

Increasing Stemness Drives Prostate Cancer Progression, Plasticity, Therapy Resistance and Poor Patient Survival

Xiaozhuo Liu^{1§*}, Eduardo Cortes^{2,3§}, Yibing Ji¹, Kathy Zhao⁴, Julian Ho⁴, Yang (Seagle) Liu⁴, Elai Davicioni⁴, Felix Y. Feng^{5†}, Joshi J. Alumkal⁶, Daniel E. Spratt⁷, Christopher J. Sweeney⁸, Han Yu², Qiang Hu², Zou Cheng⁹, Dingxiao Zhang⁹, Gurkamal Chatta¹⁰, Kent L. Nastiuk¹¹, David W. Goodrich¹, Kiera Rycaj¹, Anmbreen Jamroze¹, Jason S. Kirk¹, Igor Puzanov¹⁰, Song Liu^{2*}, Jianmin Wang^{2*}, Dean G. Tang^{1*}

¹Department of Pharmacology and Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA

²Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA

³Department of Biostatistics, University at Buffalo, Buffalo, New York, USA

⁴Veracyte Inc, San Diego, California, USA

⁵Department of Radiation Oncology, University of California, San Francisco, California, USA

⁶Department of Hematology and Oncology, Department of Internal Medicine, Rogel Cancer Center, University of Michigan, Ann Arbor, Michigan, USA

⁷University Hospitals Seidman Cancer Center and Case Western Reserve University, Cleveland, Ohio, USA

⁸South Australian Immunogenomics Cancer Institute, University of Adelaide, Adelaide, SA 5000, Australia

⁹School of Biomedical Sciences, Hunan University, Changsha, China

¹⁰Department of Medicine, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA

¹¹Department of Cancer Genetics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA

[§]These authors contributed equally to this work.

[†]We dedicate this work to the memory of Dr. Felix Y. Feng, whose contributions to this project and the field of prostate cancer research are immeasurable.

*Corresponding authors:

Xiaozhuo Liu, PhD, Department of Pharmacology and Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263 (Phone: 716-645-4988, Xiaozhuo.Liu@Roswellpark.org).

Song Liu, PhD, Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263 (Song.Liu@Roswellpark.org).

Jianmin Wang, PhD, Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263 (Jianmin.Wang@Roswellpark.org).

Dean Tang, MD, PhD, Department of Pharmacology and Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263 (Dean.Tang@Roswellpark.org).

Running title: Increasing Stemness Drives Prostate Cancer Progression.

Key words: Prostate cancer, stemness, AR signaling, aggressiveness, heterogeneity, plasticity.

Word counts: Main Text: 5328 words; Figures: 7; Supplementary Figures: 10; Supplementary Tables: 5.

37 ABSTRACT

38 Cancer progression involves loss of differentiation and acquisition of stem cell-like traits, broadly referred
 39 to as “stemness”. Here, we test whether the level of stemness, assessed by a transcriptome-derived
 40 Stemness score, can quantitatively track prostate cancer (PCa) development, progression, therapy
 41 resistance, metastasis, plasticity, and patient survival. Integrative analysis of transcriptomic data from
 42 87,183 samples across 26 datasets reveals a progressive increase in Stemness and decline in pro-
 43 differentiation androgen receptor activity (AR-A) along the PCa continuum, with metastatic castration-
 44 resistant PCa (mCRPC) exhibiting the highest Stemness and lowest AR-A. Both the general Stemness
 45 score and a newly developed 12-gene “PCa-Stem Signature” correlate with and predict poor clinical
 46 outcomes. Mechanistically, increased AR-A may promote Stemness in early-stage PCa while *MYC*
 47 amplification and bi-allelic *RB1* loss likely drive greatly elevated Stemness in mCRPC where AR-A is
 48 suppressed. Our findings establish Stemness as a robust quantitative measure of PCa aggressiveness
 49 and offer a scalable framework for PCa risk stratification.

INTRODUCTION

Malignant transformation and cancer progression are characterized by progressive loss of differentiation (or oncogenic dedifferentiation) and gain of stem cell traits (i.e., stemness), and poorly differentiated tumors are enriched in gene expression profiles of normal embryonic and adult stem cells and in cancer stem cells^{1,2,3,4,5}. However, the oncogenic dedifferentiation-based stemness concept⁶ that prostate cancer (PCa) represents abnormal organogenesis^{7,8} has not been systemically addressed and there lack quantitative measurements of the overall stemness-associated aggressiveness in PCa.

PCa is projected to claim the lives of >35,000 American men in 2025⁹. PCa incidence rates increased 3% per year from 2014 - 2019 driven by advanced PCa diagnosis and the proportion of men diagnosed with metastases has doubled from 2011 to 2019 when the next generation of androgen receptor (AR) pathway inhibitors (ARPIs) such as abiraterone (Zytiga; FDA approval in 2011) and enzalutamide (Xtandi; FDA approval in 2012) were introduced to the clinic. Significant increases in diagnosis of aggressive and advanced disease and continued high PCa mortality have been linked to therapy resistance driven by intrinsic PCa cell heterogeneity (e.g., in AR expression) and treatment-induced plasticity⁸.

The level of malignancy of primary PCa (Pri-PCa) has been traditionally assessed by pathologists using a combined Gleason Score (GS), which gauges histopathological differentiation pattern of prostatic glands. The higher the GS, the more undifferentiated and more malignant the Pri-PCa is. The GS system, although clinically the 'gold' standard in pinpointing PCa diagnosis, is only semi-quantitative, can be subjective (to individual pathologist's assessment), and is generally not applicable to treatment-failed tumors and distant metastases (which have mostly lost differentiated glandular structures). Notably, whether more advanced Pri-PCa (GS8-10), treatment failure and metastatic dissemination are quantitatively associated with increased cancer stemness remains unknown. On the other hand, AR is the master regulator of normal prostatic epithelium differentiation and represents the therapeutic target of PCa treatments including androgen-deprivation therapy (ADT) and ARPIs. Inevitably, tumors develop resistance to ADT/ARPIs leading to the state of castration-resistant PCa (CRPC), which involves many mechanisms including genetic alterations of AR and key genes in AR pathway¹⁰, lineage plasticity^{8,11,12}, and increased stemness^{8,13}. It is unclear how the global pro-differentiation AR signaling activity changes and evolves during the spectrum of PCa initiation and progression and during development of treatment resistance and metastasis. These discussions raise the need to quantify the AR signaling and stemness and closely track their dynamic evolution and inter-relationship during PCa progression.

Herein, we adapted and developed transcriptome-based gene signature scores to quantitatively assess the pro-differentiation (canonical) AR activity (AR-A) and oncogenic stemness levels (Stemness Index or

85 Score) and their inter-relationship in the PCa spectrum. Specifically, we used gene expression data
 86 collected from 87,183 samples in 26 distinct clinical cohorts encompassing the PCa continuum, i.e., from
 87 normal/benign prostate to early stage, advanced and metastatic PCa, to pre- and post-therapy PCa, and,
 88 finally, mCRPC, to track the dynamic evolution of and inter-relationship between AR-A and the Stemness
 89 Index. The results indicate that increased Stemness quantitatively tracks PCa aggressiveness, plasticity,
 90 treatment resistance and progression, and represents a poor prognostic factor for patient outcomes. Our
 91 study offers a trajectory-integrated, pan-cohort framework that defines and validates cancer Stemness as
 92 a quantifiable clinical determinant of PCa aggressiveness, plasticity, progression and therapy resistance.

RESULTS

Adapting and developing Stemness and AR-A metrics for PCa analysis.

To determine the global and dynamic changes of and inter-relationship between degree of malignancy (oncogenic dedifferentiation) and level of (luminal) differentiation during PCa development and progression (Supplementary Fig. S1A-B), we employed two transcriptome-based algorithms to annotate 26 datasets of 3,102 bulk RNA-seq and 84,081 microarray samples encompassing the PCa continuum (Supplementary Fig. S1B; Supplementary Table S1). In our studies, we focused on and compared three progressive stages: from normal prostate to localized Pri-PCa, from low-grade to high-grade Pri-PCa, and from treatment-naïve Pri-PCa to CRPC/mCRPC (Supplementary Fig. S1A). We defined “progression” as stages beyond Pri-PCa and as disease entities that are more aggressive in a comparative manner. Specifically, we adapted the Stemness Index⁶ (mRNAsi; **Stemness** for short) (Supplementary Fig. S1C) to gauge the degree of malignancy (or aggressiveness) and employed the Z-score normalized canonical (pro-differentiation) AR activity or **AR-A** (Supplementary Fig. S1D) to measure the level of differentiation. Canonical AR-A can be determined by quantifying AR-regulated transcripts under intact androgen/AR signaling conditions, and there are several such AR-regulated gene lists^{14, 15, 16, 17, 18} (Supplementary Fig. S2A-B). We validated the AR-A list based on their enrichment in prostatic epithelium differentiation, correlation of individual gene expression with AR-A signatures in clinical datasets, and overall consistency in correlation among the AR-A scores (Supplementary Fig. S2B-D). Given that Bluemn’s AR-A score¹⁴ correlated well with other scores (Supplementary Fig. S2E-G) and the 10 genes in this signature consistently represented AR targets in multiple clinical datasets (Supplementary Fig. S2D; Supplementary Table S2), we decided to use this AR-regulated gene list for our AR-A quantification.

Normal prostate epithelium shows a negative correlation between Stemness and AR-A.

The normal human prostate (NHP) comprises 2 major epithelial types: basal and luminal cells (Supplementary Fig. S3A). We annotated two studies that profiled FACS-purified NHP luminal and basal cell populations^{19, 20} and observed that the luminal cell population (Trop2⁺CD49f^{lo}) possessed higher AR mRNA levels and AR-A but lower Stemness compared to basal cells (Trop2⁺CD49f^{hi}) and, in both datasets, AR-A negatively correlated with Stemness (Supplementary Fig. S3B-C). These results support that the NHP basal cell compartment harbors primitive stem cells^{7, 19, 20, 21, 22} and AR-A measures the level of prostate epithelial differentiation.

Pri-PCa exhibits concordantly increased Stemness and AR-A, and neoadjuvant ADT (nADT) decreases both Stemness and AR-A.

In analyzing AR-A and Stemness in matched pairs of prostate tumor (T) and normal (N; tumor-adjacent benign) tissues in the TCGA-PRAD¹⁸ (n=51 pairs), Wyatt 2014²³ (n=12 pairs), and Long 2020²⁴ (n=10 pairs)

cohorts, we observed concordantly increased AR-A and Stemness (Fig. 1A-C). Accompanying elevated AR-A, AR mRNA levels were also increased in Pri-PCa compared to benign tissue (Fig. 1A-C; note that the increase in AR mRNA levels in the Long 2020 cohort approached statistical significance). The paired T/N comparisons (Fig. 1A-C) suggest that early PCa initiation and growth are characterized by concordantly increased Stemness and AR-A. To test whether AR-A might causally drive increased Stemness early during prostate tumorigenesis, we investigated 4 PCa nADT (i.e., ~2-6 months of ADT before prostatectomy or radiation) cohorts^{24, 25, 26}. In the first 3 cohorts with matched pairs of pre- and post-nADT samples (from the same patients), nADT concordantly downregulated AR-A (without decreasing AR mRNA levels) as well as the Stemness (Fig. 1D-F). In the Nastiuk cohort with non-paired post-nADT and no-nADT samples (i.e., from different patients but matched for ages and tumor grades and stages), nADT increased AR mRNA levels (Supplementary Fig. S4A) but reduced both AR-A and Stemness (Supplementary Fig. S4B-C).

Pri-PCa progression is accompanied by reduced AR-A but continually increasing Stemness.

We next assessed the dynamic changes of Stemness and AR-A during *spontaneous* (i.e., therapy-free) Pri-PCa progression (Fig. 2). To this end, we sorted out treatment-naïve tumors in TCGA-PRAD by excluding samples from patients who received adjuvant (post-surgery) hormone therapy (i.e., Pri-PCa/ADT or Tx; n=70) (Fig. 2A) and grouped the Pri-PCa samples according to increasing GS (Fig. 2B-D). When compared with the 52 benign (N) samples, all the four GS groups from Pri-PCa (GS6, GS7, GS8, and GS9/10 combined) showed higher AR-A and Stemness (Fig. 2C-D), consistent with the results in pairwise T/N comparisons (Fig. 1A-C). However, when compared to GS6 PCa, the three higher GS groups showed increasing AR mRNA levels (Fig. 2B) but decreasing AR-A (Fig. 2C) and increasing Stemness (Fig. 2D). Jonckheere-Terpstra's (J-T) trend test validated a gradual reduction in canonical (pro-differentiation) AR signaling but an increasing trend of Stemness with Pri-PCa progression (Fig. 2C-D). Notably, although we observed a modest positive correlation between AR-A and Stemness in pooled Pri-PCa samples (Pearson's $r=0.27$; Fig. 2E), this positive correlation gradually diminished with PCa progression when we categorized the samples by tumor grade (Fig. 2F). The linear regression analysis confirmed that the interaction between Stemness and AR-A significantly depended on tumor grade ($P=0.0003$), with a linear decrease in the slope of the correlation as GS increased (Fig. 2G).

Collectively, these results indicate that, in contrast to early tumorigenesis, *spontaneous PCa progression* (i.e., with increasing GS and tumor grade) is marked by a divergence between oncogenic Stemness and pro-differentiation AR-A.

Therapies and metastatic progression drive further divergence between attenuated AR-A and increased Stemness.

Next, we asked how clinical treatment and metastatic progression may impact the Stemness and AR-A by querying untreated Pri-PCa and Pri-PCa/ADT from TCGA and mCRPC in SU2C¹⁰ (Fig. 3A-C). AR-A was elevated in Pri-PCa compared to adjacent benign tissue (N) but reduced in more aggressive Pri-PCa/ADT samples, while the Stemness, in contrast, continually and persistently increased from N → Pri-PCa → Pri-PCa/ADT (Fig. 3B-C). Strikingly, mCRPC, despite markedly upregulating AR mRNA (Fig. 3A), showed the lowest AR-A but highest Stemness (Fig. 3B-C), indicating a shift from a positive correlation between AR-A and Stemness in Pri-PCa to a negative association in mCRPC (Fig. 3D).

By employing normalized RNAseq data from the PCa transcriptome atlas (Bolis 2021), which included normal prostate specimens (n = 174), Pri-PCa (n = 714) and mCRPC (n = 335)²⁷, we tracked the Stemness and AR-A across the PCa spectrum (Fig. 3E). AR-A increased from benign to Pri-PCa but declined in Pri-PCa/ADT and continued to decrease in mCRPC, reaching its lowest in mCRPC-NE²⁸ as supported by J-T test (Fig. 3E). In contrast, the Stemness exhibited a continually increasing trajectory across the PCa spectrum encompassing Normal/Benign → treatment-naïve Pri-PCa → Pri-PCa/ADT → (m)CRPC ($P < 0.0001$, J-T Trend test; Fig. 3F; Supplementary Fig. S5A). Notably, the 3 subtypes of mCRPC, i.e., AR-positive (ARPC), CRPC-NEPC and AR⁻/NE⁻ double-negative PCa (DNPC) all showed higher Stemness than Pri-PCa and also displayed increasing Stemness among the 3 subtypes (Supplementary Fig. S5B). Unexpectedly, tumor-adjacent benign prostate tissues showed much higher Stemness than normal prostate in the Genotype-Tissue Expression Project (GTEx)²⁹ (Fig. 3F; Supplementary Fig. S5A), which were all obtained from cancer-free organ donors. This surprising finding suggests that the transcriptomes of adjacent benign tissues have apparently skewed towards those of cancer samples, consistent with previous observations that adjacent benign tissues represent an intermediate state between healthy and tumor tissues³⁰. Also of interest, the Cancer Cell Line Encyclopedia (CCLE)^{31, 32} cell lines including PCa cell lines, and PCa-derived PDX³³ and xenografts³⁴ exhibited the highest Stemness (Fig. 3F; Supplementary Fig. S5A), supporting that cancer cell lines and xenografts are clonally derived from and highly enriched in cancer stem cells (CSCs).

As in our early dissection of the relationship between AR-A and Stemness during spontaneous Pri-PCa progression (Fig. 2F-G), analysis of the Bolis 2021 PCa transcriptome atlas²⁷ revealed a similar change from a positive correlation between AR-A and Stemness in Pri-PCa (blue line; Pearson's $r = 0.32$, $P < 0.0001$; Fig. 3G) to a negative correlation between the two in mCRPC (green line; Pearson's $r = -0.15$, $P = 0.005$; Fig. 3G). We further extended our Stemness and AR-A interrogation to include microarray-derived transcriptomic datasets^{35, 36}, allowing us to access data from PCa cohorts generated before the Next-generation sequencing (NGS) era. Analyzing the GRID database of over 82,000 prospectively collected biopsy samples of localized PCa from clinical use of the Veracyte Decipher test³⁵, we systematically quantified the Stemness across the NCCN Risk Groups and observed increasing Stemness along the NCCN Risk trajectory (Fig. 3H). Likewise, in the Taylor dataset³⁶, we observed that AR-A was the highest in Pri-PCa but significantly decreased in metastatic PCa (mPCa) (Fig. 3I) while the Stemness

increased progressively from benign to Pri-PCa and reached the highest levels in mPCa (Fig. 3J). Linear regression analysis confirmed a gradual loss of positive correlation between AR-A and Stemness during PCa progression, as illustrated by distinct regression lines for each stage (Fig. 3K-M).

Finally, we analyzed the dynamic changes of Stemness and AR in genetically engineered mouse models³⁷ (GEMMs) of PCa with increasing aggressiveness due to individual or combined deletion of 3 tumor suppressor genes, *Pten*, *Rb1* and *Trp53*. Briefly, *Pten*^{-/-} single knockout (SKO) prostate tumors develop around 9 weeks, and mice rarely develop metastasis with a median lifespan of 48 weeks. In contrast, the *Pten*^{-/-};*Rb1*^{-/-} double knockout (DKO) mice develop highly metastatic PCa that shortens median survival to ~38 weeks. When *Trp53* is further deleted in DKO background, the triple KO (TKO; *Pten*^{-/-};*Rb1*^{-/-};*Trp53*^{-/-}) tumors are exclusively AR⁻ NEPC and castration-resistant *de novo* with high metastatic rate and lifespan of ~16 weeks³⁷. We found that the AR-A was the highest in SKO tumors but much reduced in aggressive DKO and TKO tumors (Supplementary Fig. S5C). In contrast, Stemness was the lowest in SKO but significantly increased in DKO and TKO (Supplementary Fig. S5D). Correlation analyses revealed a positive correlation between AR-A and Stemness in SKO that turned negative in both DKO and TKO tumors (Supplementary Fig. S5E-H).

The above results, collectively, indicate that *hormone therapy and metastatic progression further drive marked increases in Stemness with concomitant decreases in pro-differentiation AR-A*.

Stemness is prognostic of poor survival in PCa patients.

We investigated global Stemness and its prognostic significance, and observed that high Stemness consistently correlated with worse patient survival across various cohorts including the Spratt 2017³⁸ ($P=0.001$, log-rank; Fig. 4A), Tosoian 2020³⁹ ($P=0.014$, log-rank; Fig. 4B), CHAARTED 2021⁴⁰ ($P=0.006$, log-rank; Fig. 4C; $P=0.149$, log-rank; Supplementary Fig. S6A), and the mCRPC (SU2C 2019⁴¹; $P=0.06$, log-rank; Supplementary Fig. S6B-C) as well as Taylor 2010³⁶ ($P=0.028$, log-rank; Supplementary Fig. S6D-E). Notably, the CHAARTED cohorts displayed the highest mean Stemness compared to Tosoian and Spratt cohorts, and the CHAARTED cohort showed the highest median Decipher Prostate Genomic Classifier^{38, 39, 40} among the three cohorts (Fig. 4D). Interestingly, like high Stemness, low AR-A was correlated with poor patient survival in the Taylor dataset ($P=0.040$, log-rank; Supplementary Fig. S6F).

Stemness-high PCa are associated with aggressive molecular subtypes and a 12-gene PCa-Stem signature prognosticates poor patient survival.

To identify key determinants of Stemness and explore differences between tumors with high versus low Stemness, we stratified Pri-PCa (TCGA-PRAD) and mCRPC (SU2C 2019) samples separately based on Stemness scores (Fig. 5A). Differential gene expression (DEG) analyses revealed significant differences between the top 33% (Stemness-high) and the bottom 33% (Stemness-low) PCa samples in both

treatment-naïve Pri-PCa (Fig. 5B) and treatment-failed mCRPC (Fig. 5C), resulting in a 12-gene “PCa-Stem signature” (Fig. 5D; Supplementary Table S2; see Methods).

Gene set enrichment analysis (GSEA) was performed to delineate the molecular differences between Stemness-low and Stemness-high PCa samples (Fig. 5E-F). GSEA revealed that Stemness-high PCa samples were enriched in gene sets associated with embryonic stem cells (ESC), as well as aggressive cancer phenotypes, such as undifferentiated cancer and metastasis, in both Pri-PCa and mCRPC (Fig. 5F). In contrast, Stemness-low PCa samples were enriched in gene signatures linked to lower aggressiveness and inflammatory pathways (Fig. 5E). Additionally, hallmark pathways such as E2F targets, MYC targets, mTORC1 signaling, and DNA repair were upregulated in Stemness-high PCa (Fig. 5F), while TGF- β signaling, TNF- α signaling, and IFN- γ responses were prominent in Stemness-low PCa (Fig. 5E).

The aggressive PAM50-LumB molecular subtype^{42, 43, 44} was significantly more prevalent in Stemness-high PCa, accounting for 74% in Pri-PCa (compared to 15% in Stemness-low PCa; Fig. 5G) and 43% in mCRPC (compared to 4% in Stemness-low PCa; Fig. 5H). The PCS1 subtype^{45, 46}, known for its association with ADT resistance and the highest risk of progression to advanced disease in comparison with PCS2 or PCS3, was predominantly enriched in both Stemness-high Pri-PCa and mCRPC (Fig. 5I-L). Notably, the PAM50-Basal subtype^{42, 43} increased in Stemness-high mCRPC but not in Pri-PCa (Fig. 5G-H), aligning with the characteristics of mCRPC, which involves lineage plasticity and acquisition of basal, mesenchymal, neural and stem-like phenotypes when developing ARPI resistance^{7, 44, 47}. Consequently, Stemness-high samples remained largely AR-dependent⁴⁸ in Pri-PCa (Fig. 5M) but transitioned to AR-independent NE subtype⁴⁴ in mCRPC (Fig. 5N). Stemness-high PCa was also characterized by high proliferation as evidenced by their enrichment in a 31-gene cell-cycle progression signature⁴⁴ (Fig. 5O-P). Finally, we examined the association between Stemness and a signature of lineage plasticity risk after Enza treatment for mCRPC¹² and found that Stemness-high tumors may be at risk for this virulent form of treatment resistance (Fig. 5Q-R).

To further assess the clinical relevance of Stemness, we analyzed the prognostic impact of our newly developed 12-gene PCa-Stem signature (Fig. 5D). We observed that higher PCa-Stem signature correlated with worse patient survival in both Pri-PCa ($P<0.0001$, log-rank; Fig. 5S-T) and mCRPC ($P<0.0001$, log-rank; Fig. 5U-V) patients. Multivariate analysis, adjusted for clinicopathologic parameters such as age, GS and stage, showed that patients with high PCa-Stem signature had significantly worse independent prognosis compared to those with low PCa-Stem signature. This was demonstrated by worse progression-free survival (PFS) in Pri-PCa (multivariable-adjusted HR=1.636 (1.014 – 2.64), $P=0.0438$; Fig. 5T) and overall survival (OS) in mCRPC (multivariable-adjusted HR=4.712 (2.011 – 11.04), $P<0.001$; Fig. 5V), respectively. Furthermore, the PCa-Stem signature steadily and continually increased across the PCa evolutionary spectrum and correlated with PCa disease progression scores as supported by trajectory inference analysis of the Bolis PCa transcriptome atlas²⁷ (Pearson’s $r=0.72$; Supplementary Fig. S7A-B). Strikingly, gene expression analysis revealed that 11 of the 12 genes (except *KLK12*) in the PCa-Stem

signature significantly correlated with the PCa disease progression ([Supplementary Fig. S7C](#)). In contrast, among the genes commonly downregulated in Stemness-high PCa samples in both Pri-PCa and mCRPC ([Supplementary Table S3](#)), some (e.g., *PTGDS*, *SPARCL1*, *CLU*, *GJA1* and *S100A4*) were involved in regulating prostate-specific glandular structures and luminal functions while others (e.g., *FBLN1*, *SFRP1*, *IGFBP6*, *GAS1* and *TIMP2*) played general roles in epithelial differentiation ([Supplementary Fig. S7D](#)).

Overall, these findings indicate that *PCa with high Stemness exhibit greater aggressiveness, a higher likelihood of developing lineage plasticity and therapy resistance, and poor survival*.

Stemness-high PCa have high tumor mutation burden (TMB) and unique genomic features.

We determined the potential differences in genomic features between Stemness-high vs. Stemness-low PCa and found that mCRPC had a higher fraction of altered genome ([Fig. 6A](#)) and higher TMB ([Fig. 6B](#)) than Pri-PCa. In both Pri-PCa and mCRPC, Stemness-high PCa displayed a higher fraction of altered genome ([Fig. 6A](#)) and TMB ([Fig. 6B](#)) than Stemness-low PCa although there was no difference in patient ages ([Fig. 6C](#)). In Pri-PCa, Stemness-high samples had a higher prevalence of *PTEN* deletion (DEL), *SPOP* and *FOXA1* mutations (MUT) and *MYC* amplification (AMP) whereas in mCRPC, Stemness-high PCa had more prevalent *RB1* DEL and *AR* MUT ([Fig. 6D-F](#)).

Globally, both Pri-PCa and mCRPC exhibited mutations in more than a dozen genes although most of these alterations were <10% ([Supplementary Fig. S8A](#); green symbols), consistent with reports that PCa is genomically heterogeneous with a broad spectrum of genomic alternations of low penetrance^{49, 50}. On the other hand, homozygous deletion (HOMDEL) events, particularly, loss of *PTEN* and *RB1*, occurred frequently in both Pri-PCa and mCRPC ([Supplementary Fig. S8A](#); blue symbols) whereas AMP events in oncogenes such as *MYC* were more prevalent in mCRPC ([Supplementary Fig. S8A](#); red symbols). Nevertheless, *PTEN* and *RB1* DEL represented among the most prevalent alterations in early-stage (GS6) PCa ([Supplementary Fig. S8B-C](#)). Analysis of genetic alterations across the spectrum of PCa progression revealed that among the 3 tumor suppressors (*RB1*, *PTEN* and *TP53*), while homozygous *RB1* DEL remained relatively constant at 13%, 6%, 10% and 10% in low-grade (GS6), high-grade (GS9/10), treated Pri-PCa and mCRPC, respectively, *PTEN* HOMDEL (9%, 23%, 30% and 26%) and *TP53* MUT (0%, 21%, 27% and 37%) continued to increase along this progression ([Supplementary Fig. S8B-C](#)). Among the 3 oncogenic events analyzed, while high levels of *TPR22* fusions (~40%) remained unchanged from low-grade to high-grade Pri-PCa to mCRPC, *MYC* (2%, 12%, 19% and 24%) and *AR* (0%, 1%, 6% and 49%) AMP continued to increase along the progression trajectory ([Supplementary Fig. S8B-C](#)). In fact, prevalent *AR* MUT (11%) were observed only in the most aggressive and Stemness-highest mCRPC ([Supplementary Fig. S8B-C](#)), suggesting potentially high noncanonical AR-A in mCRPC. Analyzing these major genomic alterations in association with PCa Stemness, we observed that *RB1* DEL, *FOXA1* (together with *SPOP*, *IDH1*, and *CHD1*) MUT, and especially *MYC* AMP were significantly associated with

increased Stemness in Pri-PCa ([Supplementary Fig. S8D](#)). In mCRPC, only *RB1* DEL exhibited a statistically significant association with higher Stemness ([Supplementary Fig. S8E](#)).

MYC further drives Stemness during spontaneous PCa progression.

AR AMP and *MUT* are exclusively treatment-induced as there were virtually no *AR* genomic alterations in treatment-naïve Pri-PCa ([Supplementary Fig. S8B-C](#)). We showed that early during PCa development increased *AR*-A may drive elevated Stemness ([Fig. 1D-F](#)); however, treatment-naïve Pri-PCa displayed continually increasing Stemness but decreasing *AR*-A ([Fig. 2](#)). What could be driving persistently high Stemness during Pri-PCa progression with decreasing *AR*-A and in the absence of *AR* genomic alterations? We found that Stemness-high Pri-PCa, compared to Stemness-low Pri-PCa, were enriched in MSigDB Hallmark *MYC*_TARGETS (*MYC*_Targets_V1 and *MYC*_Targets_V2; [Fig. 5F](#)) and had more prevalent *MYC* AMP ([Fig. 6D](#) and [6F](#)). High-grade (GS9/10) Pri-PCa also had more *MYC* AMP events than low-grade (GS6) Pri-PCa (12% vs. 2%; $P=0.006$, χ^2 test; [Supplementary Fig. S8C](#)). Importantly, *MYC* AMP was most prominently associated with increased Stemness in Pri-PCa ([Supplementary Fig. S8D](#)). These observations led us to hypothesize that *MYC* may represent a critical Stemness driver during spontaneous Pri-PCa progression (in absence of therapeutic pressure). To test this hypothesis, we quantified *MYC* signaling activity⁵¹ (*MYC*-sig; See Methods and [Supplementary Table S2](#) for gene signature information) across the PCa progression spectrum. We observed that: 1) both *MYC* mRNA and *MYC*-sig increased in Pri-PCa compared to matched benign samples ([Fig. 7A-B](#)); 2) *MYC*-sig continued to go up during Pri-PCa progression ([Fig. 7C](#)); 3) unlike nADT-induced concomitant decline in both *AR*-A and Stemness ([Fig. 1D-F](#); [Supplementary Fig. S4](#)), nADT showed no discernible effect on *MYC*-sig ([Supplementary Fig. S9A-D](#)); and 4) *MYC*-sig continued to increase across the disease spectrum with mCRPC exhibiting the highest *MYC* activity ([Fig. 7D-E](#)) and with *MYC*-sig positively correlated with the Stemness ([Fig. 7F-H](#)).

These results, together, support MYC as a Stemness driver during PCa progression.

DISCUSSION

The current project was undertaken to quantitatively co-analyze the global changes of and the interrelationship between cancer stemness and pro-differentiation AR signaling activity during PCa evolution. By employing gene expression-based general Stemness and AR-A signature scores and through developing a new 12-gene PCa-Stem signature, we show that while the differentiation-regulating AR-A steadily declines in the PCa continuum, pervasively increasing Stemness quantitatively measures PCa aggressiveness, plasticity and progression, and represents a poor prognosticator for patient outcomes. We report that Stemness-high PCa possess unique genomic features with high TMBs and are associated with aggressive molecular subtypes such as PAM50-LumB⁴²⁻⁴⁴ and PCS1^{45,46}. We further present evidence that elevated Stemness in early PCa might be driven by increased AR-A but increased MYC-sig activity and other genomic alterations likely propagate the exponentially increased Stemness in mCRPC where AR-A is suppressed.

Early studies reported that a cohort of AR-regulated genes declines in high-grade versus low-grade PCa^{52, 53} and that loss of AR expression or inhibition of AR-A promotes a stem-like phenotype^{8, 34, 54, 55}. However, the global pro-differentiation AR-A and cancer stemness have not been systematically and quantitatively assessed and co-analyzed in PCa, and it's unclear whether PCa progression is accompanied by increased stemness as in some other cancers such as glioblastoma and breast cancer⁶. Starting with systematic, comprehensive and intentionally “progression timeline-informed” collection of clinical datasets, here we integrate 26 transcriptomic datasets (covering 87,183 samples) encompassing normal prostate, benign, primary tumors, ADT-treated tumors, CRPC subtypes, metastases, and model systems. This enables a “progression-conscious” and clinically anchored stratification framework with clearly annotated clinical information across these 26 cohorts. Our pan-cohort approach enables signature validation in rare archival biopsy samples and cross-platform integration — a translational advantage not addressed by other studies. With these novel and unique dataset collection and curation and data analysis pipelines, our study demonstrates that PCa progression is accompanied by a continual increase in Stemness (score) but loss of canonical AR-A, with the two inversely correlated with each other ([Supplementary Fig. S10](#)). Importantly, the Stemness tracks treatment-induced plasticity and PCa aggressiveness, and both high global Stemness and high PCa-Stem signature prognosticate poor patient survival. Consequently, the stemness, a previously qualitative and somewhat ambiguous term that now permeates oncology field and cancer research publications, can be quantitatively measured and, in the case of PCa, the Stemness score quantitatively and positively tracks the entire spectrum of PCa development, lineage plasticity, therapy resistance and metastatic progression. To our knowledge, this is also the first comprehensive study to simultaneously quantify the oncogenic stemness and pro-differentiation AR-A across the PCa continuum.

The normal prostatic acini and ducts are lined by AR⁺ secretory luminal cells and AR⁻ basal cells together with rare neuroendocrine cells. We find that the NHP luminal cells have higher AR-A compared to basal cells, confirming AR-A as a metric for quantifying canonical AR activity and prostatic epithelial differentiation. On the other hand, basal cells have higher Stemness, consistent with earlier findings from our lab¹⁹ and others^{20, 21, 22, 56} showing the basal cell compartment harbors stem cells. Oncogenic transformation is known to confer stem-like features in the initiated cells. Indeed, compared to adjacent benign tissue, Pri-PCa exhibit elevated Stemness, which is associated with and likely driven by increased AR-A. Loss of *PTEN* and *RB1* (9% and 13%, respectively, in GS6 tumors) could also potentially contribute to increased Stemness in early prostate tumors.

During our studies, we made a surprising finding that tumor-adjacent benign tissues, the so-called ‘normal’ tissues used in most comparative studies, actually display much higher Stemness than the true normal prostate obtained from cancer-free organ donors, suggesting that these benign tissues are not normal and their transcriptomes have apparently skewed towards those of cancer samples. This important new revelation prompted us to use, in relevant subsequent studies, the GTEx-derived normal prostate tissue as the baseline comparator, which obviously strengthens the normalization standard across cohorts, improves cross-stage comparison and interpretation, and enhances the scientific rigor of our data.

In treatment-naïve Pri-PCa, tumor progression is accompanied by decreasing AR-A but persistently high Stemness (Supplementary Fig. S10). The decreasing AR-A is driven by progression-related loss of differentiation while increasing Stemness is likely driven by *MYC* amplification and increased *MYC* activity with contributions from other oncogenic events such as loss of *PTEN* and *RB1*. Long-term ADT/ARPI treatment, as expected, further attenuates AR-A but substantially upregulates the Stemness such that mCRPC have the highest Stemness, which is probably driven by continually increasing genomic alterations such as *MYC* amplification and loss of tumor suppressors. Notably, as *AR* alterations (AMP/MUT) are exclusively induced by ARPIs, noncanonical AR-A (i.e., castration-resistant AR activity) likely make significant contributions to the increased Stemness in mCRPC.

What biological correlates do our Stemness metrics measure? *First*, it measures transcriptional features and relative abundance of stem cells and CSCs. Therefore, the NHP basal/stem cells display higher Stemness than luminal cells (consistent with literature^{19, 20, 21, 22}); ARPI treatment, while shutting down AR signaling, enriches CSCs^{7, 34, 54, 55} and upregulates Stemness (this study); and PCa cell lines, xenografts and PDX, known to be clonally derived from cancer stem/progenitor cells, demonstrate the highest Stemness. *Second*, it measures cellular and lineage plasticity induced by treatment and genetic alterations. Thus, mCRPC, which harbor more abundant treatment-reprogrammed AR^{-lo} cells and PCa cells with stem-like features or in an AR-indifferent state^{8, 12, 13, 14, 16, 33, 47, 48}, manifests the highest Stemness. Our analysis

further substantiates the aggressive nature of (AR^{-lo}) CRPC-NEPC and DNPC compared to ARPC, as evidenced by their elevated PCa-Stem signature scores. This suggests a shift in the phenotypic landscape of mCRPC towards more stem-like and aggressive forms, particularly under the pressure of advanced therapies. Likewise, aggressive DKO and TKO murine prostate tumors have undergone lineage transformation and also exhibit higher Stemness than indolent SKO tumors. *Third*, the Stemness metric likely reports the high proliferative status of advanced PCa and mCRPC, as Stemness-high tumors in both Pri-PCa and mCRPC are enriched in “Cell-cycle progression” signature and most genes in the 12-gene PCa-Stem Signature are involved in mitosis and cell-cycle progression. *Finally*, the Stemness score measures PCa aggressiveness as high Stemness correlates not only with more advanced and aggressive stages of PCa but also with aggressive PAM50-LumB and PCS1 molecular subtypes.

Of clinical significance, both the 12-gene PCa-Stem signature and the global Stemness score correlate with poor patient survival, suggesting the possibility of developing the Stemness metric into a predictive biomarker to distinguish aggressive from indolent primary tumors and a prognostic indicator of unfavorable clinical outcomes in mCRPC patients. Notably, 11 of the 12 genes (except *KLK12*) in the newly developed PCa-Stem signature are significantly correlated with PCa progression ([Supplementary Fig. S7C](#)) implicating their individual contributions to the increased Stemness in advanced Pri-PCa and mCRPC. Analysis using data from The Human Protein Atlas (www.proteinatlas.org) reveals that these 11 genes are also recognized as unfavorable prognostic markers in PCa and several other cancers (Liu X and Tang, unpublished), raising the potential utility of the PCa-Stem signature as a prognostic marker not only for PCa but also for other malignancies. Coupled with earlier studies linking increased Stemness with aggressive breast and brain cancer subtypes⁶, our findings herein underscore the increased Stemness as a ‘universal’ and fundamental characteristic of cancer progression and aggressiveness.

In recent years, several studies^{27,57-60} have proposed stemness-based classifications or prognostic models using either single-cell RNA-seq (scRNA-seq), bulk RNA-seq or integrative analyses. These studies reinforce the biological relevance of stemness and highlight oncogenic pathways such as MYC activation, *RB1/PTEN* loss, and lineage plasticity in PCa progression. However, to our knowledge, none has established a quantitative, clinically validated stemness metric across a pan-cohort framework as comprehensive as ours. Of note, our 12-gene PCa-Stem signature, derived from bulk transcriptomic stratification, shows strong biological overlap with the lineage plasticity signature (LPSig) reported by Zhao *et al.* using scRNA-seq⁵⁹. Among the shared genes, *HMMR* emerges as a critical node of convergence as it has been shown that *HMMR* promotes NE differentiation and aggressive behavior in CRPC models⁵⁹. The presence of *HMMR*, along with *DLGAP5*, *AURKB*, and *BIRC5* in our PCa-Stem signature, highlights the translational potential of this gene set in targeting NEPC and other plasticity-driven CRPC subtypes. Supporting this, Zheng *et al.* used multi-omics integration and machine learning to define three stemness-

based PCa subtypes and constructed a nine-gene prognostic signature⁵⁷ that also included *HMMR* and *DLGAP5*. *KLK12*, another PCa-Stem gene, was recently identified as a unique marker for SOX9^{High} stem-like luminal epithelial cells enriched following androgen deprivation⁶¹. Additionally, our findings that *MYC* amplification, *RB1* deletion, and *PTEN* loss are enriched in Stemness-high tumors are consistent with the previously reported chromatin accessibility (ATAC-seq) patterns⁴⁸ and multi-omics spatial data⁶¹. It is worth noting that although our PCa-Stem signature was initially designed for risk stratification in mCRPC, the analytical framework is inherently flexible as the same platform can be extended to derive subtype-specific signatures (e.g., PCa-Stem-v2) tailored for distinct clinical phenotypes such as DNPC, ARPC, or NEPC. This scalability provides a foundation for personalized prognostication and therapy selection in PCa.

We recognize that the Stemness algorithm utilized in our studies was not developed specifically for PCa and therefore, its sensitivity to distinguish subtle intra-group differences remains to be fully validated. The bulk RNA-seq data used to derive Stemness and AR-A was not from pure epithelial cancer cells, a technical issue that can be potentially mitigated by scRNA-seq analysis and other Stemness pipelines such as the CytoTRACE^{62, 63} and by using cancer-specific genomic features derived, for example, from cell-free DNA⁶⁴. Finally, it will be important to determine whether and how noncanonical (castration-associated) AR signaling (driven by *AR AMP/MUT*) may contribute to the highest Stemness in mCRPC.

In conclusion, we have developed transcriptome-based gene signature scores to assess pro-differentiation AR activity (AR-A), MYC activity (MYC-sig) and oncogenic stem-cell features (general Stemness score and the 12-gene PCa-Stem signature). These scores enable us to globally profile differentiation and aggressiveness status across the PCa progression spectrum. This work advances our understanding of dynamic evolution of AR-A and cancer stemness during PCa progression and therapy resistance, and the signature scores reported herein have the potential to serve as clinical classifiers and aid in the identification of patients who may benefit from stemness-targeting therapies.

METHODS

PCa clinical datasets and data collection

In this study, we utilized a total of 26 published and in-house prostate cancer (PCa) related datasets, including various cohorts encompassing the evolutionary spectrum of PCa (Supplementary Fig. S1A-B): normal prostate, tumor-adjacent benign tissue, treatment-naïve primary PCa (Pri-PCa) with increasing GS, PCa treated with neo-adjuvant ADT (nADT) or long-term ADT, metastatic castration-resistant PCa (mCRPC), and PCa cell lines, PCa patient-derived xenografts (PDX) and xenografts. Clinical information, genomic data and gene expression profiling data were downloaded and detailed information on each dataset is listed in Supplementary Table S1. Dataset descriptions include dataset source references, PubMed IDs (PMIDs; RRID:SCR_004846), DOI links, sample type and source (e.g., normal prostate tissue, adjacent benign prostate tissue or primary culture from PCa patient, primary or metastatic prostate tumor, xenografts, cell lines), sample sizes, data types used in the current study (clinicopathologic data, RNAseq, microarray, whole exome sequencing [WES]), and public data access portals.

Data Access and Repository Information: Data for this study were retrieved from the following public repositories and databases (persistent identifiers where applicable):

1. GTEx Portal (RRID:SCR_013042; <https://gtexportal.org/>)
2. Gene Expression Omnibus (GEO; RRID:SCR_005012; <https://www.ncbi.nlm.nih.gov/geo/>)
3. cBioPortal for Cancer Genomics (RRID:SCR_014555; <https://www.cbioportal.org/>)
4. Zenodo Repository (RRID:SCR_004129; <https://zenodo.org/>)
5. Decipher GRID Database (RRID:SCR_006552; ClinicalTrials.gov identifier: NCT02609269)
6. European Genome-phenome Archive (EGA; RRID:SCR_004944; <https://ega-archive.org/>)
7. European Nucleotide Archive (ENA; RRID:SCR_006515; <https://www.ebi.ac.uk/ena/browser/home>)
8. UCSC Xena Functional Genomics Portal (RRID:SCR_018938; <https://xenabrowser.net/>)

Accession numbers and portal links for the 26 datasets described below are summarized in Supplementary Table S1.

The 26 datasets are briefly described below:

- 1) **GTEx 2013** (PMID: 23715323)²⁹: The Genotype-Tissue Expression (GTEx) Project repository collected high-throughput and clinical data of normal tissue from many organs. RNAseq data of 245 normal prostate tissue samples was downloaded from the GTEx portal (RRID:SCR_013042; ref. 29; DOI: 10.1038/ng.2653; <https://www.gtexportal.org/>, version: v8).
- 2) **Smith 2015** (PMID: 26460041)²⁰: RNAseq data of 5 prostate basal cell samples (Trop2⁺CD49f^{hi}) and corresponding 5 luminal cell samples (Trop2⁺CD49f^{lo}) FACS-purified from benign human prostate

tissue in 5 PCa patients (n=5) was retrieved from the GEO database GSE82071 (ref. 20; DOI: 10.1073/pnas.1518007112).

3) **Liu 2016** (PMID: 27926864)⁵⁶: RNAseq data of 3 FACS-purified human prostate basal cell samples (CD45⁻EpCAM⁺CD49f^{hi}CD38^{lo}) with corresponding luminal (CD45⁻EpCAM⁺CD49f^{lo}CD38^{hi}) and luminal progenitor (CD45⁻EpCAM⁺CD49f^{lo}CD38^{lo}) cell populations purified from the benign prostate tissues of 3 PCa patients (n=3) was retrieved from the supplementary data from Liu *et al.*, 2016 (ref. 56). The microarray-based transcriptomic profiles of the same samples were retrieved from the GEO database GSE89050 (ref. 56; DOI: 10.1016/j.celrep.2016.11.010).

4) **Zhang 2016** (PMID: 26924072)¹⁹: RNAseq data of 3 human prostate basal cell (Trop2⁺CD49f^{hi}) with the corresponding 3 luminal cell (Trop2⁺CD49f^{lo}) samples FACS-purified from benign prostate tissues of 3 PCa patients (n=3) was retrieved from the GEO database GSE67070 (ref. 19; DOI: 10.1038/ncomms10798).

5) **TCGA PRAD 2015** and **Pan-Cancer Atlas 2018**: The Cancer Genome Atlas (TCGA) (<https://www.cancer.gov/tcga>) repository collected high-throughput molecular and clinical data from primary cancer and matched normal/benign samples spanning 33 human cancer types, including prostate adenocarcinoma (PRAD). For each cancer type, TCGA published an initial "marker papers" summarizing analyses performed. Supplemental and associated data files supporting these "marker papers" are available through the NCI Genomic Data Commons (GDC) website (<https://gdc.cancer.gov/about-data/publications>). For prostate cancer, **TCGA PRAD 2015** (PMID: 26544944; ref. 18; DOI: 10.1016/j.cell.2015.10.025) is the marker paper on PCa published in 2015, which includes the dataset of 333 primary prostate adenocarcinomas (281 treatment-naïve Pri-PCa, 49 PCa with post-surgery adjuvant ADT, and 52 matched normal/benign prostate tissue). Genomic data were obtained from **cBioPortal for Cancer Genomics** (RRID:SCR_014555; <https://www.cbioportal.org>), specifically the TCGA-PRAD Firehose Legacy study (https://www.cbioportal.org/study/summary?id=prad_tcga). RNAseq expression data was obtained from Zhang *et al.*, 2020 (PMID: 32350277; ref. 16; DOI: 10.1038/s41467-020-15815-7).

The expanded **TCGA PRAD 2018 (Pan-Cancer Atlas)** cohort includes 494 patients with Pri-PCa (422 treatment-naïve Pri-PCa, 70 PCa with post-surgery adjuvant hormone treatment, 2 PCa with chemotherapy) and 52 matched adjacent normal/benign prostate tissue. Clinicopathologic features (age, Gleason scores, PSA levels, tumor stages) and genomic alterations data (mutation frequency, copy number alterations [CNAs], structural variants [SVs], tumor mutation burden [TMB], fraction genome altered) were retrieved from cBioPortal for Cancer Genomics PRAD PanCancer Atlas study (https://www.cbioportal.org/study/summary?id=prad_tcga_pan_can_atlas_2018) and from the Pan-Cancer Atlas companion site (welcome to the Pan-Cancer Atlas webpage: <https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html>). Survival data were obtained from Liu *et al.*, 2018⁶⁵ (PMID: 29625055; DOI: 10.1016/j.cell.2018.02.052). Normalized RNAseq data were

downloaded via NCI GDC hub at UCSC Xena Functional Genomics Portal (<https://gdc.xenahubs.net>, version 2019-07-20; DOI: 10.1038/s41587-020-0546-8).

- 6) **Rajan 2014** (PMID: 24054872)²⁵: Clinicopathologic (Gleason scores) and RNAseq data of 7 matched pre- and post-nADT PCa samples were obtained, respectively, from Table 1 in Rajan *et al.*, 2014 (ref. 25; DOI: 10.1016/j.eururo.2013.08.011) and the GEO database GSE48403.
- 7) **Sharma 2018** (PMID: 30314329)²⁶: Clinicopathologic (Gleason scores) and RNAseq data of 7 matched pre- and post-nADT PCa samples from the responder group (also defined as "Low Impact Group" in the source reference) was obtained, respectively, from supplementary Table 1 in Sharma *et al.*, 2018 (ref. 26; DOI: 10.3390/cancers10100379) and the GEO database GSE111177.
- 8) **Long 2020** (PMID: 32951005; DOI: 10.1016/j.eururo.2013.08.011)²⁴: RNAseq data of 6 matched pre- and post-nADT PCa samples was obtained from the GEO database GSE82071. Additionally, RNAseq data from 10 locally advanced PCa samples along with their paired adjacent benign prostate tissues were obtained from the GEO database GSE82071.
- 9) **Nastiuk (RPCI) 2020** (submitted)⁶⁶: Clinicopathologic data (Gleason scores) and RNAseq data of 43 Pri-PCa samples from post-nADT PCa patients and 43 tumor samples from age-, stage-, clinically-matched PCa patients without nADT was obtained from Jamroze *et al.*, 2024 (work led by Drs. K. Nastiuk and G. Chatta at the Roswell Park Comprehensive Cancer Center). RNAseq data and clinicopathologic information from the Nastiuk 2020 cohort (unpublished) were provided in-house by Dr. K. Nastiuk and are available from the corresponding author upon reasonable request.
- 10) **Gerhauser 2018** (PMID: 30537516; DOI: 10.1016/j.ccell.2018.10.016)⁶⁷: RNAseq data of 118 primary tumor samples and 9 adjacent benign prostate tissue from early-onset PCa patients was obtained from the European Genome-Phenome Archive (EGA) (<https://ega-archive.org>), accession number EGAS00001002923.
- 11) **Wyatt 2014** (PMID: 25155515; DOI: 10.1186/s13059-014-0426-y)²³: Clinicopathologic (Gleason scores) and RNAseq data of 12 primary PCa samples from treatment-naïve PCa patients and 12 adjacent benign prostate tissue were obtained from the European Nucleotide Archive (ENA) (<https://www.ebi.ac.uk/ena/browser/home>), accession number PRJEB6530.
- 12) **Spratt 2017** (PMID: 28358655; DOI: 10.1200/JCO.2016.70.2811)³⁸: Transcriptomic profiles (Microarray) and clinicopathologic data of 855 radical prostatectomy specimens from a multi-institutional study of intermediate- and high-risk localized PCa patients were obtained from Veracyte GRID (<https://decipherbio.com/grid/>).
- 13) **Tosoian 2020** (PMID: 32231245; DOI: 10.1038/s41391-020-0226-2)³⁹: Transcriptomic profiles (Microarray) and clinicopathologic data of 405 radical prostatectomy and biopsy specimens from high-risk localized PCa patients were obtained from Veracyte GRID (<https://decipherbio.com/grid/>).
- 14) **CHAARTED correlatives 2021** (PMID: 34129855; DOI: 10.1016/j.annonc.2021.06.003)⁴⁰: Transcriptomic profiles (Microarray) and clinicopathologic data of 160 biopsy specimens from a Phase

3 trial of metastatic hormone-sensitive PCa patients were obtained from the NCTN Data Archive (<https://nctn-data-archive.nci.nih.gov>), accession number **NCT00309985-D14**.

15) **Decipher GRID Biopsy** (PMID: 37060201; DOI: 10.1002/cncr.34790)³⁵: Transcriptomic profiles from 82,470 prospectively collected prostate biopsy samples were retrieved from the Decipher GRID database (RRID: SCR_006552; <https://decipherbio.com/grid>; ClinicalTrials.gov identifier: **NCT02609269**; <https://clinicaltrials.gov/ct2/show/NCT02609269>). Patient data were de-identified from clinical use of the Decipher prostate genomic classifier in accordance with the Safe Harbor method described in the HIPAA Privacy Rule 45 CFR 164.514(b) and (c) (Veracyte, San Diego, CA). The samples, comprising a large cohort of localized prostate cancer (PCa), were utilized to compare the distribution of Stemness Index values across National Comprehensive Cancer Network (NCCN) risk groups.

Clinicopathologic Breakdown and Risk Group Analysis:

Detailed clinicopathologic review revealed that the majority of the cohort presents with low-grade localized PCa. The TNM staging breakdown includes:

- **T1 stage:** Approximately 50% of samples, indicating early-stage cancer localized within the prostate.
- **T2 stage:** Approximately 8%, representing cancer confined within the prostate but more extensive than T1.
- **T3 and T4 stages:** Less than 1%, suggesting advanced local spread.
- **N0:** Approximately 98%, indicating no regional lymph node involvement.
- **N1:** Approximately 2%, suggesting regional lymph node involvement.

The Decipher GRID database (RRID:SCR_006552) categorizes patients into NCCN risk groups based on TNM stage, PSA level, and Gleason score:

- **Very Low and Low Risk:** T1–T2a, N0, M0, PSA <10 ng/mL, and Gleason score ≤6; often managed with active surveillance or less aggressive treatment.
- **Intermediate Risk – Favorable:** Typically T2b–T2c, N0, M0, PSA 10–20 ng/mL, and Gleason score 7; managed with less aggressive treatment but closer monitoring.
- **Intermediate Risk – Unfavorable:** Similar clinical features to favorable but with higher disease burden or adverse factors requiring careful monitoring.
- **High Risk:** Usually involves T3a, N0, M0, PSA >20 ng/mL, or Gleason score 8–10; often managed aggressively with surgery, radiation, and/or hormonal therapy.
- **Very High Risk:** Typically includes T3b–T4, any N, M0, and high Gleason scores or PSA levels (PSA >20 ng/mL); patients are at significant risk for metastatic progression and are treated aggressively, often with multimodal therapies.

Observations on Disease Severity and Treatment Naïvety:

Analysis of the cohort indicates that only approximately 13% of cases fall into the High and Very High risk categories (8,376 + 2,399 = 10,775 samples), suggesting that the majority (approximately 87%) are below Intermediate Risk (\leq GS7). This distribution implies that most patients were treatment-naïve and likely androgen deprivation therapy (ADT)-free due to less aggressive disease status.

- 16) **Bolis 2021** (PMID: 34857732; DOI: 10.1038/s41467-021-26840-5)²⁷: RNAseq and clinicopathologic data of 1,223 clinical samples from an integrated cohort consisting of normal prostate specimens (n = 174), Pri-PCa (n = 714) and mCRPC (n = 335) was obtained from the Zenodo repository (Record ID: 5546618; <https://zenodo.org/records/5546618>). This resource of Prostate Cancer Transcriptome Atlas (<https://prostatecanceratlas.org>) was built and integrated from the following studies/datasets: (1) Genotype-Tissue Expression Database (GTEx; PMID: 23715323; ref. 29); (2) The Cancer Genome Atlas (TCGA, TCGA-PRAD; PMID: 26544944; ref. 18); (3) Atlas of RNA-sequencing profiles of normal human tissues⁶⁸ (GSE120795; PMID: 31015567; ref. 68); (4) Integrative epigenetic taxonomy of PNPc⁶⁹ (GSE120741; PMID: 30464211; ref. 69); (5) Prognostic markers in locally advanced lymph node-negative prostate cancer (PRJNA477449); (6) The long noncoding RNA landscape of NEPC and its clinical implications⁷⁰ (PRJEB21092; PMID: 29757368; ref. 70); (7) Integrative clinical sequencing analysis of metastatic CRPC reveals a high frequency of clinical actionability (PRJNA283922; dbGaP: phs000915; PMID: 26000489; ref. 10); (8) CSER—exploring precision cancer medicine for sarcoma and rare cancers⁷¹ (PRJNA223419; dbGaP: phs000673; PMID: 28783718; ref. 71); (9) Molecular basis of NEPC (Beltran 2016; PRJNA282856; dbGaP: phs000909; PMID: 26855148; ref. 28); (10) Heterogeneity of androgen receptor splice variant-7 (AR-V7) protein expression and response to therapy in CRPC⁷² (GSE118435; PMID: 30334814; ref. 72); (11) RNAseq of human PCa cell lines and mCRPC tumor^{73, 74, 75} (GSE14750; PMIDs: 32460015; ref. 73; 33658518; ref. 74; 34244513; ref. 75 | GSE171729; PMID: 34244513; ref. 75); and (12) Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic CRPC (PRJNA520923; GEO: GSE126078; PMID: 31361600; ref. 33).
- 17) **Taylor 2010** (PMID: 20579941; DOI: 10.1016/j.ccr.2010.05.026)³⁶: Transcriptomic profiles (Microarray) and clinicopathologic (recurrence-free survival) data of 29 adjacent benign prostate tissue, 131 Pri-PCa, and 19 mPCa were obtained from the GEO database (GSE21034) and the cBioPortal for Cancer Genomics PRAD MSKCC 2010 study (https://www.cbioportal.org/study/summary?id=prad_mskcc).
- 18) **SU2C 2015** (PMID: 26000489; DOI: 10.1016/j.cell.2015.05.001)¹⁰: Clinical and genomic data of 150 mCRPC samples were obtained from the cBioPortal for Cancer Genomics PRAD SU2C 2015 study (https://www.cbioportal.org/study/summary?id=prad_su2c_2015). RNAseq data of 98 mCRPC samples was retrieved from Zhang *et al.*, 2020¹⁶ (PMID: 32350277; DOI: 10.1038/s41467-020-15815-7; ref.16).
- 19) **SU2C 2019** (PMID: 31061129; DOI: 10.1073/pnas.1902651116)⁴¹: This dataset includes clinicopathologic and genomic data for 444 mCRPC samples. RNAseq data of 266 mCRPC samples

(prepared by using poly(A) enrichment for RNAseq library construction) were obtained from the cBioPortal for Cancer Genomics PRAD SU2C 2019 study (https://www.cbioportal.org/study/summary?id=prad_su2c_2019). In parallel with TCGA-PRAD, clinicopathologic features (age, Gleason scores, PSA levels), genomic alterations (mutation frequency, CNAs, SVs, TMB, fraction genome altered), and patient outcome data (overall survival) were also retrieved from cBioPortal (RRID:SCR_014555). These datasets were used for comparative analysis between Stemness-high and Stemness-low groups, and to explore genomic drivers and clinical associations as described in the Methods.

20) **Labrecque 2019** (PMID: 31361600; DOI: 10.1172/JCI128212)³³: RNAseq data of 98 mCRPC samples and 39 PDX-LuCaP was obtained from the GEO database GSE126078.

21) **Alumkal 2020** (PMID: 32424106; DOI: 10.1073/pnas.1922207117)¹³: RNAseq data of 25 mCRPC samples before treatment with enzalutamide (ENZA) was obtained from Supplementary dataset S01 of Alumkal *et al.*, 2020 (ref. 13).

22) **Westbrook 2022** (PMID: 36109521; DOI: 10.1038/s41467-022-32701-6)¹²: RNAseq data of 42 mCRPC samples consisting of 21 matched pre/post Enzalutamide treatment samples was retrieved from Westbrook *et al.*, 2022 (ref. 12).

23) **Beltran 2016** (PMID: 26855148; DOI: 10.1038/nm.4045)²⁸: RNAseq data of mCRPC samples, including 34 castration-resistant adenocarcinoma (mCRPC-Adeno) and 15 neuroendocrine histologies (mCRPC-NEPC) samples, was obtained from cBioPortal for the Cancer Genomics (https://www.cbioportal.org/study/summary?id=nepc_wcm_2016).

24) **Goodrich 2017** (PMID: 28059767; DOI: 10.1126/science.aah4199)³⁷: RNAseq data of murine prostate tumors from genetically engineering mouse models (GEMMs) of PCa of different genotypes, including 4 SKO (single knockout: PBCre4:*Pten*^{ff}), 13 DKO (double knockout PBCre4:*Pten*^{ff}:*Rb1*^{ff}), and 6 TKO (PBCre4:*Pten*^{ff}:*Rb1*^{ff}:*Trp53*^{ff}) was obtained from the GEO database GSE90891 (*f* = a floxed allele of the indicated genes).

25) **CCLE 2018** (PMID: 22460905; DOI: 10.1038/nature11003)³¹: RNAseq data of 1,019 human cell lines were obtained from the Xena Functional Genomics Portal⁷⁶ (<https://xenabrowser.net>, version: 2018-05-30; ref. 76). Among them, eight prostate cancer cell lines — DU145 (RRID:CVCL_0105), LNCaP clone FGC (RRID:CVCL_1379), MDA PCa 2b (RRID:CVCL_4745), NCI-H660 (RRID:CVCL_0459), PC-3 (RRID:CVCL_0035), VCaP (RRID:CVCL_2235), 22Rv1 (RRID:CVCL_1045), and PRECLH (PrEC LH; RRID:CVCL_V626) — were included for analysis.

26) **XG-LNCaP** and **XG-LAPC9** (PMID: 30190514; DOI: 10.1038/s41467-018-06067-7)³⁴: RNAseq data of 12 LNCaP xenografts (XG-LNCaP) and 10 LAPC9 xenografts (XG-LAPC9) was obtained from the GEO database GSE88752.

Sex as a biological variable

All clinical datasets analyzed in this study consisted exclusively of male patients, consistent with the male-specific nature of PCa. Preclinical models, including xenografts, patient-derived xenografts, and genetically engineered mouse models (GEMMs), were also male-derived. All PCa cell lines utilized in the study were of male origin, with the exception of [Supplementary Figure S5A](#), where a broader cohort of 1,019 cell lines from the Cancer Cell Line Encyclopedia (CCLE), comprising both male- and female-derived samples across multiple cancer types, was analyzed. Analyses of CCLE data in [Supplementary Figure S5A](#) were performed without stratification by sex.

Transcriptomic data normalization and transcriptome-based signature scores

To ensure consistency and compatibility across diverse datasets, we employed normalization strategies tailored to the specific data types and analytical approaches.

Canonical AR activity score (AR-A) and MYC activity score (MYC-sig): Canonical pro-differentiation AR-A can be determined by RNA-seq based measurements of AR-regulated transcripts under intact androgen/AR signaling conditions. Similarly, MYC-sig can be determined by MYC-regulated transcripts. We employed Z-score normalization method⁷⁷ to determine AR-A and MYC-sig. Briefly, the numeric AR or MYC activity score was calculated based on a linear combination of expression values (Z-scores) of experimentally validated AR targets¹⁴ or MYC targets⁵¹ ([Supplementary Table S2](#), GSEA Molecular Signature Database 2022.1 version, MSigDB). The target gene lists were detailed in [Supplementary Table S2](#). The expression value for each AR or MYC target genes ([Supplementary Table S2](#)) was converted to Z-score by $Z = (x - \mu) / \sigma$, where x is the expression value of the target gene in a specific sample, μ is the mean and σ is the standard deviation across all samples of a gene. Finally, the combined Z-scores were summed across all genes to represent the expression score of canonical AR or MYC target genes. Analysis of gene expression signatures using Z-score normalization (AR-A and MYC-sig) was confined within individual datasets, obviating the need for cross-dataset comparisons.

To further minimize batch effects and ensure data integrity and comparability across PCa stages, normalized RNAseq data from a published integrated data framework, the Prostate Cancer Transcriptome Atlas²⁷ (<https://prostatecanceratlas.org>), was used. In this harmonized resource, raw RNAseq data with high-quality sequencing reads from different datasets were re-mapped to the human reference genome GRCh38, adjusted for library size, and normalized using the variance stabilizing transformation pipeline using DESeq2 (version 1.28.1; RRID:SCR_015687)²⁷.

Transcriptome-based quantitative measurement of oncogenic stemness (the Stemness Index): We adapted mRNA-based Stemness Index (mRNAsi) to determine the degree of oncogenic dedifferentiation (Stemness)⁶. Briefly, we computed Spearman correlations between the stemness model's weight vector and the transcriptome expression profile of the queried samples, assuming paired observations, at least ordinal scale data, and a monotonic relationship without significant outliers. Spearman correlation

calculations were performed using the ‘stats’ package (version 4.3.3) in R (version 4.3.3; RRID:SCR_001905). The resulting correlation coefficients were then linearly transformed to scale the mRNAsi scores between 0 and 1 using the following formula: Stemness Index = (Spearman correlation + 0.2619) / 0.4151. Spearman correlation is advocated for calculating the Stemness Index because it is more robust with respect to potential cross-dataset batch effects that may arise⁶.

Transcriptome-based determination of “PCa-Stem Signature”: To derive the PCa-Stem Signature, we performed single-sample gene set enrichment analysis (ssGSEA) using the Gene Set Variation Analysis (GSVA)⁷⁸ package (version 1.50.5; RRID:SCR_021058) in R (version 4.3.3; RRID:SCR_001905). Briefly, to identify robust stemness-associated genes in PCa, we performed differential gene expression (DEG) analysis between the top 33% (Stemness-high) and bottom 33% (Stemness-low) of samples, stratified by Stemness scores in both treatment-naïve primary PCa (TCGA-PRAD) and metastatic CRPC (SU2C 2019) cohorts. Genes significantly upregulated in both cohorts (fold change ≥ 2 , FDR < 0.05) were intersected to derive a 12-gene consensus signature—termed the PCa-Stem Signature ([Supplementary Table S2](#)). ssGSEA was then performed using the GSVA R package to calculate PCa-Stem scores across PCa datasets. This signature was validated in independent cohorts and assessed for clinical relevance and transcriptomic subtype enrichment, including associations with aggressive PCa subtypes and patient survival outcomes.

Integrative analysis of genomic alterations in primary PCa (TCGA) and mCRPC (SU2C)

To explore molecular and clinical features associated with Stemness, we performed integrative analysis of genomic, transcriptomic, clinicopathologic, and outcome data using cBioPortal for Cancer Genomics (RRID:SCR_014555), which provides interactive access to multidimensional cancer genomics datasets and integrated visualization tools⁷⁹. Specifically, we analyzed data from primary prostate cancer (Pri-PCa; **TCGA-PRAD 2018**, Pan-Cancer Atlas) and metastatic castration-resistant PCa (mCRPC; **SU2C 2019**) cohorts. Clinicopathologic variables (e.g., age, Gleason scores, PSA levels, tumor stage), genomic alterations (e.g., mutation frequency, TMB, fraction genome altered, CNAs, and SVs), and patient outcome data (e.g., overall survival, progression-free survival) were retrieved from cBioPortal (RRID:SCR_014555) as described in the “PCa clinical datasets and data collection” section. Transcriptomic data used for Stemness score calculation were obtained as described in the same section. For comparative analysis, mCRPC samples were stratified into top and bottom 33% by Stemness score. In TCGA, only treatment-naïve Pri-PCa samples were used for a similar 33% stratification, excluding 70 patients who had received adjuvant (post-surgery) hormone therapy to reflect spontaneous disease progression. Differential expression analysis was conducted using DESeq2⁸⁰ (version 1.42.1; RRID:SCR_015687), and gene-level and genome-wide alteration frequencies were assessed across stratified groups and disease stages (e.g.,

Pri-PCa vs. mCRPC) using OncoPrint visualizations, bar plots, and correlation analysis. Associations with clinical outcomes were evaluated as described in the “Statistical analysis” section.

Statistical analysis

Statistical analyses were performed using GraphPad Prism software (version 10.4.1 (532); RRID:SCR_002798) and R (version 4.3.3; RRID:SCR_001095).

For comparisons between two groups, independent sample or paired *t*-tests were used to compare group means, assuming normality and equal variance. For multi-group comparisons, Tukey’s range test was employed following one-way ANOVA (adjusted *p*-values listed in [Supplementary Table S4](#)). The Wilcoxon rank-sum test was applied for comparisons of altered genome fraction, TMB, or age between two groups. For differential expression analyses and comparisons involving multiple hypotheses (e.g., RNAseq-based gene comparisons), *p*-values were further adjusted using the false discovery rate (FDR) method to control for multiple testing.

Pearson’s correlation coefficient was used to evaluate linear associations between two continuous variables, assuming normally distributed paired observations. Linear regression modeling was applied to examine the relationship between two continuous variables, with the slope of the best-fit regression line reported where applicable. Spearman’s rank correlation was employed to assess monotonic relationships between continuous or ordinal variables, assuming paired observations and monotonic trends without significant outliers.

The Jonckheere-Terpstra (J-T) trend test was used to assess monotonic association between an ordinal variable and a continuous variable. Chi-square test or Fisher’s exact test was performed to evaluate the associations between categorical variables.

For the analysis of patient progression-free survival (PFS), recurrence-free survival (RFS) and overall survival (OS), the standard Kaplan-Meier method was used for the estimation of survival fractions and log-rank tests were used for group comparisons. Cox proportional hazards models were used for multivariable analysis and for assessing the prognostic significance of gene signatures (Stemness, PCa-Stem signature, etc). For the mRNA-based gene signatures (Stemness, PCa-Stem signature, AR-A), patients were stratified into high and low signature groups based on median split in cohorts including Spratt 2017 ([Fig. 4C](#)), Tosoian 2020 ([Fig. 4C](#)), CHARTED 2021 ([Supplementary Fig. S6C](#)), Taylor 2010 ([Supplementary Fig. S6C](#)), TCGA-PRAD 2018 ([Supplementary Fig. S6C](#)), and SU2C 2019 ([Supplementary Fig. S6C](#)). For certain analyses, including SU2C 2019 ([Supplementary Fig. S6C](#)) and CHARTED 2021 ([Fig. 4C](#)), patients were alternatively stratified by comparing the top 25% versus the bottom 75% of signature scores to assess whether patients with very high Stemness levels exhibited a higher risk of disease progression.

All statistical tests were two-sided, and $P < 0.05$ was considered statistically significant.

Data and Code Availability

All datasets and computational codes supporting the findings of this study are described below.

Code Availability: Custom R scripts used for data preprocessing, differential expression analysis, gene signature quantification, and survival/statistical modeling are available from the corresponding author upon reasonable request.

Data Availability: All datasets analyzed in this study are publicly available from repositories including GEO (RRID:SCR_005012), cBioPortal (RRID:SCR_014555), the NCI GDC Portal, Xena Functional Genomics Portal (RRID:SCR_018938), EGA (RRID:SCR_004944), ENA (RRID:SCR_006515), and Zenodo (RRID:SCR_004129), as detailed in [Supplementary Table S1](#).

References

1. Ben-Porath I, *et al.* An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* **40**, 499-507 (2008).
2. Pece S, *et al.* Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. *Cell* **140**, 62-73 (2010).
3. Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* **2**, 333-344 (2008).
4. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell* **24**, 65-78 (2019).
5. Loh JJ, Ma S. Hallmarks of cancer stemness. *Cell Stem Cell* **31**, 617-639 (2024).
6. Malta TM, *et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. *Cell* **173**, 338-354 e315 (2018).
7. Liu X, Li WJ, Puzanov I, Goodrich DW, Chatta G, Tang DG. Prostate cancer as a dedifferentiated organ: androgen receptor, cancer stem cells, and cancer stemness. *Essays Biochem* **66**, 291-303 (2022).
8. Tang DG. Understanding and targeting prostate cancer cell heterogeneity and plasticity. *Semin Cancer Biol* **82**, 68-93 (2022).
9. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*, (2024).
10. Robinson D, *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-1228 (2015).
11. He MX, *et al.* Transcriptional mediators of treatment resistance in lethal prostate cancer. *Nat Med* **27**, 426-433 (2021).
12. Westbrook TC, *et al.* Transcriptional profiling of matched patient biopsies clarifies molecular determinants of enzalutamide-induced lineage plasticity. *Nat Commun* **13**, 5345 (2022).
13. Alumkal JJ, *et al.* Transcriptional profiling identifies an androgen receptor activity-low, stemness program associated with enzalutamide resistance. *Proc Natl Acad Sci U S A* **117**, 12315-12323 (2020).
14. Bluemn EG, *et al.* Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. *Cancer Cell* **32**, 474-489.e476 (2017).
15. Hieronymus H, *et al.* Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* **10**, 321-330 (2006).
16. Zhang D, *et al.* Intron retention is a hallmark and spliceosome represents a therapeutic vulnerability in aggressive prostate cancer. *Nat Commun* **11**, 2089 (2020).
17. Spratt DE, *et al.* Transcriptomic Heterogeneity of Androgen Receptor Activity Defines a de novo low AR-Active Subclass in Treatment Naive Primary Prostate Cancer. *Clin Cancer Res* **25**, 6721-6730 (2019).
18. Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011-1025 (2015).
19. Zhang D, *et al.* Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. *Nat Commun* **7**, 10798 (2016).
20. Smith BA, *et al.* A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc Natl Acad Sci U S A* **112**, E6544-6552 (2015).

- 832 21. Moad M, et al. Multipotent Basal Stem Cells, Maintained in Localized Proximal Niches, Support
833 Directed Long-Ranging Epithelial Flows in Human Prostates. *Cell Rep* **20**, 1609-1622 (2017).
- 834 22. Grossmann S, et al. Development, maturation, and maintenance of human prostate inferred from
835 somatic mutations. *Cell Stem Cell* **28**, 1262-1274 e1265 (2021).
- 836 23. Wyatt AW, et al. Heterogeneity in the inter-tumor transcriptome of high risk prostate cancer.
837 *Genome Biol* **15**, 426 (2014).
- 838 24. Long X, et al. Immune signature driven by ADT-induced immune microenvironment remodeling in
839 prostate cancer is correlated with recurrence-free survival and immune infiltration. *Cell Death Dis*
840 **11**, 779 (2020).
- 841 25. Rajan P, et al. Next-generation sequencing of advanced prostate cancer treated with androgen-
842 deprivation therapy. *Eur Urol* **66**, 32-39 (2014).
- 843 26. Sharma NV, et al. Identification of the Transcription Factor Relationships Associated with
844 Androgen Deprivation Therapy Response and Metastatic Progression in Prostate Cancer.
845 *Cancers (Basel)* **10**, (2018).
- 846 27. Bolis M, et al. Dynamic prostate cancer transcriptome analysis delineates the trajectory to
847 disease progression. *Nat Commun* **12**, 7033 (2021).
- 848 28. Beltran H, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer.
849 *Nat Med* **22**, 298-305 (2016).
- 850 29. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585 (2013).
- 851 30. Aran D, et al. Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nat Commun*
852 **8**, 1077 (2017).
- 853 31. Barretina J, et al. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer
854 drug sensitivity. *Nature* **483**, 603-607 (2012).
- 855 32. Ghandi M, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*
856 **569**, 503-508 (2019).
- 857 33. Labrecque MP, et al. Molecular profiling stratifies diverse phenotypes of treatment-refractory
858 metastatic castration-resistant prostate cancer. *J Clin Invest* **129**, 4492-4505 (2019).
- 859 34. Li Q, et al. Linking prostate cancer cell AR heterogeneity to distinct castration and enzalutamide
860 responses. *Nature communications* **9**, 1-17 (2018).
- 861 35. Weiner AB, et al. A novel prostate cancer subtyping classifier based on luminal and basal
862 phenotypes. *Cancer* **129**, 2169-2178 (2023).
- 863 36. Taylor BS, et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22
864 (2010).
- 865 37. Ku SY, et al. Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis,
866 and antiandrogen resistance. *Science* **355**, 78-83 (2017).
- 867 38. Spratt DE, et al. Individual Patient-Level Meta-Analysis of the Performance of the Decipher
868 Genomic Classifier in High-Risk Men After Prostatectomy to Predict Development of Metastatic
869 Disease. *J Clin Oncol* **35**, 1991-1998 (2017).
- 870 39. Tosoian JJ, et al. Performance of clinicopathologic models in men with high risk localized prostate
871 cancer: impact of a 22-gene genomic classifier. *Prostate Cancer Prostatic Dis* **23**, 646-653
872 (2020).
- 873 40. Hamid AA, et al. Transcriptional profiling of primary prostate tumor in metastatic hormone-
874 sensitive prostate cancer and association with clinical outcomes: correlative analysis of the
875 E3805 CHARTED trial. *Ann Oncol* **32**, 1157-1166 (2021).

- 876 41. Abida W, et al. Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci U S A* **116**, 11428-11436 (2019).
- 877
- 878 42. Zhao SG, et al. Associations of Luminal and Basal Subtyping of Prostate Cancer With Prognosis
- 879 and Response to Androgen Deprivation Therapy. *JAMA Oncol* **3**, 1663-1672 (2017).
- 880 43. Zhao SG, et al. Clinical and Genomic Implications of Luminal and Basal Subtypes Across
- 881 Carcinomas. *Clin Cancer Res* **25**, 2450-2457 (2019).
- 882 44. Coleman IM, et al. Therapeutic Implications for Intrinsic Phenotype Classification of Metastatic
- 883 Castration-Resistant Prostate Cancer. *Clin Cancer Res* **28**, 3127-3140 (2022).
- 884 45. You S, et al. Integrated Classification of Prostate Cancer Reveals a Novel Luminal Subtype with
- 885 Poor Outcome. *Cancer Res* **76**, 4948-4958 (2016).
- 886 46. Yoon J, et al. A comparative study of PCS and PAM50 prostate cancer classification schemes.
- 887 *Prostate Cancer Prostatic Dis* **24**, 733-742 (2021).
- 888 47. Jamroze A, Liu X, Tang DG. Treatment-induced stemness and lineage plasticity in driving
- 889 prostate cancer therapy resistance. *Cancer Heterogeneity and Plasticity* **1**, (2024).
- 890 48. Tang F, et al. Chromatin profiles classify castration-resistant prostate cancers suggesting
- 891 therapeutic targets. *Science* **376**, eabe1505 (2022).
- 892 49. Armenia J, et al. The long tail of oncogenic drivers in prostate cancer. *Nat Genet* **50**, 645-651
- 893 (2018).
- 894 50. Haffner MC, et al. Genomic and phenotypic heterogeneity in prostate cancer. *Nat Rev Urol* **18**,
- 895 79-92 (2021).
- 896 51. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular
- 897 Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).
- 898 52. Tomlins SA, et al. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate
- 899 cancer. *Science* **310**, 644-648 (2005).
- 900 53. Stuchbery R, et al. Reduction in expression of the benign AR transcriptome is a hallmark of
- 901 localised prostate cancer progression. *Oncotarget* **7**, 31384-31392 (2016).
- 902 54. Schroeder A, et al. Loss of androgen receptor expression promotes a stem-like cell phenotype in
- 903 prostate cancer through STAT3 signaling. *Cancer Res* **74**, 1227-1237 (2014).
- 904 55. Giridhar PV, Williams K, VonHandorf AP, Deford PL, Kasper S. Constant degradation of the
- 905 androgen receptor by MDM2 conserves prostate cancer stem cell integrity. *Cancer research* **79**,
- 906 1124-1137 (2019).
- 907 56. Liu X, et al. Low CD38 Identifies Progenitor-like Inflammation-Associated Luminal Cells that Can
- 908 Initiate Human Prostate Cancer and Predict Poor Outcome. *Cell Rep* **17**, 2596-2606 (2016).
- 909 57. Zheng K, et al. Integrative multi-omics analysis unveils stemness-associated molecular subtypes
- 910 in prostate cancer and pan-cancer: prognostic and therapeutic significance. *J Transl Med* **21**, 789
- 911 (2023).
- 912 58. Zhang T, et al. Identification of a novel stemness-related signature with appealing implications in
- 913 discriminating the prognosis and therapy responses for prostate cancer. *Cancer Genet* **276-277**,
- 914 48-59 (2023).
- 915 59. Zhao F, et al. Integrated single-cell transcriptomic analyses identify a novel lineage plasticity-
- 916 related cancer cell type involved in prostate cancer progression. *EBioMedicine* **109**, 105398
- 917 (2024).
- 918 60. Zhao F, et al. Deciphering single-cell heterogeneity and cellular ecosystem dynamics during
- 919 prostate cancer progression. *bioRxiv*, (2024).

61. Bian X, et al. Integration Analysis of Single-Cell Multi-Omics Reveals Prostate Cancer Heterogeneity. *Adv Sci (Weinh)* **11**, e2305724 (2024).
62. Kirk JS, et al. Integrated single-cell analysis defines the epigenetic basis of castration-resistant prostate luminal cells. *Cell Stem Cell*, (2024).
63. Gulati GS, et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science* **367**, 405-411 (2020).
64. Chauhan PS, et al. Genomic and Epigenomic Analysis of Plasma Cell-Free DNA Identifies Stemness Features Associated with Worse Survival in Lethal Prostate Cancer. *Clin Cancer Res* **31**, 151-163 (2025).
65. Liu J, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400-416 e411 (2018).
66. Jamroze A, et al. Neoadjuvant androgen deprivation therapy initially activates and subsequently suppresses the prostate cancer tumor immune microenvironment ((Submitted 2024)).
67. Gerhauser C, et al. Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. *Cancer Cell* **34**, 996-1011 e1018 (2018).
68. Suntsova M, et al. Atlas of RNA sequencing profiles for normal human tissues. *Sci Data* **6**, 36 (2019).
69. Stelloo S, et al. Integrative epigenetic taxonomy of primary prostate cancer. *Nat Commun* **9**, 4900 (2018).
70. Ramnarine VR, et al. The long noncoding RNA landscape of neuroendocrine prostate cancer and its clinical implications. *Gigascience* **7**, (2018).
71. Robinson DR, et al. Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297-303 (2017).
72. Sharp A, et al. Androgen receptor splice variant-7 expression emerges with castration resistance in prostate cancer. *J Clin Invest* **129**, 192-208 (2019).
73. Nyquist MD, et al. Combined TP53 and RB1 Loss Promotes Prostate Cancer Resistance to a Spectrum of Therapeutics and Confers Vulnerability to Replication Stress. *Cell Rep* **31**, 107669 (2020).
74. Brady L, et al. Inter- and intra-tumor heterogeneity of metastatic prostate cancer determined by digital spatial gene expression profiling. *Nat Commun* **12**, 1426 (2021).
75. Lim Y, et al. Multiplexed functional genomic analysis of 5' untranslated region mutations across the spectrum of prostate cancer. *Nat Commun* **12**, 4217 (2021).
76. Goldman MJ, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* **38**, 675-678 (2020).
77. Lee E, Chuang HY, Kim JW, Ideker T, Lee D. Inferring pathway activity toward precise disease classification. *PLoS Comput Biol* **4**, e1000217 (2008).
78. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
79. Gao J, et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, p11 (2013).
80. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

Acknowledgements

Work in Tang lab was supported, in part, by grants from the U.S. National Institutes of Health (NIH) National Cancer Institute (NCI) R01CA237027, R01CA240290 and 2R01CA240290-06A1, grants from the U.S. Department of Defense (DOD) PC220137 and PC220273, Roswell Park Comprehensive Cancer Center and the NCI Center grant P30CA016056, Roswell Park Alliance Foundation (RPAF) and the George Decker Endowment fund. Work in Goodrich lab was support by NIH/NCI R01CA234162 and R01CA230913. Work in the labs of Goodrich, Tang and Chatta was further supported by the Prostate Cancer Foundation (PCF) Challenge Award 2022CHAL3788 (PI: Goodrich). S. Liu and J. Wang were supported by NIH/NCI U24CA232979 and U24CA274159. Work in Alumkal lab was supported by NIH/NCI R01CA251245, PCF Challenge Award, National Comprehensive Cancer Network (NCCN)/Astellas Pharma Global Development Award, Michigan Prostate SPORE (NCI P50 CA186786), and Joint Institute for Translational and Clinical Research. We thank the cloud computing resources provided by the Center for Computational Research (CCR) at the University at Buffalo (<http://hdl.handle.net/10477/79221>). We thank Mr. W. Tian and Dr. L. Yan for assistance in survival analysis, Drs. J. Yong and H. He for bioinformatics assistance and all other Tang lab members for helpful discussions and suggestions. We apologize to the colleagues whose work was not cited due to space constraint.

Author information

Authors and Affiliations

Department of Pharmacology and Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA

Xiaozhuo Liu, Yibing Ji, David W. Goodrich, Kiera Rycaj, Anmbreen Jamroze, Jason S. Kirk & Dean G. Tang

Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA

Eduardo Cortes, Han Yu, Qiang Hu, Song Liu & Jianmin Wang

Department of Biostatistics, University at Buffalo, Buffalo, New York, USA

Eduardo Cortes

Veracyte Inc, San Diego, California, USA

Kathy Zhao, Julian Ho, Yang (Seagle) Liu & Elai Davicioni

Department of Radiation Oncology, University of California, San Francisco, California, USA

Felix Y. Feng

Department of Hematology and Oncology, Department of Internal Medicine, Rogel Cancer Center, University of Michigan, Ann Arbor, Michigan, USA

Joshi J. Alumkal

University Hospitals Seidman Cancer Center and Case Western Reserve University, Cleveland, Ohio, USA

Daniel E. Spratt

South Australian Immunogenomics Cancer Institute, University of Adelaide, Adelaide, SA 5000, Australia
Christopher J. Sweeney

School of Biomedical Sciences, Hunan University, Changsha, China
Zou Cheng & Dingxiao Zhang

Department of Medicine, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA
Gurkamal Chatta & Igor Puzanov

Department of Cancer Genetics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA
Kent L. Nastiuk

Authors' Contributions

X. Liu: Conceptualization, data curation, formal analysis, statistical analysis, validation, visualization, investigation, writing—original draft, writing—review and editing, equal contribution as co-first author and co-senior author. **E. Cortes:** Data curation, formal analysis, statistical analysis, validation, investigation, writing—original draft, writing—review and editing, equal contribution as co-first author. **Y. Ji:** Study conception and design, data curation, validation, visualization. **K. Zhao:** Data curation, formal analysis, validation, visualization. **J. Ho:** Data curation, formal analysis, validation, visualization. **Y. Liu:** Data curation, formal analysis, validation, visualization. **E. Davicioni:** Data curation, investigation, writing—review and editing. **F. Feng:** Writing—review and editing. **J. Alumkal:** Writing—review and editing. **D. Spratt:** Writing—review and editing. **C. Sweeney:** Writing—review and editing. **H. Yu:** Data curation, statistical analysis, writing—review and editing. **Q. Hu:** Data curation, statistical analysis, investigation. **Z. Cheng:** Data curation, investigation, writing—review and editing. **D. Zhang:** Data curation, investigation, writing—review and editing. **G. Chatta:** Data curation, investigation, writing—review and editing. **K. Nastiuk:** Data curation, investigation, writing—review and editing. **D. Goodrich:** Obtained main funding for this project, data curation, writing—review and editing. **K. Rycaj:** Investigation, Data curation, visualization, writing—review and editing. **A. Jamroze:** Investigation, writing—review and editing. **J. Kirk:** Data curation, investigation, writing—original draft, writing—review and editing. **I. Puzanov:** Writing—review and editing. **S. Liu:** Study conception and design, supervision, writing—review and editing, funding acquisition, equal contribution as co-senior author. **J. Wang:** Study conception and design, supervision, writing—review and editing, funding acquisition, equal contribution as co-senior author. **D. Tang:** Conceptualization, supervision, writing—original draft, writing—review and editing, funding acquisition, equal contribution as co-senior author, the lead corresponding author.

Ethics declarations

Competing interests

All authors declare no competing interests relative to this work.

Figure Legends

Figure 1. *Pri-PCa shows concordant increases of AR-A and Stemness and both are decreased by nADT.*

A-C, PCa has increased AR-A as well as the Stemness compared to matched benign tissues. Pairwise comparisons of AR mRNA levels, canonical AR-A, and Stemness scores showing significant increases in primary tumors (T) as compared to the matched adjacent benign prostate tissues (N) from the same individual in TCGA-PRAD cohort (n=51; **A**), and Wyatt 2014 cohort (n=12; **B**). In the Long 2020 cohort (n=10; **C**), increases in AR-A and Stemness reached statistical significance whereas the increase in AR mRNA levels approached statistical significance ($P = 0.06$).

D-F, nADT decreases both AR-A and Stemness. Pairwise comparison of AR mRNA levels, AR-A, and Stemness showing significant decrease of both AR-A and Stemness after nADT (Post-nADT) as compared to the matched samples before nADT (Pre-nADT) from the same individual (**D**, Rajan 2014 cohort; **E**, Sharma 2018 cohort; **F**, Long 2020 cohort).

In all plots, sample sizes are indicated (see also [Supplementary Table S1](#) for cohort and RNA-seq data access information). Within the plots, results are shown as the mean \pm SD and pairs are linked with a grey line. Significance was calculated by two-tailed paired Student's *t*-test (ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$).

Figure 2. *Continual increases in Stemness with declining AR-A during spontaneous PCa progression.*

A, The pie chart showing sample sizes in different Gleason grades in the TCGA untreated Pri-PCa cohort.

B-D, Progression of treatment-naïve Pri-PCa is accompanied by decreasing AR-A but increasing Stemness. Data is presented as mean \pm SD. Shown on top are the Jonckheere-Terpstra's trend (J-T) test results (wedges).

E-F, Progression in Pri-PCa is accompanied by a gradual loss of the positive correlation between canonical AR-A and Stemness. Scatter plots illustrating the correlation between AR-A and Stemness in (**E**) combined Pri-PCa cases in TCGA with an overall linear regression model and (**F**) individual GS grades with corresponding linear regressions showing decreasing slopes from low to high GS grades by J-T test. The slopes of best-fit regression line are also shown.

G, Merged scatter plot showing the correlation between AR-A and Stemness across the 4 GS PCa with each GS represented by a distinct color. The correlation becomes progressively weaker from GS6 to GS9/10, transitioning from positive to nearly zero. The correlation coefficients (*r*) and *p*-values for each GS are as follows: GS6 ($r = 0.56$, $P < 0.0001$), GS7 ($r = 0.37$, $P < 0.0001$), GS8 ($r = 0.48$, $P = 0.022$), and GS9/10 ($r = -0.02$, $P = 0.89$). Linear regression lines are overlaid for each GS, demonstrating the decline in the correlation co-efficiency as the GS increases.

All statistical analyses were performed using R and statistical significance was assessed using one-way analysis of variance followed by Tukey's multiple-comparison test (B-D). J-T trend test in B-D and F was used to calculate the statistical significance of the trend across different groups. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant.

Figure 3. *Therapies and metastatic progression drive further divergence between Stemness and canonical AR-A with mCRPC showing the highest Stemness but the lowest AR-A.*

A-D, Hormonal treatment further drives down AR-A but significantly upregulates Stemness. **A-C,** mCRPC displays the lowest pro-differentiation AR-A but highest Stemness. Shown are AR mRNA levels (**A**), the canonical AR-A (**B**) and Stemness (**C**) in tumor-adjacent benign tissues (N), treatment-naïve Pri-PCa and advanced Pri-PCa with post-surgery adjuvant ADT (Tx) in TCGA, and mCRPC (SU2C 2015). Sample sizes are indicated in parentheses. Shown on top are the Jonckheere-Terpstra's (J-T) trend test results (wedges). **D,** Stemness inversely correlates with AR-A in mCRPC. Shown are scatter plots with best-fit linear regression lines illustrating the correlation between AR-A and Stemness with Pearson correlation coefficient (*r*), the slopes of lines and J-T test results of slopes indicated below.

E-F, PCa progression is accompanied by loss of AR-A (**E**) but gain of Stemness (**F**) across PCa states. Shown are samples representing the spectrum of PCa development and progression, from normal (organ donors in GTEx) to (tumor-adjacent) benign tissues, to (treatment-naïve) Pri-PCa, aggressive Pri-PCa treated with adjuvant ADT (Pri-PCa/ADT), mCRPC, and, finally, to (clonally-derived) PCa models including PDX, xenografts and cells lines. From Left to right: Normal (n=116): Normal prostates from GTEx; Benign (n=52): tumor-adjacent benign tissue from TCGA-PRAD; Pri-PCa (n=422): treatment-naïve localized PCa from TCGA-PRAD; Pri-PCa/ADT (n=70): advanced Pri-PCa treated with adjuvant hormone therapy from TCGA-PRAD; mCRPC (n=59): mCRPC from GSE126078; mCRPC (n=135): mCRPC from SU2C 2015 cohort; mCRPC-Adeno (n=34): adenocarcinoma mCRPC from Beltran 2016 cohort; mCRPC-NE (n=15): neuroendocrine mCRPC from Beltran 2016 cohort; mCRPC (n=42): mCRPC from Westbrook 2022; mCRPC (n=25): mCRPC from Alumkal 2020; PDX-LuCaP (n=39): PCa PDXs from GSE126078; PCa cell lines (n=8): PCa cell lines from CCLE; XG-LAPC9 (n=10) and XG-LNCaP (n=12): PCa xenografts from GSE88752. Statistical significances between groups were shown in inverted pyramid grid plot as indicated by different colors. Student's *t*-test was used to compare each group and adjusted *p*-value for multiple comparisons was done by Tukey's range test. See [Supplementary Table S4](#) for the summary of statistical comparisons. J-T trend test was used to assess the statistical significance of the trend across groups. See also [Supplementary Table S1](#) for cohort and RNA-seq data access information.

G, Scatter plots in the Bolis 2021 integrated cohort (normal n=174, Pri-PCa n=714, mCRPC n=335) showing the changing correlation between AR-A and Stemness across disease stages: positive in normal to weaker correlations in Pri-PCa to negative correlations in mCRPC. Regression lines are color-coded, and the slope of the best-fit regression line and Pearson's *r* are provided: yellow for normal (Pearson's *r*=0.42, *P* < 0.0001; slope: 43.87, 95%CI, 29.57 to 58.18); blue for Pri-PCa (*r*=0.32, *P* < 0.0001; slope: 15.04, 95%CI, 11.76 to 18.32); and green for mCRPC (*r*=-0.15, *P* = 0.005; slope: -17.25, 95%CI, -29.24 to -5.26).

H, Violin plot showing increasing Stemness among the NCCN Risk Groups containing 82470 microarray samples with localized disease in GRID. J-T trend test was used to assess the statistical significance of the trend across groups.

I-M, Increasing Stemness during PCa progression in the Taylor 2010 cohort consisting of tumor-adjacent benign tissue (Normal; n=29), Pri-PCa (n=131) and metastatic mPCa (n=19). The AR-A significantly increased in Pri-PCa vs. Normal and significantly decreased in mPCa compared to Pri-PCa (**I**) whereas the Stemness continued to increase from Benign to Pri-PCa to mPCa (**J**). **K-M**, Scatter plots illustrating the correlations between AR-A and Stemness in different contexts with corresponding linear regression lines: **K**, Positive correlation in normal (Pearson's *r*=0.69, *P* < 0.0001); **L**, a gradual loss of positive correlation between Stemness and AR-A during PCa progression (Pearson's *r*=0.18, *P* = 0.043); **M**, a loss of positive correlation between the Stemness and AR-A at more advanced mPCa stage (Pearson's *r*=0.18, *P* = 0.47).

All data in dot plots and violin plots are shown as mean ± SD. Statistical significance was determined using one-way analysis of variance (ANOVA) followed by Tukey's multiple-comparison test. Jonckheere-Terpstra's trend (J-T) test was used to calculate the statistical significance of the trend across different groups. Pearson correlation coefficients (Pearson's *r*) were used to calculate the correlations between the Stemness and AR-A. Note: ns, not significant; *, *P* < 0.05; **, *P* < 0.01; ***, *P* < 0.001; ****, *P* < 0.0001.

Figure 4. High Stemness is associated with poor patient overall survival.

A-C, Kaplan-Meier plots showing that high Stemness associates with worse overall survival in PCa patients from Spratt 2017 (**A**), Tosoian 2020 (**B**), and CHAARTED 2021 (**C**) cohorts. *P* value was determined using the log-rank test.

D, Stemness increases across patient groups with higher Decipher genomic classifier scores. The three cohorts were arranged by their median Decipher genomic classifier scores, and the J-T trend test was used to assess statistical significance of this trend.

Figure 5. Stemness-high PCa are associated with aggressive molecular subtypes and a 12-gene PCa-stem signature prognosticates poor patient survival.

A, Schematic illustrating the profiling strategy used to assess associations between Stemness and molecular features in PCa. Patients were ranked by Stemness score and stratified into Stemness-high (top 33%) and Stemness-low (bottom 33%) groups for downstream analyses. Genome-wide functional analyses were performed using transcriptomic, genomic, and clinical data.

B-C, Volcano plots showing differentially expressed genes (DEGs) between Stemness-high and Stemness-low groups in Pri-PCa (**B**) and mCRPC (**C**) cohorts. Red and blue indicate upregulated and downregulated DEGs, respectively. Genes with FDR > 0.05 are shown in gray. Sample sizes (n) of each group are indicated. See [Supplementary Table S3](#) for complete DEG lists.

D Venn diagram showing the 12-gene “PCa-Stem signature” derived from overlapping upregulated DEGs (FC ≥ 2, FDR < 0.05) in both Pri-PCa and mCRPC.

E-F, GSEA showing low Stemness is associated with immune signaling and inflammatory response (**E**) while high Stemness in both cohorts is enriched for embryonic stem cell traits, aggressiveness, DNA repair, MYC activation, and mTORC1 signaling (**F**). All enrichments are significant ($P < 0.05$, FDR < 0.05).

G-H, PAM50 subtyping showing a higher frequency of LumB subtype in Stemness-high versus Stemness-low samples in Pri-PCa (**G**) and mCRPC (**H**). (****, $P < 0.0001$, χ^2 test).

I-L, GSEA showing enrichment of the aggressive PCS1 gene expression signature in Stemness-high groups and the less aggressive PCS3 signature in Stemness-low groups in Pri-PCa (**I**, **K**) and mCRPC (**J**, **L**). The Prostate Cancer Subtype (PCS) classification system (45,46) stratifies tumors into three molecular subtypes based on transcriptomic features: PCS1 (most aggressive), PCS2 (intermediate), and PCS3 (least aggressive). See [Supplementary Table S2](#) for PCS gene signature.

M-R, GSEA showing AR signature enrichment in Pri-PCa Stemness-high (**M**), NEPC signature enrichment in mCRPC Stemness-high (**N**), and shared enrichment of cell cycle (**O**, **P**) and lineage plasticity (**Q**, **R**) programs in both cohorts. See [Supplementary Table S5](#) for gene sets.

S-V, Kaplan-Meier analyses showing that the 12-gene PCa-Stem signature predicts worse progression-free survival (PFS) in TCGA Pri-PCa (**S**) and worse overall survival (OS) in SU2C mCRPC (**U**). Multivariable Cox models adjusting for clinicopathologic variables confirm the PCa-Stem-High subgroup as independently prognostic for PFS (**T**) and OS (**V**).

Figure 6. Stemness-high PCa is associated with higher genomic instability.

A-C, Box plots showing that Stemness-high subgroup in both Pri-PCa (n = 141) and mCRPC (n = 89) displays a higher fraction of altered genome (**A**), higher TMB (**B**), and no age difference compared to Stemness-low cases (**C**). Wilcoxon rank-sum test was used to determine significance (****, $P < 0.0001$; ns, not significant). Statistical significance was determined using the two-sided Wilcoxon rank-sum test. Box plots represent the median, interquartile range (IQR), and whiskers extending to 1.5× the IQR.

D-F, Comparisons of genomic alteration frequencies (Alt. Freq) in cancer-associated genes between Stemness-high and Stemness-low subgroups in TCGA Pri-PCa (**D**, **F**) and SU2C mCRPC (**E**, **F**) cohorts. Statistical significance was determined by two-sided Fisher’s Exact test (*, $P < 0.05$).

Figure 7. MYC activation contributes to increased Stemness during spontaneous PCa progression.

A-B, Pairwise comparisons showing increased MYC mRNA levels and MYC activity (MYC-sig) in primary tumors (T) versus matched adjacent benign tissues (N) in TCGA-PRAD cohort (n=51; **A**) and Wyatt (n=12; **B**) cohorts. Data is shown as mean ± SD; paired samples are linked by grey lines. Significance was determined using two-tailed paired Student’s t-test (****, $P < 0.0001$).

C-H, MYC activity increases during PCa progression, peaking in mCRPC. In treatment-naïve Pri-PCa (**C**), MYC-sig is significantly increased in all-grade (GS) tumors as compared to tumor-adjacent benign tissues (N) with the most advanced GS9/10 tumors exhibiting the highest MYC-sig. In **D**, MYC-sig increases from normal to treatment-naïve Pri-PCa and further in advanced Pri-PCa treated with adjuvant ADT (Tx) in TCGA-PRAD, while mCRPC shows the highest MYC-sig in both SU2C 2015 (**D**) and Taylor 2010 (**E**) cohorts. Data is presented as mean \pm SD. Statistical significance was assessed by one-way ANOVA with Tukey's multiple-comparison test (**C-E**). J-T trend test was used in (C, D) to assess the significance of progressive increase. **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant. **F-H**, Scatter plots showing positive correlations between MYC activity and Stemness in the Taylor 2010 cohort across different stages. Pearson's r and P -values are indicated.

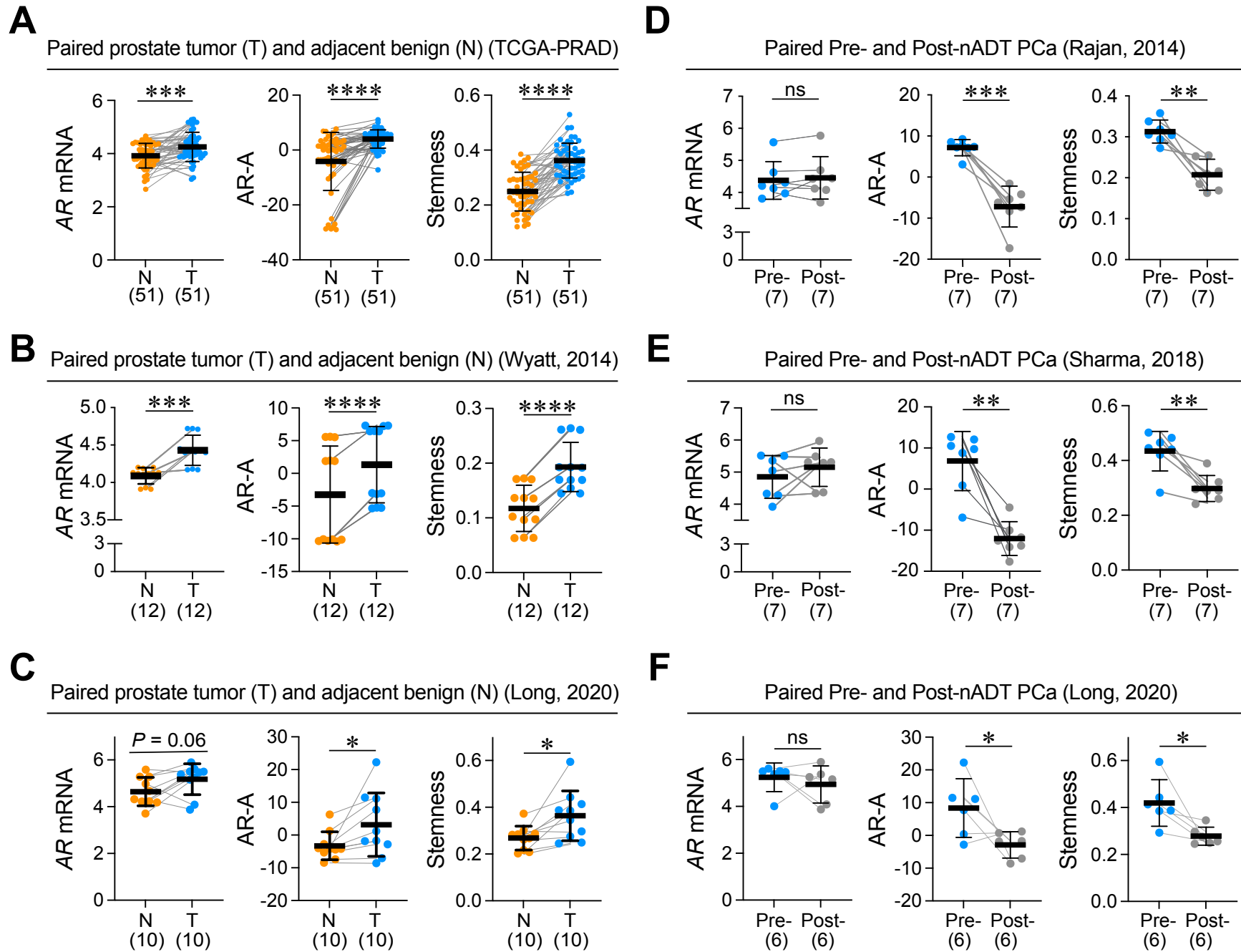


Figure 1

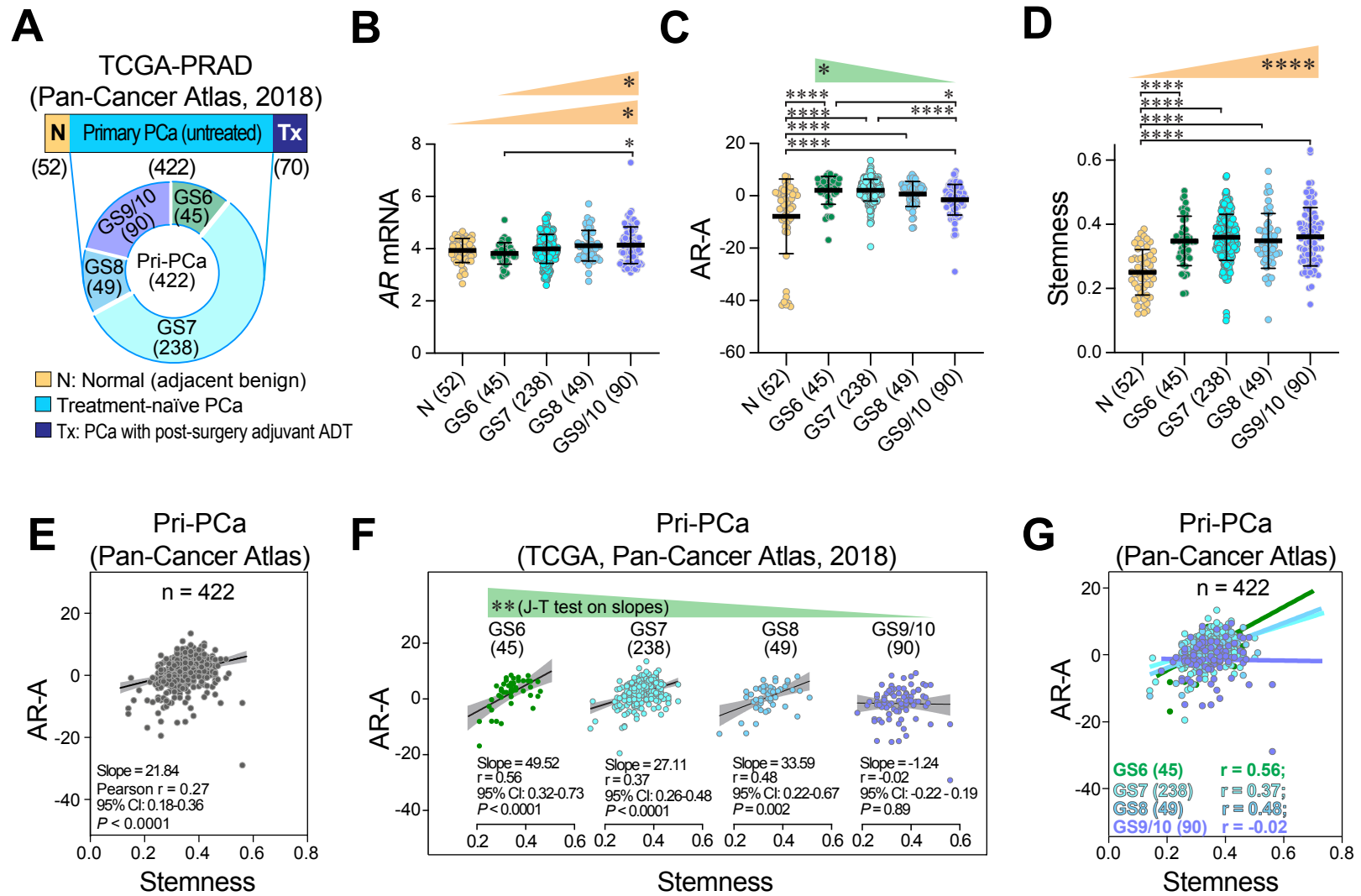


Figure 2

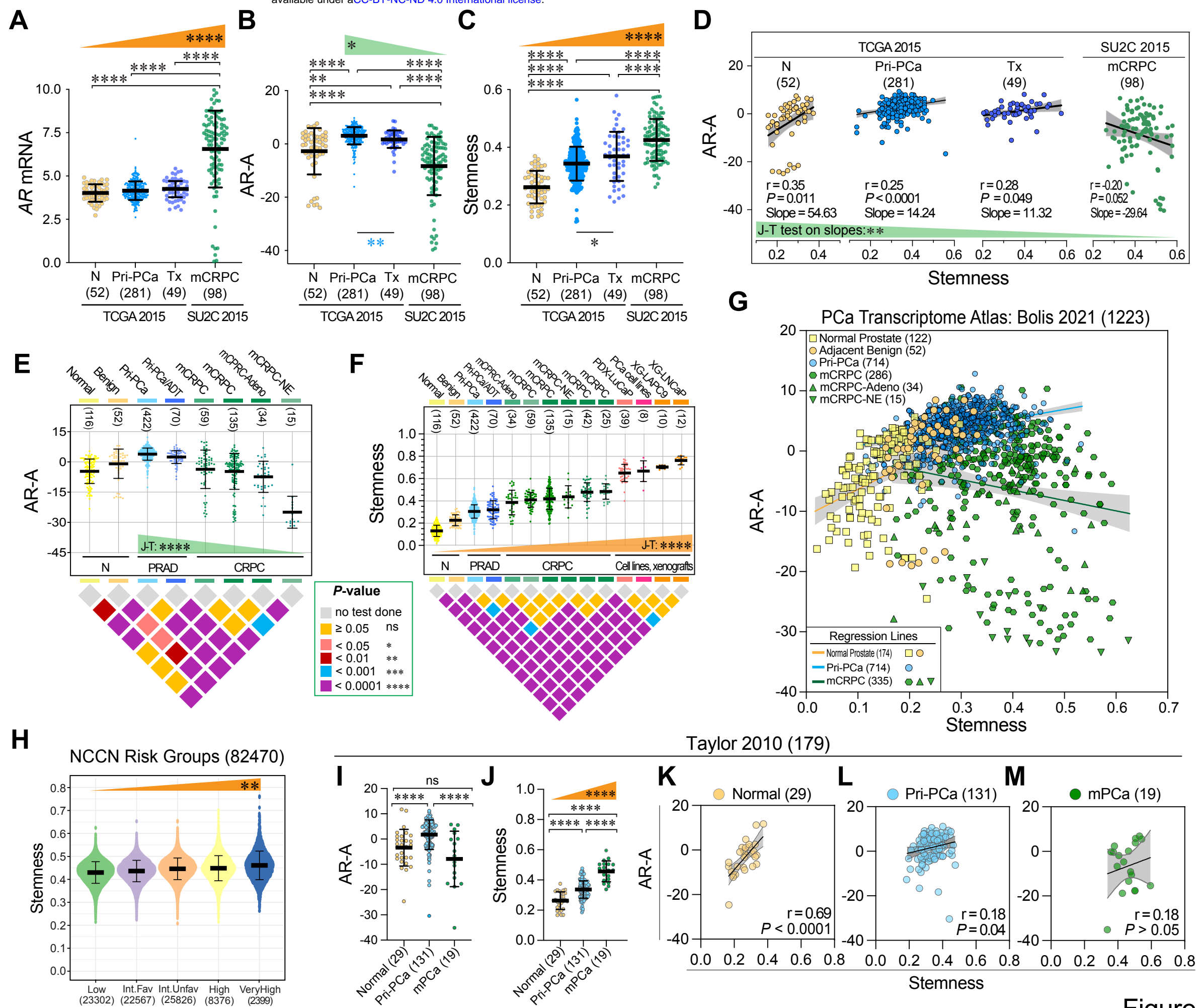


Figure 3

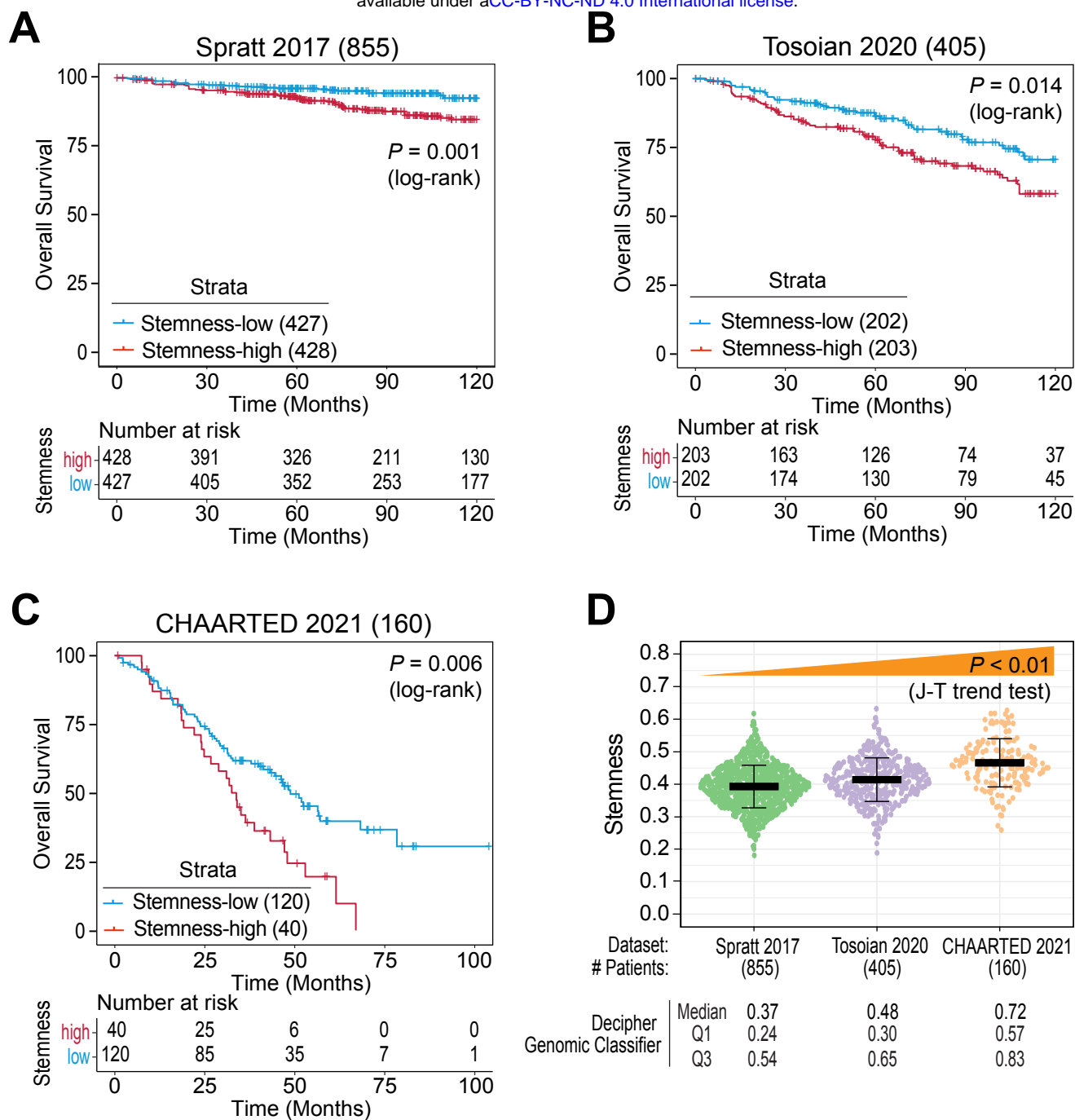


Figure 4

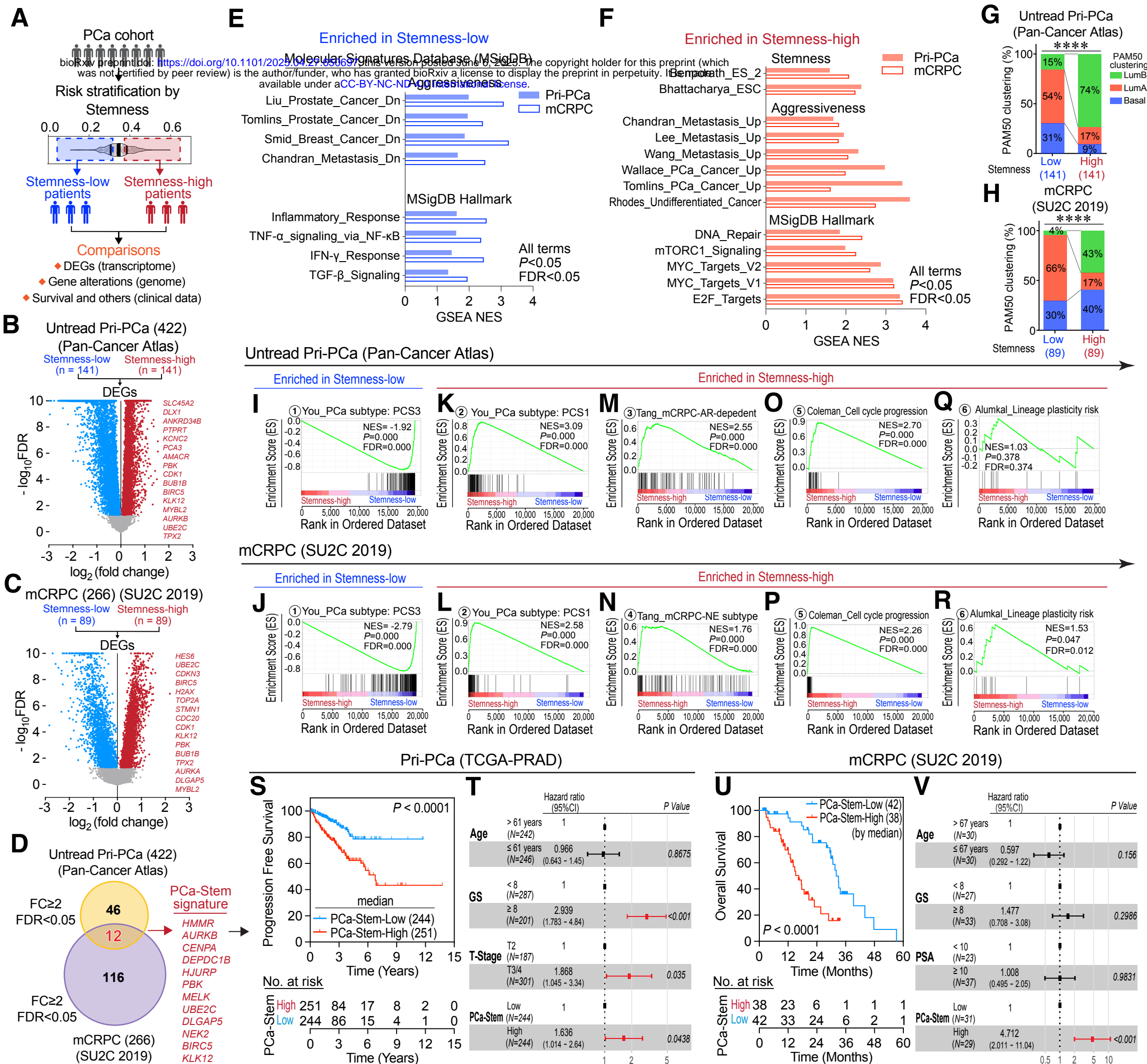


Figure 5

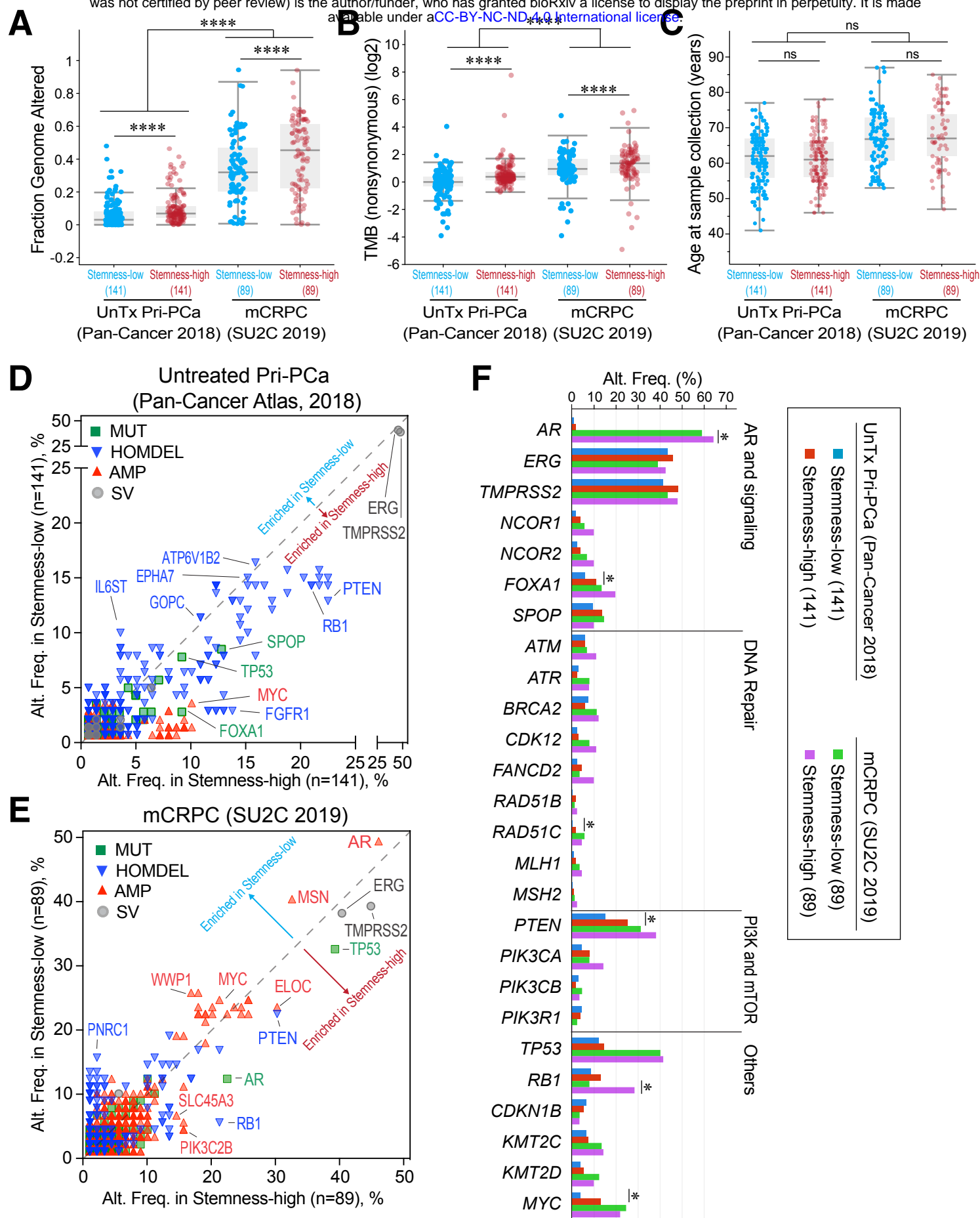


Figure 6

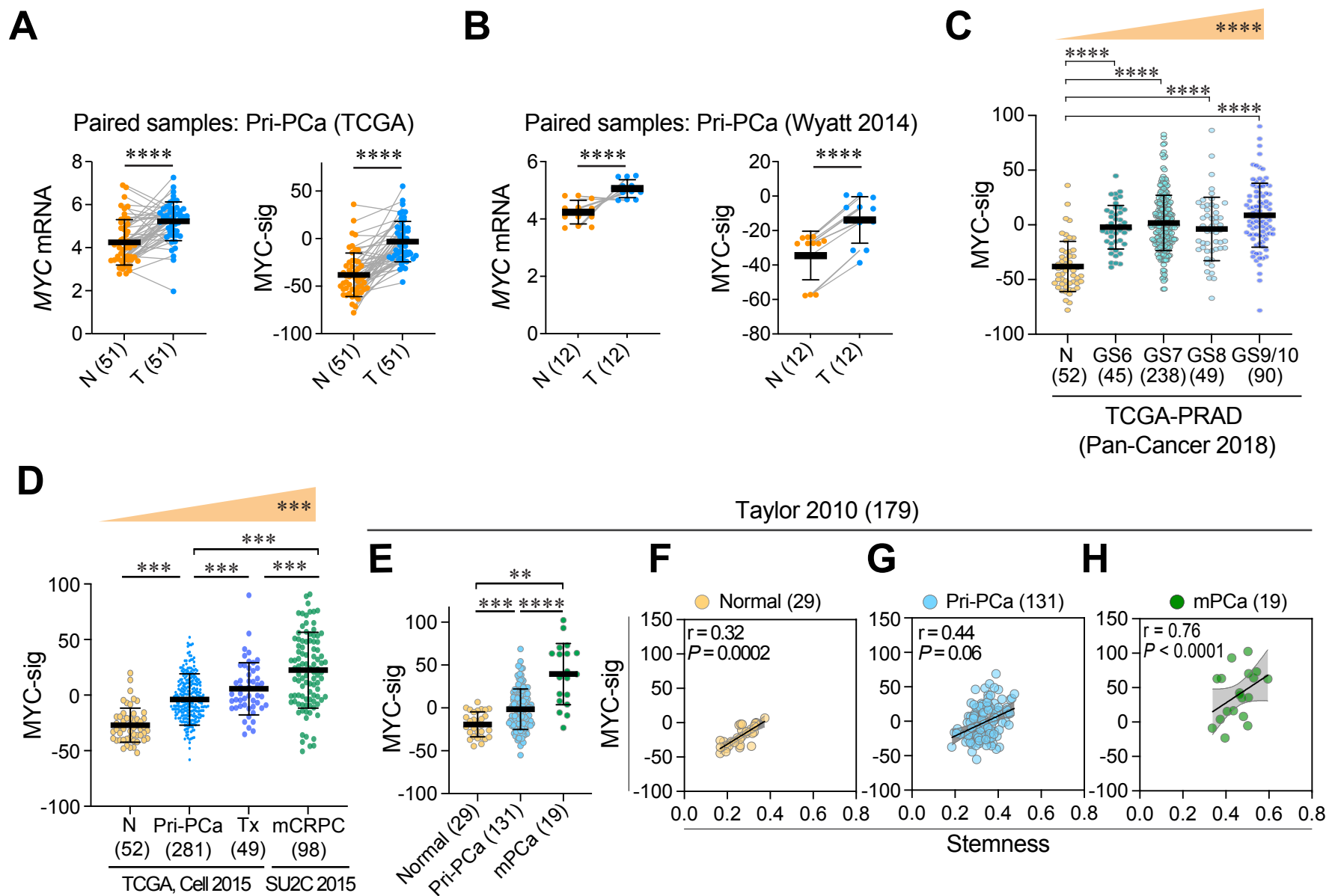


Figure 7

Supplementary Information Content

Increasing Stemness Drives Prostate Cancer Progression, Plasticity, Therapy Resistance and Poor Patient Survival

Liu X, Cortes E, et al.

Supplementary Table S1. Summary of datasets used in this study

Supplementary Table S2. Summary of gene signatures used in this study

Supplementary Table S3. DEGs identified in Stemness-high versus Stemness-low subgroups in the Pri-PCa (TCGAPRAD) cohort and in the mCRPC (SU2C 2019) cohort (*six sheets)

Supplementary Table S4. Summary of statistical significance of differences using Tukey's multiple comparisons test (*three sheets)

Supplementary Table S5. Summary of gene signatures and gene sets used in Gene Set Enrichment Analysis

Related to Fig. 1

Supplementary Fig. S1. Schematic illustration of bioinformatic datasets and signature score pipelines used for comparisons across the spectrum of PCa evolution.

Supplementary Fig. S2. Transcriptome-based signature scores for canonical (pro-differentiation) AR signaling activity (AR-A).

Supplementary Fig. S3. Prostate luminal cell compartment harbors higher AR-A but lower Stemness as compared to prostate basal cell compartment.

Supplementary Fig. S4. nADT decreases both AR-A and Stemness.

Related to Fig. 3

Supplementary Fig. S5. Increased Stemness represents a 'universal' feature of PCa progression and aggressiveness.

Related to Fig. 4

Supplementary Fig. S6. High Stemness correlates with poor patient survival.

Related to Fig. 5

Supplementary Fig. S7. PCa-Stem signature and gene expression correlate with PCa progression.

Related to Fig. 6

Supplementary Fig. S8. Unique genomic features of Stemness-high PCa.

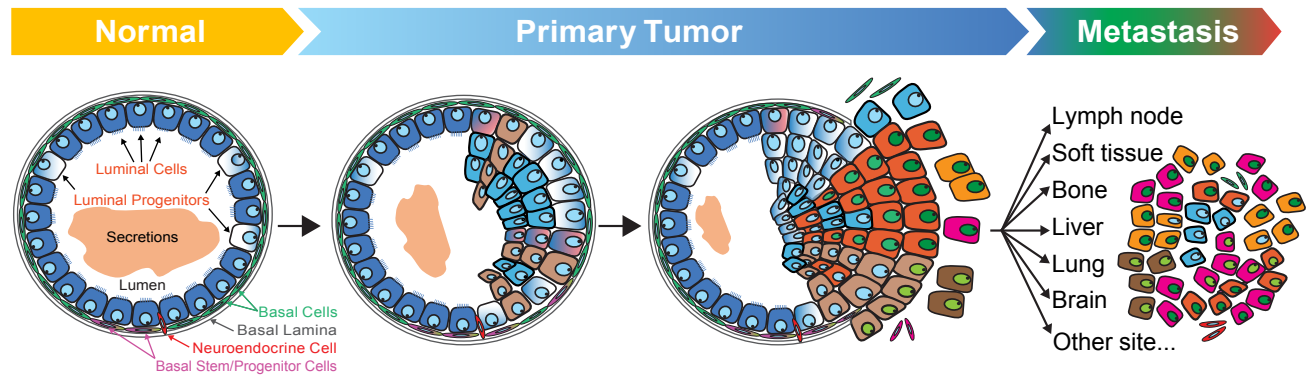
Related to Fig. 7

Supplementary Fig. S9. Lack of effect of nADT on MYC activity.

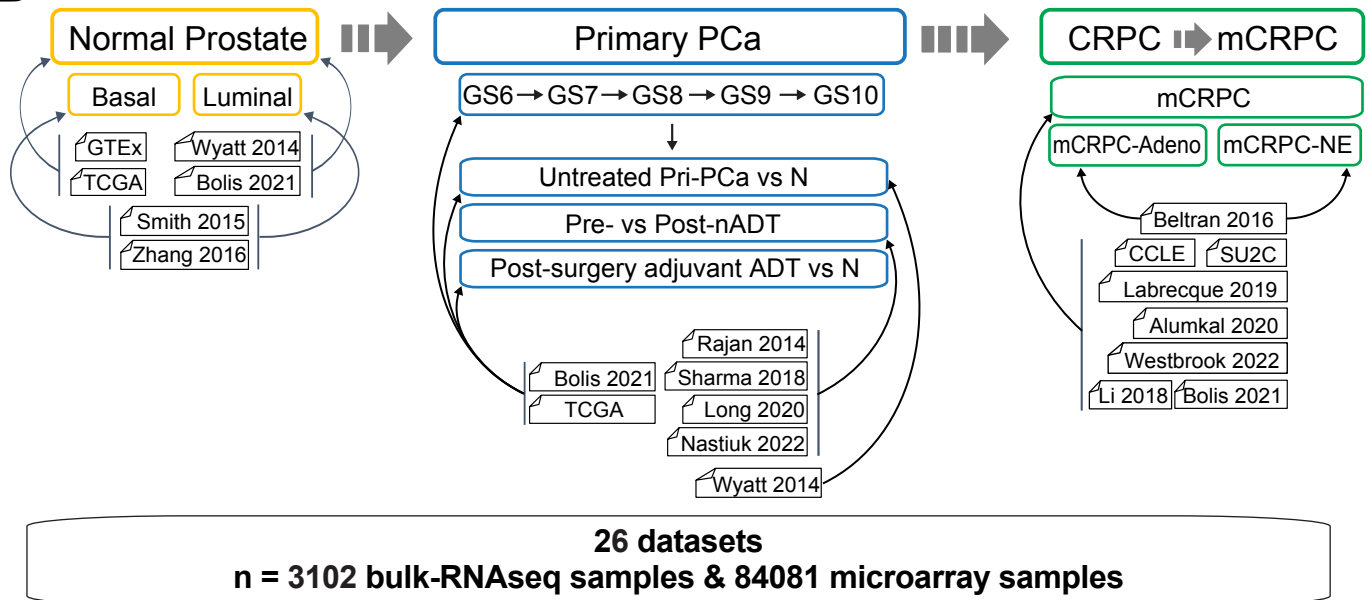
Supplementary Fig. S10. Schematic summary of PCa progression and aggressiveness accompanied by gain of oncogenic Stemness and loss of prostatic epithelium differentiation defined by canonical AR signaling.

Supplementary Figure S1

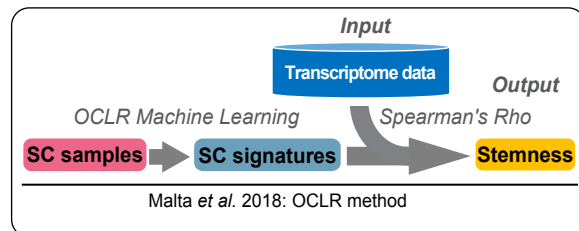
A



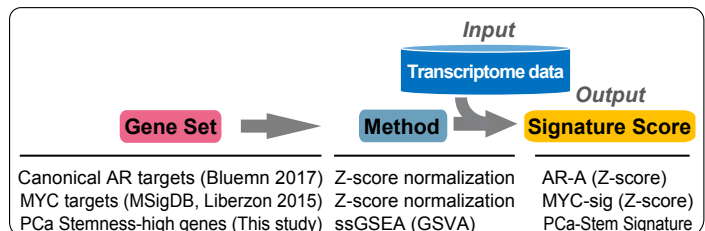
B



C



D



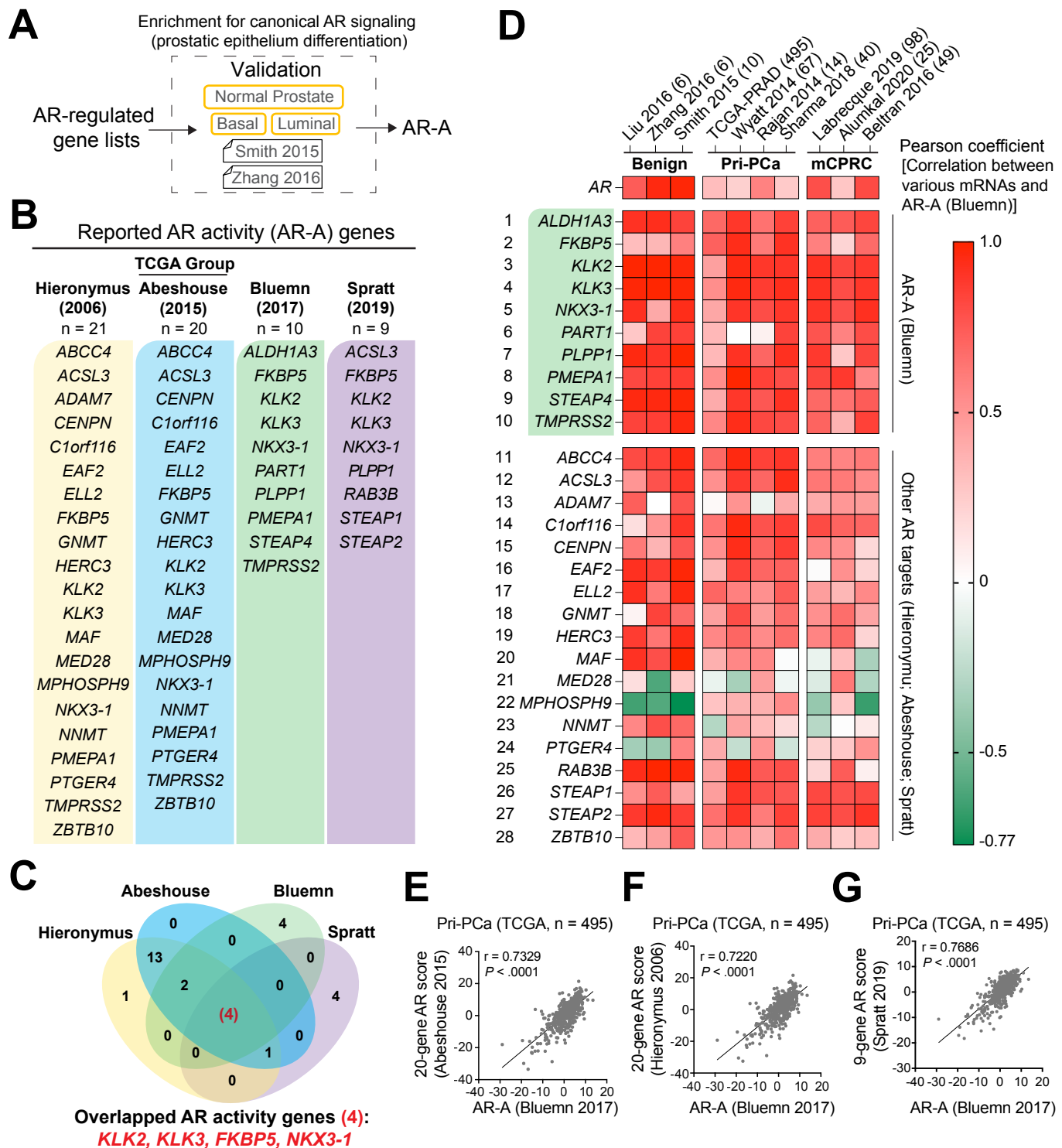
Supplementary Figure S1. Schematic illustration of bioinformatic datasets and signature score pipelines used for comparisons across the spectrum of PCa evolution.

(A-B) Schematic illustrating the spectrum of PCa development, therapy resistance, and metastatic progression (A), and related datasets used for comparisons in different contexts (B). Panel A illustrates the PCa progression continuum from Normal Prostate (Normal) through Primary PCa (Pri-PCa), to Castration-Resistant PCa (CRPC) and Metastatic CRPC (mCRPC), with each stage marked by specific changes in cellular composition and architectural organization. Panel B details the datasets utilized for comparative analyses across these stages, with arrows indicating the relationship of each dataset to specific PCa stages. In total, 26 datasets were used in the current analysis, including 3102 bulk-RNAseq samples and 84081 microarray samples. See [Supplementary Table S1](#) for dataset details. (A is adapted from ref. 7).

Detailed Description of PCa Stages:

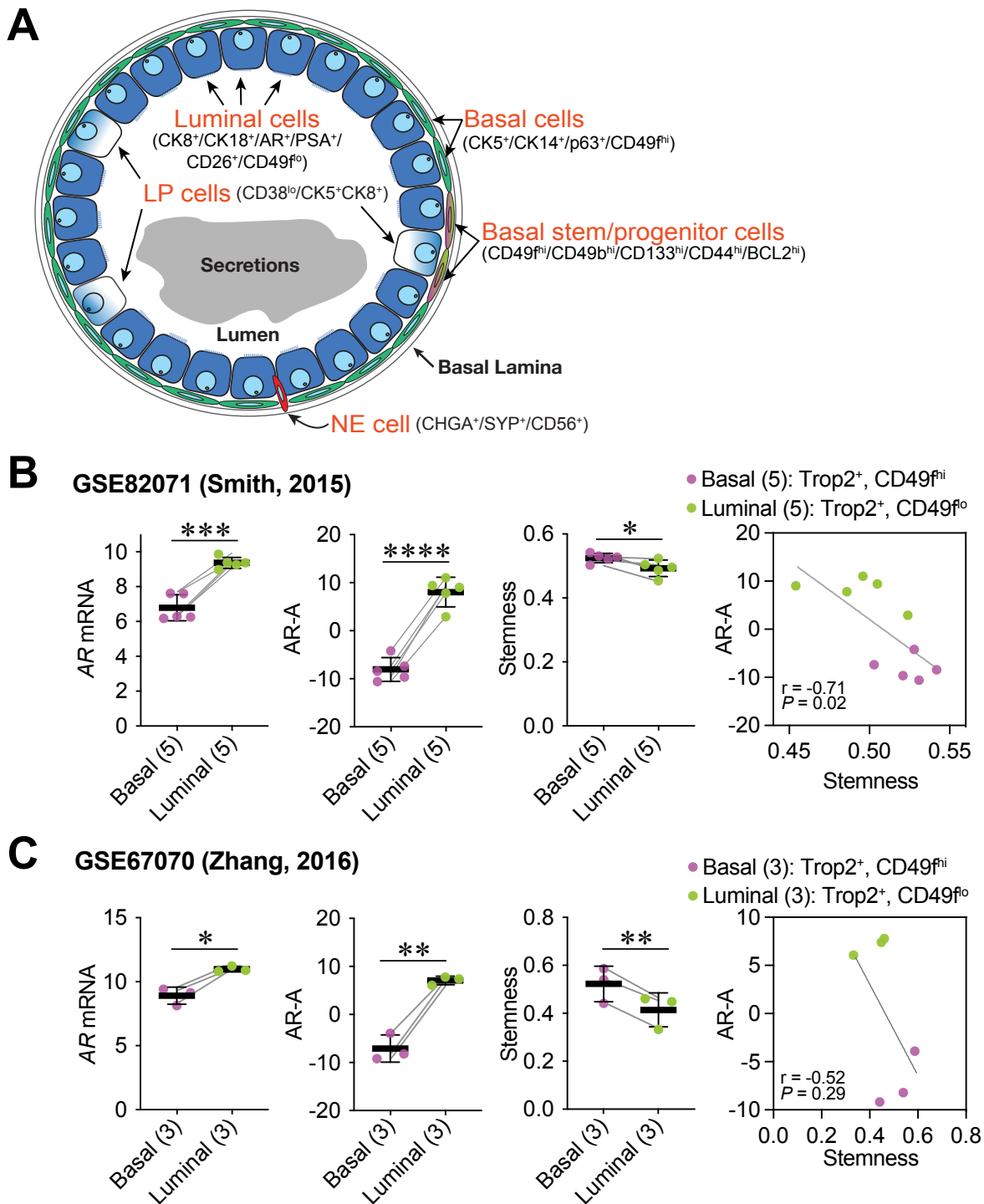
- **Normal Prostate:** Depicted as a glandular organ with a pseudostratified two-layer epithelium consisting of basal and luminal cells, interspersed with rare neuroendocrine cells. Luminal cells express cytokeratins CK8 and CK18, androgen receptor (AR), and prostate-specific antigen (PSA), and represent the primary secretory cells. Also depicted are rare luminal progenitor cells within the luminal cell layer, which represent less than 2 % of the luminal cell population and are characterized by the co-expression of CK5 and CK8 and a low level of CD38 (reviewed in ref. 8). Basal cells express CK5 and CK14, interfacing with the stroma through the basal lamina. The basal cell compartment harbors stem/progenitor cells capable of bidirectional differentiation (ref. 8). Also see an enlarged depiction of a normal human prostatic gland in [Supplementary Fig. S3A](#).
- **Localized Pri-PCa:** Progresses from well-differentiated structures (low Gleason score) to poorly differentiated configurations (high Gleason score), indicating a shift towards a more stem-like state with loss of glandular architecture. This transition involves a reduction in basal cells and an expansion of progenitor-like cells, with the combined Gleason score (GS) reflecting the architectural patterns of cancerous glands. Gleason pattern 1 represents the most well-differentiated glandular structures and Gleason pattern 5 consists of the most poorly differentiated cells lacking glandular structures. Given that PCa is often multifocal, the combined GS is the sum of the two most prevalent patterns, with $GS \leq 7$ representing low-grade PCa and $GS \geq 8$ representing high-grade PCa. Illustrated here are two representative tumor glands with a decrease/loss in basal cells and expansion of luminal progenitor-like cells.
- **Metastatic CRPC:** Features increased cellular heterogeneity and the presence of cells with stem-like and neuroendocrine-like characteristics, which drive therapy resistance and disease progression.

(C-D) Schematic presentation of transcriptome-based stemness quantification method mRNAsi (C, ref. 6), and transcriptome-based signature scores for canonical AR signaling activity levels, MYC signaling activity levels and PCa-Stem signature (D). See [Supplementary Table S2](#) for gene signature information.



Supplementary Figure S2. Transcriptome-based signature scores for canonical (pro-differentiation) AR-A.

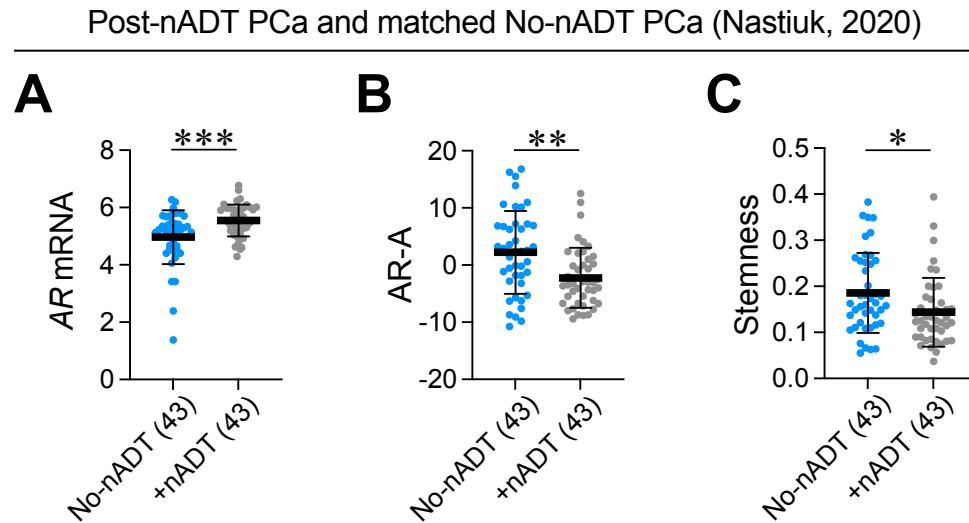
- (A) Workflow for validating canonical AR signaling activity (AR-A) signatures from different AR-regulated gene lists.
- (B-C) Comparison of AR downstream signaling targets from four studies. The 4 genes commonly shared in all four AR-A signatures are highlighted (C). Genes were listed using updated HGNC symbols (the HUGO gene nomenclature from the Human Genome Organization). Note that in the original publications, some AR-A genes were denoted by their (gene/protein) aliases, as exemplified by *CENPN* (*BM039*), *C1orf116* (*SARG*), *KLK3* (*PSA*), *PLPP1* (*PPAP2A*), and *PMEPA1* (*TMEPAI*).
- (D) AR-A signature score highly correlates with mRNA expression levels of canonical AR signaling targets in various PCa related cohorts. See [Supplementary Table S1](#) for dataset details.
- (E-G) Strong positive linear relationships between the Bluemn AR-A and the other three AR-A signatures (Hieronymus 2006; Abeshouse 2015, Spratt 2019) using primary PCa cohort (TCGA-PRAD). r indicates Pearson correlation coefficient.



Supplementary Figure S3. Prostate luminal cell compartment harbors higher AR-A but lower Stemness as compared to prostate basal cell compartment.

(A) Depiction of the major cell types in human prostatic glands. Both basal and luminal cell compartments harbor differentiated as well as less mature stem/progenitor cells. LP, luminal progenitor; NE, neuroendocrine. Cellular markers are highlighted for different cell types. Adapted from ref. 8.

(B-C) Prostate luminal cells possess higher AR-A but lower Stemness as compared to prostate basal cells, and Stemness inversely correlates with AR-A in the context of prostatic basal and luminal cells. Results are shown as mean \pm SD. The pairs are linked with a grey line. Significance was calculated by two-tailed paired Student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$). Pearson correlation coefficients (Pearson's *r*) were used to calculate the correlations between Stemness and AR-A.

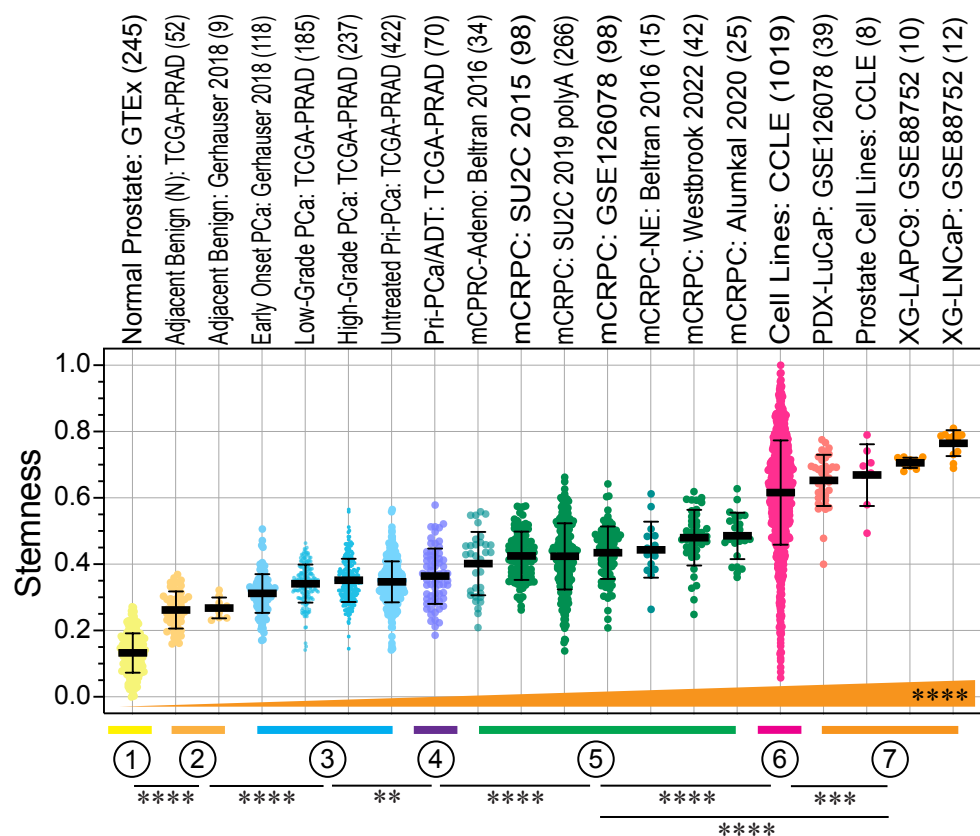


Supplementary Figure S4. nADT decreases both AR-A and Stemness.

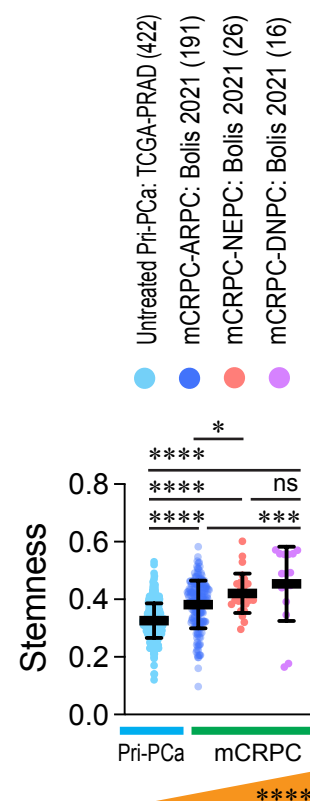
Matched comparison of AR mRNA levels (A), AR-A (B), and Stemness (C) showing significant decrease of both AR-A and Stemness in neo-adjuvant ADT (+nADT) group as compared to the samples without nADT (No-nADT) from stage-, tumor grade-, age-matched PCa individuals (RPCI Nastiuk cohort). Sample sizes are indicated. See [Supplementary Table S1](#) for cohort and RNAseq data information. Within the plots, results are shown as the mean \pm SD. Significance was calculated by two-tailed unpaired Student's *t*-test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$).

Supplementary Figure S5

A

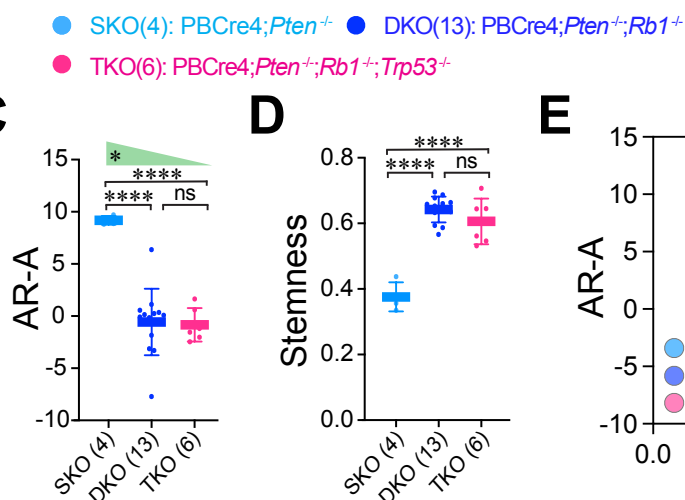


B

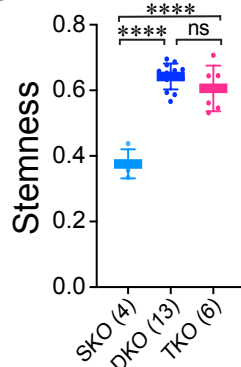


PCa GEMMs: GSE90891 (23)

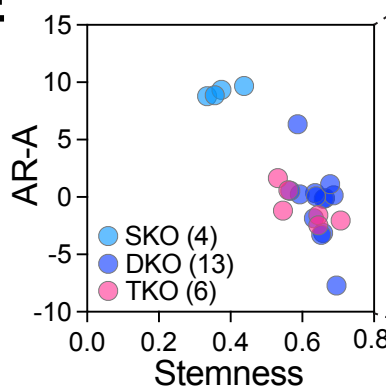
C



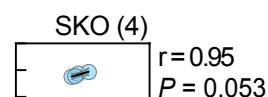
D



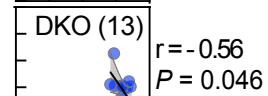
E



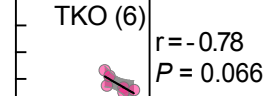
F



G



H



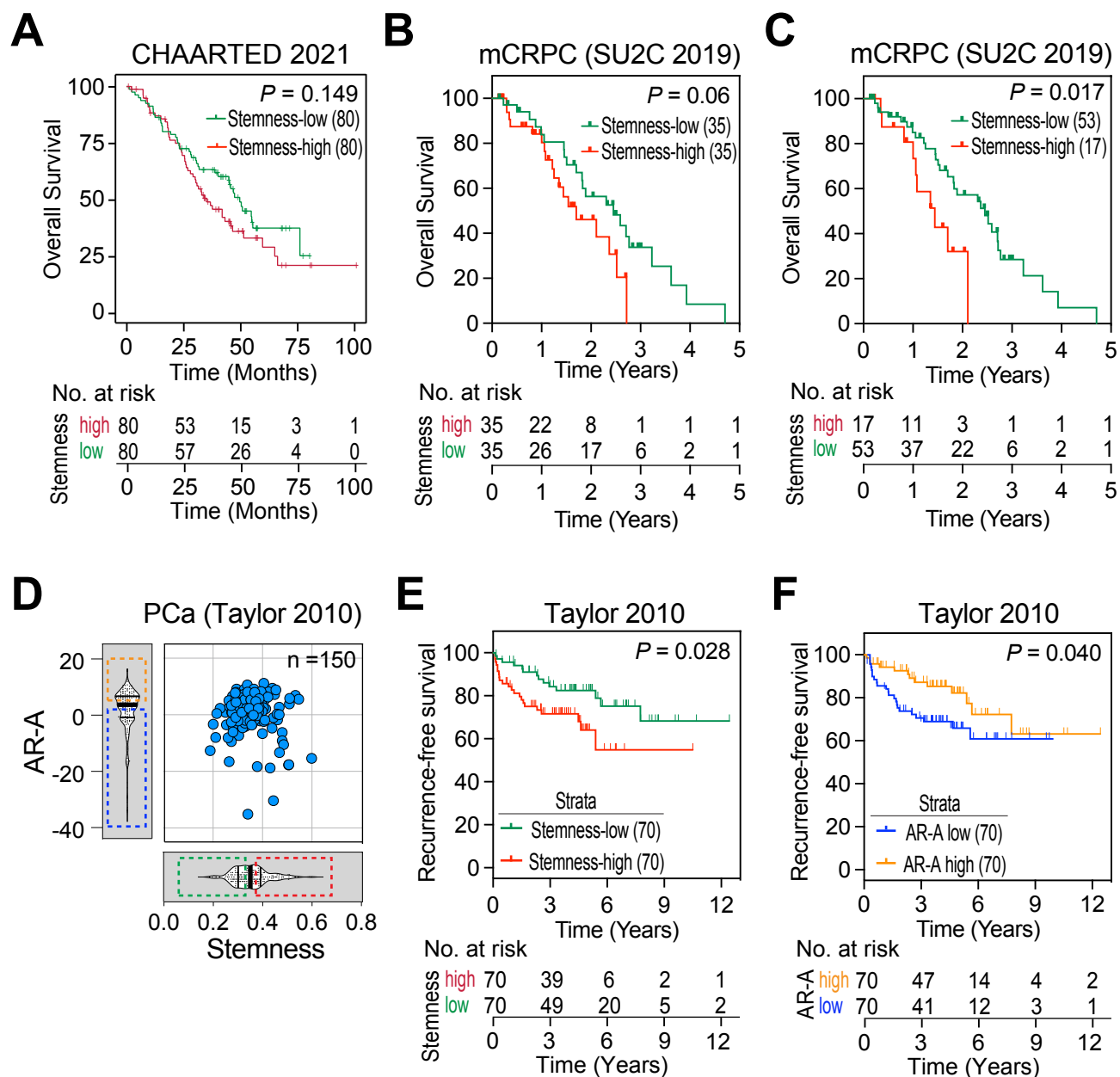
Supplementary Figure S5. Increased Stemness represents a ‘universal’ feature of PCa progression and aggressiveness.

(A) Increasing Stemness accompanies PCa progression and aggressiveness. Stemness scores were shown in various PCa cohorts representing the PCa evolutionary spectrum (from top to bottom): GTEx normal prostate → TCGA tumor-adjacent benign prostate tissue → Treatment-naïve Pri-PCa → aggressive Pri-PCa treated with adjuvant hormone therapy Pri-PCa/ADT → Treatment-failed CRPC → mCRPC → Long-term PCa models including PDX, xenografts (XG) and cell lines. Sample sizes (n) are indicated in the last parentheses. See [Supplementary Table S1](#) for dataset details and [Supplementary Table S4](#) for the statistical significance between different groups.

(B) Progressive increase in Stemness from primary PCa to advanced mCRPC subtypes. This violin plot displays increasing Stemness scores across four PCa groups, defined by treatment status and pathway activity: treatment-naïve Pri-PCa from TCGA, and three subtypes of mCRPC categorized by AR and NE pathway activities - AR pathway-active (ARPC), neuroendocrine (NEPC), and double-negative (DNPC). These three subtypes were delineated based on established signatures (Bluemn *et al.*, 2017, ref. 14; Labrecque *et al.*, 2019, ref. 33). Displayed from left to right are: Untreated Pri-PCa (n = 422), mCRPC-ARPC (191), mCRPC-NEPC (26), and mCRPC-DNPC (16). Pri-PCa samples originated from TCGA, while mCRPC samples were derived from studies by Sharp *et al.* (2019, ref. 67, GSE118435), Labrecque *et al.* (2019, ref. 33, GSE126078), Nyquist *et al.* (2020, ref. 68, GSE147250), Lim *et al.* (2021, ref. 70, GSE171729), and Beltran *et al.* (2016, ref. 28, PRJNA282856; dbGaP: phs000909). Normalized RNAseq data were sourced from the Bolis 2021 integrated cohort (ref. 27). For dataset details, see [Supplementary Table S1](#).

(C-H) PCa progression in GEMMs is characterized by decreasing AR-A and increasing Stemness in aggressive DKO and TKO tumors. The Stemness and AR-A were analyzed in distinct genetically engineering mouse models (GEMMs) of PCa due to individual or combined deletion of 3 tumor suppressor genes, *Pten*, *Rb1* and *Trp53* (Ku *et al.*, 2017, ref. 37). Briefly, *Pten*^{-/-} single knockout (SKO) prostate tumors develop around 9 weeks, and mice rarely develop metastasis with a median lifespan of 48 weeks. In contrast, the *Pten*^{-/-};*Rb1*^{-/-} double knockout (DKO) mice develop highly metastatic PCa that shortens median survival to ~38 weeks. DKO tumors are initially castration-sensitive but eventually become castration-resistant and, notably, castration-resistant DKO tumors turn into an AR^{-lo} phenotype. When *Trp53* is further deleted in DKO background, the triple KO (TKO; *Pten*^{-/-};*Rb1*^{-/-};*Trp53*^{-/-}) tumors are exclusively AR⁻ NEPC and castration-resistant *de novo* with high metastatic rate and lifespan of ~16 weeks. AR-A was dramatically reduced (C) but Stemness was significantly increased (D) in the aggressive DKO and TKO PCa compared to indolent SKO prostate tumors. E-H, Scatter plots illustrating the correlations between AR-A and Stemness in combined cases in PCa GEMMs (E), SKO (F, Pearson's $r=0.95$, $P=0.053$), DKO (G, Pearson's $r=-0.56$, $P=0.053$), and TKO (H, Pearson's $r=-0.78$, $P=0.066$).

All data in dot plots and violin plots are shown as mean ± SD. Statistical significance was determined using one-way analysis of variance (ANOVA) followed by Tukey's multiple-comparison test. Jonckheere-Terpstra's trend (J-T) test was used to calculate the statistical significance of the trend across different groups. Pearson correlation coefficients (Pearson's r) were used to calculate the correlations between the Stemness and AR-A. Note: ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$.



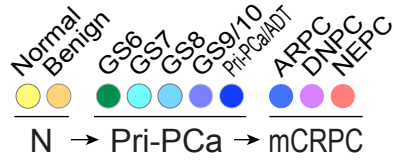
Supplementary Figure S6. High Stemness correlates with poor patient survival.

(A-C) High Stemness correlates with poorer overall survival in CHAARTED 2021 cohort (A) or SU2C mCRPC cohort (B-C). The CHAARTED 2021 cohort was divided into Stemness-high ($n = 80$) and Stemness-low ($n = 80$) using Stemness median split (A). The SU2C 2019 mCRPC cohort was divided into Stemness-high and Stemness-low using Stemness median split (B) or the analysis was done by comparing the 25% of highest Stemness samples with the 75% of lowest Stemness samples (C). P -value was determined using the Log-Rank test.

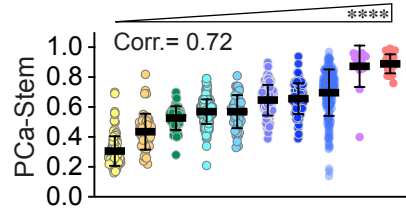
(D-F) High Stemness and low AR-A correlate with worse recurrence-free survival in PCa patients in the Taylor 2010 PCa cohort. Stemness (E) or AR-A (F) was used to fractionate Taylor cohort for Kaplan-Meier analysis. The Taylor patient cohort was divided into Stemness-high ($n = 70$) and Stemness-low ($n = 70$) using Stemness median split (E) or AR-A high ($n = 70$) and AR-A low ($n = 70$) using AR-A median split (F) followed by the Kaplan-Meier analyses. Note: there are 150 PCa samples with transcriptome data in the Taylor cohort but only 140 PCa samples with survival data. P -value was determined using the Log-Rank test.

Supplementary Figure S7

A

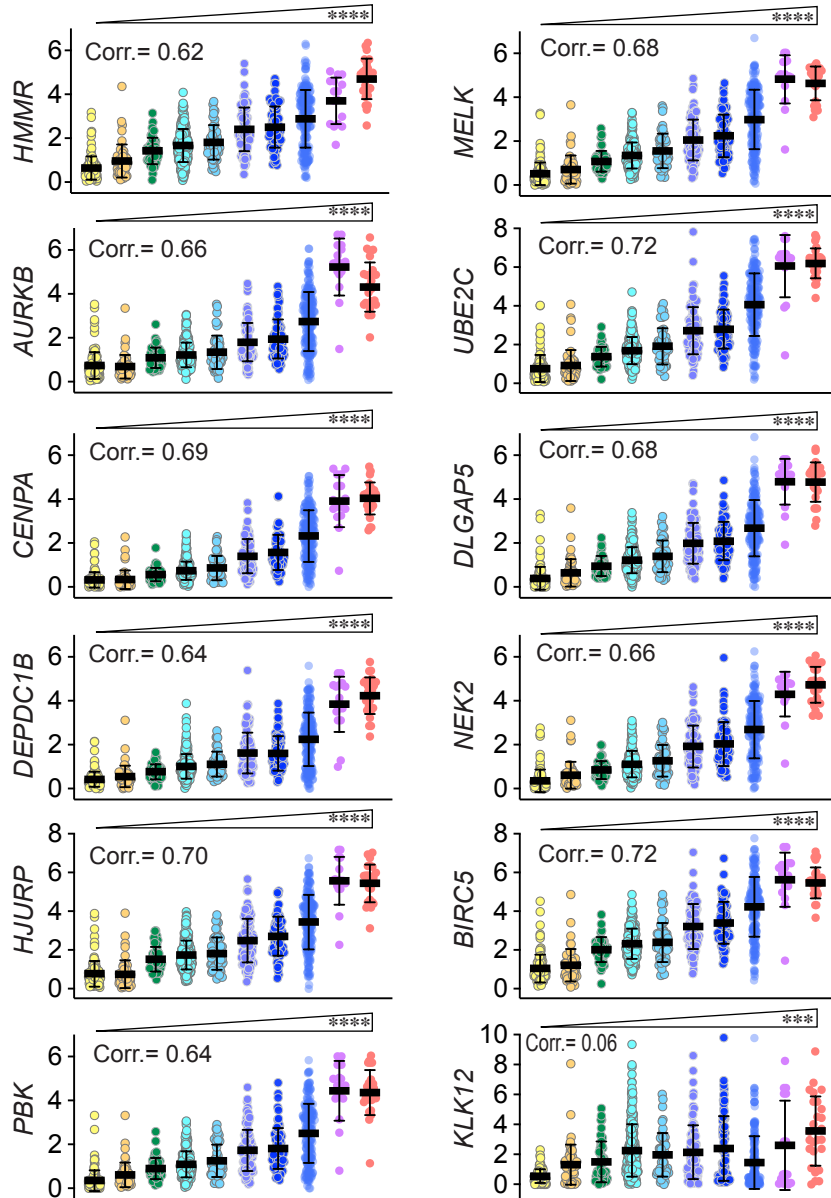


B



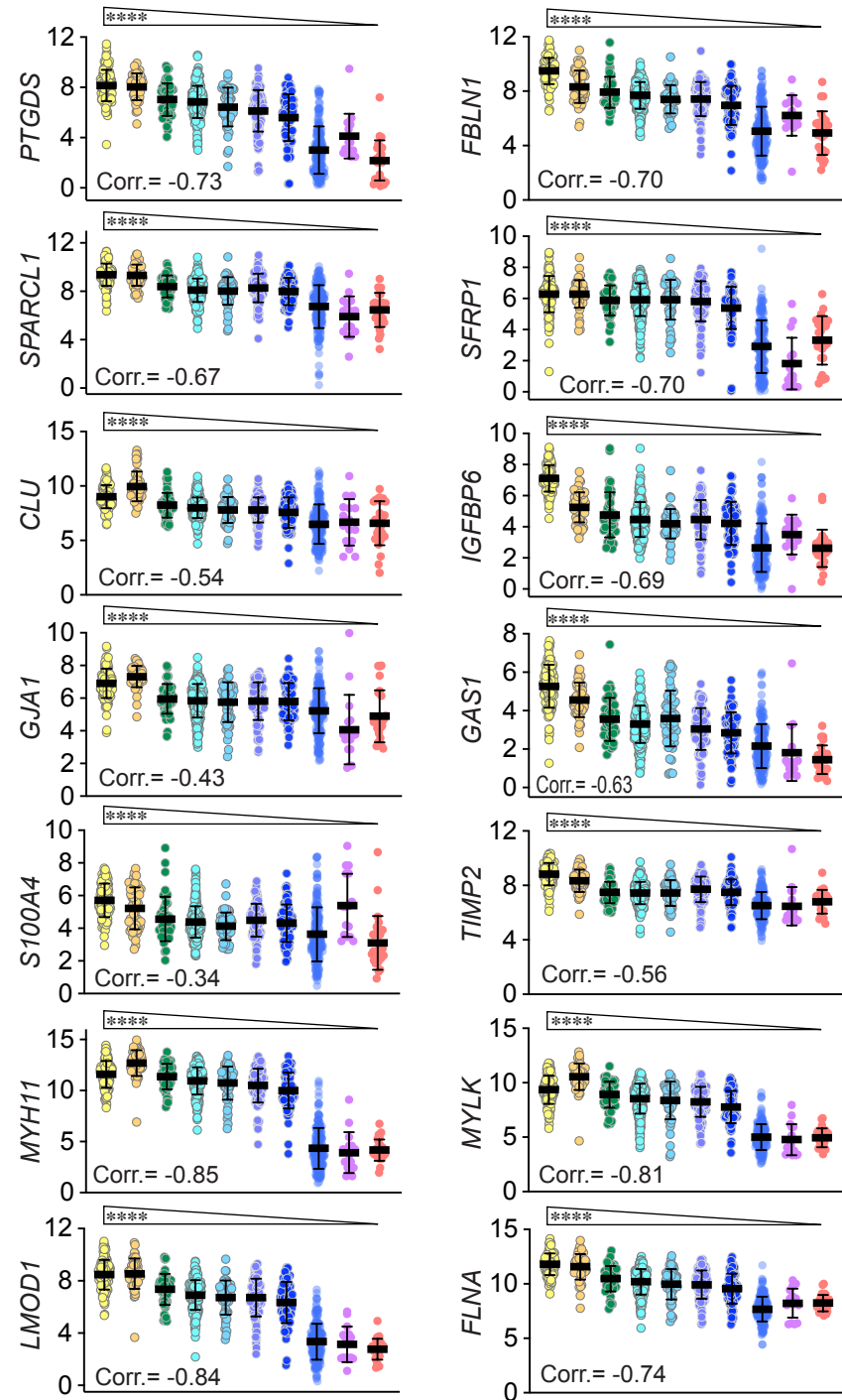
C

Genes Commonly Upregulated in Stemness-high



D

Genes Commonly Downregulated in Stemness-high



Supplementary Figure S7. PCa-Stem signature and gene expression correlate with PCa progression.

(A) Schematic of PCa continuum, from normal prostate (N) through various stages of primary PCa (Pri-PCa) to treatment-failed mCRPC. Sample groups include:

- Two subgroups in N (Normal/Benign), including GTEx normal prostate (n=116) and TCGA tumor-adjacent benign prostate tissue (n=52).
- Five subgroups in Pri-PCa, including GS6 (n=45), GS7 (n=238), GS8 (n=49), combined GS9 and GS10 (GS9/10; n=90), and aggressive Pri-PCa treated with post-surgery adjuvant hormone therapy (Pri-PCa/ADT; n=70).
- Three subtypes of mCRPC, including AR-active PCa (ARPC; n=191), double-negative AR-null/neuroendocrine-null PCa (DNPC; n=16), and neuroendocrine PCa (NEPC; n=26).

*Note that the color-coded entities of normal/benign prostate and progressive PCa stages in panel A also represent the legend for all dot plots in panels B-D.

(B) Dot plot showing the increasing 12-gene PCa-Stem signature scores positively correlate with the PCa progression described in (A).

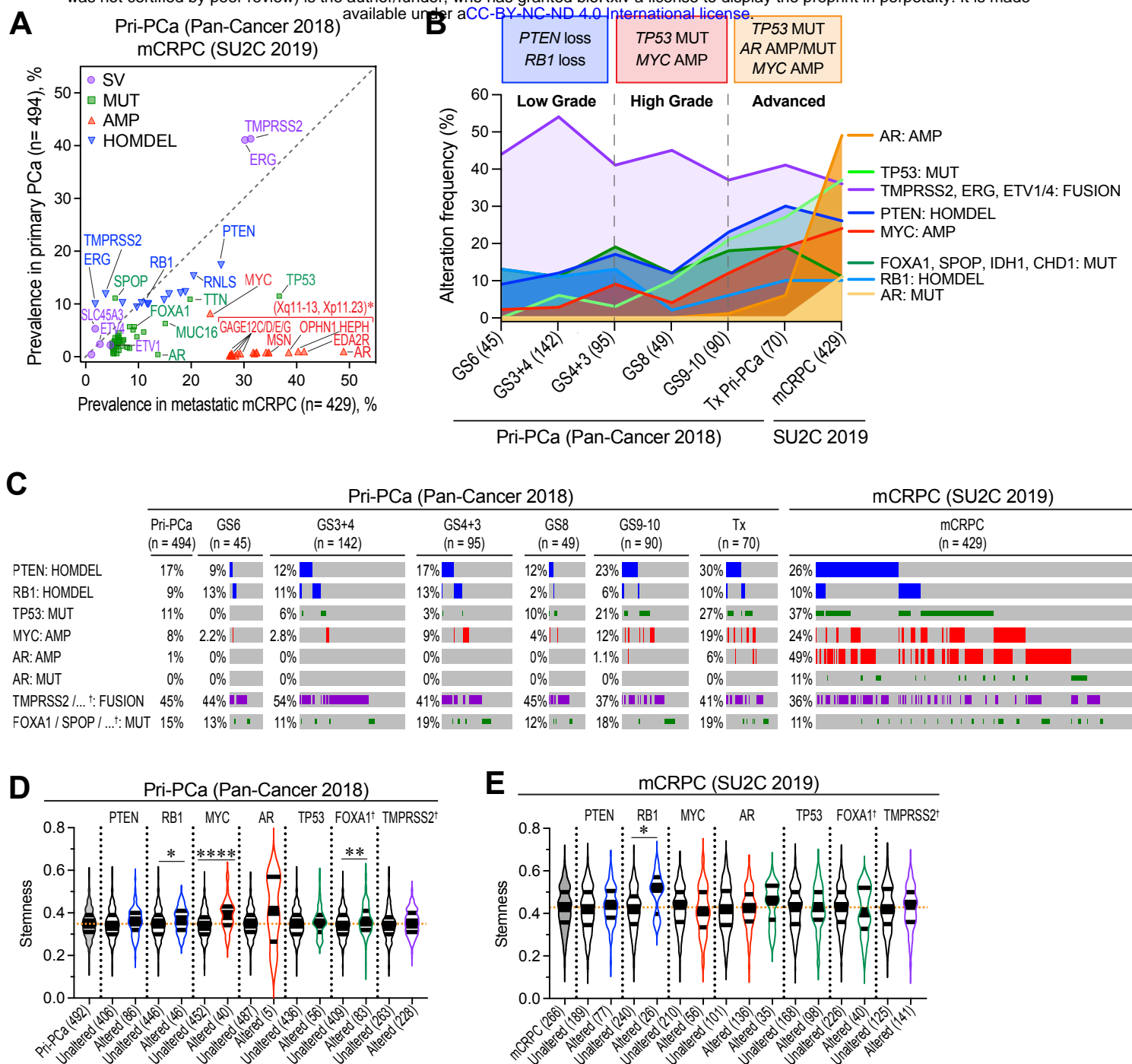
(C) Dot plots showing that the mRNA levels of 11 of the 12 PCa-Stem signature genes (i.e., *HMMR*, *AURKB*, *CENPA*, *DEPDC1B*, *HJURP*, *PBK*, *MELK*, *UBE2C*, *DLGAP5*, *NEK2*, and *BIRC5*) except *KLK12* (see [Figure 5D](#); [Supplementary Table S3](#)) exhibited stage-related increases and positively correlated with PCa progression.

(D) Dot plots showing that the mRNA levels for 14 genes commonly downregulated in Stemness-high groups (i.e., *PTGDS*, *SPARCL1*, *CLU*, *GJA1*, *S100A4*, *MYH11*, *LMOD1*, *FBLN1*, *SFRP1*, *IGFBP6*, *GAS1*, *TIMP2*, *MYLK*, and *FLNA*; see [Supplementary Table S3](#)) all individually displayed stage-related downregulation and negatively correlated with PCa progression.

For all dot plots, the Pearson's correlation coefficients (Corr.) are shown for the relationships between PCa progression scores (derived from trajectory inference analysis in the PCa transcriptome atlas; Bolis 2021; ref.27) and either PCa-Stem signature scores (B) or individual gene expression levels (C, D).

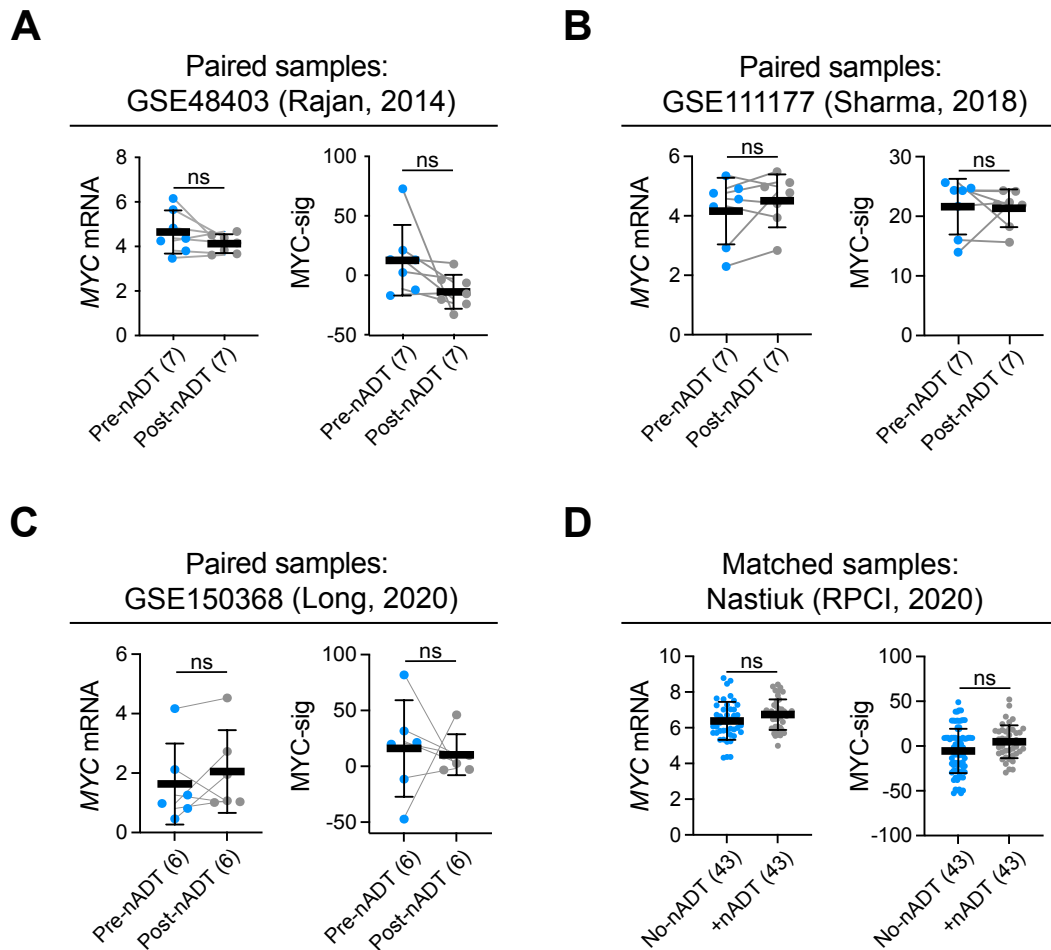
Dot plots display the mean \pm SD. Trends across groups were assessed using the Jonckheere-Terpstra (J-T) trend test (wedges above). ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant.

Data Source: Normalized RNA-seq data across the stages (shown in the legend in A) were obtained from the integrated transcriptomic framework of PCa (Bolis 2021; ref.27).



Supplementary Figure S8. Unique genomic features of Stemness-high PCa.

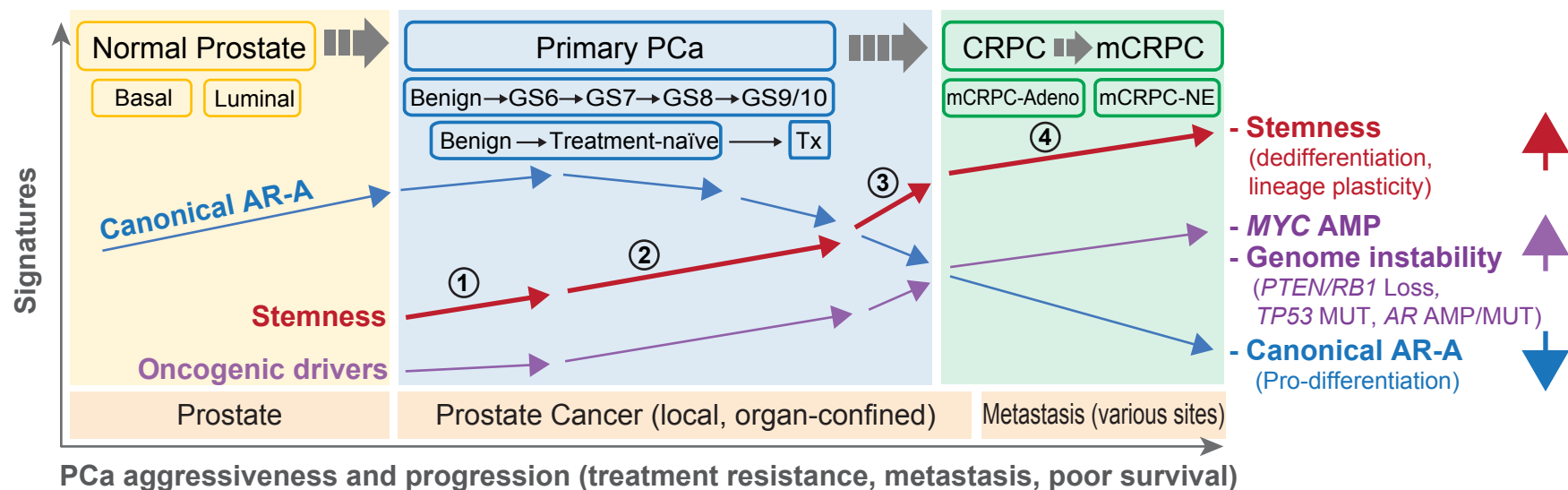
- (A)** Scatterplot of observed frequencies of genomic alterations in cancer genes reported in Pri-PCa (TCGA) versus mCRPC (SU2C) with data retrieved from cBioPortal. SV (purple), structural variants/fusions; MUT (green), mutation; AMP (red), amplification in DNA copy numbers (i.e., copy number alterations (CNA) ≥ 2); HOMDEL (blue), homozygous deletion or deep deletion.
- (B)** Genomic alteration frequency plot showing that (early-stage) low-grade PCa have prevalent *PTEN* and *RB1* loss and high-grade PCa have frequent *TP53* MUT and *MYC* AMP whereas advanced PCa have prevalent *TP53* MUT, *MYC* AMP and *AR* AMP/MUT.
- (C)** cBioPortal Oncoprint showing the frequencies of major genetic alteration events along the PCa progression continuum.
- (D-E)** Association of major genomic alterations in PCa with Stemness. Grey, no alterations; Blue, HOMDEL; Green, MUT; Red, AMP; Purple, Fusion. *TP53: MUT / ...: FUSION* (in C) or *TP53: MUT* (in D-E) stands for combined fusion events in *TP53*, *ERG*, *ETV1* and *ETV4*; *FOXA1 / SPOP / ...: MUT* (in C) or *FOXA1* (in D-E) represent combined mutation events in *FOXA1*, *SPOP*, *IDH1* and *CHD1*. Note that data presented represent the major mutation subtypes in Pri-PCa (TCGA PRAD cohort) reported in Cancer Genome Atlas Research Network 2015 (Abeshouse *et al.*, 2015, ref. 18) and alterations (mutations, structural variants and copy number) of unknown significance were excluded. Within the violin plots, the center lines represent median values, and box edges are 75th and 25th percentiles. Significance was calculated by Mann-Whitney test (*, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$).



Supplementary Figure S9. Lack of effect of nADT on MYC activity.

Matched comparisons showing no significant differences in both *MYC* mRNA and MYC-sig after neo-adjuvant ADT (Post-nADT) as compared to the matched samples before nADT (Pre-nADT) from the same individuals in **A**, Rajan 2014 cohort; **B**, Sharma 2018 cohort; and **C**, Long 2020 cohort or in **D**, the RPCI Chatta/Nastiuk cohort (unpublished) comprising PCa samples from patients who received nADT (+nADT) and the patients who did not receive ADT (no-ADT) but otherwise matched (with the +nADT patients) for ages and tumor stage and grade.

Sample sizes are indicated in parentheses. Within the plots, results are shown as the mean \pm SD. The pairs are linked with a grey line. Significance was calculated by two-tailed paired (A-C) or unpaired (D) Student's *t*-test (ns, not significant; *, $P < 0.05$). See [Supplementary Table S2](#) for MYC-sig gene signature information.



Supplementary Figure S10. Schematic summary of PCa progression and aggressiveness accompanied by gain of oncogenic Stemness and loss of prostatic epithelium differentiation defined by canonical AR signaling.

In the current study, we quantified the oncogenic dedifferentiation state (Stemness; indicated by red arrows) and canonical pro-differentiation AR signaling activity (AR-A; blue arrows) using transcriptomic data (3102 RNA-seq and 84081 microarray samples) collected from 26 datasets encompassing the evolutionary spectrum of PCa development and progression – normal prostate, early-stage and advanced primary tumors, metastatic castration-resistant PCa, and PCa subjected to long-term or neo-adjuvant androgen-deprivation therapy (ADT). Our analysis reveals a continual increase in the Stemness at all PCa progression stages, i.e., (1) PCa initiation and early tumor growth (compared with normal or matched benign prostate tissues); (2) natural tumor progression (treatment-naïve); and (3) long-term ADT to treatment resistance and metastasis. In contrast, the pro-differentiation AR-A displays a bell-shape trajectory with an increase in early-stage tumors followed by steady decreases in advanced tumors and the lowest in mCRPC. Our results also demonstrate that the increased AR-A represents a driving force of elevated Stemness in early PCa but tumor progression is accompanied by a divergence between decreasing AR-A and increasing Stemness, and clinical therapies such as ADT further diversify AR-A and Stemness. In fact, canonical AR-A shows an inverse correlation with the Stemness in treatment-failed mCRPC, and the Stemness tracks faithfully with the treatment-induced reprogramming. On the other hand, akin to the accompanying continually increasing Stemness, PCa progression is characterized by continually elevating MYC amplifications and activity, as well as increasing genome instability associated with deletion of tumor suppressors and AR amplifications and mutations (indicated by purple arrows). Importantly, high Stemness correlates with poor patient survival suggesting that Stemness is prognostic of patient outcomes, and increased Stemness represents a ‘universal’ feature of PCa progression and aggressiveness.