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

Xiaozhuo Liu[1§*], Eduardo Cortes[2,3§], Yibing Ji[1], Kathy Zhao[4], Anmbreen Jamroze[1], Julian Ho[4], Yang Liu[4], Elai Davicioni[4], Felix Y. Feng[5†], Joshi J. Alumkal[6], Daniel E. Spratt[7], Christopher J. Sweeney[8], Han Yu[2], Qiang Hu[2], Cheng Zou[9], Dingxiao Zhang[9], Kevin Lin[10], Yue Lu[10], Gurkamal Chatta[11], Kent L. Nastiuk[12], David W. Goodrich[1], Kiera Rycaj[1], Jason S. Kirk[1], Igor Puzanov[11], Song Liu[2*], Jianmin Wang[2*], Dean G. Tang[1,13*#]

[1]Department of Pharmacology and Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA
[2]Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA
[3]Department of Biostatistics, University at Buffalo, Buffalo, NY 14263, USA
[4]Veracyte Inc, San Diego, CA 92121, USA
[5]Department of Radiation Oncology, University of California, San Francisco, California, USA
[6]Department of Internal Medicine, Department of Hematology and Oncology, Rogel Cancer Center, University of Michigan, Ann Arbor, MI 48109, USA
[7]University Hospitals Seidman Cancer Center and Case Western Reserve University, Cleveland, Ohio, USA
[8]South Australian Immunogenomics Cancer Institute, University of Adelaide, Adelaide, SA 5000, Australia
[9]School of Biomedical Sciences, Hunan University, Changsha, China
[10]Department of Epigenetics and Molecular Carcinogenesis, the University of Texas M.D Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030, USA
[11]Department of Medicine, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA
[12]Department of Cancer Genetics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA
[13]Experimental Therapeutics Graduate Program, University at Buffalo and Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, 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 (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).

[#]Lead contact

Running title: Increasing Stemness Drives Prostate Cancer Progression.

Key words: Prostate cancer, stemness, AR signaling, heterogeneity, plasticity, aggressiveness, *RB1*, MYC.

Word counts: Main Text: 6689 words; Figures: 8; Supplementary Figures: 12; Supplementary Tables: 6.

Abstract

Background

Cancer progression is often accompanied by dedifferentiation and acquisition of stem cell-like properties (stemness). In prostate cancer (PCa), lineage plasticity and therapy resistance remain major clinical challenges, yet a unified quantitative transcriptomic framework connecting stemness, androgen receptor (AR) signaling, castration resistance, and disease progression across the PCa continuum is lacking.

Methods

We performed an integrative analysis of 87,192 transcriptomic data from 27 preclinical and clinical datasets spanning the PCa continuum—from normal prostate and treatment-naïve primary PCa (Pri-PCa) to PCa treated with neoadjuvant ADT (nADT) and metastatic castration-resistant PCa (mCRPC). Tumor stemness was quantified using a transcriptome-derived mRNAsi Stemness Index (Stemness for short), and a 12-gene PCa-Stem signature was developed to capture PCa-specific stemness. Canonical AR activity (c_AR-A) and castration-reprogrammed AR activity (cr_AR-A), *RB1*-loss, *PTEN*-loss, and MYC activity signature scores were analyzed across cohorts, with survival assessed in multiple datasets. Functional validations included MYC knockdown RNA-seq data analysis in LNCaP cells and siRNA-mediated depletion of representative PCa-Stem genes in androgen-independent LAPC4 (LAPC4-AI) cells.

Results

The Stemness score and c_AR-A increased concordantly during early prostate tumorigenesis but diverged with PCa progression: as Gleason grade increased, c_AR-A declined while Stemness continually increased. mCRPC exhibited the highest Stemness and lowest c_AR-A, a pattern recapitulated in *Pten*/*Rb1*/*Trp53*-deficient mouse models. Both global Stemness score and the PCa-Stem signature were enriched in aggressive PAM50-LumB and PCS1 subtypes, associated with the proliferative and lineage plasticity programs, and predicted poor survival. Depletion of representative PCa-Stem genes (*HMMR*, *PBK*, *AURKB*) suppressed proliferation, invasion and organoid formation in LAPC4-AI cells. Mechanistically, MYC activity, cr_AR-A and *RB1*-loss transcriptomic signature were consistently associated with and drove pervasively increasing Stemness during PCa progression. Shared mitotic regulators linked cr_AR-A, *RB1*-loss, MYC activity, and PCa-Stem to mitotic control and therapy-resistant proliferation.

Conclusions

The Stemness scores reported herein quantitatively capture PCa aggressiveness, plasticity, treatment resistance and progression, and prognosticates poor patient survival. Therapy-reprogrammed AR activity (cr_AR-A), *RB1* loss, and MYC activation together reinforce the high-Stemness state and therapy resistance in mCRPC. Collectively, our work establishes a trajectory-integrated and scalable transcriptomic framework that defines cancer Stemness as a quantifiable clinical determinant of PCa progression, plasticity, progression, therapy resistance and patient survival.

Background

Malignant transformation and cancer progression are characterized by progressive loss of differentiation (or oncogenic dedifferentiation) and gain of stem cell-like traits (i.e., stemness). Poorly differentiated tumors are enriched in gene expression profiles of normal embryonic and adult stem cells and in cancer stem cells [1-5]. However, the oncogenic dedifferentiation-based stemness concept [6] 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 >36,000 American men in 2026 [9]. 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 plasticity [8].

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.

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 pathway [10], lineage plasticity [8, 11, 12], and increased stemness [8, 13]. Under therapeutic pressure, AR binding shifts toward noncanonical targets (i.e., reprogramming of AR cistrome), resulting in extensive reconfiguration of the AR transcriptional program [14-17]. These observations suggest a progressive divergence of AR signaling from its canonical, pro-differentiation role during tumor evolution. Thus, quantifying canonical AR activity across disease stages should provide a tractable means to capture this adaptive trajectory. However, how such changes in AR signaling relate to global stemness and tumor aggressiveness remains unclear.

These discussions raise the need for a quantitative framework that simultaneously captures AR signaling and stemness to reveal how their dynamic interplays contribute to PCa aggressiveness, plasticity,

111 and therapy resistance. Here, we adapted and developed transcriptome-based gene signature scores to
112 quantitatively assess the canonical AR activity (c_AR-A) and oncogenic stemness levels (Stemness score;
113 Stemness for short) and their inter-relationship in the PCa spectrum. Specifically, we used gene expression
114 data collected from 87,192 samples in 27 distinct pre-clinical datasets and clinical cohorts encompassing
115 the PCa continuum, i.e., from normal/benign prostate to early stage and advanced PCa, to pre- and post-
116 therapy PCa, and, finally, metastatic CRPC (mCRPC), to systematically track the dynamic evolution of and
117 inter-relationship between c_AR-A and the Stemness Index. Our findings demonstrate that Stemness
118 quantitatively tracks PCa progression and adverse clinical outcomes, establishing a trajectory-integrated,
119 pan-cohort framework that links pervasively increasing Stemness as a defining hallmark of PCa
120 aggressiveness and therapy resistance, and poor patient prognosis.

121 **Methods**

122 **I. PCa clinical datasets and data collection**

123 In this study, we utilized a total of 27 published and in-house datasets, including various cohorts
124 encompassing the evolutionary spectrum of PCa (Fig. S1A-B): normal prostate, tumor-adjacent benign
125 tissue, treatment-naïve primary PCa (Pri-PCa) with increasing GS, PCa treated with neo-adjuvant ADT
126 (nADT) or long-term ADT, mCRPC, and PCa cell lines, PCa patient-derived xenografts (PDX) and
127 xenografts. Clinical information, genomic data and gene expression profiling data were downloaded and
128 detailed information on each dataset is listed in Table S1. Dataset descriptions include dataset source
129 references, PubMed IDs (PMIDs; RRID:SCR_004846), DOI links, sample type and source, sample sizes,
130 data types used in the current study (clinicopathologic data, RNAseq, microarray, whole exome
131 sequencing [WES]), and public data access portals.

132

133 Data Access and Repository Information: Data for this study were retrieved from the following public
134 repositories and databases:

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

143 Accession numbers and portal links for the 27 datasets described below are summarized in Table S1.

144

145 Brief Descriptions of the 27 Datasets:

146 1) **GTEx 2013** (PMID: 23715323; DOI: 10.1038/ng.2653) [18]: The Genotype-Tissue Expression (GTEx)
147 Project repository collected high-throughput and clinical data of normal tissue from many organs.
148 RNAseq data of 245 normal prostate tissue samples was downloaded from the GTEx portal
149 (RRID:SCR_013042; https://www.gtexportal.org/, version: v8).

150 2) **Smith 2015** (PMID: 26460041; DOI: 10.1073/pnas.1518007112) [19]: RNAseq data of 5 prostate basal
151 cell (Trop2^{+}CD49f^{hi}) and corresponding 5 luminal cell (Trop2^{+}CD49f^{lo}) samples FACS-purified from
152 benign human prostate tissue in 5 PCa patients (n=5) was retrieved from the GEO database
153 GSE82071.

154 3) **Liu 2016** (PMID: 27926864; DOI: 10.1016/j.celrep.2016.11.010) [20]: RNAseq data of 3 FACS-purified
155 human prostate basal cell (CD45^{-}EpCAM^{+}CD49f^{hi}CD38^{lo}) and corresponding luminal cell (CD45^{-}

156 EpCAM^{+}CD49f^{lo}CD38^{hi}) and luminal progenitor cell (CD45^{-}EpCAM^{+}CD49f^{lo}CD38^{lo}) populations from
157 the benign prostate tissues of 3 PCa patients (n=3) was retrieved from the supplementary data of the
158 indicated reference. The microarray-based transcriptomic profiles of the same samples were retrieved
159 from the GEO database GSE89050.

160 4) **Zhang 2016** (PMID: 26924072; DOI: 10.1038/ncomms10798) [21]: RNAseq data of 3 human prostate
161 basal cell (Trop2^{+}CD49f^{hi}) with the corresponding 3 luminal cell (Trop2^{+}CD49f^{lo}) samples FACS-
162 purified from benign prostate tissues of 3 PCa patients (n=3) was retrieved from the GEO database
163 GSE67070.

164 5) **TCGA PRAD 2015** [22] and **Pan-Cancer Atlas 2018**: The Cancer Genome Atlas (TCGA)
165 (https://www.cancer.gov/tcga) repository collected high-throughput molecular and clinical data from
166 primary cancer and matched normal/benign samples spanning 33 human cancer types, including
167 prostate adenocarcinoma (PRAD). For each cancer type, TCGA published an initial "marker paper"
168 summarizing analyses performed. Supplemental and associated data files supporting these "marker
169 papers" are available through the NCI Genomic Data Commons (GDC) website
170 (https://gdc.cancer.gov/about-data/publications). For PCa, **TCGA PRAD 2015** [22] (PMID: 26544944;
171 DOI: 10.1016/j.cell.2015.10.025) is the marker paper on PCa published in 2015, which included the
172 dataset of 333 primary prostate adenocarcinomas (281 treatment-naïve Pri-PCa, 49 PCa with post-
173 surgery adjuvant ADT, and 52 matched normal/benign prostate tissue). Genomic data were obtained
174 from **cBioPortal for Cancer Genomics** (RRID:SCR_014555; https://www.cbioportal.org), specifically
175 the TCGA-PRAD Firehose Legacy study (https://www.cbioportal.org/study/summary?id=prad_tcga).
176 RNAseq expression data was obtained from Zhang *et al*. [23] (PMID: 32350277; DOI: 10.1038/s41467-
177 020-15815-7).

178 The expanded **TCGA PRAD 2018 (Pan-Cancer Altas)** cohort included 494 patients with Pri-PCa
179 (422 treatment-naïve Pri-PCa, 70 PCa with post-surgery adjuvant hormone treatment, 2 PCa with
180 chemotherapy) and 52 matched adjacent normal/benign prostate tissue. Clinicopathologic features
181 (age, Gleason scores, PSA levels, tumor stages) and genomic alterations data (mutation frequency,
182 copy number alterations [CNAs], structural variants [SVs], tumor mutation burden [TMB], fraction
183 genome altered) were retrieved from cBioPortal for Cancer Genomics PRAD PanCancer Atlas study
184 (https://www.cbioportal.org/study/summary?id=prad_tcga_pan_can_atlas_2018) and from the Pan-
185 Cancer Atlas companion site (https://www.cell.com/pb-
186 assets/consortium/pancanceratlas/pancani3/index.html). Survival data were obtained from Liu *et al*.
187 [24] (PMID: 29625055; DOI: 10.1016/j.cell.2018.02.052). Normalized RNAseq data were downloaded
188 via NCI GDC hub at UCSC Xena Functional Genomics Portal [25] (https://gdc.xenahubs.net, version
189 2019-07-20; PMID: 32444850; DOI: 10.1038/s41587-020-0546-8).

6) **Rajan 2014** (PMID: 24054872; DOI: 10.1016/j.eururo.2013.08.011) [26]: Clinicopathologic (Gleason scores) and RNAseq data of 7 matched pre- and post-nADT PCa samples were obtained from Table 1 of the indicated reference and the GEO database GSE48403.

7) **Sharma 2018** (PMID: 30314329; DOI: 10.3390/cancers10100379) [27]: 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 from supplementary Table S1 of the indicated reference and the GEO database GSE111177.

8) **Long 2020** (PMID: 32951005; DOI: 10.1016/j.eururo.2013.08.011) [28]: 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** (PMID: 41646678; DOI: 10.64898/2026.01.10.26343859) [29]: 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 Drs. K. Nastiuk and G. Chatta at the Roswell Park Comprehensive Cancer Center.

10) **Gerhauser 2018** (PMID: 30537516; DOI: 10.1016/j.ccell.2018.10.016) [30]: 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) [31]: 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 supplementary Table S1 of the indicated reference and 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) [32]: 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) [33]: 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) [34]: 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 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 PCa with 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 - 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 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 (≤GS7). This distribution implies that most patients were treatment-naïve and likely ADT-free due to less aggressive disease status.

16) **Bolis 2021** (PMID: 34857732; DOI: 10.1038/s41467-021-26840-5) [36]: 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 PCa Transcriptome Atlas (https://prostatecanceratlas.org) was built and integrated from the following studies/datasets: (1) Genotype-Tissue Expression Database [18] (GTEx; PMID: 23715323); (2) The Cancer Genome Atlas [22] (TCGA, TCGA-PRAD; PMID: 26544944); (3) Atlas of RNA-sequencing profiles of normal human tissues [37] (GSE120795; PMID: 31015567); (4) Integrative epigenetic taxonomy of PNPCa [38] (GSE120741; PMID: 30464211); (5) Prognostic markers in locally advanced lymph node-negative PCa (PRJNA477449); (6) The long noncoding RNA landscape of NEPC and its clinical implications [39] (PRJEB21092; PMID: 29757368); (7) Integrative clinical sequencing analysis of metastatic CRPC [10] (PRJNA283922; dbGaP: phs000915; PMID: 26000489); (8) CSER—exploring precision cancer medicine for sarcoma and rare cancers [40] (PRJNA223419; dbGaP: phs000673; PMID: 28783718); (9) Molecular basis of NEPC [41] (Beltran 2016; PRJNA282856; dbGaP: phs000909; PMID: 26855148); (10) Heterogeneity of androgen receptor splice variant-7 (AR-V7) protein expression and response to therapy in CRPC [42] (GSE118435; PMID: 30334814); (11) RNAseq of human PCa cell lines and mCRPC [43-45] (GSE14750; PMIDs: 32460015; 33658518; 34244513 | GSE171729; PMID: 34244513); and (12) Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic CRPC [46] (PRJNA520923; GEO: GSE126078; PMID: 31361600).

17) **Taylor 2010** (PMID: 20579941; DOI: 10.1016/j.ccr.2010.05.026) [47]: 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*. [23] (PMID: 32350277; DOI: 10.1038/s41467-020-15815-7).

19) **SU2C 2019** (PMID: 31061129; DOI: 10.1073/pnas.1902651116) [48]: 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) [46]: 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 was obtained from Supplementary dataset S01 of the indicated reference.

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 of the indicated reference.

23) **Beltran 2016** (PMID: 26855148; DOI: 10.1038/nm.4045) [41]: RNAseq data of mCRPC samples, including 34 castration-resistant adenocarcinoma (mCRPC-Adeno) and 15 mCRPC with neuroendocrine histologies (mCRPC-NEPC), 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) [49]: RNAseq data of murine prostate tumors from genetically engineered mouse models (GEMMs) of PCa of different genotypes, including 4 SKO (single knockout: PBCre4;*Pten*$^{f/f}$), 13 DKO (double knockout PBCre4;*Pten*$^{f/f}$*:Rb1*$^{f/f}$), and 6 TKO (PBCre4;*Pten*$^{f/f}$*:Rb1*$^{f/f}$*:Trp53*$^{f/f}$) was obtained from the GEO database GSE90891 (*f* = a floxed allele of the indicated genes).

25) **CCLE 2018** (PMID: 22460905; DOI: 10.1038/nature11003) [50]: RNAseq data of 1,019 human cell lines were obtained from the Xena Functional Genomics Portal [25] (https://xenabrowser.net, version: 2018-05-30). Among them, eight PCa 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) [51]: RNAseq data of 12 LNCaP xenografts (XG-LNCaP) and 10 LAPC9 xenografts (XG-LAPC9) was obtained from the GEO database GSE88752.

27) **LNCaP (MYC knockdown series)** (PMID: 32681068; DOI: 10.1038/s41388-020-01389-7) [52]: RNAseq data for the LNCaP MYC-knockdown system were obtained from the GEO database (GSE135942). This dataset included non-targeting control samples (n=3) and MYC-knockdown samples (n=6) generated using two shMYC constructs (shMYC-512 and shMYC-637) targeting the MYC open reading frame (ORF), each performed in triplicate. The shMYC-627 construct, which targets the 3′UTR, was excluded from analysis.

<u>Sex as a biological variable:</u>

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 GEMMs were also male-derived. All PCa cell lines analyzed in the study were of male origin, with the exception of Figure S5A, where a broader cohort of 1,019 cell lines from the CCLE, comprising both male- and female-derived samples across multiple cancer types, was analyzed. Analyses of CCLE data in Figure S5A were performed without stratification by sex.

II. Transcriptomic data normalization and transcriptome-based signature quantification

To ensure consistency across diverse transcriptomic datasets, all data acquisition and normalization, and signature quantification (scoring) were performed using methods tailored to data type and analytical scope. All analyses were conducted within each dataset, thereby avoiding inter-study batch effects. For comparisons across pathological stages (e.g., normal/benign → Pri-PCa → mCRPC), we used only harmonized, pre-integrated transcriptomic datasets, including the PCa Transcriptome Atlas (Bolis 2021) [36] and Taylor 2010 [47], which provide variance-stabilized, library size–adjusted expression matrices normalized across disease stages.

In this project, we adopted, derived, and utilized multiple transcriptomic gene signatures representing PCa differentiation, stemness, oncogenic signaling, and tumor-suppressor loss. Signatures analyzed included: (1) the Stemness Index, (2) Canonical AR activity (c_AR-A; 10 genes), (3) MYC signaling activity (MYC-sig; 58 genes), (4) the PCa-Stem Signature (12 genes), (5) castration-reprogrammed AR activity (cr_AR-A; 63 genes), (6) cr_AR-Growth_Wang (8 genes), (7) *RB1*-loss signature (158 genes), and (8) *PTEN*-loss signature (45 genes).

<u>Transcriptome-based quantitative measurement of oncogenic stemness (the Stemness Index):</u>

We adapted mRNA-based **Stemness Index** (mRNAsi) [6] to determine the degree of oncogenic dedifferentiation and to quantify the aggressiveness in PCa. Briefly, we computed Spearman correlations between the stemness model's weight vector and the transcriptome expression profile of the queried samples. Spearman correlation calculations were performed using the 'stats' package 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 was adopted for calculating the Stemness Index because of its robustness to non-normal expression distributions and its reduced susceptibility to cross-dataset batch effects inherent in large-scale integrative analyses.

<u>Canonical AR activity and MYC activity:</u>

Canonical AR activity (**c_AR-A**) was determined by RNA-seq based measurements of AR-regulated transcripts under intact androgen/AR signaling conditions. Similarly, the MYC signaling (**MYC-sig**) activity was determined by MYC-regulated transcripts. We employed Z-score normalization [53] to quantify c_AR-

367 A and MYC-sig. Briefly, the numeric AR or MYC activity score was calculated based on a linear
368 combination of expression values (Z-scores) of experimentally validated AR target genes (10 genes) [54]
369 or MYC target genes (58 genes; GSEA Molecular Signature Database 2022.1 version, MSigDB) [55]. The
370 target gene lists were detailed in Table S2. The expression value for each AR or MYC target gene was
371 converted to Z-score by $Z = (x - \mu)/\sigma$, where x is the expression value of the target gene in a specific sample,
372 μ is the mean and σ is the standard deviation across all samples of a gene. Finally, the combined Z-scores
373 were summed across all genes to represent the expression score of canonical AR or MYC target genes.
374 All Z-score analyses were confined to individual datasets to avoid cross-study normalization artifacts.
375 For cross-stage analyses using harmonized datasets (e.g., the PCa Transcriptome Atlas; ref. [36]),
376 c_AR-A and MYC-sig were additionally quantified using single-sample gene set enrichment analysis
377 (ssGSEA) to enable direct comparison with other gene set–based signatures.
378 PCa-Stem Signature (the 12-gene PCa-specific stemness signature):
379 A PCa–specific stemness signature was constructed by identifying genes consistently enriched in high-
380 stemness tumors across distinct disease states. Differential expression analysis compared the top 33%
381 (Stemness-high) versus bottom 33% (Stemness-low) samples in treatment-naïve Pri-PCa (TCGA-PRAD)
382 and mCRPC (SU2C 2019) cohorts. Genes upregulated in both cohorts (fold change $\geqslant$ 2, FDR < 0.05;
383 Table S3) were intersected to generate the 12-gene **PCa-Stem signature** (Table S2). Per-sample
384 enrichment scores were quantified by ssGSEA.
385 Castration reprogrammed AR activity (cr_AR-A) signatures:
386 The 63-gene **cr_AR-A** (Table S2) signature was derived by integrating differential chromatin analyses of
387 AR and H3K27ac ChIP-seq performed in CRPC versus Pri-PCa samples (GSE130408) [16] with
388 differential transcriptional analysis comparing CRPC and Pri-PCa samples (GSE21032) [47], thereby
389 identifying genes exhibiting concordant CRPC-enriched AR binding, enhancer activation, and
390 transcriptional up-regulation.
391 The **cr_AR-Growth_Wang** (8 genes; Table S2) signature was adopted from Wang *et al*. [14] and
392 consists of AR-bound, enhancer-validated growth regulators identified in androgen-independent LNCaP-
393 abl cells, representing a proliferation-oriented, castration-reprogrammed AR transcriptional program. Both
394 AR-reprogrammed signatures were quantified by ssGSEA.
395 Tumor-suppressor-loss signatures:
396 Transcriptional signatures reflecting functional loss of key tumor suppressors were analyzed using
397 previously validated gene sets. The ***RB1*-loss signature** (158 genes; Table S2), derived from Ertel *et al.*
398 [56], captures E2F-driven transcriptional deregulation associated with RB1 pathway inactivation. The
399 ***PTEN*-loss signature** (45 genes; Table S2), derived from Liu *et al.* [57], represents transcriptional
400 consequences of *PTEN* loss. Both signatures were quantified using ssGSEA.
401 ssGSEA implementation:

402 ssGSEA was performed using the Gene Set Variation Analysis (GSVA) [58] package (version 1.50.5;
403 RRID:SCR_021058) in R (version 4.3.3; RRID:SCR_001905) to quantify per-sample enrichment scores.
404 All ssGSEA analyses were conducted within individual datasets, using rank-based enrichment to minimize
405 sensitivity to expression scale differences. For analyses spanning disease stages, ssGSEA was applied
406 to the harmonized transcriptomic datasets to minimize batch effects. All gene sets are provided in Table
407 S2.
408
409 **III. Integrative analysis of genomic alterations in Pri-PCa (TCGA) and mCRPC (SU2C)**
410 To explore molecular and clinical features associated with the PCa stemness, we performed integrative
411 analysis of genomic, transcriptomic, clinicopathologic, and outcome data using cBioPortal for Cancer
412 Genomics (RRID:SCR_014555), which provides interactive access to multidimensional cancer genomics
413 datasets and integrated visualization tools [59]. Specifically, we analyzed data from Pri-PCa (TCGA-PRAD
414 2018, Pan-Cancer Atlas) and mCRPC (SU2C 2019) cohorts. Clinicopathologic variables (e.g., age,
415 Gleason scores, PSA levels, tumor stage), genomic alterations (e.g., mutation frequency, TMB, fraction
416 genome altered, CNAs, and SVs), and patient outcome data (e.g., overall survival, progression-free
417 survival) were retrieved from cBioPortal (RRID:SCR_014555) as described in the "PCa clinical datasets
418 and data collection" section. Transcriptomic data used for the Stemness Score calculation were obtained
419 as described in the same section. For comparative analysis, mCRPC samples were stratified into top and
420 bottom 33% by the Stemness Scores. In TCGA, only treatment-naïve Pri-PCa samples were used for a
421 similar 33% stratification, excluding 70 patients who received adjuvant (post-surgery) hormone therapy to
422 reflect spontaneous disease progression. Differential expression analysis was conducted using DESeq2
423 [60] (version 1.42.1; RRID:SCR_015687), and gene-level and genome-wide alteration frequencies were
424 assessed across stratified groups and disease stages (e.g., Pri-PCa vs. mCRPC) using OncoPrint
425 visualizations, bar plots, and correlation analysis. Associations with clinical outcomes were evaluated as
426 described in the "Statistical analysis" section.
427
428 **IV. Functional characterization of representative PCa-Stem signature genes in castration-resistant**
429 **LAPC4-AI cells**
430 Cell culture:
431 The androgen-independent (AI) LAPC4 (LAPC4-AI) cells were derived as described [61] and cultured at
432 37°C with 5% CO_2 in IMDM (Thermo Fisher, Cat# 21056023) supplemented with charcoal-stripped serum
433 and 1× penicillin/streptomycin. Cells were routinely passaged to maintain exponential growth and
434 confirmed to be mycoplasma-free prior to experimentation.
435 siRNA transfection:
436 LAPC4-AI cells at ~70% confluence were transfected with Accell SMARTpool siRNAs targeting *HMMR*,
437 *AURKB*, or *PBK*, or a non-targeting control (NTC) siRNA (Horizon Discovery). The following reagents were

438 used: Accell Human *HMMR* siRNA SMARTpool (E-010409-00-0005); Accell Human *PBK* siRNA
439 SMARTpool (E-005390-00-0005); Accell Human *AURKB* siRNA SMARTpool (E-003326-00-0005). siRNA
440 complexes were prepared in serum-free medium and added to cells at 1 μM final concentration per
441 manufacturer's protocol. Media were replaced with complete medium after incubation for 24 h. Knockdown
442 efficiency and functional assays were performed at the indicated time points.

443 RNA isolation and qRT-PCR:

444 Total RNA was isolated 48 h post-transfection using the PicoPure RNA Isolation Kit (Applied Biosystems,
445 Cat# KIT0204). RNA (1 μg) was reverse-transcribed using SuperScript™ VILO™ Master Mix (Invitrogen,
446 Cat# 11755050). Quantitative real-time PCR (qRT-PCR) was performed using TaqMan probe-based
447 assays on a QuantStudio™ Real-Time PCR System (Applied Biosystems). Relative expression levels were
448 calculated using the ΔΔCt method and normalized to GAPDH, with values expressed relative to NTC
449 controls. Data were presented as mean ± SD from technical triplicates.

450 Trans-well Matrigel invasion assays:

451 At 48 h post-transfection, cells were harvested and equal numbers were seeded into Matrigel-coated trans-
452 well inserts (Corning, Cat# 354480) in serum-free medium. Medium containing serum was placed in the
453 lower chamber as a chemoattractant. After 24 h, non-invading cells were removed, and invaded cells on
454 the underside of the membrane were fixed and stained with DAPI. Images were acquired using a Keyence
455 microscope. Invaded cells were quantified by counting nuclei in multiple non-overlapping fields per insert.
456 Data were presented as mean ± SD from technical replicates.

457 Colony formation assays:

458 At 24 h post-transfection, LAPC4-AI cells were replated at clonal density (500 cells per well in 6-well plates)
459 and cultured for 14 days. Colonies were fixed and stained with crystal violet with representative images
460 shown. Where indicated, colonies were quantified using a consistent size threshold across conditions.

461 Organoid formation and real-time cytotoxicity assays:

462 At 24 h after siRNA transfection, LAPC4-AI cells were seeded for 3D culture in organoid medium
463 supplemented with 5% Matrigel [61]. Real-time cytotoxicity assays were initiated 24 h later and monitored
464 for 72 h using a cell-impermeant dead-cell dye (Thermo Fisher, Cat# L7013). Fluorescence was recorded
465 at defined intervals to assess cell death kinetics. Fluorescent signals were background-corrected using
466 cell-free wells. Data were presented as mean ± SD from technical triplicates.

467

468 **V. Statistical analyses**

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

471 For comparisons between two groups, independent sample or paired *t*-tests were used to compare
472 group means, assuming normality and equal variance. For multi-group comparisons, Tukey's range test
473 was employed following one-way ANOVA (adjusted p-values listed in 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. For analyses involving proportions across ordered disease categories, the Cochran–Armitage trend test was used to test for a linear trend in alteration proportions across GS6, GS9–10, Pri-PCa/ADT, and mCRPC. 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 (the Stemness Index, PCa-Stem signature, etc). For the mRNA-based gene signatures (the Stemness Index, PCa-Stem signature, c_AR-A), patients were stratified into high and low signature groups based on median split in cohorts including Spratt 2017 (Fig. 4A), Tosoian 2020 (Fig. 4B), SU2C 2019 (Fig. S6A), and Taylor 2010 (Fig. S6D-E). For some analyses, including CHAARTED 2021 (Fig. 4C) and SU2C 2019 (Fig. S6B), 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.

For the receiver operating characteristic (ROC) analysis, ROC curves and corresponding area under the curve (AUC) values were computed to evaluate the discriminative performance of the 12-gene PCa-Stem signature and the general Stemness Index. ROC analyses were performed using the pROC package in R (version 4.3.3), with AUC estimated by the non-parametric DeLong method. Classifications included (i) high-grade (GS > 7) versus low-grade (GS $\leqslant$ 7) Pri-PCa, and (ii) mCRPC versus treatment-naïve Pri-PCa. All ROC analyses were performed within the PCa Transcriptome Atlas to avoid inter-dataset batch effects.

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

VI. Data and Code Availability

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, the UCSC Xena Functional Genomics Portal (RRID:SCR_018938), EGA (RRID:SCR_004944), ENA (RRID:SCR_006515), and Zenodo (RRID:SCR_004129), as detailed in Table S1. No new sequencing or genomic data were generated.

Code availability: All custom R scripts used for quantification of Stemness Score (i.e., mRNAsi), PCa-Stem signature ssGSEA scoring, Jonckheere–Terpstra trend testing, multivariable Cox modeling, and regression analysis are available on GitHub: https://github.com/DaltonLiu2026/ProstateCancer-Stemness-Scripts. A permanently archived, citable version of this repository has been deposited in Zenodo: https://zenodo.org/records/15620990. The GitHub/Zenodo repository includes all scripts, example input/output files, a README, dependency list, and a code-reproducibility overview.

521 **Results**

522

523 **Adapting and developing the stemness and c_AR-A index scores for PCa analysis.**
524 To determine the global and dynamic changes of and inter-relationship between the degree of malignancy
525 (oncogenic dedifferentiation) and level of (luminal) differentiation during PCa development and progression
526 (Fig. S1A-B), we employed transcriptome-based algorithms to annotate 27 datasets of 3,111 bulk RNA-
527 seq and 84,081 microarray samples encompassing the PCa continuum (Fig. S1B; Table S1). In our studies,
528 we focused on and compared three progressive stages: from normal prostate to localized Pri-PCa, from
529 low-grade to high-grade Pri-PCa, and from treatment-naïve Pri-PCa to CRPC/mCRPC (Fig. S1A). We
530 defined "progression" as stages beyond Pri-PCa and as disease entities that are more aggressive in a
531 comparative manner. Specifically, we adapted the Stemness Index [6] (mRNAsi; **Stemness** for short) (Fig.
532 S1C) to gauge the degree of malignancy (or aggressiveness) and employed the Z-score normalized
533 canonical AR activity or **c_AR-A** (Fig. S1D) to measure the level of differentiation. Canonical c_AR-A can
534 be determined by quantifying AR-regulated transcripts under intact androgen/AR signaling conditions, and
535 there are several such AR-regulated gene lists [22, 23, 54, 62, 63] (Fig. S2A-B). We validated the c_AR-A
536 list based on their enrichment in prostatic epithelium differentiation, correlation of individual gene
537 expression with c_AR-A signatures in clinical datasets, and overall consistency in correlation among the
538 AR-A scores (Fig. S2B-D). Given that Bluemn's c_AR-A score [54] correlated well with other scores (Fig.
539 S2E-G) and the 10 genes in this signature consistently represented AR targets in multiple clinical datasets
540 (Fig. S2D; Table S2), we decided to use this AR-regulated gene list for our c_AR-A quantification.

541

542 **Normal prostate epithelium shows a negative correlation between Stemness and c_AR-A.**
543 The normal human prostate (NHP) comprises 2 major epithelial types: basal and luminal cells (Fig. S3A).
544 We annotated two studies that profiled FACS-purified NHP luminal and basal cell populations [19, 21] and
545 observed that the luminal cell population (Trop2^{+}CD49f^{lo}) possessed higher *AR* mRNA levels and c_AR-A
546 but lower Stemness compared to basal cells (Trop2^{+}CD49f^{hi}) and, in both datasets, c_AR-A negatively
547 correlated with Stemness (Fig. S3B-C). These results support that the NHP basal cell compartment
548 harbors primitive stem cells [7, 19, 21, 64, 65] and that c_AR-A measures the level of prostate epithelial
549 differentiation.

550

551 **Pri-PCa exhibits concordantly increased Stemness and c_AR-A, and neoadjuvant ADT (nADT)**
552 **decreases both Stemness and c_AR-A.**
553 In analyzing c_AR-A and the Stemness in matched pairs of prostate tumor (T) and normal (N; tumor-
554 adjacent benign) tissues in the TCGA-PRAD [22] (n=51 pairs), Wyatt 2014 [31] (n=12 pairs), and Long
555 2020 [28] (n=10 pairs) cohorts, we observed concordantly increased c_AR-A and Stemness (Fig. 1A-C).

556 Accompanying elevated c_AR-A, *AR* mRNA levels were also increased in Pri-PCa compared to benign
557 tissue (Fig. 1A-C; note that increase in *AR* mRNA levels in the Long 2020 cohort approached statistical
558 significance). The paired T/N comparisons (Fig. 1A-C) suggest that early PCa growth is characterized by
559 concordantly increased Stemness and c_AR-A. To test whether c_AR-A might causally drive increased
560 Stemness during early prostate tumorigenesis, we investigated 4 PCa nADT (i.e., ~2-6 months of ADT
561 before prostatectomy or radiation) cohorts [26-28]. In the first 3 cohorts with matched pairs of pre- and
562 post-nADT samples (from same patients), nADT concordantly downregulated c_AR-A (without decreasing
563 *AR* mRNA levels) as well as the Stemness (Fig. 1D-F). In the Nastiuk cohort with non-paired post-nADT
564 and no-nADT samples (i.e., from different patients but matched for ages and tumor grades and stages),
565 nADT increased *AR* mRNA levels (Fig. S4A) but reduced both c_AR-A and Stemness (Fig. S4B-C).
566
567 **Pri-PCa progression is accompanied by reduced c_AR-A but continually increasing Stemness.**
568 We next assessed the dynamic changes of Stemness and c_AR-A during *spontaneous* (i.e., therapy-free)
569 Pri-PCa progression (Fig. 2). To this end, we sorted out treatment-naïve tumors in TCGA-PRAD by
570 excluding samples from patients who received adjuvant (post-surgery) hormone therapy (i.e., Pri-PCa/ADT
571 or Tx; n=70) (Fig. 2A) and grouped the Pri-PCa samples according to increasing GS (Fig. 2B-D). When
572 compared with the 52 benign (N) samples, all the four GS groups from Pri-PCa (GS6, GS7, GS8, and
573 GS9/10 combined) showed higher c_AR-A and Stemness (Fig. 2C-D), consistent with the results in
574 pairwise T/N comparisons (Fig. 1A-C). However, when compared to GS6 PCa, the three higher GS groups
575 showed increasing *AR* mRNA levels (Fig. 2B) but decreasing c_AR-A (Fig. 2C) and increasing Stemness
576 (Fig. 2D). Jonckheere-Terpstra's (J-T) trend test validated a gradual reduction in canonical AR signaling
577 but an increasing trend of Stemness with Pri-PCa progression (Fig. 2C-D). Notably, although we observed
578 a modest positive correlation between c_AR-A and Stemness in pooled Pri-PCa samples (Pearson's
579 r=0.27; Fig. 2E), this positive correlation gradually diminished with PCa progression when we categorized
580 the samples by tumor grade (Fig. 2F). The linear regression analysis confirmed that the interaction
581 between Stemness and c_AR-A significantly depended on tumor grade (P=0.0003), with a linear decrease
582 in the slope of the correlation as GS increased (Fig. 2G).
583 Collectively, these results indicate that, in contrast to early tumorigenesis, *spontaneous PCa*
584 *progression (i.e., with increasing GS and tumor grade) is marked by a divergence between oncogenic*
585 *Stemness and pro-differentiation c_AR-A*.
586
587 **Therapies and metastatic progression drive further divergence between attenuated c_AR-A and**
588 **increased Stemness.**
589 Next, we asked how clinical treatment and metastatic progression may impact the Stemness and c_AR-A
590 by querying untreated Pri-PCa and Pri-PCa/ADT from TCGA and mCRPC in SU2C [10] (Fig. 3A-C). c_AR-
591 A was elevated in Pri-PCa compared to adjacent benign tissue (N) but reduced in more aggressive Pri-

592 PCa/ADT samples, while the Stemness, in contrast, continually and persistently increased from N → Pri-
593 PCa → Pri-PCa/ADT (Fig. 3B-C). Strikingly, mCRPC, despite markedly upregulating *AR* mRNA (Fig. 3A),
594 showed the lowest c_AR-A but highest Stemness (Fig. 3B-C), *indicating a shift from a positive correlation*
595 *between* c_*AR-A and Stemness in Pri-PCa to a negative association in mCRPC* (Fig. 3D).
596 By employing normalized RNAseq data from the Bolis 2021 PCa transcriptome atlas [36], which
597 included normal prostate specimens (n = 174), Pri-PCa (n = 714) and mCRPC (n = 335), we tracked the
598 Stemness and c_AR-A across the PCa spectrum (Fig. 3E). c_AR-A increased from benign to Pri-PCa but
599 declined in Pri-PCa/ADT and continued to decrease in mCRPC, reaching its lowest in mCRPC-NE [41] as
600 supported by J-T test (Fig. 3E). In contrast, the Stemness exhibited a continually increasing trajectory
601 across the PCa spectrum encompassing Normal/Benign → treatment-naïve Pri-PCa → Pri-PCa/ADT →
602 (m)CRPC (P<0.0001, J-T Trend test; Fig. 3F; Fig. S5A). Notably, the 3 subtypes of mCRPC, i.e., AR-
603 positive (ARPC), CRPC-NEPC and AR^-/NE^- double-negative PCa (DNPC) all showed higher Stemness
604 than Pri-PCa and displayed increasing Stemness among the 3 subtypes (Fig. S5B). Unexpectedly, tumor-
605 adjacent benign prostate tissues showed much higher Stemness than normal prostate in the Genotype-
606 Tissue Expression Project (GTEx) [18] (Fig. 3F; Fig. S5A), which were all obtained from cancer-free organ
607 donors. This surprising finding suggests that the transcriptomes of adjacent benign tissues have apparently
608 skewed towards those of cancer samples, consistent with previous observations that adjacent benign
609 tissues represent an intermediate state between healthy and tumor tissues [66]. Also of interest, the
610 Cancer Cell Line Encyclopedia (CCLE) [50, 67] cell lines including PCa cell lines, and PCa-derived PDX
611 [46] and xenografts (XG) [51] exhibited the highest Stemness (Fig. 3F; Fig. S5A), supporting that cancer
612 cell lines and xenografts are clonally derived from and highly enriched in cancer stem cells (CSCs).
613 As in our early dissection of the relationship between c_AR-A and Stemness during spontaneous Pri-
614 PCa progression (Fig. 2F-G), analysis of the Bolis 2021 PCa transcriptome atlas [36] revealed a similar
615 change from a positive correlation between c_AR-A and Stemness in Pri-PCa (blue line; Pearson's r=0.32,
616 P<0.0001; Fig. 3G) to a negative correlation between the two in mCRPC (green line; Pearson's r=-0.15,
617 P=0.005; Fig. 3G). We further extended our Stemness and c_AR-A interrogation to include microarray-
618 derived transcriptomic datasets [35, 47], allowing us to access data from PCa cohorts generated before
619 the next-generation sequencing (NGS) era. Analyzing the GRID database of over 82,000 prospectively
620 collected biopsy samples of localized PCa from clinical use of the Veracyte Decipher test [35], we
621 systematically quantified the Stemness across the NCCN Risk Groups and observed increasing Stemness
622 along the NCCN Risk trajectory (Fig. 3H). Likewise, in the Taylor dataset [47], we observed that c_AR-A
623 was the highest in Pri-PCa but significantly decreased in metastatic PCa (mPCa) (Fig. 3I) while the
624 Stemness increased progressively from benign to Pri-PCa and reached the highest levels in mPCa (Fig.
625 3J). Linear regression analysis confirmed a gradual loss of positive correlation between c_AR-A and
626 Stemness during PCa progression, as illustrated by distinct regression lines for each stage (Fig. 3K-M).

627 Finally, we analyzed the changes of Stemness and c_AR-A in genetically engineered mouse models
628 (GEMMs) of PCa with increasing aggressiveness due to individual or combined deletion of 3 tumor
629 suppressor genes, *Pten*, *Rb1* and *Trp53* [49] (Fig. S5C-H). Briefly, *Pten*$^{-/-}$ single knockout (SKO) prostate
630 tumors develop around 9 weeks, and mice rarely develop metastasis with a median lifespan of 48 weeks.
631 In contrast, the *Pten*$^{-/-}$;*Rb1*$^{-/-}$ double knockout (DKO) mice develop highly metastatic PCa that shortens
632 median survival to ~38 weeks whereas the triple KO (TKO; *Pten*$^{-/-}$;*Rb1*$^{-/-}$;*Trp53*$^{-/-}$) tumors are exclusively
633 AR^{-} NEPC and castration-resistant *de novo* with high metastatic rate and lifespan of ~16 weeks [49]. We
634 found that the c_AR-A was the highest in SKO tumors but much reduced in aggressive DKO and TKO
635 tumors (Fig. S5C). In contrast, Stemness was the lowest in SKO but significantly increased in DKO and
636 TKO (Fig. S5D). Correlation analyses revealed a positive correlation between c_AR-A and Stemness in
637 SKO that turned negative in both DKO and TKO tumors (Fig. S5E-H).
638 The above results, collectively, indicate that *hormone therapy and metastatic progression further*
639 *drive marked increases in Stemness with concomitant decreases in* c_*AR-A*.
640
641 **The general Stemness score is prognostic of poor survival in PCa patients.**
642 We investigated global Stemness and its prognostic significance, and observed that high Stemness
643 consistently correlated with worse patient survival across various cohorts including the Spratt 2017 [32]
644 (Fig. 4A), Tosoian 2020 [33] (Fig. 4B), CHAARTED 2021 [34] (Fig. 4C), and the mCRPC (SU2C 2019 [48];
645 Fig. S6A-B) as well as Taylor 2010 [47] (Fig. S6C-D). Notably, the CHAARTED cohorts displayed the
646 highest mean Stemness compared to Tosoian and Spratt cohorts, and the highest median Decipher
647 Prostate Genomic Classifier [32-34] among the three cohorts (Fig. 4D). Interestingly, like high Stemness,
648 low c_AR-A was correlated with poor patient survival in the Taylor dataset (Fig. S6E).
649
650 **Stemness-high PCa are associated with aggressive molecular subtypes and a new 12-gene PCa-**
651 **Stem signature prognosticates poor patient survival.**
652 To identify key determinants of Stemness and explore differences between tumors with high versus low
653 Stemness scores, we stratified Pri-PCa (TCGA-PRAD) and mCRPC (SU2C 2019) samples separately
654 based on Stemness scores (Fig. 5A). Differential gene expression (DEG) analyses revealed significant
655 differences between the top 33% (Stemness-high) and the bottom 33% (Stemness-low) PCa samples in
656 both treatment-naïve Pri-PCa (Fig. 5B) and treatment-failed mCRPC (Fig. 5C), resulting in 12 shared
657 genes that we considered as the "**PCa-Stem signature**" (Fig. 5D; Table S2; see Methods).
658 Gene set enrichment analysis (GSEA) revealed that Stemness-high PCa samples were enriched in
659 gene sets associated with embryonic stem cells (ESC) as well as aggressive cancer phenotypes such as
660 undifferentiated cancer and metastasis in both Pri-PCa and mCRPC whereas Stemness-low PCa samples
661 were enriched in gene signatures linked to lower aggressiveness and inflammatory pathways (Fig. 5E-F).
662 Additionally, hallmark pathways such as E2F targets, MYC targets, mTORC1 signaling, and DNA repair

663 were upregulated in Stemness-high PCa (Fig. 5F), while TGF-β signaling, TNF-α signaling, and IFN-γ
664 responses were prominent in Stemness-low PCa (Fig. 5E).

665 The aggressive PAM50-LumB molecular subtype [68-70] was significantly more prevalent in
666 Stemness-high PCa, accounting for 74% in Pri-PCa (compared to 15% in Stemness-low PCa; Fig. 5G)
667 and 43% in mCRPC (compared to 4% in Stemness-low PCa; Fig. 5H). The PAM50-Basal subtype [68, 69]
668 was also increased in Stemness-high mCRPC but not in Pri-PCa (Fig. 5G-H), aligning with the
669 characteristics of mCRPC, which involves lineage plasticity and acquisition of basal, mesenchymal, neural
670 and stem-like phenotypes when developing ARPI resistance [7, 70, 71]. The PCS1 subtype [72, 73], known
671 for its association with ADT resistance and the highest risk of progression to advanced disease in
672 comparison with PCS2 or PCS3, was enriched in both Stemness-high Pri-PCa and mCRPC whereas
673 PCS3 signature was enriched in Stemness-low Pri-PCa and mCRPC (Fig. 5I-L). Moreover, in PCa
674 subtypes epigenetically stratified by chromatin accessibility [74], Stemness-high samples remained largely
675 AR-dependent in Pri-PCa (Fig. 5M) but transitioned to AR-independent NE subtype in mCRPC (Fig. 5N).
676 Of interest, Stemness-high PCa was characterized by high proliferation in both Pri-PCa and mCRPC as
677 evidenced by their enrichment in a 31-gene cell-cycle progression signature [70] (Fig. 5O-P; Table S5).
678 Finally, we examined the association between Stemness and a signature of lineage plasticity risk in
679 mCRPC after Enza treatment [12] and found that Stemness-high tumors was at risk for this virulent form
680 of treatment resistance (Fig. 5Q-R).

681 To further assess the clinical relevance of Stemness, we analyzed the prognostic impact of our newly
682 developed **12-gene PCa-Stem signature** (Fig. 5D). We observed that higher PCa-Stem signature
683 correlated with worse patient survival in both Pri-PCa (P<0.0001, log-rank; Fig. 5S-T) and mCRPC
684 (P<0.0001, log-rank; Fig. 5U-V) patients. Multivariate analysis, adjusted for clinicopathologic parameters
685 such as age, GS and stage, showed that patients with high PCa-Stem signature had significantly worse
686 independent prognosis compared to those with low PCa-Stem signature. This was demonstrated by worse
687 progression-free survival (PFS) in Pri-PCa (multivariable-adjusted HR=1.636 (1.014 – 2.64), P=0.0438;
688 Fig. 5T) and overall survival (OS) in mCRPC (multivariable-adjusted HR=4.712 (2.011 – 11.04), P<0.001;
689 Fig. 5V), respectively. Furthermore, the PCa-Stem signature steadily and continually increased across the
690 PCa evolutionary spectrum and correlated with PCa disease progression scores as supported by trajectory
691 inference analysis of the Bolis PCa transcriptome atlas [36] (Pearson's r=0.72; Fig. S7A-B). Strikingly,
692 gene expression analysis revealed that 11 of the 12 genes (except *KLK12*) in the PCa-Stem signature
693 significantly correlated with the PCa disease progression (Fig. S7C). In contrast, among the genes
694 commonly downregulated in Stemness-high PCa samples in both Pri-PCa and mCRPC (Table S3), some
695 (e.g., *PTGDS*, *SPARCL1*, *CLU*, *GJA1* and *S100A4*) were involved in regulating prostate-specific glandular
696 structures and luminal functions while others (e.g., *FBLN1*, *SFRP1*, *IGFBP6*, *GAS1* and *TIMP2*) played
697 general roles in epithelial differentiation (Fig. S7D).

We further compared the 12-gene PCa-Stem signature with the general Stemness Score using ROC analysis and observed that PCa-Stem signature demonstrated enhanced discriminative power in distinguishing high-grade (GS > 7) from low-grade (GS ≤ 7) primary tumors (AUC = 0.662; Fig. S7E) and mCRPC versus treatment-naïve Pri-PCa (AUC = 0.865; Fig. S7F). These findings indicate that while the global Stemness Score effectively captures tumor dedifferentiation and plasticity, the PCa-Stem signature provides added PCa–specific sensitivity for identifying aggressive and advanced disease states.

Finally, we employed siRNAs to knock down 3 representative genes of the PCa-Stem signature, i.e., *HMMR* (encoding hyaluronan-mediated motility receptor), *AURKB* (encoding Aurora Kinase B), and *PBK* (encoding PDZ-binding kinase) (Fig. S8A) and analyzed their impact on the aggressiveness and stemness (Fig. S8B–F) in androgen-independent LAPC4 (LAPC4-AI) cells [51, 61]. The results revealed that knockdown of all 3 genes inhibited cell invasion (Fig. S8B–C) and single cell-based colony formation (Fig. S8D) whereas knockdown of AURKB and PBK, two mitotic kinases, significantly inhibited the establishment of live, stemness-associated organoids (Fig. S8E–F).

Taken together, these findings indicate that *PCa with high Stemness exhibit increased aggressiveness, lineage plasticity, and therapy resistance, and is associated with poor patient survival.*

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

To explore potential genomic differences between Stemness-high and Stemness-low prostate tumors, we first compared global genomic instability and found that mCRPC exhibited a greater fraction of altered genome (Fig. 6A) and higher TMB (Fig. 6B) than Pri-PCa. Furthermore, within both Pri-PCa and mCRPC, the Stemness-high tumors displayed a significantly higher fraction of altered genome (Fig. 6A) and TMB (Fig. 6B) than Stemness-low tumors, despite no differences in patient age between the groups (Fig. 6C). These findings support a strong association between oncogenic Stemness and genomic instability. We next examined the distribution of recurrent cancer-associated alterations in Stemness-defined subgroups across disease states (Fig. 6D–E). In treatment-naïve Pri-PCa, Stemness-high tumors were enriched for tumor-suppressor lesions (including *PTEN* and *RB1* deletions (DEL) and *TP53* mutations (MUT)), and alterations in known PCa drivers such as *SPOP* and *FOXA1* MUT and focal *MYC* amplification (AMP) (Fig. 6D). In contrast, the Stemness-high mCRPC exhibited markedly increased frequencies of therapy-selected drivers such as AR alterations (*AR* MUT, *AR* AMP; Fig. 6F). Furthermore, PI3K-pathway lesions, including *PIK3C2B* AMP (Fig. 6E), were enriched in Stemness-high mCRPC in contrast to their low prevalence in Pri-PCa. Several additional amplified loci—including *ELK4*, *RECQL4*, *SLC45A3*, and *ELOC*—were present in both disease states but showed substantially higher frequencies in mCRPC (Fig. 6D-E). Notably, copy-number gain of *SLC45A3*, an androgen-regulated gene, was enriched in Stemness-high tumors and increased in frequency in mCRPC (Fig. 6E). These results support selective amplification of prostate lineage–associated loci under ARPI treatment and during progression to mCRPC.

733 Globally, both Pri-PCa and mCRPC exhibited mutations in more than a dozen genes although most
734 of these alterations were <10% (Fig. S9A; green symbols), consistent with reports that PCa is genomically
735 heterogeneous with a broad spectrum of genomic alternations of low penetrance [75, 76]. On the other
736 hand, homozygous deletion (HOMDEL) events, particularly, loss of *PTEN* and *RB1*, occurred frequently in
737 both Pri-PCa and mCRPC (Fig. S9A; blue symbols) whereas AMP events in oncogenes such as *MYC*
738 were more prevalent in mCRPC (Fig. S9A; red symbols). Nevertheless, *PTEN* and *RB1* HOMDEL
739 represented among the most prevalent alterations in early-stage (GS6) PCa (Fig. S9B-D). Analysis of
740 genetic alterations across the spectrum of PCa progression revealed that among the 3 tumor suppressors
741 (*RB1, PTEN* and *TP53*), while *RB1* HOMDEL remained relatively constant at 13%, 6%, 10% and 10% in
742 low-grade (GS6), high-grade (GS9/10), Pri-PCa/ADT and mCRPC, respectively, *PTEN* HOMDEL (9%,
743 23%, 30% and 26%) and *TP53* MUT (0%, 21%, 27% and 37%) continued to increase along this
744 progression (Fig. S9B-D). Among the 3 oncogenic events analyzed, while high levels of *TMPRSS2* fusions
745 (~40%) remained unchanged from low-grade Pri-PCa to mCRPC, *MYC* (2%, 12%, 19% and 24%) and *AR*
746 (0%, 1%, 6% and 49%) AMP continued to increase along the progression trajectory (Fig. S9B-C, E). In
747 fact, prevalent *AR* MUT (11%) were observed only in the most aggressive and Stemness-highest mCRPC
748 (Fig. S9B-C). Analyzing these major genomic alterations in association with PCa Stemness, we observed
749 that *RB1* HOMDEL, *FOXA1* (together with *SPOP*, *IDH1*, and *CHD1*) MUT, and especially *MYC* AMP were
750 significantly associated with increased Stemness in Pri-PCa (Fig. S9F) whereas in mCRPC, only *RB1*
751 HOMDEL exhibited a statistically significant association with higher Stemness (Fig. S9G).
752
753 **MYC drives Stemness during PCa progression.**
754 The above studies (Fig. 6; Fig. S9) provided genomic correlates of PCa stemness. *In following studies, we*
755 *explored 1) MYC signaling, 2) noncanonical (i.e., castration-induced) AR activity, and 3) tumor suppressor*
756 *loss as potential drivers of persistently increasing Stemness during PCa progression*. We showed that
757 during early PCa growth, increased c_AR-A may drive elevated Stemness (Fig. 1D-F); however, treatment-
758 naïve Pri-PCa displayed continually increasing Stemness but decreasing c_AR-A (Fig. 2). As *AR* AMP and
759 MUT are exclusively treatment-induced and there were virtually no *AR* genomic alterations in treatment-
760 naïve Pri-PCa (Fig. S9B-C, E), what could be driving high Stemness during Pri-PCa progression? We
761 found that Stemness-high Pri-PCa, compared to Stemness-low Pri-PCa, were enriched in MSigDB
762 Hallmark MYC_TARGETS (Fig. 5F) and had more prevalent *MYC* AMP (Fig. 6D and 6F). High-grade
763 (GS9/10) Pri-PCa also had more *MYC* AMP events than low-grade (GS6) Pri-PCa, and in fact, *MYC* AMP
764 continued to increase along the PCa continuum (Fig. S9B-C, E) and was associated with increased
765 Stemness in Pri-PCa (Fig. S9F). These observations led us to hypothesize that MYC may represent a
766 critical Stemness driver during both spontaneous Pri-PCa progression and mCRPC transition. To test this
767 hypothesis, we quantified MYC signaling activity [55] (**MYC-sig**; See Methods and Table S2 for gene
768 signature information) across the PCa progression spectrum. We observed that: 1) both *MYC* mRNA and

769 MYC-sig increased in Pri-PCa compared to matched benign samples (Fig. 7A-B); 2) MYC-sig continued
770 to go up during Pri-PCa progression (Fig. 7C); 3) unlike nADT-induced concomitant decline in both c_AR-
771 A and Stemness (Fig. 1D-F; Fig. S4), nADT showed no discernible effect on MYC-sig (Fig. S10A-D); and
772 4) MYC-sig continued to increase across the disease spectrum with mCRPC exhibiting the highest MYC
773 activity (Fig. 7D-E) and with MYC-sig positively correlated with the Stemness (Fig. 7F-H).
774 To further test whether MYC functionally promotes stemness, we examined an independent RNA-
775 seq dataset generated in LNCaP cells with MYC knockdown (KD) (GSE135942) [52]. Silencing MYC using
776 two ORF-targeting shMYC constructs led to a marked reduction in MYC-sig (Fig. S10E), confirming
777 effective suppression of the MYC transcriptional output. Importantly, MYC-KD significantly reduced the
778 Stemness score (Fig. S10F), thus supporting a direct role for MYC in maintaining the stem-like
779 transcriptional state. Strikingly, MYC-KD also increased c_AR-A (Fig. S10G), consistent with the reported
780 inverse relationship between pro-differentiation AR signaling and oncogenic MYC activity [77, 78].
781 *These results, together, support MYC as a Stemness driver during PCa progression*.
782
783 **Castration-reprogrammed AR activity (cr_AR-A) significantly rises and aligns with stem-like**
784 **programs during PCa progression.**
785 Having established MYC as a driver of Stemness during PCa progression, we next asked how therapeutic
786 pressure remodels AR signaling to sustain or amplify the stem-like programs. Androgen deprivation and
787 AR-pathway inhibition remodel the AR cistrome and redirect AR toward non-canonical survival/plasticity
788 programs [14-17]. To capture this shift, we derived the **cr_AR-A** by integrating AR and H3K27ac ChIP-seq
789 data from Pomerantz *et al.* (GSE130408; [16]) with mRNA profiling data from Taylor *et al.* (GSE21032;
790 [47]) (Fig. S11A; Table S2). We intersected genes that exhibited CRPC-enriched AR binding, enhancer
791 activation, and transcriptional up-regulation to obtain a 63-gene set representing AR-regulated outputs
792 preferentially active in mCRPC (Fig. S11A; Table S6).
793 As expected, cr_AR-A increased sharply in androgen-independent (AI) versus androgen-
794 dependent (AD) LNCaP xenografts, whereas the canonical AR program (c_AR-A) declined (Fig. S11B).
795 Genome-browser views confirmed that enhancer-level reprogramming [16] for *AR*, *UBE2C*, *CDK1*, *MELK*,
796 *BIRC5*, and many other oncogenic and proliferation-associated genes displayed stronger AR and H3K27ac
797 peaks in CRPC than in Pri-PCa (Fig. S11C; data not shown), while *CNN1* and *TNS1* showed the opposite
798 pattern (Fig. S11D). To further benchmark the cr_AR-A, we compared it with the Wang "**cr_AR-Growth**"
799 8-gene set [14] curated from AR-bound, enhancer-validated and AI LNCaP-abl growth drivers (e.g.,
800 *UBE2C*, *CDK1*, *CDC20*) (Fig. S11E). Despite distinct derivations, the two signatures, with only one
801 overlapped gene (i.e., *CDK1*), were highly concordant ($r = 0.93$, $P < 0.0001$; Fig. S11F–H) and both
802 increased from normal/benign to Pri-PCa to mCRPC (Fig. 8A; S11I), underscoring a shared transcriptional
803 logic linking therapy-reprogrammed AR signaling with stemness-associated proliferative activity.

Together, these data reveal a state-dependent bifurcation of AR programs: the c_AR-A tracks differentiation in androgen-intact contexts whereas the cr_AR-A couples AR signaling to stem-like aggressiveness under therapeutic pressure, establishing cr_AR-A as a transcriptomic bridge between AR adaptation and PCa evolution.

***RB1* and *PTEN* losses differentially associate with Stemness along the disease continuum.**

Finally, we examined whether transcriptomic signatures associated with tumor suppressor losses may also reinforce therapy-driven Stemness. Prior genomic analyses (Fig. 6; Fig. S9) showed that, in Pri-PCa, *RB1* DEL, *FOXA1* MUT, and especially *MYC* AMP associate with elevated Stemness whereas *PTEN* loss is prevalent early and persists into advanced disease and that in mCRPC, *RB1* DEL significantly correlated with high Stemness. To further assess these trends transcriptionally, we quantified the *RB1*-loss and *PTEN*-loss gene signatures described by Ertel *et al.* [56] and Liu *et al.* [57], respectively. We found that the *RB1*-loss signature increased progressively from benign to Pri-PCa to mCRPC (Fig. 8B) whereas the *PTEN*-loss signature rose sharply from normal prostate to Pri-PCa but plateaued thereafter, peaking in aggressive Pri-PCa/ADT and remaining high in mCRPC (Fig. 8C). These analyses suggest that *PTEN* loss might function as an early and persistent driver of spontaneous progression whereas *RB1* loss represents a more prominent contributor to Stemness and proliferative reprogramming in therapy-resistant disease.

Convergence of cr_AR-A, *RB1*-loss, and MYC signaling with PCa-Stem and global Stemness score

Building on the above analyses showing progressive increase in cr_AR-A (Fig. 8A) and distinct dynamics of *RB1*- and *PTEN*-loss signatures (Fig. 8B-C), we next examined how these programs relate quantitatively to the general Stemness score and PCa-Stem signature (Fig. 8D-P). Consistent with its progressive rise across disease states, the PCa-Stem score increased from normal prostate and Pri-PCa to mCRPC (Fig. 8D), broadly tracking the trajectory of cr_AR-A and *RB1*-loss (Fig. 8A-B). Across the Bolis integrated cohort (n = 1,223), the cr_AR-A correlated strongly with both general Stemness (Fig. 8E; see Fig. 8L for symbol definitions) and the PCa-Stem (Fig. 8I). In parallel, the LNCaP-abl-derived cr_AR-Growth_Wang signature also correlated with the two Stemness indices (Fig. S11J-K). Similarly, the *RB1*-loss signature correlated with the global Stemness ($r = 0.74$, $P < 0.0001$; Fig. 8F) and even more strongly with PCa-Stem ($r = 0.91$, $P < 0.0001$; Fig. 8J) whereas the *PTEN*-loss signature exhibited a more modest association with the two Stemness indices (Fig. 8G, K).

We next integrated MYC signaling into this framework. Consistent with our earlier MYC analyses (Fig. 7), MYC-sig increased along the PCa evolution continuum in the Bolis integrated PCa cohort (Fig. S11M) and correlated strongly with both the global Stemness (Fig. S11N) and PCa-Stem (Fig. S11O). Of interest, despite this strong transcriptional convergence, the MYC-sig and PCa-Stem showed no direct gene overlap (Fig. S11P), indicative of their convergence at the level of coordinated downstream proliferative programs rather than shared gene membership.

Across the same cohort, PCa-Stem displayed the strongest positive correlation with the general Stemness (Fig. 8H), consistent with the PCa-Stem being derived from Stemness-stratified DEGs (Fig. 5B-D). Gene-overlap analyses further supported convergence among these programs. The PCa-Stem shared *MELK* with the 63-gene cr_AR-A signature (Fig. 8M) and shared *UBE2C* with the 8-gene cr_AR-Growth_Wang (Fig. S11L). PCa-Stem also showed broader overlap with the *RB1*-loss signature (Fig. 8N), including *AURKB*, *BIRC5*, *CENPA*, *DLGAP5*, and *NEK2*, genes largely involved in cell-cycle progression and mitotic regulation. In contrast, no direct gene overlap was observed between *PTEN*-loss and PCa-Stem (Fig. 8O). When signatures were integrated (Fig. 8P), the proliferative core was most evident between cr_AR-A and the *RB1*-loss signature (i.e., *ANLN*, *RRM2*, *MCM4*, *TYMS*, *PLK1*, *CCNE2*, and *CCNB1*), supporting convergent selection for cell-cycle machinery characteristic of stem-like, self-renewing tumor cells.

852 **Discussion**

853 The current project was undertaken to quantitatively co-analyze the global changes of and interrelationship
854 between cancer stemness and canonical AR signaling activity during PCa evolution. By employing gene
855 expression-based general Stemness and c_AR-A signature scores and through developing a new 12-gene
856 PCa-Stem signature, we show that while differentiation-associated c_AR-A steadily declines in the PCa
857 continuum, progressively increasing Stemness quantitatively tracks PCa aggressiveness, plasticity and
858 progression, and represents a poor prognosticator for patient outcomes. Stemness-high PCa possess
859 unique genomic features including elevated TMBs and are transcriptionally enriched in aggressive
860 molecular subtypes such as PAM50-LumB [68-70] and PCS1 [72, 73]. Our data further suggest that
861 elevated Stemness in early PCa is causally associated with increased c_AR-A while castration-
862 reprogrammed AR-A (cr_AR-A), MYC activation and tumor suppressor alterations (especially *RB1* loss)
863 likely propagate the markedly increased Stemness observed in mCRPC (where c_AR-A is suppressed).
864 Early studies reported that AR-regulated genes decline in high-grade Pri-PCa [79, 80] and inhibition
865 of AR signaling promotes stem-like phenotypes [8, 51, 81, 82]. However, global pro-differentiation AR
866 activity and oncogenic stemness have not been systematically and quantitatively co-analyzed across the
867 PCa continuum. By integrating 27 transcriptomic datasets encompassing normal prostate, tumor-adjacent
868 benign tissues, treatment-naïve Pri-PCa, nADT-treated tumors, (m)CRPC subtypes and model systems,
869 we demonstrate that PCa progression is accompanied by continual increase in Stemness and progressive
870 decline in canonical c_AR-A, with the two inversely correlated (Fig. S12). Importantly, both the global
871 Stemness and PCa-Stem signature prognosticate poor patient survival, establishing Stemness as a
872 quantifiable and clinically relevant indicator of PCa progression. To our knowledge, this is the first
873 comprehensive study to simultaneously quantify the oncogenic stemness and c_AR-A across the PCa
874 continuum. Our pan-cohort approach enables signature validation in rare archival biopsy samples and
875 cross-platform integration — a translational advantage not addressed by other studies.
876 The normal prostatic epithelium consists of AR^+ secretory luminal cells and AR^- basal cells. We find
877 that NHP luminal cells have higher c_AR-A whereas basal cells display higher Stemness, consistent with
878 earlier findings from our lab [21] and others [19, 20, 64, 65] showing stem/progenitor cell properties within
879 the basal compartment. Compared to adjacent benign tissue, Pri-PCa exhibit elevated Stemness, which
880 is associated with and likely driven by increased c_AR-A. Early genomic alterations including *PTEN* and
881 *RB1* loss in GS6 tumors (9% and 13%, respectively; Fig. S9D) may further contribute to this initial rise in
882 Stemness. Strikingly, tumor-adjacent benign tissues, the so-called 'normal' tissues used in most
883 comparative studies, displayed higher Stemness than the true normal prostate obtained from cancer-free
884 organ donors in GTEx, indicating that the commonly used "benign" controls may already exhibit
885 transcriptional shifts toward cancer-associated programs.

886 Our studies suggest that the drivers of increased Stemness may differ between spontaneous Pri-
887 PCa progression and mCRPC transition. In treatment-naïve Pri-PCa, where tumor progression is
888 characterized by a gradual decline in c_AR-A while Stemness shows a continual elevation (Fig. S12),
889 increased Stemness may likely be driven by oncogenic events such as *MYC* amplification and activation,
890 with additional contributions from early loss of tumor-suppressors such as *PTEN* and *RB1*. On the other
891 hand, in mCRPC, which develops under sustained therapeutic pressure from long-term ADT and ARPI
892 treatment, the c_AR-A is further suppressed while Stemness reaches its highest levels, This paradoxical
893 increase in Stemness despite reduced c_AR-A coincides with emergence of the cr_AR-A, which is
894 enriched for cell-cycle and proliferation-associated genes and correlates tightly with *RB1* loss and MYC
895 activation (Fig. S12). Thus, unlike in Pri-PCa, the high Stemness in mCRPC appears to be reinforced by
896 AR functional reprogramming that cooperates with genomic instability to sustain aggressive, proliferative
897 and therapy-resistant stem-like states.

898 What biological correlates do our Stemness metrics measure? ***First***, it measures transcriptional
899 features and relative abundance of stem cells and CSCs. Therefore, the NHP basal/stem cells display
900 higher Stemness than luminal cells (consistent with literature [19, 21, 64, 65]); ARPI treatment, while
901 shutting down the c_AR-A, enriches CSCs [7, 51, 81, 82] and upregulates Stemness (this study); and PCa
902 cell lines, xenografts and PDX, known to be clonally derived from cancer stem/progenitor cells,
903 demonstrate the highest Stemness. ***Second***, it measures cellular and lineage plasticity [8] induced by
904 treatment and genetic alterations. Thus, mCRPC, which harbor more abundant treatment-reprogrammed
905 $AR^{-/lo}$ cells and PCa cells with stem-like features or in an AR-indifferent state [8, 12, 13, 23, 46, 54, 71, 74],
906 manifests the highest Stemness. Our analysis further substantiates the aggressive nature of ($AR^{-/lo}$) CRPC-
907 NEPC and DNPC compared to ARPC, as evidenced by their elevated Stemness scores (Fig. S5B). This
908 suggests a shift in the phenotypic landscape of mCRPC towards more stem-like and aggressive forms.
909 Likewise, aggressive DKO and TKO murine prostate tumors have undergone lineage transformation and
910 exhibit higher Stemness than indolent SKO tumors. ***Third***, the Stemness metrics likely report the high
911 proliferative status of advanced PCa and mCRPC, as Pri-PCa and mCRPC with high global Stemness
912 both are enriched in "Cell-cycle progression" signature (Fig. 5O-P) while the majority of PCa-Stem genes
913 encode mitotic regulators (Fig. 5D). ***Finally***, the Stemness metrics quantitatively measure PCa
914 aggressiveness as high Stemness correlates not only with more advanced and aggressive stages of PCa
915 but also with aggressive molecular subtypes and poor survival.

916 A key implication of our analysis is that stemness in PCa represents a dynamic cell-state continuum
917 rather than a fixed CSC population. Studies in normal adult tissues have demonstrated that regenerative
918 capacity is maintained through the coexistence of quiescent reserve stem cells and fast-cycling active stem
919 cells, with cells reversibly transitioning between these states in response to stress or injury [83]. Applying
920 this framework to PCa, our data suggest that both the general Stemness Index and PCa-Stem signature
921 preferentially reflect a fast-cycling, proliferation-enriched stem-like state. This framework explains why

Stemness progressively increases despite declining c_AR-A associated with cellular maturation: therapeutic pressure and accumulating genomic alterations select for proliferating and plastic cell states rather than a static CSC compartment. Our functional validation provides support for this proliferation-driven cancer stemness - depletion of mitotic kinases *AURKB* and *PBK* markedly suppressed invasion, clonogenic growth, and 3D organoid viability (Fig. S8). Notably, this proliferative stem-like state mirrors transcriptional programs captured by cr_AR-A and cr_AR-growth signatures, both enriched for cell-cycle genes and strongly correlated with global Stemness and PCa-Stem scores. Together, these data support convergence on a fast-cycling, therapy-adapted stem-like state in aggressive PCa, consistent with recently reported convergence of CSC states across tumor types [84].

Of clinical significance, both the general Stemness Index and the 12-gene PCa-Stem signature correlate with poor patient survival, suggesting the possibility of developing the Stemness metrics 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 PCa-Stem genes correlate with PCa progression (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 subtypes [6], our findings herein underscore the increased Stemness as a 'universal' and fundamental characteristic of cancer progression and aggressiveness. Beyond its ability to track tumor dedifferentiation and lineage plasticity across the PCa continuum, the PCa-Stem signature demonstrated improved disease-focused classification compared with the pan-cancer Stemness Index (Fig. S7E-F). This does not diminish the biological value of the general Stemness Index; rather, PCa-Stem complements it by providing increased PCa–specific sensitivity.

In recent years, several studies [36, 85-88] have proposed stemness-based PCa 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 quantitative, clinically validated Stemness metrics 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-seq [87]. 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 models [87]. 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

958 stemness-based PCa subtypes and constructed a nine-gene prognostic signature [85] that also included
959 *HMMR* and *DLGAP5*. *KLK12*, another PCa-Stem gene, was recently identified as a unique marker for
960 SOX9^{High} stem-like luminal epithelial cells enriched following androgen deprivation [89]. Additionally, our
961 findings that *MYC* amplification, *RB1* deletion, and *PTEN* loss are enriched in Stemness-high tumors are
962 consistent with the previously reported chromatin accessibility (ATAC-seq) patterns [74] and multi-omics
963 spatial data [89]. It is worth noting that although our PCa-Stem signature was initially designed for risk
964 stratification in mCRPC, the analytical framework is inherently flexible as the same platform can be
965 extended to derive subtype-specific signatures.

966 We recognize that the Stemness algorithm utilized in our studies was not developed specifically for
967 PCa and therefore, its sensitivity to distinguish subtle intra-group differences remains to be fully validated.
968 The bulk RNA-seq data used to derive Stemness, c_AR-A and other signature scores was not from pure
969 epithelial cancer cells, a technical issue that can be potentially mitigated by scRNA-seq analysis and other
970 Stemness pipelines such as the CytoTRACE [90, 91] and by using cancer-specific genomic features
971 derived, for example, from cell-free DNA [92].

972 In conclusion, we establish transcriptome-based signature scores to assess pro-differentiation AR
973 activity, MYC signaling, and oncogenic stemness across the PCa continuum. Stemness progressively
974 increases during spontaneous PCa progression and becomes much heightened in mCRPC through AR
975 reprogramming, *RB1*-loss associated transcriptional programs, and MYC activation. These findings
976 provide a quantitative framework linking PCa progression, oncogenic dedifferentiation, proliferative
977 reprogramming, lineage plasticity and therapy resistance, and identify stemness-associated transcriptional
978 programs as potential clinical vulnerabilities. The signature scores reported herein have the potential to
979 serve as clinical classifiers and aid in the identification of patients who may benefit from stemness-targeting
980 therapies.
981

982 **References**

983 1. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, et al. An embryonic stem cell-
984 like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet.
985 2008;40(5):499-507.

986 2. Pece S, Tosoni D, Confalonieri S, Mazzarol G, Vecchi M, Ronzoni S, et al. Biological and
987 molecular heterogeneity of breast cancers correlates with their cancer stem cell content. Cell.
988 2010;140(1):62-73.

989 3. Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes
990 guides creation of epithelial cancer stem cells. Cell Stem Cell. 2008;2(4):333-44.

991 4. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic Plasticity: Driver
992 of Cancer Initiation, Progression, and Therapy Resistance. Cell Stem Cell. 2019;24(1):65-78.

993 5. Loh JJ, Ma S. Hallmarks of cancer stemness. Cell Stem Cell. 2024;31(5):617-39.

994 6. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, et al. Machine
995 Learning Identifies Stemness Features Associated with Oncogenic Dedifferentiation. Cell.
996 2018;173(2):338-54 e15.

997 7. Liu X, Li WJ, Puzanov I, Goodrich DW, Chatta G, Tang DG. Prostate cancer as a dedifferentiated
998 organ: androgen receptor, cancer stem cells, and cancer stemness. Essays Biochem. 2022;66(4):291-
999 303.

1000 8. Tang DG. Understanding and targeting prostate cancer cell heterogeneity and plasticity. Semin
1001 Cancer Biol. 2022;82:68-93.

1002 9. Siegel RL, Kratzer TB, Wagle NS, Sung H, Jemal A. Cancer statistics, 2026. CA Cancer J Clin.
1003 2026;76(1):e70043.

1004 10. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative
1005 clinical genomics of advanced prostate cancer. Cell. 2015;161(5):1215-28.

1006 11. He MX, Cuoco MS, Crowdis J, Bosma-Moody A, Zhang Z, Bi K, et al. Transcriptional mediators of
1007 treatment resistance in lethal prostate cancer. Nat Med. 2021;27(3):426-33.

1008 12. Westbrook TC, Guan X, Rodansky E, Flores D, Liu CJ, Udager AM, et al. Transcriptional profiling
1009 of matched patient biopsies clarifies molecular determinants of enzalutamide-induced lineage plasticity.
1010 Nat Commun. 2022;13(1):5345.

1011 13. Alumkal JJ, Sun D, Lu E, Beer TM, Thomas GV, Latour E, et al. Transcriptional profiling identifies
1012 an androgen receptor activity-low, stemness program associated with enzalutamide resistance. Proc Natl
1013 Acad Sci U S A. 2020;117(22):12315-23.

1014 14. Wang Q, Li W, Zhang Y, Yuan X, Xu K, Yu J, et al. Androgen receptor regulates a distinct
1015 transcription program in androgen-independent prostate cancer. Cell. 2009;138(2):245-56.

1016 15. Pomerantz MM, Li F, Takeda DY, Lenci R, Chonkar A, Chabot M, et al. The androgen receptor
1017 cistrome is extensively reprogrammed in human prostate tumorigenesis. Nat Genet. 2015;47(11):1346-
1018 51.

1019 16. Pomerantz MM, Qiu X, Zhu Y, Takeda DY, Pan W, Baca SC, et al. Prostate cancer reactivates
1020 developmental epigenomic programs during metastatic progression. Nat Genet. 2020;52(8):790-9.

1021 17. Formaggio N, Rubin MA, Theurillat JP. Loss and revival of androgen receptor signaling in
1022 advanced prostate cancer. Oncogene. 2021;40(7):1205-16.

1023 18. Consortium GT. The Genotype-Tissue Expression (GTEx) project. Nat Genet. 2013;45(6):580-5.

1024 19. Smith BA, Sokolov A, Uzunangelov V, Baertsch R, Newton Y, Graim K, et al. A basal stem cell
1025 signature identifies aggressive prostate cancer phenotypes. Proc Natl Acad Sci U S A.
1026 2015;112(47):E6544-52.

1027 20. Liu X, Grogan TR, Hieronymus H, Hashimoto T, Mottahedeh J, Cheng D, et al. Low CD38
1028 Identifies Progenitor-like Inflammation-Associated Luminal Cells that Can Initiate Human Prostate
1029 Cancer and Predict Poor Outcome. Cell Rep. 2016;17(10):2596-606.

1030 21. Zhang D, Park D, Zhong Y, Lu Y, Rycaj K, Gong S, et al. Stem cell and neurogenic gene-
1031 expression profiles link prostate basal cells to aggressive prostate cancer. Nat Commun. 2016;7:10798.

1032 22. Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. Cell.
1033 2015;163(4):1011-25.

1034 23. Zhang D, Hu Q, Liu X, Ji Y, Chao HP, Liu Y, et al. Intron retention is a hallmark and spliceosome
1035 represents a therapeutic vulnerability in aggressive prostate cancer. Nat Commun. 2020;11(1):2089.

1036 24. Liu J, Lichtenberg T, Hoadley KA, Poisson LM, Lazar AJ, Cherniack AD, et al. An Integrated
1037 TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. Cell.
1038 2018;173(2):400-16 e11.

1039 25. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, et al. Visualizing and
1040 interpreting cancer genomics data via the Xena platform. Nat Biotechnol. 2020;38(6):675-8.

1041 26. Rajan P, Sudbery IM, Villasevil ME, Mui E, Fleming J, Davis M, et al. Next-generation sequencing
1042 of advanced prostate cancer treated with androgen-deprivation therapy. Eur Urol. 2014;66(1):32-9.

1043 27. Sharma NV, Pellegrini KL, Ouellet V, Giuste FO, Ramalingam S, Watanabe K, et al. Identification
1044 of the Transcription Factor Relationships Associated with Androgen Deprivation Therapy Response and
1045 Metastatic Progression in Prostate Cancer. Cancers (Basel). 2018;10(10).

1046 28. Long X, Hou H, Wang X, Liu S, Diao T, Lai S, et al. Immune signature driven by ADT-induced
1047 immune microenvironment remodeling in prostate cancer is correlated with recurrence-free survival and
1048 immune infiltration. Cell Death Dis. 2020;11(9):779.

1049 29. Jamroze A, Zhang R, Ahuja K, Deng L, Jatwani K, Nguyen U, et al. The temporal dynamics of the
1050 immune response to neoadjuvant androgen deprivation therapy suggests a window-of-opportunity for
1051 checkpoint inhibitor therapy in prostate cancer. medRxiv. 2026. [Preprint]. 2026 Jan
1052 13:2026.01.10.26343859. doi: 10.64898/2026.01.10.26343859. PMID: 41646678.

1053 30. Gerhauser C, Favero F, Risch T, Simon R, Feuerbach L, Assenov Y, et al. Molecular Evolution of
1054 Early-Onset Prostate Cancer Identifies Molecular Risk Markers and Clinical Trajectories. Cancer Cell.
1055 2018;34(6):996-1011 e8.

1056 31. Wyatt AW, Mo F, Wang K, McConeghy B, Brahmbhatt S, Jong L, et al. Heterogeneity in the inter-
1057 tumor transcriptome of high risk prostate cancer. Genome Biol. 2014;15(8):426.

1058 32. Spratt DE, Yousefi K, Deheshi S, Ross AE, Den RB, Schaeffer EM, et al. Individual Patient-Level
1059 Meta-Analysis of the Performance of the Decipher Genomic Classifier in High-Risk Men After
1060 Prostatectomy to Predict Development of Metastatic Disease. J Clin Oncol. 2017;35(18):1991-8.

1061 33. Tosoian JJ, Birer SR, Jeffrey Karnes R, Zhang J, Davicioni E, Klein EE, et al. Performance of
1062 clinicopathologic models in men with high risk localized prostate cancer: impact of a 22-gene genomic
1063 classifier. Prostate Cancer Prostatic Dis. 2020;23(4):646-53.

1064 34. Hamid AA, Huang HC, Wang V, Chen YH, Feng F, Den R, et al. Transcriptional profiling of
1065 primary prostate tumor in metastatic hormone-sensitive prostate cancer and association with clinical
1066 outcomes: correlative analysis of the E3805 CHAARTED trial. Ann Oncol. 2021;32(9):1157-66.

1067 35. Weiner AB, Liu Y, Hakansson A, Zhao X, Proudfoot JA, Ho J, et al. A novel prostate cancer
1068 subtyping classifier based on luminal and basal phenotypes. Cancer. 2023;129(14):2169-78.

1069 36. Bolis M, Bossi D, Vallerga A, Ceserani V, Cavalli M, Impellizzieri D, et al. Dynamic prostate
1070 cancer transcriptome analysis delineates the trajectory to disease progression. Nat Commun.
1071 2021;12(1):7033.

1072 37. Suntsova M, Gaifullin N, Allina D, Reshetun A, Li X, Mendeleeva L, et al. Atlas of RNA
1073 sequencing profiles for normal human tissues. Sci Data. 2019;6(1):36.

1074 38. Stelloo S, Nevedomskaya E, Kim Y, Schuurman K, Valle-Encinas E, Lobo J, et al. Integrative
1075 epigenetic taxonomy of primary prostate cancer. Nat Commun. 2018;9(1):4900.

1076 39. Ramnarine VR, Alshalalfa M, Mo F, Nabavi N, Erho N, Takhar M, et al. The long noncoding RNA
1077 landscape of neuroendocrine prostate cancer and its clinical implications. Gigascience. 2018;7(6).

1078 40. Robinson DR, Wu YM, Lonigro RJ, Vats P, Cobain E, Everett J, et al. Integrative clinical
1079 genomics of metastatic cancer. Nature. 2017;548(7667):297-303.

1080 41. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, et al. Divergent clonal evolution of
1081 castration-resistant neuroendocrine prostate cancer. Nat Med. 2016;22(3):298-305.

1082 42. Sharp A, Coleman I, Yuan W, Sprenger C, Dolling D, Rodrigues DN, et al. Androgen receptor
1083 splice variant-7 expression emerges with castration resistance in prostate cancer. J Clin Invest.
1084 2019;129(1):192-208.

1085 43. Nyquist MD, Corella A, Coleman I, De Sarkar N, Kaipainen A, Ha G, et al. Combined TP53 and
1086 RB1 Loss Promotes Prostate Cancer Resistance to a Spectrum of Therapeutics and Confers
1087 Vulnerability to Replication Stress. Cell Rep. 2020;31(8):107669.

1088 44. Brady L, Kriner M, Coleman I, Morrissey C, Roudier M, True LD, et al. Inter- and intra-tumor
1089 heterogeneity of metastatic prostate cancer determined by digital spatial gene expression profiling. Nat
1090 Commun. 2021;12(1):1426.

1091 45. Lim Y, Arora S, Schuster SL, Corey L, Fitzgibbon M, Wladyka CL, et al. Multiplexed functional
1092 genomic analysis of 5' untranslated region mutations across the spectrum of prostate cancer. Nat
1093 Commun. 2021;12(1):4217.

1094 46. Labrecque MP, Coleman IM, Brown LG, True LD, Kollath L, Lakely B, et al. Molecular profiling
1095 stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. J
1096 Clin Invest. 2019;129(10):4492-505.

1097 47. Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y, Carver BS, et al. Integrative genomic
1098 profiling of human prostate cancer. Cancer Cell. 2010;18(1):11-22.

1099 48. Abida W, Cyrta J, Heller G, Prandi D, Armenia J, Coleman I, et al. Genomic correlates of clinical
1100 outcome in advanced prostate cancer. Proc Natl Acad Sci U S A. 2019;116(23):11428-36.

1101 49. Ku SY, Rosario S, Wang Y, Mu P, Seshadri M, Goodrich ZW, et al. Rb1 and Trp53 cooperate to
1102 suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. Science.
1103 2017;355(6320):78-83.

1104 50. Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, et al. The Cancer Cell
1105 Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. Nature.
1106 2012;483(7391):603-7.

1107 51. Li Q, Deng Q, Chao H-P, Liu X, Lu Y, Lin K, et al. Linking prostate cancer cell AR heterogeneity
1108 to distinct castration and enzalutamide responses. Nature communications. 2018;9(1):1-17.

1109 52. Whitlock NC, Trostel SY, Wilkinson S, Terrigino NT, Hennigan ST, Lake R, et al. MEIS1 down-
1110 regulation by MYC mediates prostate cancer development through elevated HOXB13 expression and AR
1111 activity. Oncogene. 2020;39(34):5663-74.

1112 53. Lee E, Chuang HY, Kim JW, Ideker T, Lee D. Inferring pathway activity toward precise disease
1113 classification. PLoS Comput Biol. 2008;4(11):e1000217.

1114 54. Bluemn EG, Coleman IM, Lucas JM, Coleman RT, Hernandez-Lopez S, Tharakan R, et al.
1115 Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. Cancer
1116 Cell. 2017;32(4):474-89.e6.

1117 55. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular
1118 Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015;1(6):417-25.

1119 56. Ertel A, Dean JL, Rui H, Liu C, Witkiewicz AK, Knudsen KE, et al. RB-pathway disruption in
1120 breast cancer: differential association with disease subtypes, disease-specific prognosis and therapeutic
1121 response. Cell Cycle. 2010;9(20):4153-63.

1122 57. Liu D, Augello MA, Grbesa I, Prandi D, Liu Y, Shoag JE, et al. Tumor subtype defines distinct
1123 pathways of molecular and clinical progression in primary prostate cancer. J Clin Invest. 2021;131(10).

1124 58. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and
1125 RNA-seq data. BMC Bioinformatics. 2013;14:7.

1126 59. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative analysis of
1127 complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal. 2013;6(269):pl1.

1128 60. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq
1129 data with DESeq2. Genome Biol. 2014;15(12):550.

1130 61. Jamroze A, Liu X, Hou S, Li WJ, Yu H, Tracz A, et al. Single-cell imaging analysis, therapeutic
1131 modeling, and a Phase Ib trial establish BCL-2 as a target across heterogeneous and plastic CRPC
1132 subtypes. Signal Tranduct Targeted Ther. 2026. *In press.*

1133 62. Hieronymus H, Lamb J, Ross KN, Peng XP, Clement C, Rodina A, et al. Gene expression
1134 signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators.
1135 Cancer Cell. 2006;10(4):321-30.

1136 63. Spratt DE, Alshalalfa M, Fishbane N, Weiner AB, Mehra R, Mahal BA, et al. Transcriptomic
1137 Heterogeneity of Androgen Receptor Activity Defines a de novo low AR-Active Subclass in Treatment
1138 Naive Primary Prostate Cancer. Clin Cancer Res. 2019;25(22):6721-30.

1139 64. Moad M, Hannezo E, Buczacki SJ, Wilson L, El-Sherif A, Sims D, et al. Multipotent Basal Stem
1140 Cells, Maintained in Localized Proximal Niches, Support Directed Long-Ranging Epithelial Flows in
1141 Human Prostates. Cell Rep. 2017;20(7):1609-22.

1142 65. Grossmann S, Hooks Y, Wilson L, Moore L, O'Neill L, Martincorena I, et al. Development,
1143 maturation, and maintenance of human prostate inferred from somatic mutations. Cell Stem Cell.
1144 2021;28(7):1262-74 e5.

1145 66. Aran D, Camarda R, Odegaard J, Paik H, Oskotsky B, Krings G, et al. Comprehensive analysis of
1146 normal adjacent to tumor transcriptomes. Nat Commun. 2017;8(1):1077.

1147 67. Ghandi M, Huang FW, Jane-Valbuena J, Kryukov GV, Lo CC, McDonald ER, 3rd, et al. Next-
1148 generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019;569(7757):503-8.

1149 68. Zhao SG, Chang SL, Erho N, Yu M, Lehrer J, Alshalalfa M, et al. Associations of Luminal and
1150 Basal Subtyping of Prostate Cancer With Prognosis and Response to Androgen Deprivation Therapy.
1151 JAMA Oncol. 2017;3(12):1663-72.

1152 69. Zhao SG, Chen WS, Das R, Chang SL, Tomlins SA, Chou J, et al. Clinical and Genomic
1153 Implications of Luminal and Basal Subtypes Across Carcinomas. Clin Cancer Res. 2019;25(8):2450-7.

1154 70. Coleman IM, DeSarkar N, Morrissey C, Xin L, Roudier MP, Sayar E, et al. Therapeutic
1155 Implications for Intrinsic Phenotype Classification of Metastatic Castration-Resistant Prostate Cancer.
1156 Clin Cancer Res. 2022;28(14):3127-40.

1157 71. Jamroze A, Liu X, Tang DG. Treatment-induced stemness and lineage plasticity in driving
1158 prostate cancer therapy resistance. Cancer Heterogeneity and Plasticity. 2024;1(1).

1159 72. You S, Knudsen BS, Erho N, Alshalalfa M, Takhar M, Al-Deen Ashab H, et al. Integrated
1160 Classification of Prostate Cancer Reveals a Novel Luminal Subtype with Poor Outcome. Cancer Res.
1161 2016;76(17):4948-58.

1162 73. Yoon J, Kim M, Posadas EM, Freedland SJ, Liu Y, Davicioni E, et al. A comparative study of PCS
1163 and PAM50 prostate cancer classification schemes. Prostate Cancer Prostatic Dis. 2021;24(3):733-42.

1164 74. Tang F, Xu D, Wang S, Wong CK, Martinez-Fundichely A, Lee CJ, et al. Chromatin profiles
1165 classify castration-resistant prostate cancers suggesting therapeutic targets. Science.
1166 2022;376(6596):eabe1505.

1167 75. Armenia J, Wankowicz SAM, Liu D, Gao J, Kundra R, Reznik E, et al. The long tail of oncogenic
1168 drivers in prostate cancer. Nat Genet. 2018;50(5):645-51.

1169 76. Haffner MC, Zwart W, Roudier MP, True LD, Nelson WG, Epstein JI, et al. Genomic and
1170 phenotypic heterogeneity in prostate cancer. Nat Rev Urol. 2021;18(2):79-92.

1171 77. Qiu X, Boufaied N, Hallal T, Feit A, de Polo A, Luoma AM, et al. MYC drives aggressive prostate
1172 cancer by disrupting transcriptional pause release at androgen receptor targets. Nat Commun.
1173 2022;13(1):2559.

1174 78. Guo H, Wu Y, Nouri M, Spisak S, Russo JW, Sowalsky AG, et al. Androgen receptor and MYC
1175 equilibration centralizes on developmental super-enhancer. Nat Commun. 2021;12(1):7308.

1176 79. Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, et al. Recurrent fusion
1177 of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science. 2005;310(5748):644-8.

1178 80. Stuchbery R, Macintyre G, Cmero M, Harewood LM, Peters JS, Costello AJ, et al. Reduction in
1179 expression of the benign AR transcriptome is a hallmark of localised prostate cancer progression.
1180 Oncotarget. 2016;7(21):31384-92.

1181 81. Schroeder A, Herrmann A, Cherryholmes G, Kowolik C, Buettner R, Pal S, et al. Loss of
1182 androgen receptor expression promotes a stem-like cell phenotype in prostate cancer through STAT3
1183 signaling. Cancer Res. 2014;74(4):1227-37.

1184 82. Giridhar PV, Williams K, VonHandorf AP, Deford PL, Kasper S. Constant degradation of the
1185 androgen receptor by MDM2 conserves prostate cancer stem cell integrity. Cancer research.
1186 2019;79(6):1124-37.

1187 83. Li L, Clevers H. Coexistence of quiescent and active adult stem cells in mammals. Science.
1188 2010;327(5965):542-5.

1189 84. Taylor MA, Kandyba E, Halliwill K, Delrosario R, Khoroshkin M, Goodarzi H, et al. Stem-cell
1190 states converge in multistage cutaneous squamous cell carcinoma development. Science.
1191 2024;384(6699):eadi7453.

1192 85. Zheng K, Hai Y, Xi Y, Zhang Y, Liu Z, Chen W, et al. Integrative multi-omics analysis unveils
1193 stemness-associated molecular subtypes in prostate cancer and pan-cancer: prognostic and therapeutic
1194 significance. J Transl Med. 2023;21(1):789.

1195 86. Zhang T, Li J, Dai J, Yuan F, Yuan G, Chen H, et al. Identification of a novel stemness-related
1196 signature with appealing implications in discriminating the prognosis and therapy responses for prostate
1197 cancer. Cancer Genet. 2023;276-277:48-59.

1198 87. Zhao F, Zhang T, Wei J, Chen L, Liu Z, Jin Y, et al. Integrated single-cell transcriptomic analyses
1199 identify a novel lineage plasticity-related cancer cell type involved in prostate cancer progression.
1200 EBioMedicine. 2024;109:105398.

1201 88. Zhao F, Zeng J, Chen C, Zhao X, Zhang T, Thomas GV, et al. Deciphering single-cell
1202 heterogeneity and cellular ecosystem dynamics during prostate cancer progression. bioRxiv. 2024.

1203 89. Bian X, Wang W, Abudurexiti M, Zhang X, Ma W, Shi G, et al. Integration Analysis of Single-Cell
1204 Multi-Omics Reveals Prostate Cancer Heterogeneity. Adv Sci (Weinh). 2024;11(18):e2305724.

1205 90. Kirk JS, Wang J, Long M, Rosario S, Tracz A, Ji Y, et al. Integrated single-cell analysis defines
1206 the epigenetic basis of castration-resistant prostate luminal cells. Cell Stem Cell. 2024;31(8):1203-21 e7.

1207 91. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, et al. Single-cell
1208 transcriptional diversity is a hallmark of developmental potential. Science. 2020;367(6476):405-11.

1209 92. Chauhan PS, Alahi I, Sinha S, Ledet EM, Mueller R, Linford J, et al. Genomic and Epigenomic
1210 Analysis of Plasma Cell-Free DNA Identifies Stemness Features Associated with Worse Survival in
1211 Lethal Prostate Cancer. Clin Cancer Res. 2025;31(1):151-63.

1212

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 gratefully acknowledge Dr. X. Qiu for kindly providing supporting data for AR cistrome, H3K27ac profiles and transcriptomic data in prostate cancer tissue samples (Pomerantz *Nat Genet* 2020); the cloud computing resources provided by the Center for Computational Research (CCR) at the University at Buffalo (http://hdl.handle.net/10477/79221); 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 14263, USA
Xiaozhuo Liu, Yibing Ji, Anmbreen Jamroze, David W. Goodrich, Kiera Rycaj, Jason S. Kirk & Dean G. Tang

Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, New York 14263, USA
Eduardo Cortes, Han Yu, Qiang Hu, Song Liu & Jianmin Wang

Department of Biostatistics, University at Buffalo, Buffalo, New York 14263, USA
Eduardo Cortes

Veracyte Inc, San Diego, California 92121, 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 48109, 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
Cheng Zou & Dingxiao Zhang

Department of Epigenetics and Molecular Carcinogenesis, the University of Texas M.D Anderson Cancer Center, 1515 Holcombe Blvd., Houston, TX 77030, USA
Kevin Lin & Yue Lu

Department of Medicine, Roswell Park Comprehensive Cancer Center, Buffalo, New York 14263, USA
Gurkamal Chatta & Igor Puzanov

Department of Cancer Genetics, Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA
Kent L. Nastiuk

Experimental Therapeutics Graduate Program, University at Buffalo and Roswell Park Comprehensive Cancer Center, Buffalo, NY 14263, USA
Dean G. Tang

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. **A. Jamroze**: Data curation, investigation, writing–review and editing. **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. **C. Zou**: Data curation, investigation, writing–review and editing. **D. Zhang**: Data curation, investigation, writing–review and editing. **K. Lin**: Data curation and formal analysis. **Y. Lu**: Data curation and formal analysis. **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. **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, editing and finalization, 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.

1290 **Figure Legends**

1291 **Figure 1.** *Pri-PCa shows concordant increases of c_AR-A and Stemness and both are decreased by nADT.*

1292 **A-C,** PCa has increased c_AR-A as well as the Stemness compared to matched benign tissues. Pairwise
1293 comparisons of *AR* mRNA levels, canonical c_AR-A, and Stemness scores showing significant increases
1294 in primary tumors (T) as compared to the matched adjacent benign prostate tissues (N) from the same
1295 individual in TCGA-PRAD cohort (n=51; **A**), and Wyatt 2014 cohort (n=12; **B**). In the Long 2020 cohort
1296 (n=10; **C**), increases in c_AR-A and Stemness reached statistical significance whereas the increase in
1297 *AR* mRNA levels approached statistical significance ($P = 0.06$).

1298 **D-F,** nADT decreases both c_AR-A and Stemness. Pairwise comparison of *AR* mRNA levels, c_AR-A, and
1299 Stemness showing significant decrease of both c_AR-A and Stemness after nADT (Post-nADT) as
1300 compared to the matched samples before nADT (Pre-nADT) from the same individual (**D**, Rajan 2014
1301 cohort; **E**, Sharma 2018 cohort; **F**, Long 2020 cohort).

1302 In all plots, sample sizes are indicated (see also Table S1 for cohort and RNA-seq data access information).
1303 Within the plots, results are shown as the mean ± SD and pairs are linked with a grey line. Significance
1304 was calculated by two-tailed paired Student's *t*-test (ns, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P <$
1305 0.001; ****, $P < 0.0001$).

1306

1307 **Figure 2.** *Continual increases in Stemness with declining c_AR-A during spontaneous PCa progression*.

1308 **A,** The pie chart showing sample sizes in different Gleason grades in the TCGA untreated Pri-PCa cohort.

1309 **B-D,** Progression of treatment-naïve Pri-PCa is accompanied by decreasing c_AR-A but increasing
1310 Stemness. Data is presented as mean ± SD. Shown on top are the Jonckheere-Terpstra's trend (J-T)
1311 test results (wedges).

1312 **E-F,** Progression in Pri-PCa is accompanied by a gradual loss of the positive correlation between canonical
1313 c_AR-A and Stemness. Scatter plots illustrating the correlation between c_AR-A and Stemness in (**E**)
1314 combined Pri-PCa cases in TCGA with an overall linear regression model and (**F**) individual GS grades
1315 with corresponding linear regressions showing decreasing slopes from low to high GS grades by J-T test.
1316 The slopes of best-fit regression line are also shown.

1317 **G,** Merged scatter plot showing the correlation between c_AR-A and Stemness across the 4 GS PCa with
1318 each GS represented by a distinct color. The correlation becomes progressively weaker from GS6 to
1319 GS9/10, transitioning from positive to nearly zero. The correlation coefficients (r) and p-values for each
1320 GS are as follows: GS6 ($r = 0.56$, $P < 0.0001$), GS7 ($r = 0.37$, $P < 0.0001$), GS8 ($r = 0.48$, $P = 0.022$),
1321 and GS9/10 ($r = -0.02$, $P = 0.89$). Linear regression lines are overlaid for each GS, demonstrating the
1322 decline in the correlation co-efficiency as the GS increases.

1323 All statistical analyses were performed using R and statistical significance was assessed using one-way
1324 analysis of variance followed by Tukey's multiple-comparison test (B-D). J-T trend test in B-D and F was

1325 used to calculate the statistical significance of the trend across different groups. *, $P < 0.05$; **, $P < 0.01$;
1326 ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant.
1327
1328 **Figure 3.** *Therapies and metastatic progression drive further divergence between Stemness and canonical*
1329 *c_AR-A with mCRPC showing the highest Stemness but the lowest c_AR-A*.

1330 **A-D,** Hormonal treatment further drives down c_AR-A but significantly upregulates Stemness. **A-C**,
1331 mCRPC displays the lowest pro-differentiation c_AR-A but highest Stemness. Shown are *AR* mRNA
1332 levels (**A**), c_AR-A (**B**) and Stemness (**C**) in tumor-adjacent benign tissues (N), treatment-naïve Pri-PCa
1333 and advanced Pri-PCa with post-surgery adjuvant ADT (Tx) in TCGA, and mCRPC (SU2C 2015).
1334 Sample sizes are indicated in parentheses. Shown on top are the Jonckheere-Terpstra's (J-T) trend test
1335 results (wedges). **D,** Stemness inversely correlates with c_AR-A in mCRPC**.** Shown are scatter plots with
1336 best-fit linear regression lines illustrating the correlation between c_AR-A and Stemness with Pearson
1337 correlation coefficient (r), the slopes of lines and J-T test results of slopes indicated below.

1338 **E-F,** PCa progression is accompanied by loss of c_AR-A (**E**) but gain of Stemness (**F**) across PCa states.
1339 Shown are samples representing the spectrum of PCa development and progression, from normal (organ
1340 donors in GTEx) to (tumor-adjacent) benign tissues, to (treatment-naïve) Pri-PCa, aggressive Pri-PCa
1341 treated with adjuvant ADT (Pri-PCa/ADT), mCRPC, and, finally, to (clonally-derived) PCa models
1342 including PDX, xenografts (XG) and cells lines. From Left to right: Normal (n=116): Normal prostates
1343 from GTEx; Benign (n=52): tumor-adjacent benign tissue from TCGA-PRAD; Pri-PCa (n=422):
1344 treatment-naïve localized PCa from TCGA-PRAD; Pri-PCa/ADT (n=70): advanced Pri-PCa treated with
1345 adjuvant hormone therapy from TCGA-PRAD; mCRPC (n=59): mCRPC from GSE126078; mCRPC
1346 (n=135): mCRPC from SU2C 2015 cohort; mCRPC-Adeno (n=34): adenocarcinoma mCRPC from
1347 Beltran 2016 cohort; mCRPC-NE (n=15): neuroendocrine mCRPC from Beltran 2016 cohort; mCRPC
1348 (n=42): mCRPC from Westbrook 2022; mCRPC (n=25): mCRPC from Alumkal 2020; PDX-LuCaP (n=39):
1349 PCa PDXs from GSE126078; PCa cell lines (n=8): PCa cell lines from CCLE; XG-LAPC9 (n=10) and
1350 XG-LNCaP (n=12): PCa xenografts from GSE88752. Statistical significances between groups were
1351 shown in inverted pyramid grid plot as indicated by different colors. Student's *t*-test was used to compare
1352 each group and adjusted p-value for multiple comparisons was done by Tukey's range test. See Table
1353 S4 for the summary of statistical comparisons. J-T trend test was used to assess the statistical
1354 significance of the trend across groups. See also Table S1 for cohort and RNA-seq data access
1355 information.

1356 **G,** Scatter plots in the Bolis 2021 integrated cohort (normal n=174, Pri-PCa n=714, mCRPC n=335)
1357 showing the changing correlation between c_AR-A and Stemness across disease stages: positive in
1358 normal to weaker correlations in Pri-PCa to negative correlations in mCRPC. Regression lines are color-
1359 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 c_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 c_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 c_AR-A during PCa progression (Pearson's r=0.18, *P* = 0.043); **M,** a loss of positive correlation between the Stemness and c_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 c_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 newly developed 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

1393 (top 33%) and Stemness-low (bottom 33%) groups for downstream analyses. Genome-wide functional
1394 analyses were performed using transcriptomic, genomic, and clinical data.

1395 **B-C,** Volcano plots showing differentially expressed genes (DEGs) between Stemness-high and
1396 Stemness-low groups in Pri-PCa (**B**) and mCRPC (**C**) cohorts. Red and blue indicate upregulated and
1397 downregulated DEGs, respectively. Genes with FDR > 0.05 are shown in gray. Sample sizes (n) of each
1398 group are indicated. See Table S3 for complete DEG lists.

1399 **D,** Venn diagram showing the 12-gene "PCa-Stem signature" derived from commonly upregulated DEGs
1400 (FC ≥ 2, FDR < 0.05) in Pri-PCa and mCRPC.

1401 **E-F,** GSEA showing low Stemness is associated with immune signaling and inflammatory response (**E**)
1402 while high Stemness in both cohorts is enriched for embryonic stem cell traits, aggressiveness, DNA
1403 repair, MYC activation, and mTORC1 signaling (**F**). All enrichments are significant (P < 0.05, FDR <
1404 0.05).

1405 **G-H,** PAM50 subtyping showing a higher frequency of aggressive LumB subtype in Stemness-high versus
1406 Stemness-low samples in Pri-PCa (**G**) and mCRPC (**H**). (****, P < 0.0001, χ^2 test).

1407 **I-L,** GSEA showing enrichment of the aggressive PCS1 signature in Stemness-high groups and the less
1408 aggressive PCS3 signature in Stemness-low groups in Pri-PCa (**I**, **K**) and mCRPC (**J**, **L**). The Prostate
1409 Cancer Subtype (PCS) classification system (45,46) stratifies tumors into three molecular subtypes
1410 based on transcriptomic features: PCS1 (most aggressive), PCS2 (intermediate), and PCS3 (least
1411 aggressive). See Table S2 for PCS gene signature.

1412 **M-R,** GSEA showing AR signature enrichment in Pri-PCa Stemness-high (**M**), NEPC signature enrichment
1413 in mCRPC Stemness-high (**N**), and shared enrichment of cell cycle (**O**, **P**) and lineage plasticity (**Q**, **R**)
1414 programs in both cohorts. See Table S5 for gene sets.

1415 **S-V,** Kaplan-Meier analyses showing that the 12-gene PCa-Stem signature predicts worse progression-
1416 free survival (PFS) in TCGA Pri-PCa (**S**) and worse overall survival (OS) in SU2C mCRPC (**U**).
1417 Multivariable Cox models adjusting for clinicopathologic variables confirm the high PCa-Stem signature
1418 score as an independent prognostic for PFS (**T**) and OS (**V**).

1419

1420 **Figure 6.** *Stemness-high PCa is associated with higher genomic instability.*

1421 **A-C,** Box plots showing that Stemness-high subgroup in both Pri-PCa (n = 141) and mCRPC (n = 89)
1422 displays a higher fraction of altered genome (**A**) and higher TMB (**B**), but no age difference (**C**) compared
1423 to Stemness-low cases. Wilcoxon rank-sum test was used to determine significance (****, P < 0.0001;
1424 ns, not significant). Box plots represent the median, interquartile range (IQR), and whiskers extending to
1425 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 signaling 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 ± 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 8. *Castration-reprogrammed AR activity (cr_AR-A) and RB1-loss signature contribute to increased Stemness in mCRPC.*

A–D, The cr_AR-A and genomic *RB1*-loss signature increase progressively from normal and Pri-PCa to mCRPC. Shown are cr_AR-A (**A**) and *RB1*-loss (**B**), *PTEN*-loss (**C**) and PCa-Stem **(D)** signature scores across the disease spectrum using the Bolis integrated PCa cohort: normal prostates (GTEx, n=116), tumor-adjacent benign tissues (TCGA-PRAD, n=52), treatment-naïve Pri-PCa (TCGA-PRAD, n=422), aggressive Pri-PCa treated with adjuvant ADT (Pri-PCa/ADT) (n=70), and two mCRPC cohorts including GSE126078 mCRPC (n=59) and SU2C 2015 (n=135). Sample sizes are indicated in parentheses. Shown at the bottom are Jonckheere–Terpstra (J-T) trend-test results (wedges).

E–H, Scatter plots illustrating correlations between cr_AR-A (**E**), *RB1*-loss (**F**), *PTEN*-loss (**G**), PCa-Stem (**H**) scores with Stemness across Bolis integrated cohort (total n=1,223; including normal (n=174), Pri-PCa (n=714), and mCRPC (n=335)). Pearson's correlation coefficients (*r*) and *P* values are shown. Details of symbol colors and shapes are shown in (**L**).

I–L, Inter-signature relationships among cr_AR-A, *RB1*-loss, *PTEN*-loss, and PCa-Stem signatures in the Bolis integrated cohort (total n=1,223, normal n=174, Pri-PCa n=714, mCRPC n=335). Correlation plots

demonstrate strong positive correlations between cr_AR-A and PCa-Stem (**I**), and between *RB1*-loss and PCa-Stem (**J**), and a moderate but significant correlation between *PTEN*-loss and PCa-Stem (**K**). Details of symbol colors and shapes are shown in (**L**).

M-P, Venn diagrams showing overlapping genes among cr_AR-A (**M, P**), *RB1*-loss (**N, P**), *PTEN*-loss (**O**), and PCa-Stem signatures, highlighting shared components (e.g., *AURKB, BIRC5, CENPA, PLK1, ANLN)* involved mitotic regulation and the self-renewing proliferation typical of stem-like tumor cells.

All dot plots data are shown as mean ± SD. Statistical significance among groups was determined using one-way ANOVA followed by Tukey's multiple-comparison test, and J-T trend tests were used to assess significance across disease states. Pearson correlation coefficients (r) were calculated for all pairwise associations. *ns*, not significant; *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$. See Table S4 for full statistical summaries and data sources.

Fig. 1

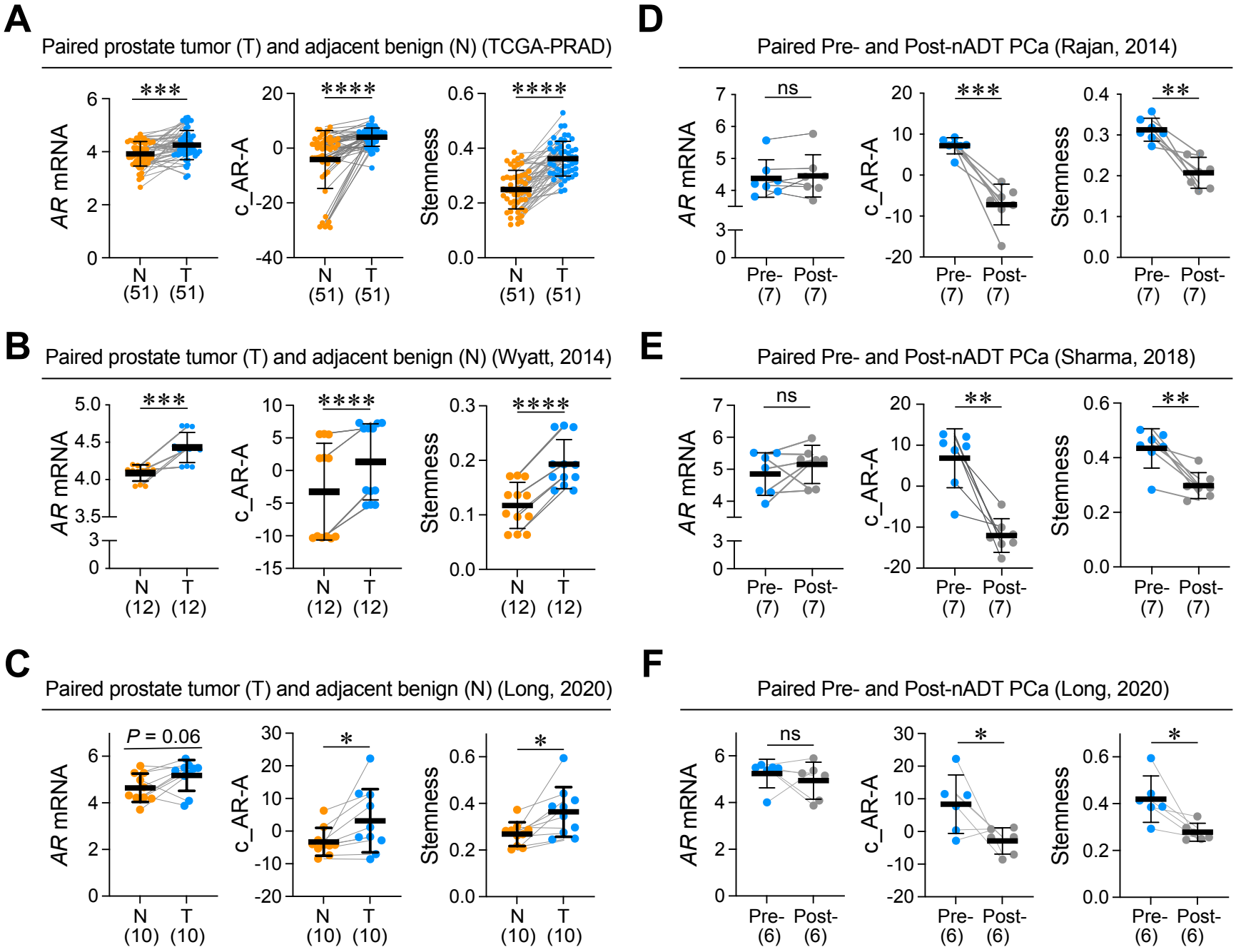

A
Paired prostate tumor (T) and adjacent benign (N) (TCGA-PRAD)
AR mRNA
c_AR-A
Stemness
***
****
****
N (51)
T (51)
B
Paired prostate tumor (T) and adjacent benign (N) (Wyatt, 2014)
***
****
****
N (12)
T (12)
C
Paired prostate tumor (T) and adjacent benign (N) (Long, 2020)
P = 0.06
*
*
N (10)
T (10)
D
Paired Pre- and Post-nADT PCa (Rajan, 2014)
ns
***
**
Pre- (7)
Post- (7)
E
Paired Pre- and Post-nADT PCa (Sharma, 2018)
ns
**
**
Pre- (7)
Post- (7)
F
Paired Pre- and Post-nADT PCa (Long, 2020)
ns
*
*
Pre- (6)
Post- (6)

Fig. 2

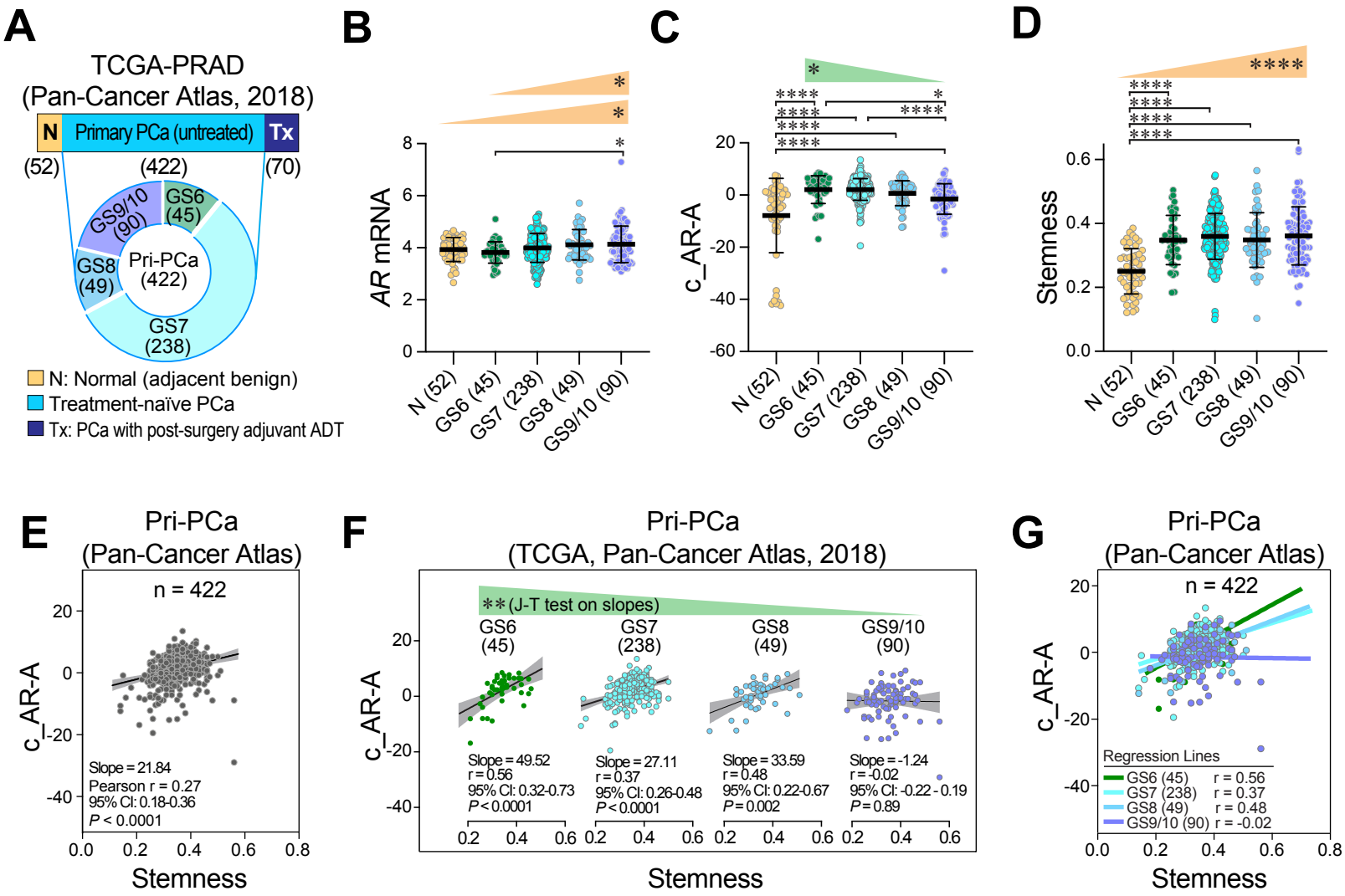

A
TCGA-PRAD
(Pan-Cancer Atlas, 2018)
N
Primary PCa (untreated)
Tx
(52)
(422)
(70)
GS6 (45)
GS7 (238)
GS8 (49)
GS9/10 (90)
Pri-PCa (422)
N: Normal (adjacent benign)
Treatment-naïve PCa
Tx: PCa with post-surgery adjuvant ADT
B
AR mRNA
N (52)
GS6 (45)
GS7 (238)
GS8 (49)
GS9/10 (90)
C
c_AR-A
D
Stemness
E
Pri-PCa
(Pan-Cancer Atlas)
n = 422
c_AR-A
Slope = 21.84
Pearson r = 0.27
95% CI: 0.18-0.36
P < 0.0001
Stemness
F
Pri-PCa
(TCGA, Pan-Cancer Atlas, 2018)
**(J-T test on slopes)
GS6 (45)
GS7 (238)
GS8 (49)
GS9/10 (90)
Slope = 49.52
r = 0.56
95% CI: 0.32-0.73
P < 0.0001
Slope = 27.11
r = 0.37
95% CI: 0.26-0.48
P < 0.0001
Slope = 33.59
r = 0.48
95% CI: 0.22-0.67
P = 0.002
Slope = -1.24
r = -0.02
95% CI: -0.22 - 0.19
P = 0.89
Stemness
G
Pri-PCa
(Pan-Cancer Atlas)
n = 422
c_AR-A
Regression Lines
GS6 (45) r = 0.56
GS7 (238) r = 0.37
GS8 (49) r = 0.48
GS9/10 (90) r = -0.02
Stemness

Fig. 3

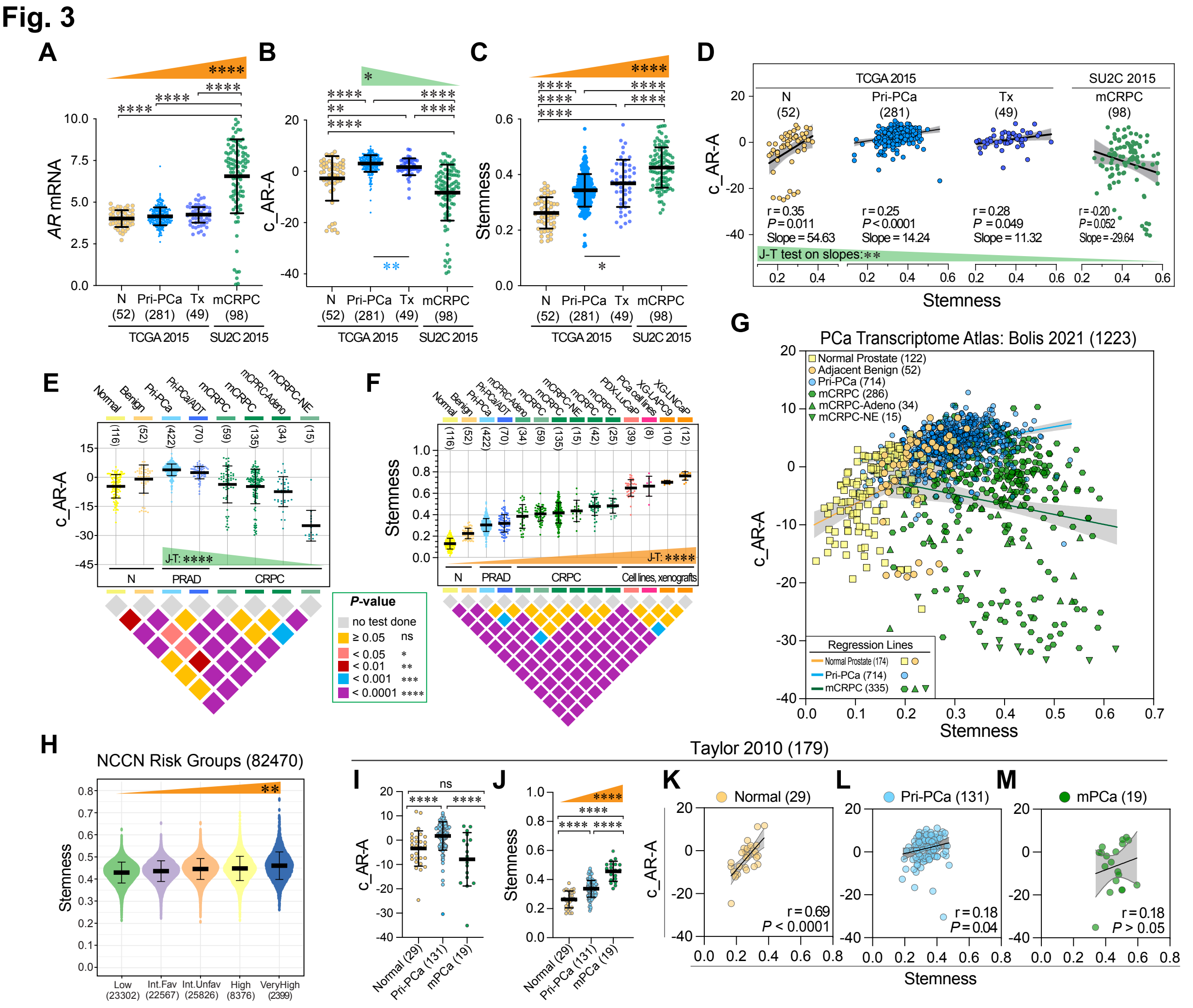

A
AR mRNA
N (52)
Pri-PCa (281)
Tx (49)
mCRPC (98)
TCGA 2015
SU2C 2015
B
c_AR-A
C
Stemness
D
TCGA 2015
SU2C 2015
c_AR-A
r = 0.35
P = 0.011
Slope = 54.63
r = 0.25
P < 0.0001
Slope = 14.24
r = 0.28
P = 0.049
Slope = 11.32
r = -0.20
P = 0.052
Slope = -29.64
J-T test on slopes: **
Stemness
E
Normal
Benign
Pri-PCa
Pri-PCa/ADT
mCRPC
mCRPC
mCRPC-Adeno
mCRPC-NE
c_AR-A
J-T: ****
N
PRAD
CRPC
F
Stemness
J-T: ****
Cell lines, xenografts
P-value
no test done
≥ 0.05 ns
< 0.05 *
< 0.01 **
< 0.001 ***
< 0.0001 ****
G
PCa Transcriptome Atlas: Bolis 2021 (1223)
Normal Prostate (122)
Adjacent Benign (52)
Pri-PCa (714)
mCRPC (286)
mCRPC-Adeno (34)
mCRPC-NE (15)
Regression Lines
Normal Prostate (174)
Pri-PCa (714)
mCRPC (335)
c_AR-A
Stemness
H
NCCN Risk Groups (82470)
Stemness
Low (23302)
Int.Fav (22567)
Int.Unfav (25826)
High (8376)
VeryHigh (2399)
Taylor 2010 (179)
I
c_AR-A
Normal (29)
Pri-PCa (131)
mPCa (19)
J
Stemness
K
Normal (29)
r = 0.69
P < 0.0001
L
Pri-PCa (131)
r = 0.18
P = 0.04
M
mPCa (19)
r = 0.18
P > 0.05
Stemness

Fig. 4

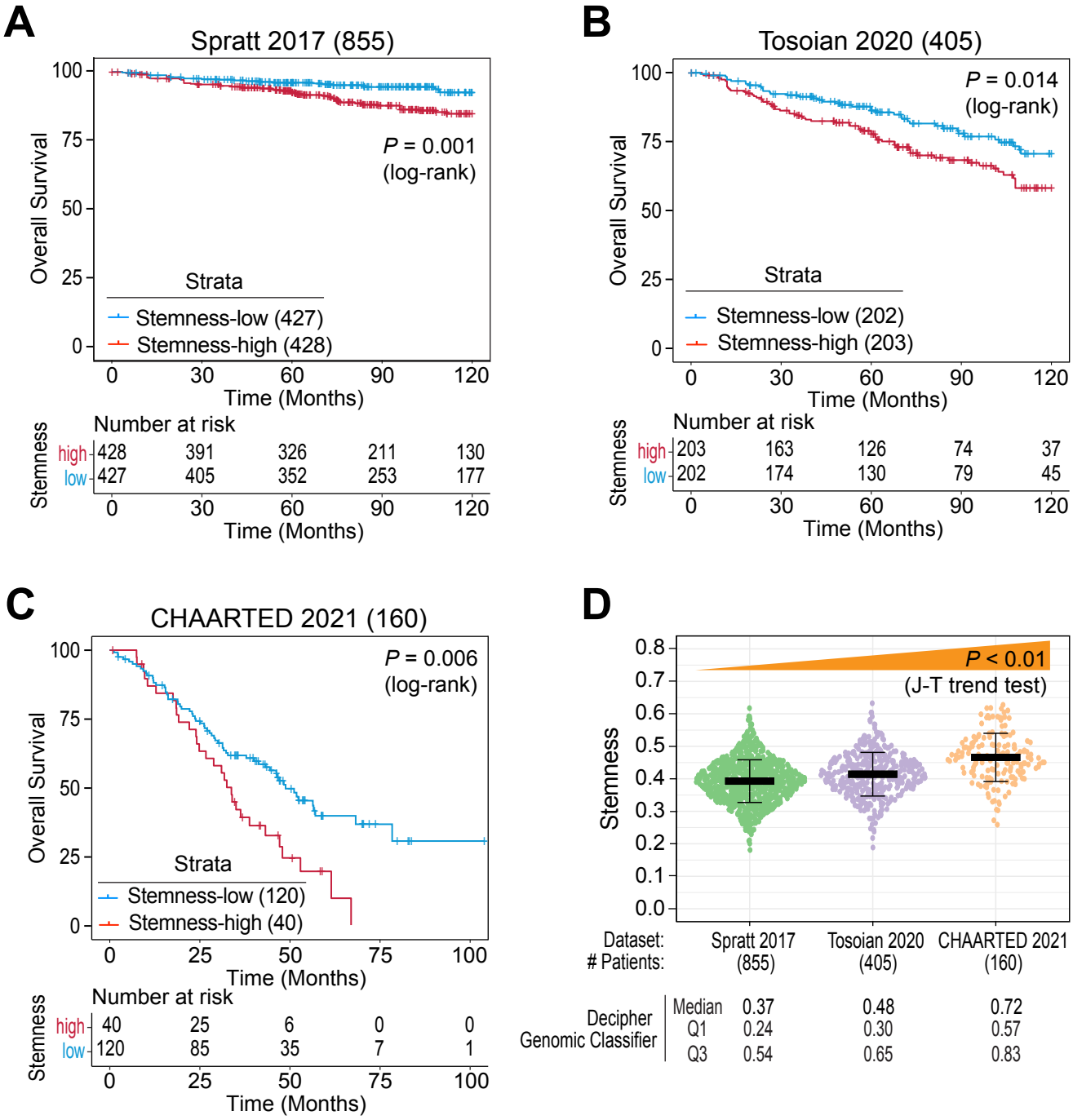

A
Spratt 2017 (855)
Overall Survival
P = 0.001
(log-rank)
Strata
Stemness-low (427)
Stemness-high (428)
Time (Months)
Number at risk
Stemness
high 428 391 326 211 130
low 427 405 352 253 177
B
Tosoian 2020 (405)
P = 0.014
(log-rank)
Stemness-low (202)
Stemness-high (203)
high 203 163 126 74 37
low 202 174 130 79 45
C
CHAARTED 2021 (160)
P = 0.006
(log-rank)
Stemness-low (120)
Stemness-high (40)
high 40 25 6 0 0
low 120 85 35 7 1
D
Stemness
P < 0.01
(J-T trend test)
Dataset:
Patients:
Spratt 2017 (855)
Tosoian 2020 (405)
CHAARTED 2021 (160)
Decipher Genomic Classifier
Median 0.37 0.48 0.72
Q1 0.24 0.30 0.57
Q3 0.54 0.65 0.83

Fig. 5

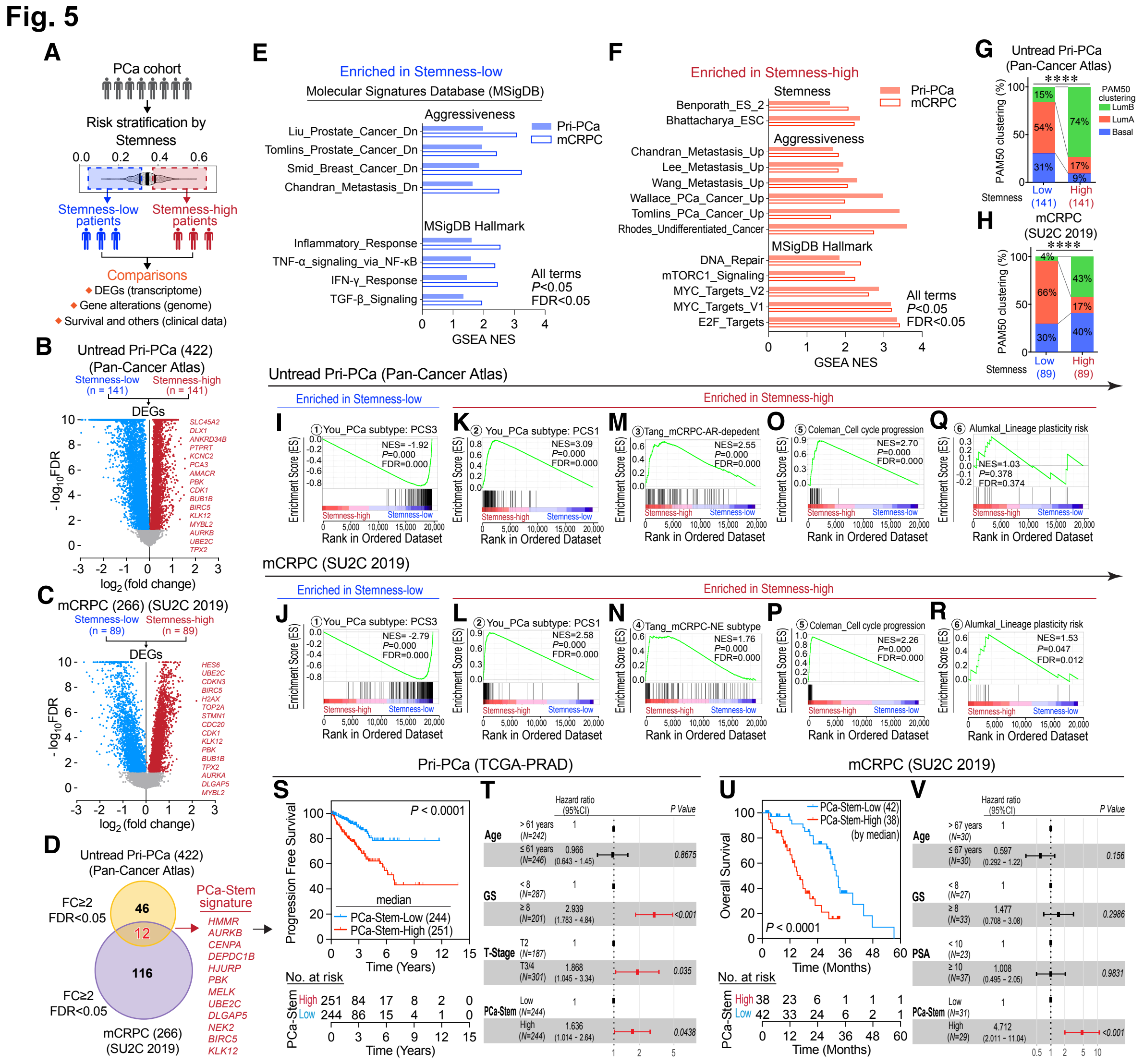

A
PCa cohort
Risk stratification by Stemness
0.0 0.2 0.4 0.6
Stemness-low patients
Stemness-high patients
Comparisons
DEGs (transcriptome)
Gene alterations (genome)
Survival and others (clinical data)
B
Untread Pri-PCa (422) (Pan-Cancer Atlas)
Stemness-low (n = 141)
Stemness-high (n = 141)
DEGs
SLC45A2
DLX1
ANKRD34B
PTPRT
KCNC2
PCA3
AMACR
PBK
CDK1
BUB1B
BIRC5
KLK12
MYBL2
AURKB
UBE2C
TPX2
- log10FDR
log2 (fold change)
C
mCRPC (266) (SU2C 2019)
Stemness-low (n = 89)
Stemness-high (n = 89)
DEGs
HES6
UBE2C
CDKN3
BIRC5
H2AX
TOP2A
STMN1
CDC20
CDK1
KLK12
PBK
BUB1B
TPX2
AURKA
DLGAP5
MYBL2
- log10FDR
log2 (fold change)
D
Untread Pri-PCa (422) (Pan-Cancer Atlas)
FC≥2 FDR<0.05
46
12
116
FC≥2 FDR<0.05
mCRPC (266) (SU2C 2019)
PCa-Stem signature
HMMR
AURKB
CENPA
DEPDC1B
HJURP
PBK
MELK
UBE2C
DLGAP5
NEK2
BIRC5
KLK12
E
Enriched in Stemness-low
Molecular Signatures Database (MSigDB)
Aggressiveness
Liu_Prostate_Cancer_Dn
Tomlins_Prostate_Cancer_Dn
Smid_Breast_Cancer_Dn
Chandran_Metastasis_Dn
MSigDB Hallmark
Inflammatory_Response
TNF-α_signaling_via_NF-κB
IFN-γ_Response
TGF-β_Signaling
Pri-PCa
mCRPC
All terms
P<0.05
FDR<0.05
GSEA NES
F
Enriched in Stemness-high
Stemness
Benporath_ES_2
Bhattacharya_ESC
Aggressiveness
Chandran_Metastasis_Up
Lee_Metastasis_Up
Wang_Metastasis_Up
Wallace_PCa_Cancer_Up
Tomlins_PCa_Cancer_Up
Rhodes_Undifferentiated_Cancer
MSigDB Hallmark
DNA_Repair
mTORC1_Signaling
MYC_Targets_V2
MYC_Targets_V1
E2F_Targets
Pri-PCa
mCRPC
All terms
P<0.05
FDR<0.05
GSEA NES
G
Untread Pri-PCa (Pan-Cancer Atlas)
****
PAM50 clustering
LumB
LumA
Basal
PAM50 clustering (%)
15%
54%
31%
74%
17%
9%
Stemness
Low (141)
High (141)
H
mCRPC (SU2C 2019)
****
4%
66%
30%
43%
17%
40%
Low (89)
High (89)
Untread Pri-PCa (Pan-Cancer Atlas)
Enriched in Stemness-low
Enriched in Stemness-high
I
① You_PCa subtype: PCS3
NES= -1.92
P=0.000
FDR=0.000
K
② You_PCa subtype: PCS1
NES=3.09
P=0.000
FDR=0.000
M
③ Tang_mCRPC-AR-depedent
NES=2.55
P=0.000
FDR=0.000
O
⑤ Coleman_Cell cycle progression
NES=2.70
P=0.000
FDR=0.000
Q
⑥ Alumkal_Lineage plasticity risk
NES=1.03
P=0.378
FDR=0.374
Enrichment Score (ES)
Stemness-high
Stemness-low
Rank in Ordered Dataset
mCRPC (SU2C 2019)
Enriched in Stemness-low
Enriched in Stemness-high
J
① You_PCa subtype: PCS3
NES= -2.79
P=0.000
FDR=0.000
L
② You_PCa subtype: PCS1
NES=2.58
P=0.000
FDR=0.000
N
④ Tang_mCRPC-NE subtype
NES=1.76
P=0.000
FDR=0.000
P
⑤ Coleman_Cell cycle progression
NES=2.26
P=0.000
FDR=0.000
R
⑥ Alumkal_Lineage plasticity risk
NES=1.53
P=0.047
FDR=0.012
Pri-PCa (TCGA-PRAD)
S
P < 0.0001
median
PCa-Stem-Low (244)
PCa-Stem-High (251)
Progression Free Survival
Time (Years)
No. at risk
PCa-Stem
High 251 84 17 8 2 0
Low 244 86 15 4 1 0
T
Hazard ratio (95%CI)
P Value
Age
> 61 years (N=242) 1
≤ 61 years (N=246) 0.966 (0.643 - 1.45) 0.8675
GS
< 8 (N=287) 1
≥ 8 (N=201) 2.939 (1.783 - 4.84) <0.001
T-Stage
T2 (N=187) 1
T3/4 (N=301) 1.868 (1.045 - 3.34) 0.035
PCa-Stem
Low (N=244) 1
High (N=244) 1.636 (1.014 - 2.64) 0.0438
mCRPC (SU2C 2019)
U
PCa-Stem-Low (42)
PCa-Stem-High (38)
(by median)
P < 0.0001
Overall Survival
Time (Months)
No. at risk
PCa-Stem
High 38 23 6 1 1 1
Low 42 33 24 6 2 1
V
Hazard ratio (95%CI)
P Value
Age
> 67 years (N=30) 1
≤ 67 years (N=30) 0.597 (0.292 - 1.22) 0.156
GS
< 8 (N=27) 1
≥ 8 (N=33) 1.477 (0.708 - 3.08) 0.2986
PSA
< 10 (N=23) 1
≥ 10 (N=37) 1.008 (0.495 - 2.05) 0.9831
PCa-Stem
Low (N=31) 1
High (N=29) 4.712 (2.011 - 11.04) <0.001

Fig. 6

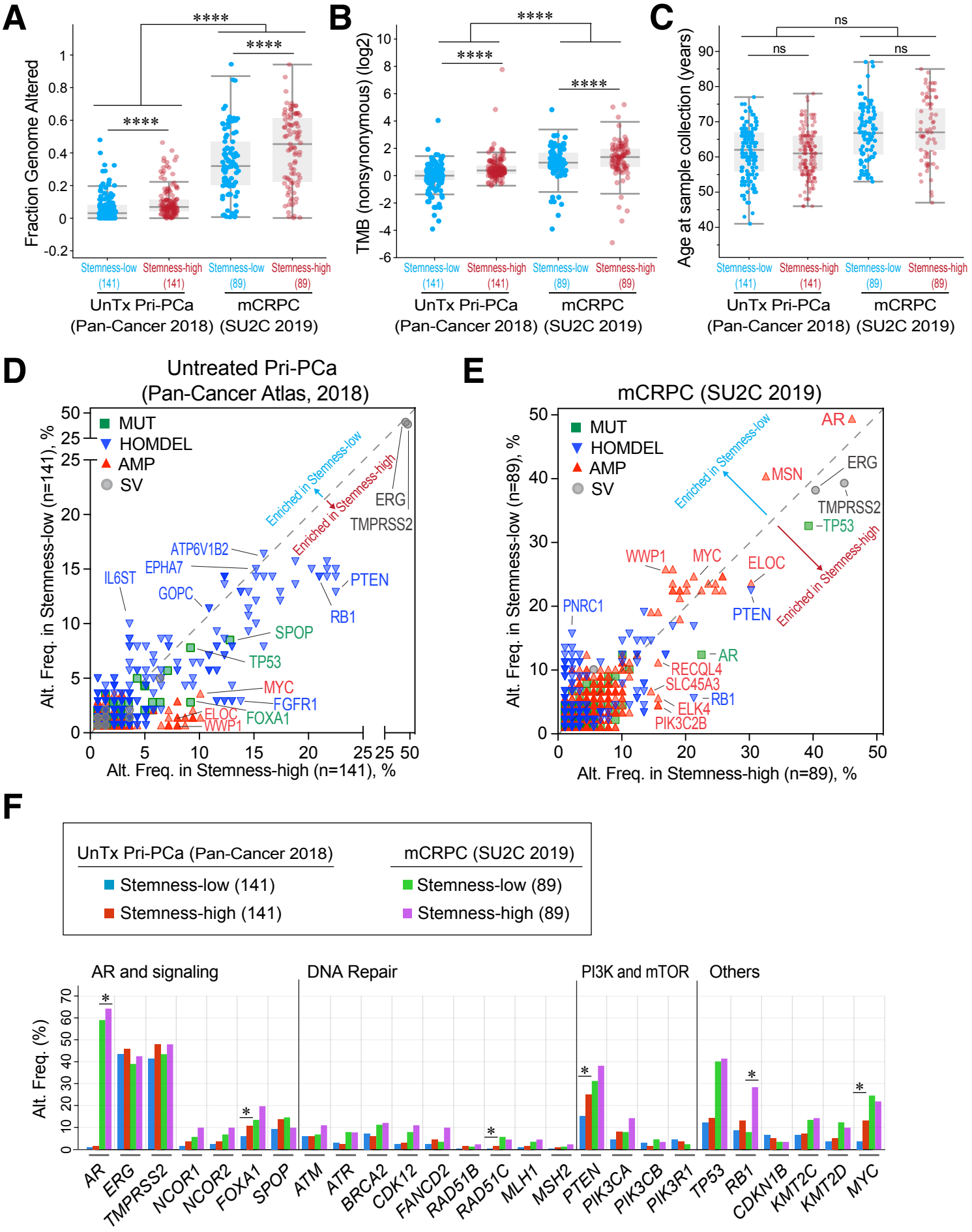

A
Fraction Genome Altered
****
Stemness-low (141)
Stemness-high (141)
Stemness-low (89)
Stemness-high (89)
UnTx Pri-PCa (Pan-Cancer 2018)
mCRPC (SU2C 2019)
B
TMB (nonsynonymous) (log2)
C
Age at sample collection (years)
ns
D
Untreated Pri-PCa (Pan-Cancer Atlas, 2018)
MUT
HOMDEL
AMP
SV
Enriched in Stemness-low
Enriched in Stemness-high
ERG
TMPRSS2
ATP6V1B2
EPHA7
IL6ST
GOPC
PTEN
RB1
SPOP
TP53
MYC
FGFR1
ELOC
WWP1
FOXA1
Alt. Freq. in Stemness-low (n=141), %
Alt. Freq. in Stemness-high (n=141), %
E
mCRPC (SU2C 2019)
AR
MSN
ERG
TMPRSS2
TP53
WWP1
MYC
ELOC
PTEN
PNRC1
RECQL4
SLC45A3
RB1
ELK4
PIK3C2B
Alt. Freq. in Stemness-low (n=89), %
Alt. Freq. in Stemness-high (n=89), %
F
UnTx Pri-PCa (Pan-Cancer 2018)
Stemness-low (141)
Stemness-high (141)
mCRPC (SU2C 2019)
Stemness-low (89)
Stemness-high (89)
AR and signaling
DNA Repair
PI3K and mTOR
Others
Alt. Freq. (%)
AR
ERG
TMPRSS2
NCOR1
NCOR2
FOXA1
SPOP
ATM
ATR
BRCA2
CDK12
FANCD2
RAD51B
RAD51C
MLH1
MSH2
PTEN
PIK3CA
PIK3CB
PIK3R1
TP53
RB1
CDKN1B
KMT2C
KMT2D
MYC

Fig. 7

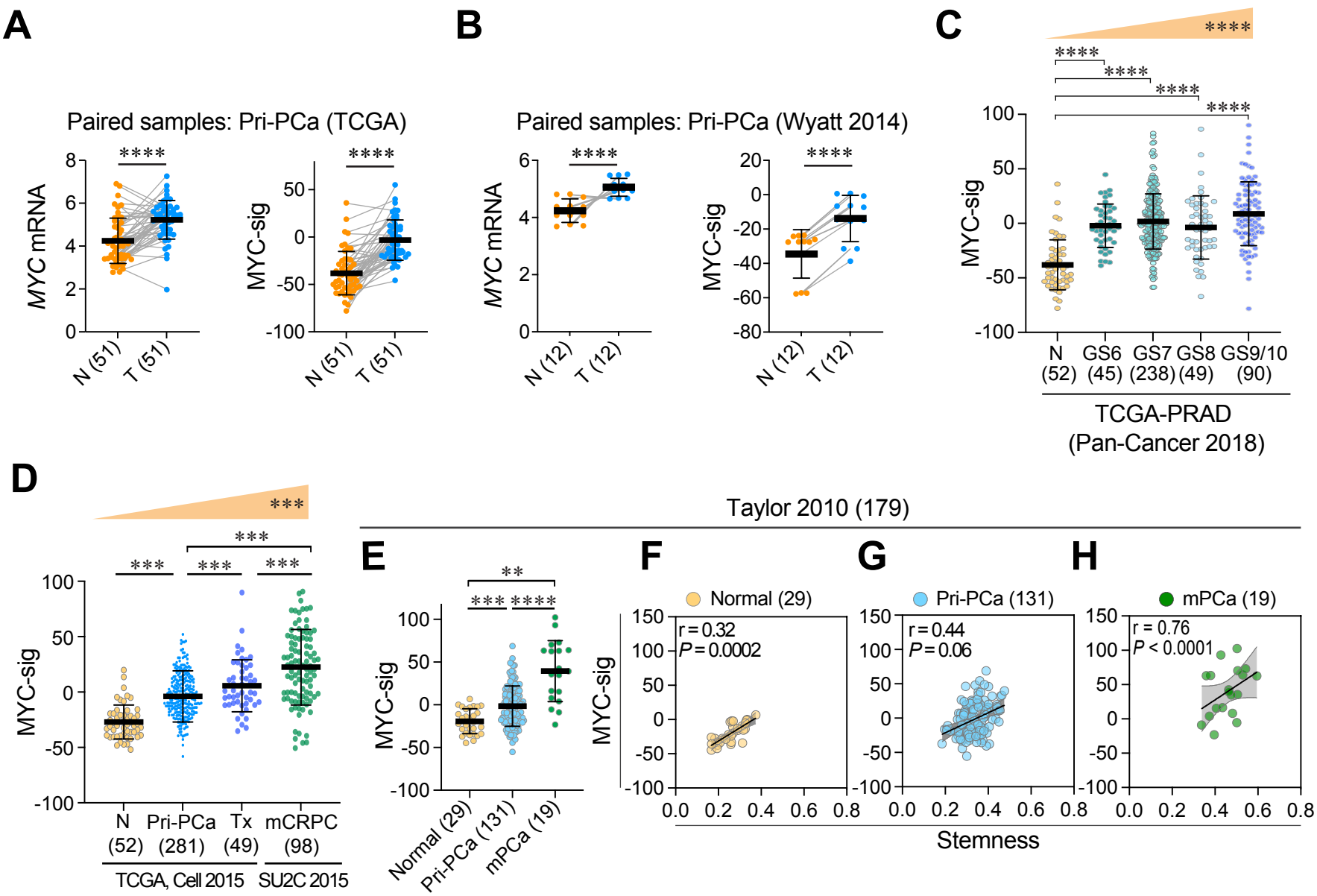
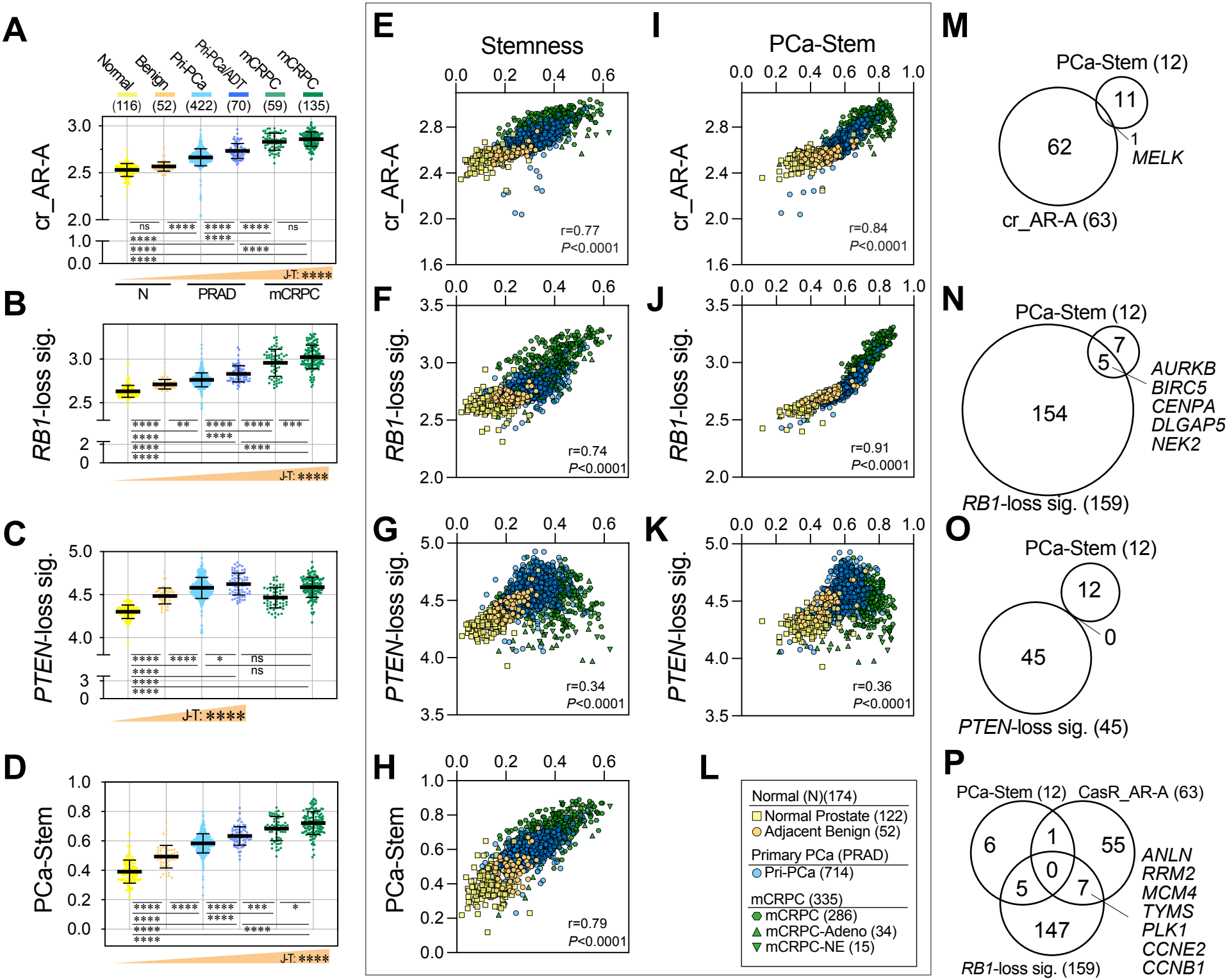

A
Paired samples: Pri-PCa (TCGA)
MYC mRNA
MYC-sig
****
N (51)
T (51)
B
Paired samples: Pri-PCa (Wyatt 2014)
MYC mRNA
MYC-sig
****
N (12)
T (12)
C
MYC-sig
****
N (52)
GS6 (45)
GS7 (238)
GS8 (49)
GS9/10 (90)
TCGA-PRAD
(Pan-Cancer 2018)
D
MYC-sig
***
N (52)
Pri-PCa (281)
Tx (49)
mCRPC (98)
TCGA, Cell 2015
SU2C 2015
Taylor 2010 (179)
E
MYC-sig
**
***
****
Normal (29)
Pri-PCa (131)
mPCa (19)
F
Normal (29)
r = 0.32
P = 0.0002
G
Pri-PCa (131)
r = 0.44
P = 0.06
H
mPCa (19)
r = 0.76
P < 0.0001
MYC-sig
Stemness

Fig. 8


A
Normal (116)
Benign (52)
Pri-PCa (422)
Pri-PCa/ADT (70)
mCRPC (59)
mCRPC (135)
cr_AR-A
J-T: ****
N
PRAD
mCRPC
B
RB1-loss sig.
C
PTEN-loss sig.
D
PCa-Stem
E
Stemness
cr_AR-A
r=0.77
P<0.0001
F
RB1-loss sig.
r=0.74
P<0.0001
G
PTEN-loss sig.
r=0.34
P<0.0001
H
PCa-Stem
r=0.79
P<0.0001
I
PCa-Stem
cr_AR-A
r=0.84
P<0.0001
J
RB1-loss sig.
r=0.91
P<0.0001
K
PTEN-loss sig.
r=0.36
P<0.0001
L
Normal (N)(174)
Normal Prostate (122)
Adjacent Benign (52)
Primary PCa (PRAD)
Pri-PCa (714)
mCRPC (335)
mCRPC (286)
mCRPC-Adeno (34)
mCRPC-NE (15)
M
PCa-Stem (12)
11
1
MELK
62
cr_AR-A (63)
N
PCa-Stem (12)
7
5
AURKB
BIRC5
CENPA
DLGAP5
NEK2
154
RB1-loss sig. (159)
O
PCa-Stem (12)
12
0
45
PTEN-loss sig. (45)
P
PCa-Stem (12)
CasR_AR-A (63)
6
1
55
0
5
7
147
ANLN
RRM2
MCM4
TYMS
PLK1
CCNE2
CCNB1
RB1-loss sig. (159)