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], Julian Ho[4], Yang (Seagle) 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], Zou Cheng[9], Dingxiao Zhang[9], Gurkamal Chatta[10], Kent L. Nastiuk[11], David W. Goodrich[1], Kiera Rycaj[1], Anmbreen Jamroze[1], Jason S. Kirk[1], Igor Puzanov[10], Song Liu[2*], Jianmin Wang[2*], Dean G. Tang[1*]

[1]Department of Pharmacology and Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA
[2]Department of Biostatistics and Bioinformatics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA
[3]Department of Biostatistics, University at Buffalo, Buffalo, New York, USA
[4]Veracyte Inc, San Diego, California, USA
[5]Department of Radiation Oncology, University of California, San Francisco, California, USA
[6]Department of Hematology and Oncology, Department of Internal Medicine, Rogel Cancer Center, University of Michigan, Ann Arbor, Michigan, 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 Medicine, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA
[11]Department of Cancer Genetics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA

[§]These authors contributed equally to this work.

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

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

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

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

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

Running title: Increasing Stemness Drives Prostate Cancer Progression.

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

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

ABSTRACT

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

INTRODUCTION

Malignant transformation and cancer progression are characterized by progressive loss of differentiation (or oncogenic dedifferentiation) and gain of stem cell traits (i.e., stemness), and poorly differentiated tumors are enriched in gene expression profiles of normal embryonic and adult stem cells and in cancer stem cells[1, 2, 3, 4, 5]. However, the oncogenic dedifferentiation-based stemness 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 >35,000 American men in 2025[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. On the other hand, AR is the master regulator of normal prostatic epithelium differentiation and represents the therapeutic target of PCa treatments including androgen-deprivation therapy (ADT) and ARPIs. Inevitably, tumors develop resistance to ADT/ARPIs leading to the state of castration-resistant PCa (CRPC), which involves many mechanisms including genetic alterations of *AR* and key genes in AR pathway[10], lineage plasticity[8, 11, 12], and increased stemness[8, 13]. It is unclear how the global pro-differentiation AR signaling activity changes and evolves during the spectrum of PCa initiation and progression and during development of treatment resistance and metastasis. These discussions raise the need to quantify the AR signaling and stemness and closely track their dynamic evolution and inter-relationship during PCa progression.

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

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

93 **RESULTS**
94

95 **Adapting and developing Stemness and AR-A metrics for PCa analysis.**
96 To determine the global and dynamic changes of and inter-relationship between degree of malignancy
97 (oncogenic dedifferentiation) and level of (luminal) differentiation during PCa development and progression
98 (Supplementary Fig. S1A-B), we employed two transcriptome-based algorithms to annotate 26 datasets
99 of 3,102 bulk RNA-seq and 84,081 microarray samples encompassing the PCa continuum (Supplementary
100 Fig. S1B; Supplementary Table S1). In our studies, we focused on and compared three progressive stages:
101 from normal prostate to localized Pri-PCa, from low-grade to high-grade Pri-PCa, and from treatment-naïve
102 Pri-PCa to CRPC/mCRPC (Supplementary Fig. S1A). We defined "progression" as stages beyond Pri-
103 PCa and as disease entities that are more aggressive in a comparative manner. Specifically, we adapted
104 the Stemness Index[6] (mRNAsi; **Stemness** for short) (Supplementary Fig. S1C) to gauge the degree of
105 malignancy (or aggressiveness) and employed the Z-score normalized canonical (pro-differentiation) AR
106 activity or **AR-A** (Supplementary Fig. S1D) to measure the level of differentiation. Canonical AR-A can be
107 determined by quantifying AR-regulated transcripts under intact androgen/AR signaling conditions, and
108 there are several such AR-regulated gene lists[14, 15, 16, 17, 18] (Supplementary Fig. S2A-B). We validated the
109 AR-A list based on their enrichment in prostatic epithelium differentiation, correlation of individual gene
110 expression with AR-A signatures in clinical datasets, and overall consistency in correlation among the AR-
111 A scores (Supplementary Fig. S2B-D). Given that Bluemn's AR-A score[14] correlated well with other scores
112 (Supplementary Fig. S2E-G) and the 10 genes in this signature consistently represented AR targets in
113 multiple clinical datasets (Supplementary Fig. S2D; Supplementary Table S2), we decided to use this AR-
114 regulated gene list for our AR-A quantification.
115

116 **Normal prostate epithelium shows a negative correlation between Stemness and AR-A.**
117 The normal human prostate (NHP) comprises 2 major epithelial types: basal and luminal cells
118 (Supplementary Fig. S3A). We annotated two studies that profiled FACS-purified NHP luminal and basal
119 cell populations[19, 20] and observed that the luminal cell population (Trop2^{+}CD49f^{lo}) possessed higher *AR*
120 mRNA levels and AR-A but lower Stemness compared to basal cells (Trop2^{+}CD49f^{hi}) and, in both datasets,
121 AR-A negatively correlated with Stemness (Supplementary Fig. S3B-C). These results support that the
122 NHP basal cell compartment harbors primitive stem cells[7, 19, 20, 21, 22] and AR-A measures the level of
123 prostate epithelial differentiation.
124

125 **Pri-PCa exhibits concordantly increased Stemness and AR-A, and neoadjuvant ADT (nADT)**
126 **decreases both Stemness and AR-A.**
127 In analyzing AR-A and Stemness in matched pairs of prostate tumor (T) and normal (N; tumor-adjacent
128 benign) tissues in the TCGA-PRAD[18] (n=51 pairs), Wyatt 2014[23] (n=12 pairs), and Long 2020[24] (n=10 pairs)

129 cohorts, we observed concordantly increased AR-A and Stemness (Fig. 1A-C). Accompanying elevated
130 AR-A, *AR* mRNA levels were also increased in Pri-PCa compared to benign tissue (Fig. 1A-C; note that
131 the increase in *AR* mRNA levels in the Long 2020 cohort approached statistical significance). The paired
132 T/N comparisons (Fig. 1A-C) suggest that early PCa initiation and growth are characterized by
133 concordantly increased Stemness and AR-A. To test whether AR-A might causally drive increased
134 Stemness early during prostate tumorigenesis, we investigated 4 PCa nADT (i.e., ~2-6 months of ADT
135 before prostatectomy or radiation) cohorts[24, 25, 26]. In the first 3 cohorts with matched pairs of pre- and post-
136 nADT samples (from the same patients), nADT concordantly downregulated AR-A (without decreasing *AR*
137 mRNA levels) as well as the Stemness (Fig. 1D-F). In the Nastiuk cohort with non-paired post-nADT and
138 no-nADT samples (i.e., from different patients but matched for ages and tumor grades and stages), nADT
139 increased *AR* mRNA levels (Supplementary Fig. S4A) but reduced both AR-A and Stemness
140 (Supplementary Fig. S4B-C).
141
142 **Pri-PCa progression is accompanied by reduced AR-A but continually increasing Stemness.**
143 We next assessed the dynamic changes of Stemness and AR-A during *spontaneous* (i.e., therapy-free)
144 Pri-PCa progression (Fig. 2). To this end, we sorted out treatment-naïve tumors in TCGA-PRAD by
145 excluding samples from patients who received adjuvant (post-surgery) hormone therapy (i.e., Pri-PCa/ADT
146 or Tx; n=70) (Fig. 2A) and grouped the Pri-PCa samples according to increasing GS (Fig. 2B-D). When
147 compared with the 52 benign (N) samples, all the four GS groups from Pri-PCa (GS6, GS7, GS8, and
148 GS9/10 combined) showed higher AR-A and Stemness (Fig. 2C-D), consistent with the results in pairwise
149 T/N comparisons (Fig. 1A-C). However, when compared to GS6 PCa, the three higher GS groups showed
150 increasing *AR* mRNA levels (Fig. 2B) but decreasing AR-A (Fig. 2C) and increasing Stemness (Fig. 2D).
151 Jonckheere-Terpstra's (J-T) trend test validated a gradual reduction in canonical (pro-differentiation) AR
152 signaling but an increasing trend of Stemness with Pri-PCa progression (Fig. 2C-D). Notably, although we
153 observed a modest positive correlation between AR-A and Stemness in pooled Pri-PCa samples
154 (Pearson's r=0.27; Fig. 2E), this positive correlation gradually diminished with PCa progression when we
155 categorized the samples by tumor grade (Fig. 2F). The linear regression analysis confirmed that the
156 interaction between Stemness and AR-A significantly depended on tumor grade (P=0.0003), with a linear
157 decrease in the slope of the correlation as GS increased (Fig. 2G).
158 Collectively, these results indicate that, in contrast to early tumorigenesis, *spontaneous PCa*
159 *progression (i.e., with increasing GS and tumor grade) is marked by a divergence between oncogenic*
160 *Stemness and pro-differentiation AR-A*.
161
162 **Therapies and metastatic progression drive further divergence between attenuated AR-A and**
163 **increased Stemness.**

164 Next, we asked how clinical treatment and metastatic progression may impact the Stemness and AR-A by
165 querying untreated Pri-PCa and Pri-PCa/ADT from TCGA and mCRPC in SU2C[10] (Fig. 3A-C). AR-A was
166 elevated in Pri-PCa compared to adjacent benign tissue (N) but reduced in more aggressive Pri-PCa/ADT
167 samples, while the Stemness, in contrast, continually and persistently increased from N → Pri-PCa → Pri-
168 PCa/ADT (Fig. 3B-C). Strikingly, mCRPC, despite markedly upregulating *AR* mRNA (Fig. 3A), showed the
169 lowest AR-A but highest Stemness (Fig. 3B-C), *indicating a shift from a positive correlation between AR-*
170 *A and Stemness in Pri-PCa to a negative association in mCRPC* (Fig. 3D).

171 By employing normalized RNAseq data from the PCa transcriptome atlas (Bolis 2021), which
172 included normal prostate specimens (n = 174), Pri-PCa (n = 714) and mCRPC (n = 335)[27], we tracked the
173 Stemness and AR-A across the PCa spectrum (Fig. 3E). AR-A increased from benign to Pri-PCa but
174 declined in Pri-PCa/ADT and continued to decrease in mCRPC, reaching its lowest in mCRPC-NE[28] as
175 supported by J-T test (Fig. 3E). In contrast, the Stemness exhibited a continually increasing trajectory
176 across the PCa spectrum encompassing Normal/Benign → treatment-naïve Pri-PCa → Pri-PCa/ADT →
177 (m)CRPC (P<0.0001, J-T Trend test; Fig. 3F; Supplementary Fig. S5A). Notably, the 3 subtypes of mCRPC,
178 i.e., AR-positive (ARPC), CRPC-NEPC and AR^-/NE^- double-negative PCa (DNPC) all showed higher
179 Stemness than Pri-PCa and also displayed increasing Stemness among the 3 subtypes (Supplementary
180 Fig. S5B). Unexpectedly, tumor-adjacent benign prostate tissues showed much higher Stemness than
181 normal prostate in the Genotype-Tissue Expression Project (GTEx)[29] (Fig. 3F; Supplementary Fig. S5A),
182 which were all obtained from cancer-free organ donors. This surprising finding suggests that the
183 transcriptomes of adjacent benign tissues have apparently skewed towards those of cancer samples,
184 consistent with previous observations that adjacent benign tissues represent an intermediate state
185 between healthy and tumor tissues[30]. Also of interest, the Cancer Cell Line Encyclopedia (CCLE)[31, 32] cell
186 lines including PCa cell lines, and PCa-derived PDX[33] and xenografts[34] exhibited the highest Stemness
187 (Fig. 3F; Supplementary Fig. S5A), supporting that cancer cell lines and xenografts are clonally derived
188 from and highly enriched in cancer stem cells (CSCs).

189 As in our early dissection of the relationship between AR-A and Stemness during spontaneous Pri-
190 PCa progression (Fig. 2F-G), analysis of the Bolis 2021 PCa transcriptome atlas[27] revealed a similar
191 change from a positive correlation between AR-A and Stemness in Pri-PCa (blue line; Pearson's r=0.32,
192 P<0.0001; Fig. 3G) to a negative correlation between the two in mCRPC (green line; Pearson's r=-0.15,
193 P=0.005; Fig. 3G). We further extended our Stemness and AR-A interrogation to include microarray-
194 derived transcriptomic datasets[35, 36], allowing us to access data from PCa cohorts generated before the
195 Next-generation sequencing (NGS) era. Analyzing the GRID database of over 82,000 prospectively
196 collected biopsy samples of localized PCa from clinical use of the Veracyte Decipher test[35], we
197 systematically quantified the Stemness across the NCCN Risk Groups and observed increasing Stemness
198 along the NCCN Risk trajectory (Fig. 3H). Likewise, in the Taylor dataset[36], we observed that AR-A was
199 the highest in Pri-PCa but significantly decreased in metastatic PCa (mPCa) (Fig. 3I) while the Stemness

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

203 Finally, we analyzed the dynamic changes of Stemness and AR in genetically engineered mouse
204 models[37] (GEMMs) of PCa with increasing aggressiveness due to individual or combined deletion of 3
205 tumor suppressor genes, *Pten*, *Rb1* and *Trp53*. Briefly, *Pten*$^{-/-}$ single knockout (SKO) prostate tumors
206 develop around 9 weeks, and mice rarely develop metastasis with a median lifespan of 48 weeks. In
207 contrast, the *Pten*$^{-/-}$;*Rb1*$^{-/-}$ double knockout (DKO) mice develop highly metastatic PCa that shortens
208 median survival to ~38 weeks. When *Trp53* is further deleted in DKO background, the triple KO (TKO;
209 *Pten*$^{-/-}$;*Rb1*$^{-/-}$;*Trp53*$^{-/-}$) tumors are exclusively AR^{-} NEPC and castration-resistant *de novo* with high
210 metastatic rate and lifespan of ~16 weeks[37]. We found that the AR-A was the highest in SKO tumors but
211 much reduced in aggressive DKO and TKO tumors (Supplementary Fig. S5C). In contrast, Stemness was
212 the lowest in SKO but significantly increased in DKO and TKO (Supplementary Fig. S5D). Correlation
213 analyses revealed a positive correlation between AR-A and Stemness in SKO that turned negative in both
214 DKO and TKO tumors (Supplementary Fig. S5E-H).

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

217

218 **Stemness is prognostic of poor survival in PCa patients.**

219 We investigated global Stemness and its prognostic significance, and observed that high Stemness
220 consistently correlated with worse patient survival across various cohorts including the Spratt 2017[38]
221 (P=0.001, log-rank; Fig. 4A), Tosoian 2020[39] (P=0.014, log-rank; Fig. 4B), CHAARTED 2021[40] (P=0.006,
222 log-rank; Fig. 4C; P=0.149, log-rank; Supplementary Fig. S6A), and the mCRPC (SU2C 2019[41]; P=0.06,
223 log-rank; Supplementary Fig. S6B-C) as well as Taylor 2010[36] (P=0.028, log-rank; Supplementary Fig.
224 S6D-E). Notably, the CHAARTED cohorts displayed the highest mean Stemness compared to Tosoian
225 and Spratt cohorts, and the CHAARTED cohort showed the highest median Decipher Prostate Genomic
226 Classifier[38, 39, 40] among the three cohorts (Fig. 4D). Interestingly, like high Stemness, low AR-A was
227 correlated with poor patient survival in the Taylor dataset (P=0.040, log-rank; Supplementary Fig. S6F).

228

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

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

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

237 Gene set enrichment analysis (GSEA) was performed to delineate the molecular differences between
238 Stemness-low and Stemness-high PCa samples (Fig. 5E-F). GSEA revealed that Stemness-high PCa
239 samples were enriched in gene sets associated with embryonic stem cells (ESC), as well as aggressive
240 cancer phenotypes, such as undifferentiated cancer and metastasis, in both Pri-PCa and mCRPC (Fig.
241 5F). In contrast, Stemness-low PCa samples were enriched in gene signatures linked to lower
242 aggressiveness and inflammatory pathways (Fig. 5E). Additionally, hallmark pathways such as E2F targets,
243 MYC targets, mTORC1 signaling, and DNA repair were upregulated in Stemness-high PCa (Fig. 5F), while
244 TGF-β signaling, TNF-α signaling, and IFN-γ responses were prominent in Stemness-low PCa (Fig. 5E).

245 The aggressive PAM50-LumB molecular subtype[42, 43, 44] was significantly more prevalent in
246 Stemness-high PCa, accounting for 74% in Pri-PCa (compared to 15% in Stemness-low PCa; Fig. 5G)
247 and 43% in mCRPC (compared to 4% in Stemness-low PCa; Fig. 5H). The PCS1 subtype[45, 46], known for
248 its association with ADT resistance and the highest risk of progression to advanced disease in comparison
249 with PCS2 or PCS3, was predominantly enriched in both Stemness-high Pri-PCa and mCRPC (Fig. 5I-L).
250 Notably, the PAM50-Basal subtype[42, 43] increased in Stemness-high mCRPC but not in Pri-PCa (Fig. 5G-
251 H), aligning with the characteristics of mCRPC, which involves lineage plasticity and acquisition of basal,
252 mesenchymal, neural and stem-like phenotypes when developing ARPI resistance[7, 44, 47]. Consequently,
253 Stemness-high samples remained largely AR-dependent[48] in Pri-PCa (Fig. 5M) but transitioned to AR-
254 independent NE subtype[44] in mCRPC (Fig. 5N). Stemness-high PCa was also characterized by high
255 proliferation as evidenced by their enrichment in a 31-gene cell-cycle progression signature[44] (Fig. 5O-P).
256 Finally, we examined the association between Stemness and a signature of lineage plasticity risk after
257 Enza treatment for mCRPC[12] and found that Stemness-high tumors may be at risk for this virulent form of
258 treatment resistance (Fig. 5Q-R).

259 To further assess the clinical relevance of Stemness, we analyzed the prognostic impact of our newly
260 developed 12-gene PCa-Stem signature (Fig. 5D). We observed that higher PCa-Stem signature
261 correlated with worse patient survival in both Pri-PCa (P<0.0001, log-rank; Fig. 5S-T) and mCRPC
262 (P<0.0001, log-rank; Fig. 5U-V) patients. Multivariate analysis, adjusted for clinicopathologic parameters
263 such as age, GS and stage, showed that patients with high PCa-Stem signature had significantly worse
264 independent prognosis compared to those with low PCa-Stem signature. This was demonstrated by worse
265 progression-free survival (PFS) in Pri-PCa (multivariable-adjusted HR=1.636 (1.014 – 2.64), P=0.0438;
266 Fig. 5T) and overall survival (OS) in mCRPC (multivariable-adjusted HR=4.712 (2.011 – 11.04), P<0.001;
267 Fig. 5V), respectively. Furthermore, the PCa-Stem signature steadily and continually increased across the
268 PCa evolutionary spectrum and correlated with PCa disease progression scores as supported by trajectory
269 inference analysis of the Bolis PCa transcriptome atlas[27] (Pearson's r=0.72; Supplementary Fig. S7A-B).
270 Strikingly, gene expression analysis revealed that 11 of the 12 genes (except *KLK12*) in the PCa-Stem

271 signature significantly correlated with the PCa disease progression (Supplementary Fig. S7C). In contrast,
272 among the genes commonly downregulated in Stemness-high PCa samples in both Pri-PCa and mCRPC
273 (Supplementary Table S3), some (e.g., *PTGDS*, *SPARCL1*, *CLU*, *GJA1* and *S100A4*) were involved in
274 regulating prostate-specific glandular structures and luminal functions while others (e.g., *FBLN1*, *SFRP1*,
275 *IGFBP6*, *GAS1* and *TIMP2*) played general roles in epithelial differentiation (Supplementary Fig. S7D).
276 Overall, these findings indicate that *PCa with high Stemness exhibit greater aggressiveness, a higher*
277 *likelihood of developing lineage plasticity and therapy resistance, and poor survival.*
278
279 **Stemness-high PCa have high tumor mutation burden (TMB) and unique genomic features.**
280 We determined the potential differences in genomic features between Stemness-high vs. Stemness-low
281 PCa and found that mCRPC had a higher fraction of altered genome (Fig. 6A) and higher TMB (Fig. 6B)
282 than Pri-PCa. In both Pri-PCa and mCRPC, Stemness-high PCa displayed a higher fraction of altered
283 genome (Fig. 6A) and TMB (Fig. 6B) than Stemness-low PCa although there was no difference in patient
284 ages (Fig. 6C). In Pri-PCa, Stemness-high samples had a higher prevalence of *PTEN* deletion (DEL),
285 *SPOP* and *FOXA1* mutations (MUT) and *MYC* amplification (AMP) whereas in mCRPC, Stemness-high
286 PCa had more prevalent *RB1* DEL and *AR* MUT (Fig. 6D-F).
287 Globally, both Pri-PCa and mCRPC exhibited mutations in more than a dozen genes although most
288 of these alterations were <10% (Supplementary Fig. S8A; green symbols), consistent with reports that
289 PCa is genomically heterogeneous with a broad spectrum of genomic alternations of low penetrance[49, 50].
290 On the other hand, homozygous deletion (HOMDEL) events, particularly, loss of *PTEN* and *RB1*, occurred
291 frequently in both Pri-PCa and mCRPC (Supplementary Fig. S8A; blue symbols) whereas AMP events in
292 oncogenes such as *MYC* were more prevalent in mCRPC (Supplementary Fig. S8A; red symbols).
293 Nevertheless, *PTEN* and *RB1* DEL represented among the most prevalent alterations in early-stage (GS6)
294 PCa (Supplementary Fig. S8B-C). Analysis of genetic alterations across the spectrum of PCa progression
295 revealed that among the 3 tumor suppressors (*RB1, PTEN* and *TP53*), while homozygous *RB1* DEL
296 remained relatively constant at 13%, 6%, 10% and 10% in low-grade (GS6), high-grade (GS9/10), treated
297 Pri-PCa and mCRPC, respectively, *PTEN* HOMDEL (9%, 23%, 30% and 26%) and *TP53* MUT (0%, 21%,
298 27% and 37%) continued to increase along this progression (Supplementary Fig. S8B-C). Among the 3
299 oncogenic events analyzed, while high levels of *TMPRSS2* fusions (~40%) remained unchanged from low-
300 grade to high-grade Pri-PCa to mCRPC, *MYC* (2%, 12%, 19% and 24%) and *AR* (0%, 1%, 6% and 49%)
301 AMP continued to increase along the progression trajectory (Supplementary Fig. S8B-C). In fact, prevalent
302 *AR* MUT (11%) were observed only in the most aggressive and Stemness-highest mCRPC
303 (Supplementary Fig. S8B-C), suggesting potentially high noncanonical AR-A in mCRPC. Analyzing these
304 major genomic alterations in association with PCa Stemness, we observed that *RB1* DEL, *FOXA1*
305 (together with *SPOP*, *IDH1*, and *CHD1*) MUT, and especially *MYC* AMP were significantly associated with

306 increased Stemness in Pri-PCa (Supplementary Fig. S8D). In mCRPC, only *RB1* DEL exhibited a
307 statistically significant association with higher Stemness (Supplementary Fig. S8E).
308
309 **MYC further drives Stemness during spontaneous PCa progression.**
310 *AR* AMP and MUT are exclusively treatment-induced as there were virtually no *AR* genomic alterations in
311 treatment-naïve Pri-PCa (Supplementary Fig. S8B-C). We showed that early during PCa development
312 increased AR-A may drive elevated Stemness (Fig. 1D-F); however, treatment-naïve Pri-PCa displayed
313 continually increasing Stemness but decreasing AR-A (Fig. 2). What could be driving persistently high
314 Stemness during Pri-PCa progression with decreasing AR-A and in the absence of *AR* genomic alterations?
315 We found that Stemness-high Pri-PCa, compared to Stemness-low Pri-PCa, were enriched in MSigDB
316 Hallmark MYC_TARGETS (MYC_Targets_V1 and MYC_Targets_V2; Fig. 5F) and had more prevalent
317 *MYC* AMP (Fig. 6D and 6F). High-grade (GS9/10) Pri-PCa also had more *MYC* AMP events than low-
318 grade (GS6) Pri-PCa (12% vs. 2%; P=0.006, χ^2 test; Supplementary Fig. S8C). Importantly, *MYC* AMP
319 was most prominently associated with increased Stemness in Pri-PCa (Supplementary Fig. S8D). These
320 observations led us to hypothesize that MYC may represent a critical Stemness driver during spontaneous
321 Pri-PCa progression (in absence of therapeutic pressure). To test this hypothesis, we quantified MYC
322 signaling activity[51] (MYC-sig; See Methods and Supplementary Table S2 for gene signature information)
323 across the PCa progression spectrum. We observed that: 1) both *MYC* mRNA and MYC-sig increased in
324 Pri-PCa compared to matched benign samples (Fig. 7A-B); 2) MYC-sig continued to go up during Pri-PCa
325 progression (Fig. 7C); 3) unlike nADT-induced concomitant decline in both AR-A and Stemness (Fig. 1D-
326 F; Supplementary Fig. S4), nADT showed no discernible effect on MYC-sig (Supplementary Fig. S9A-D);
327 and 4) MYC-sig continued to increase across the disease spectrum with mCRPC exhibiting the highest
328 MYC activity (Fig. 7D-E) and with MYC-sig positively correlated with the Stemness (Fig. 7F-H).
329 *These results, together, support MYC as a Stemness driver during PCa progression*.

330 **DISCUSSION**

331 The current project was undertaken to quantitatively co-analyze the global changes of and the
332 interrelationship between cancer stemness and pro-differentiation AR signaling activity during PCa
333 evolution. By employing gene expression-based general Stemness and AR-A signature scores and
334 through developing a new 12-gene PCa-Stem signature, we show that while the differentiation-regulating
335 AR-A steadily declines in the PCa continuum, pervasively increasing Stemness quantitatively measures
336 PCa aggressiveness, plasticity and progression, and represents a poor prognosticator for patient outcomes.
337 We report that Stemness-high PCa possess unique genomic features with high TMBs and are associated
338 with aggressive molecular subtypes such as PAM50-LumB[42-44] and PCS1[45,46]. We further present evidence
339 that elevated Stemness in early PCa might be driven by increased AR-A but increased MYC-sig activity
340 and other genomic alterations likely propagate the exponentially increased Stemness in mCRPC where
341 AR-A is suppressed.
342
343 Early studies reported that a cohort of AR-regulated genes declines in high-grade versus low-grade Pri-
344 PCa[52, 53] and that loss of AR expression or inhibition of AR-A promotes a stem-like phenotype[8, 34, 54, 55].
345 However, the global pro-differentiation AR-A and cancer stemness have not been systematically and
346 quantitatively assessed and co-analyzed in PCa, and it's unclear whether PCa progression is accompanied
347 by increased stemness as in some other cancers such as glioblastoma and breast cancer[6]. Starting with
348 systematic, comprehensive and intentionally "progression timeline-informed" collection of clinical datasets,
349 here we integrate 26 transcriptomic datasets (covering 87,183 samples) encompassing normal prostate,
350 benign, primary tumors, ADT-treated tumors, CRPC subtypes, metastases, and model systems. This
351 enables a "progression-conscious" and clinically anchored stratification framework with clearly annotated
352 clinical information across these 26 cohorts. Our pan-cohort approach enables signature validation in rare
353 archival biopsy samples and cross-platform integration — a translational advantage not addressed by other
354 studies. With these novel and unique dataset collection and curation and data analysis pipelines, our study
355 demonstrates that PCa progression is accompanied by a continual increase in Stemness (score) but loss
356 of canonical AR-A, with the two inversely correlated with each other (Supplementary Fig. S10). Importantly,
357 the Stemness tracks treatment-induced plasticity and PCa aggressiveness, and both high global Stemness
358 and high PCa-Stem signature prognosticate poor patient survival. Consequently, the stemness, a
359 previously qualitative and somewhat ambiguous term that now permeates oncology field and cancer
360 research publications, can be quantitatively measured and, in the case of PCa, the Stemness score
361 quantitatively and positively tracks the entire spectrum of PCa development, lineage plasticity, therapy
362 resistance and metastatic progression. To our knowledge, this is also the first comprehensive study to
363 simultaneously quantify the oncogenic stemness and pro-differentiation AR-A across the PCa continuum.
364

365 The normal prostatic acini and ducts are lined by AR^+ secretory luminal cells and AR^- basal cells together
366 with rare neuroendocrine cells. We find that the NHP luminal cells have higher AR-A compared to basal
367 cells, confirming AR-A as a metric for quantifying canonical AR activity and prostatic epithelial
368 differentiation. On the other hand, basal cells have higher Stemness, consistent with earlier findings from
369 our lab[19] and others[20, 21, 22, 56] showing the basal cell compartment harbors stem cells. Oncogenic
370 transformation is known to confer stem-like features in the initiated cells. Indeed, compared to adjacent
371 benign tissue, Pri-PCa exhibit elevated Stemness, which is associated with and likely driven by increased
372 AR-A. Loss of *PTEN* and *RB1* (9% and 13%, respectively, in GS6 tumors) could also potentially contribute
373 to increased Stemness in early prostate tumors.
374
375 During our studies, we made a surprising finding that tumor-adjacent benign tissues, the so-called 'normal'
376 tissues used in most comparative studies, actually display much higher Stemness than the true normal
377 prostate obtained from cancer-free organ donors, suggesting that these benign tissues are not normal and
378 their transcriptomes have apparently skewed towards those of cancer samples. This important new
379 revelation prompted us to use, in relevant subsequent studies, the GTEx-derived normal prostate tissue as
380 the baseline comparator, which obviously strengthens the normalization standard across cohorts,
381 improves cross-stage comparation and interpretation, and enhances the scientific rigor of our data.
382
383 In treatment-naïve Pri-PCa, tumor progression is accompanied by decreasing AR-A but persistently high
384 Stemness (Supplementary Fig. S10). The decreasing AR-A is driven by progression-related loss of
385 differentiation while increasing Stemness is likely driven by *MYC* amplification and increased MYC activity
386 with contributions from other oncogenic evens such as loss of *PTEN* and *RB1*. Long-term ADT/ARPI
387 treatment, as expected, further attenuates AR-A but substantially upregulates the Stemness such that
388 mCRPC have the highest Stemness, which is probably driven by continually increasing genomic alterations
389 such as *MYC* amplification and loss of tumor suppressors. Notably, as *AR* alterations (AMP/MUT) are
390 exclusively induced by ARPIs, noncanonical AR-A (i.e., castration-resistant AR activity) likely make
391 significant contributions to the increased Stemness in mCRPC.
392
393 What biological correlates do our Stemness metrics measure? *First*, it measures transcriptional features
394 and relative abundance of stem cells and CSCs. Therefore, the NHP basal/stem cells display higher
395 Stemness than luminal cells (consistent with literature[19, 20, 21, 22]); ARPI treatment, while shutting down AR
396 signaling, enriches CSCs[7, 34, 54, 55] and upregulates Stemness (this study); and PCa cell lines, xenografts
397 and PDX, known to be clonally derived from cancer stem/progenitor cells, demonstrate the highest
398 Stemness. *Second*, it measures cellular and lineage plasticity induced by treatment and genetic alterations.
399 Thus, mCRPC, which harbor more abundant treatment-reprogrammed $AR^{-/lo}$ cells and PCa cells with stem-
400 like features or in an AR-indifferent state[8, 12, 13, 14, 16, 33, 47, 48], manifests the highest Stemness. Our analysis

401 further substantiates the aggressive nature of ($AR^{-/lo}$) CRPC-NEPC and DNPC compared to ARPC, as
402 evidenced by their elevated PCa-Stem signature scores. This suggests a shift in the phenotypic landscape
403 of mCRPC towards more stem-like and aggressive forms, particularly under the pressure of advanced
404 therapies. Likewise, aggressive DKO and TKO murine prostate tumors have undergone lineage
405 transformation and also exhibit higher Stemness than indolent SKO tumors. *Third*, the Stemness metric
406 likely reports the high proliferative status of advanced PCa and mCRPC, as Stemness-high tumors in both
407 Pri-PCa and mCRPC are enriched in "Cell-cycle progression" signature and most genes in the 12-gene
408 PCa-Stem Signature are involved in mitosis and cell-cycle progression. *Finally*, the Stemness score
409 measures PCa aggressiveness as high Stemness correlates not only with more advanced and aggressive
410 stages of PCa but also with aggressive PAM50-LumB and PCS1 molecular subtypes.
411
412 Of clinical significance, both the 12-gene PCa-Stem signature and the global Stemness score correlate
413 with poor patient survival, suggesting the possibility of developing the Stemness metric into a predictive
414 biomarker to distinguish aggressive from indolent primary tumors and a prognostic indicator of unfavorable
415 clinical outcomes in mCRPC patients. Notably, 11 of the 12 genes (except *KLK12*) in the newly developed
416 PCa-Stem signature are significantly correlated with PCa progression (Supplementary Fig. S7C)
417 implicating their individual contributions to the increased Stemness in advanced Pri-PCa and mCRPC.
418 Analysis using data from The Human Protein Atlas (www.proteinatlas.org) reveals that these 11 genes are
419 also recognized as unfavorable prognostic markers in PCa and several other cancers (Liu X and Tang,
420 unpublished), raising the potential utility of the PCa-Stem signature as a prognostic marker not only for
421 PCa but also for other malignancies. Coupled with earlier studies linking increased Stemness with
422 aggressive breast and brain cancer subtypes[6], our findings herein underscore the increased Stemness as
423 a 'universal' and fundamental characteristic of cancer progression and aggressiveness.

424 In recent years, several studies[27,57-60] have proposed stemness-based classifications or prognostic models
425 using either single-cell RNA-seq (scRNA-seq), bulk RNA-seq or integrative analyses. These studies
426 reinforce the biological relevance of stemness and highlight oncogenic pathways such as MYC activation,
427 *RB1*/*PTEN* loss, and lineage plasticity in PCa progression. However, to our knowledge, none has
428 established a quantitative, clinically validated stemness metric across a pan-cohort framework as
429 comprehensive as ours. Of note, our 12-gene PCa-Stem signature, derived from bulk transcriptomic
430 stratification, shows strong biological overlap with the lineage plasticity signature (LPSig) reported by Zhao
431 *et al*. using scRNA-seq[59]. Among the shared genes, *HMMR* emerges as a critical node of convergence as
432 it has been shown that *HMMR* promotes NE differentiation and aggressive behavior in CRPC models[59].
433 The presence of *HMMR*, along with *DLGAP5*, *AURKB*, and *BIRC5* in our PCa-Stem signature, highlights
434 the translational potential of this gene set in targeting NEPC and other plasticity-driven CRPC subtypes.
435 Supporting this, Zheng *et al*. used multi-omics integration and machine learning to define three stemness-

436 based PCa subtypes and constructed a nine-gene prognostic signature[57] that also included *HMMR* and
437 *DLGAP5*. *KLK12*, another PCa-Stem gene, was recently identified as a unique marker for SOX9^{High} stem-
438 like luminal epithelial cells enriched following androgen deprivation[61]. Additionally, our findings that *MYC*
439 amplification, *RB1* deletion, and *PTEN* loss are enriched in Stemness-high tumors are consistent with the
440 previously reported chromatin accessibility (ATAC-seq) patterns[48] and multi-omics spatial data[61]. It is worth
441 noting that although our PCa-Stem signature was initially designed for risk stratification in mCRPC, the
442 analytical framework is inherently flexible as the same platform can be extended to derive subtype-specific
443 signatures (e.g., PCa-Stem-v2) tailored for distinct clinical phenotypes such as DNPC, ARPC, or NEPC.
444 This scalability provides a foundation for personalized prognostication and therapy selection in PCa.

445 We recognize that the Stemness algorithm utilized in our studies was not developed specifically for PCa
446 and therefore, its sensitivity to distinguish subtle intra-group differences remains to be fully validated. The
447 bulk RNA-seq data used to derive Stemness and AR-A was not from pure epithelial cancer cells, a
448 technical issue that can be potentially mitigated by scRNA-seq analysis and other Stemness pipelines such
449 as the CytoTRACE[62, 63] and by using cancer-specific genomic features derived, for example, from cell-free
450 DNA[64]. Finally, it will be important to determine whether and how noncanonical (castration-associated) AR
451 signaling (driven by *AR* AMP/MUT) may contribute to the highest Stemness in mCRPC.
452
453 In conclusion, we have developed transcriptome-based gene signature scores to assess pro-differentiation
454 AR activity (AR-A), MYC activity (MYC-sig) and oncogenic stem-cell features (general Stemness score
455 and the 12-gene PCa-Stem signature). These scores enable us to globally profile differentiation and
456 aggressiveness status across the PCa progression spectrum. This work advances our understanding of
457 dynamic evolution of AR-A and cancer stemness during PCa progression and therapy resistance, and the
458 signature scores reported herein have the potential to serve as clinical classifiers and aid in the
459 identification of patients who may benefit from stemness-targeting therapies.
460

461 # METHODS

462 ## PCa clinical datasets and data collection

463 In this study, we utilized a total of 26 published and in-house prostate cancer (PCa) related datasets,
464 including various cohorts encompassing the evolutionary spectrum of PCa (Supplementary Fig. S1A-B):
465 normal prostate, tumor-adjacent benign tissue, treatment-naïve primary PCa (Pri-PCa) with increasing GS,
466 PCa treated with neo-adjuvant ADT (nADT) or long-term ADT, metastatic castration-resistant PCa
467 (mCRPC), and PCa cell lines, PCa patient-derived xenografts (PDX) and xenografts. Clinical information,
468 genomic data and gene expression profiling data were downloaded and detailed information on each
469 dataset is listed in Supplementary Table S1. Dataset descriptions include dataset source references,
470 PubMed IDs (PMIDs; RRID:SCR_004846), DOI links, sample type and source (e.g., normal prostate tissue,
471 adjacent benign prostate tissue or primary culture from PCa patient, primary or metastatic prostate tumor,
472 xenografts, cell lines), sample sizes, data types used in the current study (clinicopathologic data, RNAseq,
473 microarray, whole exome sequencing [WES]), and public data access portals.
474
475 Data Access and Repository Information: Data for this study were retrieved from the following public
476 repositories and databases (persistent identifiers where applicable):
477 1. GTEx Portal (RRID:SCR_013042; https://gtexportal.org/)
478 2. Gene Expression Omnibus (GEO; RRID:SCR_005012; https://www.ncbi.nlm.nih.gov/geo/)
479 3. cBioPortal for Cancer Genomics (RRID:SCR_014555; https://www.cbioportal.org/)
480 4. Zenodo Repository (RRID:SCR_004129; https://zenodo.org/)
481 5. Decipher GRID Database (RRID:SCR_006552; ClinicalTrials.gov identifier: NCT02609269)
482 6. European Genome-phenome Archive (EGA; RRID:SCR_004944; https://ega-archive.org/)
483 7. European Nucleotide Archive (ENA; RRID:SCR_006515; https://www.ebi.ac.uk/ena/browser/home)
484 8. UCSC Xena Functional Genomics Portal (RRID:SCR_018938; https://xenabrowser.net/)
485 Accession numbers and portal links for the 26 datasets described below are summarized in Supplementary
486 Table S1.
487
488 The 26 datasets are briefly described below:
489 1) **GTEx 2013** (PMID: 23715323)[29]: The Genotype-Tissue Expression (GTEx) Project repository collected
490 high-throughput and clinical data of normal tissue from many organs. RNAseq data of 245 normal
491 prostate tissue samples was downloaded from the GTEx portal (RRID:SCR_013042; ref. 29; DOI:
492 10.1038/ng.2653; https://www.gtexportal.org/, version: v8).
493 2) **Smith 2015** (PMID: 26460041)[20]: RNAseq data of 5 prostate basal cell samples ($Trop2^{+}CD49f^{hi}$) and
494 corresponding 5 luminal cell samples ($Trop2^{+}CD49f^{lo}$) FACS-purified from benign human prostate

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

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

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

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

The expanded **TCGA PRAD 2018 (Pan-Cancer Altas)** cohort includes 494 patients with Pri-PCa (422 treatment-naïve Pri-PCa, 70 PCa with post-surgery adjuvant hormone treatment, 2 PCa with chemotherapy) and 52 matched adjacent normal/benign prostate tissue. Clinicopathologic features (age, Gleason scores, PSA levels, tumor stages) and genomic alterations data (mutation frequency, copy number alterations [CNAs], structural variants [SVs], tumor mutation burden [TMB], fraction genome altered) were retrieved from cBioPortal for Cancer Genomics PRAD PanCancer Atlas study (https://www.cbioportal.org/study/summary?id=prad_tcga_pan_can_atlas_2018) and from the Pan-Cancer Atlas companion site (welcome to the Pan-Cancer Atlas webpage: https://www.cell.com/pb-assets/consortium/pancanceratlas/pancani3/index.html). Survival data were obtained from Liu *et al*., 2018[65] (PMID: 29625055; DOI: 10.1016/j.cell.2018.02.052). Normalized RNAseq data were

531 downloaded via NCI GDC hub at UCSC Xena Functional Genomics Portal (https://gdc.xenahubs.net,
532 version 2019-07-20; DOI: 10.1038/s41587-020-0546-8).
533 6) **Rajan 2014** (PMID: 24054872)[25]: Clinicopathologic (Gleason scores) and RNAseq data of 7 matched
534 pre- and post-nADT PCa samples were obtained, respectively, from Table 1 in Rajan *et al*., 2014 (ref.
535 25; DOI: 10.1016/j.eururo.2013.08.011) and the GEO database GSE48403.
536 7) **Sharma 2018** (PMID: 30314329)[26]: Clinicopathologic (Gleason scores) and RNAseq data of 7 matched
537 pre- and post-nADT PCa samples from the responder group (also defined as "Low Impact Group" in
538 the source reference) was obtained, respectively, from supplementary Table 1 in Sharma *et al*., 2018
539 (ref. 26; DOI: 10.3390/cancers10100379) and the GEO database GSE111177.
540 8) **Long 2020** (PMID: 32951005; DOI: 10.1016/j.eururo.2013.08.011)[24]: RNAseq data of 6 matched pre-
541 and post-nADT PCa samples was obtained from the GEO database GSE82071. Additionally, RNAseq
542 data from 10 locally advanced PCa samples along with their paired adjacent benign prostate tissues
543 were obtained from the GEO database GSE82071.
544 9) **Nastiuk (RPCI) 2020** (submitted)[66]: Clinicopathologic data (Gleason scores) and RNAseq data of 43
545 Pri-PCa samples from post-nADT PCa patients and 43 tumor samples from age-, stage-, clinically-
546 matched PCa patients without nADT was obtained from Jamroze *et al*., 2024 (work led by Drs. K.
547 Nastiuk and G. Chatta at the Roswell Park Comprehensive Cancer Center). RNAseq data and
548 clinicopathologic information from the Nastiuk 2020 cohort (unpublished) were provided in-house by
549 Dr. K. Nastiuk and are available from the corresponding author upon reasonable request.
550 10) **Gerhauser 2018** (PMID: 30537516; DOI: 10.1016/j.ccell.2018.10.016)[67]: RNAseq data of 118 primary
551 tumor samples and 9 adjacent benign prostate tissue from early-onset PCa patients was obtained from
552 the European Genome-Phenome Archive (EGA) (https://ega-archive.org), accession number
553 EGAS00001002923.
554 11) **Wyatt 2014** (PMID: 25155515; DOI: 10.1186/s13059-014-0426-y)[23]: Clinicopathologic (Gleason
555 scores) and RNAseq data of 12 primary PCa samples from treatment-naïve PCa patients and 12
556 adjacent benign prostate tissue were obtained from the European Nucleotide Archive (ENA)
557 (https://www.ebi.ac.uk/ena/browser/home), accession number PRJEB6530.
558 12) **Spratt 2017** (PMID: 28358655; DOI: 10.1200/JCO.2016.70.2811)[38]: Transcriptomic profiles
559 (Microarray) and clinicopathologic data of 855 radical prostatectomy specimens from a multi-
560 institutional study of intermediate- and high-risk localized PCa patients were obtained from Veracyte
561 GRID (https://decipherbio.com/grid/).
562 13) **Tosoian 2020** (PMID: 32231245; DOI: 10.1038/s41391-020-0226-2)[39]: Transcriptomic profiles
563 (Microarray) and clinicopathologic data of 405 radical prostatectomy and biopsy specimens from high-
564 risk localized PCa patients were obtained from Veracyte GRID (https://decipherbio.com/grid/).
565 14) **CHAARTED correlatives 2021** (PMID: 34129855; DOI: 10.1016/j.annonc.2021.06.003)[40]:
566 Transcriptomic profiles (Microarray) and clinicopathologic data of 160 biopsy specimens from a Phase

567 3 trial of metastatic hormone-sensitive PCa patients were obtained from the NCTN Data Archive
568 (https://nctn-data-archive.nci.nih.gov), accession number **NCT00309985-D14**.
569 15) **Decipher GRID Biopsy** (PMID: 37060201; DOI: 10.1002/cncr.34790)[35]: Transcriptomic profiles from
570 82,470 prospectively collected prostate biopsy samples were retrieved from the Decipher GRID
571 database (RRID: SCR_006552; https://decipherbio.com/grid; ClinicalTrials.gov identifier:
572 **NCT02609269;** https://clinicaltrials.gov/ct2/show/NCT02609269). Patient data were de-identified from
573 clinical use of the Decipher prostate genomic classifier in accordance with the Safe Harbor method
574 described in the HIPAA Privacy Rule 45 CFR 164.514(b) and (c) (Veracyte, San Diego, CA). The
575 samples, comprising a large cohort of localized prostate cancer (PCa), were utilized to compare the
576 distribution of Stemness Index values across National Comprehensive Cancer Network (NCCN) risk
577 groups.
578 Clinicopathologic Breakdown and Risk Group Analysis:
579 Detailed clinicopathologic review revealed that the majority of the cohort presents with low-grade
580 localized PCa. The TNM staging breakdown includes:
581 • **T1 stage:** Approximately 50% of samples, indicating early-stage cancer localized within
582 the prostate.
583 • **T2 stage:** Approximately 8%, representing cancer confined within the prostate but more
584 extensive than T1.
585 • **T3 and T4 stages:** Less than 1%, suggesting advanced local spread.
586 • **N0:** Approximately 98%, indicating no regional lymph node involvement.
587 • **N1:** Approximately 2%, suggesting regional lymph node involvement.
588 The Decipher GRID database (RRID:SCR_006552) categorizes patients into NCCN risk groups
589 based on TNM stage, PSA level, and Gleason score:
590 • **Very Low and Low Risk:** T1–T2a, N0, M0, PSA <10 ng/mL, and Gleason score ≤6; often
591 managed with active surveillance or less aggressive treatment.
592 • **Intermediate Risk – Favorable:** Typically T2b–T2c, N0, M0, PSA 10–20 ng/mL, and
593 Gleason score 7; managed with less aggressive treatment but closer monitoring.
594 • **Intermediate Risk – Unfavorable:** Similar clinical features to favorable but with higher
595 disease burden or adverse factors requiring careful monitoring.
596 • **High Risk:** Usually involves T3a, N0, M0, PSA >20 ng/mL, or Gleason score 8–10; often
597 managed aggressively with surgery, radiation, and/or hormonal therapy.
598 • **Very High Risk:** Typically includes T3b–T4, any N, M0, and high Gleason scores or PSA
599 levels (PSA >20 ng/mL); patients are at significant risk for metastatic progression and are treated
600 aggressively, often with multimodal therapies.
601

602 Observations on Disease Severity and Treatment Naïvety:
603 Analysis of the cohort indicates that only approximately 13% of cases fall into the High and Very High
604 risk categories (8,376 + 2,399 = 10,775 samples), suggesting that the majority (approximately 87%)
605 are below Intermediate Risk (≤GS7). This distribution implies that most patients were treatment-naïve
606 and likely androgen deprivation therapy (ADT)-free due to less aggressive disease status.
607 16) **Bolis 2021** (PMID: 34857732; DOI: 10.1038/s41467-021-26840-5)[27]: RNAseq and clinicopathologic
608 data of 1,223 clinical samples from an integrated cohort consisting of normal prostate specimens
609 (n = 174), Pri-PCa (n = 714) and mCRPC (n = 335) was obtained from the Zenodo repository (Record
610 ID: 5546618; https://zenodo.org/records/5546618). This resource of Prostate Cancer Transcriptome
611 Atlas (https://prostatecanceratlas.org) was built and integrated from the following studies/datasets: (1)
612 Genotype-Tissue Expression Database (GTEx; PMID: 23715323; ref. 29); (2) The Cancer Genome
613 Atlas (TCGA, TCGA-PRAD; PMID: 26544944; ref. 18); (3) Atlas of RNA-sequencing profiles of normal
614 human tissues[68] (GSE120795; PMID: 31015567; ref. 68); (4) Integrative epigenetic taxonomy of
615 PNPCa[69] (GSE120741; PMID: 30464211; ref. 69); (5) Prognostic markers in locally advanced lymph
616 node-negative prostate cancer (PRJNA477449); (6) The long noncoding RNA landscape of NEPC and
617 its clinical implications[70] (PRJEB21092; PMID: 29757368; ref. 70); (7) Integrative clinical sequencing
618 analysis of metastatic CRPC reveals a high frequency of clinical actionability (PRJNA283922; dbGaP:
619 phs000915; PMID: 26000489; ref. 10); (8) CSER—exploring precision cancer medicine for sarcoma
620 and rare cancers [71](PRJNA223419; dbGaP: phs000673; PMID: 28783718; ref. 71); (9) Molecular basis
621 of NEPC (Beltran 2016; PRJNA282856; dbGaP: phs000909; PMID: 26855148; ref. 28); (10)
622 Heterogeneity of androgen receptor splice variant-7 (AR-V7) protein expression and response to
623 therapy in CRPC[72] (GSE118435; PMID: 30334814; ref. 72); (11) RNAseq of human PCa cell lines and
624 mCRPC tumor[73, 74, 75] (GSE14750; PMIDs: 32460015; ref. 73; 33658518; ref. 74; 34244513; ref. 75 |
625 GSE171729; PMID: 34244513; ref. 75); and (12) Molecular profiling stratifies diverse phenotypes of
626 treatment-refractory metastatic CRPC (PRJNA520923; GEO: GSE126078; PMID: 31361600; ref. 33).
627 17) **Taylor 2010** (PMID: 20579941; DOI: 10.1016/j.ccr.2010.05.026)[36]: Transcriptomic profiles (Microarray)
628 and clinicopathologic (recurrence-free survival) data of 29 adjacent benign prostate tissue, 131 Pri-
629 PCa, and 19 mPCa were obtained from the GEO database (GSE21034) and the cBioPortal for Cancer
630 Genomics PRAD MSKCC 2010 study (https://www.cbioportal.org/study/summary?id=prad_mskcc).
631 18) **SU2C 2015** (PMID: 26000489; DOI: 10.1016/j.cell.2015.05.001)[10]: Clinical and genomic data of 150
632 mCRPC samples were obtained from the cBioPortal for Cancer Genomics PRAD SU2C 2015 study
633 (https://www.cbioportal.org/study/summary?id=prad_su2c_2015). RNAseq data of 98 mCRPC
634 samples was retrieved from Zhang *et al.*, 2020[16] (PMID: 32350277; DOI: 10.1038/s41467-020-15815-
635 7; ref.16).
636 19) **SU2C 2019** (PMID: 31061129; DOI: 10.1073/pnas.1902651116)[41]: This dataset includes
637 clinicopathologic and genomic data for 444 mCRPC samples. RNAseq data of 266 mCRPC samples

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

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

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

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

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

658 24) **Goodrich 2017** (PMID: 28059767; DOI: 10.1126/science.aah4199)[37]: RNAseq data of murine prostate
659 tumors from genetically engineering mouse models (GEMMs) of PCa of different genotypes, including
660 4 SKO (single knockout: PBCre4:*Pten*$^{f/f}$), 13 DKO (double knockout PBCre4:*Pten*$^{f/f}$*:Rb1*$^{f/f}$), and 6 TKO
661 (PBCre4:*Pten*$^{f/f}$*:Rb1*$^{f/f}$*:Trp53*$^{f/f}$) was obtained from the GEO database GSE90891 (*f* = a floxed allele of
662 the indicated genes).

663 25) **CCLE 2018** (PMID: 22460905; DOI: 10.1038/nature11003)[31]: RNAseq data of 1,019 human cell lines
664 were obtained from the Xena Functional Genomics Portal[76] (https://xenabrowser.net, version: 2018-
665 05-30; ref. 76). Among them, eight prostate cancer cell lines — DU145 (RRID:CVCL_0105), LNCaP
666 clone FGC (RRID:CVCL_1379), MDA PCa 2b (RRID:CVCL_4745), NCI-H660 (RRID:CVCL_0459),
667 PC-3 (RRID:CVCL_0035), VCaP (RRID:CVCL_2235), 22Rv1 (RRID:CVCL_1045), and PRECLH
668 (PrEC LH; RRID:CVCL_V626) — were included for analysis.

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

672

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

674 All clinical datasets analyzed in this study consisted exclusively of male patients, consistent with the male-
675 specific nature of PCa. Preclinical models, including xenografts, patient-derived xenografts, and genetically
676 engineered mouse models (GEMMs), were also male-derived. All PCa cell lines utilized in the study were
677 of male origin, with the exception of Supplementary Figure S5A, where a broader cohort of 1,019 cell lines
678 from the Cancer Cell Line Encyclopedia (CCLE), comprising both male- and female-derived samples
679 across multiple cancer types, was analyzed. Analyses of CCLE data in Supplementary Figure S5A were
680 performed without stratification by sex.
681
682 **Transcriptomic data normalization and transcriptome-based signature scores**
683 To ensure consistency and compatibility across diverse datasets, we employed normalization strategies
684 tailored to the specific data types and analytical approaches.
685 Canonical AR activity score (AR-A) and MYC activity score (MYC-sig): Canonical pro-differentiation AR-A
686 can be determined by RNA-seq based measurements of AR-regulated transcripts under intact
687 androgen/AR signaling conditions. Similarly, MYC-sig can be determined by MYC-regulated transcripts.
688 We employed Z-score normalization method[77] to determine AR-A and MYC-sig. Briefly, the numeric AR or
689 MYC activity score was calculated based on a linear combination of expression values (Z-scores) of
690 experimentally validated AR targets[14] or MYC targets[51] (Supplementary Table S2, GSEA Molecular
691 Signature Database 2022.1 version, MSigDB). The target gene lists were detailed in Supplementary Table
692 S2. The expression value for each AR or MYC target genes (Supplementary Table S2) was converted to
693 Z-score by $Z = (x - \mu)/\sigma$, where x is the expression value of the target gene in a specific sample, μ is the
694 mean and σ is the standard deviation across all samples of a gene. Finally, the combined Z-scores were
695 summed across all genes to represent the expression score of canonical AR or MYC target genes. Analysis
696 of gene expression signatures using Z-score normalization (AR-A and MYC-sig) was confined within
697 individual datasets, obviating the need for cross-dataset comparisons.
698 To further minimize batch effects and ensure data integrity and comparability across PCa stages,
699 normalized RNAseq data from a published integrated data framework, the Prostate Cancer Transcriptome
700 Atlas[27] (https://prostatecanceratlas.org), was used. In this harmonized resource, raw RNAseq data with
701 high-quality sequencing reads from different datasets were re-mapped to the human reference genome
702 GRCh38, adjusted for library size, and normalized using the variance stabilizing transformation pipeline
703 using DESeq2 (version 1.28.1; RRID:SCR_015687)[27].
704
705 Transcriptome-based quantitative measurement of oncogenic stemness (the Stemness Index): We
706 adapted mRNA-based Stemness Index (mRNAsi) to determine the degree of oncogenic dedifferentiation
707 (Stemness)[6]. Briefly, we computed Spearman correlations between the stemness model's weight vector
708 and the transcriptome expression profile of the queried samples, assuming paired observations, at least
709 ordinal scale data, and a monotonic relationship without significant outliers. Spearman correlation

710 calculations were performed using the 'stats' package (version 4.3.3) in R (version 4.3.3;
711 RRID:SCR_001905). The resulting correlation coefficients were then linearly transformed to scale the
712 mRNAsi scores between 0 and 1 using the following formula: Stemness Index = (Spearman correlation +
713 0.2619) /0.4151. Spearman correlation is advocated for calculating the Stemness Index because it is more
714 robust with respect to potential cross-dataset batch effects that may arise[6].
715
716 Transcriptome-based determination of "PCa-Stem Signature": To derive the PCa-Stem Signature, we
717 performed single-sample gene set enrichment analysis (ssGSEA) using the Gene Set Variation Analysis
718 (GSVA)[78] package (version 1.50.5; RRID:SCR_021058) in R (version 4.3.3; RRID:SCR_001905). Briefly,
719 to identify robust stemness-associated genes in PCa, we performed differential gene expression (DEG)
720 analysis between the top 33% (Stemness-high) and bottom 33% (Stemness-low) of samples, stratified by
721 Stemness scores in both treatment-naïve primary PCa (TCGA-PRAD) and metastatic CRPC (SU2C 2019)
722 cohorts. Genes significantly upregulated in both cohorts (fold change ≥ 2, FDR < 0.05) were intersected to
723 derive a 12-gene consensus signature—termed the PCa-Stem Signature (Supplementary Table S2).
724 ssGSEA was then performed using the GSVA R package to calculate PCa-Stem scores across PCa
725 datasets. This signature was validated in independent cohorts and assessed for clinical relevance and
726 transcriptomic subtype enrichment, including associations with aggressive PCa subtypes and patient
727 survival outcomes.
728
729 **Integrative analysis of genomic alterations in primary PCa (TCGA) and mCRPC (SU2C)**
730 To explore molecular and clinical features associated with Stemness, we performed integrative analysis of
731 genomic, transcriptomic, clinicopathologic, and outcome data using cBioPortal for Cancer Genomics
732 (RRID:SCR_014555), which provides interactive access to multidimensional cancer genomics datasets
733 and integrated visualization tools[79]. Specifically, we analyzed data from primary prostate cancer (Pri-PCa;
734 **TCGA-PRAD 2018**, Pan-Cancer Atlas) and metastatic castration-resistant PCa (mCRPC; **SU2C 2019**)
735 cohorts. Clinicopathologic variables (e.g., age, Gleason scores, PSA levels, tumor stage), genomic
736 alterations (e.g., mutation frequency, TMB, fraction genome altered, CNAs, and SVs), and patient outcome
737 data (e.g., overall survival, progression-free survival) were retrieved from cBioPortal (RRID:SCR_014555)
738 as described in the "PCa clinical datasets and data collection" section. Transcriptomic data used for
739 Stemness score calculation were obtained as described in the same section. For comparative analysis,
740 mCRPC samples were stratified into top and bottom 33% by Stemness score. In TCGA, only treatment-
741 naïve Pri-PCa samples were used for a similar 33% stratification, excluding 70 patients who had received
742 adjuvant (post-surgery) hormone therapy to reflect spontaneous disease progression. Differential
743 expression analysis was conducted using DESeq2[80] (version 1.42.1; RRID:SCR_015687), and gene-level
744 and genome-wide alteration frequencies were assessed across stratified groups and disease stages (e.g.,

745 Pri-PCa vs. mCRPC) using OncoPrint visualizations, bar plots, and correlation analysis. Associations with
746 clinical outcomes were evaluated as described in the "Statistical analysis" section.
747
748 **Statistical analysis**
749 Statistical analyses were performed using GraphPad Prism software (version 10.4.1 (532);
750 RRID:SCR_002798) and R (version 4.3.3; RRID:SCR_001095).
751 For comparisons between two groups, independent sample or paired *t*-tests were used to compare
752 group means, assuming normality and equal variance. For multi-group comparisons, Tukey's range test
753 was employed following one-way ANOVA (adjusted p-values listed in Supplementary Table S4). The
754 Wilcoxon rank-sum test was applied for comparisons of altered genome fraction, TMB, or age between
755 two groups. For differential expression analyses and comparisons involving multiple hypotheses (e.g.,
756 RNAseq-based gene comparisons), p-values were further adjusted using the false discovery rate (FDR)
757 method to control for multiple testing.
758 Pearson's correlation coefficient was used to evaluate linear associations between two continuous
759 variables, assuming normally distributed paired observations. Linear regression modeling was applied to
760 examine the relationship between two continuous variables, with the slope of the best-fit regression line
761 reported where applicable. Spearman's rank correlation was employed to assess monotonic relationships
762 between continuous or ordinal variables, assuming paired observations and monotonic trends without
763 significant outliers.
764 The Jonckheere-Terpstra (J-T) trend test was used to assess monotonic association between an
765 ordinal variable and a continuous variable. Chi-square test or Fisher's exact test was performed to evaluate
766 the associations between categorical variables.
767 For the analysis of patient progression-free survival (PFS), recurrence-free survival (RFS) and
768 overall survival (OS), the standard Kaplan-Meier method was used for the estimation of survival fractions
769 and log-rank tests were used for group comparisons. Cox proportional hazards models were used for
770 multivariable analysis and for assessing the prognostic significance of gene signatures (Stemness, PCa-
771 Stem signature, etc). For the mRNA-based gene signatures (Stemness, PCa-Stem signature, AR-A),
772 patients were stratified into high and low signature groups based on median split in cohorts including Spratt
773 2017 (Fig. 4C), Tosoian 2020 (Fig. 4C), CHAARTED 2021 (Supplementary Fig. S6C), Taylor 2010
774 (Supplementary Fig. S6C), TCGA-PRAD 2018 (Supplementary Fig. S6C), and SU2C 2019
775 (Supplementary Fig. S6C). For certain analyses, including SU2C 2019 (Supplementary Fig. S6C) and
776 CHAARTED 2021 (Fig. 4C), patients were alternatively stratified by comparing the top 25% versus the
777 bottom 75% of signature scores to assess whether patients with very high Stemness levels exhibited a
778 higher risk of disease progression.
779 All statistical tests were two-sided, and $P < 0.05$ was considered statistically significant.
780

781 **Data and Code Availability**

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

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

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

790 **References**

791 1. Ben-Porath I*, et al.* An embryonic stem cell-like gene expression signature in poorly differentiated
792 aggressive human tumors. *Nat Genet* **40**, 499-507 (2008).

793 2. Pece S*, et al.* Biological and molecular heterogeneity of breast cancers correlates with their
794 cancer stem cell content. *Cell* **140**, 62-73 (2010).

795 3. Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes
796 guides creation of epithelial cancer stem cells. *Cell Stem Cell* **2**, 333-344 (2008).

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

799 5. Loh JJ, Ma S. Hallmarks of cancer stemness. *Cell Stem Cell* **31**, 617-639 (2024).

800 6. Malta TM*, et al.* Machine Learning Identifies Stemness Features Associated with Oncogenic
801 Dedifferentiation. *Cell* **173**, 338-354 e315 (2018).

802 7. Liu X, Li WJ, Puzanov I, Goodrich DW, Chatta G, Tang DG. Prostate cancer as a dedifferentiated
803 organ: androgen receptor, cancer stem cells, and cancer stemness. *Essays Biochem* **66**, 291-303
804 (2022).

805 8. Tang DG. Understanding and targeting prostate cancer cell heterogeneity and plasticity. *Semin*
806 *Cancer Biol* **82**, 68-93 (2022).

807 9. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin*, (2024).

808 10. Robinson D*, et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215-1228
809 (2015).

810 11. He MX*, et al.* Transcriptional mediators of treatment resistance in lethal prostate cancer. *Nat Med*
811 **27**, 426-433 (2021).

812 12. Westbrook TC*, et al.* Transcriptional profiling of matched patient biopsies clarifies molecular
813 determinants of enzalutamide-induced lineage plasticity. *Nat Commun* **13**, 5345 (2022).

814 13. Alumkal JJ*, et al.* Transcriptional profiling identifies an androgen receptor activity-low, stemness
815 program associated with enzalutamide resistance. *Proc Natl Acad Sci U S A* **117**, 12315-12323
816 (2020).

817 14. Bluemn EG*, et al.* Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained
818 through FGF Signaling. *Cancer Cell* **32**, 474-489.e476 (2017).

819 15. Hieronymus H*, et al.* Gene expression signature-based chemical genomic prediction identifies a
820 novel class of HSP90 pathway modulators. *Cancer Cell* **10**, 321-330 (2006).

821 16. Zhang D*, et al.* Intron retention is a hallmark and spliceosome represents a therapeutic
822 vulnerability in aggressive prostate cancer. *Nat Commun* **11**, 2089 (2020).

823 17. Spratt DE*, et al.* Transcriptomic Heterogeneity of Androgen Receptor Activity Defines a de novo
824 low AR-Active Subclass in Treatment Naive Primary Prostate Cancer. *Clin Cancer Res* **25**, 6721-
825 6730 (2019).

826 18. Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell*
827 **163**, 1011-1025 (2015).

828 19. Zhang D*, et al.* Stem cell and neurogenic gene-expression profiles link prostate basal cells to
829 aggressive prostate cancer. *Nat Commun* **7**, 10798 (2016).

830 20. Smith BA*, et al.* A basal stem cell signature identifies aggressive prostate cancer phenotypes.
831 *Proc Natl Acad Sci U S A* **112**, E6544-6552 (2015).

832 21. Moad M*, et al.* Multipotent Basal Stem Cells, Maintained in Localized Proximal Niches, Support
833 Directed Long-Ranging Epithelial Flows in Human Prostates. *Cell Rep* **20**, 1609-1622 (2017).

834 22. Grossmann S*, et al.* Development, maturation, and maintenance of human prostate inferred from
835 somatic mutations. *Cell Stem Cell* **28**, 1262-1274 e1265 (2021).

836 23. Wyatt AW*, et al.* Heterogeneity in the inter-tumor transcriptome of high risk prostate cancer.
837 *Genome Biol* **15**, 426 (2014).

838 24. Long X*, et al.* Immune signature driven by ADT-induced immune microenvironment remodeling in
839 prostate cancer is correlated with recurrence-free survival and immune infiltration. *Cell Death Dis*
840 **11**, 779 (2020).

841 25. Rajan P*, et al.* Next-generation sequencing of advanced prostate cancer treated with androgen-
842 deprivation therapy. *Eur Urol* **66**, 32-39 (2014).

843 26. Sharma NV*, et al.* Identification of the Transcription Factor Relationships Associated with
844 Androgen Deprivation Therapy Response and Metastatic Progression in Prostate Cancer.
845 *Cancers (Basel)* **10**, (2018).

846 27. Bolis M*, et al.* Dynamic prostate cancer transcriptome analysis delineates the trajectory to
847 disease progression. *Nat Commun* **12**, 7033 (2021).

848 28. Beltran H*, et al.* Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer.
849 *Nat Med* **22**, 298-305 (2016).

850 29. Consortium GT. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585 (2013).

851 30. Aran D*, et al.* Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nat Commun*
852 **8**, 1077 (2017).

853 31. Barretina J*, et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer
854 drug sensitivity. *Nature* **483**, 603-607 (2012).

855 32. Ghandi M*, et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature*
856 **569**, 503-508 (2019).

857 33. Labrecque MP*, et al.* Molecular profiling stratifies diverse phenotypes of treatment-refractory
858 metastatic castration-resistant prostate cancer. *J Clin Invest* **129**, 4492-4505 (2019).

859 34. Li Q*, et al.* Linking prostate cancer cell AR heterogeneity to distinct castration and enzalutamide
860 responses. *Nature communications* **9**, 1-17 (2018).

861 35. Weiner AB*, et al.* A novel prostate cancer subtyping classifier based on luminal and basal
862 phenotypes. *Cancer* **129**, 2169-2178 (2023).

863 36. Taylor BS*, et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11-22
864 (2010).

865 37. Ku SY*, et al.* Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis,
866 and antiandrogen resistance. *Science* **355**, 78-83 (2017).

867 38. Spratt DE*, et al.* Individual Patient-Level Meta-Analysis of the Performance of the Decipher
868 Genomic Classifier in High-Risk Men After Prostatectomy to Predict Development of Metastatic
869 Disease. *J Clin Oncol* **35**, 1991-1998 (2017).

870 39. Tosoian JJ*, et al.* Performance of clinicopathologic models in men with high risk localized prostate
871 cancer: impact of a 22-gene genomic classifier. *Prostate Cancer Prostatic Dis* **23**, 646-653
872 (2020).

873 40. Hamid AA*, et al.* Transcriptional profiling of primary prostate tumor in metastatic hormone-
874 sensitive prostate cancer and association with clinical outcomes: correlative analysis of the
875 E3805 CHAARTED trial. *Ann Oncol* **32**, 1157-1166 (2021).

41. Abida W*, et al.* Genomic correlates of clinical outcome in advanced prostate cancer. *Proc Natl Acad Sci U S A* **116**, 11428-11436 (2019).

42. Zhao SG*, et al.* Associations of Luminal and Basal Subtyping of Prostate Cancer With Prognosis and Response to Androgen Deprivation Therapy. *JAMA Oncol* **3**, 1663-1672 (2017).

43. Zhao SG*, et al.* Clinical and Genomic Implications of Luminal and Basal Subtypes Across Carcinomas. *Clin Cancer Res* **25**, 2450-2457 (2019).

44. Coleman IM*, et al.* Therapeutic Implications for Intrinsic Phenotype Classification of Metastatic Castration-Resistant Prostate Cancer. *Clin Cancer Res* **28**, 3127-3140 (2022).

45. You S*, et al.* Integrated Classification of Prostate Cancer Reveals a Novel Luminal Subtype with Poor Outcome. *Cancer Res* **76**, 4948-4958 (2016).

46. Yoon J*, et al.* A comparative study of PCS and PAM50 prostate cancer classification schemes. *Prostate Cancer Prostatic Dis* **24**, 733-742 (2021).

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

48. Tang F*, et al.* Chromatin profiles classify castration-resistant prostate cancers suggesting therapeutic targets. *Science* **376**, eabe1505 (2022).

49. Armenia J*, et al.* The long tail of oncogenic drivers in prostate cancer. *Nat Genet* **50**, 645-651 (2018).

50. Haffner MC*, et al.* Genomic and phenotypic heterogeneity in prostate cancer. *Nat Rev Urol* **18**, 79-92 (2021).

51. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).

52. Tomlins SA*, et al.* Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. *Science* **310**, 644-648 (2005).

53. Stuchbery R*, et al.* Reduction in expression of the benign AR transcriptome is a hallmark of localised prostate cancer progression. *Oncotarget* **7**, 31384-31392 (2016).

54. Schroeder A*, et al.* Loss of androgen receptor expression promotes a stem-like cell phenotype in prostate cancer through STAT3 signaling. *Cancer Res* **74**, 1227-1237 (2014).

55. Giridhar PV, Williams K, VonHandorf AP, Deford PL, Kasper S. Constant degradation of the androgen receptor by MDM2 conserves prostate cancer stem cell integrity. *Cancer research* **79**, 1124-1137 (2019).

56. Liu X*, et al.* Low CD38 Identifies Progenitor-like Inflammation-Associated Luminal Cells that Can Initiate Human Prostate Cancer and Predict Poor Outcome. *Cell Rep* **17**, 2596-2606 (2016).

57. Zheng K*, et al.* Integrative multi-omics analysis unveils stemness-associated molecular subtypes in prostate cancer and pan-cancer: prognostic and therapeutic significance. *J Transl Med* **21**, 789 (2023).

58. Zhang T*, et al.* Identification of a novel stemness-related signature with appealing implications in discriminating the prognosis and therapy responses for prostate cancer. *Cancer Genet* **276-277**, 48-59 (2023).

59. Zhao F*, et al.* Integrated single-cell transcriptomic analyses identify a novel lineage plasticity-related cancer cell type involved in prostate cancer progression. *EBioMedicine* **109**, 105398 (2024).

60. Zhao F*, et al.* Deciphering single-cell heterogeneity and cellular ecosystem dynamics during prostate cancer progression. *bioRxiv*, (2024).

920 61. Bian X*, et al.* Integration Analysis of Single-Cell Multi-Omics Reveals Prostate Cancer
921 Heterogeneity. *Adv Sci (Weinh)* **11**, e2305724 (2024).

922 62. Kirk JS*, et al.* Integrated single-cell analysis defines the epigenetic basis of castration-resistant
923 prostate luminal cells. *Cell Stem Cell*, (2024).

924 63. Gulati GS*, et al.* Single-cell transcriptional diversity is a hallmark of developmental potential.
925 *Science* **367**, 405-411 (2020).

926 64. Chauhan PS*, et al.* Genomic and Epigenomic Analysis of Plasma Cell-Free DNA Identifies
927 Stemness Features Associated with Worse Survival in Lethal Prostate Cancer. *Clin Cancer Res*
928 **31**, 151-163 (2025).

929 65. Liu J*, et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality
930 Survival Outcome Analytics. *Cell* **173**, 400-416 e411 (2018).

931 66. Jamroze A*, et al.* Neoadjuvant androgen deprivation therapy initially activates and subsequently
932 suppresses the prostate cancer tumor immune microenvironment ((Submitted 2024)).

933 67. Gerhauser C*, et al.* Molecular Evolution of Early-Onset Prostate Cancer Identifies Molecular Risk
934 Markers and Clinical Trajectories. *Cancer Cell* **34**, 996-1011 e1018 (2018).

935 68. Suntsova M*, et al.* Atlas of RNA sequencing profiles for normal human tissues. *Sci Data* **6**, 36
936 (2019).

937 69. Stelloo S*, et al.* Integrative epigenetic taxonomy of primary prostate cancer. *Nat Commun* **9**, 4900
938 (2018).

939 70. Ramnarine VR*, et al.* The long noncoding RNA landscape of neuroendocrine prostate cancer and
940 its clinical implications. *Gigascience* **7**, (2018).

941 71. Robinson DR*, et al.* Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297-303
942 (2017).

943 72. Sharp A*, et al.* Androgen receptor splice variant-7 expression emerges with castration resistance
944 in prostate cancer. *J Clin Invest* **129**, 192-208 (2019).

945 73. Nyquist MD*, et al.* Combined TP53 and RB1 Loss Promotes Prostate Cancer Resistance to a
946 Spectrum of Therapeutics and Confers Vulnerability to Replication Stress. *Cell Rep* **31**, 107669
947 (2020).

948 74. Brady L*, et al.* Inter- and intra-tumor heterogeneity of metastatic prostate cancer determined by
949 digital spatial gene expression profiling. *Nat Commun* **12**, 1426 (2021).

950 75. Lim Y*, et al.* Multiplexed functional genomic analysis of 5' untranslated region mutations across
951 the spectrum of prostate cancer. *Nat Commun* **12**, 4217 (2021).

952 76. Goldman MJ*, et al.* Visualizing and interpreting cancer genomics data via the Xena platform. *Nat*
953 *Biotechnol* **38**, 675-678 (2020).

954 77. Lee E, Chuang HY, Kim JW, Ideker T, Lee D. Inferring pathway activity toward precise disease
955 classification. *PLoS Comput Biol* **4**, e1000217 (2008).

956 78. Hanzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and
957 RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).

958 79. Gao J*, et al.* Integrative analysis of complex cancer genomics and clinical profiles using the
959 cBioPortal. *Sci Signal* **6**, pl1 (2013).

960 80. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq
961 data with DESeq2. *Genome Biol* **15**, 550 (2014).

962

Acknowledgements

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

Author information

Authors and Affiliations

Department of Pharmacology and Therapeutics, Roswell Park Comprehensive Cancer Center, Buffalo, New York, USA
Xiaozhuo Liu, Yibing Ji, David W. Goodrich, Kiera Rycaj, Anmbreen Jamroze, Jason S. Kirk & Dean G. Tang

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

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

Veracyte Inc, San Diego, California, USA
Kathy Zhao, Julian Ho, Yang (Seagle) Liu & Elai Davicioni

Department of Radiation Oncology, University of California, San Francisco, California, USA
Felix Y. Feng

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

University Hospitals Seidman Cancer Center and Case Western Reserve University, Cleveland, Ohio, USA
Daniel E. Spratt

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

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

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

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

Authors' Contributions

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

Ethics declarations

Competing interests

All authors declare no competing interests relative to this work.

1031 ## Figure Legends

1032 **Figure 1.** *Pri-PCa shows concordant increases of AR-A and Stemness and both are decreased by nADT.*
1033 **A-C,** PCa has increased AR-A as well as the Stemness compared to matched benign tissues. Pairwise
1034 comparisons of *AR* mRNA levels, canonical AR-A, and Stemness scores showing significant increases
1035 in primary tumors (T) as compared to the matched adjacent benign prostate tissues (N) from the same
1036 individual in TCGA-PRAD cohort (n=51; **A**), and Wyatt 2014 cohort (n=12; **B**). In the Long 2020 cohort
1037 (n=10; **C**), increases in AR-A and Stemness reached statistical significance whereas the increase in *AR*
1038 mRNA levels approached statistical significance ($P = 0.06$).
1039 **D-F,** nADT decreases both AR-A and Stemness. Pairwise comparison of *AR* mRNA levels, AR-A, and
1040 Stemness showing significant decrease of both AR-A and Stemness after nADT (Post-nADT) as
1041 compared to the matched samples before nADT (Pre-nADT) from the same individual (**D**, Rajan 2014
1042 cohort; **E**, Sharma 2018 cohort; **F**, Long 2020 cohort).
1043 In all plots, sample sizes are indicated (see also Supplementary Table S1 for cohort and RNA-seq data
1044 access information). Within the plots, results are shown as the mean ± SD and pairs are linked with a grey
1045 line. Significance was calculated by two-tailed paired Student's *t*-test (ns, not significant; *, $P < 0.05$; **, P
1046 < 0.01; ***, $P < 0.001$; ****, $P < 0.0001$).
1047
1048
1049 **Figure 2.** *Continual increases in Stemness with declining AR-A during spontaneous PCa progression*.
1050 **A,** The pie chart showing sample sizes in different Gleason grades in the TCGA untreated Pri-PCa cohort.
1051 **B-D,** Progression of treatment-naïve Pri-PCa is accompanied by decreasing AR-A but increasing
1052 Stemness. Data is presented as mean ± SD. Shown on top are the Jonckheere-Terpstra's trend (J-T)
1053 test results (wedges).
1054 **E-F,** Progression in Pri-PCa is accompanied by a gradual loss of the positive correlation between canonical
1055 AR-A and Stemness. Scatter plots illustrating the correlation between AR-A and Stemness in (**E**)
1056 combined Pri-PCa cases in TCGA with an overall linear regression model and (**F**) individual GS grades
1057 with corresponding linear regressions showing decreasing slopes from low to high GS grades by J-T test.
1058 The slopes of best-fit regression line are also shown.
1059 **G,** Merged scatter plot showing the correlation between AR-A and Stemness across the 4 GS PCa with
1060 each GS represented by a distinct color. The correlation becomes progressively weaker from GS6 to
1061 GS9/10, transitioning from positive to nearly zero. The correlation coefficients (r) and p-values for each
1062 GS are as follows: GS6 ($r = 0.56$, $P < 0.0001$), GS7 ($r = 0.37$, $P < 0.0001$), GS8 ($r = 0.48$, $P = 0.022$),
1063 and GS9/10 ($r = -0.02$, $P = 0.89$). Linear regression lines are overlaid for each GS, demonstrating the
1064 decline in the correlation co-efficiency as the GS increases.
1065 All statistical analyses were performed using R and statistical significance was assessed using one-way
1066 analysis of variance followed by Tukey's multiple-comparison test (B-D). J-T trend test in B-D and F was
1067 used to calculate the statistical significance of the trend across different groups. *, $P < 0.05$; **, $P < 0.01$;
1068 ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant.
1069
1070
1071 **Figure 3.** *Therapies and metastatic progression drive further divergence between Stemness and canonical*
1072 *AR-A with mCRPC showing the highest Stemness but the lowest AR-A*.
1073 **A-D,** Hormonal treatment further drives down AR-A but significantly upregulates Stemness. **A-C**, mCRPC
1074 displays the lowest pro-differentiation AR-A but highest Stemness. Shown are *AR* mRNA levels (**A**), the
1075 canonical AR-A (**B**) and Stemness (**C**) in tumor-adjacent benign tissues (N), treatment-naïve Pri-PCa
1076 and advanced Pri-PCa with post-surgery adjuvant ADT (Tx) in TCGA, and mCRPC (SU2C 2015).
1077 Sample sizes are indicated in parentheses. Shown on top are the Jonckheere-Terpstra's (J-T) trend test
1078 results (wedges). **D,** Stemness inversely correlates with AR-A in mCRPC**.** Shown are scatter plots with
1079 best-fit linear regression lines illustrating the correlation between AR-A and Stemness with Pearson
1080 correlation coefficient (r), the slopes of lines and J-T test results of slopes indicated below.

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

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

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

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

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

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

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.

1132
1133
1134 **Figure 5.** *Stemness-high PCa are associated with aggressive molecular subtypes and a 12-gene PCa-*
1135 *stem signature prognosticates poor patient survival.*

1136 **A,** Schematic illustrating the profiling strategy used to assess associations between Stemness and
1137 molecular features in PCa. Patients were ranked by Stemness score and stratified into Stemness-high
1138 (top 33%) and Stemness-low (bottom 33%) groups for downstream analyses. Genome-wide functional
1139 analyses were performed using transcriptomic, genomic, and clinical data.

1140 **B-C,** Volcano plots showing differentially expressed genes (DEGs) between Stemness-high and
1141 Stemness-low groups in Pri-PCa (**B**) and mCRPC (**C**) cohorts. Red and blue indicate upregulated and
1142 downregulated DEGs, respectively. Genes with FDR > 0.05 are shown in gray. Sample sizes (n) of each
1143 group are indicated. See Supplementary Table S3 for complete DEG lists.

1144 **D** Venn diagram showing the 12-gene "PCa-Stem signature" derived from overlapping upregulated DEGs
1145 (FC ≥ 2, FDR < 0.05) in both Pri-PCa and mCRPC.

1146 **E-F,** GSEA showing low Stemness is associated with immune signaling and inflammatory response (**E**)
1147 while high Stemness in both cohorts is enriched for embryonic stem cell traits, aggressiveness, DNA
1148 repair, MYC activation, and mTORC1 signaling (**F**). All enrichments are significant ($P < 0.05$, FDR <
1149 0.05).

1150 **G-H,** PAM50 subtyping showing a higher frequency of LumB subtype in Stemness-high versus Stemness-
1151 low samples in Pri-PCa (**G**) and mCRPC (**H**). (****, $P < 0.0001$, χ^2 test).

1152 **I-L,** GSEA showing enrichment of the aggressive PCS1 gene expression signature in Stemness-high
1153 groups and the less aggressive PCS3 signature in Stemness-low groups in Pri-PCa (**I**, **K**) and mCRPC
1154 (**J**, **L**). The Prostate Cancer Subtype (PCS) classification system (45,46) stratifies tumors into three
1155 molecular subtypes based on transcriptomic features: PCS1 (most aggressive), PCS2 (intermediate),
1156 and PCS3 (least aggressive). See Supplementary Table S2 for PCS gene signature.

1157 **M-R,** GSEA showing AR signature enrichment in Pri-PCa Stemness-high (**M**), NEPC signature enrichment
1158 in mCRPC Stemness-high (**N**), and shared enrichment of cell cycle (**O**, **P**) and lineage plasticity (**Q**, **R**)
1159 programs in both cohorts. See Supplementary Table S5 for gene sets.

1160 **S-V,** Kaplan-Meier analyses showing that the 12-gene PCa-Stem signature predicts worse progression-
1161 free survival (PFS) in TCGA Pri-PCa (**S**) and worse overall survival (OS) in SU2C mCRPC (**U**).
1162 Multivariable Cox models adjusting for clinicopathologic variables confirm the PCa-Stem-High subgroup
1163 as independently prognostic for PFS (**T**) and OS (**V**).

1164
1165
1166 **Figure 6.** *Stemness-high PCa is associated with higher genomic instability.*

1167 **A-C,** Box plots showing that Stemness-high subgroup in both Pri-PCa (n = 141) and mCRPC (n = 89)
1168 displays a higher fraction of altered genome (**A**), higher TMB (**B**), and no age difference compared to
1169 Stemness-low cases (**C**). Wilcoxon rank-sum test was used to determine significance (****, $P < 0.0001$;
1170 ns, not significant). Statistical significance was determined using the two-sided Wilcoxon rank-sum test.
1171 Box plots represent the median, interquartile range (IQR), and whiskers extending to 1.5× the IQR.

1172 **D-F,** Comparisons of genomic alteration frequencies (Alt. Freq) in cancer-associated genes between
1173 Stemness-high and Stemness-low subgroups in TCGA Pri-PCa (**D**, **F**) and SU2C mCRPC (**E**, **F**) cohorts.
1174 Statistical significance was determined by two-sided Fisher's Exact test (*, $P < 0.05$).

1175
1176
1177 **Figure 7.** *MYC activation contributes to increased Stemness during spontaneous PCa progression.*

1178 **A-B,** Pairwise comparisons showing increased *MYC* mRNA levels and MYC activity (MYC-sig) in primary
1179 tumors (T) versus matched adjacent benign tissues (N) in TCGA-PRAD cohort (n=51; **A**) and Wyatt
1180 (n=12; **B**) cohorts. Data is shown as mean ± SD; paired samples are linked by grey lines. Significance
1181 was determined using two-tailed paired Student's *t*-test (****, $P < 0.0001$).

1182 **C-H,** MYC activity increases during PCa progression, peaking in mCRPC. In treatment-naïve Pri-PCa (**C**),
1183 MYC-sig is significantly increased in all-grade (GS) tumors as compared to tumor-adjacent benign
1184 tissues (N) with the most advanced GS9/10 tumors exhibiting the highest MYC-sig. In **D**, MYC-sig
1185 increases from normal to treatment-naïve Pri-PCa and further in advanced Pri-PCa treated with adjuvant
1186 ADT (Tx) in TCGA-PRAD, while mCRPC shows the highest MYC-sig in both SU2C 2015 (**D**) and Taylor
1187 2010 (**E**) cohorts. Data is presented as mean ± SD. Statistical significance was assessed by one-way
1188 ANOVA with Tukey's multiple-comparison test (**C-E**). J-T trend test was used in (C, D) to assess the
1189 significance of progressive increase. **, $P < 0.01$; ***, $P < 0.001$; ****, $P < 0.0001$; ns, not significant. **F-**
1190 **H,** Scatter plots showing positive correlations between MYC activity and Stemness in the Taylor 2010
1191 cohort across different stages. Pearson's r and P-values are indicated.

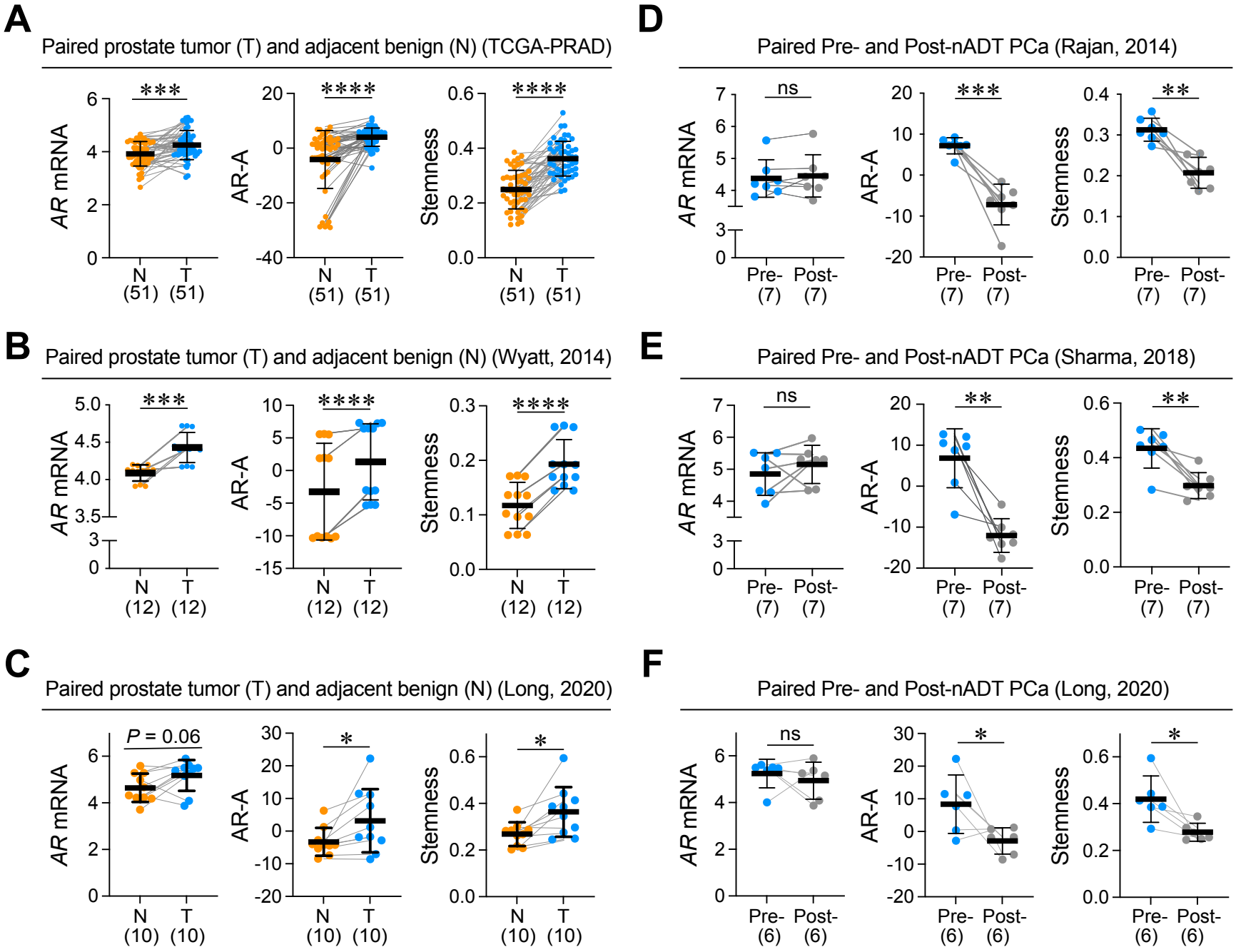

A
Paired prostate tumor (T) and adjacent benign (N) (TCGA-PRAD)
AR mRNA
AR-A
Stemness
***
****
****
N (51)
T (51)
B
Paired prostate tumor (T) and adjacent benign (N) (Wyatt, 2014)
***
****
****
N (12)
T (12)
C
Paired prostate tumor (T) and adjacent benign (N) (Long, 2020)
P = 0.06
*
*
N (10)
T (10)
D
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)

Figure 1

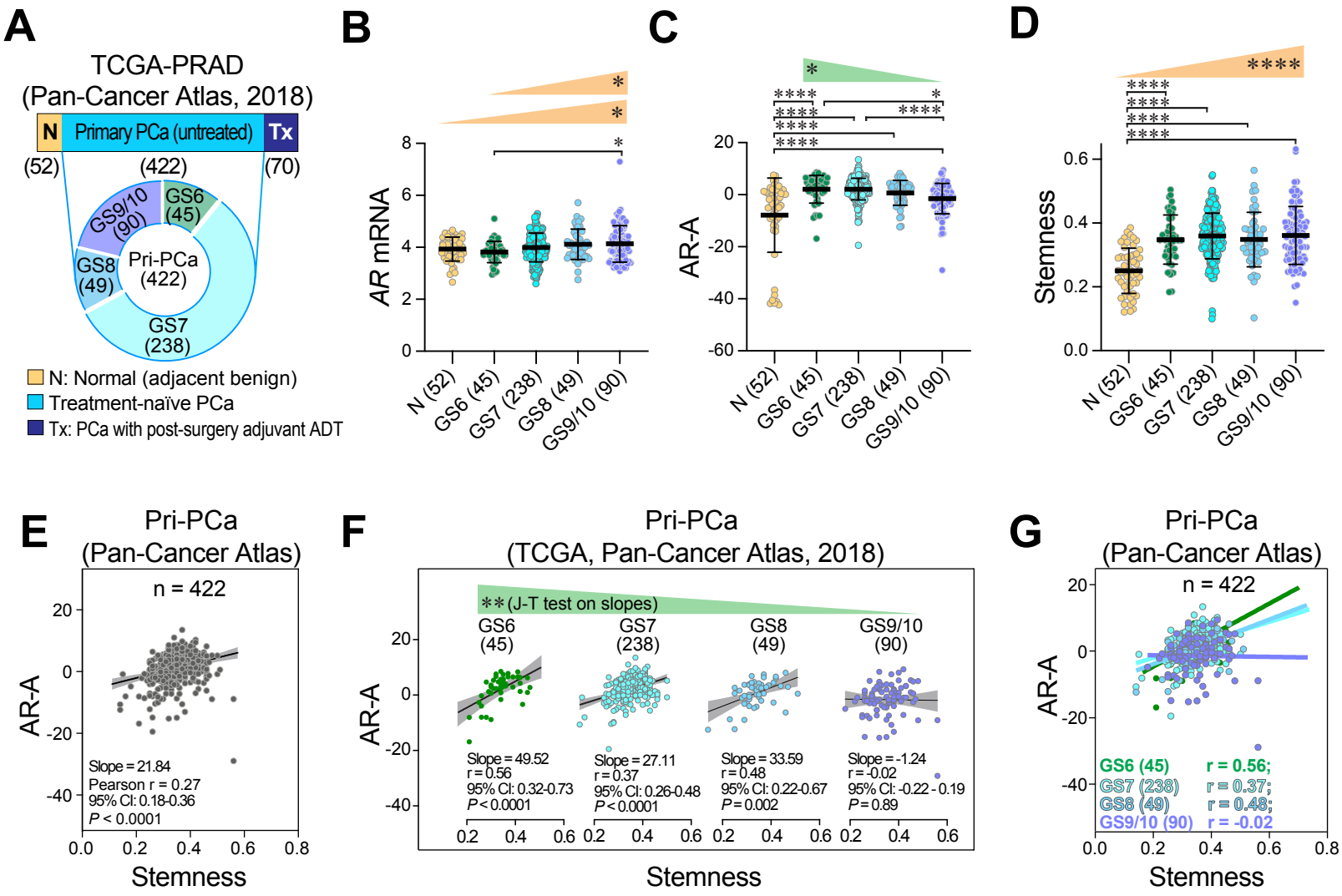

A
TCGA-PRAD
(Pan-Cancer Atlas, 2018)
N
Primary PCa (untreated)
Tx
(52)
(422)
(70)
GS9/10 (90)
GS6 (45)
GS8 (49)
Pri-PCa (422)
GS7 (238)
N: Normal (adjacent benign)
Treatment-naïve PCa
Tx: PCa with post-surgery adjuvant ADT
B
AR mRNA
N (52)
GS6 (45)
GS7 (238)
GS8 (49)
GS9/10 (90)
C
AR-A
D
Stemness
E
Pri-PCa
(Pan-Cancer Atlas)
n = 422
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
AR-A
GS6 (45) r = 0.56;
GS7 (238) r = 0.37;
GS8 (49) r = 0.48;
GS9/10 (90) r = -0.02
Stemness

Figure 2

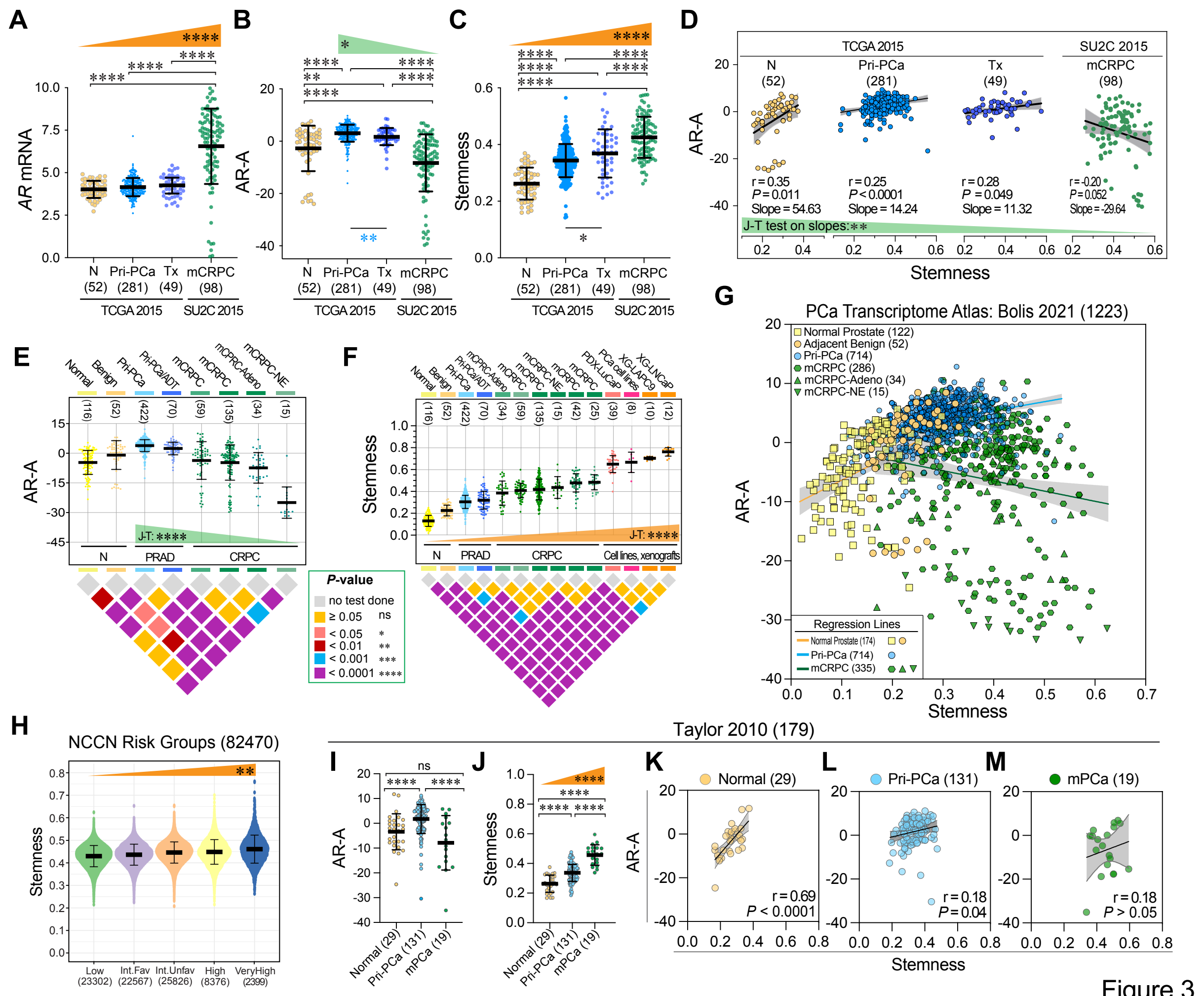

A
AR mRNA
N (52)
Pri-PCa (281)
Tx (49)
mCRPC (98)
TCGA 2015
SU2C 2015
B
AR-A
C
Stemness
D
TCGA 2015
SU2C 2015
r = 0.35
P = 0.011
Slope = 54.63
r = 0.25
P < 0.0001
Slope = 14.24
r = 0.28
P = 0.049
Slope = 11.32
r = -0.20
P = 0.052
Slope = -29.64
J-T test on slopes: **
E
Normal
Benign
Pri-PCa
Pri-PCa/ADT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

Figure 3

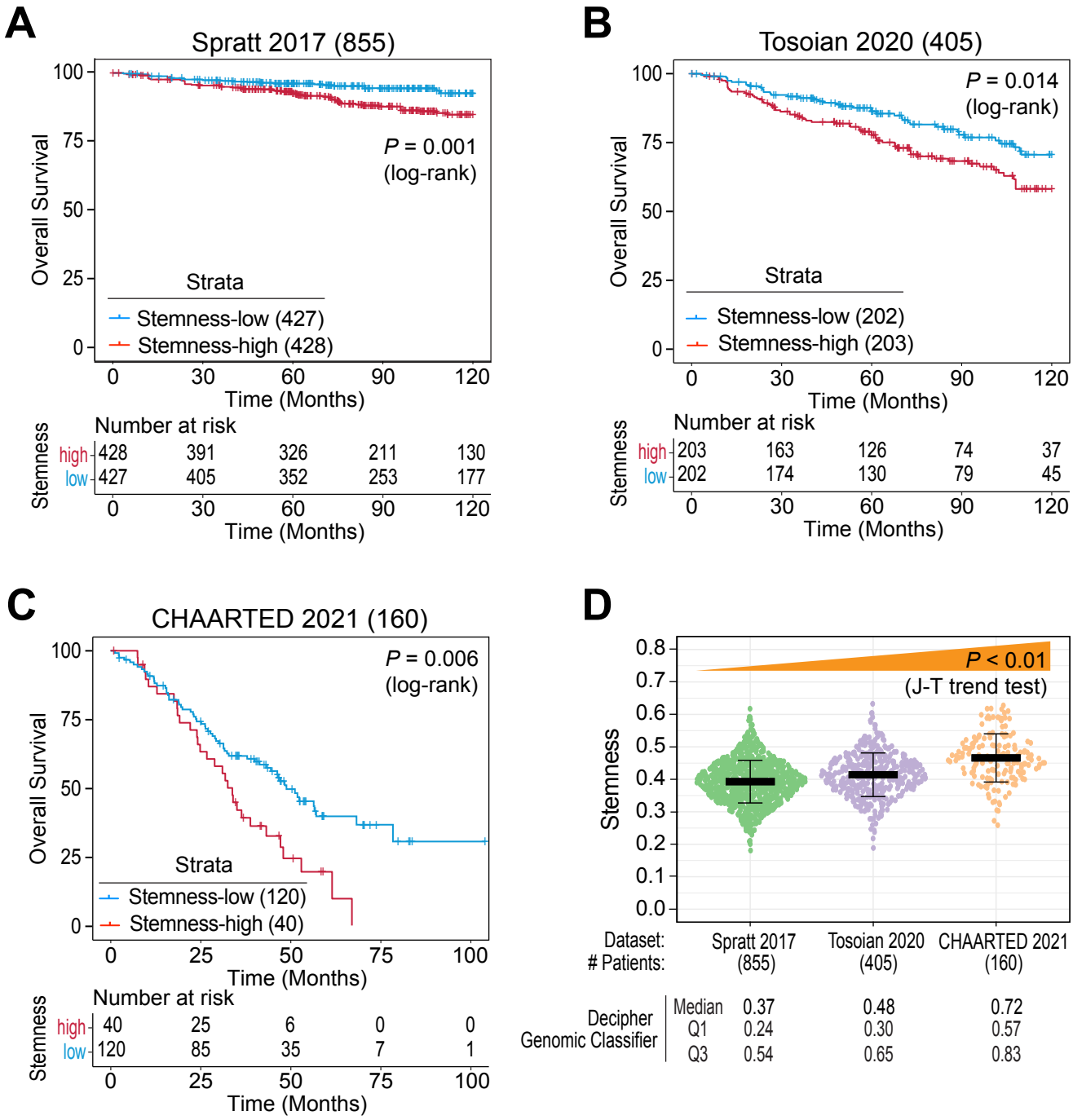

A
Spratt 2017 (855)
Overall Survival
P = 0.001
(log-rank)
Strata
Stemness-low (427)
Stemness-high (428)
Time (Months)
Number at risk
Stemness
high 428 391 326 211 130
low 427 405 352 253 177
B
Tosoian 2020 (405)
P = 0.014
(log-rank)
Stemness-low (202)
Stemness-high (203)
high 203 163 126 74 37
low 202 174 130 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
P < 0.01
(J-T trend test)
Stemness
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

Figure 4

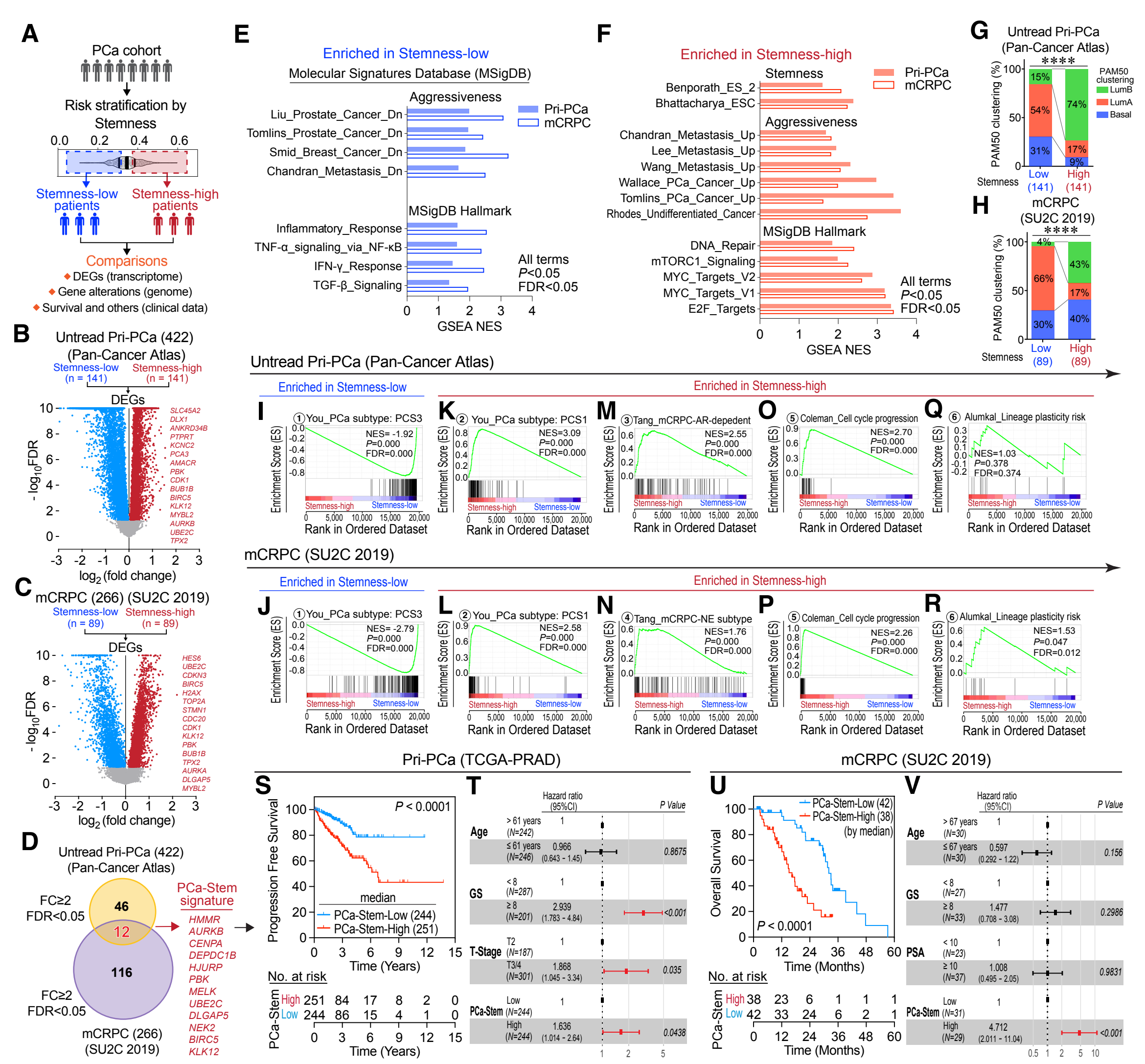

A
PCa cohort
Risk stratification by Stemness
0.0 0.2 0.4 0.6
Stemness-low patients
Stemness-high patients
Comparisons
DEGs (transcriptome)
Gene alterations (genome)
Survival and others (clinical data)
B
Untread Pri-PCa (422) (Pan-Cancer Atlas)
Stemness-low (n = 141)
Stemness-high (n = 141)
DEGs
SLC45A2 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 (%)
PAM50 clustering
LumB
LumA
Basal
Stemness
Low (141)
High (141)
H
mCRPC (SU2C 2019)
PAM50 clustering (%)
Stemness
Low (89)
High (89)
Untread Pri-PCa (Pan-Cancer Atlas)
Enriched in Stemness-low
Enriched in Stemness-high
I You_PCa subtype: PCS3 NES= -1.92 P=0.000 FDR=0.000
K You_PCa subtype: PCS1 NES=3.09 P=0.000 FDR=0.000
M Tang_mCRPC-AR-depedent NES=2.55 P=0.000 FDR=0.000
O Coleman_Cell cycle progression NES=2.70 P=0.000 FDR=0.000
Q 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 You_PCa subtype: PCS3 NES= -2.79 P=0.000 FDR=0.000
L You_PCa subtype: PCS1 NES=2.58 P=0.000 FDR=0.000
N Tang_mCRPC-NE subtype NES=1.76 P=0.000 FDR=0.000
P Coleman_Cell cycle progression NES=2.26 P=0.000 FDR=0.000
R Alumkal_Lineage plasticity risk NES=1.53 P=0.047 FDR=0.012
Pri-PCa (TCGA-PRAD)
S
Progression Free Survival
P < 0.0001
median
PCa-Stem-Low (244)
PCa-Stem-High (251)
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
Overall Survival
PCa-Stem-Low (42)
PCa-Stem-High (38)
(by median)
P < 0.0001
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

Figure 5

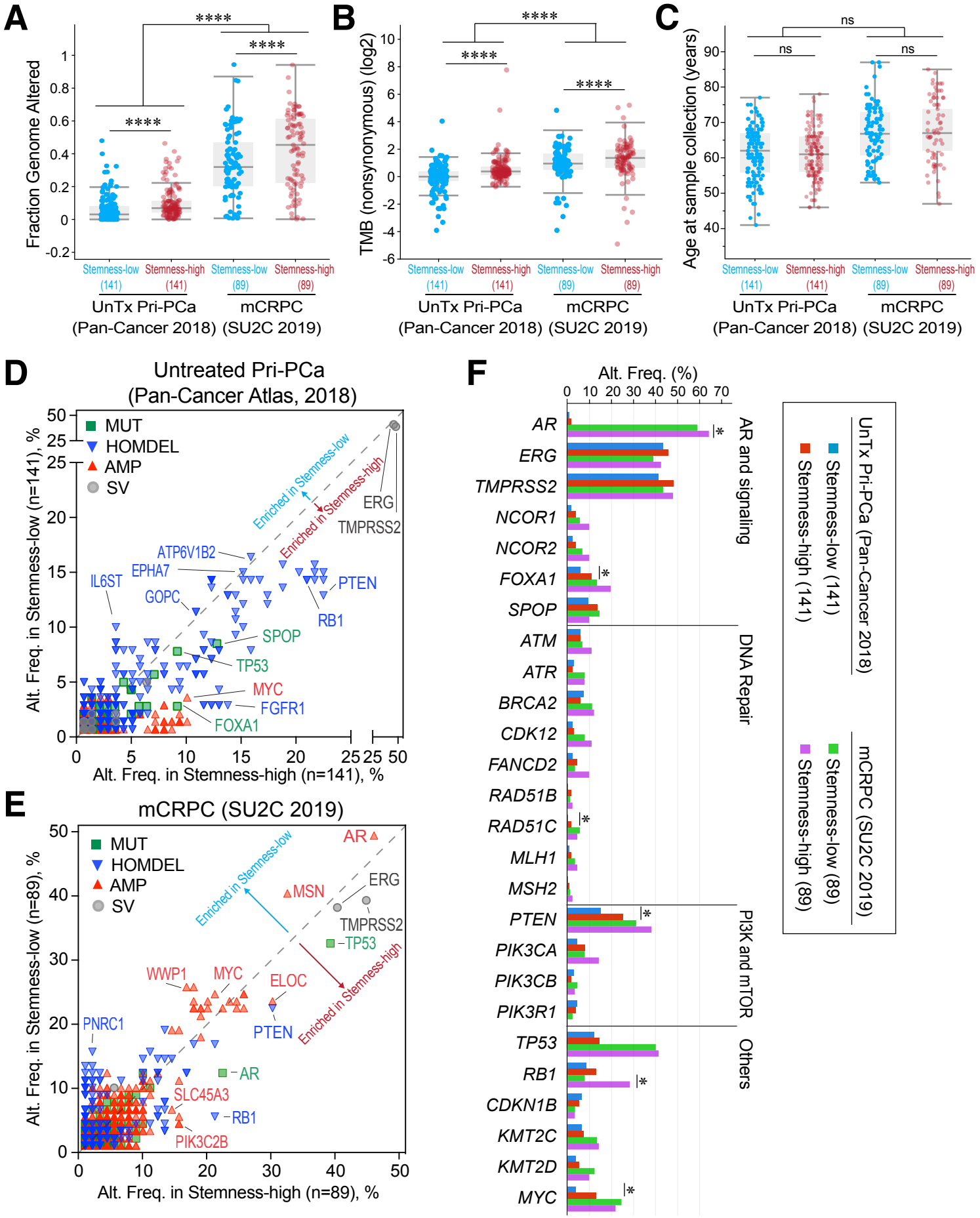

A
Fraction Genome Altered
****
****
****
Stemness-low (141)
Stemness-high (141)
Stemness-low (89)
Stemness-high (89)
UnTx Pri-PCa (Pan-Cancer 2018)
mCRPC (SU2C 2019)
B
TMB (nonsynonymous) (log2)
****
****
****
C
Age at sample collection (years)
ns
ns
ns
D
Untreated Pri-PCa (Pan-Cancer Atlas, 2018)
MUT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

Figure 6

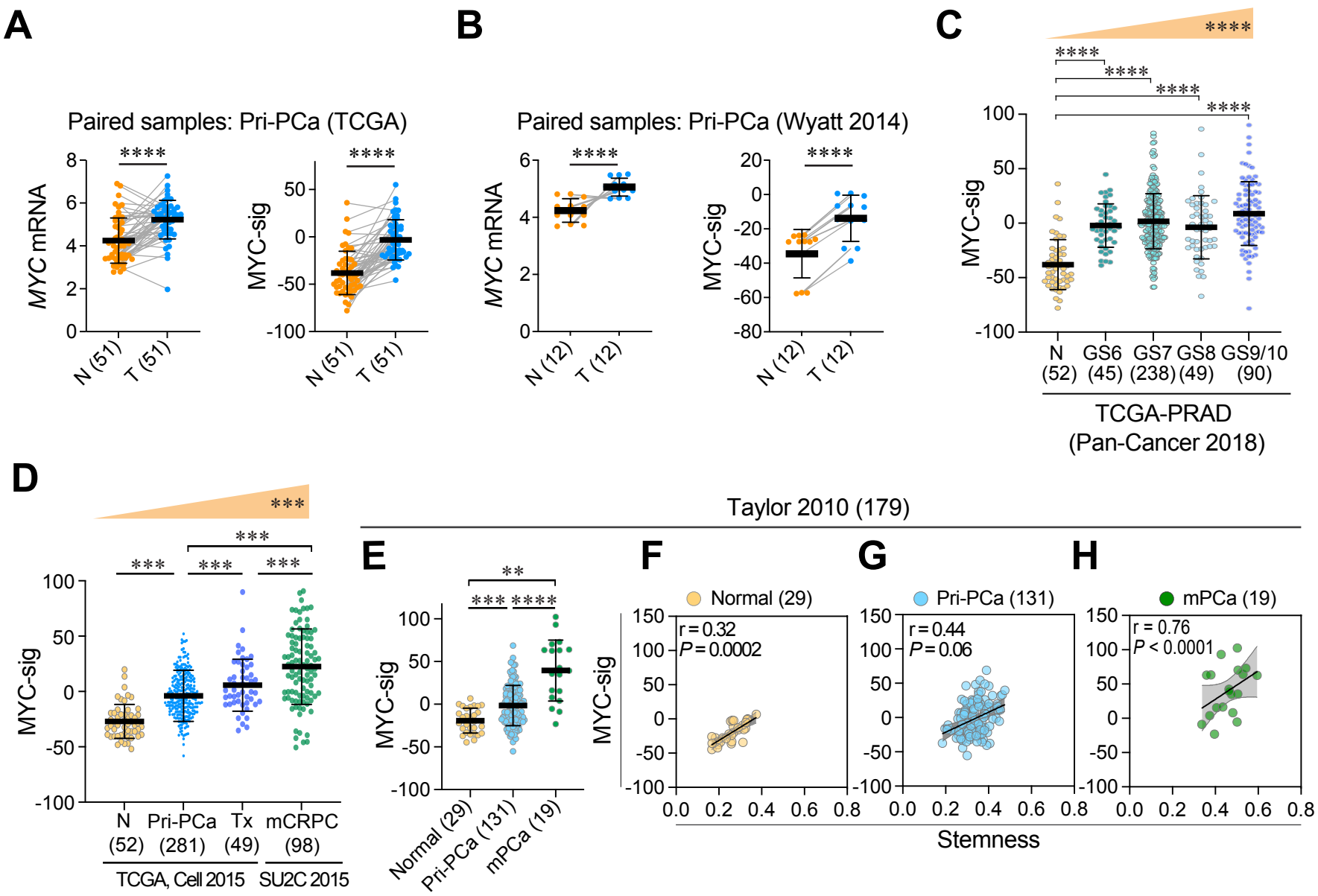

A
Paired samples: Pri-PCa (TCGA)
MYC mRNA
MYC-sig
****
N (51)
T (51)
B
Paired samples: Pri-PCa (Wyatt 2014)
MYC mRNA
MYC-sig
****
N (12)
T (12)
C
MYC-sig
****
N (52)
GS6 (45)
GS7 (238)
GS8 (49)
GS9/10 (90)
TCGA-PRAD
(Pan-Cancer 2018)
D
MYC-sig
***
N (52)
Pri-PCa (281)
Tx (49)
mCRPC (98)
TCGA, Cell 2015
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

Figure 7

Supplementary Information Content

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

Liu X, Cortes E, et al.

Supplementary Table S1. Summary of datasets used in this study
Supplementary Table S2. Summary of gene signatures used in this study
Supplementary Table S3. DEGs identified in Stemness-high versus Stemness-low subgroups in the Pri-PCa (TCGAPRAD) cohort and in the mCRPC (SU2C 2019) cohort (*six sheets)
Supplementary Table S4. Summary of statistical significance of differences using Tukey's multiple comparisons test (*three sheets)
Supplementary Table S5. Summary of gene signatures and gene sets used in Gene Set Enrichment Analysis

Related to Fig. 1
Supplementary Fig. S1. Schematic illustration of bioinformatic datasets and signature score pipelines used for comparisons across the spectrum of PCa evolution.
Supplementary Fig. S2. Transcriptome-based signature scores for canonical (pro-differentiation) AR signaling activity (AR-A).
Supplementary Fig. S3. Prostate luminal cell compartment harbors higher AR-A but lower Stemness as compared to prostate basal cell compartment.

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

Related to Fig. 3
Supplementary Fig. S5. Increased Stemness represents a 'universal' feature of PCa progression and aggressiveness.

Related to Fig. 4
Supplementary Fig. S6. High Stemness correlates with poor patient survival.

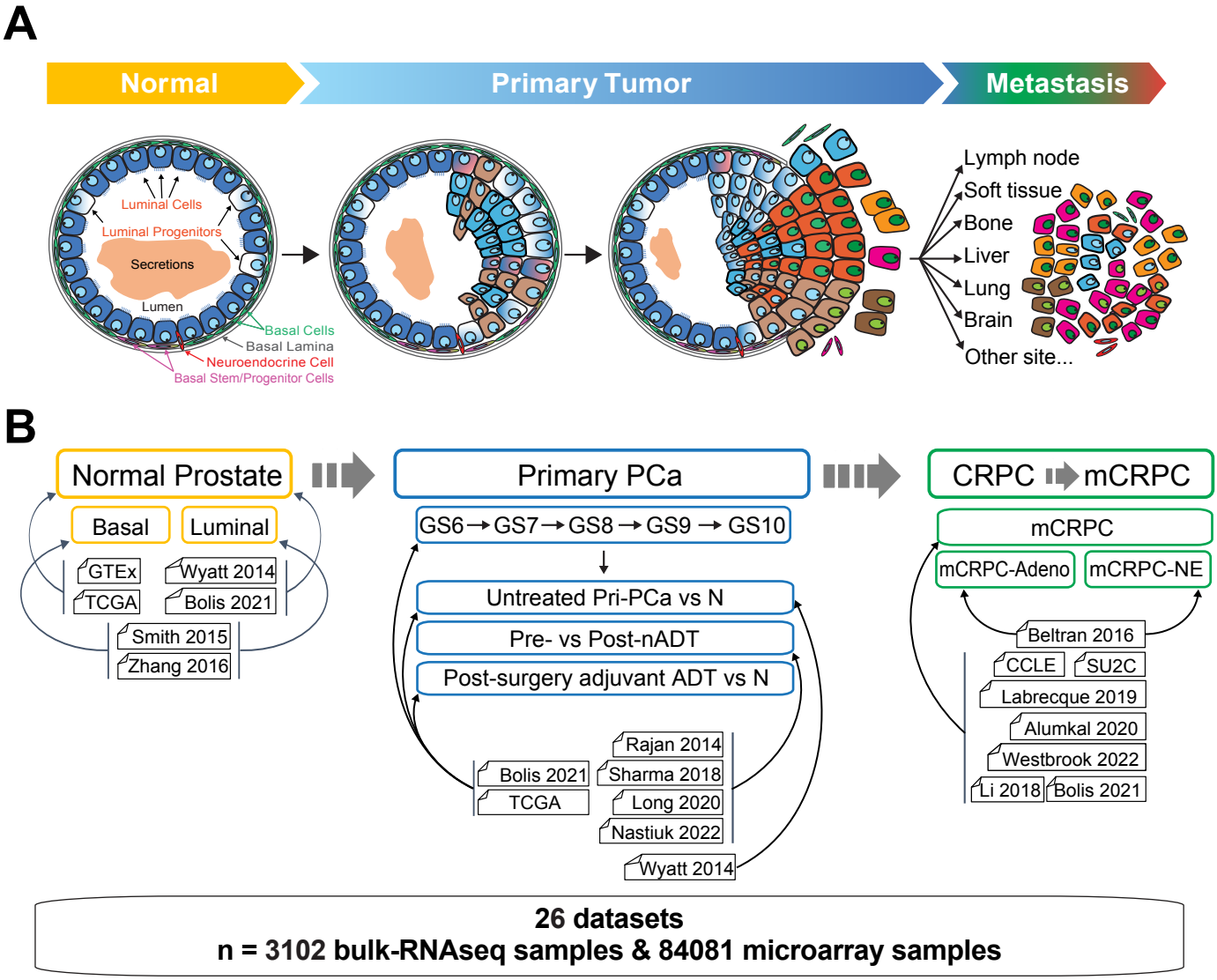
Related to Fig. 5
Supplementary Fig. S7. PCa-Stem signature and gene expression correlate with PCa progression.

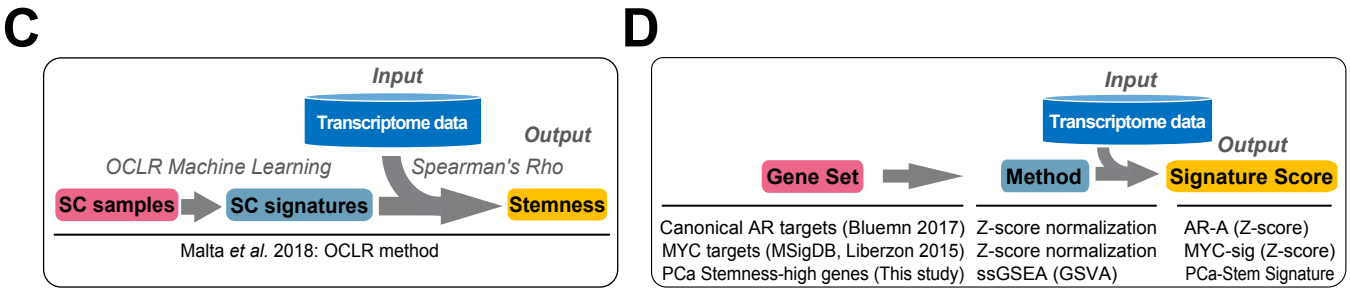
Related to Fig. 6
Supplementary Fig. S8. Unique genomic features of Stemness-high PCa.

Related to Fig. 7
Supplementary Fig. S9. Lack of effect of nADT on MYC activity.

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

Supplementary Figure S1





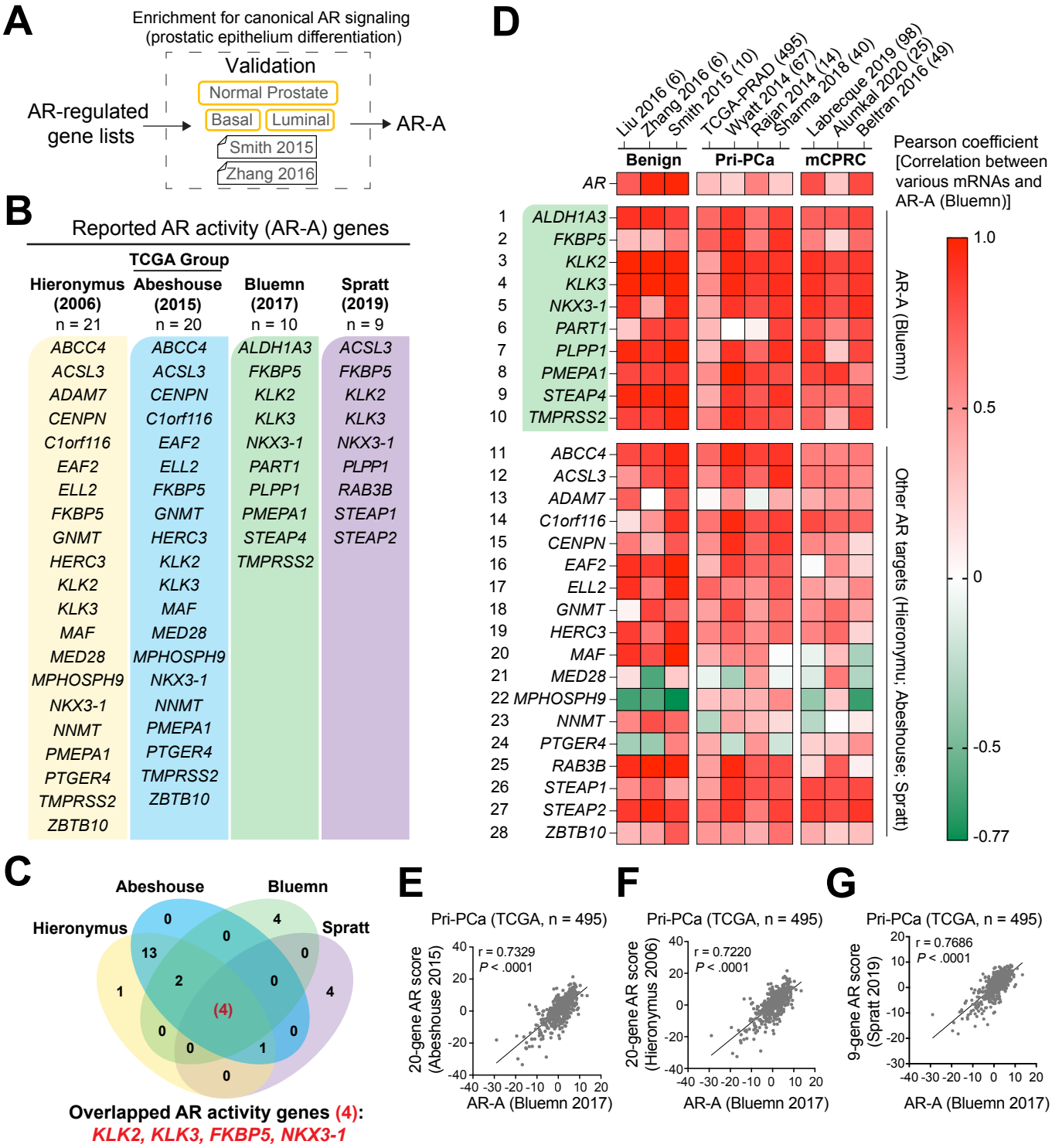
Supplementary Figure S1. Schematic illustration of bioinformatic datasets and signature score pipelines used for comparisons across the spectrum of PCa evolution.

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

Detailed Description of PCa Stages:

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

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



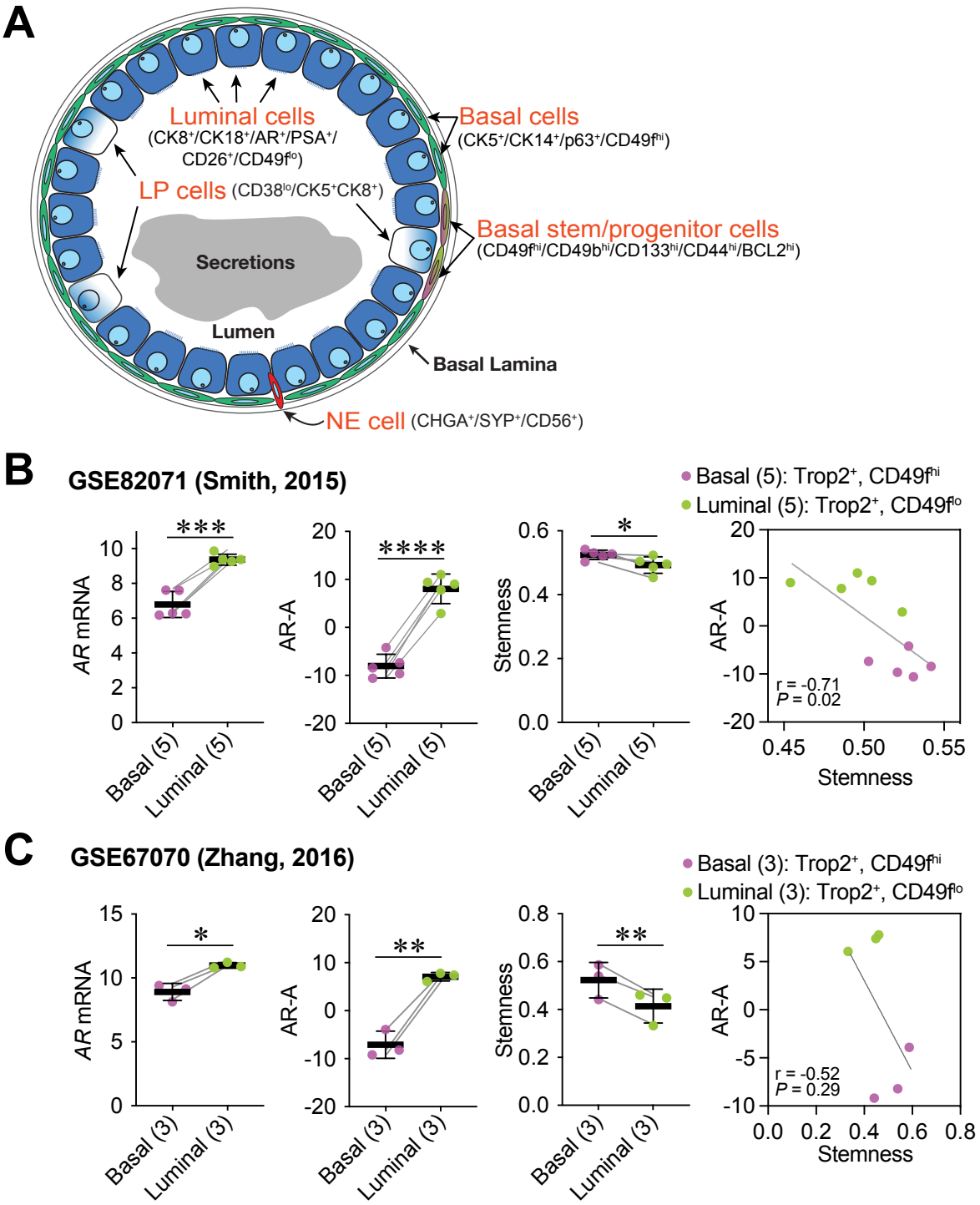
Supplementary Figure S2. Transcriptome-based signature scores for canonical (pro-differentiation) AR-A.

(A) Workflow for validating canonical AR signaling activity (AR-A) signatures from different AR-regulated gene lists.

(B-C) Comparison of AR downstream signaling targets from four studies. The 4 genes commonly shared in all four AR-A signatures are highlighted (C). Genes were listed using updated HGNC symbols (the HUGO gene nomenclature from the Human Genome Organization). Note that in the original publications, some AR-A genes were denoted by their (gene/protein) aliases, as exemplified by *CENPN (BM039), C1orf116 (SARG), KLK3 (PSA), PLPP1 (PPAP2A), and PMEPA1 (TMEPAI)*.

(D) AR-A signature score highly correlates with mRNA expression levels of canonical AR signaling targets in various PCa related cohorts. See Supplementary Table S1 for dataset details.

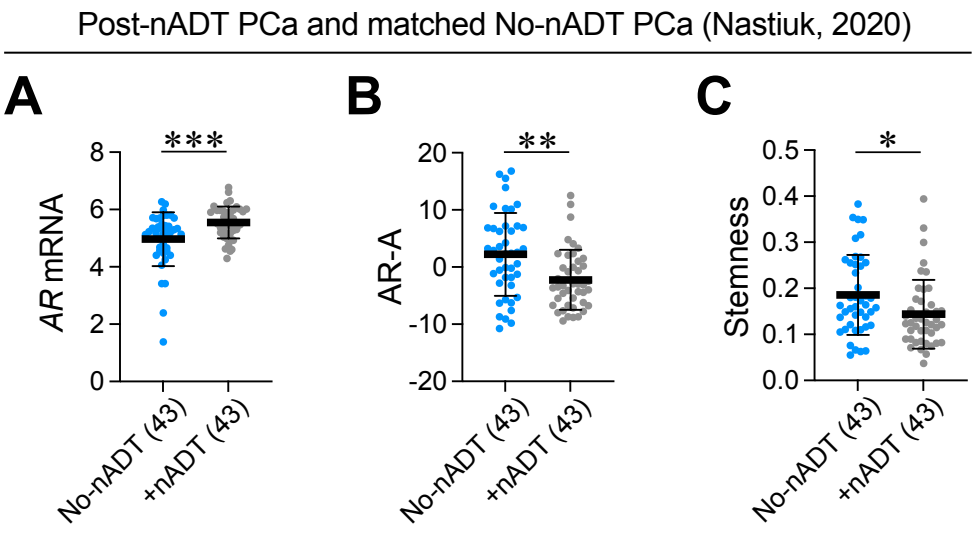
(E-G) Strong positive linear relationships between the Bluemn AR-A and the other three AR-A signatures (Hieronymus 2006; Abeshouse 2015, Spratt 2019) using primary PCa cohort (TCGA-PRAD). r indicates Pearson correlation coefficient.



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

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

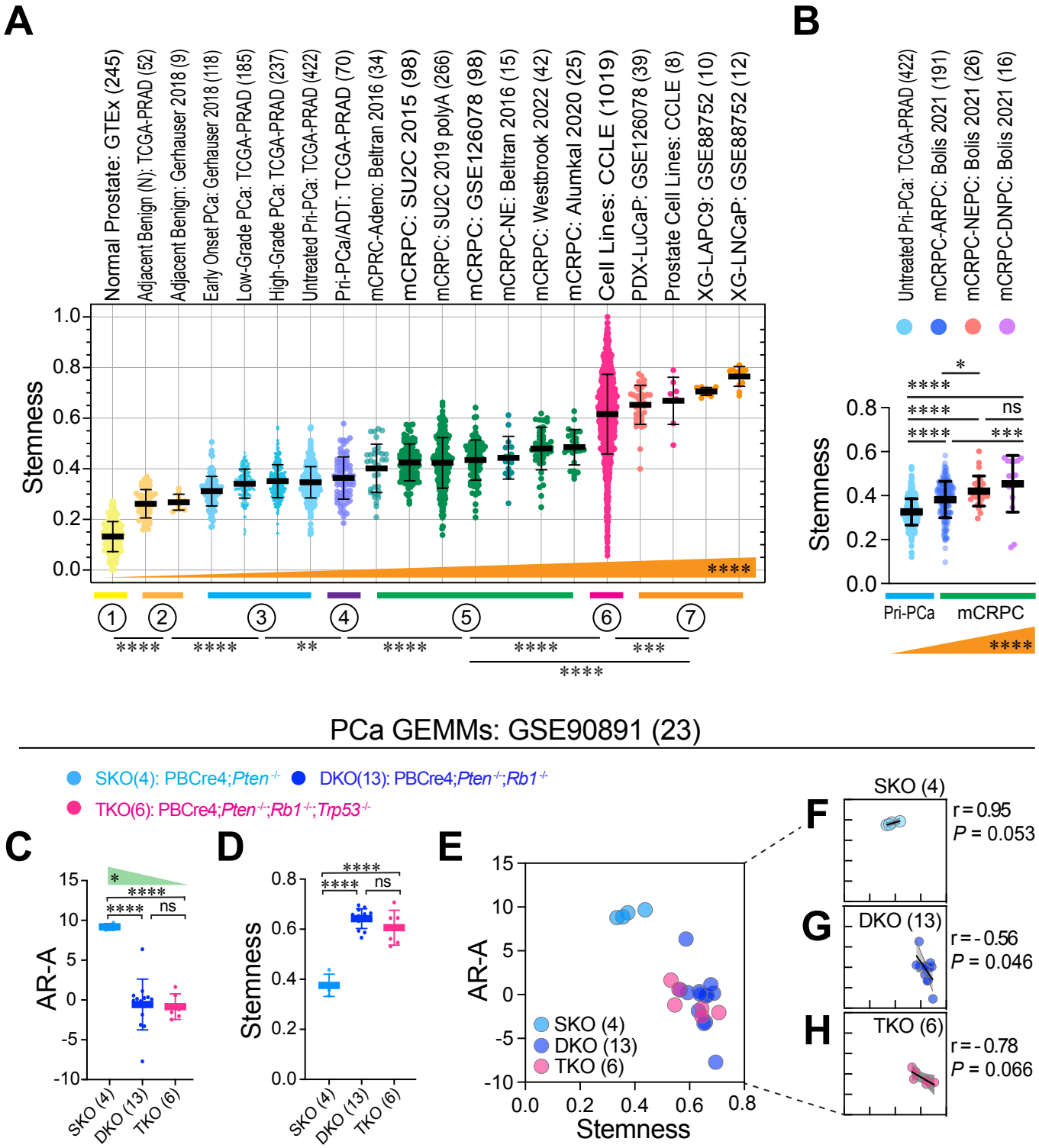
(B-C) Prostate luminal cells possess higher AR-A but lower Stemness as compared to prostate basal cells, and Stemness inversely correlates with AR-A in the context of prostatic basal and luminal cells. Results are shown as mean ± SD. The pairs are linked with a grey line. Significance was calculated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



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

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

Supplementary Figure S5



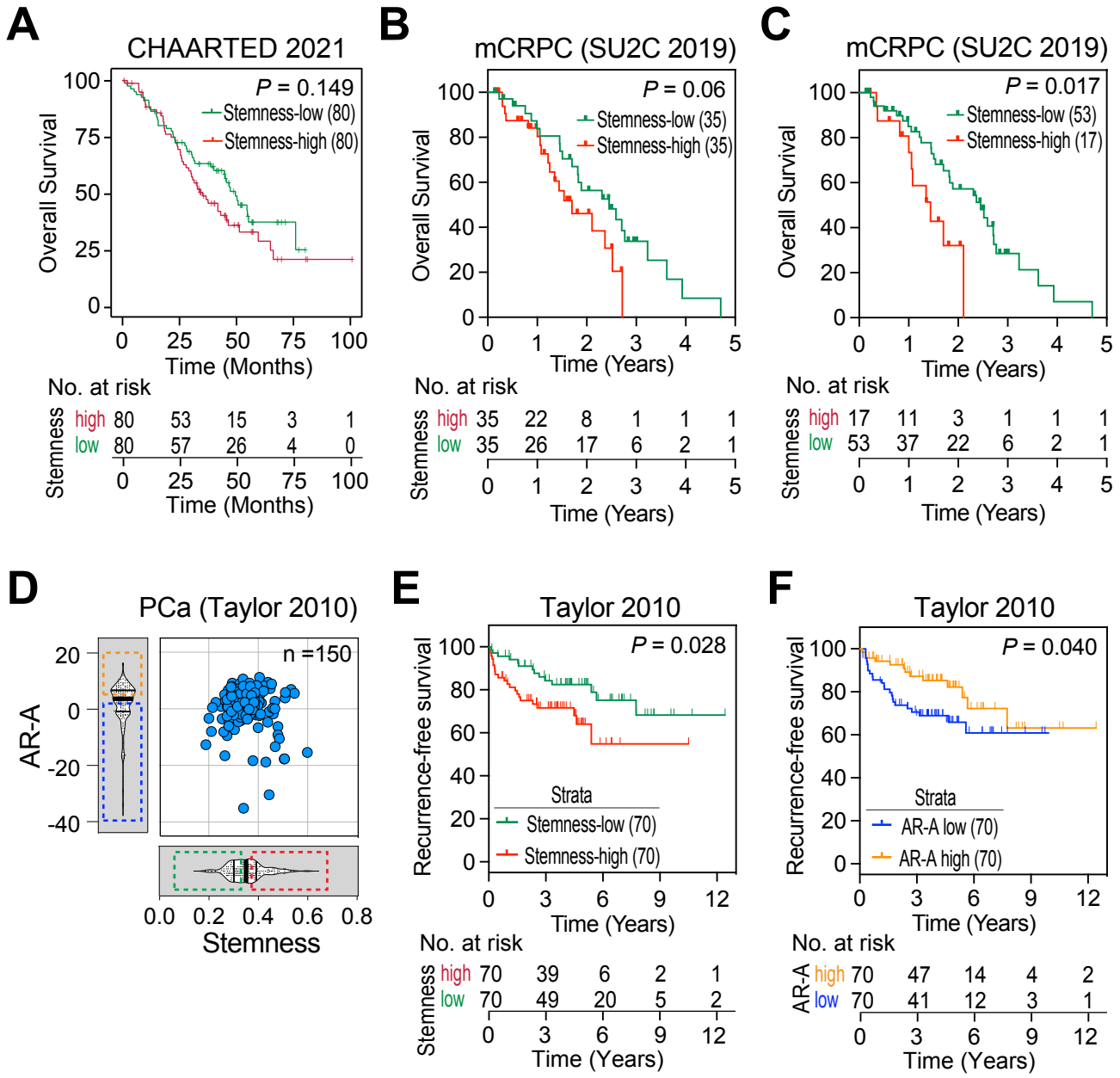
Supplementary Figure S5. **Increased Stemness represents a 'universal' feature of PCa progression and aggressiveness.**

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

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

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

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

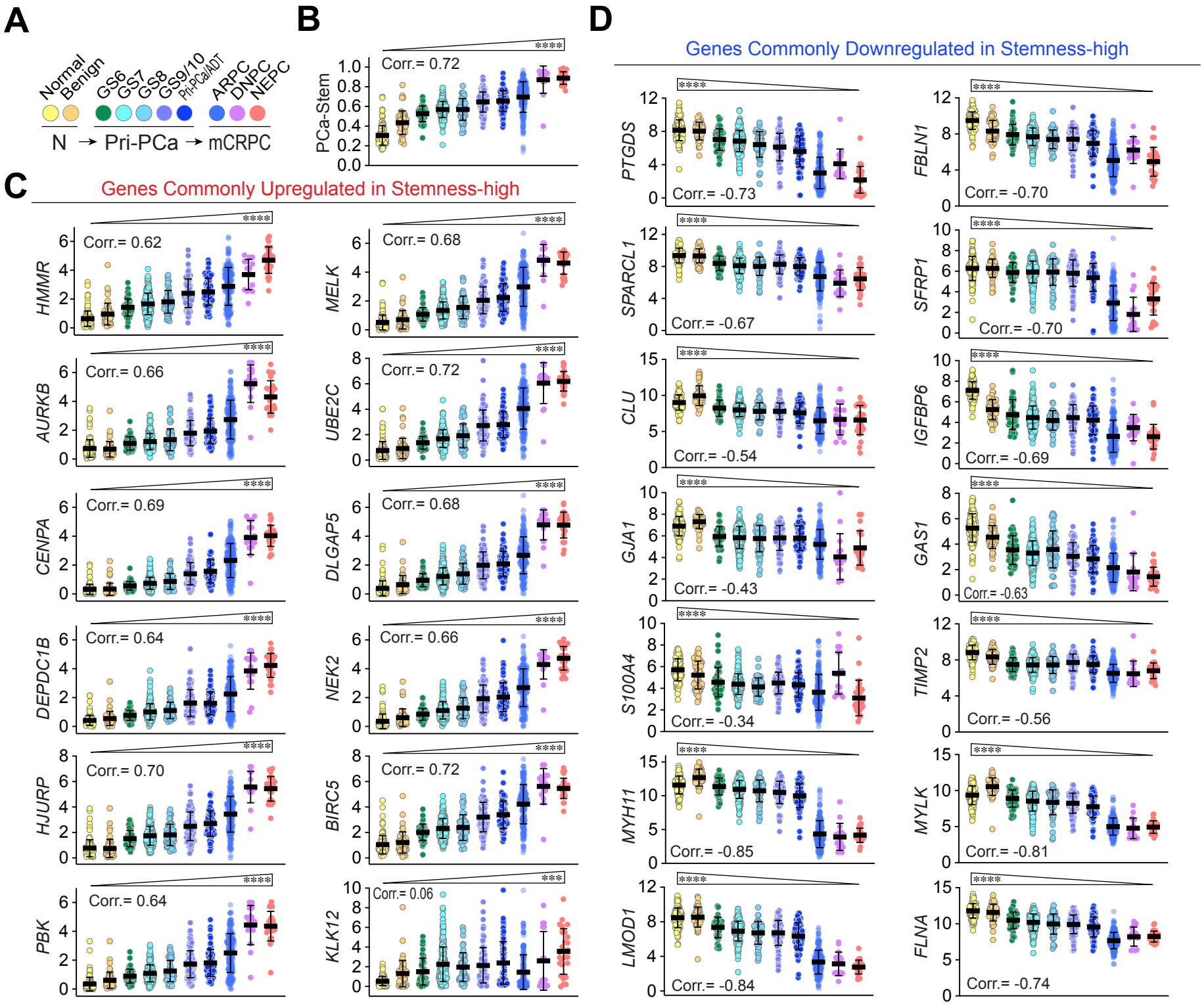


Supplementary Figure S6. High Stemness correlates with poor patient survival.

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

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

Supplementary Figure S7



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

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

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

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

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

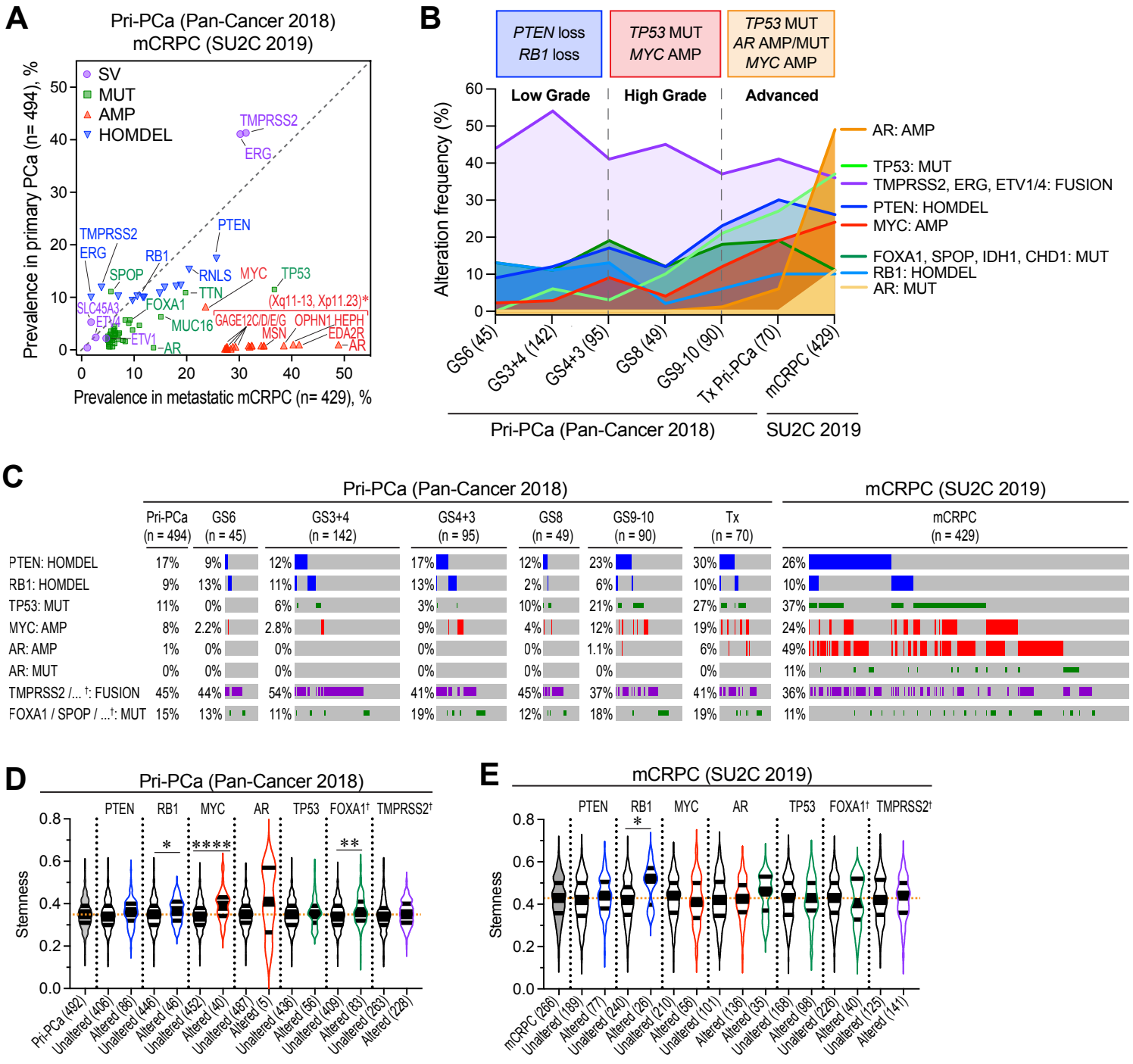
(C) Dot plots showing that the mRNA levels of 11 of the 12 PCa-Stem signature genes (i.e., *HMMR, AURKB, CENPA, DEPDC1B, HJURP, PBK, MELK, UBE2C, DLGAP5, NEK2,* and *BIRC5)* except *KLK12* (see Figure 5D; Supplementary Table S3) exhibited stage-related increases and positively correlated with PCa progression.

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

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

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

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



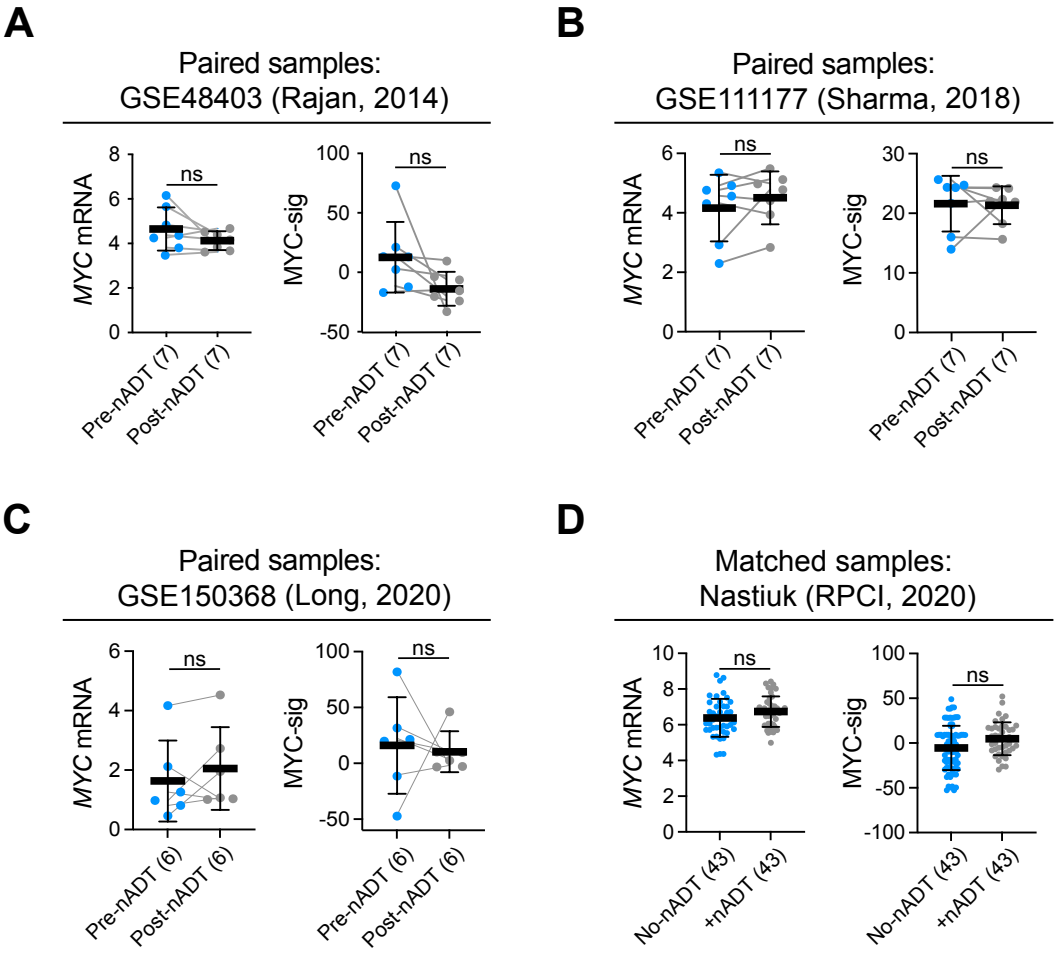
Supplementary Figure S8. Unique genomic features of Stemness-high PCa.

(A) Scatterplot of observed frequencies of genomic alterations in cancer genes reported in Pri-PCa (TCGA) versus mCRPC (SU2C) with data retrieved from cBioPortal. SV (purple), structural variants/fusions; MUT (green), mutation; AMP (red), amplification in DNA copy numbers (i.e., copy number alterations (CNA) ≥ 2); HOMDEL (blue), homozygous deletion or deep deletion.

(B) Genomic alteration frequency plot showing that (early-stage) low-grade PCa have prevalent *PTEN* and *RB1* loss and high-grade PCa have frequent *TP53* MUT and *MYC* AMP whereas advanced PCa have prevalent *TP53* MUT, *MYC* AMP and *AR* AMP/MUT.

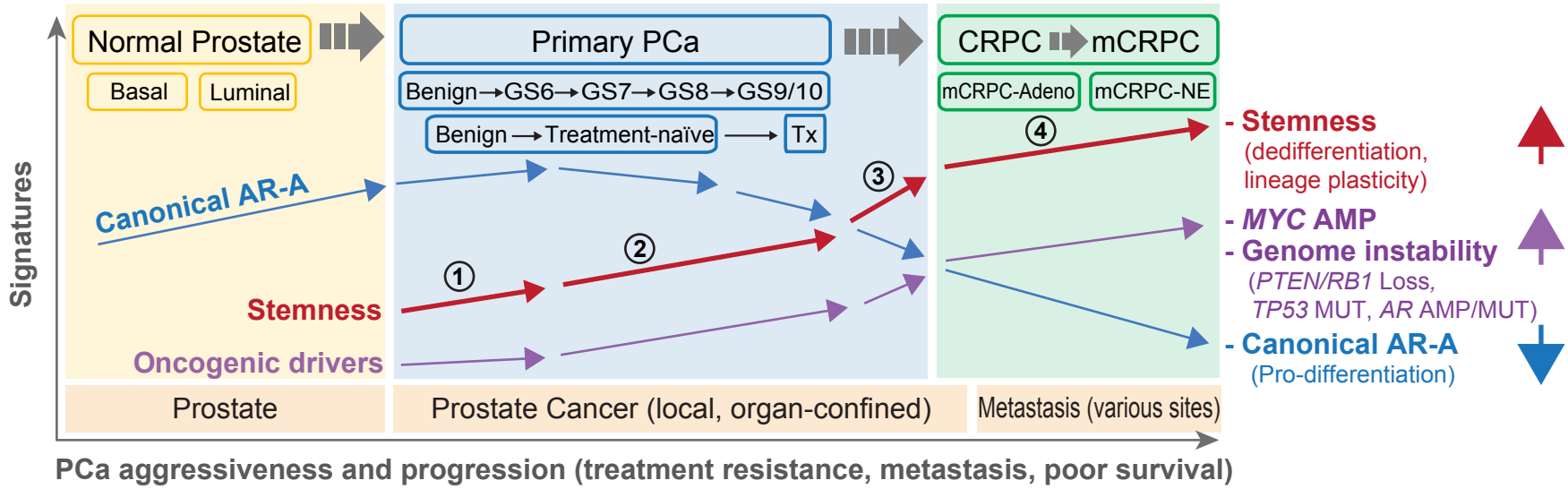
(C) cBioPortal Oncoprint showing the frequencies of major genetic alteration events along the PCa progression continuum.

(D-E) Association of major genomic alterations in PCa with Stemness. Grey, no alterations; Blue, HOMDEL; Green, MUT; Red, AMP; Purple, Fusion. TMPRSS2 /...†: FUSION (in C) or TMPRSS2† (in D-E) stands for combined fusion events in *TMPRSS2*, *ERG*, *ETV1* and *ETV4*; FOXA1 / SPOP /...†: MUT (in C) or FOXA1† (in D-E) represent combined mutation events in *FOXA1*, *SPOP*, *IDH1* and *CHD1*. Note that data presented represent the major mutation subtypes in Pri-PCa (TCGA PRAD cohort) reported in Cancer Genome Atlas Research Network 2015 (Abeshouse *et al.*, 2015, ref. 18) and alterations (mutations, structural variants and copy number) of unknown significance were excluded. Within the violin plots, the center lines represent median values, and box edges are 75th and 25th percentiles. Significance was calculated by Mann-Whitney test (*, $P < 0.05$; **, $P < 0.01$; ****, $P < 0.0001$).



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

Matched comparisons showing no significant differences in both *MYC* mRNA and MYC-sig after neo-adjuvant ADT (Post-nADT) as compared to the matched samples before nADT (Pre-nADT) from the same individuals in **A**, Rajan 2014 cohort; **B**, Sharma 2018 cohort; and **C**, Long 2020 cohort or in **D**, the RPCI Chatta/Nastiuk cohort (unpublished) comprising PCa samples from patients who received nADT (+nADT) and the patients who did not receive ADT (no-ADT) but otherwise matched (with the +nADT patients) for ages and tumor stage and grade.
Sample sizes are indicated in parentheses. Within the plots, results are shown as the mean ± SD. The pairs are linked with a grey line. Significance was calculated by two-tailed paired (A-C) or unpaired (D) Student's *t*-test (ns, not significant; *, $P < 0.05$). See Supplementary Table S2 for MYC-sig gene signature information.



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

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