bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

A cancer stem cell population underlies a multi-lineage phenotype and drug resistance in prostate cancer

Michael L. Beshiri[1], Brian J. Capaldo[1], Ross Lake[1], Anson T. Ku[1], Danielle Burner[1], Caitlin M. Tice[1], Crystal Tran[1], Julianna Kostas[1], Aian Neil Alilin[1], JuanJuan Yin[1], Supreet Agarwal[1], Samantha A. Morris[2,3,4], Fatima H. Karzai[5], Tamara L. Lotan[6], William L. Dahut[5], Adam G. Sowalsky[1], Kathleen Kelly[1*]

[1]Laboratory of Genitourinary Cancer Pathogenesis, Center for Cancer Research, National Cancer Institute, NIH, Bethesda, Maryland.

[2]Department of Developmental Biology, Washington University School of Medicine in St Louis, St Louis, MO, USA.

[3]Department of Genetics, Washington University School of Medicine in St Louis, St Louis, MO, USA.

[4]Center of Regenerative Medicine, Washington University School of Medicine in St Louis, St Louis, MO, USA.

[5]Genitourinary Malignancies Branch, National Cancer Institute, NIH, Bethesda, Maryland.

[6]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA.

*Corresponding author. Email: kellyka@mail.nih.gov

SUMMARY

To resist lineage-dependent therapies, cancer cells adopt a plastic stem-like state, leading to phenotypic heterogeneity. Here we dissect the cellular origins of such heterogeneity in a metastatic castration-resistant prostate cancer (CRPC) patient-derived adenocarcinoma organoid model displaying a range of luminal and neuroendocrine phenotypes and driven by mutations in cell cycle (*CDKN1B*) and epigenetic (*ARID1A,* and *ARID1B)* regulators. As shown by lineage tracing, metastatic tumor heterogeneity originated from distinct subclones of infrequent stem/progenitor cells that each produced a full distribution of differentiated lineage markers, suggesting multiclonal evolution to a relatively stable bipotential state. Single cell ATAC-seq analyses revealed the co-occurrence of transcription factor activities associated with multiple disparate lineages in the stem/progenitors: WNT and RXR stem factors, AR and FOXA1 luminal epithelial drivers, and NR2F1 and ASCL1 neural factors. Inhibition of AR in combination with AURKA but not EZH2 blocked tumor growth. These data provide insight into the origins and dynamics of cancer cell plasticity and stem targeted therapy.

INTRODUCTION

Targeted therapies are designed to attack cancer cells through specific molecular pathways to maximize impact and minimize general toxicity to the patient. Cancer cells can develop resistance to targeted therapies through a process of transdifferentiation where drug-sensitive

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

tumor cells modify their lineage to acquire an alternate cellular identity that is not dependent on the targeted pathway for survival (Beltran et al., 2016; Ping Mu et al., 2017; Quintanal-Villalonga et al., 2020; Sheng Yu Ku et al., 2017). Transition from an adenocarcinoma (AC) to neuroendocrine (NE) lineage is common in multiple epithelial cancers including lung and prostate (Balanis et al., 2019; Quintanal-Villalonga et al., 2020). In metastatic castration-resistant prostate cancer (mCRPC), a decrease in luminal epithelial identity upon treatment with potent AR pathway inhibitors (ARPIs) occurs in ~20% of cases (Bluemn et al., 2017; Quintanal-Villalonga et al., 2020; Rahul Aggarwal et al., 2018). The spectrum of mCRPC phenotypes encompassed by the term lineage-plasticity is broad and thought to exist along a continuum (Labrecque et al., 2019). Common phenotypes include the AR^-/NE^+ small cell neuroendocrine prostate cancer (scNEPC) that is frequently driven by the loss of both *RB1* and *TP53* (Beltran et al., 2016); an AR^-/NE^- double-negative subtype shown to bypass AR-dependence through FGF/MAPK signaling (Bluemn et al., 2017); and an AR^+/NE^+ combined (amphicrine) adenocarcinoma lineage that gains NE features while maintaining AR activity at least in part through downregulation of RE1 silencing transcription factor (REST) activity (Labrecque et al., 2019).

These lineage-plastic subtypes of mCRPC are not static or homogeneous - multiple subpopulations can exist in a patient tumor, yet the dynamic relationship of the subpopulation structure is not well understood (Cejas et al., 2021; Labrecque et al., 2019). Epigenetic mechanisms underlying lineage-plasticity in cancer can instill resistance without genetic clonal selection (Fennell et al., 2021) and minor subpopulations that are not easily detected in bulk may command an outsized role in growth and resistance (Sharma et al., 2010). Distinct hierarchical phenotypes existing at variable frequencies within a tumor often respond differently to the selective pressure of a given treatment (Sharma et al., 2010). Effective therapies against mCRPC that has undergone a lineage-switch are not available and the complex heterogeneity of the phenotypes represents a challenge that is not easily overcome. A detailed characterization and mapping of subpopulation hierarchy and the molecular drivers that govern it is needed to identify cellular points of therapeutic vulnerability.

Questions regarding the state(s) of transition from AC prostate cancer (ACPC) to various forms of lineage-plastic mCRPC remain outstanding. These include 1) the role of cancer stem cells, 2) acquired transcriptional regulators, 3) genetic drivers other than *RB1* and *TP53* loss, and 4) subpopulation heterogeneity in the evolution of plasticity and the response to subsequent therapy. A significant barrier in the field continues to be a lack of representative preclinical models. PDX models of NEPC are available but do not represent the intra-tumoral heterogeneity observed in patients, and are just one genetically-defined subtype (Cejas et al., 2021). AR^+/NE^+ combined lineage patient-derived models present a singular kind of resource to study the dynamic interplay between the two lineage states. The presence of both AC and NE lineages allows the dynamic interplay and cellular states bridging the two to be examined and perturbed. Here we used patient-derived organoid models of AR^+/NE^+ mCRPC harboring mutations in *ARID1A* and *ARID1B*, that capture the phenotypic and genetic heterogeneity observed in the patient tumor. We used these models to identify the existence of bipotential stem-like/progenitor subpopulations underlying growth and phenotypic heterogeneity, and to uncover a molecular vulnerability in the stem cells that can be effectively targeted to block tumor growth.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

RESULTS

Patient-derived organoids with mutations in BAF core complex components demonstrate lineage-plasticity and NE differentiation

We established a set of patient-derived organoid models designated NCI-PC35-1 and NCI-PC35-2 (Beshiri et al., 2018) (PC35-1 and PC35-2) from two spatially-separated needle biopsies of an mCRPC lymph node metastasis that was histologically AC with islands of NE marker-expressing cells (Figures 1A and S1A). There was no evidence of neuroendocrine markers in the primary tumor (Figure S1B). The organoids reflected the pathology of the metastatic tumor with a range of AC and/or NE-marked populations (Figure 1B). Whole-genome sequencing (WGS) phylogenetic analysis revealed that PC35-1 and PC35-2 arose from a common ancestor in the primary tumor featuring genomic mutations with high oncogenic potential: a deep deletion of *CDKN1B,* a frameshift mutation of *ARID1A* and a small deletion in *ARID1B* (Figures 1C, S2A and S2B). ARID1A/B are core components of the BAF complex, and a reduction of ARID1A and ARID1B as we observed (Figure S2A), has been shown to drive carcinogenesis and neural developmental disorders (Hui Shi, 2020; Jung et al., 2017). The *RB1*, *TP53* and *PTEN* loci were intact and expressed (Figure S2C). The two PC35 models shared driver mutations and similar phenotypes but represented divergent clonal populations within the heterogenous tumor. Phylogenetic analysis showed little geographic co-mingling as PC35-1 and PC35-2, which demonstrated 77% and 97% exclusive subclonal genomic variants, respectively. Although we identified an apparent metastasis-specific tandem duplication of the AR enhancer, no additional known driver mutations were found in the metastatic clones, suggesting epigenetic regulation of lineage plasticity, consistent with mutations in the BAF complex (Figures 1C and S2D). Here we have captured a dynamic, multi-lineage phenotype of a patient tumor in an experimentally tractable model, enabling molecular and cellular investigation of naturally-occurring lineage-plasticity.

PC35 organoids are composed of cells from luminal epithelial, neuroendocrine, or stem cell lineages, which display differential proliferation capacity

All possible combinations of AR and NE marker (CHGA) expression status were observed in individual cells: (1) AR-pos/CHGA-low-neg; (2) AR-neg/CHGA-high; (3) AR-pos/CHGA-high; (4) AR-neg/CHGA-low-neg (Figure 1D). Further, mapping AR-activity and NE signature scores of bulk RNA-seq data relative to other ACPC and NEPC models and clinical samples placed PC35-1/2 at the NEPC-adjacent edge of the ACPC cluster, consistent with the pathology and an early evolutionary step in ACPC lineage switching (Figure 1E). Of note, EdU pulse-chase assays showed that $AR^{POS}/CHGA^{Lo/NEG}$ cells proliferated faster than $CHGA^{Hi}$ cells (Figure 2F), yet the balance of AR:CHGA positive cells remained stable over several generations (Figure 2G). PC35-1/2 grew slowly relative to two scNEPC (LuCaPs 145.2 and 173.1) and two ACPC (PC44 and PC155) patient-derived organoid models (Figure S2E). Strikingly, in PC35-1/2 only a fraction of the total cells (~25%) divided in the time it took for the whole population to double (Figure 1H), indicating that a minor population of cells underwent multiple divisions while the majority of cells were not proliferative. By contrast, all cells divided in the scNEPC models, and most but not all cells in the two ACPC models (Figure 1H). These observations in PC35-1/2 imply a constitutive program of multi-lineage commitment coupled with subpopulation-restricted growth.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

To better understand the heterogeneity and subpopulation dynamics of the organoid models, we performed single-cell RNA-sequencing (scRNA-seq). By this method we identified two major clusters (I, II) in PC35-1 and three (I, II, and III) in PC35-2, with a heterogeneous range of lineage phenotypes (Figure 2A). The heterogeneity of the major clusters was delineated into subclusters designated with Greek letters (Figure 2A) that could be distinguished by phenotype based on AR, NE, and proliferation (PRLF) signature scores (Figures 2B–2D and S3A), lineage marker expression (Figures S3B and S3C) and cell cycle profile (Figure S3D). Additionally, we performed RNA-velocity analysis (La Manno et al., 2018) on the data to infer temporal states of differentiation (Figure S4). Based on the overlap with the signature scores, lineage marker expression, RNA-velocity vector patterns and cell cycle state, the subclusters from (Figure 2A) were categorized with respect to lineage and differentiation status. α clusters showed features of stem/progenitor cells (s/p), including a high proliferation score, strong enrichment of proliferation and stem marker gene expression (TK1, EZH2, AURKA, HES4), long RNA velocity vectors showing a multidirectional pattern, and a prominent G2/M transcriptional profile. α' in PC35-2 has most features of the stem/progenitors that begin to decline at the distal end proximal to the β cluster, likely as they transition to a more differentiated state. β and δ clusters were the most adenocarcinoma-like, featuring a high AR signature score, low NE score, strong enrichment of ACPC marker gene expression (AR, KLK3), no enrichment of the G2/M state, a low proliferation score compared to the progenitor group, and a more uniform direction of RNA velocity vectors. The γ subclusters showed the most neuroendocrine-like differentiation with a high NE score, a low AR signature score, strong enrichment of NE marker gene expression (CHGA, SCG2), no enrichment of the G2/M state, a low proliferation score, and a uniform direction of short RNA velocity vectors. The ε subcluster of PC35-1was heterogeneous for all features and did not have a clear lineage. Within the PC35 models, PC35-1 major cluster II was an exception as it lacked a high degree of heterogeneity and scored uniformly NE-high/AR-low and PRLF low. In contrast to PC35-1/2, the ACPC model PC44 and NEPC model LuCaP 145.2 were homogenous in their respective lineages and PRLF scores, indicating a more equal proliferative potential for all/most cells in the population (Figures 2B-2D). In summary, heterogeneity of PC35-1/2 captured from the patient tumor is spread across and within distinct major clusters with associated proliferative subpopulations, demonstrating both inter- and intra-clonal lineage-plasticity.

To validate the scRNA-seq analysis we performed quantitative single-molecule RNA-FISH using selected markers on PC35-1/2 organoid-derived cells. In agreement with the scRNA-seq results, we found that ACPC lineage marker-positive cells were largely distinct from NEPC marker-positive cells (Figure S5A). Where cells were double positive, they tended to show reduced expression of one or both markers. EZH2 and TK1 marked the same population of EdU-positive, dividing cells (Figures S5A-S5C). Double staining for mRNA and protein of selected lineage markers confirmed a strong positive correlation between mRNA and protein (Figure S5D). The combined results of the scRNA-seq and RNA-FISH analyses allowed us to finely resolve and map the phenotypes. The data showed that PC35-1 and PC35-2 were composed of subpopulations that were generally similar in their transcriptional profiles, but still maintained discernably unique identities.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Simultaneous tracking of lineage and clonal identity with single-cell resolution identifies self-renewing stem-like subpopulations with differentiation potential

To address the dynamic plasticity and the hierarchical structure of the subpopulations, we used the "CellTagging" method of combinatorial indexing by expressed barcodes read-out by scRNA-seq, to simultaneously track cellular origin and phenotypic identity within growing organoids (Biddy et al., 2018). We analyzed single time-point experiments of different durations after tagging to identify clonal expansion of sibling cells, allowing determination of the clonal relationships for lineage-marked populations. A separation of four weeks between tagging and harvest captured uniquely tagged clones in all major clusters. For both PC35-1 and PC35-2 all tagged clonal sibling cells that were associated with any given major cluster (I, II, III) were exclusive to that cluster, verifying the clonality of each cluster. The location of tagged sibling cells is graphically depicted on the UMAPs for PC35-1/2 with black lines connecting clones (Figures 3A, 3B). Within each major cluster a disproportionate number of the tagged clones were located entirely within the α stem/progenitor subclusters (Figures 3A and 3B), indicating that α was self-renewing. However, tagged clones in subclusters-α also spanned across the differentiated subclusters within the same major cluster, demonstrating differentiation. The more differentiated states of the β and γ subclusters showed reduced internal cellular replication, suggesting that they resulted from the differentiation of subclusters-α (Figures 3A and 3B). These data demonstrate in a near patient biopsy-derived sample the existence of cancer stem/progenitor cells (subclusters-α), which maintain distinct clonal populations (major clusters) of dynamically differentiating heterogeneous ACPC and NEPC phenotypes.

In contrast to the clonal dynamics observed in the PC35 models, the clones captured in LuCaP 145.2 NEPC organoids were indicative of a widely proliferative population. We detected numerous clones both within and across most clusters that did not show directionality of expansion (Figure 3C). The ACPC model PC44 exhibited a disproportionately high number of clones associated with two small clusters; however, unlike the PC35 models, there was no evidence of self-renewal within those two clusters and like LuCaP 145.2, sibling cells were widely dispersed within and across nearly all clusters (Figure 3D). Therefore, while we cannot rule out the existence of progenitor populations in these models, division is not restricted to a specific subpopulation.

Unique combinations of transcription factor activities are linked to divergent phenotypes

One possible explanation for the discrete clonality of the major clusters in the PC35 models is that cluster-specific genetic events led to distinct phenotypes, although we were unable to identify subclonal driver mutations by WGS analyses. We analyzed our scRNA-seq data using CopyKAT to identify clonal subpopulations based on genomic copy number variation (CNV) and associated this genomic substructure with the phenotypically-defined major clusters (Gao et al., 2021). We found a combination of contributions to the different phenotypes of the clusters: those which were independent of CNV clonal patterns (PC35-1 UMAP clusters I and II, and PC35-2 cluster I) and those attributed to unique or closely related clonal genotypes (PC35-2 UMAP clusters II and III) (Figures 4A, S6A and S6B). These data suggest that there are fixed differentiation patterns for pre-existing clonal populations in addition to common pathways of lineage differentiation. Genetic clones did not show lineage bias toward phenotypic subclusters (α – ε), confirming ongoing dynamic, heterogeneous differentiation (Figures S6A and S6B).

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

To investigate the role of epigenetic regulation on the phenotype of the PC35 models, we performed single-cell ATAC-seq. Clustering by genome-wide chromatin accessibility yielded three clusters (***1, 2, 3***) in both PC35-1 and PC35-2 (Figure 4B). To look for transcription factors (TFs) that may be responsible for the differing phenotypes among the clusters, we performed an analysis of inferred TF activity. Clusters-***3*** in both models were distinguished as the most neuroendocrine-like, exhibiting a relative absence of REST activity and high activity scores for TFs such as NRF1, HES4 and ONECUT2 (Figures 4C and 4D), similar to previously described NE models (Balanis et al., 2019). Clusters ***1*** and ***2*** in PC35-1/2 demonstrated unique but highly overlapping combinations of transcription factors contributing to stem cell, luminal epithelial, and neural phenotypes. Additionally, Clusters ***1*** and ***2*** of both models could be partitioned into two pairs of subclusters *1.1*, *1.2* and *2.1*, *2.2* (Figure 4B). Inferred TF activities in subclusters *1.1* and *2.1,* were consistent with a stem-like phenotype and included WNT pathway effectors such as *TCF7* and *TCF7L2*, and retinoid X receptors (Figures 4B-4D) and co-occurred with TFs determining luminal epithelial (FOXA1, AR, NR3C1) and neural (NR2F1, ASCL1) lineages. Considering the adenocarcinoma origin of the tumor, these data suggest the gain of stem cell and NE lineage determining TFs while some luminal TFs remain active. There were no remarkable TF activities gained in Clusters *1.2* and *2.2* compared to the stemlike *1.1* and *2.1.* On the contrary, differentiation was mostly associated with reduced TF activity found in the stem-like clusters (Figures 4B-4D); however, it is possible that the plasticity-associated heterogeneity across the differentiating population obscured TF patterns. These data denote a model of plasticity whereby variable activity of a TF program across stem-like/progenitor clones resulted in distinct cellular phenotypes that share to differing degrees features of ACPC and NEPC lineages, in addition to a more complete switch to a neuroendocrine program in a small subpopulation of cells. Thus, it appears that stochastic epigenetic processes acting on cancer stem cells contribute to lineage differentiation.

Targeting both AR pathway dependent and independent compartments of the stem/progenitor subpopulations inhibits *in vitro* and *in vivo* tumor growth

The existence of multiple identifiable clones propagated by stem/progenitor cells enables analysis into the heterogeneity of intratumoral resistance mechanisms as well as cancer stem cell targeted therapeutics. Although CRPC implies a loss of AR-targeted responsiveness, the presence of potentially disparate resistance mechanisms across multiple clones presents an important clinical challenge when discontinuing AR suppression therapy. We treated PC35-1 and PC35-2 organoids with enzalutamide, quantified cell numbers, and found a partial response in both models, concordant with the notion of a subpopulation-specific dependence on AR signaling (Figure 5A). PC35-1 showed a greater than two-fold reduction after treatment, while PC35-2 showed a less than 30% decrease. We then performed RNA-FISH in combination with EdU to quantify subpopulation-specific changes due to enzalutamide treatment (Figures 5B-5D). Congruent with the different overall response observed in bulk, we found that enzalutamide caused a >10-fold reduction to proliferating $AR^{POS}EdU^{POS}$ cells in PC35-1 while the same population in PC35-2 showed only a small decrease (Figure 5C). The SCG2-positive, neuroendocrine-like, populations in both PC35-1/2 were insensitive to enzalutamide (Figure 5D). To determine whether resistance mapped to a specific subcluster of AR-positive cells, we identified *MAP3K5*/*ASK1* as a top differentially expressed gene marking cluster III of PC35-2 (Figure 5E). Proliferating $MAP3K5^{POS}$ and AR^{POS} $MAP3K5^{POS}$ (double-positive) cells were resistant to enzalutamide, but AR^{POS} $MAP3K5^{NEG}$ cells were depleted two-fold after treatment

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

(Figure 5F). This result was unexpected given that PC35-2 cluster III had a strong AR signature score. MAP3K5 is an upstream regulator of *NR3C1*, which was co-enriched in cluster III (Figure 5G) (Perez Kerkvliet et al., 2020). *NR3C1* expression is a well-established ARPI resistance mechanism that leads to expression of some AR-regulated genes (Arora et al., 2013). These data demonstrate plasticity-mediated, clonal variability, selected within a patient, leading to partial ARPI resistance within a population of mCRPC tumor cells.

Although AR^{POS} cells made up a proportion of the EdU^{POS} progenitor population, $\geq$ 50% of stem/progenitor cells were AR-negative (Figure 5C, -Enza columns). To specifically address the stem/progenitor population, we identified multiple druggable targets as highly enriched in the stem-like subclusters-α of the PC35 models, including EZH2, AURKA, and the Notch pathway (Figures S3B and S3C) and targeted them with CPI-1205, Alisertib, or Compound E, (EZH2i, AURKAi and Notchi respectively). For comparison, we included the chemotherapeutic agent carboplatin, which is used as a late line of therapy in mCRPC. After initial dose response determinations by two-week assays, we observed a heterogeneous response potentially indicative of a subpopulation-specific drug sensitivity/resistance (Figure S7A). We then treated the PC35 organoids with AURKAi, EZH2i, Notchi, carboplatin, or DMSO for six weeks with concentrations that were selected from the middle of the plateau of the dose-response curves. In both organoid models, the AURKAi caused a nearly 10-fold decrease in cell number compared to DMSO while the other drug conditions resulted in only minor reductions (Figure 6A). We tracked the effect of AURKAi, EZH2i, and carboplatin relative to DMSO with single-cell resolution using RNA-FISH/EdU combined assays. Subpopulations were identified by marker gene expression: AR to mark ACPC lineage; SCG2 to mark NEPC lineage; TK1, EZH2 and AURKA and EdU incorporation to mark stem-like/progenitors. We found that the AURKAi specifically depleted the stem-like/progenitor subpopulation while carboplatin had no effect (Figures 6B and 6C). The percentage of AR^{POS} cells within the EdU^{POS} population decreased from 40% to about 15% (Figure S7B), demonstrating significant sensitivity of the AR^{POS} stem cell population. We hypothesize that AURKAi-resistant AR^{POS} cells may represent either a more differentiated transit amplifying AR^{POS} population or partial intrinsic resistance.

Although EZH2 has been shown to regulate a transcriptional program driving a lineage-switch away from differentiated adenocarcinoma in *RB1*$^{-/-}$,*TP53*$^{-/-}$ models (Davies et al., 2021; Ping Mu et al., 2017; Sheng Yu Ku et al., 2017), we observed only an insignificant increase in AR-positivity in both PC35-1 and PC35-2 in the EZH2i condition (Figure S7C), and we did not see upregulation of EZH2 or phosho-EZH2 upon enzalutamide treatment (Figure S7D), suggesting context-dependence for EZH2-driven mechanisms observed in RB1/TP53 loss models. Together these results indicate that the stem-like/progenitor subpopulation can be directly targeted by AURKAi to block growth and imply the existence of a residual AR^{POS} proliferative population with potential sensitivity to AR inhibition.

To evaluate how effective inhibition of AURKA and/or AR is at blocking tumor growth *in vivo*, we treated PC35-1 organoid-derived xenograft tumors for nine weeks with either alisertib (half standard dose), castration, alisertib combined with castration, or vehicle. Castration or the low dose of alisertib alone caused a 50% decrease of tumor growth that was not statistically significant. However, the combination treatment rapidly and dramatically blocked tumor growth (Figure 6D). In week-nine tumors, castration caused a strong increase in cytoplasmic and

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

decrease in nuclear AR, as well as increased expression of the NE marker, synaptophysin (Figure S8A). Consistent with the effects on tumor growth, the strong BrdU incorporation observed in the control was decreased in all the treated conditions, reaching the lowest level in the combination treatment (Figures S8A and S8B). These results demonstrate that the subpopulation-specific vulnerabilities that we identified in patient-derived organoids can be exploited to yield impactful results on tumor growth *in vivo*.

DISCUSSION

Prostate cancer is dependent upon AR signaling for growth and survival (Huggins, 1972). Tumors are exquisitely responsive to AR-inhibition upon initial treatment; however, relapse in the form of castration-resistant disease is incurable despite a continued dependence on the AR pathway (Attard et al., 2008; Scher and Sawyers, 2005). Second generation AR pathway inhibitors such as enzalutamide and abiraterone used to treat CRPC effectively block AR signaling, but still ultimately fail (Attard et al., 2008; Scher et al., 2010). Resistance frequently occurs through mechanisms that bypass AR signaling, including lineage-switching and alternative receptor activity such as the glucocorticoid receptor (GR) coded by *NR3C1*(Buttigliero et al., 2015). A full appreciation of the molecular and cellular mechanisms contributing to lineage switching and resistance has been hampered by a lack of tractable, preclinical models representing the phenotypic complexity of tumors.

In a set of patient-derived organoid models of the AR^+/NE^+ phenotype featuring mutations in *ARID1A* and *ARID1B* subunits of the BAF chromatin-remodeling complex, we report the existence of clonally-distinct cancer stem/progenitor subpopulations as the source of growth and phenotypic heterogeneity. The stem/progenitor subpopulations demonstrated co-occurring transcription factor activities associated with both luminal epithelial and neuroendocrine lineages. Consistent with this duality of TF activity, the heterogeneity observed at the single-cell level exhibited all permutations of luminal epithelial and NE lineage marker expression. Surprisingly, the least proliferative population was the most neuroendocrine-like. This finding is contrary to the increased growth rate of AR^- NEPC driven by *RB1* and *TP53* loss, but consistent with less aggressive NE tumors including gastroenteropancreatic NE neoplasms, breast cancer with neuroendocrine differentiation, and pulmonary NE carcinoids that are frequently driven by mutations in *ARID1A* (Cros et al., 2021; Marchio et al., 2017; Puccini et al., 2020).

Diversified and labile transcriptional programs within a heterogenous tumor cell population can rapidly confer clonal fitness in the face of therapeutic pressure (Bolis et al., 2021; Davies et al., 2021; Fennell et al., 2021; Taavitsainen et al., 2021). In our models clonally-determined lineage distributions were partially explained by pre-existing genomic alterations. In addition, we observed that highly similar but variable patterns of TF activity across cancer stem cell clones, in the absence of additional identifiable driver mutations, produced overlapping spectra of lineage phenotypes. This observation highlights the underlying complexity driving phenotypic heterogeneity in tumors and suggests contributions from both genetic and epigenetic evolution.

Although EZH2 has been implicated as an epigenetic factor mediating the loss of plasticity-associated AR independence using various genetic models of the LNCaP cell line and in RB1 deficient mouse models, this near patient model demonstrated relatively little phenotypic and no

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

growth-related response to EZH2 inhibition. This finding suggests that a fuller understanding of context-dependent EZH2 activity is needed to use EZH2 inhibitors selectively in patients.

In addition to lineage-switching, we found ARPI resistance at the subpopulation level mediated by other known mechanisms, such as high expression of *NR3C1*. This observation indicates that multiple different paths to resistance are employed by cancer cells within the same tumor, underscoring the challenges in the development of curative treatments.

The existence of a stem-like/progenitor subpopulation as the seedbed of growth in a tumor would have great potential as a point at which to direct therapeutic intervention. We found Aurora Kinase A, a regulator of mitotic progression, stem cell self-renewal, and asymmetric division (David M Glover, 1995; Eterno et al., 2016; Wang et al., 2019), to be expressed and restricted to the stem/progenitors. Importantly, inhibition of AURKA in the organoid models caused a strong and specific depletion of the stem/progenitor pool that blocked growth of the entire heterogeneous population.

We also found that targeting self-renewal within the cancer stem cell population blocked tumor growth *in vivo*. A half dose regimen of alisertib, once-daily (see methods) instead of the twice-daily standard, reduced growth of the tumor by 50%. This low dose combined with castration resulted in 90% tumor growth inhibition consistent with the continued expression of AR-dependent target genes in the stem/progenitor cells. These data suggest that the clinical application of ARPI combined with alisertib may be useful for treating mCRPC displaying lineage plasticity.

Here we demonstrate the existence of a minor stem/progenitor subpopulation representing a singular vulnerability within the larger heterogeneous tumor cell population. These data suggest that potentially responsive tumors may be overlooked because key subpopulations are obscured in the heterogeneity of the tumor. Therefore, an effort to identify important minor populations, as we have shown here, may better inform treatment decisions by identifying responsive tumors that would otherwise appear to be poor candidates.

REFERENCES

Arora, V.K., Schenkein, E., Murali, R., Subudhi, S.K., Wongvipat, J., Balbas, M.D., Shah, N., Cai, L., Efstathiou, E., Logothetis, C.*, et al.* (2013). Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor blockade. Cell *155*, 1309-1322.

Attard, G., Reid, A.H., Yap, T.A., Raynaud, F., Dowsett, M., Settatree, S., Barrett, M., Parker, C., Martins, V., Folkerd, E.*, et al.* (2008). Phase I clinical trial of a selective inhibitor of CYP17, abiraterone acetate, confirms that castration-resistant prostate cancer commonly remains hormone driven. J Clin Oncol *26*, 4563-4571.

Balanis, N.G., Sheu, K.M., Esedebe, F.N., Patel, S.J., Smith, B.A., Park, J.W., Alhani, S., Gomperts, B.N., Huang, J., Witte, O.N.*, et al.* (2019). Pan-cancer Convergence to a Small-Cell Neuroendocrine Phenotype that Shares Susceptibilities with Hematological Malignancies. Cancer Cell *36*, 17-34 e17.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Beltran, H., Prandi, D., Mosquera, J.M., Benelli, M., Puca, L., Cyrta, J., Marotz, C., Giannopoulou, E., Chakravarthi, B.V., Varambally, S.*, et al.* (2016). Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. Nat Med *22*, 298-305.

Beshiri, M.L., Tice, C.M., Tran, C., Nguyen, H.M., Sowalsky, A.G., Agarwal, S., Jansson, K.H., Yang, Q., McGowen, K.M., Yin, J.*, et al.* (2018). A PDX/Organoid Biobank of Advanced Prostate Cancers Captures Genomic and Phenotypic Heterogeneity for Disease Modeling and Therapeutic Screening. Clin Cancer Res *24*, 4332-4345.

Biddy, B.A., Kong, W., Kamimoto, K., Guo, C., Waye, S.E., Sun, T., and Morris, S.A. (2018). Single-cell mapping of lineage and identity in direct reprogramming. Nature.

Bluemn, E.G., Coleman, I.M., Lucas, J.M., Coleman, R.T., Hernandez-Lopez, S., Tharakan, R., Bianchi-Frias, D., Dumpit, R.F., Kaipainen, A., Corella, A.N.*, et al.* (2017). Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. Cancer Cell *32*, 474-489 e476.

Bolis, M., Bossi, D., Vallerga, A., Ceserani, V., Cavalli, M., Impellizzieri, D., Di Rito, L., Zoni, E., Mosole, S., Elia, A.R.*, et al.* (2021). Dynamic prostate cancer transcriptome analysis delineates the trajectory to disease progression. Nat Commun *12*, 7033.

Buttigliero, C., Tucci, M., Bertaglia, V., Vignani, F., Bironzo, P., Di Maio, M., and Scagliotti, G.V. (2015). Understanding and overcoming the mechanisms of primary and acquired resistance to abiraterone and enzalutamide in castration resistant prostate cancer. Cancer Treat Rev *41*, 884-892.

Cejas, P., Xie, Y., Font-Tello, A., Lim, K., Syamala, S., Qiu, X., Tewari, A.K., Shah, N., Nguyen, H.M., Patel, R.A.*, et al.* (2021). Subtype heterogeneity and epigenetic convergence in neuroendocrine prostate cancer. Nat Commun *12*, 5775.

Cros, J., Theou-Anton, N., Gounant, V., Nicolle, R., Reyes, C., Humez, S., Hescot, S., Thomas de Montpreville, V., Guyetant, S., Scoazec, J.Y.*, et al.* (2021). Specific Genomic Alterations in High-Grade Pulmonary Neuroendocrine Tumours with Carcinoid Morphology. Neuroendocrinology *111*, 158-169.

David M Glover, M.H.L., Doris A McLean, Huw Parry (1995). Mutations in aurora prevent centrosome separation leading to the formation of monopolar spindles. Cell *81*, 95-105.

Davies, A., Nouruzi, S., Ganguli, D., Namekawa, T., Thaper, D., Linder, S., Karaoglanoglu, F., Omur, M.E., Kim, S., Kobelev, M.*, et al.* (2021). An androgen receptor switch underlies lineage infidelity in treatment-resistant prostate cancer. Nat Cell Biol.

Eterno, V., Zambelli, A., Villani, L., Tuscano, A., Manera, S., Spitaleri, A., Pavesi, L., and Amato, A. (2016). AurkA controls self-renewal of breast cancer-initiating cells promoting wnt3a stabilization through suppression of miR-128. Scientific reports *6*, 28436.

Fennell, K.A., Vassiliadis, D., Lam, E.Y.N., Martelotto, L.G., Balic, J.J., Hollizeck, S., Weber, T.S., Semple, T., Wang, Q., Miles, D.C.*, et al.* (2021). Non-genetic determinants of malignant clonal fitness at single-cell resolution. Nature.

Gao, R., Bai, S., Henderson, Y.C., Lin, Y., Schalck, A., Yan, Y., Kumar, T., Hu, M., Sei, E., Davis, A.*, et al.* (2021). Delineating copy number and clonal substructure in human tumors from single-cell transcriptomes. Nat Biotechnol *39*, 599-608.

Huggins, C.a.H., C.V. (1972). Studies on prostatic cancer: I. The effect of castration, of estrogen and of androgen injection on serum phosphatases in metastatic carcinoma of the prostate. CA: A Cancer Journal for Clinicians *22*, 232-240.

Hui Shi, T.T., Brian J. Abraham, Adam D. Durbin, Mark W. Zimmerman, Cigall Kadoch, A. Thomas Look (2020). ARID1A loss in neuroblastoma promotes the adrenergic-to-mesenchymal transition by regulating enhancer-mediated gene expression. SCIENCE ADVANCES *6*, 1-12.

Jung, E.M., Moffat, J.J., Liu, J., Dravid, S.M., Gurumurthy, C.B., and Kim, W.Y. (2017). Arid1b haploinsufficiency disrupts cortical interneuron development and mouse behavior. Nat Neurosci *20*, 1694-1707.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lonnerberg, P., Furlan, A.*, et al.* (2018). RNA velocity of single cells. Nature *560*, 494-498.

Labrecque, M.P., Coleman, I.M., Brown, L.G., True, L.D., Kollath, L., Lakely, B., Nguyen, H.M., Yang, Y.C., da Costa, R.M.G., Kaipainen, A.*, et al.* (2019). Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. J Clin Invest *129*, 4492-4505.

Marchio, C., Geyer, F.C., Ng, C.K., Piscuoglio, S., De Filippo, M.R., Cupo, M., Schultheis, A.M., Lim, R.S., Burke, K.A., Guerini-Rocco, E.*, et al.* (2017). The genetic landscape of breast carcinomas with neuroendocrine differentiation. J Pathol *241*, 405-419.

Perez Kerkvliet, C., Dwyer, A.R., Diep, C.H., Oakley, R.H., Liddle, C., Cidlowski, J.A., and Lange, C.A. (2020). Glucocorticoid receptors are required effectors of TGFbeta1-induced p38 MAPK signaling to advanced cancer phenotypes in triple-negative breast cancer. Breast Cancer Res *22*, 39.

Ping Mu, Z.Z., Matteo Benelli, Wouter R. Karthaus, Elizabeth Hoover,, Chi-Chao Chen, J.W., Sheng-Yu Ku, Dong Gao, Zhen Cao,, Neel Shah, E.J.A., Wassim Abida, Philip A. Watson, Davide Prandi,, Chun-Hao Huang, E.d.S., Scott W. Lowe, Leigh Ellis,, Himisha Beltran, M.A.R., David W. Goodrich,, and Francesca Demichelis, C.L.S. (2017). SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. Science *355*, 84-88.

Puccini, A., Poorman, K., Salem, M.E., Soldato, D., Seeber, A., Goldberg, R.M., Shields, A.F., Xiu, J., Battaglin, F., Berger, M.D.*, et al.* (2020). Comprehensive Genomic Profiling of Gastroenteropancreatic Neuroendocrine Neoplasms (GEP-NENs). Clin Cancer Res *26*, 5943-5951.

Quintanal-Villalonga, A., Chan, J.M., Yu, H.A., Pe'er, D., Sawyers, C.L., Sen, T., and Rudin, C.M. (2020). Lineage plasticity in cancer: a shared pathway of therapeutic resistance. Nat Rev Clin Oncol *17*, 360-371.

Rahul Aggarwal, J.H., Joshi J. Alumkal, Li Zhang, Felix Y. Feng, George V. Thomas, Alana S. Weinstein,, Verena Friedl, C.Z., Owen N. Witte, Paul Lloyd, Martin Gleave, Christopher P. Evans, Jack Youngren,, Tomasz M. Beer, M.R., Christopher K. Wong, Lawrence True, Adam Foye, Denise Playdle, Charles J., Ryan, P.L., Kim N. Chi, Vlado Uzunangelov, Artem Sokolov, Yulia Newton, Himisha Beltran, Francesca, and Demichelis, M.A.R., Joshua M. Stuart, and Eric J. Small (2018). Clinical and Genomic Characterization of Treatment-Emergent Small-Cell Neuroendocrine Prostate Cancer: A Multi-institutional Prospective Study. JOURNAL OF CLINICAL ONCOLOGY *36*, 2492-2503.

Scher, H.I., Beer, T.M., Higano, C.S., Anand, A., Taplin, M.-E., Efstathiou, E., Rathkopf, D., Shelkey, J., Yu, E.Y., Alumkal, J.*, et al.* (2010). Antitumour activity of MDV3100 in castration-resistant prostate cancer: a phase 1–2 study. The Lancet *375*, 1437-1446.

Scher, H.I., and Sawyers, C.L. (2005). Biology of progressive, castration-resistant prostate cancer: directed therapies targeting the androgen-receptor signaling axis. J Clin Oncol *23*, 8253-8261.

Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A.*, et al.* (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell *141*, 69-80.

Sheng Yu Ku, S.R., Yanqing Wang, Ping Mu, Mukund Seshadri,, Zachary W. Goodrich, M.M.G., David P. Labbé,, Eduardo Cortes Gomez, J.W., Henry W. Long, Bo Xu, Myles Brown,, and Massimo Loda, C.L.S., Leigh Ellis, David W. Goodrich (2017). Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. Science *355*, 78-83.

Taavitsainen, S., Engedal, N., Cao, S., Handle, F., Erickson, A., Prekovic, S., Wetterskog, D., Tolonen, T., Vuorinen, E.M., Kiviaho, A.*, et al.* (2021). Single-cell ATAC and RNA sequencing reveal pre-existing and persistent cells associated with prostate cancer relapse. Nat Commun *12*, 5307.

Wang, Y.X., Feige, P., Brun, C.E., Hekmatnejad, B., Dumont, N.A., Renaud, J.M., Faulkes, S., Guindon, D.E., and Rudnicki, M.A. (2019). EGFR-Aurka Signaling Rescues Polarity and Regeneration Defects in Dystrophin-Deficient Muscle Stem Cells by Increasing Asymmetric Divisions. Cell Stem Cell *24*, 419-432 e416.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

ACKNOWLEDGMENTS

The authors wish to express their gratitude to the patients and the families of the patients who contributed to this study. We would like to thank the LGCP Microscopy Core at the NCI/CCR and we would like to thank the CCR Single Cell Analysis Facility. Sequencing was performed with the CCR Genomics Core. This work utilized the computational resources of the NIH HPC Biowulf cluster (http://hpc.nih.gov). We would like to thank A. Zoubeidi for providing the EZH2 phospho-T350 antibody. We thank D. Takeda, G. Merlino, J. Shern, and M. Shen for reviewing the manuscript.

FUNDING

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Center for Cancer Research.

Prostate Cancer Foundation (Young Investigator Awards to M.L.B. and A.G.S.)

Department of Defense Prostate Cancer Research Program (W81XWH-16-1-0433 to A.G.S)

Support from CCR Single Cell Analysis Facility was funded by FNLCR Contract HHSN261200800001E.

AUTHOR CONTRIBUTIONS

Conceptualization: M.L.B., K.K.

Methodology: M.L.B., S.A.M., S.A., F.H.K., W.L.D., K.K.

Investigation: M.L.B., R.L., A.G.S., D.B., C.M.T., C.T., J.K., J.Y., A.N.A.

Formal analysis: M.L.B., B.J.C., A.T.K., K.K., R.L., A.G.S., J.Y., T.L.L.

Visualization: M.L.B., B.J.C., A.T.K.

Project administration: M.L.B., K.K.

Supervision: K.K.

Writing: M.L.B., K.K.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

MATERIALS AND METHODS

Histology

Formaldehyde-fixed tissue and organoid sections were embedded in paraffin blocks. Sections were cut, mounted on slides, and put through steps of graded alcohol deparaffinization. Steam antigen retrieval was performed for fifteen minutes (DAKO 1699) followed by washes in PBS/0.1% Tween-20 (PBST) 3x five minutes. Sections were blocked in Background Buster (Innovex NB306) for 40 minutes and then incubated overnight in primary antibody at 4ºC. The next day the slides were washed three times with PBST and then incubated with a biotinylated secondary antibody for 30 minutes at room temperature. Antibody staining was developed with 3, 3' diaminobenzidine (DAB) and counterstained with hematoxylin. Slides were imaged using a Zeiss Axioscan.Z1 microscope with a plan-apochromat 20x NA 0.8 objective.
ODX tumor sections were processed and imaged as described above. The sections were stained using an Intellipath FLX autostainer (Biocare Medical). Quantification of BrdU was done using the Indica Labs HALO v3.3 software running the CytoNuclear v2.0.9 algorithm. The optical density threshold for "weak" labeling was 0.377 "strong" was set at 1.331. Data was plotted using GraphPad Prism v8.
The IHC for the biopsy tissue sections was done in the NIH Clinical Center Pathology lab.

Organoid culture

Organoids were established and cultured according to our previously described methods and culture conditions (*9*). Patients provided informed consent, and samples were procured from the NIH Clinical Center under NIH Institutional Review Board approval in accordance with U.S. Common Rule. NCI-PC35-1, NCI-PC35-2, and LuCaP 145.2 organoids were grown in PrEN -p38i/-NAC media conditions. NCI-PC44 organoids were grown in PrEN -p38.

DNA extraction

DNA was extracted from the primary prostate tumor. Sections at 5 μM thick from the paraffin block of radical prostatectomy tissue were cut onto slides but not mounted and then stained with H&E. Tumor tissue from five sections was macrodissected and combined into one tube. Adjacent normal prostate tissue from 19 unmounted sections was combined into another tube. DNA was extracted using the QIAmp DNA FFPE Tissue Kit (Qiagen 56404). The protocol was modified to include the following steps: (1) Incubated overnight with shaking in Buffer ATL with proteinase K. (2) An additional wash step with 80% ethanol prior to elution. (3) The Qiagen ATE buffer was replaced with Low TE buffer (Applied Biosystems 4389764), pre-heated to 55ºC and applied to the column for ten minutes. The DNA was quantified with Quant-iT Picogreen (Invitrogen P11495).

DNA was extracted from NCI-PC35-1 and NCI-PC35-2 organoids using an AllPrep DNA/RNA Mini Kit (Qiagen 80204) according to the manufacturer's protocol for animal cells. Qiashredder columns (Qiagen 79656) were used for the homogenization step.

DNA whole-genome sequencing

1 μg of genomic DNA was fragmented (Covaris), end-repaired, and assembled into paired-end libraries using the Illumina TruSeq DNA Library Preparation Kit. Libraries were sequenced with 150 cycles paired-end (2 × 150) on an Illumina HiSeq 4000. Per-lane FASTQ pairs were

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

trimmed using Trimmomatic version 0.39 and aligned to hg19 using BWA-MEM version 0.7.17. PCR duplicates were marked using the SPARK implementation of GATK MarkDuplicates version 4.1.4.1 with PICARD SetNmMdAndUqTags. Base quality score recalibration was performed using the SPARK implementation of GATK BQSRPipeline. Lane-level BAM files were merged using PICARD MergeSamFiles and GATK MarkDuplicates was run a second time with PICARD SetNmMdAndUqTags. A normal saliva sample was sequenced to a mean depth of 32.8× coverage. The tumor samples were sequenced to a mean depth of 54.5× coverage (range: 40.4× to 78.6×).

Somatic mutation calling

MuTect2 in GATK 4.1.3.0 was used in single-sample mode to generate VCF files for each normal BAM with the disable-read-filter set to MateOnSameContigOrNoMappedMateReadFilter and max-mnp-distance set to 0. A panel-of-normals was generated using GATK GenomicsDBImport with merge-input-intervals set to true and GATK CreateSomaticPanelOfNormals. MuTect2 was next run in paired mode with each tumor sample BAM matched to its benign normal BAM from the same type of sample (FFPE or fresh) and run with the panel-of-normals (pon), filtering in real-time against mutations observed in gnomAD, and with disable-read-filter set to MateOnSameContigOrNoMappedMateReadFilter. GATK GetPileupSummaries (filtering on ExAC sites) and GATK CalculateContamination were used on each tumor BAM for filtering raw MuTect2 calls using GATK FilterMutectCalls. Finally, 8-OxoG and FFPE filtering was performed, first using GATK CollectSequencingArtifactMetrics on each tumor BAM and passing its output GATK FilterByOrientationBias with artifact-modes set to G/T and C/T. Mutations were annotated using Oncotator.

Somatic copy number alteration calling

A joint set of copy number alterations and their clonal prevalence was determined using both GATK 4.1.3.0 and TitanCNA version 1.23.1 from whole-genome sequencing data. Using GATK, denoising was performed separately for FFPE and fresh tissues, first applying GATK CollectReadCounts for each tumor and normal BAM, and assembling a panel of normals using CreateReadCountPanelOfNormals. GATK DenoiseReadCounts was run on each tumor or normal sample using the appropriate panel of normals. GATK CollectAllelicCounts was run on each sample BAM for high-confidence 1000 Genomes Phase 1 SNP sites. Segmented copy number ratios were then calculated by using GATK ModelSegments, using denoised copy ratios for both matched tumor and normal as well as the allelic counts for each tumor sample. GATK CallCopyRatioSegments identified each region of gain or loss, per sample. TitanCNA was run using R version 3.6 on chromosomes 1-22 and X with 10kb intervals.

Tumor phylogenetic analysis

Phylogenetic tree estimation was performed using PhyloWGS version 1.0. Prior to tree evolution, mutations input was optimized as follows: 1) MuTect2 output multi-sample VCF files were filtered to tumor-only; 2) A floating depth cutoff was applied so that mutations in a single sample must be greater than 70% of the average depth of that sample from the same patient; 3) A hard filter of 90% strand bias was imposed; 4) A combined list of all mutations for all samples from each individual were compiled with a hard filter at 10% variant allele fraction (VAF); mutations less than 10% VAF were recovered from other samples provided they were >10% VAF in at least one sample. Copy number input was optimized as follows: 1) 1-bp segments

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

were removed from the joint output of TitanCNA and GATK; 2) high-level amplification and deep deletion events filtered from TitanCNA but present in output from the ichorCNA module were reintegrated into the .SEG file output when overlapping with GATK calls. PhyloWGS inputs per-patient were prepared using the create_phylowgs_inputs script joining each individual VCF (vardict) and CNV sample into a single set of SSM and CNV data. The corresponding SSM, CNV and parameters JSON files were then run using the multievolve script for parallel tree generation across 40 chains, using 1000 burn-in Markov chain Monte Carlo (MCMC) samples and 2500 fit MCMC iterations for a total of 100,000 potential tree structures. After tree generation, mutation and tree JSON files from the write_results script were parsed to select the tree with the lowest (most negative) log likelihood score. The best scoring tree was pruned to conservatively decrease the number of major subclones. If any given node did not have at least 5 SNVs or SSMs assigned to it, it was merged with its sibling node with the greatest number of events. If that node had no siblings, it was merged with its most immediate ancestral node, unless it was a direct descendent of the germ/normal node with no descendants, in which case it was eliminated. The subclonal composition of each node was determined by the average clonal prevalence of SSMs/CNVs assigned to each node and their relative proportion in each sequenced tumor sample.

Immunoblots

5 x 10^5 cells from dissociated organoids were lysed in lysis buffer (50 mM Tris (pH 8) + EDTA (10 mM) + 1% SDS) with protease and phosphatase inhibitors. Protein concentration was determined using a BCA assay (Pierce 23227). 10 µg of protein was loaded onto 4-20% Mini_PROTEAN TGX gels (Bio-Rad 456-1094) or 4-20% Mini_PROTEAN TGX Stain-Free gels (Bio-Rad 4568091). Semi-dry transfer was done with a Bio-RadTrans Blot Turbo apparatus for 30 minutes using Trans-Blot Turbo 5x Transfer Buffer (Bio-Rad10026938) except for ARID1A and ARID1B overnight - wet transfers were done. Membranes were blocked for 1 hour in 5% BSA. Overnight incubations with the primary antibodies were done at 4ºC while rocking. Secondary antibody incubations were done for one hour at room temperature while rocking. Blots were developed with Clarity Western ECL Substrate (Bio-Rad 170-5061) and visualized on a Bio-Rad ChemiDoc Touch Imaging System.

Immunofluorescent staining

Organoids were dissociated and re-plated in 2D on 16-well chamber slides (Nunc 178599) coated with 75 µg/ml poly-D-lysine (Millipore A-003-E) followed by 3% Matrigel (Corning 356231). Cells were fixed for 10 minutes in 4% formaldehyde, then rinsed three times with PBS. Cells were permeabilized and blocked for one hour in PBS/5% goat serum/0.3% Triton-X 100. The cells were then incubated in primary antibody diluted in PBS/0.5% BSA overnight at 4ºC. The cells were then washed 5x fifteen minutes at room temperature in PBST and incubated with fluorochrome-conjugate secondary antibody for one hour, followed by 5x fifteen-minute washes. Coverslips were mounted with Fluoro-Gel II + DAPI (Electron Microscopy Sciences 17985-50). Slides were imaged using a Zeiss Axioscan.Z1 microscope with a plan-apochromat 20x NA 0.8 objective and a Colibri 7 LED light source. Quantification of IF images was done using the Indica Labs HALO v3.3 software running the CytoNuclear FL v2.0.12 algorithm.

Proliferation assays

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Organoids were dissociated then replated in 3D in 96 well plates. Each time-point was plated in five well replicates and incubated overnight. All time-points were then quantified at the indicated day with CellTiter Glo 3D (Promega G9682) and luminescence was measured using a Tecan infinite M200 Pro plate reader. The average fold change for each time-point relative to day-0 was calculated. Three independent experiments were performed.

EdU-incorporation assays

Twenty-four-hour pulse: organoids were dissociated then replated in 3D overnight. The next morning 10 μM EdU (Invitrogen C10338) was added to the cultures for 24 hours. The organoids were then either immediately collected and replated in 2D for staining and imaging or they were maintained in culture for a chase period and collected at the appropriate time-point. EdU staining was performed according to the manufacturers protocol for most assays except the combination EdU/RNA-FISH assays where the following modifications were made: 1) BSA was not used in the wash buffers. 2) The incubation time in the Click-iT reaction cocktail was reduced to five minutes. Imaging was performed as describe above for immunofluorescence. Quantification of EdU was done using the Indica Labs HALO v3.3 software running either the CytoNuclear FL v2.0.12 algorithm or FISH-IF v1.2.2 algorithm. Cells were counted as EdU-positive above a minimum fluorescence value of 2,000.

Long-term incorporation assays: organoids were dissociated then replated in 3D overnight. Culture media containing 10 μM EdU was added and replaced every twenty-four hours until the organoids were collected at the appropriate time-points.

PCA plot of AR v NE score WCM cohort

Raw FASTQ files were accessed from dbGaP phs000909.v.p1 and reanalyzed using the nextflow core RNA seq pipeline v1.0. Following the methods described in Beltran et al. (*2*), a reference AR sample was generated by using the gene expression values for genes in the AR signature from a series of three LNCaP samples sequenced at NCI/CCR. A reference neuroendocrine sample was generated by averaging the expression of neuroendocrine genes across the neuroendocrine samples from the Weill Cornell Medicine (WCM) cohort. The AR score was defined as the correlation of the expression of the sample with the AR reference sample. The integrated NEPC score is defined as the correlation between the sample and the reference neuroendocrine sample.

scRNA-seq

Organoids growing in 3D in Matrigel and culture media in a 12-well plate were collected from the Matrigel by adding 1 mg/ml Dispase (Gibco 17105-041) to the culture for two hours and transferred to Eppendorf tubes. The organoids were pelleted by centrifuge and dissociated in 100 μl of TrypLE (Gibco 12605-028) + 100 μg/ml of DNAse-I (Sigma Aldrich DN25) for 20 minutes at 37ºC with mechanical agitation every five minutes by pipette, using low retention tips. One ml of Advanced DMEM/F12 (Invitrogen 12634-02898) + 10 μM Y-27632 ROCK inhibitor (Stemcell Technologies 72307) was added to neutralize the TrypLE. The cells were then passed through a 30 μM cell strainer (Miltenyi Biotec 130-098-458) and assessed for viability and doublets before being pelleted and washed 3x in buffer (PBS + 0.04% BSA + Y-27632 (10 μM)). The cells were then counted and loaded onto the 10x Genomics Chromium platform using the 3' v3.0 gene expression chemistry. Preparation of libraries were performed

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

according to vendor recommendations. Single cell libraries were sequenced on either an Illumina NextSeq 500/550 instrument or an Illumina NextSeq 2000 instrument. Data was processed using the 10x Genomics cellranger pipeline to demultiplex reads and then align those reads to the GRCh38 reference genome. Gene barcode matricies were generated using the cellranger pipeline from 10x Genomics aligned against grch38. An in-house single cell processing pipeline was used to standardize analysis across all samples which follows the methodology laid out in the Bioconductor single cell analysis book. Gene barcode matricies were read into R and doublets were detected and removed using scDblFinder. Additional quality control was applied using the scran and scatter packages, using the addPerCellQC function and filtering out cells that were identified as outliers using the isOutlier function for mitochondrial gene content, lower number of reads, and lower number of detected genes. Initial dimensional reduction was performed using GLMPCA from the scry package on all genes in the experiment. UMAPs were generated from three independent experiments for PC35-1 and PC35-2 and two experiments for LuCaP 145.2. Mutual nearest neighbor correction was performed to correct for batch effects on the principal components, and the corrected top 30 principal components were used to generate the UMAP. For PC44, UMAP was performed on the top 30 principal components from one experiment. Monocle3's graph-based clustering using leiden community detection with a q value cutoff of 0.05 was used to identify clusters and larger partitions. Marker gene detection was performed using the score markers function from scater. Cell cycle state was inferred using cyclone.

scRNA-seq signature scores

Signature scores for individual cells were generated by running PCA on batch corrected and normalized expression values from all single cell RNA sequencing samples using only the genes in published signatures. The AR and neuroendocrine signatures were created using the Beltran et al. (*2*) signatures, and the proliferation signature was generated using the gene list from Balanis et al. (*5*). The signature value is the loading for a particular cell from the first principal component.

RNA velocity

RNA velocity was calculated independently on each sample using the default settings in velocyto. RNA velocity vectors were generated using batch corrected principal components to embed on the UMAP.

CellTag analysis

Organoids were collected and dissociated to single cells for transduction with a lentiviral library of CellTags. The CellTag library (CTL) was prepared according to Biddy et al. (*13*). Lentivirus was made by transfecting Lenti-X 293T cells (Clontech 632180) with CTL plasmids plus psPAX2 and VSV-G packaging plasmids using Lipofectamine 2000 (Invitrogen 11668019). The transfection mix was applied to the cells for six hours then removed and replaced with lentiviral collection media: DMEM + 10% FBS(HyClone) + 1.1% BSA + HEPES (10 mM) + sodium pyruvate (10 mM) + Primocin (Invivogen ant-pm-1). The lentivirus was collected in two batches at 48 and 72 hours and pooled together, then spun for 5 minutes at 1000 x g to pellet debris. The supernatant was then passed through a 0.45 μM PES membrane filter. The lentivirus was concentrated 100-fold by ultracentrifuge: four hours at 4ºC at 20,000 x g with low acceleration and then resuspended in PBS, aliquoted and stored at -80ºC. For transduction, 5 x 10^5 cells were combined with 3.5 μl of lentivirus and 2 μl of LentiBOOST (Sirion Biotech) in 2 ml of culture

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

media. The cells/lentivirus were transferred to one well of a 6-well plate coated with 3% Matrigel and centrifuged at 1,000 x g, low acceleration, for 90 minutes at 32ºC. The plate was then incubated overnight at 37ºC. The next morning the cells were detached from the plate with TrypLE, collected and counted, then re-plated in 3D in multiple wells of a 24-well plate at different concentrations ranging from $5 \times 10^3 – 1 \times 10^5$ in order to maximize recovery of the targeted 10,000 - 15,000 cells desired for loading onto the 10x Genomics platform. The cells were kept in culture for four weeks, changing the media twice/week. Each well of organoids was then collected and processed as described above in the "scRNA-seq" paragraph of this methods section. After counting, we determined that 5×10^4 cells/well yielded the ideal 15,000 cells after processing. 15,000 single cells were loaded onto the 10x Genomics Chromium platform as described above. Single cell libraries were sequenced as described above. Raw single cell FASTQs were aligned to a custom reference including the EGFP construct used in the vector for the cell tags (*13*). Reads were filtered to include only sequences that aligned to EGFP. The CellTagR package was used with barcode correction relying on starcode to call clones. Cells were considered clones if they shared at least two celltags and their jaccard similarity exceeded 0.7 as specified in the documentation. To project clones onto UMAP embeddings, segments were drawn between cells that were called clones.

RNA-FISH

Organoids were dissociated and 75,000 cells were replated overnight in 2D on 12 mm round #1 coverglass (Electron Microscopy Sciences 72231-01) coated with 75 µg/ml Poly-D-lysine followed by 3% Matrigel. The cells were washed in PBS, fixed in 4% formaldehyde for 10 minutes at room temperature, and finally washed twice in PBS. The cells were permeabilized in 70% ethanol for at least one hour at 4ºC, the ethanol was removed, and Wash Buffer A (Biosearch Technologies SMF-WA1-60) was added and incubated at room temperature for five minutes. For staining, the Stellaris RNA-FISH probes, diluted in Hybridization Buffer (Biosearch Technologies SMF-HB1-10) plus 10% Deionized Formamide (Millipore 4610), were added to the cells and incubated overnight in a humidified chamber at 37ºC. The cells were washed in Wash Buffer A for 30 minutes at 37ºC in the dark, then counter-stained with 5 ng/ml DAPI diluted in Wash Buffer A in the dark at 37ºC for 30 minutes. The cells were washed in Wash Buffer B (Biosearch Technologies SMF-WB1-20) for 5 minutes at room temperature in the dark, the cover glass was mounted onto a slide with ProLong Gold antifade reagent (Invitrogen P36934), allowed to dry and stored at -20ºC in the dark. For RNA-FISH/EdU and RNA-FISH/IF combined assays, the RNA-FISH hybridization was done first up to/including the Wash Buffer B step. The cells were rinsed twice in PBS and stained for EdU or stained with antibodies for IF. For EdU incorporation/staining, see the above methods section for details. For IF, the blocking step was excluded, and antibodies were diluted in PBS. Imaging was done with a Nikon Ti2 microscope equipped with a CFI Plan-Apochromat 60x NA 1.4 oil immersion objective, Lumencor Sola SE 365 FISH light engine, and Photometrics Prime BSI sCMOS camera. A maximum intensity projection was created from a 3.6 µM 13 step Z stack for each field of view. Quantification of RNA-FISH and combined assay images was done using Indica Labs HALO v3.3 software running the FISH-IF v1.2.2 algorithm. For RNA-FISH scatter plots, total FISH counts were plotted. For the RNA-FISH/IF combined assays, total FISH counts were plotted against raw IF intensity values. For the RNA-FISH/EdU drug-treated assays, a minimum threshold of five spots (transcripts) per cell was set to call a cell positive for a given FISH marker. Cells were counted as EdU-positive above a minimum fluorescence value of 2,000. The

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

EZH2 probe set was ordered from the Stellaris Design Ready Probe Sets (Biosearch Technologies VSMF-2123-5). All other RNA-FISH probe sets were custom designed using the Stellaris Probe Designer tool at the biosearchtech.com website, and QC'd for specificity using the UCSC genome browser BLAT function. The custom designed RNA-FISH probe-set sequence information is in Table S1.

CopyKAT

Copy number variation was computed using CopyKat *(14)* with default setting and *cell.line* mode enabled. Briefly, raw counts from scRNAseq experiments were used as input to CopyKat. CopyKat clusters were generated by unsupervised hierarchical clustering of the CNV results using the function *hclust* with ward.D linkage function on the cell distance matrix computed using the *dist* function calculated with "euclidean" method. Copykat clusters were assigned based on the number of UMAP clusters using *cutree* function.

scATAC-seq

Organoids were collected and dissociated as described above in the "scRNA-seq" paragraph of this methods section. Single cell suspensions of 2.5 x 10^5 cells were spun down and resuspended in 100 µl of cold ATAC lysis buffer (10 mM Tris(pH 7.4) + 10 mM NaCl + 3 mM MgCl2 + 1% BSA + 0.1% Tween-20), pipetted up/down 10x, incubated on ice for five minutes, and finally pipetted an additional 5x before adding 1 ml of ATAC Wash Buffer (10 mM Tris(pH 7.4) + 10 mM NaCl + 3 mM MgCl2 + 1% BSA + 0.1% Tween-20 + 0.1% NP40 + 0.01% digitonin). The cells were then pelleted at 500 x g for 5 minutes at 4ºC. All of the wash buffer was removed, and the nuclei were resuspended in 50 µl of Nuclei Buffer (10x Genomics PN-2000153/2000207). Single nuclei suspensions were transposed before being partitioned on the 10x Genomics Chromium platform using the Single Cell ATAC v1.1 chemistry (10x Genomics). Preparation of libraries were performed according to vendor recommendations.
Single cell atac sequencing was processed using the cellranger scatac pipeline from 10x Genomics. Additional analysis was performed using the ArchR library using 250,000 features for the latent semantic indexing. Inferred transcription factor activity was generated using the method included in ArchR for generating ChromVAR deviations Z scores. The score markers function was applied and performs a competitive ranking of features using three statistical tests, Welch's t test, Wilcoxon rank sum test, and a binominal test. Features that were ranked highly in all three tests were considered.

Dose-response assays

Organoids were dissociated then replated in 3D at 3,000 cells/well in 384 well plates. Drugs were prepared by two-fold serial dilutions starting at 10 µM and spanning 11 concentrations, plus an additional vehicle control. All treatments were done in replicates of five. The cells were treated twice per week for two weeks, then quantified with CellTiter Glo 3D and luminescence was measured using a Tecan infinite M200 Pro plate reader. Data is shown as an average of three independent experiments.

Xenograft tumor study

The animal study was performed according to the protocol approved by the NCI-Bethesda Animal Care and Use Committee. The organoid-derived xenograft (ODX) model was established initially from NCI-PC35-1 organoids subcutaneously injected in NOD scid gamma (NSG) mice,

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

and subsequently maintained by serial passage of tumor fragments in NSG mice. For the experiment, 2 mm tumor fragments were implanted subcutaneously in NSG mice. When the tumors reached an average size of 0.3 cm^3 the mice were randomized into four treatment groups of five mice/group. Mice in the castrated groups were castrated by orchiectomy concurrent with the start of drug treatment. Mice were drugged once daily, five days/week by oral gavage with 30 mg/kg of alisertib suspended in vehicle (10% 2-hydroxypropyl-β-cyclodextrin, 1% sodium bicarbonate in water). Mice in the vehicle control group were treated on the same schedule. Tumor volumes were measured twice/week. The study was terminated after nine weeks when the control group reached the maximum allowable burden of 2 cm^3. Tumors were harvested and fixed in 4% formaldehyde overnight then transferred to 70% ethanol.

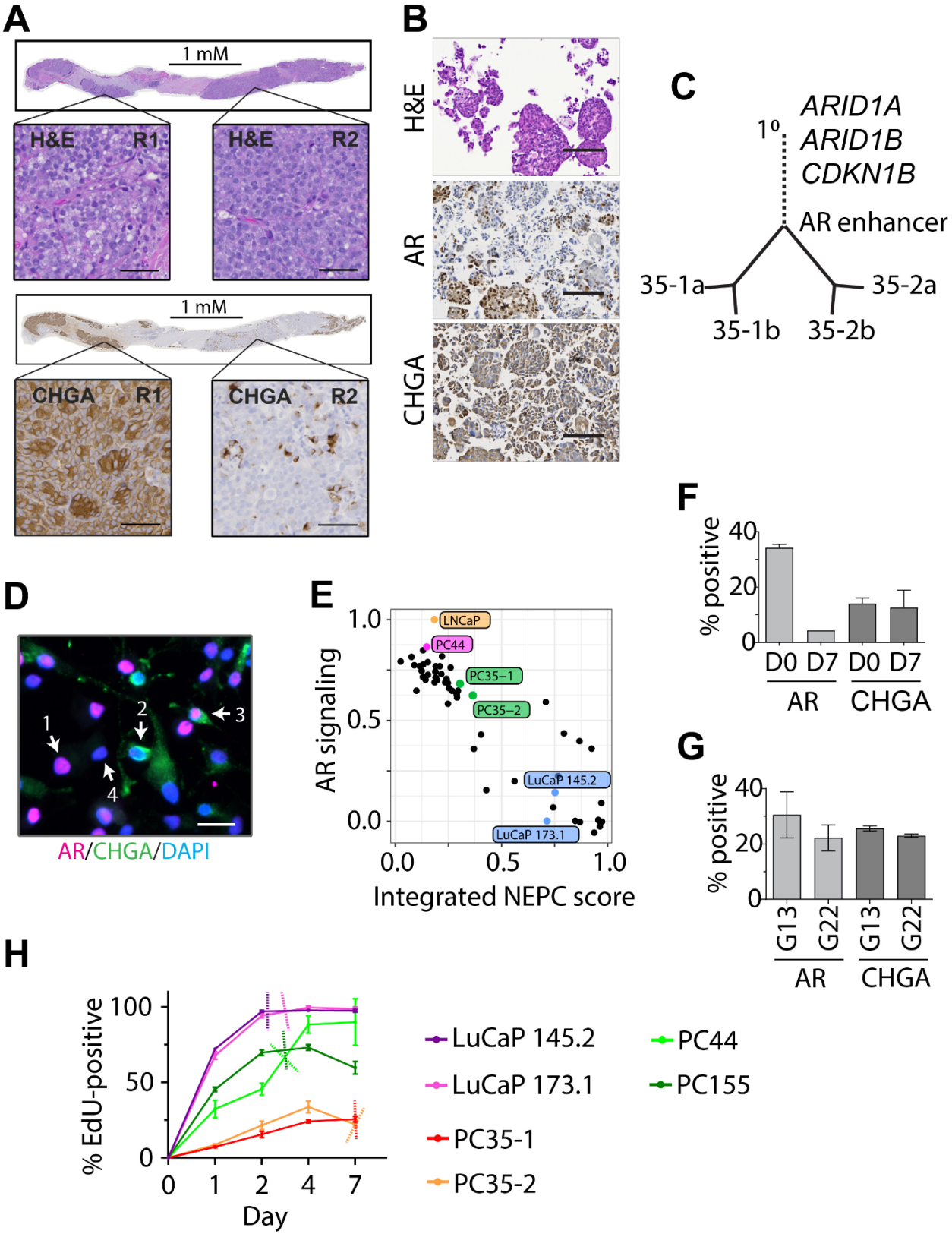
Antibodies

Target	**Company**	**Catalog number**	**Assay**
AR	Abcam	ab133273	IF/Western
AR	Cell Signaling	5153	IHC
ARID1A	Cell Signaling	12354	Western
ARID1B	Abcam	ab57461	Western
BrdU	Abcam	ab6326	IHC
CHGA	Invitrogen	MA5-13096	IF/IHC
EZH2	Cell Signaling	5246	Western
EZH2 phospho-S21	Bethyl Laboratories	IHC-00388	Western
EZH2 phospho-T345	Active Motif	61242	Western
EZH2 phospho-T350	Gift from Amina Zoubeidi	NA	Western
p27	BD Biosciences	610241	Western
p53	Cell Signaling	2524	Western
PTEN	Cell Signaling	9188	Western
RB	Cell Signaling	9309	Western
SYP	Agilent	M731529-2	IHC

Data availability

The sequence information for all RNA-FISH probe sets is located in Supplementary Information Table 1. The WGS, scRNA-seq, and scATAC-seq data have been deposited in ###.

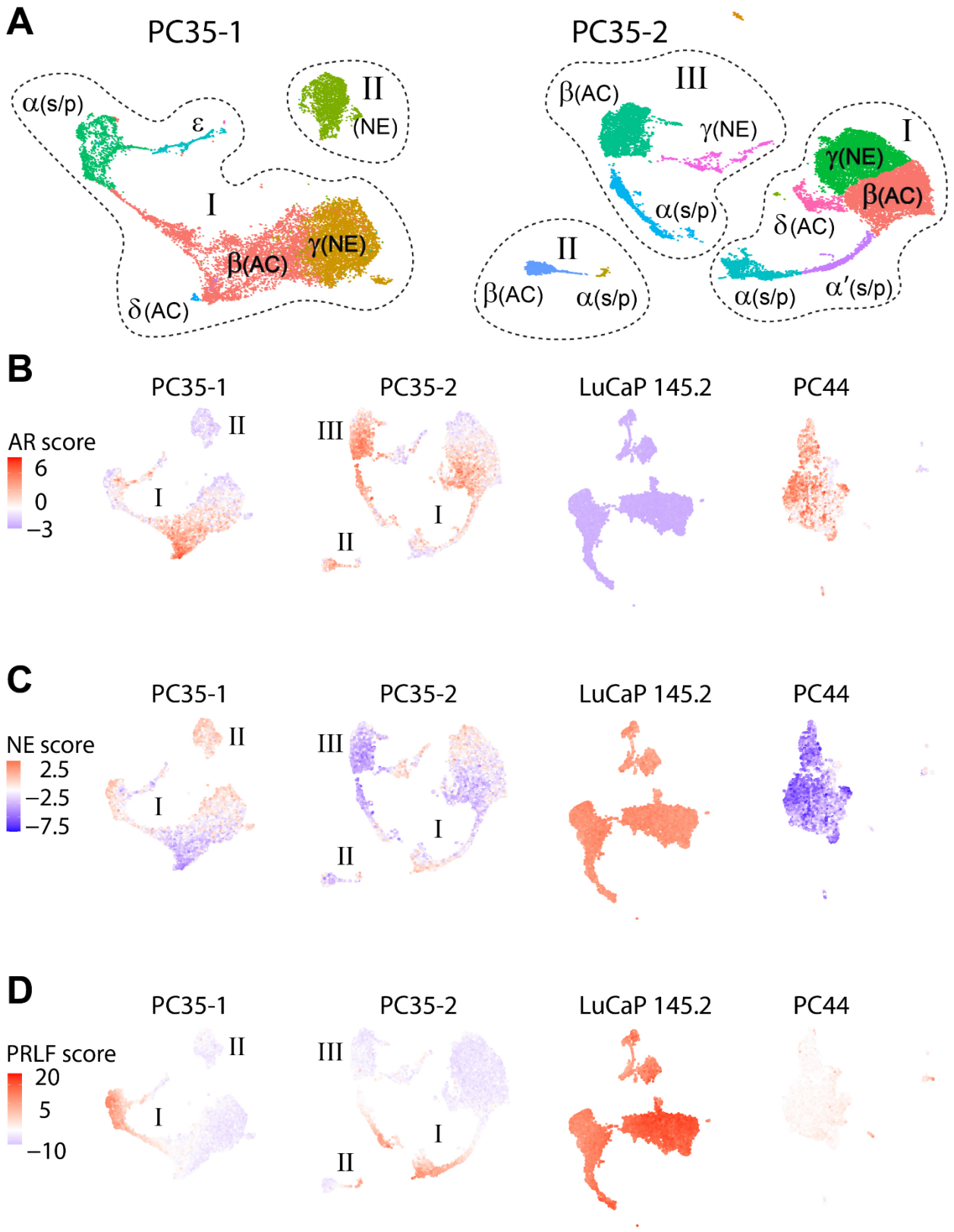
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Figure 1. Patient-derived organoid models of mCRPC capture and maintain genetic and phenotypic heterogeneity (A) Serial sections of tumor biopsy tissue stained with hematoxylin and eosin (H&E) or the indicated antibodies. Magnified views of Region-1 (R1) and Region-2 (R2) are shown. Scale bars, 50 μM. (B) Patient biopsy derived organoid sections stained with H&E or the indicated antibodies. Scale bars, 200 μM. (C) Phylogenetic tree of primary tumor and metastasis derived organoids. Primary prostate tumor, 1°. PC35-1 subclones - 35-1a, 35-1b. PC35-2 subclones - 35-2a, 35-2b. Significant genetic events indicated at positions in the tree where they originated. (D) PC35-1 organoids were dissociated to single cells and stained by immunofluorescence (IF) with antibodies against the indicated proteins and DAPI. Each of four phenotypes is indicated by a number and arrow. 1 = $AR^{POS}/CHGA^{Lo/NEG}$; 2 = $AR^{NEG}/CHGA^{Hi}$; 3 = $AR^{POS}/CHGA^{Hi}$; 4 = $AR^{NEG}/CHGA^{Lo/NEG}$. (E) The Weill Cornell Medicine cohort of mCRPC (black filled circles) and the indicated organoid models or cell line are plotted by AR signaling score and NEPC score. (F) IF combined with 5-ethynyl-2'-deoxyuridine (EdU) pulse-chase assay of PC35-2. Organoids were pulsed with EdU (10μM) for 24 hours (D0) and chased for seven days (D7). The organoids were dissociated and quantified as single cells and plotted as percent-positive of total cells. (G) IF staining of CT35-1 organoids with antibodies against the indicated proteins. Generation 13 and 22 organoids were dissociated and quantified as single cells then plotted as percent-positive of total cells. (H) Continuous EdU-incorporation assay for the indicated organoid lines. The graph shows the percentage of EdU-positive cells of the over time. Dashed lines mark the approximate day of population doubling for each. Bar and line graphs are plotted as the mean of three independent experiments. Error bars represent ± standard error of the mean (SEM).

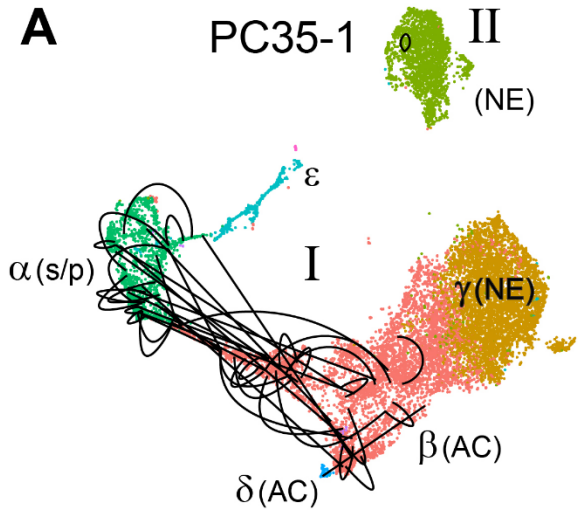
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



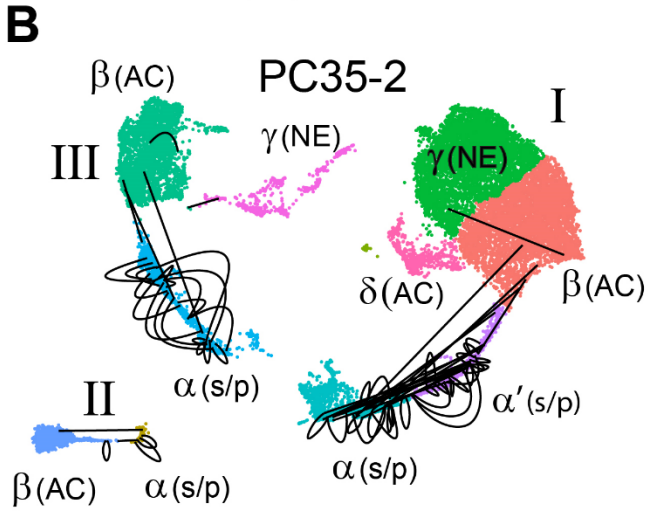
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Figure 2. Single-cell transcriptomics identifies lineage-distinct heterogeneity in the PC35 organoid models (A) scRNA-seq transcriptomic profiles of PC35-1 and PC35-2organoids plotted as UMAPs. Major clusters are circled and labeled with roman numerals (I, II, II). Subclusters are colored and annotated with Greek letters (α – ε) and phenotype designations: stem/progenitor (s/p), adenocarcinoma-like (AC), or neuroendocrine-like (NE). (B) AR and (C) neuroendocrine signature scores for each cell determined by principal component analysis (PCA) using published gene sets from Beltran et al. Loadings from the first principal component for each cell are projected onto the UMAPs from (A) and UMAPs plotted for LuCaP 145.2 and PC44 scRNA-seq transcriptomic data. (D) Proliferation score determined as in (B) and (C). The proliferation gene set was derived from Balanis et al.

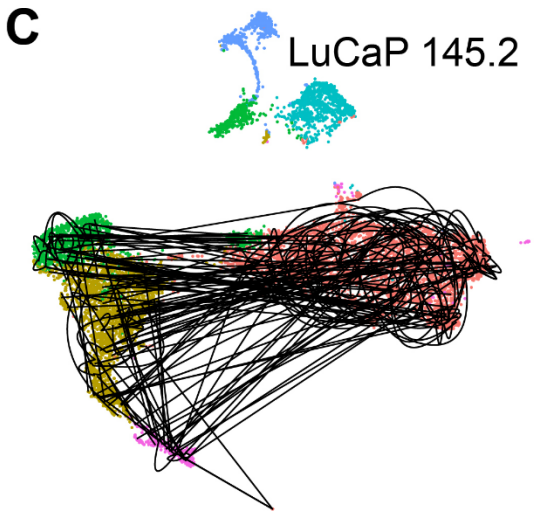
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

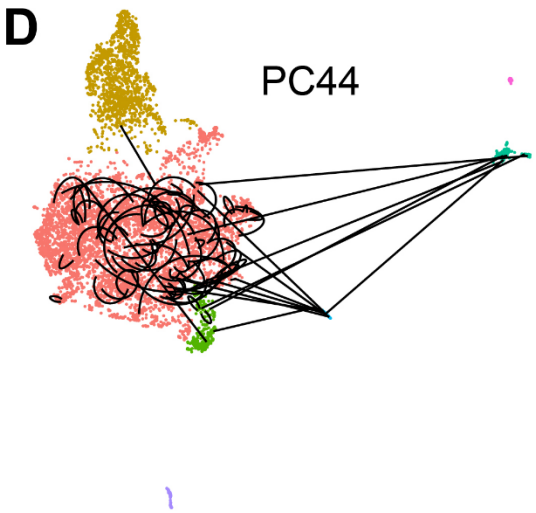


Major cluster	Subcluster	Number of cells/ subcluster	Number of CellTag clones	Percent CellTag clones
I	**α**	826	35	**4.2**
	β	1803	47	2.6
	γ	1566	0	0.0
	δ	22	1	4.5
	ε	361	0	0.0
II	——	625	2	0.3



Major cluster	Subcluster	Number of cells/ subcluster	Number of CellTag clones	Percent CellTag clones
I	**α**	729	36	**4.9**
	α'	298	41	**13.8**
	β	1472	6	0.4
	γ	2173	1	0.05
	δ	235	0	0.0
II	**α**	57	5	**8.8**
	β	334	4	1.2
III	**α**	571	41	**7.2**
	β	1099	7	0.6
	γ	312	1	0.3

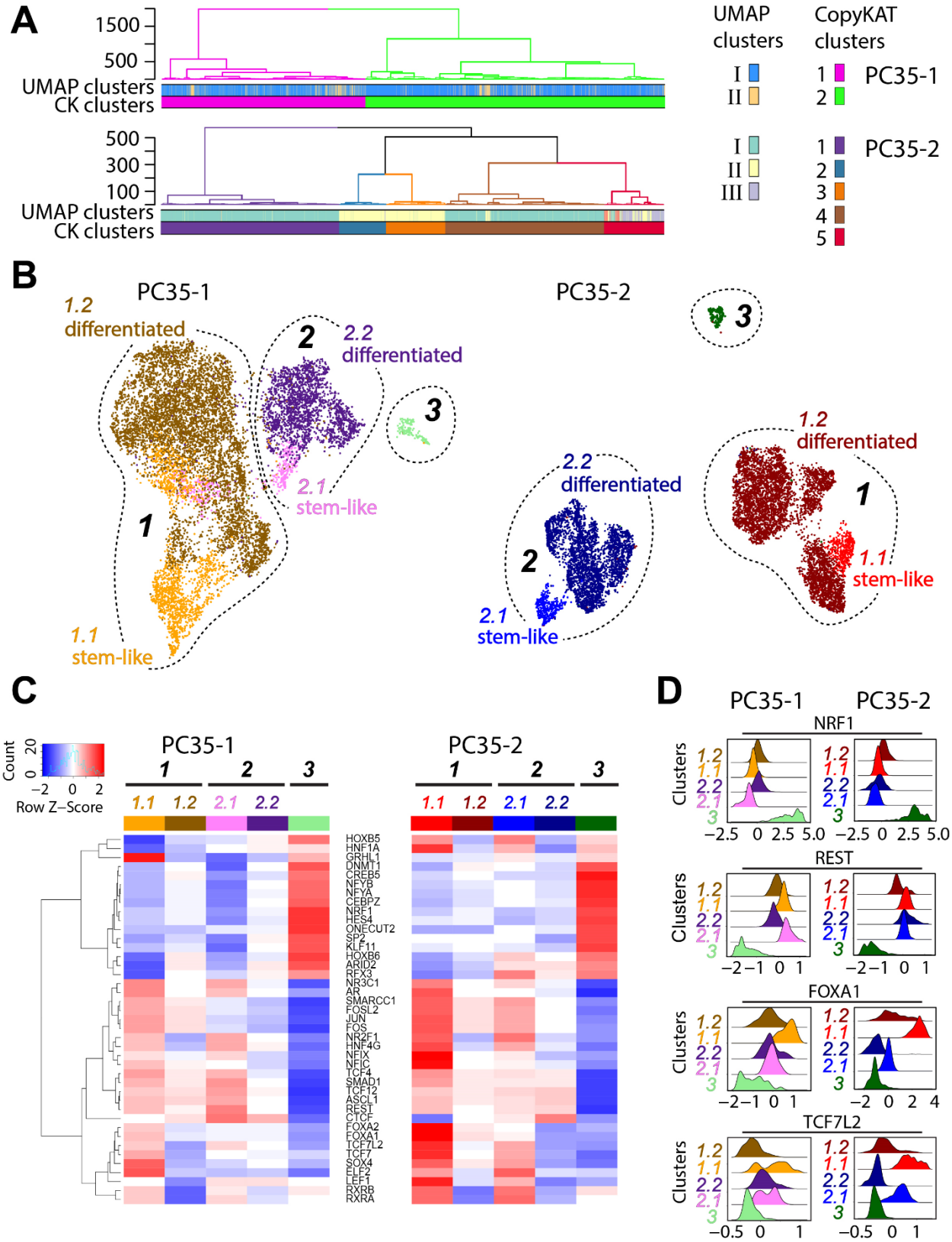




bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Figure 3. Single-cell combinatorial barcoding identifies lineage-distinct and stem-like/progenitor subpopulations (A-D) CellTag lineage-tracing analysis. (A) UMAP major clusters I and II of PC35-1 and (B) I, II, and III of PC35-2 are shown. The major clusters are further divided into annotated subclusters (α-ε). Each cell in a clonal population (≥ 2 cells expressing the same combination of barcode IDs), is connected by a black line. Self-renewing clones that exist in the same subcluster are connected by curved lines, differentiating clones that span at least two subclusters are connected by straight lines. The tables to the right show the quantification of cells/subcluster, CellTagged clones/subcluster, and the percentage of CellTagged clones/subcluster. (C) Clonal connections as in (A) and (B) mapped onto LuCaP 145.2 and (D) PC44 UMAPs.

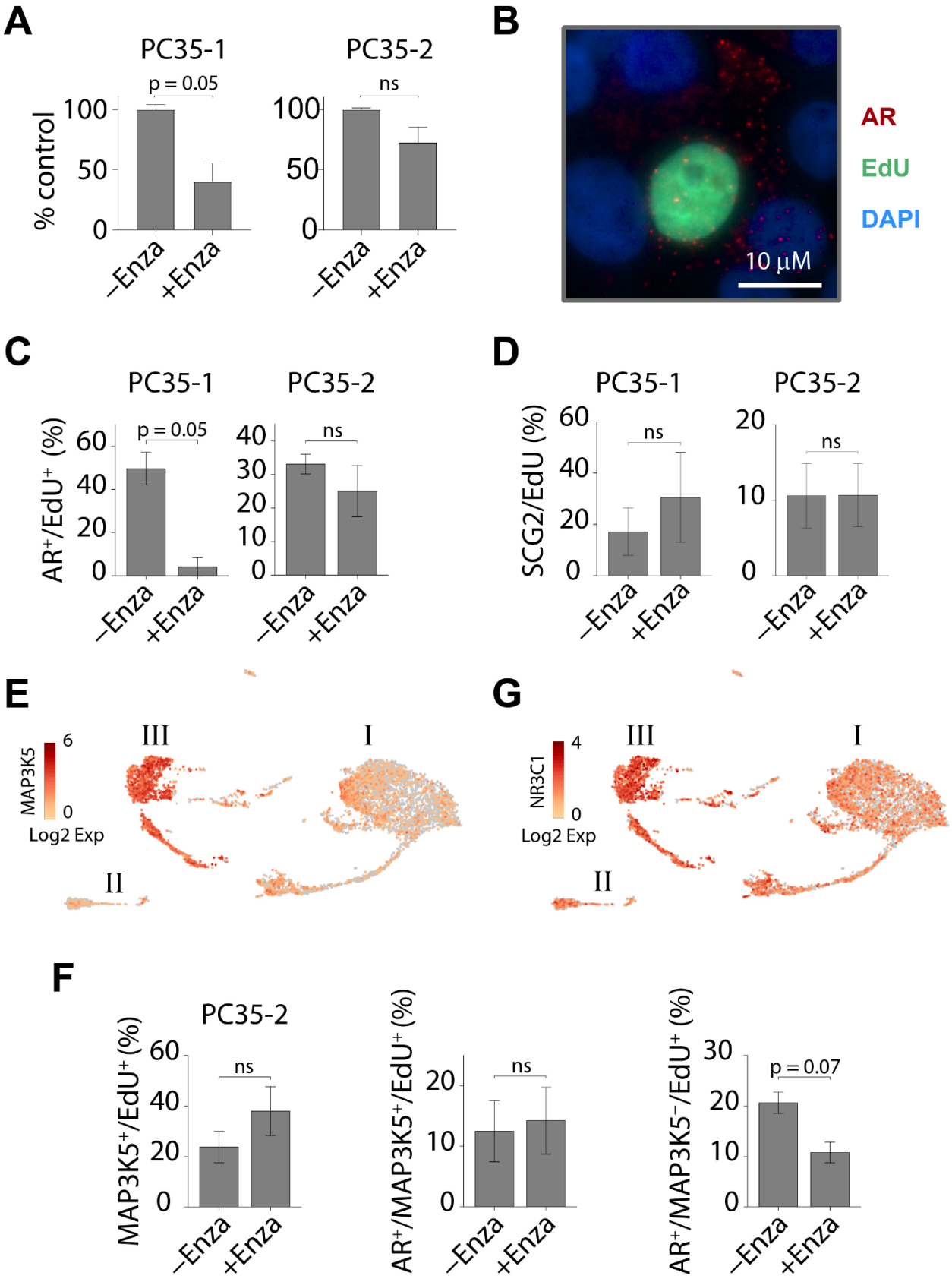
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Figure 4. Distinct states of chromatin accessibility and transcription factor activities are associated with NEPC and stem-like progenitor subpopulations in the PC35 organoids (A) Comparison of subpopulations defined by their scRNA-seq transcriptional profile (UMAP clusters), to subclones defined by genomic CNV (CopyKAT clusters) determined using the same scRNA-seq data. Results for both PC35-1 (top) and PC35-2 (bottom) are shown. Dendrograms are colored according to CopyKAT cluster number and show the hierarchical relationships among the CopyKAT clusters. Heatmaps directly below the dendrograms show the distribution of cells from UMAP major clusters I, II, and III throughout the CopyKAT clusters. Each cell is represented by a vertical line colored according to the UMAP cluster (top row) or the CopyKAT cluster (bottom row) to which it belongs and sorted by CopyKAT cluster. PC35-2 contains two additional minor UMAP clusters lacking differentially-expressed genes that were not annotated here. (B) UMAPs of global chromatin accessibility for PC35-1 and PC35-2. Major clusters are annotated as *1*, *2*, *3*. Clusters *1* and *2* are partitioned into two additional subclusters, stem-like and differentiated. (C) Heatmaps show inferred transcription factor (TF) activities of the listed TFs for each of the UMAP clusters/subclusters in PC35-1 and PC35-2. The heatmaps are colored by deviations z-scores for each row. Deviations z scores are the inferred transcription factor activity score. Deviations is the measure of how different the accessibility profile of a cell is for a particular transcription factor compared to the average accessibility profile for the entire dataset. Z scores are the z scored deviations. Selection of TF activity to be visualized was accomplished using the score markers function (see Methods) on the deviations z scores to identify which TFs showed the most enrichment when comparing across lineages and development. Transcription factors shown were determined to be expressed and selected from a list of the top fifty most deviant TFs. (D) Inferred transcription factor activity density plots for each cluster population. Deviations z scores are shown on the x-axis for the TF indicated at the top. Density estimates are represented along the y-axis and broken down by cluster/subcluster.

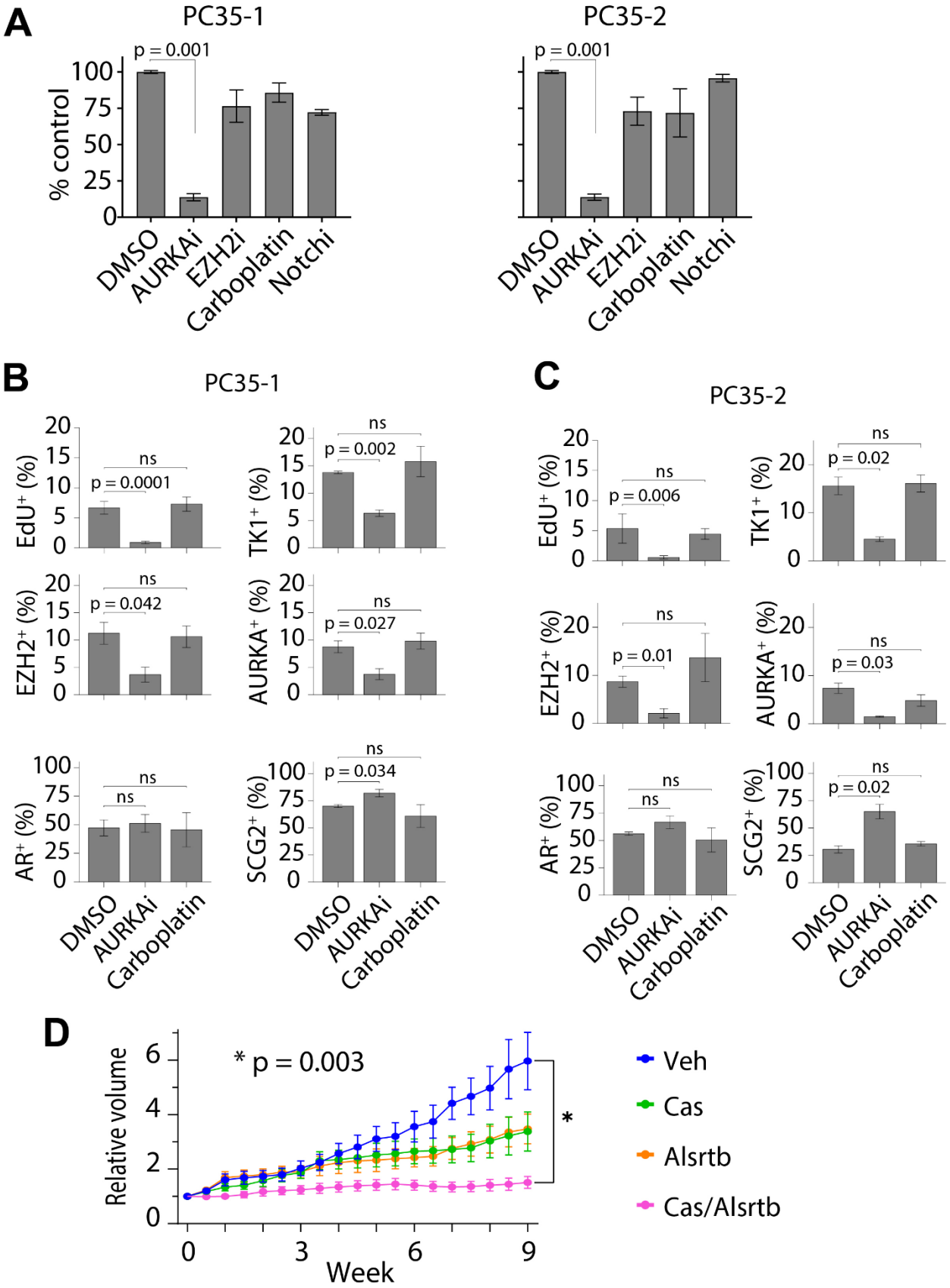
bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Figure 5. PC35 organoids show subpopulation-specific sensitivity to AR inhibition (A) PC35-1 and PC35-2 organoids treated for six weeks with enzalutamide (10 μM). Relative cell numbers were quantified with CellTiter Glo 3D and plotted relative to the control. (B) A representative image of a combination RNA-FISH/EdU assay on PC35-1 organoids. The organoids were pulsed for 24 hours with EdU prior to collection and then dissociated and replated in 2D on cover slips and stained for AR and EdU. (C) PC35-1 and PC35-2 organoids were treated for six weeks with enzalutamide (10 μM). Organoids were pulsed with 10 μM EdU for 24 hours prior to collection, then dissociated and replated in 2D on cover slips and stained for AR expression by RNA-FISH. EdU incorporation status (positive or negative) was determined for each cell (see Methods). The data was plotted as the percentage of EdU-positive cells that also expressed AR in each treatment condition. (D) PC35-1 and PC35-2 organoids were treated as in (C) and stained for SCG2 expression by RNA-FISH. EdU incorporation status (positive or negative) was determined for each cell. The data was plotted as the percentage of EdU-positive cells that also expressed SCG2 in each treatment condition. (E) PC35-2 UMAPs showing Log2 expression of MAP3K5. (F) PC35-2 organoids were treated and stained for marker expression and EdU incorporation as in (C). The data was plotted as the percentage of EdU-positive cells that also expressed MAP3K5 (left), both AR and MAP3K5 (center), or AR but not MAP3K5 (right). (G) PC35-2 UMAPs showing Log2 expression of NR3C1. Bar graphs are plotted as the mean of three (A) or two (C, D, F) independent experiments. Error bars, ± SEM. P-values were calculated using the student's t-test, two-tailed, unpaired.

bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.



bioRxiv preprint doi: https://doi.org/10.1101/2022.03.24.484651; this version posted March 27, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. This article is a US Government work. It is not subject to copyright under 17 USC 105 and is also made available for use under a CC0 license.

Figure 6. The stem-like/progenitor subpopulation is vulnerable to AURKA inhibition (A) Drug assays. Organoids were treated twice weekly for six weeks with 500 nM AURKAi, 500 nM EZH2i, 500 nM carboplatin, 1 μM Notchi, or 0.02% DMSO-treated controls. Quantification was done by dissociating the organoids and manually counting the cells. The quantified values for each condition were plotted relative to the DMSO controls. (B) PC35-1 and (C) PC35-2 organoids treated for six weeks as in (A) with AURKAi, carboplatin, or DMSO, then pulsed with 10 μM EdU, 24 hours prior to collection. The indicated marker expression for each cell was determined by RNA-FISH. EdU incorporation status (positive or negative) was determined for each cell. Data was plotted as the percentage of cells that were positive for a given marker or EdU for the three treatment conditions. (D) Relative change in tumor volume for PC35-1 organoid-derived xenografts (ODXs) during nine weeks of the indicated treatments. Tumor volume was calculated as an average of the replicates. The change in volume was calculated relative to the "0" time-point. Vehicle n = 5 mice; castration n = 5 mice; alisertib n = 4 mice; castration + alisertib n = 5 mice. Bar graphs are plotted as the mean of three independent experiments. Error bars, ± SEM. P-values were calculated using the student's t-test, two-tailed, unpaired.