Prostate Cancer Patient Stratification by Molecular Signatures in the Veterans Precision Oncology Data Commons

Kyle M. Hernandez [1,2] *, Aarti Venkat [1,2] *, Danne C. Elbers [3,4] *, John R. Bihn [3], Mary T. Brophy [3,5], Nhan V. Do [3,5], Jennifer La [3,4], Qiong Liu [2], Andrew Prokhorenkov [2], Noah Metoki-Shlubsky [2], Feng-Chi Sung [3], Channing J. Paller [6] **, Nathanael R. Fillmore [3,4,7] **, and Robert L. Grossman [1,2,8] **+

* Co-first authors
** Co-last authors
+ Corresponding author

[1] Department of Medicine, University of Chicago, Chicago, IL, USA

[2] Center for Translational Data Science, University of Chicago, Chicago, IL, USA

[3] VA Cooperative Studies Program, VA Boston Healthcare System, MA, USA

[4] Harvard Medical School, Boston, MA, USA

[5] Boston University School of Medicine, Boston, MA, USA

[6] The Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA

[7] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA

[8] Open Commons Consortium, Chicago, IL, USA

Running title: Prostate cancer stratification by molecular signatures

Abstract

Veterans are at an increased risk for prostate cancer, a disease with extraordinary clinical and molecular heterogeneity, compared to the general population. Yet, little is known about the underlying molecular heterogeneity within the veteran population, and its impact on patient management and treatment. Using clinical and targeted tumor sequencing data from the national Veterans Affairs health system, we conducted a retrospective cohort study on 45 patients with advanced prostate cancer in the Veterans Precision Oncology Data Commons (VPODC), most of whom were metastatic castration resistant. We characterized the mutational burden in this cohort and conducted unsupervised clustering analysis to stratify patients by molecular alterations. Veterans with prostate cancer exhibited a mutational landscape broadly similar to prior studies, including KMT2A and NOTCH1 mutations associated with neuroendocrine prostate cancer phenotype, previously reported to be enriched in veterans. We also identified several potential novel mutations in *PTEN, MSH6, VHL, SMO* and *ABL1*. Hierarchical clustering analysis revealed two subgroups containing therapeutically targetable molecular features with novel mutational signatures distinct from those reported in the Catalogue of Somatic Mutations in Cancer database. The clustering

approach presented in this study can potentially be used to clinically stratify patients based on their distinct mutational profiles and identify actionable somatic mutations for precision oncology.

INTRODUCTION

Prostate cancer remains the second most common cause of cancer deaths among men in the United States (Siegel et al. 2019). For most men with prostate cancer, definitive local therapy is curative, but for 20-40% of patients the disease relapses (Paller et al. 2013) and progresses to metastatic castration resistant prostate cancer (mCRPC) with a median overall survival of 2-3 years (Armstrong et al. 2020; Tagawa et al. 2022). Since 2010 the therapeutic options for mCRPC have increased quickly, extending survival. New treatments include androgen receptor (AR) signaling inhibitors (enzalutamide, abiraterone, and darolutamide), taxanes (cabazitaxel), radioisotopes (Radium223 and PSMA Lu177) and immunotherapy (sipuleucel-T and pembrolizumab)(Teo et al. 2019). Several genetically targeted therapies have also been approved, including pembrolizumab for prostate cancer patients with microsatellite instability-high (MSI-H) or mismatch repair deficient (dMMR) cancer (Sokolova and Cheng 2020), and poly(adenosine diphosphate-ribose) polymerase (PARP) inhibitors olaparib, and rucaparib for patients with mutations in DNA-repair genes, especially *BRCA2* (Marshall et al. 2019). There are many other drugs in development to target new pathways, that provide a potential for a more precision-medicine oriented approach (Yamada and Beltran 2021; Frantzi et al. 2020; Sayegh et al. 2022).

Metastatic prostate cancer has extraordinary heterogeneity in terms of molecular alterations, and this remains a substantial barrier to optimizing patient care and precision oncology for this population (Haffner et al. 2021). Both molecular diversity and heterogeneity point to a major unmet need to understand the molecular biomarkers that can better inform treatment decisions, as well as guide future drug development. A previous study has explored the complex genomic landscape of mCRPC using targeted genomic panels and shown that such patients harbor clinically actionable somatic and germline molecular alterations in PIK3CA/B, RSPO, RAF, APC, β-catenin, ZBTB16, as well as in genes underlying DNA repair pathways (Dessel et al. 2019). Additionally, certain mutations, such as BRCA1, BRCA2, PALB2, CHEK2 and ATM, are more enriched in mCRPC than in primary localized prostate cancer samples, reflecting the unique biology of cancer progression and castration resistance. (Robinson et al. 2015; Henríquez et al. 2021)

Veteran Affairs (VA) patients are at increased risk for developing prostate cancer compared to the general population, with a prostate cancer incidence rate ratio of 2 compared to the non-military population (Zhu et al. 2009). The reason for this increased incidence is unclear, but factors include increased screening rates, exposure scenarios unique to the military such as depleted uranium and Agent Orange, and high rates of

STIs (Dennis et al. 2009). In this retrospective study, we use existing clinical data from the VA healthcare system available in VPODC combining clinical data with targeted sequencing (Elbers et al. 2020).

A previous study investigated the genomic landscape of metastatic solid tumors from a sample of the Veterans Health Administration population as part of the National Precision Oncology program, but their focus was not specific to prostate cancers and the sample size for advanced castration resistance was limited (Poonnen et al. 2019a). We show that the VA cohort displays clinically actionable mutations previously reported and potential novel targets. Moreover, we show that unsupervised clustering revealed two subpopulations with novel mutational signatures distinct from existing signatures in the COSMIC database (Tate et al. 2018).

RESULTS

Somatic targeted sequencing and clinical data from advanced prostate cancer patients available in the VPODC were analyzed using three capture kits. We investigated mutations in genes that were targeted by all capture kits (N=45) and excluded samples with evidence of contamination and poor coverage (N=1), resulting in a total of N=44 samples (Supplementary Table 1, Supplementary Figures 1 and 2). We only considered genes that were targeted by all three capture kits to exclude kit-specific clustering artifacts (Supplementary Figure 3). As described in more detail below, we classified the veterans cohort into two subgroups using hierarchical clustering, and identified two *de novo* SBS mutational signatures that did not exhibit strong similarity to existing COSMIC SBS signatures (Alexandrov et al. 2020) (Methods, Supplementary Figure 4).

Clinical Presentation of the VA Cohort

Table 1 summarizes clinical details, including age at biopsy, biopsy site of sequenced tissue, Eastern Cooperative Oncology Group performance scale and death status (Appendix). Samples were selected as part of the VISN 1 precision oncology program, the VA New England health care system, and prioritized for inclusion of patients in the New England region that had advanced disease (Elbers et al. 2020; Fiore et al. 2016) (Methods).

Most patients were identified as metastatic stage (N=41, 91%), and N=29 (64%) were castrate resistant (Table 1). The metastatic biopsies taken from prostate were confirmed by the imaging results either recorded directly or summarized in clinical notes provided by the oncologist. N=14 (31%) samples were biopsied from sites other than the prostate. The median pathology estimated tumor purity of this cohort was 70% (range, 25% to 90%) across all samples. A small subset of samples were of low tumor purity (<35%, N=5) (Supplementary Table 2). The targeted sequencing libraries were sequenced to a mean target coverage of 717.15X (Supplementary Table 3).

Somatic Mutations Landscape and TMPRSS2:ERG Gene Fusion in Prostate Cancer

Consistent with previous analysis on metastatic solid tumors in Veterans, we found somatic mutations in classic genes such as TP53, BRCA2, ATM, PALB2 and NOTCH1, with TP53 mutations being the most frequent (Figure 1) (Robinson et al. 2015; Henríquez et al. 2021; Poonnen et al. 2019b) . The TMPRSS2:ERG gene fusion, which is recognized as one of the most common gene fusions found in prostate cancer patients, was identified in 15 of 44 patients (34%) (Wang et al. 2017), slightly lower than a previously reported estimate of 42.6% in metastatic castrate resistant prostate cancers (Dessel et al. 2019). We confirmed the presence of the gene fusion through a visualization of the drastic decline in depth of coverage around the breakpoint (Supplementary Figure 4).

Mutation-driven Clustering and Survival Analyses

First, we used the VCF files to characterize unique mutational signatures in the cohort based on previously reported COSMIC mutational signatures (Manders et al. 2022)**.** The distribution of known COSMIC signatures varied within a sample, and no single signature dominated across samples (Supplementary Figure 5). Two *de novo* SBS signatures, defined as mutational signature 1 and mutational signature 2 were identified, reflecting a few unique mutational processes in the cohort (Methods, Supplementary Figure 6). A heatmap of pairwise cosine similarities between known COSMIC mutational signatures and mutational signature 1 and 2 show a moderate level of similarity with COSMIC signature SBS5, but no existing signature exceeded the *de novo* threshold of 0.85 (Methods, Supplementary Figure 7). SBS5 has been previously found to be ubiquitous among various cancer types, including prostate cancer (Alexandrov et al. 2020; Petljak et al. 2019)

Next, we performed unsupervised hierarchical clustering on the genomic features extracted from targeted sequencing data to classify the cohort into subgroups (Methods). Two distinct subgroups were identified: Cluster A (N=26) and Cluster B (N=18) (Figure 2A). A principal component analysis (PCA) showed the separation between these two subgroups using the first three components, with PC1 and PC2 explaining 20.98% and 15.99% of the total variance, respectively (Supplementary Figure 8).

We inspected the relative frequencies of SBS and counts of indel events and calculated a ratio of transitions (Ti) to transversions (Tv) from SBS frequencies (Figures 2B, 2C, 2D). Cluster A had a higher Ti/Tv ratio than Cluster B (x^2=4.32, df=1, P=0.04), suggesting the possibility of either difference in the DNA damage mechanism or in DNA

repair pathways between the two groups (Figure 2C). In Cluster B, we noticed an enrichment of mutational signature 2 than in Cluster A (x^2=19.1, df=1, P=1.3e-05) (Figure 2E). In addition to the patterns of unique somatic mutations of individuals in the two clusters, any differences in sites of biopsies between could also give rise to these differences. Specifically, there were 10 cases with biopsies from outside the prostate that belong to cluster A, while only 3 belong to cluster B (Supplementary Table 4). The contribution of non-metastatic tumors to the clustering pattern appears minimal, with one non-metastatic prostate biopsy found in Cluster A and Cluster B, each. The TMPRSS2-ERG fusion, a high frequency fusion gene in prostate cancers, was identified in 34% of the patients (Figure 2F).

To explore if there are any differences in survival based on the two clusters, we conducted a Kaplan-Meier survival analysis. Since any actionable changes in care could be taken after obtaining the sequencing results, we considered the time interval between obtaining sequencing results until death. This analysis suggested a slight difference in survival time, but the results were insignificant (Cluster A 200 days vs Cluster B 238 days, P value=0.64) (Figure 3A). An alternate clinical event that is relevant to advanced prostate cancers is the date of diagnosis of metastatic prostate cancer, but these dates are not available in VPODC. Instead, the date of initial diagnosis of prostate cancer is available, and reanalysis of survival data using these dates yielded insignificant results (data not shown).

Mutational Analyses of Clusters

While the low sample size of mutations and our cohort precludes a significant comparison of mutations between clusters A and B, we examined which mutations were shared or unique to a cluster in a purely descriptive analysis. We found that while most non-synonymous mutations were common between Cluster A and Cluster B and reflected well known oncogenes such as TP53, MET, PALB2 and BRCA2, some mutations were unique to one cluster (Figure 3B). Cluster B included PTEN, a common mutation in castration resistant prostate cancer, (Myint et al. 2021) and MSH6, a mismatch repair gene seen in 1% of prostate cancer cases. (Bratslavsky et al. 2021) It also included VHL which is a critical part of the androgen ubiquination pathway and therefore necessary for many drugs that target increased AR receptor ubiquination pathway. (Kregel et al. 2020) Other genes in Cluster B that can be targeted by agonists and inhibitors include FGFR2, JAK3 and ALK (Figure 3B, Table 2).

Cluster A, on the other hand, included a mutation in GNAS, a mutation which has previously been reported as unique to mCRPC, and mutually exclusive with AR mutations. (Robinson et al. 2015) GNAS has no clinically approved drugs or clinical trials, however, it does have several drugs that can potentially act as antagonists (Table 2). Additionally, Cluster A included a mutation in SMO, a critical gene in the hedgehog

signaling pathway with several potential inhibitors. Hedgehog signaling in the tumor microenvironment can be induced by androgen deprivation and can drive steroidogenesis from benign stromal cells within a prostate tumor (Lubik et al. 2017). An SMO antagonist was able to suppress the development of castration resistant prostate cancer in a xenograft model (Li et al. 2018).

DISCUSSION

Clinically, prostate cancer is an extraordinarily heterogenous disease. Some patients have slow growing local disease that only requires observation, while others have highly lethal disease even after radical treatments such as surgery, radiation, and castration. This heterogeneity is also apparent at the molecular level, with prostate cancer showing enormous diversity in terms of genetic architecture and intra-tumoral diversity. (Wei et al. 2017) While it has been demonstrated that genetic signatures predict clinical outcomes more accurately than to traditional factors such as tumor stage, PSA level, and Gleason score, (Network et al. 2015; Group et al. 2015) it remains a challenge to optimally stratify prostate cancer patients based on molecular features.

The first line therapy for metastatic prostate cancer is androgen deprivation therapy and often results in tumor regression. However, it also has the potential to induce castration-resistance through mechanisms ranging from AR amplification to increasing androgen receptor sensitivity from mutations, and activation of pathways that bypass the androgen receptor pathway. (Group et al. 2015) Understanding the molecular evolution and clinical trajectories of prostate cancer remains a challenge and an opportunity, as does connecting this information to clinical care.

Here we conduct a descriptive analysis of prostate cancer mutations in a small cohort of veterans for whom clinical and genomic data were available in VPODC. We find distinct mutational signatures with clinically actionable targets such as those for SMO, ABL1, MET, ALK, FGFR2 and JAK3 (Table 2). Additionally, we identified mutations in SMAD4/TGFβ (NCT02452008), ROS1, PTEN, EGFR and BRAF where therapies are currently being explored in the MATCH Trial (NCT02465060). We also identified mutations in GNAS that warrant further clinical exploration if confirmed in another larger cohort. While there was no statistically significant difference in survival between the subpopulations, the small numbers of patients limit any conclusions about the clinical differences in outcomes. Additionally, sequencing tumor-normal matched pairs will allow for a reliable analysis of MSI, homologous recombination deficiency, tumor mutational burden and copy number variants, features that were not evaluated in this study. The clustering approach used shows that the different mutational profiles have potential to clinically stratify patients and guide treatment options.

While not necessarily representative of the entire VA population, our analysis of 45 patients yielded insights consistent with the previous study on prostate cancers in the VA population. We found coding mutations in genes associated with neuroendocrine

prostate cancers that were shared by both Clusters A and B, such as KMT2A and NOTCH1, but not in PIK3CA or KRAS (Figure 3B) (Poonnen et al. 2019b). This may be due to the unique biology and diversity of advanced metastatic prostate cancers and castration status in veterans but more follow up work is needed to test this hypothesis given the small cohort size and rarity of mutations in our dataset. The VA cohort has several advantages because it represents a population with diverse ancestry that is offered routine screening and equal access to care. A significant challenge of previous studies has been accessing populations with sufficient diversity to tease out the different contributions of varied ancestry and access to care (Gong et al. 2022). It remains to be seen if demographic or environmental exposures account for the increased risk of prostate cancer amongst a military population above and beyond what would be expected from increased screening. Thus, it is possible that tumors in military veterans may have a slightly different evolutionary history compared to a broader population, giving rise to their unique mutational profile.

METHODS

Data Sources and Population

This retrospective cohort study was conducted using data from the Veterans Affairs Precision Oncology Data Repository (VA-PODR) available in the VPODC (Elbers et al. 2020). The VA-PODR includes information from four sources: electronic health record data from the Corporate Data Warehouse (CDW), manually curated data on cancer cases from the VA Cancer Registry System (VACRS), imaging data, and targeted tumor sequencing information. The study's cohort includes 45 men with prostate cancer for whom targeted tumor sequencing data is available in the VPODC. Data in the VPODC is collected as secondary use data, obtained as part of routine clinical care. Some missing values in VPODC are the result of not all data being properly recorded in the Electronic health record (EHR) or Cancer Data Warehouse and Cancer Registry. Genomic data utilized was sequenced as part of clinical care and is made available through consent under the Precision Oncology Data Repository (PODR) (Elbers et al. 2020). All tissues sequenced were Formalin-Fixed Paraffin-Embedded (FFPE) tissue specimens. Targeted tumor sequencing was provided through PGDx. Tumor sequencing capture kits were updated as needed during routine clinical care and several versions used included PGDx CancerSelect 125 (N=34), PGDx CancerSELECT 203 (N=6), and PGDx CancerSELECT 88 (N=5).

Sequence Processing and Quality Control

Raw sequencing reads were processed following standard somatic mutation workflows (Zhang et al. 2021). Briefly, reads were aligned to GRCh38 reference available from the Genomic Data Commons (GDC; https://api.gdc.cancer.gov/data/254f697d-310d-4d7d-a27b-27fbf767a834) with BWA 0.7.15-r1142-dirty (Li 2013), PCR duplicates were detected with Picard 2.18.11-SNAPSHOT (http://broadinstitute.github.io/picard/), and

base recalibration was applied by GATK v4.0.7.0 (Auwera and O'Connor 2020). The processed alignments were further evaluated for somatic mutations by GATK MuTect2 v4.1.2 (Auwera and O'Connor 2020) in tumor-only mode utilizing a panel of normals (GDC UUID: 6c4c4a48-3589-4fc0-b1fd-ce56e88c06e4) (Zhang et al. 2021). Variants remaining after applying filters from MuTect2 were annotated by Variant Effect Predictor (VEP) v84 (McLaren et al. 2016) using the GDC VEP cache (https://api.gdc.cancer.gov/data/8b9278b3-1e0c-430a-aae5-a944428401c0). Picard v2.22.4 was used to evaluate capture kit-specific coverage to exclude any outliers based on coverage using Hsmetrics and variants located outside of the kit-specific targeted region were removed. Targeted gene lists from each capture kit were intersected to generate a list of common genes. This list was used to include only on-target variants that are present in one of the common genes among capture kits. Cross sample contamination was assessed with GATK, using a 10% threshold. One sample exceeded this threshold and was excluded from subsequent analysis.

Structural Variant Detection and Annotation

Manta structural variant caller v1.6.0 was used to detect somatic structural variants (SVs) and insertion deletion mutations(indels) from mapped paired-end sequencing reads (Chen et al. 2016). The predicted SVs were then filtered to reduce false positives using the following criteria: (1) spanning paired-end reads >= 10 or split reads >= 10; and (2) both ends of the inspected SV located within the capture kit intervals. The filtered SVs were then used to detect potential gene fusions using the R package StructuralVariantAnnotation v3.15, (Cameron et al. 2022) with a specific focus on the TMPRSS2:ERG gene fusion. The identified TMPRSS2:ERG gene fusions were then manually investigated with the Arriba v2.3.0 draw fusion.R script (Uhrig et al. 2021).

Genomic Features Extraction

The variant call format (VCF) files of all samples across 3 capture kits were consolidated into a single Mutation Annotation Format (MAF) file using a custom script (https://github.com/uc-cdis/vpodc-prostate-cancer-pub). We then further filtered out the variants with tumor read depth < 20. Using VCF files, we computed four genomic features of interest: (1) frequency of six single base substitution (SBS) (C>A, C>G, C>T, T>A, T>C, T>G), (2) small indel frequencies, (3) *de novo* mutational signature contributions, defined as those with less than 85% similarity with previously reported 96 COSMIC SBS and (4) the presence/absence of the TMPRSS2:ERG gene fusion. The mutational signatures were generated using the R packages MutationalPatterns v3.4.1 (Blokzijl et al. 2018) and NMF v0.24.0 (Gaujoux and Seoighe 2010). Two *de novo* SBS mutational signatures were detected, and the absolute contribution of each signature was subsequently extracted for each sample.

Unsupervised Hierarchical Clustering

Unsupervised clustering of samples was performed based on the extracted genomic features using Euclidean distance and Ward's minimum variance method in a

hierarchical cluster analysis. (Murtagh and Legendre 2014) Cluster assignments were determined by cutting the dendogram using a Euclidean distance of 15 as a threshold and used for further analysis.

Matching Genes to Therapy

To investigate therapeutically targetable genes in each cluster we used the drug gene interaction database (https://www.dgidb.org/). (Freshour et al. 2020)

Survival Analysis

We explored survival from the time when sequencing results were received to the date of death or, for right censored patients (N=2), the date of last follow-up. Patients with negative time to the event values (N=2) were removed from the analysis. Kaplan-Meier (KM) survival analysis was performed using the R package survival v3.4.0 (https://CRAN.R-project.org/package=survival) and survival plots were generated using the survminer R package v0.4.9 (https://CRAN.R-project.org/package=survminer).

ADDITIONAL INFORMATION

Data Deposition and Access

Data analyzed in this study is available at https://vpodc.data-commons.org. There are restrictions on the availability of data due to security and privacy considerations. Please refer to the previous VPODC publication for data access guidelines (Elbers et al. 2020)

Ethics Statement

Additional written consent was not obtained as all data used in this study was retrieved under Precision Oncology Data Repository agreements and its IRB approval(s), as described in Elbers et al., 2020 (section: 'Regulatory Considerations'). For this study only de-identified data, housed at the Veterans Precision Oncology Data Commons per Data Use Agreement with the Precision Oncology Data Repository, was used.

Acknowledgements

This work was supported in part by the Open Commons Consortium and the Center for Translational Data Science, University of Chicago. The views expressed are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the U.S. government.

Author Contributions

Study concept and design: Kyle M. Hernandez, Aarti Venkat, Danne Elbers, Mary Brophy, Nhan Do, Channing Paller, Nate Fillmore and Robert L. Grossman

Data collection: Danne Elbers, John Bihn, Mary Brophy, Nhan Do, Jennifer La, Andrew Prokhorenkov, Feng-Chi Sung, and Nate Fillmore

Analysis and interpretation of results: Kyle M Hernandez, Aarti Venkat, Channing Paller, and Qiong Liu

Draft manuscript preparation: Kyle M. Hernandez, Aarti Venkat, Danne Elbers, John Bihn, Qiong Liu, Channing Paller, Nate Fillmore, and Robert L. Grossman

System development, operation and management of VPODC: Kyle M. Hernandez, Noah Metoki-Shlubsky and Robert L. Grossman

Funding

This work was supported in part by the VA Office of Research and Development, Cooperative Studies Program (#CSP2010) and by NIH/NCI P30 CA006973 (W81XWH-22-2-0024).

REFERENCES

Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Ng AWT, Wu Y, Boot A, Covington KR, Gordenin DA, Bergstrom EN, et al. 2020. The repertoire of mutational signatures in human cancer. *Nature* **578**: 94–101.

Armstrong AJ, Lin P, Tombal B, Saad F, Higano CS, Joshua AM, Parli T, Rosbrook B, Os S van, Beer TM. 2020. Five-year Survival Prediction and Safety Outcomes with Enzalutamide in Men with Chemotherapy-naïve Metastatic Castration-resistant Prostate Cancer from the PREVAIL Trial. *Eur Urol* **78**: 347–357.

Auwera GAV der, O'Connor BD. 2020. *Genomics in the cloud: using Docker, GATK, and WDL in Terra*. O'Reilly Media.

Blokzijl F, Janssen R, Boxtel R van, Cuppen E. 2018. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* **10**: 33.

Bratslavsky G, Decker B, Jacob JM, Necchi A, Spiess PE, Grivas P, Lin DI, Ramkissoon SH, Severson EA, Huang RSP, et al. 2021. Genomic landscape of MSH6-mutated clinically advanced castrate-resistant prostate cancer (mCRPC). *J Clin Oncol* **39**: 5062–5062.

Cameron DL, Dong R, Papenfuss AT. 2022. StructuralVariantAnnotation: a R/Bioconductor foundation for a caller-agnostic structural variant software ecosystem. *Bioinformatics* **38**: 2046–2048.

Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, Cox AJ, Kruglyak S, Saunders CT. 2016. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics* **32**: 1220–1222.

Dennis LK, Coughlin JA, McKinnon BC, Wells TS, Gaydos CA, Hamsikova E, Gray GC. 2009. Sexually Transmitted Infections and Prostate Cancer among Men in the U.S. Military. *Cancer Epidemiology Prev Biomarkers* **18**: 2665–2671.

Dessel LF van, Riet J van, Smits M, Zhu Y, Hamberg P, Heijden MS van der, Bergman AM, Oort IM van, Wit R de, Voest EE, et al. 2019. The genomic landscape of metastatic castration-resistant prostate cancers reveals multiple distinct genotypes with potential clinical impact. *Nat Commun* **10**: 5251.

Elbers DC, Fillmore NR, Sung F-C, Ganas SS, Prokhorenkov A, Meyer C, Hall RB, Ajjarapu SJ, Chen DC, Meng F, et al. 2020. The Veterans Affairs Precision Oncology Data Repository, a Clinical, Genomic, and Imaging Research Database. *Patterns* **1**: 100083.

Fiore LD, Brophy MT, Turek S, Kudesia V, Ramnath N, Shannon C, Ferguson R, Pyarajan S, Fiore MA, Hornberger J, et al. 2016. The VA Point-of-Care Precision Oncology Program: Balancing Access with Rapid Learning in Molecular Cancer Medicine. *Biomark Cancer* **8**: BIC.S37548.

Frantzi M, Hupe MC, Merseburger AS, Schanstra JP, Mischak H, Latosinska A. 2020. Omics Derived Biomarkers and Novel Drug Targets for Improved Intervention in Advanced Prostate Cancer. *Diagnostics* **10**: 658.

Freshour SL, Kiwala S, Cotto KC, Coffman AC, McMichael JF, Song JJ, Griffith M, Griffith OL, Wagner AH. 2020. Integration of the Drug–Gene Interaction Database (DGIdb 4.0) with open crowdsource efforts. *Nucleic Acids Res* **49**: D1144–D1151.

Gaujoux R, Seoighe C. 2010. A flexible R package for nonnegative matrix factorization. *Bmc Bioinformatics* **11**: 367–367.

Gong T, Jaratlerdsiri W, Jiang J, Willet C, Chew T, Patrick SM, Lyons RJ, Haynes A-M, Pasqualim G, Brum IS, et al. 2022. Genome-wide interrogation of structural variation reveals novel African-specific prostate cancer oncogenic drivers. *Genome Med* **14**: 100.

Group IPU, Gundem G, Loo PV, Kremeyer B, Alexandrov LB, Tubio JMC, Papaemmanuil E, Brewer DS, Kallio HML, Högnäs G, et al. 2015. The Evolutionary History of Lethal Metastatic Prostate Cancer. *Nature* **520**: 353–357.

Haffner MC, Zwart W, Roudier MP, True LD, Nelson WG, Epstein JI, Marzo AMD, Nelson PS, Yegnasubramanian S. 2021. Genomic and phenotypic heterogeneity in prostate cancer. *Nat Rev Urol* **18**: 79–92.

Henríquez I, Roach M, Morgan TM, Bossi A, Gómez JA, Abuchaibe O, Couñago F. 2021. Current and Emerging Therapies for Metastatic Castration-Resistant Prostate Cancer (mCRPC). *Biomed* **9**: 1247.

Kregel S, Wang C, Han X, Xiao L, Fernandez-Salas E, Bawa P, McCollum BL, Wilder-Romans K, Apel IJ, Cao X, et al. 2020. Androgen receptor degraders overcome common resistance mechanisms developed during prostate cancer treatment. *Neoplasia New York N Y* **22**: 111–119.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *Arxiv*.

Li N, Truong S, Nouri M, Moore J, Nakouzi NA, Lubik AA, Buttyan R. 2018. Non-canonical activation of hedgehog in prostate cancer cells mediated by the interaction of transcriptionally active androgen receptor proteins with Gli3. *Oncogene* **37**: 2313–2325.

Lubik AA, Nouri M, Truong S, Ghaffari M, Adomat HH, Corey E, Cox ME, Li N, Guns ES, Yenki P, et al. 2017. Paracrine sonic hedgehog signaling contributes significantly to acquired steroidogenesis in the prostate tumor microenvironment. *Int J Cancer* **140**: 358–369.

Manders F, Brandsma AM, Kanter J de, Verheul M, Oka R, Roosmalen MJ van, Roest B van der, Hoeck A van, Cuppen E, Boxtel R van. 2022. MutationalPatterns: the one stop shop for the analysis of mutational processes. *BMC Genom* **23**: 134.

Marshall CH, Sokolova AO, McNatty AL, Cheng HH, Eisenberger MA, Bryce AH, Schweizer MT, Antonarakis ES. 2019. Differential Response to Olaparib Treatment Among Men with Metastatic

Castration-resistant Prostate Cancer Harboring BRCA1 or BRCA2 Versus ATM Mutations. *Eur Urol* **76**: 452–458.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol* **17**: 122.

Murtagh F, Legendre P. 2014. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J Classif* **31**: 274–295.

Myint ZW, Allison DB, Ellis CS. 2021. A Case Report of Metastatic Castration-Resistant Prostate Cancer Harboring a PTEN Loss. *Frontiers Oncol* **11**: 731002.

Network TCGAR, Abeshouse A, Ahn J, Akbani R, Ally A, Amin S, Andry CD, Annala M, Aprikian A, Armenia J, et al. 2015. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**: 1011–1025.

Paller CJ, Antonarakis ES, Eisenberger MA, Carducci MA. 2013. Management of Patients with Biochemical Recurrence After Local Therapy for Prostate Cancer. *Hematology Oncol Clin North Am* **27**: 1205–1219.

Petljak M, Alexandrov LB, Brammeld JS, Price S, Wedge DC, Grossmann S, Dawson KJ, Ju YS, Iorio F, Tubio JMC, et al. 2019. Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis. *Cell* **176**: 1282-1294.e20.

Poonnen P, Duffy J, Hintze BJ, Shukla M, Brettin TS, Conrad NR, Yoo H, Guertin CM, Looney JA, Vashistha V, et al. 2019a. Genomic analysis of metastatic solid tumors in veterans: Findings from the VHA National Precision Oncology Program. *J Clin Oncol* **37**: 3074–3074.

Poonnen PJ, Duffy JE, Hintze B, Shukla M, Brettin TS, Conrad NR, Yoo H, Guertin C, Looney JA, Vashistha V, et al. 2019b. Genomic Analysis of Metastatic Solid Tumors in Veterans: Findings From the VHA National Precision Oncology Program. *Jco Precis Oncol* **3**: 1–13.

Robinson D, Van Allen EM, Wu Y-M, Schultz N, Lonigro RJ, Mosquera J-M, Montgomery B, Taplin M-E, Pritchard CC, Attard G, et al. 2015. Integrative Clinical Genomics of Advanced Prostate Cancer. *Cell* **162**: 454.

Sayegh N, Swami U, Agarwal N. 2022. Recent Advances in the Management of Metastatic Prostate Cancer. *Jco Oncol Pract* **18**: 45–55.

Siegel RL, Miller KD, Jemal A. 2019. Cancer statistics, 2019. *Ca Cancer J Clin* **69**: 7–34.

Sokolova AO, Cheng HH. 2020. Genetic Testing in Prostate Cancer. *Curr Oncol Rep* **22**: 5.

Tagawa ST, Ramaswamy K, Huang A, Mardekian J, Schultz NM, Wang L, Sandin R, Lechpammer S, George DJ. 2022. Correction: Survival outcomes in patients with chemotherapy-naive metastatic castration-resistant prostate cancer treated with enzalutamide or abiraterone acetate. *Prostate Cancer P D* **25**: 804–804.

Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, Creatore C, Dawson E, et al. 2018. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**: gky1015-.

Teo MY, Rathkopf DE, Kantoff P. 2019. Treatment of Advanced Prostate Cancer. *Annu Rev Med* **70**: 479–499.

Uhrig S, Ellermann J, Walther T, Burkhardt P, Fröhlich M, Hutter B, Toprak UH, Neumann O, Stenzinger A, Scholl C, et al. 2021. Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Res* **31**: gr.257246.119.

Wang Z, Wang Y, Zhang J, Hu Q, Zhi F, Zhang S, Mao D, Zhang Y, Liang H. 2017. Significance of the TMPRSS2:ERG gene fusion in prostate cancer. *Mol Med Rep* **16**: 5450–5458.

Wei L, Wang J, Lampert E, Schlanger S, DePriest AD, Hu Q, Gomez EC, Murakam M, Glenn ST, Conroy J, et al. 2017. Intratumoral and Intertumoral Genomic Heterogeneity of Multifocal Localized Prostate Cancer Impacts Molecular Classifications and Genomic Prognosticators. *Eur Urol* **71**: 183–192.

Yamada Y, Beltran H. 2021. The treatment landscape of metastatic prostate cancer. *Cancer Lett* **519**: 20–29.

Zhang Z, Hernandez K, Savage J, Li S, Miller D, Agrawal S, Ortuno F, Staudt LM, Heath A, Grossman RL. 2021. Uniform genomic data analysis in the NCI Genomic Data Commons. *Nat Commun* **12**: 1226.

Zhu K, Devesa SS, Wu H, Zahm SH, Jatoi I, Anderson WF, Peoples GE, Maxwell LG, Granger E, Potter JF, et al. 2009. Cancer Incidence in the U.S. Military Population: Comparison with Rates from the SEER Program. *Cancer Epidemiology Prev Biomarkers* **18**: 1740–1745.

FIGURE LEGENDS

Figure 1. Top 20 mutated genes in the VA advanced prostate cancer cohort (N=33). N=11 samples lacked mutations in these genes and were excluded from the plot. Patients were ordered based on the frequency of the observed top 20 mutated genes. ECOG, Eastern Cooperative Oncology Group.

Figure 2. (A) Unsupervised clustering revealed 2 main Clusters: Cluster A and Cluster B (B) Relative frequency of SBS changes (C) transition (Ti) to transversion (Tv) ratio (D) InDel counts (E) Absolute contribution of *de novo* mutational signatures 1 and 2 (F) the presence/absence of TMPRSS2:ERG gene fusion

Figure 3. (A). The time from sequencing results to death (or censoring) was compared using Kaplan-Meier curves. Number at risk table shows patients who have not yet experienced the death event or censored at certain time point, (B) Venn diagram shows shared and unique genes in clusters A and B

Table 1

Characteristics		VPODC Advanced Prostate Cancer Cohort
Count		45
Demographics		
	Age at Results	68.4 (65.6-74.1)
	Race	
	White	36
	Black or African American	6
	Declined to Answer	2
	Unknown	1
	Ethnicity	
	Hispanic or Latino	1
	Not Hispanic or Latino	40
	Declined to Answer	2
	Unknown	2
Metastatic Status	Yes	41
	No	3
	Unknown	1
Castrate Resistant	Yes	29
	No	4
	Unknown	12
ECOG	0-1	29
	2-4	12
	Unknown	4
Biopsy Site	Prostate	31
	Lymph Node	5
	Bone	1
	Lung	1
	Soft Tissue	1
	Bladder	1
	Gastrointestinal Tract	1
	Liver	2
	Brain	1
	Unknown	1
Tumor Purity (%)	Min	25.0
	Mean	62.4
	Median	70.0
	Max	90.0
	Unknown(n)	7
Deceased Status	Yes	43
	No	2

Table 2

Gene	Action	Drug(s)
		Vismogedib
		Pategedib
SMO	antagonist/inhibitors	Taladegib
GNAS	unknown	Dobutamine
		Bostunib
		Nilotinib
ABL1	inhibitors	Ponatinib
MET	inhibitors	Tepotinib
ALK	inhibitors	Brigatinib
FGFR2	agonist	Palifermin
JAK3	inhibitors	Decernotinib

TP53
ATM
NOTCH1
FANCA
BRCA2
GNAS
IDH1
MET
PALB2
ROS1
BRCA1
CTNNB1
ERBB4
JAK2
KMT2A
PMS2
RB1
SMAD4
TSC2
ABL1

Capture Kit
Biopsy Site
ECOG
Metastatic
Castrate Resistant

0
5
10
Sample_Count

Variant Classification
- Missense_Mutation
- Frame_Shift_Ins
- Splice_Site
- Nonsense_Mutation
- Splice_Region
- In_Frame_Del
- Frame_Shift_Del

Capture Kit
- CancerSelect 125
- CancerSelect 203
- CancerSelect 88

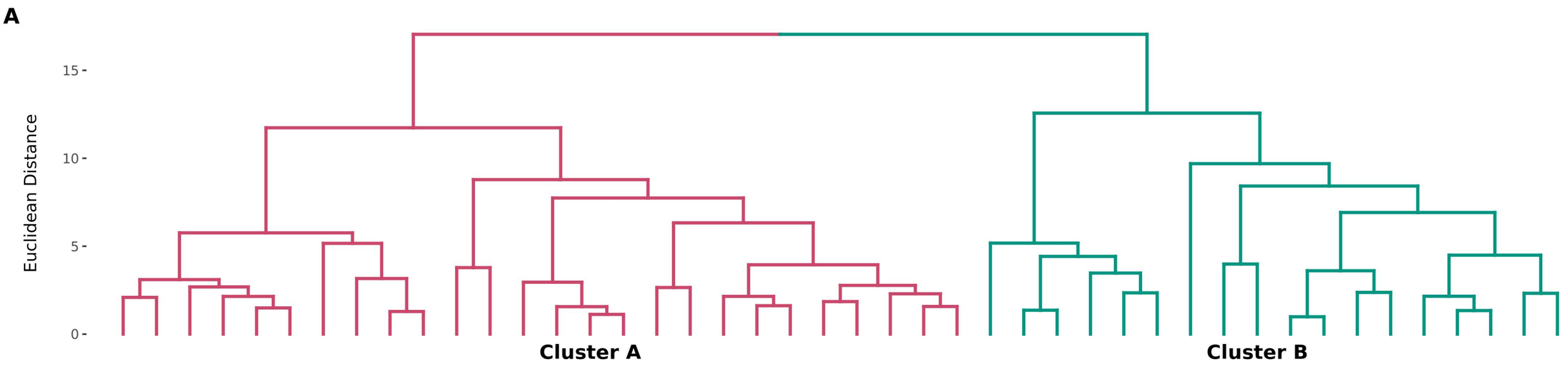
Biopsy Site
- Bladder
- GI Tract
- Liver
- Lung
- Lymph Node
- Prostate
- Soft Tissue
- Unknown

ECOG
- 3
- 2
- 1
- 0

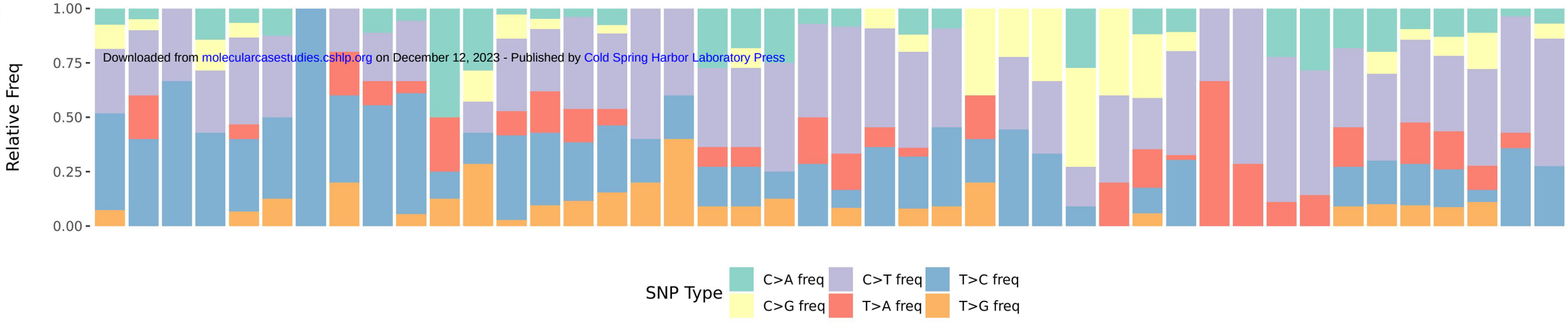
Metastatic
- No
- Yes

Castrate Resistant
- No
- Yes

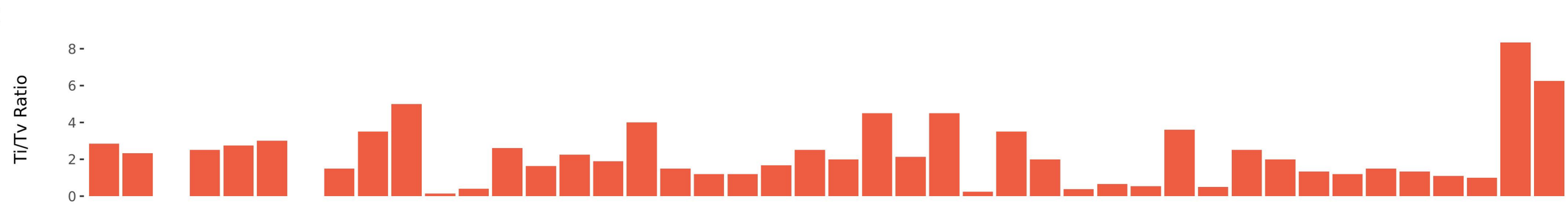
A



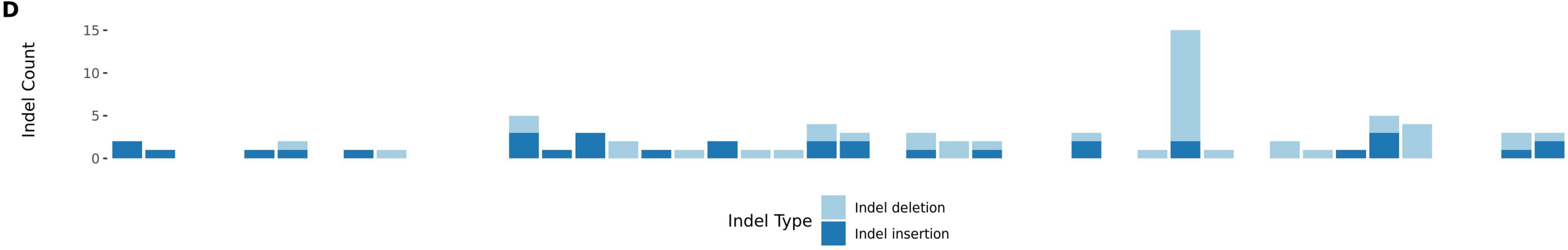
B



C



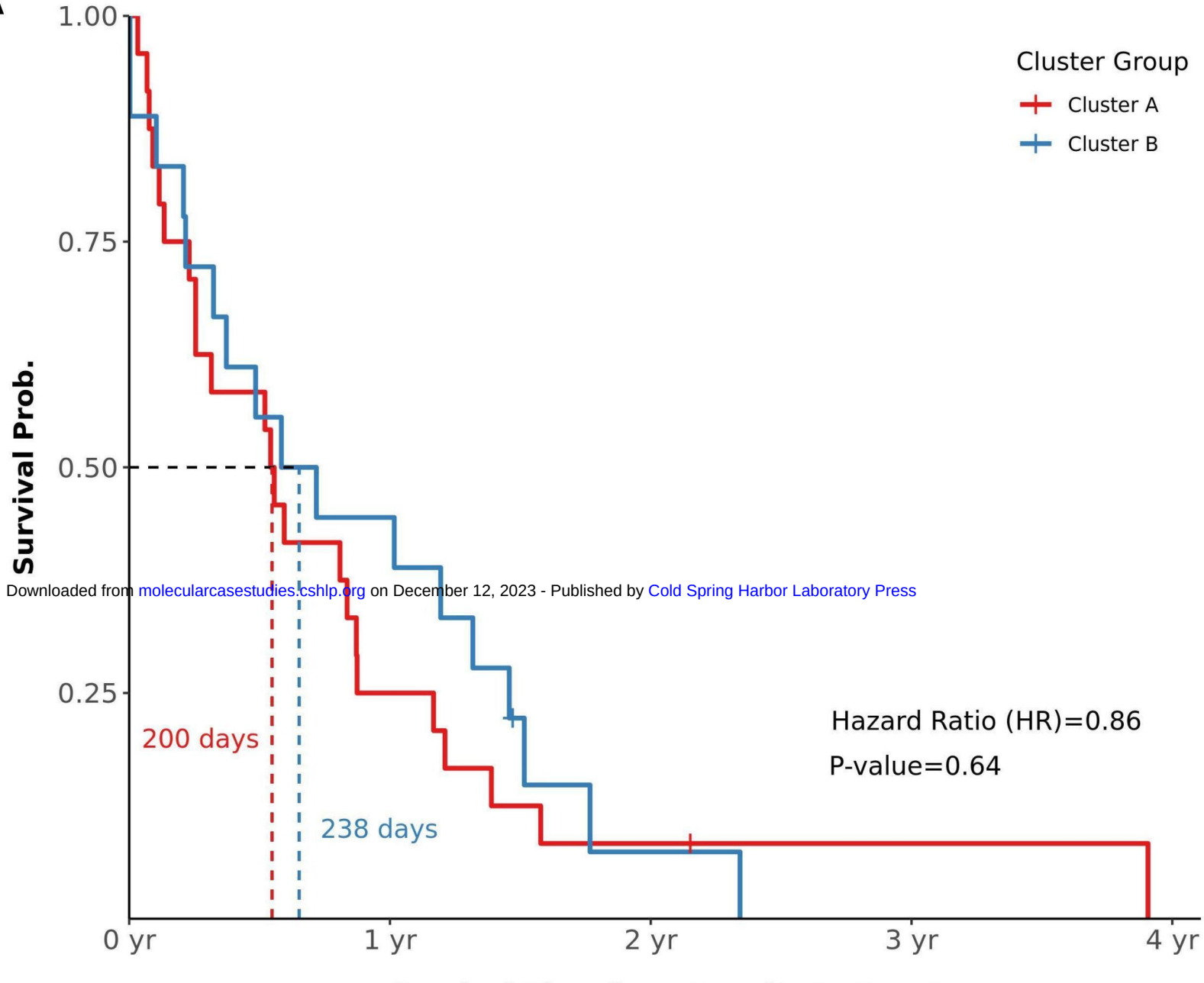
D

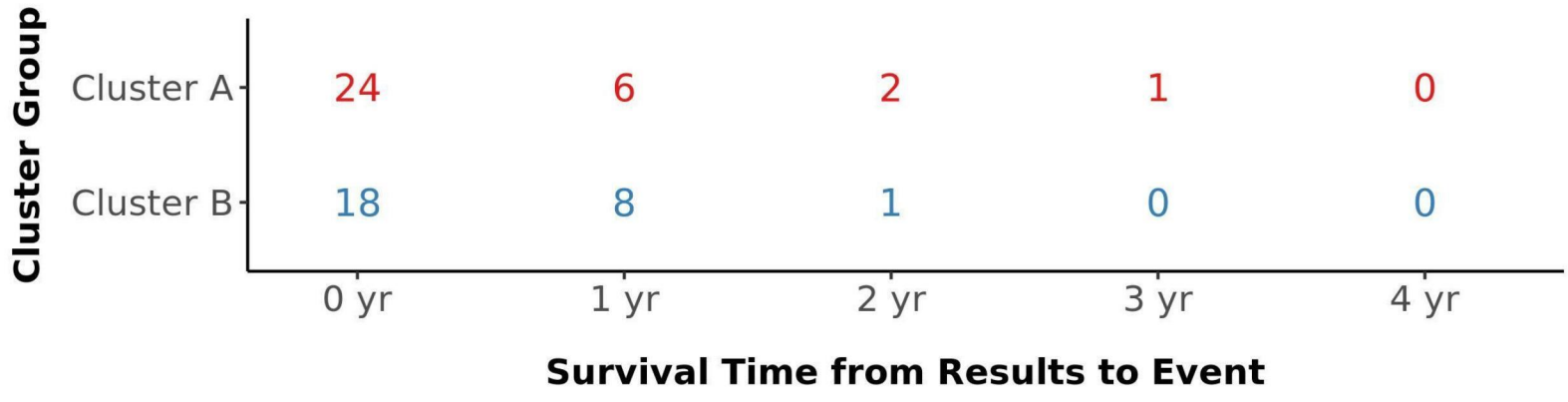


E

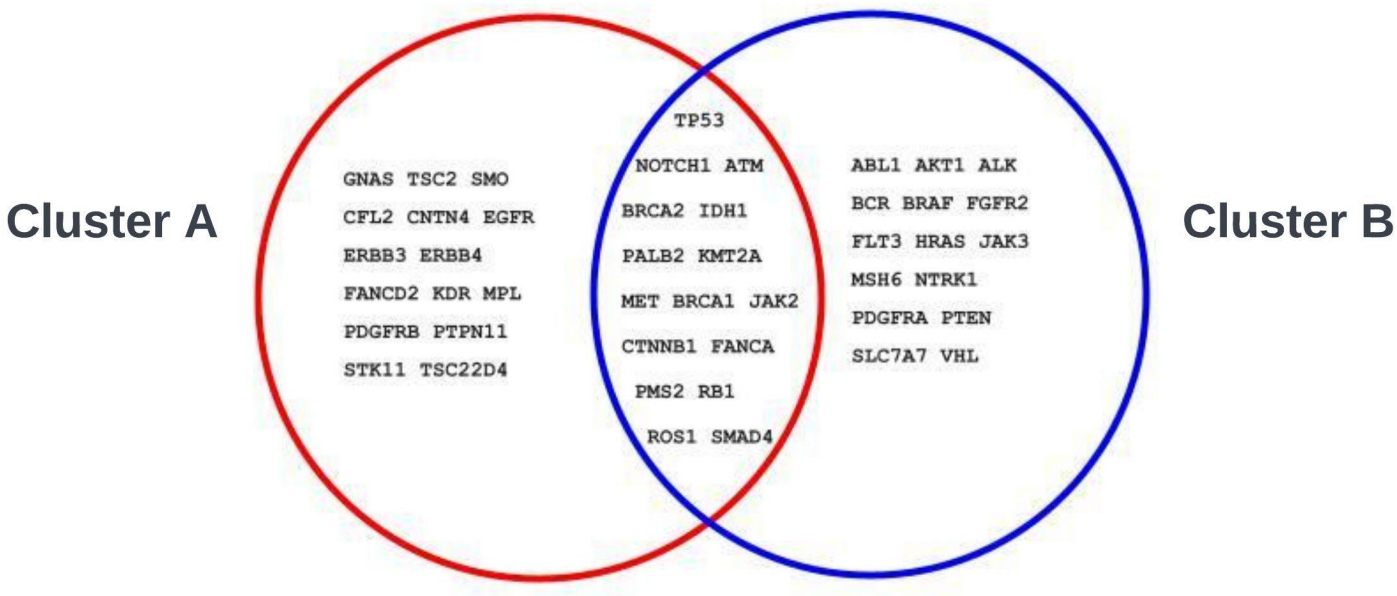
F

A





B



Prostate Cancer Patient Stratification by Molecular Signatures in the Veterans Precision Oncology Data Commons

Kyle M. Hernandez, Aarti Venkat, Danne C Elbers, et al.

Cold Spring Harb Mol Case Stud published online November 22, 2023
Access the most recent version at doi:10.1101/mcs.a006298