).

Next, we sought to identify a specific set of gene alterations associated with the development of mCRPC. Using a combination of feature selection and supervised machine learning, we found a set of 16 genes to be most highly associated with mCRPC (Fig. 1C, 1D). *AR* was the most significant predictor of mCRPC, in accordance with previous studies [21]. Supervised ML algorithms that were trained on these 16 genes showed improved performance in comparison with the same algorithms trained on the entire panel (PPV $=$ 94% vs. 88%, specificity $=$ 91% vs. 86%, AUC $=$ 0.74 vs. 0.74) (Fig. 1C, 1D; supplemental online Fig. 1B–1D). By performing Gene Ontology enrichment analysis, we found that receptor tyrosine kinases (RTK), mitogen-activated protein kinase (MAPK) signaling, phosphoinositide 3 kinase (PI3K) signaling, and G1/S cell cycle were significantly enriched for these mCRPC-associated GAs (Fig. 1E). Supervised ML algorithms trained using GA counts in these pathways showed a similarly high performance for classifying samples as mCRPC or mCSPC (supplemental online Fig. 1E). Therefore, GAs in these pathways distinguish mCRPC from mCSPC with high specificity and positive predictive value.

Next, we assessed whether recurrent patterns existed among mCRPC-associated GAs. In order to identify such patterns, we used Bayesian networks to discover statistical dependencies between GAs. The presence of GAs in *AR* increased the likelihood of mCRPC phenotype, as well as the likelihood of harboring GAs in *KIT*, *MET*, *BRAF*, and *PIK3CA* (Fig. 2A). Positive statistical interdependencies between genes in the RTK, PI3K, MAPK, and G1/S pathways were numerous, notably between *BRAF* and *PIK3CA* (Fig. 2A). More than half of patients with mCRPC harbored GAs in these pathways, which were predominantly gene amplifications (Fig. 2B–E).

At a pathway level, the probability of having an mCRPC phenotype was significantly associated with harboring GAs in the *AR* gene (Relative risk [RR]$=$ 2.32) (Fig. 3A, 3B). The presence of GAs in *AR* also increased the likelihood of harboring GAs in the G1/S pathway (RR $=$ 3.88 $\pm$ 1.03) (Fig. 3A, 3C). In turn, harboring GAs in the G1/S pathway significantly increased the likelihood of harboring GAs in the MAPK (RR $=$ 3.40 $\pm$ 1.14), PI3K (RR $=$ 3.43 $\pm$ 1.32), and RTK (RR $=$ 7.12 $\pm$ 1.90) signaling pathways (Fig. 3C, 3D). Increased numbers of GAs in AR were significantly associated with the mCRPC phenotype (supplemental online Fig. 2A, 2B). Having developed mCRPC was accompanied by increased numbers of GAs in both PI3K and G1S (supplemental online Fig. 2A–D). Finally, increased numbers of GAs in G1/S were also positively associated with increased numbers of GAs in RTK and MAPK signaling (supplemental online Fig. 2A, 2E, 2F). Drugs that are capable of targeting many of these GAs have been approved in other indications (supplemental online Table 2). In sum, mCRPC is enriched in targetable genomic alterations in RTK, MAPK, PI3K, and G1/S signaling pathways.

Discussion

In summary, we applied ML algorithms to cfDNA NGS data to discover targetable patterns of GAs associated with the progression from mCSPC to mCRPC. Unsupervised ML algorithms revealed recurrent patterns of GAs in the RTK, PI3K, MAPK, and G1/S signaling pathways that were associated with mCRPC. Supervised ML algorithms could robustly distinguish samples from patients with mCRPC from samples from patients with mCSPC using these GAs. These pathways are not targeted by currently approved therapies for mCRPC and should be explored as targets in possible combination therapies.

Our results provide direct clinical evidence showing that progression from mCSPC to mCRPC is associated with an

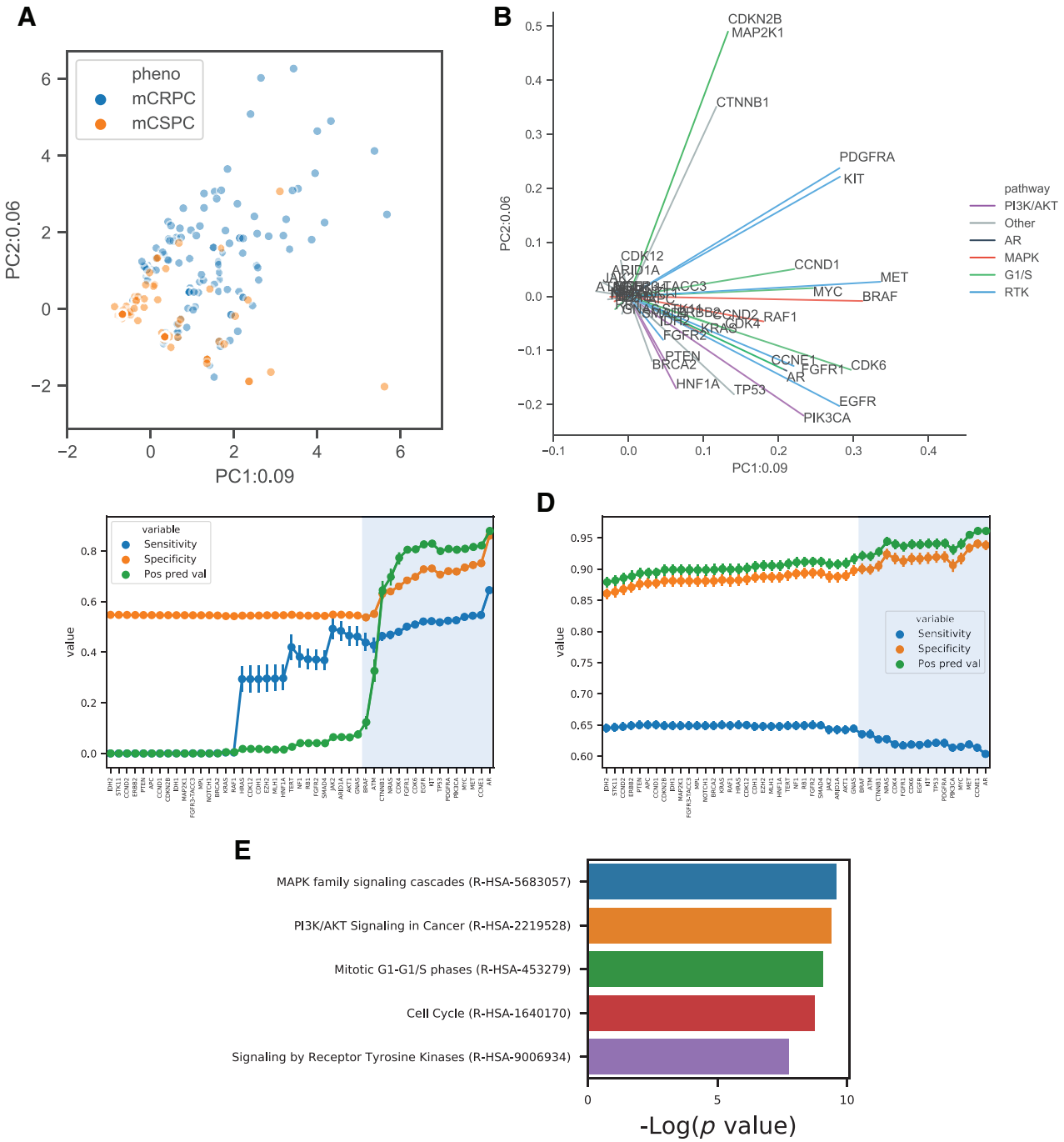


Figure 1. Genomic alterations (GAs) associated with mCRPC. **(A):** Principal components analysis shows clusters of cell-free DNA samples that correspond to patients with mCRPC and patients with mCSPC. **(B):** GAs that contribute to the formation of mCRPC and mCSPC clusters plotted in principle component space. **(C):** The association between gene sets and mCRPC using forward feature selection. Each label on the *x*-axis represents the performance of support vector classification after a gene was included. Vertical bands represent the 95th percentile confidence intervals. The top 16 genes are highlighted in blue. **(D):** The association between gene sets and mCRPC using backward feature selection. Each label on the *x*-axis represents the performance of support vector classification after a gene was eliminated. Vertical bands represent the 95th percentile confidence intervals. The top 16 genes are highlighted in blue. **(E):** Pathways significantly enriched for mCRPC-associated GAs were determined using Gene Ontology enrichment analysis.
Abbreviations: MAPK, mitogen-activated protein kinase; mCRPC, metastatic castration-resistant prostate cancer; mCSPC, metastatic castration-sensitive prostate cancer; RTK, receptor tyrosine kinase.

accumulation of GAs in both MAPK and PI3K signaling. Although previous studies have described GAs in PI3K and MAPK in prostate cancer, they did not show an association with progression from mCSPC to mCRPC [22–25]. Another study showed that MAPK gene expression was upregulated in mCRPC in comparison to primary localized prostate

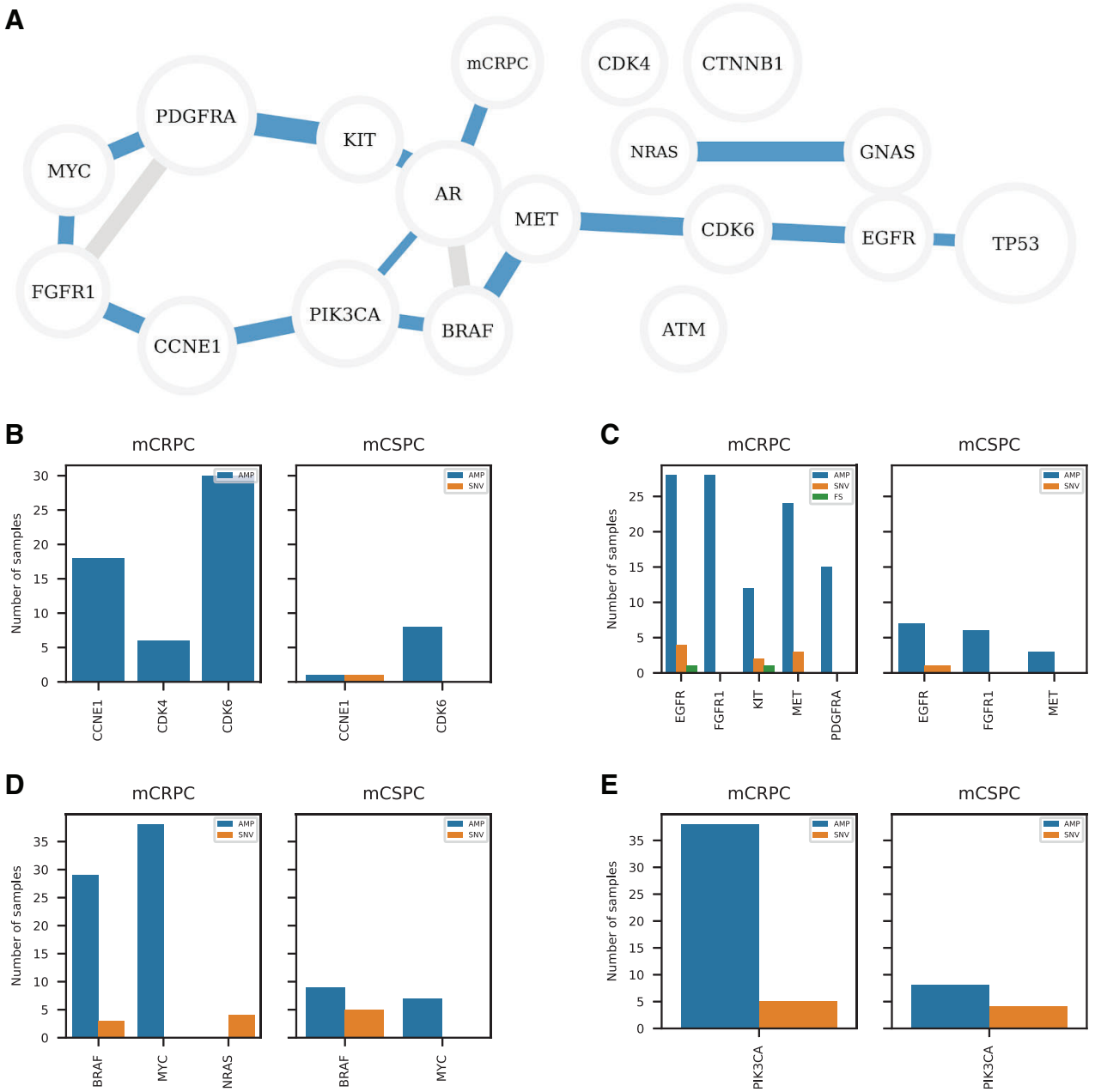
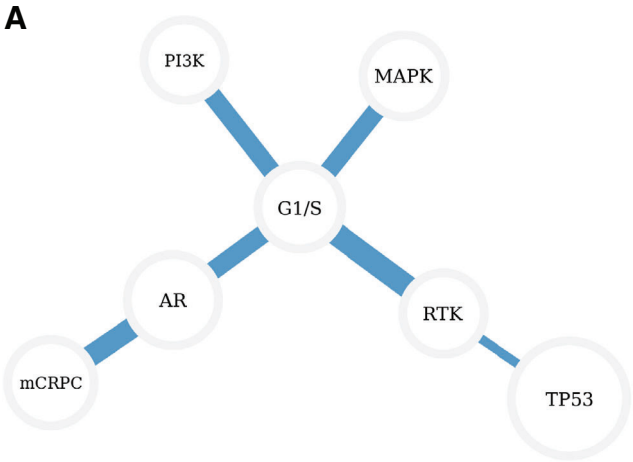


Figure 2. Gene-level patterns of genomic alterations (GAs) associated with mCRPC. **(A):** A Bayesian network of mCRPC-associated GAs. GAs are represented as nodes on the graph. The size of each node is proportional to the marginal likelihood, or frequency, of observing a GA. Blue edges represent a positive conditional probability between two GAs. Gray edges represent positive conditional probability between two GAs upon moralization of the Bayesian network. No edges are drawn between two GAs if they are statistically independent. The width of an edge is proportional to the averaged bidirectional relative risk between two GAs. **(B–E):** Alteration counts for each gene in the mCRPC and mCSPC cohorts separated by pathway.
Abbreviations: AMP, amplification; mCRPC, metastatic castration-resistant prostate cancer; mCSPC, metastatic castration-sensitive prostate cancer; SNV, single nucleotide variant.

cancers, but not mCSPC [26]. In our cohort, mCRPC was enriched for amplifications of RTK genes (i.e., *PDGFRA, KIT, EGFR, MET*, and *FGFR*) that signal through both MAPK and PI3K pathways [27]. We also show that RTK gene amplifications are positively associated with GAs in MAPK, PI3K, and G1/S signaling. Moreover, our data indicate that 13.5% of patients with mCRPC had GAs in both MAPK and PI3K signaling genes. We also observed positive statistical dependencies between *PIK3CA* and *BRAF* GAs. Preclinical models have shown a compensatory relationship between PI3K and MAPK signaling in prostate cancer [28], and that simultaneous inhibition of PI3K and MAPK signaling can inhibit mCRPC [29–31]. MAPK signaling can be targeted with-MAPK/ERK kinase (MEK) inhibitors, such as trametinib and cobimetinib, which are approved for *BRAF*-mutated melanoma and *KRAS/BRAF*-mutated colorectal cancer [32]. Although anecdotal, a heavily pretreated patient with mCRPC treated with trametinib elicited a durable clinical

A



B

Alterations	RR mCRPC given alt	Standard deviation
PI3K	1.24	0.18
AR	2.34	0.20
MAPK	1.15	0.06
G1/S	1.47	0.12
RTK	1.37	0.17
TP53	1.15	0.11

C

Alterations	RR G1/S given alt	Standard deviation
PI3K	2.94	0.90
AR	3.88	1.03
MAPK	3.12	0.95
RTK	5.83	1.27
TP53	1.64	0.41
mCRPC	2.21	0.45

D

Alterations	RR alt given G1/S	Standard deviation
PI3K	3.43	1.32
AR	3.38	0.60
MAPK	3.40	1.14
RTK	7.12	1.90
TP53	1.41	0.16
mCRPC	1.48	0.12

Figure 3. Pathway-level patterns of genomic alteration associated with mCRPC. **(A):** A Bayesian network of mCRPC-associated genomic alterations (GAs). The presence of GAs in each pathway are represented as nodes on the graph. Blue edges represent a positive conditional probability of harboring GAs in two pathways. No edges are drawn between pathways if the probability of harboring GAs in the pathways are statistically independent. The width of an edge is proportional to the averaged bidirectional relative risk between two GAs. **(B):** The relative risk of GAs in a pathway if a sample is mCRPC versus mCSPC. **(C):** The relative risk of GAs in a pathway if a sample has GAs in G1/S. **(D):** The relative risk of GAs in G1/S if a sample has GAs in other pathways.
Abbreviations: mCRPC, metastatic castration-resistant prostate cancer; mCSPC, metastatic castration-sensitive prostate cancer; RR, relative risk.

response [26]. PI3K/AKT signaling inhibitors, including buparlisib and dactolisib, have shown limited efficacy in clinical trials [29, 33]. The compensatory effects of MAPK signaling may explain this finding. Our results provide new clinical evidence that provides rationale for investigating concomitant PI3K and MAPK inhibition in clinical trials for mCRPC.

In our cohort, mCRPC was enriched for gene amplifications in G1/S, which signal downstream of MAPK [34]. Importantly, we found that the presence of amplifications in G1/S increased the likelihood of harboring GAs in RTK and MAPK. This association could be explained by selective pressure or a survival advantage, suggesting that targeting of MAPK signaling could be viable for mCRPC with G1/S amplifications. Specifically, we found that mCRPC was enriched for amplifications in genes that encode the CyclinD-Cdk4/6 complexes (i.e., *CDK4* and *CDK6*) [34]. Direct inhibitors of CDK4/6 for mCRPC are in phase II trials and clinical data are limited [35]. However, many patients are excluded because Rb loss-of-function mediates resistance

against CDK4/6 inhibitors [36] and occurs frequently in prostate cancer [37]. Direct upregulation of MAPK signaling has also been shown to contribute to CDK4/6 inhibitor resistance [38, 39]. MAPK signaling regulates cyclinD1 expression and post-translationally regulates CyclinD-Cdk4/6 complex assembly [34]. A preclinical study showed that prostate cancer cells with Rb loss–mediated resistance to G1/S inhibitors had increased MAPK activity and were sensitized to MEK inhibition [38]. In our cohort, mCRPC was also enriched for amplifications of *CCNE1*, which encodes CyclinE [40]. Cdk2, in association with CyclinE, regulates G1/S transition and S phase progression [34, 40]. Cdk2 must be transported to the nucleus in order to be activated, and its transport is dependent on MAPK activity [41, 42]. In preclinical models, MAPK inhibition decreased Cdk2-cyclin E activation by decreasing nuclear localization and phosphorylation of Cdk2 [42]. Together, our results provide clinical evidence to support further investigation of MAPK inhibitors for patients with mCRPC with G1/S gene amplifications.

Recent publications have indicated circulating tumor DNA (ctDNA) fraction is increasingly recognized as a prognostic biomarker for patients with mCRPC [43, 44]. Guardant does not include total ctDNA fraction in their patient reports. However, we estimated the total ctDNA fraction using the allelic frequencies provided in the patient reports. Our findings did not detect a significant difference in ctDNA fraction between the mCSPC and mCRPC settings. It is important to note that in this study we were not investigating whether ctDNA fraction was associated with patient outcomes; rather, the focus was in identifying gene aberrations that were associated with disease progression from the mCSPC to mCRPC settings.

Finally, these findings demonstrate the utility of our novel machine learning–based analysis of cancer genomics data obtained from circulating cfDNA. We showed that a combination of supervised and unsupervised ML algorithms could provide biologically interpretable, potentially actionable insights into genomic signatures associated with clinical phenotypes. In principle, the methods employed here could be used to identify genomic signatures associated with any clinical phenotype of interest (e.g., treatment response, prognosis, diagnosis). The same methods could be applied to diverse investigations in medical oncology, which could yield novel research insights as well as improved companion diagnostics.

Our study has several limitations. A major limitation of this work lies on the Guardant cfDNA assay not detecting copy number loss, which includes key deletions in mCRPC such as *PTEN* loss, *RB1* loss, and loss of DNA damage repair genes. Additionally, the assay used in our study does not test for all genes of interest in metastatic prostate cancer, including *SPOP*, *ETS* family gene fusions, and *AR* splice variants. The absence of these genes from the panel may explain why we obtain high positive predictive (94%) and specificity (91%) values on these data, but lower sensitivity (64%) and negative predictive values (62%). The retrospective nature of our study means that unknown exposures could act as confounding variables; however, this is representative of real-world clinical settings. Although the strength of this study is inclusion of clinically annotated patient samples, the limitation is that patients with mCSPC and mCRPC were unmatched.

Conclusion

GAs, primarily amplifications in *AR*, RTK, MAPK, PI3K, and G1/S signaling pathways, are enriched in metastatic castration-resistant prostate cancer. Identification of these pathways may provide an avenue for developing novel therapeutic combination strategies for these men. More specifically, MEK inhibitors alone or in combination with AKT inhibitors warrant further investigation in patients with amplifications in the MAPK pathway. Finally, our machine learning framework for cancer genomics facilitates the discovery of mutational signatures associated with clinical phenotypes, provides insight into novel therapeutic strategies, has potential as a companion diagnostic for treatment selection, and is applicable to diverse investigations in clinical oncology.

Acknowledgments

Research reported in this publication was supported by the National Cancer Institute of the National Institutes of Health under Award Number P30CA042014 and to N. Agarwal (3P30CA042014-25S2). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We thank Dr. Gabor Marth for critically reviewing the manuscript and helping to improve its content. We acknowledge Richard B. Lanman's role toward interpretation of the data.

Author Contributions

Conception/design: Edwin Lin, Mark Yandell, Neeraj Agarwal

Provision of study material or patients: Benjamin L. Maughan, Deepika Sirohi, Guru Sonpavde, Umang Swami, Manish Kohli, Neeraj Agarwal, Sergiusz Wesolowski, Thereasa Rich, Oliver Sartor

Collection and/or assembly of data: Neeraj Agarwal, Andrew W. Hahn, Roberto H. Nussenzveig, Benjamin L. Maughan, Taylor R. McFarland, Nityam Rathi, Oliver Sartor, Guru Sonpavde, Umang Swami

Data analysis and interpretation: Edwin Lin, Andrew W. Hahn, Roberto H. Nussenzveig, Sergiusz Wesolowski, Benjamin L. Maughan, Taylor R. McFarland, Nityam Rathi, Oliver Sartor, Guru Sonpavde, Umang Swami, Thereasa Rich, Mark Yandell, Neeraj Agarwal

Manuscript writing: Edwin Lin, Andrew W. Hahn, Roberto H. Nussenzveig, Sergiusz Wesolowski, Benjamin L. Maughan, Taylor R. McFarland, Nityam Rathi, Oliver Sartor, Guru Sonpavde, Umang Swami, Thereasa Rich, Mark Yandell, Neeraj Agarwal, Manish Kohli

Final approval of manuscript: Edwin Lin, Andrew W. Hahn, Roberto H. Nussenzveig, Sergiusz Wesolowski, Nicolas Sayegh, Benjamin L. Maughan, Taylor Ryan McFarland, Nityam Rathi, Deepika Sirohi, Guru Sonpavde, Umang Swami, Manish Kohli, Thereasa Rich, Oliver Sartor, Mark Yandell, Neeraj Agarwal

Disclosures

Edwin Lin: Tempus (C/A); **Roberto H. Nussenzveig:** Tempus (C/A); **Benjamin L. Maughan:** AVEO Oncology, Janssen, Astellas, Bristol-Myers Squibb, Clovis, Tempus, Merck, Exelixis, Bayer Oncology, Peloton Therapeutics (SAB), Exelixis, Bavarian-Nordic, Clovis, Bristol-Myers Squibb (RF); **Guru Sonpavde:** Bristol-Myers Squibb, Genentech, EMD Serono, Merck, Sanofi, Seattle Genetics/Astellas, AstraZeneca, Exelixis, Janssen, Bicycle Therapeutics, Pfizer, Immunomedics/Gilead, Scholar Rock, G1 Therapeutics (SAB), Sanofi, AstraZeneca, Immunomedics/Gilead, QED, Predicine,

Bristol-Myers Squibb (RF—Inst), Bristol-Myers Squibb, Bavarian Nordic, Seattle Genetics, QED, G1 Therapeutics (Other—steering committee—unpaid), AstraZeneca, EMD Serono, Debiopharm (Other—steering committee—paid), Mereo (Other—safety monitoring committee), Bristol-Myers Squibb, AstraZeneca (Other—travel), Physicians Education Resource, OncLive, Research to Practice, Medscape (H); **Oliver Sartor:** Advanced Accelerator Applications, Astellas, AstraZeneca, Bayer, Blue Earth Diagnostics, Bavarian Nordic, Bristol-Myers Squibb, Clarity Pharmaceuticals, Clovis, Constellation, Dendreon, EMD Serono, Fusion, Isotopen Technologien Meunchen, Janssen, Myovant, Myriad, Noria Therapeutics, Inc., Novartis, Noxopharm, Progenics, POINT Biopharma, Pfizer, Sanofi, Tenebio, Telix, Theragnostics (C/A), Advanced Accelerator Applications, AstraZeneca, Bayer, Constellation, Dendreon, Endocyte, Invitae, Janssen, Merck, Progenics, Sanofi, SOTIO (RF); **Mark Yandell:** Fabric Genomics, IDbyDNA Inc. (C/A); **Neeraj Agarwal:** Pfizer, Exelixis, Merck, Argos, EMD Serono, Eisai, Bayer, Novartis, Genentech, Bristol-Myers Squibb, AstraZeneca, Medivation, Clovis, Foundation One, Astellas, Eli Lilly & Co., Nektar, Active Biotech, Bavarian Nordic, Calithera, Celldex, GlaxoSmithKline, Immunomedics, Janssen, Merck, New Link Genetics, Prometheus, Rexahn, Sanofi, Takeda, Tracon (C/A). The other authors indicated no financial relationships.

(C/A) Consulting/advisory relationship; (RF) Research funding; (E) Employment; (ET) Expert testimony; (H) Honoraria received; (OI) Ownership interests; (IP) Intellectual property rights/inventor/patent holder; (SAB) Scientific advisory board

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7–30.

2. Halabi S, Lin CY, Kelly WK et al. Updated prognostic model for predicting overall survival in first-line chemotherapy for patients with metastatic castration-resistant prostate cancer. J Clin Oncol 2014;32:671–677.

3. Kyriakopoulos, CE, Chen YH, Carducci MA et al. Chemohormonal therapy in metastatic hormone-sensitive prostate cancer: Long-Term survival analysis of the randomized phase III E3805 CHAARTED trial. J Clin Oncol 2018;36: 1080–1087.

4. Agarwal N, Di Lorenzo G, Sonpavde G et al. New agents for prostate cancer. Ann Oncol 2014; 25:1700–1709.

5. Francini E, Gray KP, Shaw GK et al. Impact of new systemic therapies on overall survival of patients with metastatic castration-resistant prostate cancer in a hospital-based registry. Prostate Cancer Prostatic Dis 2019;22:420–427.

6. Swami U, Agarwal N. Improvement in overall survival with Apalutamide, Darolutamide and Enzalutamide in patients with non-metastatic castration-resistant prostate cancer. Cancer Treat Res Commun 2020;25:100205.

7. de Bono JS, De Giorgi U, Rodrigues DN et al. Randomized phase II study evaluating Akt blockade with ipatasertib, in combination with abiraterone, in patients with metastatic prostate cancer with and without PTEN loss. Clin Cancer Res 2019;25:928–936.

8. Nakazawa M, Paller C, Kyprianou N. Mechanisms of therapeutic resistance in prostate cancer. Curr Oncol Rep 2017;19:13.

9. Bronkhorst AJ, Ungerer V, Holdenrieder S. The emerging role of cell-free DNA as a molecular marker for cancer management. Biomol Detect Quantif 2019;17:100087.

10. Libbrecht MW, Noble WS. Machine learning applications in genetics and genomics. Nat Rev Genet 2015;16:321–332.

11. Scher HI, Halabi S, Tannock I et al. Design and end points of clinical trials for patients with progressive prostate cancer and castrate levels of testosterone: recommendations of the Prostate Cancer Clinical Trials Working Group. J Clin Oncol 2008;26:1148–1159.

12. Lanman RB, Mortimer SA, Zill OA et al. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. PLoS One 2015;10:e0140712.

13. Odegaard JI, Vincent JJ, Mortimer S et al. Validation of a plasma-based comprehensive cancer genotyping assay utilizing orthogonal tissue- and plasma-based methodologies. Clin Cancer Res 2018;24:3539–3549.

14. Bamford S, Dawson E, Forbes S et al., The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. Br J Cancer 2004;91: 355–358.

15. Gao J, Aksoy BA, Dogrusoz U et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 2013;6:pl1.

16. UniProt Consortium. The universal protein resource (UniProt). Nucleic Acids Res 2008;36 (Database issue):D190–195.

17. Olivier M, Eeles R, Hollstein M et al. The IARC TP53 database: New online mutation analysis and recommendations to users. Hum Mutat 2002;19:607–614.

18. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res 2010;11:2079–2107.

19. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res 2019;47:D330–D338.

20. Ashburner M, Ball CA, Blake JA et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 2000;25: 25–29.

21. Feng Q, He B. Androgen receptor signaling in the development of castration-resistant prostate cancer. Front Oncol 2019;9:858.

22. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. Cell 2015;163:1011–1025.

23. Armenia J. Wankowicz SAM, Liu D et al. The long tail of oncogenic drivers in prostate cancer. Nat Genet 2018;50:645–651.

24. Hieronymus H, Schultz N, Gopalan A et al. Copy number alteration burden predicts prostate cancer relapse. Proc Natl Acad Sci USA 2014;111: 11139–11144.

25. Robinson D, Van Allen EM, Wu YM et al. Integrative clinical genomics of advanced prostate cancer. Cell 2015;161:1215–1228.

26. Nickols NG, Nazarian R, Zhao SG et al. MEK-ERK signaling is a therapeutic target in metastatic castration resistant prostate cancer. Prostate Cancer Prostatic Dis 2019;22:531–538.

27. Katz M, Amit I, Yarden Y. Regulation of MAPKs by growth factors and receptor tyrosine kinases. Biochim Biophys Acta 2007; 1773:1161–1176.

28. Gao H, Ouyang X, Banach-Petrosky WA et al. Combinatorial activities of Akt and B-Raf/Erk signaling in a mouse model of androgen-independent prostate cancer. Proc Natl Acad Sci USA 2006;103:14477–14482.

29. Crumbaker M, Khoja L, Joshua AM. AR Signaling and the PI3K Pathway in Prostate Cancer. Cancers (Basel) 2017;9:34.

30. Kinkade CW, Castillo-Martin M, Puzio-Kuter A et al. Targeting AKT/mTOR and ERK MAPK signaling inhibits hormone-refractory prostate cancer in a preclinical mouse model. J Clin Invest 2008;118:3051–3064.

31. Toren P, Kim S, Johnson F et al. Combined AKT and MEK pathway blockade in pre-clinical models of enzalutamide-resistant prostate cancer. PLoS One 2016;11:e0152861.

32. Grimaldi AM, Simeone E, Festino L et al. MEK Inhibitors in the treatment of metastatic melanoma and solid tumors. Am J Clin Dermatol 2017;18:745–754.

33. Massard C, Chi KN, Castellano D et al. Phase Ib dose-finding study of abiraterone acetate plus buparlisib (BKM120) or dactolisib (BEZ235) in patients with castration-resistant prostate cancer. Eur J Cancer 2017;76:36–44.

34. Zhang W, Liu HT. MAPK signal pathways in the regulation of cell proliferation in mammalian cells. Cell Res 2002;12:9–18.

35. Palmbos, PL, Daignault-Newton S, Tomlins SA et al. A randomized phase II study of androgen deprivation therapy with or without palbociclib in RB-positive metastatic hormone-sensitive prostate cancer. Clin Cancer Res 2021; 16:3017–3027.

36. Schettini F, De Santo I, Rea CG et al. CDK 4/6 inhibitors as single agent in advanced solid tumors. Front Oncol 2018;8:608.

37. Aparicio A, Den RB, Knudsen KE. Time to stratify? The retinoblastoma protein in castrate-resistant prostate cancer. Nat Rev Urol 2011;8: 562–568.

38. de Leeuw R, McNair C, Schiewer MJ et al. MAPK reliance via acquired CDK4/6 inhibitor resistance in cancer. Clin Cancer Res 2018;24:4201–4214.

39. Pandey K, An HJ, Kim SK et al. Molecular mechanisms of resistance to CDK4/6 inhibitors in breast cancer: A review. Int J Cancer 2019;145: 1179–1188.

40. Möröy T, Geisen C. Cyclin E. Int J Biochem Cell Biol 2004;36:1424–1439.

41. Blanchard DA, Mouhamad S, Auffredou MT et al. Cdk2 associates with MAP kinase in vivo and its nuclear translocation is dependent on

MAP kinase activation in IL-2-dependent Kit 225 T lymphocytes. Oncogene 2000;19:4184–4189.

42. Correze C, Blondeau JP, Pomerance M. p38 mitogen-activated protein kinase contributes to cell cycle regulation by cAMP in FRTL-5 thyroid cells. Eur J Endocrinol 2005;153:123–133.

43. Annala M, Fu S, Bacon JVW et al. Cabazitaxel versus abiraterone or enzalutamide in poor prognosis metastatic castration-resistant prostate cancer: A multicentre, randomised, open-label, phase II trial. Ann Oncol 2021;32:896–905.

44. Choudhury AD, Werner L, Francini E et al. Tumor fraction in cell-free DNA as a biomarker in prostate cancer. JCI Insight 2018;3:e122109.

See http://www.TheOncologist.com for supplemental material available online.