AWARD NUMBER: W81XWH-20-1-0644

TITLE: Driver Gene Networks of Genomic Instability in Prostate Cancer Progression

PRINCIPAL INVESTIGATOR: Sungyong You PhD

CONTRACTING ORGANIZATION: Cedars-Sinai Medical Center, Los Angeles, CA

REPORT DATE: September 2021

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
September 2021	Annual	15Aug2020 – 14 Aug2021

4. TITLE AND SUBTITLE

Driver Gene Networks of Genomic Instability in Prostate Cancer Progression

5a. CONTRACT NUMBER
W81XWH-20-1-0644

5b. GRANT NUMBER
GRANT12937731

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Sungyong You

E-Mail: Sungyong.You@cshs.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

CEDARS-SINAI MEDICAL CENTER
8700 BEVERLY BLVD
LOS ANGELES CA 90048-1804

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Development Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The less stable the cancer gene is, the faster cancer worsens. This can be used as an important measure of cancer's characteristics in various human cancers. By comparing and analyzing tumors at the genetic or gene expression level that is highly unstable in cancerous cells, we can identify changes in factors that are important to the various genetic manifestations that result in increased genomic instability (GI). This talk will describe the GI driver networks that can distinguish disease subtypes. Progress was made in four key points in the first year of the funding period. I have 1) developed a novel method for GI score computation to identify PCs with highly altered genome, 2) identified the genes (PCGI-E) significantly correlated with GI events, 3) Identified MRs that are relevant to PCGI-E gene signature regulation, and 4) developed PCGI TRN classifier that can identify the PCs with poor survival outcomes.

15. SUBJECT TERMS

Genomic Instability, Driver Network, Prostate Cancer, Transcriptome, Circulating Tumor Cell

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	21	19b. TELEPHONE NUMBER *(include area code)*

TABLE OF CONTENTS

1. INTRODUCTION

Prostate cancer (PC) is a major leading cause of death in men in US. Genomic rearrangement, copy number changes, and increased mutational burden are well characterized features that correlate highly with PC progression. However, drivers of genomic instability (GI) and their relationships in PC progression are elusive. Specific oncogene mutations such as SPOP and TP53 mutations can alter gene expression and promoting other genomic abnormalities. Recent studies suggest that tumors harboring genomic alterations are significantly associated with gene expression phenotypes linked with PC subtypes and loss/gain of function mutations of genetic drivers have been associated with a characteristic transcriptional phenotype.

This project is testing the hypotheses that *transcriptional programs governing the disease progression are distinct between PCGI high and low tumors and that the activity of GI driver networks allows us to make clinical predictions about prognosis and likely treatment sensitivity.* To this end we identified and assessed the GI driver network concept through systems biology approach to available PC genomic and transcriptomic profiles from multiple cohorts. During the second year of the funding period, we will identify common and distinct transcriptional programs between SPOP and TP53 driven drivers' networks.

Our specific aims have **not** modified from those stated in the original application.

Specific Aim 1. Determine whether the transcriptional programs are distinct between PCGI high and low tumors.
Specific Aim 2. Determine common and distinct transcriptional programs between the two mutually exclusive GI driver networks.
Specific Aim 3. Test whether the GI driver networks evolve in pre/post treatment using PC transcriptome.

2. KEYWORDS

Genomic Instability, Driver Network, Prostate Cancer, Transcriptome, Circulating Tumor Cell

3. ACCOMPLISHMENTS

What were the major goals of the project?
Research Goal 1: Identify transcriptional signature (PCGI-E) describing the specific transcriptional programs in each PCGI high and low tumors.
Milestones:
1) PCGI scores for all the tumor samples.
2) The most significant genes as a PCGI-E signature in PC genomic alterations.
Target months: 9
Percentage of completion: 100%

Research Goal 2: Develop the PCGI TRN classifier to predict metastatic progression and assess clinical association with PCGI TRN classifier.
Milestones:
1) A TRN describing the interaction of the PCGI dependent TFs and their targets in the DEGs.
2) A PCGI TRN classifier.
3) A PCGI TRN classifier correlated with clinical outcomes.
Target months: 12
Percentage of completion: 100%

What was accomplished under these goals?

We have developed a novel method to quantify degree of genomic alterations in a single PC sample using five different types of GI events, including broad and focal CNAs, genomic rearrangements, MSI, and somatic mutational burden. Using this, we identified potential links from the genetic alterations to secondary oncogenic driver activations resulting in a large number of transcriptional changes, thereby

Major accomplishments of this funding cycle include:

1) We identified Identify transcriptional signature (PCGI-E) describing the specific transcriptional programs in each PCGI high and low tumors.

PCGI computations in TCGA PRAD, MET440 and Taylor et al. cohorts: We have developed a novel scoring method of GI through summation of broad and focal copy number alterations (CNAs), microsatellite instability (MSI), tumor mutational burden (TMB), and genomic rearrangements (fusion). Categories of genomic alteration events were identified in individual tumor samples and summed in order to generate a numerical PCGI score after scaling by Min-Max normalization method. We have obtained genomic alteration profiles of TCGA[1], MET440[2] and Taylor et al.[3] cohorts due to the limited availability of the raw data from public databases and applied modified version of PCGI (pseudo-PCGI; here after pPCGI) computation method, which does not include MSI in the score computation because it is weakly correlated to the overall PCGI score with less variability (**Figure 1A**). To assess the use of pPCGI instead of PCGI, we have checked the correlation between PCGI and pPCGI scores in TCGA PRAD cohort. As shown in **Figure 1B**, it has strong correlation (Spearman's rho = 0.85, P<0.001).

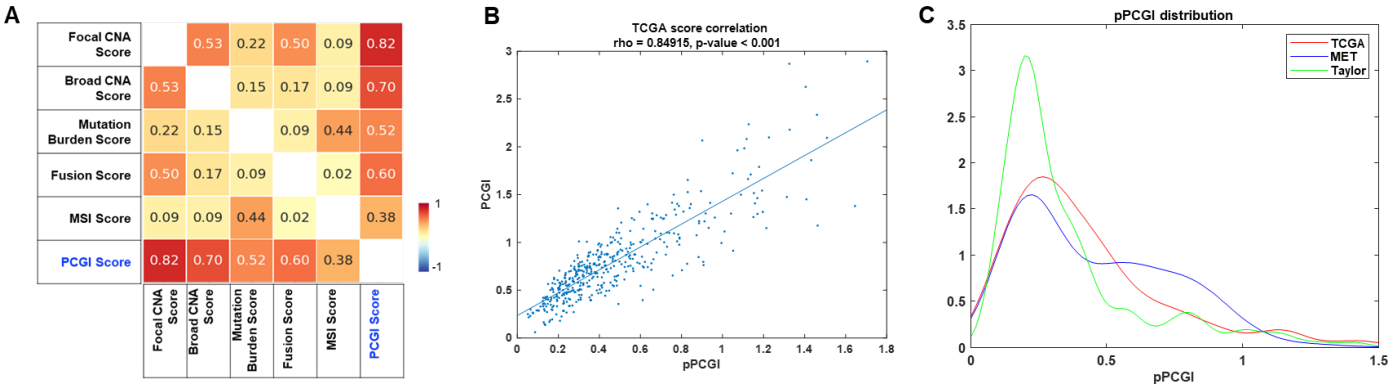

Figure 1. PCGI and pPCGI scores in 3 PC cohorts. A. Correlation between original PCGI score and individual genomic alterations. Correlation was measured by using Spearman's method. B. Comparison of original PCGI and pseudo-PCGI (pPCGI) scores in TCGA PRAD cohort. C. pPCGI distribution of TCGA, MET440, and Taylor et al. cohorts.

Given this, we have computed pPCGI scores for all 3 cohorts including TCGA, MET440, and Taylor et al. Of note, two cohorts MET440 and Taylor et al. including metastases have more samples with higher pPCGI score, while TCGA has a smaller number of samples with high pPCGI due to the samples are from primary prostate tissues (**Figure 1C**). This suggests that the higher PCGI is the more metastases.

Identification of significantly differentially expressed genes between PCGI high and low tumors: In order to identify differentially regulated genes by the genomic alterations. We first divided the patients three groups PCGI high, intemediate, and low groups at upper and lower tiles of the pPCGI score in each cohort. We then compared PCGI high and low groups to select the most significantly differentially expressed genes between the tumors highly genomically altered and less altered. For this, integrated hypothesis tesing was employed. Briefly, we computed T, median difference, and Rank-Sum stats. Significance levels of individual stats were obtained from individual null distribution and combined them suing Stouffer's method[4], resulting in combined p-values. Multiple correction with BH method was performed with the combined p-values. Differentially expressed genes (DEGs) were selected with false discovery rate (FDR) < 0.05 and log2-fold-change 1 (TCGA)

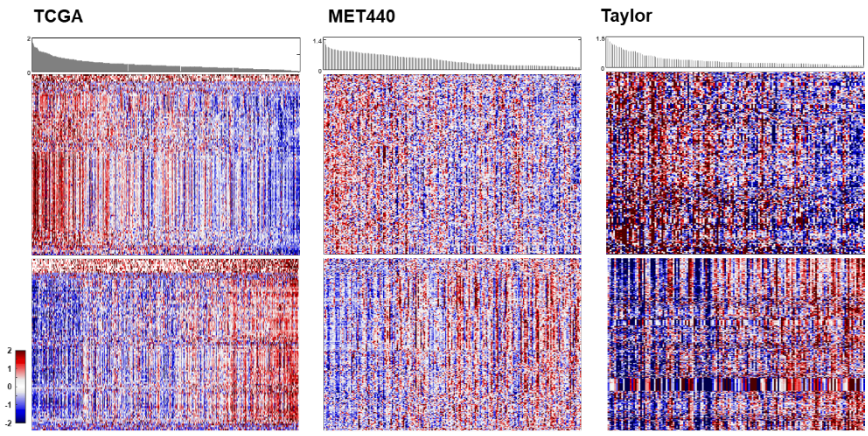

Figure 2. Heatmap of z-score normalized gene expression that are differentially expressed between pPCGI high and low. Top color bars indicate the pPCGI score of each data set. The differentially expressed genes indicate up-regulated genes (up panel) and down-regulated genes (down panel) between pPCGI high and low groups in TCGA (left), MET440(middle), and Taylor et al., (right). Each column in the heatmap is an individual sample.

or 0.58 (MET440 and Taylor et al.) according to deviation of the log2-fold-change distribution. Finally, we have identified 107, 193, and 201 up-regulated genes in pPCGI high tumors compared to pPCGI low tumors in TCGA, MET440, and Taylor et al., respectivley. We also identified 99, 234, and 337 down-regulated genes in the same comparisons in TCGA, MET440, and Taylor et al., respecively. The heatmaps display the differential expression pattern of the genes according to the PCGI scores (**Figure 2**).

Comparing GI driven transcriptional changes between TCGA, MET440, and Taylor et al. cohorts: We then compared those DEGs using Venn Diagram to see the common and distinct DEGs between the cohorts. As shown in Figure 3. Notably, there is less overlap of the DEGs between the 3 cohorts. Only 10 up-regulated genes

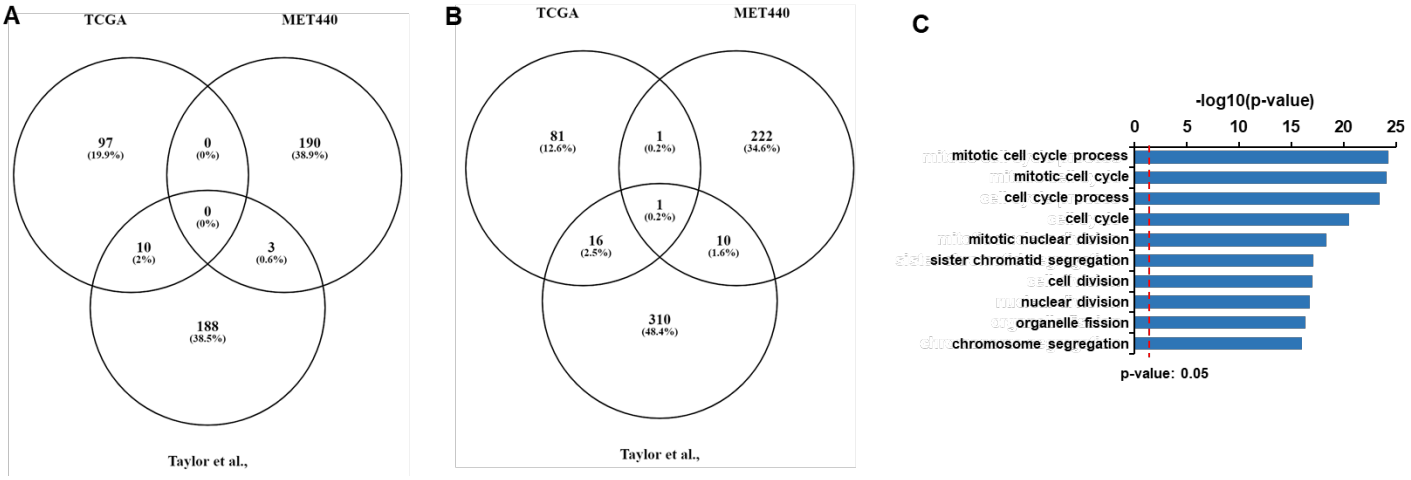

Figure 3. The Venn Diagrams of up- and down-regulated genes between pPCGI high and low tumors. A. Up-regulated genes. B. Down-regulated genes. C. Bar plot displays enrichment of the up-regulated genes.

are overlapped between TCGA and Taylor et al. (**Figure 3A**). Down-regulated genes are also a smaller overlapping gene between the cohorts (**Figure 3B**). However, functional enrichment of the DEGs using DAVID software[5] revealed that these genes are mostly associated with cell cycle, chromatin segregation and mitosis as shown in **Figure 1C**. This indicates that the most correlated genes with GI in PC are associated with cell cycle progression and chromatin regulations. These DEGs is defined as PCGI-E gene signature representing transcriptional changes by GI in PC.

2) We developed the PCGI TRN classifier to predict metastatic progression and assess clinical association with PCGI TRN classifier.

Master regulator analysis (MRA) of the DEGs: Given the DEGs, we performed MRA to identify the most relevant transcription factors (TFs) of the differential expression. For this, we have used in-house TF-target relationship database consisting of 799,657 TF-target interactions. Individual TF significance was tested by using null distribution based on 10,000 random sampling of target genes. Finally, we identified significant TFs for each cohort with significance level (p-value) < 0.01 and minimum number of targets ≥ 10.

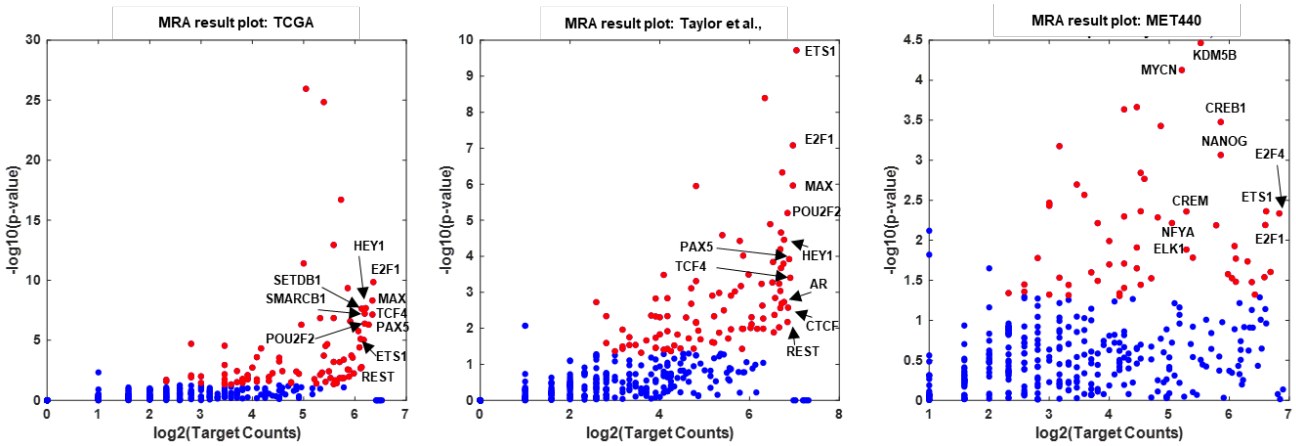

Figure 4. MRA results from TCGA (left), Taylor et al. (middle), and MET440 (right). Red dot indicates significant TFs and blue dot indicates insignificant TFs. The top 10 TFs were labeled with gene symbol.

TFs were sorted by a ratio of number of targets in the DEGs over total number of targets and top 10 TFs were selected from the sorted list of TFs (**Figure 4**).

Common and distinct master regulators (MRs) and their functional associations: To assess whether transcriptional drivers driven by genomic alterations are different between the cohorts, we compared the top10 TFs from TCGA, MET440, and Taylor et al. cohorts. We found that 8 MRs are overlapped between TCGA and

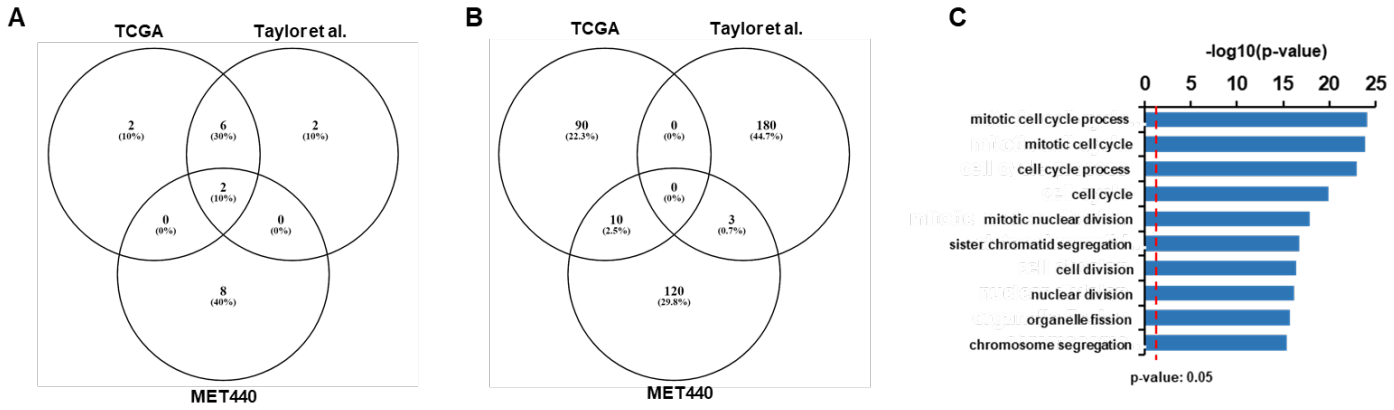

Figure 5. Top10 MRs and their functional association using their target genes. A. Venn Diagram depicts large overlaps of MRs between TCGA and Taylor et al. B. Venn Diagram depicts overlap of MRs' target genes. C. Bar plot displays enriched cellular processes based on the MRs' target genes.

Taylor et al., while MET440 has only 2 MRs overlapped (**Figure 5A**). We further checked whether their target genes are overlapped like MRs overlap, however a small overlap was found (**Figure 5B**). Despite a small overlap of the target genes, their cellular processes are highly overlapped and enriched in cell cycle, chromatin segregation, and mitosis (**Figure 5C**) consistent with the DEG enrichment (**Figure 3C**). Such a high overlap between TCGA and Taylor cohorts could be explained by the sample origin. Those samples are from primary tissue in prostate, pertaining prostate tissue specific transcriptional regulations. However, MET440 samples are from distance tissues of metastases and its transcriptional regulation should be different from prostate lineage gene expression. However, it is still an interesting part of the process involved in cell cycle and chromatin regulations.

Reconstruction of the transcriptional regulatory network (TRN) models as a PCGI TRN network: We have identified a total 20 MRs through MRA of the 3 cohorts. We then reconstruct MRs network model describing the interactions between the MRs with distinct number of targets in different sample context such as primary vs. metastasis or prostate vs. non-prostate. Protein-protein interaction information was obtained from STRING database[6] and construct network model with log2-fold-change between PCGI high and low tumors, as well as number of targets in the DEGs for each cohort. As shown in Figure 6, TCGA and Taylor cohorts from primary prostate tumor tissues have similar target enrichment and log2-fold-change pattern, while MET440 has very different target enrichment. These TRN model shows several AR interactors including ETS1 and KDM5B. ETS1

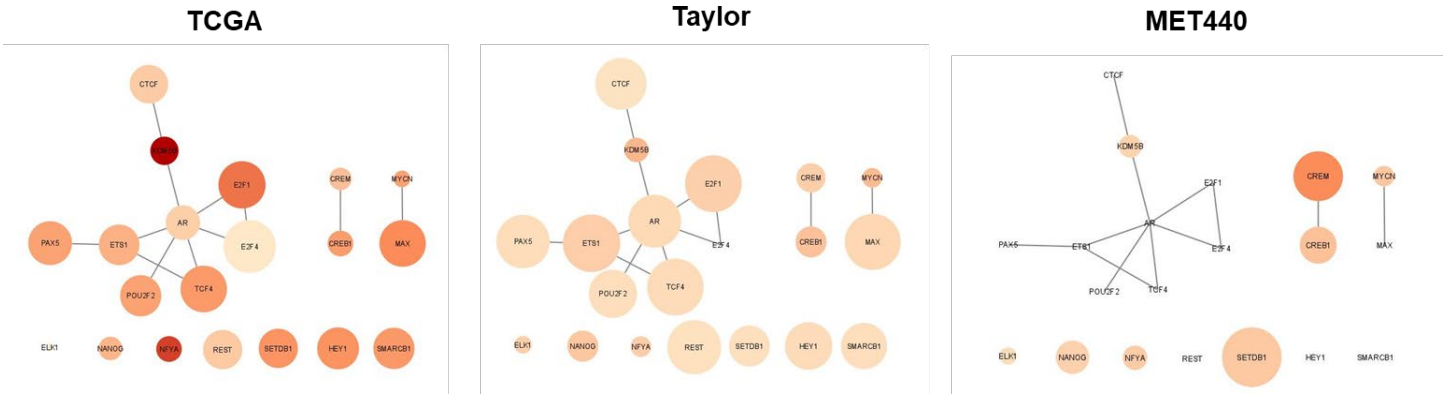

Figure 6. PCGI TRN models describing MRs interactions and their target gene regulation in each cohort. Node color: -log10(p-value). Node size: number of targets.

expression is highest in high-grade tumors and increased ETS1 expression and transcriptional activity promotes

an aggressive and castrate-resistant phenotype in immortalized PC cells[7]. Histone Demethylase KDM5B (a.k.a. JARID1B) was found to be overexpressed in PC[8] and it suppresses tumor suppress genes such as TP53[9]. We also found that NFYA and SETDB1 are active in all 3 cohorts which are associated with mutant TP53[10] or disruption of TP53[11]. In addition to these, WNT and Notch signaling downstream regulators, TCF4, NANOG, HEY1, CREM and CREB1 are highly active in PCGI high tumors. WNT and Notch signaling interlinked to promote prostate progenitor cell proliferations and differentiation[12]. In prostate cancer these MRs might associated with increased cancer cell plasticity, resulting in resistant to chemotherapy and/or radiotherapy[13]. Collectively, MR networks illustrates active transcriptional programs associated largely with 1) increase of castrate-resistant phenotype, 2) inhibition of tumor suppressors, and 3) lineage plasticity.

Clinical outcome association with PCGI TRN classifier score: Using the PCGI TRN, we computed PCGI TNR score and compared with PCGI-E signature score (**Figure 7A**). PCGI TRN score was calculated using unweighted Z-score method[14] with MRs and their target gene expression profiles. The same method was used to compute PCGI-E score with the PCGI-E signature (**Figure 3**). The PCGI TRN score was defined as a PCGI TRN classifier to discriminate PCGI TRN activation high and low tumors. For this analysis, Decipher GRID cohort[15] was employed and was grouped by using PCGI TRN score and PCGI-E score at median values (Figure 7A). Of note, PCGI TRN activity (TFA) high & PCGI-E high tumors exhibit poor clinical outcome with fast metastatic progression (**Figure 7B**). Cox proportional hazard regression analysis showed significant survival difference between TFA-high & PCGI-E high and others (**Figure 7C**; HR=1.43, CI=1.26-1.63). This demonstrates that our PCGI TNR classifier identifies aggressive PCs.

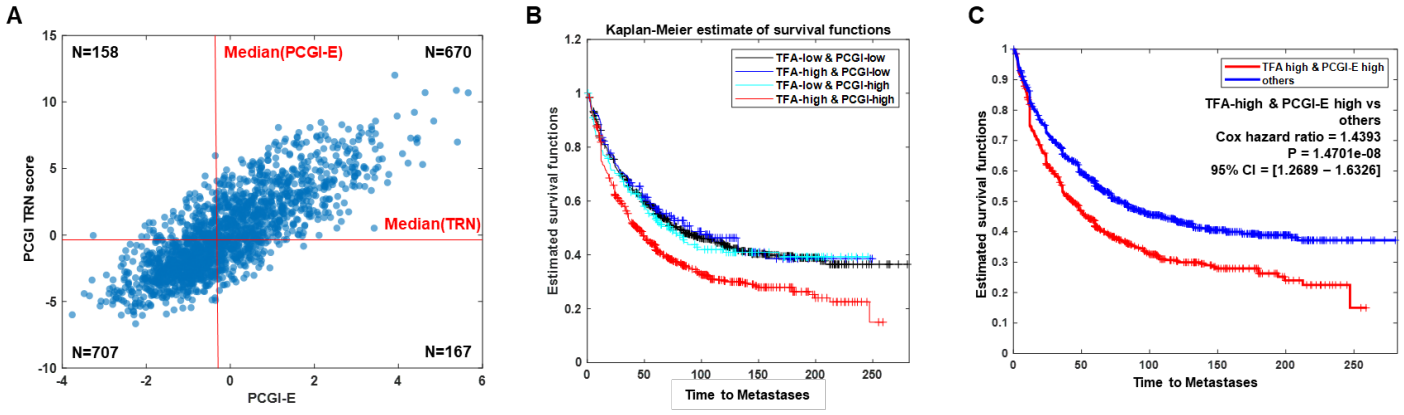

Figure 7. PCGI TRN score distribution and its clinical association. A. Scatter plot shows distribution of PCGI TRN score by PCGI-E score. B. Kaplan-Meier curves depicts distinct survival outcomes between TRN activity (TFA) high & PCGI-high tumors and other tumors. C. Significant survival difference of TFA-high & PCGI-E high versus others was shown.

4) Key research accomplishments

- Development of a novel method for GI score computation to identify PCs with highly altered genome.
- Identified the genes (PCGI-E) significantly correlated with GI events.
- Identified MRs that are relevant to PCGI-E gene signature regulation.
- Development of PCGI TRN classifier that can identify the PCs with poor survival outcomes.

5) Conclusion

A novel GI scoring method (pPCGI) facilatates identification of downstream transcriptional programs driven by highly altered genome and PCGI TNR classifier can sepearte tumors with poor clinical outcomes. This result also suggests that GI dependent transcriptional network assocaited with aggressive disease phenotype such as metatstatic progression.

6) Other achievement

We furthe developed a simplied method to compute GI score using precomputed genomic alteration profiles from cBioPortal dasebase with significantly high concordance with original PCGI score (**Figure 1B**). This allow us to access many other data with less computational cost to identify prostate tumors with high genomic alterations.

7) Reference

1. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
2. Wu, Y.-M. *et al.* Inactivation of CDK12 Delineates a Distinct Immunogenic Class of Advanced Prostate Cancer. *Cell* **173**, 1770-1782.e14 (2018).
3. Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
4. Hwang, D. *et al.* A data integration methodology for systems biology. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 17296–17301 (2005).
5. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2009).
6. Szklarczyk, D. *et al.* The STRING database in 2021: customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* **49**, D605–D612 (2021).
7. Smith, A. M. *et al.* ETS1 transcriptional activity is increased in advanced prostate cancer and promotes the castrate-resistant phenotype. *Carcinogenesis* **33**, 572–580 (2012).
8. Jose, A. *et al.* Histone Demethylase KDM5B as a Therapeutic Target for Cancer Therapy. *Cancers* **12**, E2121 (2020).
9. Shen, X. *et al.* JARID1B modulates lung cancer cell proliferation and invasion by regulating p53 expression. *Tumour Biol. J. Int. Soc. Oncodevelopmental Biol. Med.* **36**, 7133–7142 (2015).
10. Di Agostino, S. *et al.* Gain of function of mutant p53: the mutant p53/NF-Y protein complex reveals an aberrant transcriptional mechanism of cell cycle regulation. *Cancer Cell* **10**, 191–202 (2006).
11. Ogawa, S. *et al.* SETDB1 Inhibits p53-Mediated Apoptosis and Is Required for Formation of Pancreatic Ductal Adenocarcinomas in Mice. *Gastroenterology* **159**, 682-696.e13 (2020).
12. Shahi, P., Seethammagari, M. R., Valdez, J. M., Xin, L. & Spencer, D. M. Wnt and Notch pathways have interrelated opposing roles on prostate progenitor cell proliferation and differentiation. *Stem Cells Dayt. Ohio* **29**, 678–688 (2011).
13. Takebe, N. *et al.* Targeting Notch, Hedgehog, and Wnt pathways in cancer stem cells: clinical update. *Nat. Rev. Clin. Oncol.* **12**, 445–464 (2015).
14. Levine, D. M. *et al.* Pathway and gene-set activation measurement from mRNA expression data: the tissue distribution of human pathways. *Genome Biol.* **7**, R93 (2006).
15. Tomlins, S. A. *et al.* Characterization of 1577 primary prostate cancers reveals novel biological and clinicopathologic insights into molecular subtypes. *Eur. Urol.* **68**, 555–567 (2015).

What opportunities for training and professional development has the project provided?

At this stage in the funding cycle, I have published a paper early this year in *Prostate Cancer and Prostatic Disease* as a senior author. This study describes a systematic comparison of PC subtyping methods, PCS and PAM50. We found that PCS is outperform in identification of tumors with fast metastatic progression.

How were the results disseminated to communities of interest?

An early version of this work has been presented at School of Life Science Online Seminar in GIST Korea on June 3rd. These findings are being prepared for publication now.

What do you plan to do during the next reporting period to accomplish the goals?

A major objective of the second year of funding period is to determine common and distinct transcriptional programs between the two mutually exclusive GI driver networks. For this, we will 1) assess mutual exclusivity of GI driver networks driven by TP53 and SPOP; 2) identify transcriptional signatures of the novel GI drivers in the networks; and 3) reconstruct integrated networks describing interactions of GI drivers and downstream transcriptional programs identified in the first year.

4. IMPACT

What was the impact on the development of the principal discipline(s) of the project?

I have made an important conceptual and clinically relevant advance by developing a novel method of characterizing PC using genomic and transcriptomic profiles. Consequently, this project is high impact and high reward, with potentially immediate opportunities to alter clinical practice if the classification scheme can be shown to have clinical utility. This new PC classification scheme I developed might improve prognostication of PC and enable the development of subtype-specific therapies and companion diagnostics. Using computational

modeling, I have also identified 20 MRs which appears to be highly active in CRPC/Met tumors, but which are associated with castrate-resistant phenotype, 2) inhibition of tumor suppressors, and 3) lineage plasticity, and therefore represents a comprehensive catalog of MRs in PC with highly altered genome.

What was the impact on other disciplines?
Nothing to Report.

What was the impact on technology transfer?
Nothing to Report.

What was the impact on society beyond science and technology?
Nothing to Report.

5. CHANGES/PROBLEMS

Changes in approach and reasons for change
Nothing to Report.

Actual or anticipated problems or delays and actions or plans to resolve them
Nothing to Report.

Changes that had a significant impact on expenditures
Nothing to Report.

Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents
Nothing to Report.

6. PRODUCTS:

Publications, conference papers, and presentations

Journal publications.
Yoon, J., Kim, M., Posadas, EM., Freedland, SJ., Liu, Y., Davicioni, E., Den, RB., Trock, BJ., Karnes, RJ., Klein, EA., Freeman, MR., **You, S**., "A comparative study of PCS and PAM50 prostate cancer classification schemes". Prostate Cancer and Prostatic Diseases, 2021 Feb 2;10.1038/s41391-021-00325-4. (2021). PMID: 33531653.
Acknowledgement of federal support (Yes)

Books or other non-periodical, one-time publications.
Nothing to Report.

Other publications, conference papers, and presentations.
Lecture:
You S, Driver gene networks of genomic instability in prostate cancer progression. Gwangju Institute of Science and Technology (GIST), Korea. School of Life Science Online Seminar. June 3rd, 2021.

Website(s) or other Internet site(s)
Nothing to Report.

Technologies or techniques
Nothing to Report.

Inventions, patent applications, and/or licenses
Nothing to Report.

Other Products
Nothing to Report.

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name:	Sungyong You
Project Role:	Principal Investigator
Researcher Identifier:	yousung1
Nearest person month worked:	1.8
Contribution to Project:	Dr. You has supervised Drs. Minhyung Kim and Yeonjoo Lee to perform the genomic and transcriptomic data analysis. In addition, Dr. You wrote report and paper and provide lecture to disseminate this study result.
Funding Support:	The Urology Care Foundation Research Scholar Program

Name:	Minhyung Kim
Project Role:	Co-Investigator
Researcher Identifier:	MINHYUNGK
Nearest person month worked:	3.6
Contribution to Project:	Dr. Kim performed differential expression analysis and survival analysis.
Funding Support:	W81XWH-20-1-0644 PI: **You** R01 CA246304-01 PIs: Agopian/Tseng/**You**

Name:	Yeonjoo Lee
Project Role:	Co-Investigator
Researcher Identifier:	LEEYEO
Nearest person month worked:	8.4
Contribution to Project:	Dr. Lee performed PCGI computation, differential gene expression analysis, TRN network modeling, TRN score computation.
Funding Support:	W81XWH-20-1-0644 PI: **You**

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?
Nothing to Report.

What other organizations were involved as partners?
Nothing to Report.

8. SPECIAL REPORTING REQUIREMENTS: Nothing to Report.

9. APPENDICES: Published paper in *Prostate Cancer and Prostatic Diseases*

Prostate Cancer and Prostatic Diseases
https://doi.org/10.1038/s41391-021-00325-4

ARTICLE

Clinical Research

A comparative study of PCS and PAM50 prostate cancer classification schemes

Junhee Yoon[1] · Minhyung Kim[1] · Edwin M. Posadas[2,3] · Stephen J. Freedland[1,4] · Yang Liu[5] · Elai Davicioni[5] · Robert B. Den[6] · Bruce J. Trock[7] · R. Jeffrey Karnes[8] · Eric A. Klein[9] · Michael R. Freeman[1,10,11] · Sungyong You[1,10]

Received: 25 May 2020 / Revised: 20 December 2020 / Accepted: 15 January 2021

Abstract

Background Two prostate cancer (PC) classification methods based on transcriptome profiles, a de novo method referred to as the "Prostate Cancer Classification System" (PCS) and a variation of the established PAM50 breast cancer algorithm, were recently proposed. Both studies concluded that most human PC can be assigned to one of three tumor subtypes, two categorized as luminal and one as basal, suggesting the two methods reflect consistency in underlying biology. Despite the similarity, differences and commonalities between the two classification methods have not yet been reported.

Methods Here, we describe a comparison of the PCS and PAM50 classification systems. PCS and PAM50 signatures consisting of 37 (PCS37) and 50 genes, respectively, were used to categorize 9,947 PC patients into PCS and PAM50 classes. Enrichment of hallmark gene sets and luminal and basal marker gene expression were assessed in the same datasets. Finally, survival analysis was performed to compare PCS and PAM50 subtypes in terms of clinical outcomes.

Results PCS and PAM50 subtypes show clear differential expression of PCS37 and PAM50 genes. While only three genes are shared in common between the two systems, there is some consensus between three subtype pairs (PCS1 versus Luminal B, PCS2 versus Luminal A, and PCS3 versus Basal) with respect to gene expression, cellular processes, and clinical outcomes. PCS categories displayed better separation of cellular processes and luminal and basal marker gene expression compared to PAM50. Although both PCS1 and Luminal B tumors exhibited the worst clinical outcomes, outcomes between aggressive and less aggressive subtypes were better defined in the PCS system, based on larger hazard ratios observed.

Conclusion The PCS and PAM50 classification systems are similar in terms of molecular profiles and clinical outcomes. However, the PCS system exhibits greater separation in multiple clinical outcomes and provides better separation of prostate luminal and basal characteristics.

These authors contributed equally: Junhee Yoon, Minhyung Kim

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41391-021-00325-4.

✉ Sungyong You
Sungyong.You@cshs.org

[1] Department of Surgery, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[2] Urologic Oncology Program & Uro-Oncology Research Program, Cedars-Sinai Cancer, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[3] Division of Oncology, Department of Medicine, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[4] Division of Urology, Department of Surgery, Veteran Affairs Healthcare System, Durham, NC, USA

[5] Decipher Biosciences Inc., San Diego, CA, USA

[6] Department of Radiation Oncology, Jefferson Medical College of Thomas Jefferson University, Philadelphia, PA, USA

[7] James Buchanan Brady Urological Institute, Johns Hopkins Hospital, Baltimore, MD, USA

[8] Department of Urology, Mayo Clinic, Rochester, MN, USA

[9] Glickman Urological and Kidney Institute, Cleveland Clinic, Cleveland, OH, USA

[10] Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[11] Department of Medicine, University of California, Los Angeles, CA, USA

Published online: 02 February 2021

Introduction

There has been much recent progress in the use of prostate cancer (PC) genomics to identify drivers of aggressive disease. These advances have spurred development of commercially available genomic classifiers that can identify aggressive tumors at high risk of progression to metastatic and/or castration resistant PC (CRPC) [1–7]. Despite this technology, there is no universally accepted or widely used molecular subtyping system for PC, unlike other cancers such as breast cancer, in which luminal A, luminal B, and basal subclassification are commonplace and clinically meaningful. Despite these limitations, PC subtypes have been proposed based on genomic criteria, such as various somatic alterations in chromatin sequence, (e.g., TMPRSS-ERG fusions [8, 9]) and androgen receptor amplification [3, 4]. The Cancer Genome Atlas identified several genomic PC subtypes, referred to as ERG, ETV1, ETV4, FLI1, SPOP, FOXA1, IDH1, and Other [7]. Tomlins et al. [10] described four subtypes based on gene expression (ERG+, ETV+, SPINK1+, and Triple Negative (ERG−/ETS−/SPINK1−)). However, the clinical applicability of these PC genomic classifiers has been limited [11, 12].

Many reports have proposed gene expression signatures as a means of tumor classification [13–18]. Two transcriptome-based classification methods were recently reported that categorize PC into three subtypes [19, 20]. Our group described the prostate cancer classification system (PCS), an integrated approach employing activation signatures of 14 pathways associated with PC biology to interrogate a virtual cohort of 1321 clinical samples [19]. Two luminal subtypes (PCS1 and PCS2) and one basal subtype (PCS3) were described in this report. PCS1 tumors exhibited the poorest clinical outcomes, including increased risk of metastatic progression, PC-specific mortality, and overall survival. In contrast, no significant differences in clinical outcomes were observed between PCS2 and PCS3; however, both PCS1 and PCS3 tumors were enriched in bone metastases in comparison to PCS2 [21]. In the initial report, we validated the PCS scheme with ten independent patient cohorts and 19 laboratory models of PC. In line with this, PCS1-specific genes were highly expressed in androgen receptor signaling inhibitor (ARSI) resistant PC [19] and this was independently validated by our recent study using a novel circulating tumor cell RNA assay system [22]. The initial You et al. study [19] was the first to categorize PC into only three subtypes using genomic approaches.

Zhao et al. [20] applied a variation of the widely used PAM50 breast cancer classifier to a large PC transcriptome dataset. PAM50 classifies breast cancer into luminal A (LumA), luminal B (LumB), HER2, Normal-like, and Basal subgroups [23]. The adaptation of PAM50 to PC focused on luminal and basal PC phenotypes and disregarded the HER2 and Normal-like subtypes described in the standard PAM50 system. Similar to the PCS system, PAM50 revealed that transcriptome data alone can divide PCs into only three subtypes, LumA, LumB, and Basal. These subgroups were associated with varied clinical behaviors. LumB exhibited the worst survival rate, while LumA and Basal subtypes exhibited similar clinical outcomes with better survival rates. In comparison to other classification schemes, which identified as many as seven PC subtypes [1–7], it is intriguing that both PCS and PAM50 concluded that PCs can be categorized into two distinct luminal and one basal subtype using only transcriptome data.

To date, no comparison has been made between these two different three-category classification systems. To address this gap, here we present a comparison of the PCS and PAM50 systems. We applied the PCS and PAM50 methods to two large PC transcriptome datasets: (1) the Prostate Cancer Transcriptome Atlas (PCTA) and (2) the Decipher GRID^{TM} database (GRID) (Supplementary Table 1). The PCTA is a virtual cohort consisting of 1,321 PC transcriptome profiles. These data were used to develop the PCS [19]. The GRID is a cohort consisting of 8626 PC transcriptome profiles, a subset of which was employed in the Zhao et al. study [20]. We hypothesized that the PCS and PAM50 systems have many similarities as well as differences in terms of molecular profiles and clinical outcomes. We thus performed a comparative analysis assessing three different measures: (1) gene expression patterns, (2) pathway associations, and (3) correlation with clinical outcomes.

Methods

Statistical analysis

To examine the association between clinical outcomes and PCS and PAM50 categories in the tumors from the GRID dataset, we used Kaplan–Meier survival analysis and Cox proportional hazards regression analysis with the following outcomes: biochemical recurrence (BCR), metastasis (Met), and PC-specific mortality (PCSM). To test whether PCS or PAM50 classification was prognostic independent of other clinical variables, multivariable analyses were performed adjusting for pathological grade, which was categorized into three groups with Gleason sum score of less than 7, equal to 7, and more than 7. All analyses were conducted using Python (version 2.7). $p < 0.05$ was considered statistically significant. Detailed information of transcriptome datasets and analysis procedures used in this study are described in Supplementary Methods.

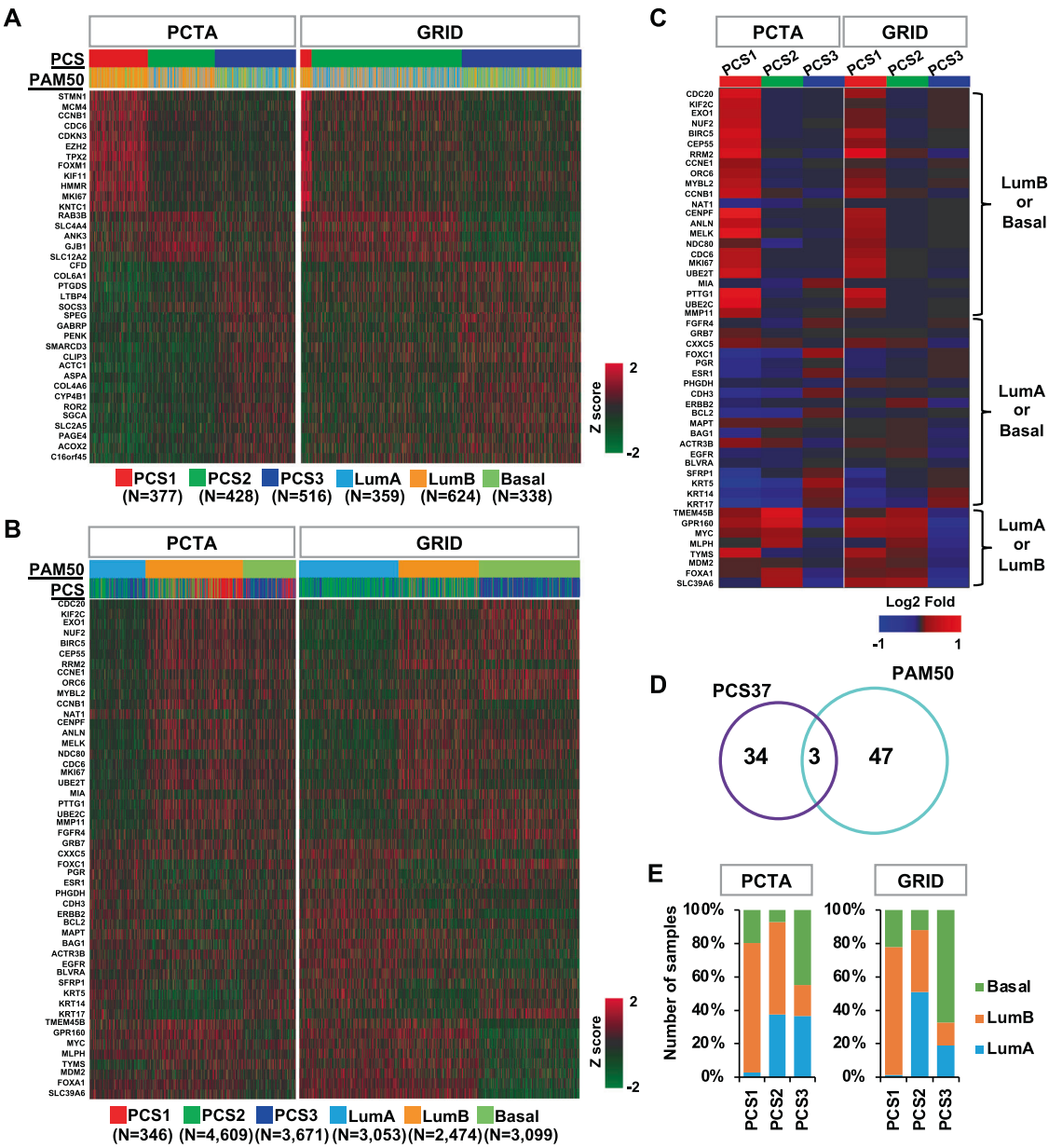

Fig. 1 PCS37 and PAM50 genes in the PCTA and the GRID datasets. The heatmap depicts differential expression of PCS37 and PAM50 in the PCTA and GRID datasets based on **A** PCS and **B** PAM50 grouping. **C** The heatmap shows differential expression of PAM50 genes in the PCTA and GRID datasets based on PCS grouping. **D** The Venn diagram of PCS37 and PAM50. **E** The distribution of PAM50 subtypes across three PCS subtypes in the PCTA and GRID datasets. PCS prostate cancer subtype, LumA Luminal A subtype, LumB Luminal B subtype, Basal Basal subtype.

Results

Conserved gene expression patterns between PCS and PAM50

We applied the PCS and PAM50 classifiers to assign the tumors from the PCTA cohort ($n = 1{,}321$) [19] and the GRID cohort ($n = 8{,}626$) into PCS1-3 and LumA/B and Basal subtypes. Differential expression patterns of PCS37 genes [19] were displayed by PCS classes with the PCTA and the GRID datasets (Fig. 1A and Supplementary Fig. 1), showing distinct expression patterns of individual PCS class-specific genes. For example, 12 PCS1-specific genes (STMN1, MCM4, CCNB1, CDC6, CDKN3, EZH2, TPX2, FOXM1, KIF11, HMMR, MKI67, and KNTC1) are only highly expressed in PCS1 tumors (Fig. 1A). The GRID contains 4% (346/8,626) PCS1, the most aggressive PCS subtype, as most specimens were primary PCs. In the PCTA cohort, containing 260 metastatic PCs, 26% (377/1,321) of patients were assigned to PCS1. In total, 66% (176/260) of

metastatic PCs belong to this PCS1 subtype. LumB, the most aggressive PAM50 subtype, accounts for 29% (2,474/8,626) of the GRID cohort and 47% (624/1,321) of the PCTA cohort. Of 260 metastatic PCs in the PCTA cohort, 170 (65%) patients were assigned LumB.

We then assessed PAM50 gene expression by PAM50 classes with both datasets (Fig. 1B). These 50 genes also showed distinct expression patterns among the classes. However, their expression was not unique to a specific subtype. For example, some genes that were highly expressed in LumB also showed high expression in Basal. In addition, some genes (e.g., NAT1, GRB7, MAPT) did not show a clear differential expression pattern among the PAM50 classes. Both PCS37 and PAM50 displayed clear separation into three PC classes. We further assessed the expression pattern of PAM50 genes by the PCS classifier (Fig. 1C). Of note, genes with high expression in LumB or Basal were highly expressed in PCS1. These genes exhibit mostly low expression in LumA.

Given that both classification methods resolve all PCs into three subtypes, it is interesting that only three genes (CCNB1, CDC6, and MKI67) overlap between the two systems (Fig. 1D). These three genes show high expression in both PCS1 and LumB. We observed all PCS categories in all classification groups as defined by PAM50 in both PCTA and GRID (Fig. 1E). We found a high frequency of PCS2 in LumA, but not PCS1 and PCS3. LumB was enriched for PCS1, while Basal subtype but not LumA or LumB was enriched for PCS3. This was validated by visualizing the distribution of the tumors from each category and their overlaps using PCS37 and PAM50 gene expression (Supplementary Fig. 2A). We then defined centroids of each categories and estimated the pairwise distances between the PCS and PAM50 categories (Supplementary Fig. 2B and Supplementary Table 2). Consistent with the enrichment of sample numbers shown in Fig. 1E, three subtype pairs (PCS1 versus LumB, PCS2 versus LumA, and PCS3 versus Basal) exhibit the shortest distance between centroids. Taken together, three subtype pairs (PCS1 versus LumB, PCS2 versus LumA, and PCS3 versus Basal) show comparable enrichment in terms of number of tumors commonly included in both subtypes, despite the small number of genes shared by PCS37 and PAM50.

Pairwise overlap of PCS and PAM50 classes in cellular functions

We computed enrichment scores using the GSEA method [24] with hallmark gene sets [25] for all the PCS and PAM50 categories. In the PCTA, PCS1 samples were highly enriched with E2F targets and G2M checkpoint; PCS2 samples were enriched with androgen response, fatty acid metabolism, and cholesterol homeostasis; and PCS3 was enriched with IL6-JAK-STAT3 signaling and KRAS signaling (UP) (Fig. 2A). PAM50-classified samples in the PCTA showed an overall similar enrichment result with PCS classification (Fig. 2A). LumA samples were enriched in fatty acid metabolism, androgen response, cholesterol homeostasis. LumB samples were enriched with the hallmark gene sets of PCS1, such as E2F targets and G2M checkpoint. Basal samples were enriched in the same hallmark gene sets as PCS3 (Fig. 2B). In the GRID, PCS1, and PCS2 were enriched in E2F targets, G2M checkpoint, fatty acid metabolism, androgen response, and cholesterol homeostasis, but PCS3 did not show any significant enrichment of any of these hallmark gene sets. PAM50-classified samples in the GRID showed the same enrichment pattern. LumB exhibited higher ES of E2F targets and G2M checkpoint than LumA, but the highest score was androgen response in LumB-enriched hallmark gene sets (Supplementary Fig. 3A). The highest enrichment score in LumA was androgen response, similar to PCS2 (Supplementary Fig. 3B).

Luminal and basal marker expression in PCS and PAM50 classifications

We assessed expression levels of luminal and basal cell marker genes in the PCTA and the GRID datasets [19, 26]. Expression of these marker genes was highly concordant with PCS classification in both the PCTA (Fig. 3A) and the GRID (Fig. 3B), consistent with our prior results in the initial PCS report [19]. In the PAM50 classes, LumB and Basal subtypes display clear high and low expression of luminal and basal marker genes, respectively. However, LumA tumors exhibit mixed marker gene expression, indicating a composite luminal and basal expression phenotype. This result suggests that the PCS classification scheme is superior in separating prostatic luminal and basal gene expression.

PCS1 and LumB are associated with the worst survival outcomes

We examined relationships between subtype category and the clinical outcomes of PCSM, BCR, and Met using the GRID. PCS1 within the PCS categories (Fig. 4A) and LumB within the PAM50 categories (Fig. 4B) were the most aggressive for all three metrics (PCSM, BCR, and Met). Cox proportional hazard analysis was performed using the two classification systems. In univariable analysis using the PCS scheme, the PCS1 subtype exhibited the highest hazard ratio (HR) (PCSM: 3.91 [2.78–5.52], BCR: 2.48 [1.34–4.08], Met: 3.91 [2.78–5.52]). Similarly, the LumB subtype exhibited the highest HR (PCSM: 1.85 [1.32–2.61], BCR: 1.54 [1.01–2.29], Met: 2.04 [1.58–2.55])

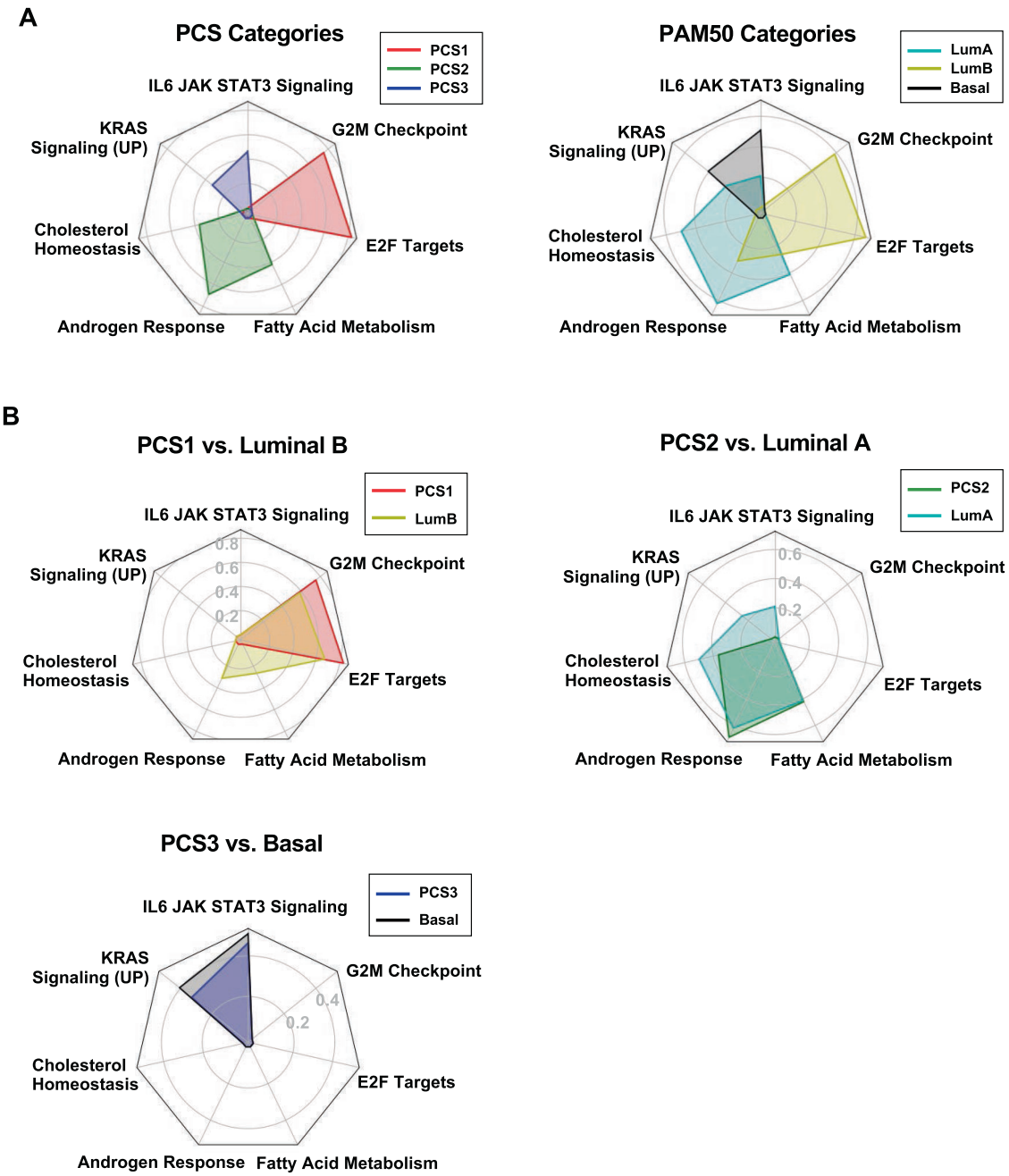

Fig. 2 Enriched cellular processes of PCS and PAM50 subtypes in the PCTA. A Enriched seven hallmark gene sets of PCS (left) and PAM50 (right) categories were displayed with radar chart in the PCTA cohort. **B** Radar chart illustrates overlaps of enriched hallmark gene sets by PCS1 versus LumB, PCS2 versus LumA, and PCS3 versus Basal in the PCTA cohort. LumA Luminal A subtype, LumB Luminal B subtype, Basal Basal subtype.

(Fig. 4C). In all combinations of pairwise comparisons, except for LumB compared to LumA in BCR (HR = 3.41 [2.40–4.87]), PCS1 exhibited higher HRs compared to LumB (Supplementary Table 3). We also performed multivariable analysis of PCS or PAM50 adjusting for Gleason score (Table 1). After the adjustment, PCS still showed larger HR (PCSM: 3.76 [2.17–6.25], BCR: 2.03 [1.39–2.96], Met: 3.38 [2.26–5.06]) compared to PAM50 (PCSM: 1.51 [0.96–2.38], BCR: 1.43 [1.14–1.79], Met: 1.73 [1.30–2.29]). Importantly, we found that controlling for Gleason score did not dramatically affect the HRs in either PCS or PAM50, indicating both are largely independent of Gleason grade for predicting patient outcomes. Of note, both multivariable PCS and PAM50 models with Gleason score exhibited the same concordance index at 0.75. We further checked the association of each subtype and Decipher score [27], which is designed to predict metastasis risk after radical prostatectomy. The Decipher

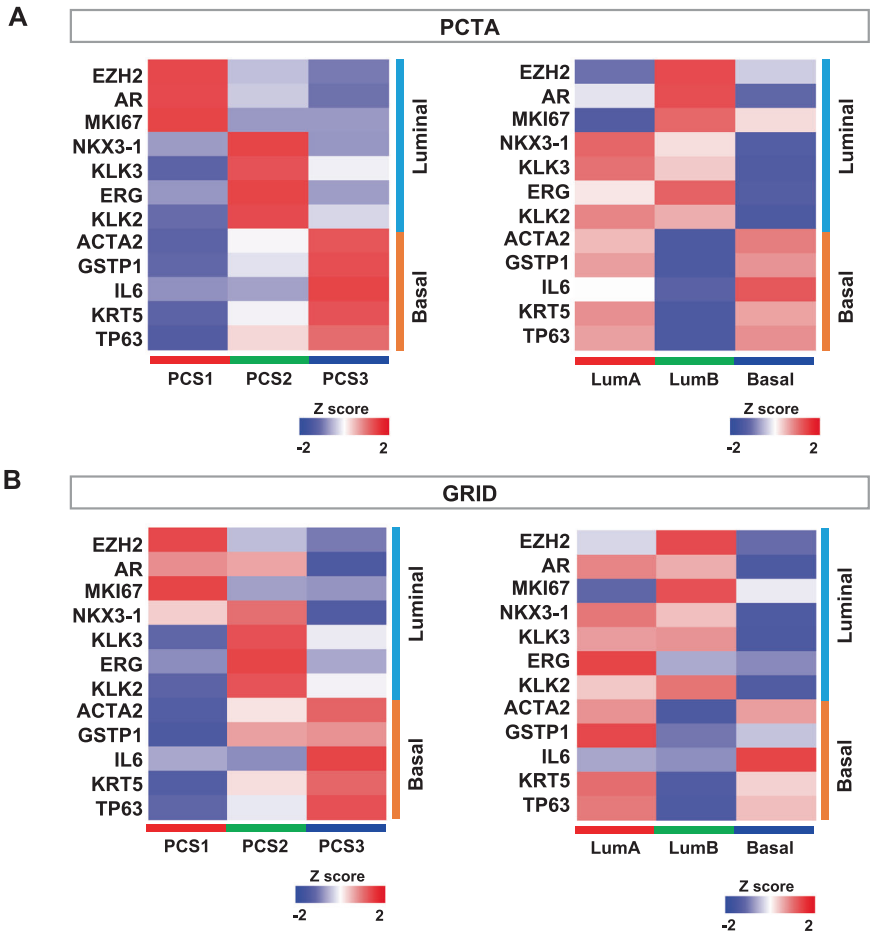

Fig. 3 Gene expression of Luminal and Basal cell markers in the PCTA and the GRID. The heatmap displays expression of the Luminal and Basal cell marker genes in the PCTA (**A**) and the GRID (**B**).

score was significantly higher in PCS1 and LumB compared to other subtypes (Supplementary Fig. 4A, B). Multivariable analysis of Decipher score after adjustment of Gleason score was performed to compare with PCS and PAM50. Decipher score exhibited the highest HRs and significant p values in PCSM, BCR, and Met (Supplementary Fig. 4C).

Discussion

In this study, we performed a comprehensive cross-comparison of subtype assignments obtained from the PCS and PAM50 classification methods. We found that three subtype pairs (PCS1 versus LumB, PCS2 versus LumA, and PCS3 versus Basal) show conserved gene expression patterns in the tumors included in both subtypes. Distance metrics of subtype centroids followed by visualization showed a similarity between these three pairs of subtypes. Despite the fact that only three genes overlap between the PCS37 and PAM50 signatures, these three subtype pairs exhibit very similar enrichment of cellular processes, as well as the distance of the centroids between the subtypes (e.g., PCS1 and LumB). This result suggests high concordance of the two subtypes in cellular functions and clinical phenotypes. Finally, a further comparison was made in clinical outcomes with PCSM, BCR, and Met, demonstrating that PCS1 and LumB subtypes are the most aggressive tumors with the worst survival outcomes in each categorization scheme.

The two classification methods likely reflect underlying luminal and basal biology as described in basic PC studies [19, 20]. PCS was originally designed as an unbiased classification scheme, with 14 PC relevant pathway activation signatures that revealed two luminal and one basal phenotype, while the PAM50 as applied to PC was biased toward luminal and basal phenotypes at the outset [20]. As such, it is interesting that two distinct PC categorization schemes converged to similar conclusions about the manner in which PCs can be grouped from transcriptome data alone. As shown in Fig. 3, both PCS and PAM50 exhibit differential expression of luminal and basal marker genes. However, PCS classes show consistent luminal and basal marker gene expression with clear separation between PCS1/2 and PCS3, respectively, while LumA tumors have high expression of basal marker genes, which are highly expressed in Basal tumors as well. This suggests that PCS classification more accurately identifies prostate luminal and basal phenotypes expressed within the cancer specimens.

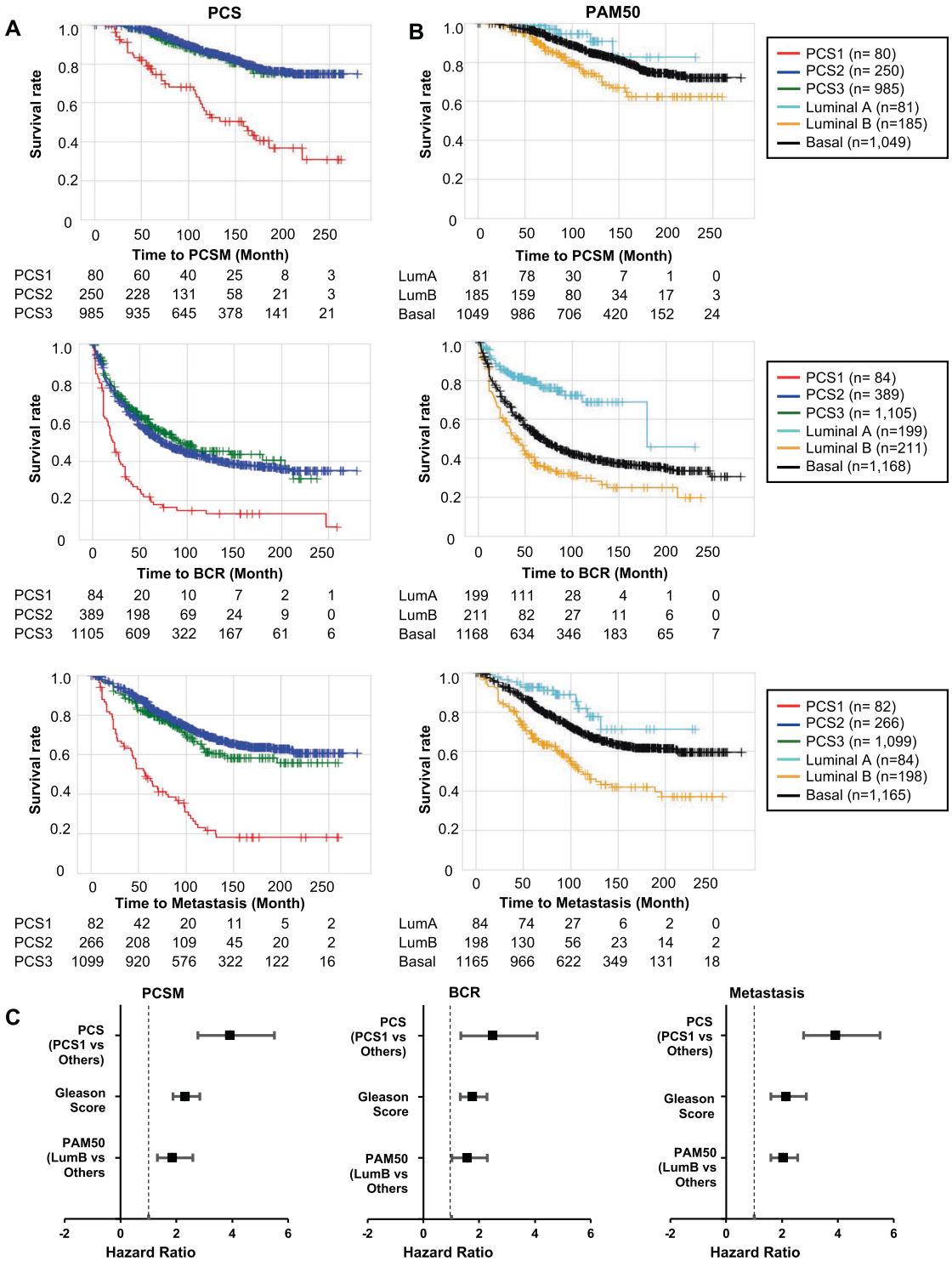

Fig. 4 Clinical outcomes in distinct PCS and PAM50 subtypes in the GRID. Kaplan–Meier survival curves shows differential clinical outcome association of PCS (**A**) and PAM50 (**B**) categorization. Tables below the KM plot represent the number at risk. (PCSM prostate cancer-specific survival, BCR biochemical recurrence, Met metastasis.) **C** Forest plots display hazard ratios of PCS, PAM50, and Gleason grade against PCSM, BCR, and Met. (GS Gleason score).

The development of classification schema such as PCS and PAM50 in PC represents an important evolution in the field. Rather than considering an isolated biomarker or limited groups of biomarkers, these approaches focus on patterns of gene expression that ultimately relate to key biological drivers of the disease process. Thus far, these classifiers have shown

Table 1 Multivariable Cox proportional hazard regression analysis for PCS and PAM50 classification with clinical parameters.

Multivariable analysis with PCS				Multivariable analysis with PAM50			
Variable	Hazard ratio	*p* value	C-index	Variable	Hazard ratio	*p* value	C-index
PCSM							
PCS	3.76 (2.17–6.52)	<0.001	0.75	PAM50	1.51 (0.96–2.38)	0.073	0.75
Gleason score	2.16 (1.74–2.66)	<0.001		Gleason score	2.25 (1.83–2.78)	<0.001	
BCR							
PCS	2.03 (1.39–2.96)	<0.001	0.66	PAM50	1.43 (1.14–1.79)	0.002	0.67
Gleason score	1.71 (1.55–1.88)	<0.001		Gleason score	1.69 (1.54–1.86)	<0.001	
Met							
PCS	3.38 (2.26–5.06)	<0.001	0.71	PAM50	1.73 (1.30–2.29)	<0.001	0.72
Gleason score	2.03 (1.79–2.31)	<0.001		Gleason score	2.04 (1.80–2.32)	<0.001	

Multivariate analysis of PCS or PAM50 classification system with Gleason score was performed.

the capacity to enhance clinical prediction and prognostication beyond conventional clinical criteria used in practice today. Emerging data with PAM50 and PCS point toward their potential utility in personalizing the approach to PC care, particularly with the use of systemic therapies such as ARSIs (e.g., abiraterone, enzalutamide, apalutamide, and darolutamide) and taxanes (e.g., docetaxel and cabazitaxel). Analysis of samples from the SPARTAN clinical trial revealed clinically important differences in behavior in response to apalutamide, based on PAM50 classification. In SPARTAN, PAM50 Basal patients seemed to have a greater absolute benefit from ADT and apalutamide, suggesting these patients may benefit from intensification with ARSIs [28]. Consistent with this, analysis of PCS1 and LumB specimens in the GRID from patients receiving ADT showed higher rates of PCSM, BCR, and Met compared to PCS2/3 or LumA/Basal (Supplementary Fig. 5). However, absence of information on ADT response in the GRID prohibits analysis of relative performance of the PCS and PAM50 classifications with respect to this clinical variable. While encouraging, prospective validations of these data are needed prior to more widespread deployment of these approaches. Studies are now underway that will provide important insights into this question, including the ongoing NRG-GU-006/BALANCE study (NCT03371719), which used PAM50 as a stratification variable, and has now finished accrual and is expected to yield results within the next 3 years.

While tissue-based classification will remain an important standard in PC, the clinical behavior of this disease makes obtaining tissue samples from patients with metastatic disease challenging. Given the potential importance of these tools in advanced PC/metastatic CRPC, additional means of deploying these technologies is crucial. Fortunately, given recent technical advances, it will become possible to conduct genomic classifications using liquid biopsies [22]. These blood-based tools will also require prospective validation. Liquid biopsies are included in ongoing studies, including NRG-GU-006/BALANCE. These initial studies will provide a foundational experience for use of PCS and PAM50 in clinical practice.

In order for these classification schemes to be useful in a clinical setting, it is critical that they offer actionable prognostic information. We observed a disparate percentage of subtype assignments within the study cohorts. The GRID consists of a prospective cohort ($n = 7{,}000$) and a retrospective cohort ($n = 1{,}626$) as shown in Supplementary Fig. 1. In the GRID, only 3.7% of the prospective cohort and 5% of the retrospective cohort are classified as PCS1, whereas the PCTA cohort contains 19% of primary tumors classified as PCS1. The likely explanation for this difference is that the GRID and the PCTA contain different proportions of high grade tumors or tumors with distinct metastatic potential. However, these three cohorts include a similar percentage of high risk patients (Gleason score > 7), and PCS1 tumors exhibit the worst prognosis in low grade tumors (Gleason score $\leqq 7$) based on our study [19]. This indicates that the PCS categorization is a variable independent variable of Gleason grade. On the other hand, the proportion of PCS1 (6% in our previous study [19]) and the incidence rate of metastases were similar, with about 3–6% of newly diagnosed PC cases with metastasis [29, 30]. The PCS1 assignment rate can therefore vary according to metastatic potential independent of pathological grade.

It is also well known that dataset composition and choices for reference construction affect subsequent subtype calling. Of note, standard PAM50 classification used in Zhao et al. and the present study is profoundly affected by the composition of the sample cohort used for reference construction. Zhao et al. conducted median centering to the GRID retrospective and prospective cohorts separately prior to applying the PAM50 algorithm. Thus, the percentage of PAM50 categories was assigned different median values from each cohort. In this study, we applied single median

values based on the GRID, combining retrospective and prospective cohorts because the median value approaches the actual population mean as the sample number increases. In future studies, it will be necessary to introduce a method for improving classification robustness by applying appropriate reference construction [31].

PC is one of the malignancies most affected by genetic factors [32], and a range of genomic alterations and structural variations associated with the clinical outcomes have been described [3, 7, 33]. However, the present study is based only on transcriptome data, and did not consider genetic or structural changes associated with the cancers in the individual PCS and PAM50 categories. Thus, our report identified differences and similarities between the two classification schemes but did not make comparisons with other PC subtyping systems. In the future, comparative analysis with other classification methods using multi-omics data [3, 7, 34] will complement the work described here.

In conclusion, PCS and PAM50 present a new lens through which PC may be viewed. There are important similarities in these signatures despite obvious differences in origin and performance (based on existing datasets or samples). Prospective validation of these tools is underway and will help to clarify how they may be most effectively deployed in clinical practice.

Acknowledgements This research was supported by National Cancer Institute (R01-CA220327, R01-CA218356, P01-CA098912, P01-CA233452, U01-CA198900), Department of Defense Prostate Cancer Research Program (PC151088, PC171066, PC180541, PC190482), Steven Spielberg Discovery Fund in Prostate Cancer Research, St. Anthony Prostate Cancer Research Fund, CD McKinnon Memorial Fund for Neuroendocrine Prostate Cancer, Michael & Patricia Berns Family Prostate Cancer Research Fund, Prostate Cancer Foundation Stuart Rahr Young Investigator Award, Prostate Cancer Foundation VALOR Challenge Award.

Compliance with ethical standards

Conflict of interest YL and ED are employees of Decipher Biosciences.

References

1. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010;18:11–22.
2. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, et al. Punctuated evolution of prostate cancer genomes. Cell. 2013;153:666–77.
3. Grasso CS, Wu YM, Robinson DR, Cao X, Dhanasekaran SM, Khan AP, et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature. 2012;487:239–43.
4. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015;161:1215–28.
5. Spratt DE, Dai DLY, Den RB, Troncoso P, Yousefi K, Ross AE, et al. Performance of a prostate cancer genomic classifier in predicting metastasis in men with prostate-specific antigen persistence postprostatectomy. Eur Urol. 2018;74:107–14.
6. van Dessel LF, van Riet J, Smits M, Zhu Y, Hamberg P, van der Heijden MS, et al. The genomic landscape of metastatic castration-resistant prostate cancers reveals multiple distinct genotypes with potential clinical impact. Nat Commun. 2019;10:5251.
7. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. Cell. 2015;163:1011–25.
8. Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera JM, Setlur S, et al. TMPRSS2: ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. Cancer Res. 2006;66:8337–41.
9. Tomlins SA, Laxman B, Dhanasekaran SM, Helgeson BE, Cao X, Morris DS, et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature. 2007;448:595–9.
10. Tomlins SA, Alshalalfa M, Davicioni E, Erho N, Yousefi K, Zhao S, et al. Characterization of 1577 primary prostate cancers reveals novel biological and clinicopathologic insights into molecular subtypes. Eur Urol. 2015;68:555–67.
11. Al Hussein Al Awamlh B, Shoag JE. Genomics and risk stratification in high-risk prostate cancer. Nat Rev Urol. 2019;16:641–2.
12. Fraser M, Rouette A. Prostate cancer genomic subtypes. Adv Exp Med Biol. 2019;1210:87–110.
13. Markert EK, Mizuno H, Vazquez A, Levine AJ. Molecular classification of prostate cancer using curated expression signatures. Proc Natl Acad Sci USA. 2011;108:21276–81.
14. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406:747–52.
15. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 2008;455:1061–8.
16. Howard LE, Zhang J, Fishbane N, Hoedt AM, Klaassen Z, Spratt DE, et al. Validation of a genomic classifier for prediction of metastasis and prostate cancer-specific mortality in African-American men following radical prostatectomy in an equal access healthcare setting. Prostate Cancer Prostatic Dis. 2019;23:419–28.
17. Erho N, Crisan A, Vergara IA, Mitra AP, Ghadessi M, Buerki C, et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. PLoS ONE. 2013;8:e66855.
18. Luca BA, Brewer DS, Edwards DR, Edwards S, Whitaker HC, Merson S, et al. DESNT: a poor prognosis category of human prostate cancer. Eur Urol Focus. 2018;4:842–50.
19. You S, Knudsen BS, Erho N, Alshalalfa M, Takhar M, Al-Deen Ashab H, et al. Integrated classification of prostate cancer reveals

a novel luminal subtype with poor outcome. Cancer Res. 2016;76: 4948–58.
20. Zhao SG, Chang SL, Erho N, Yu M, Lehrer J, Alshalalfa M, et al. Associations of luminal and basal subtyping of prostate cancer with prognosis and response to androgen deprivation therapy. JAMA Oncol. 2017;3:1663–72.
21. You S, Freeman MR. A systems approach to prostate cancer classification-response. Cancer Res. 2017;77:7133–5.
22. Jan YJ, Yoon J, Chen JF, Teng PC, Yao N, Cheng S, et al. A circulating tumor cell-RNA assay for assessment of androgen receptor signaling inhibitor sensitivity in metastatic castration-resistant prostate cancer. Theranostics. 2019;9:2812–26.
23. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol. 2009;27:1160–7.
24. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005;102:15545–50.
25. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1:417–25.
26. He Y, Johnson DT, Yang JS, Wu H, You S, Yoon J, et al. Loss of the tumor suppressor, Tp53, enhances the androgen receptor-mediated oncogenic transformation and tumor development in the mouse prostate. Oncogene. 2019;38:6507–20.
27. Ross AE, Feng FY, Ghadessi M, Erho N, Crisan A, Buerki C, et al. A genomic classifier predicting metastatic disease progression in men with biochemical recurrence after prostatectomy. Prostate Cancer Prostatic Dis. 2014;17:64–69.
28. Feng F, Thomas S, Gormley M, Lopez-Gitlitz A, Yu MK, Cheng S, et al. Identifying molecular determinants of response to apalutamide (APA) in patients (pts) with nonmetastatic castration-resistant prostate cancer (nmCRPC) in the SPARTAN study. Cancer Res 2019;79:13_Supplement.
29. Kelly SP, Anderson WF, Rosenberg PS, Cook MB. Past, current, and future incidence rates and burden of metastatic prostate cancer in the United States. Eur Urol Focus. 2018;4:121–7.
30. Weiner AB, Matulewicz RS, Eggener SE, Schaeffer EM. Increasing incidence of metastatic prostate cancer in the United States (2004–2013). Prostate Cancer Prostatic Dis. 2016; 19:395–7.
31. Cascianelli S, Molineris I, Isella C, Masseroli M, Medico E. Machine learning for RNA sequencing-based intrinsic subtyping of breast cancer. Sci Rep. 2020;10:14071.
32. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer–analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med. 2000;343:78–85.
33. Boysen G, Barbieri CE, Prandi D, Blattner M, Chae SS, Dahija A, et al. SPOP mutation leads to genomic instability in prostate cancer. Elife. 2015;4:e09207.
34. Ross-Adams H, Lamb AD, Dunning MJ, Halim S, Lindberg J, Massie CM, et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: a discovery and validation cohort study. EBioMedicine. 2015;2:1133–44.