

# Dynamic prostate cancer transcriptome analysis delineates the trajectory to disease progression.

**Marco Bolis** (✉ [marco.bolis@ior.usi.ch](mailto:marco.bolis@ior.usi.ch))

Institute of Oncology Research

**Daniela Bossi**

Institute of Oncology Research

**Arianna Vallerga**

Institute of Oncology Research

**Valentina Ceserani**

Institute of Oncology Research

**Manuela Cavalli**

Institute of Oncology Research

**Laura di Rito**

Institute of Oncology Research

**Eugenio Zoni**

Oncology Institute of Southern Switzerland

**Simone Mosole**

Institute of Oncology Research

**Andrea Rinaldi**

Institute of Oncology Research

**Ricardo Pereira Mestre**

Oncology Institute of Southern Switzerland

**Eugenia D'Antonio**

Oncology Institute of Southern Switzerland

**Matteo Ferrari**

Ente Ospedaliero Cantonale

**Flavio Stoffel**

Ente Ospedaliero Cantonale

**Fernando Jermini**

Ente Ospedaliero Cantonale

**Silke Gillessen Sommer**

Ospedale San Giovanni Bellinzona

**Lukas Bubendorf**

Institute for Pathology, University Hospital Basel <https://orcid.org/0000-0001-5970-1803>

**Peter Schraml**

University Hospital Zurich

**Holger Moch**

University Hospital of Zurich

**Martin Spahn**

Lindenhofspital Bern, Prostate Center Bern

**George Thalmann**

Bern University Hospital

**Marianna Kruithof-de Julio**

Department of Biomedical Research, University of Bern <https://orcid.org/0000-0002-6085-7706>

**Mark Rubin**

University of Bern <https://orcid.org/0000-0002-8321-9950>

**Jean-Philippe Theurillat**

Institute of Oncology Research

---

**Article**

**Keywords:** prostate cancer, transcriptome analysis, disease progression

**Posted Date:** March 15th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-296396/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Nature Communications on December 1st, 2021. See the published version at <https://doi.org/10.1038/s41467-021-26840-5>.

# Abstract

Comprehensive genomic studies have delineated key driver mutations linked to disease progression for most cancers. However, corresponding transcriptional changes remain largely elusive because of the bias associated with cross-study analysis. Here, we overcome these hurdles and generate a comprehensive prostate cancer transcriptome atlas that describes the roadmap to tumor progression in an unprecedented qualitative and quantitative manner. Most cancers follow a uniform trajectory characterized by upregulation of polycomb-repressive-complex-2, G2-M checkpoints, and M2 macrophage polarization. Using patient-derived xenograft models, we functionally validate our observations and add single-cell resolution. Thereby, we show that tumor progression occurs through transcriptional adaptation rather than a selection of pre-existing cancer cell clusters. Moreover, we determine at the single-cell level how inhibition of EZH2 - the top upregulated gene along the trajectory – reverts tumor progression and macrophage polarization. Finally, a user-friendly web-resource is provided that enables the investigation of dynamic transcriptional perturbations linked to disease progression.

## Introduction

Many decades of research have established the fundamental understanding of cancer as an anarchistic proliferation and dissemination of cells caused by acquired mutations in key driver genes <sup>1</sup>. During the last decade, the most common cancer types have been extensively characterized for alterations in the tumor DNA sequence <sup>2</sup>. While these studies have been initially conducted on primary cancer tissues, more recent clinical studies have also included biopsies from metastatic disease <sup>3-9</sup>. Because of the binary nature of DNA sequence alterations (mutated versus non-mutated), mutation frequencies can be readily compared across studies and enable the nomination of drivers intimately linked to disease progression and outcome <sup>10,11</sup>. That said, the plethora of complex genetic alterations largely complicates a quantitative assessment of the transformed phenotype.

The assessment of gene expression may provide a more complete and quantitative measure of the biological processes related to disease progression. Most transcriptomic studies have been thus far conducted on primary tumors <sup>12,13</sup>. However, multiple efforts have been dedicated in recent years to the characterization of metastatic disease for a few tumor types, including prostate cancer, opening the possibility to assess transcriptional changes along with disease progression in a systematic manner <sup>11,14-17</sup>.

Nevertheless, this approach requires the accurate integration of multiple data sets across studies to overcome the issue of introducing dataset-specific features, often referred to as batch effects. The substantial amount of non-biological artifacts introduced both by RNA sequencing libraries generation techniques and by the exploitation of different quantification algorithms are among the reasons why to date, there are no studies that attempt to nominate a trajectory of prostate cancer disease progression by inferring dynamic transcriptional changes from a large integrated cohort.

Here, we provide a framework to overcome these issues and enable the accurate quantitative integration of RNA sequencing data from over 1000 clinical tissues ranging from normal prostate tissue to primary prostate cancer and metastatic CRPC. The harmonized prostate cancer transcriptome atlas provides a unique resource to mine transcriptional changes related to different disease stages. Using this resource, we characterize the trajectory to disease progression and functionally validate our findings in patient-derived xenograft models at the single-cell level. Finally, we show how our prostate cancer transcriptome atlas can infer or validate new therapeutic avenues for cancer patients.

## Results

### Generation of the Prostate Cancer Transcriptome Atlas

To nominate gene expression changes related to disease progression, we re-processed and integrated high-throughput transcriptional data sets from 13 different studies, constituting thus far the most comprehensive compendium of the disease (Supplementary Fig. 1.A & Supp.Table 1)<sup>11,16-24</sup>. The resulting principal component analysis (PCA) showed that samples' position at a given disease stage largely overlapped with another regardless of their origin. In contrast, samples from distinct disease stages differed in localization (Fig. 1A). An appreciable "batch effect" related to the hybrid capture sequencing technique was detected and subsequently corrected (Supplementary Fig. 1B).

Gene set enrichment analysis (GSEA) of the first two principal components (PC) revealed that PC1 correlated with enhanced proliferation while PC2 anti-correlated with canonical AR-signaling (Supplementary Fig. 1C-D). Moreover, PC3 separated cancers harboring truncal mutations in SPOP & FOXA1 from the ones harboring gene fusions involving ETS family transcription factors (Supplementary Fig. 1E)<sup>25-28</sup>.

### Trajectory analysis quantifies the path to disease progression

We applied trajectory inference analysis to characterize disease progression. The approach identified the path to disease progression and assigned a pseudo-time to each sample that describes the advancements along this specific path (Fig. 1B). Subsequently, we assessed corresponding gene expression changes (Fig. 1C). Among the most up-regulated genes, we noticed key genes encoding for chromatin remodelers, which mediate gene silencing during development, such as DNA methyltransferases (DNMTs) and members of the polycomb-repressive-complex-2 (PRC2)<sup>29</sup>. Most importantly, the PRC2 member EZH2 emerged as the top up-regulated gene, corroborating its previously suggested role in disease progression (Fig. 1C, Supplementary Fig. 1F)<sup>15,30,31</sup>. Besides, among the most up-regulated genes, we noted AR-regulated genes that promote G2-M cell cycle progression, while AR-regulated differentiation genes were suppressed, as expected (Fig. 1D)<sup>32-34</sup>.

The progression path indicates that most prostate cancers evolve from normal tissue by continuously increasing AR-signaling (PC2). Then, under androgen deprivation therapy, the tumors progress to castration-resistant prostate cancer (CRPC) by increasing cell cycle genes and eventually de-differentiate to AR-negative disease with or without neuroendocrine features (NEPC) (Fig. 1E). Notably, the transcriptional changes correlated well with the protein level changes in an independent set of primary and CRPC samples (Fig. 1F)<sup>35</sup>. Because EZH2 was not assessed in this dataset, we ascertained its upregulation with disease progression on a tissue microarray of 33 primary and matched CRPC samples (Supplementary Fig. 1G)<sup>36</sup>.

Next, we evaluated whether genomic alterations in driver genes correlate with disease progression. We noted a significant correlation of point mutations in PIK3CA, TP53, FOXA1, KMT2C, and PTEN with progression in primary tumors and FOXA1 in the metastatic counterpart (Fig. 1G). In primary tumors, we also noticed a positive correlation with MYC copy number and an inverse correlation with deletions of RB1, PTEN, and TP53, as expected. In contrast, in CRPC/NEPC samples, only RB1 loss seemed to correlate well with increased progression (Fig. 1H, Supplementary Fig. 1H, Supplementary Fig. 1I).

Finally, we assessed transcriptional changes in key immune pathways throughout tumor progression along the trajectory. It has been widely appreciated during recent years that cancer growth is supported by changes in the tumor microenvironment, such as the polarization of macrophages from an M1- towards M2-like phenotype<sup>37,38</sup>. Indeed, we noticed a potent downregulation of pro-inflammatory M1 markers and an increased shift towards M2-associated pro-tumorigenic effectors (Fig. 1C,1I,1J). Interestingly, CD24 – a potent “don’t eat me” signal for M1 macrophages – was associated with progression as well<sup>39</sup>.

### **Integration of prostate cancer models in the transcriptome analysis**

We next set out to further functionally validate our findings related to disease progression in eight established human prostate cancer cell lines and six patient-derived xenografts (PDX) models originating either from a surgically, carried off primary prostate cancer (PNPCa)<sup>40</sup> or CRPC (LuCaP-23.1, -35, -78, -145, -147)<sup>41</sup>. To this end, the transcriptional fingerprint of all models clustered towards the outer layer of the progression trajectory (Fig. 2A & Supplementary Fig. 2A, B).

As expected, the PCA positioning of cell lines and the PDX models along the trajectory was mostly associated with the originating disease stage and the dependence on androgens (Supplementary Fig. 2C). The hormone-naïve PNPCa model was placed first, followed by the CRPC-derived models, positioned progressively according to their decreasing levels of AR-dependency. Finally, we observe the AR-negative (PC3, DU145) and neuroendocrine models (NCI-H660, LuCaP-145.2), which are located at the end of the route (Fig. 2A & Supplementary Fig. 2A, B). As expected, we also noted a corresponding upregulation of key proteins related to polycomb complexes (EZH2, SUZ12 EED), DNA methylation (DNMT1, DNMT3A/B), and G2-M cell cycle progression (Fig. 2B).

Multiple castration-resistant sublines of cell lines and PDX models have been generated over the last decades, enabling us to further functionally validate the disease progression trajectory in an isogenic

system<sup>42</sup>. Indeed, we found that all sublines progressed on the trajectory (Fig. 2C & Supplementary Fig. 2D-F). Most notably, the LTL-331 PDX model displayed a gradual transcriptional progression from late-stage primary prostate cancer to AR-negative, neuroendocrine disease within a timeframe of 32 weeks (Fig. 2C & Supplementary Fig. 2D)<sup>43</sup>. At the molecular level, we also noted an increase in key proteins linked to the trajectory in LNCaP xenograft tumors upon tumor recurrence after castration (Fig. 2D). Altogether, the data suggest that progression along the trajectory can be recapitulated in human cell line and PDX models.

The *ex vivo* culture of prostate cancer cells has been traditionally a major challenge. That said, the adjustment of the 3D organoid culture system for prostate cancer has enabled the *ex vivo* culture of PDX-derived cells and the generation of new prostate cancer organoid lines<sup>44,45</sup>. We wondered if the transcriptional output of *ex vivo* cultures would mirror the corresponding PDX models *in vivo*. In general, we found that *ex vivo* organoid cultures displayed a more progressed transcriptional output compared to the corresponding *in vivo* models (Fig. 2E). In agreement, the AR-dependency was also largely diminished (Fig. 2F & Supplementary Fig. 2C). This observation could be further validated when androgen-dependent LNCaP cells in standard 2D were cultured in the 3D organoid condition (Fig. 2F). Of note, the standard 2D culture matched better the corresponding xenograft model concerning the position on the progression trajectory (Fig. 2E). In aggregate, the data may suggest that the advances in culturing prostate cancer cells using the organoid system may come at the expense of transformation towards a more progressed and aggressive androgen-independent state.

### **Single-cell resolution to the trajectory**

We performed single-cell RNA sequencing (scRNAseq) of most aforementioned PDX models *in vivo* to interrogate the individual cells' distribution along the trajectory of disease progression. In each case, normal mouse stromal cells were identified and separated from human tumor cells (Fig. 3A, Supplementary Fig. 3A-D). When comparing the merged single-cell data with the previously generated bulk RNA sequencing data, we noticed in each case an excellent concordance between the position of both data points on the PCA plot, suggesting that our single-cell data is sufficiently similar to allow the integration into the pan-prostate cancer transcriptome cohort (Fig. 3B, Supplementary Fig. 3E-H).

Subsequently, we interrogated each PDX for the existence of separate subpopulations using the *Seurat* workflow<sup>46</sup> (Fig. 3A, Supplementary Fig. 3A-D) and integrated the data into the PCA plot (see Method section). Overall, single cells of the various subpopulations within a given PDX model did not greatly differ in their position to the trajectory and displayed relatively little overlap across PDX models (Fig. 3B, Supplementary Fig. 3E-L). As expected, subpopulations in cell cycle progression (i.e. S and G2M phase) positioned higher on the trajectory (Fig. 3B, Supplementary Fig. 3E-P). That said, the PDX model LuCaP-35 showed a wider distribution of subpopulations along the trajectory with distinct features linked to S and G2M phase (H1-3 versus H4, 6), respectively, raising the possibility of being composed of two major, biologically diverse tumor clones (Supplementary Fig. 3G, K, O).

Subsequently, we assessed if and how these subpopulations would evolve during progression to androgen-independence. For this purpose, we took advantage of the LuCaP-147 PDX tumor model that quickly develops castration resistance and compared the single-cell transcriptional profiles before and after castration (Fig. 3C). Upon regrowth, there was no major difference in the position and abundance of previously identified subpopulations (Fig. 3D). Instead, we noticed a concordant shift along the trajectory for each of the clusters h1-7, which was characterized by a shutdown of canonical AR signaling and upregulation of pro-proliferative MYC target genes, among others (Fig. 3E, F). Altogether, the data suggest that resistance to castration in this setting occurs likely through reprogramming of the entire tumor cell population instead of a clonal selection of a particular cluster.

Subsequently, we wondered if the induction of resistance may be paralleled by changes in the tumor microenvironment. Indeed, after castration, we observed an increase in the abundance of tumor-associated macrophages that displayed a change in polarization from M1- to M2-like features (Fig. 3G, H). In line with this, we also observed a gradual reduction of TNA alpha signaling and inflammatory signatures – key features of M1 macrophages – in PDX models with increasing pseudo-time along the trajectory (Fig. 3I). The results agree with the expression changes of M1- and M2-related transcripts along the trajectory of disease progression described earlier in Fig. 1. Taken together, the data illustrates how bulk transcriptional changes related to disease progression can help to shed light on the emergence of androgen-independent prostate cancer at the single-cell level.

### **Co-targeting AR and EZH2 delays tumor progression**

Because EZH2 emerged as a top-upregulated transcript within the trajectory of disease progression and had been shown to promote androgen-independence<sup>15,30,47,48</sup>, we set out to investigate if co-targeting AR and EZH2 may prevent or substantially delay disease progression. Indeed, we noted a dramatic change in the transcriptional output program of LNCaP cells when treated with the EZH2 inhibitor GSK126 under androgen deprived culture conditions in charcoal-stripped serum (CSS) (Fig. 4A). Previously detected LNCaP subpopulations (h1-6, h8) formed a new subpopulation (h7), suggesting a nearly-complete rewiring of transcription, up-regulation of AR signaling, reduction of E2F-related cell cycle genes, and reversion of progression on the trajectory (Fig. 4B & Supplementary Fig. 4A-C). In line with this, we noticed a strong reduction in colony formation when androgen-dependent LNCaP, VCaP, and LAPC4 cells were subjected to CCS and treated with GSK126, while forced expression of EZH2 was sufficient to promote colony formation in the same setting (Supplementary Fig. 4D).

Next, we tested if our observations would also translate into an *in vivo* setting. For this purpose, we injected LNCaP cells into the flank of immune-compromised mice and treated the emerging xenograft tumors with castration alone or in combination with three weeks of GSK126. In both cases, the tumors fully regressed. While the tumors of castrated mice regrew with a latency of around four weeks, GSK126 co-treated tumors took more than twice as much time to re-initiate tumor growth (Fig. 4C).

We subsequently performed scRNA-seq on the emerging tumors pre- and post-castration to investigate transcriptional changes on tumor and stromal cell subpopulations. As noted previously for LuCaP-147,

we found no major change in the tumor cell subpopulations (i.e., h1-6) that adapted to castration (Fig. 4D). Because GSK126 treatment *in vivo* had been stopped for three months before harvesting the tumors, the transcriptional changes in the tumor cells appeared less striking than in the aforementioned cell culture setting (Fig. 4A, D). That said, we observed after GSK126 co-treatment a relative increase in tumor cell numbers of cluster h6 – the least progressed cluster on the trajectory that also displayed the highest AR mRNA levels (Fig. 4D & Supplementary Fig. 4E-G). This cluster showed a further increase in AR signaling and a reversion of disease progression after GSK126 treatment (Fig. 4E & Supplementary Fig. 4H).

Finally, we assessed if pharmacologic inhibition of EZH2 may also affect the polarization of tumor-associated macrophages. In line with our previous findings, LNCaP xenograft-associated macrophages increased in numbers and displayed a shift towards M2-like polarization in tumors adapted to castration (Fig. 4D, F, G). Strikingly, we found a pronounced reduction of preferentially M2-like macrophages in GSK126-pretreated tumors, suggesting that GSK126-mediated changes on the tumor microenvironment may have contributed as well to the delayed regrowth of LNCaP xenografts (Fig. 4H, I). In aggregate, the data suggest a rationale for joint targeting of AR and EZH2 in prostate cancer because the latter reverts tumor cell progression towards a more androgen-dependent state and at the same time counteracts adaptive changes in macrophages that are intimately linked to disease progression.

## Discussion

In the present study, we combine transcriptional profiles of prostate cancers at various disease stages to a comprehensive prostate cancer transcriptome atlas with negligible study-related interference (i.e. “batch effects”). Mining the atlas reveals a rather uniform trajectory towards disease progression from normal prostate, primary, and metastatic castration-resistant prostate cancer. The trajectory is characterized by a gradual upregulation of genes related to EZH2-mediated polycomb signaling and cell cycle progression, most notably G2M checkpoints and mitotic spindle genes. The latter may provide an explanation why taxanes (i.e. docetaxel, cabazitaxel) which disrupt microtubule function during cell division, remain a cornerstone of prostate cancer treatment in the hormone-sensitive and castration-resistant metastatic setting<sup>49–56</sup>.

EZH2 has been previously described to be critically involved in prostate cancer as an activator of AR signaling<sup>30</sup>. It is also a key component of polycomb repressor complex 2-mediated gene silencing – a developmental pathway implicated in de-differentiation and prostate cancer progression<sup>15,29,57,58</sup>. In agreement with the latter, we find EZH2 the top-upregulated gene in the progression trajectory along with other PRC2 members. In line with a function in driving disease progression and de-differentiation towards the loss of AR expression, we demonstrate how EZH2 inhibition reverts the transcriptional output of prostate cancer cells along the progression trajectory. The findings may have important implications for the treatment of prostate cancer patients in a hormone-naïve or early CRPC because it may prevent the de-differentiation of cancer cells as an escape-mechanism to AR-directed therapeutic interventions.



In line with previous reports, we noticed along the trajectory a change of macrophage polarization from inflammatory M1 to pro-tumorigenic M2<sup>37,38</sup>. Our findings further underscore the anti-tumor potential of pharmacologically re-educating macrophages towards M1. Strikingly, castration was sufficient in our PDX models to increase the number of macrophages and induce a change toward M2-polarization after a relatively short period, suggesting that therapeutic interventions per se may be at least in part the underlying cause. Importantly, in the same setting, inhibition of EZH2 substantially blocked the castration-induced polarization change towards M2, uncovering a thus far underappreciated role for EZH2 in macrophage polarization another rationale towards co-targeting AR and EZH2 in prostate cancer.

It is mostly unknown how disease progression in prostate cancer emerges at the single-cell level. Using a series of PDX models reflecting different progression stages from hormone-naïve to AR-negative late-stage disease enabled the addition of single-cell resolution to the progression trajectory. Our results suggest that resistance to androgen deprivation may occur through transcriptional adaptation of tumor cells towards a more progressed state. In line with this, a recent study has proposed that prostate regeneration (a process that shares many molecular features with prostate cancer progression) is driven by nearly all persisting luminal cells, not just by rare stem cells<sup>59</sup>. That said, in our study, we have used a relatively uniform xenograft tumor model that has been already derived from CRPC and thus adapt swiftly to castration in mice. Conceivably, resistance to androgen receptor inhibition over a longer period may also involve the selection of stem-cell-like subpopulations irrespective of the presence of genetic drivers of CRPC (e.g., AR amplification or point mutations)<sup>60–64</sup>.

We provide a web-based interface for the research community to facilitate the mining of the prostate cancer transcriptome atlas, called the PCaProfiler (<https://www.pcaprofiler.com>). Using this resource, we readily identify, for example, that a subpopulation of very advanced prostate cancer tissues expresses high levels of IL23A, a cytokine recently described to mediate castration resistance in prostate cancer<sup>65</sup>. Interestingly, correlating the IL23A expression with genomic features in our webtool identifies a tight association of IL23A expression with gains and amplification of its receptor IL23R. Such insights may be important for patient selection/stratification for anti-IL23 targeting monoclonal antibodies under clinical development (i.e. NCT04458311).

The PCaProfiler will also allow the pseudo-time annotation of new cancer transcriptomes. In a clinical trial setting, this information may enable identifying anti-tumor responses within a certain subset of patients with a given degree of disease progression. In a preclinical setting, the atlas may also help researchers to choose the corresponding model system that reflects the particular disease stage under investigation. Of note, in this regard, we have already annotated the pseudo-time for the most frequently used prostate cancer cell lines (see PCaProfiler). Alternatively, the PCaProfiler may enable researchers to verify and optimize the *ex vivo* culture condition so that it best mirrors the *in vivo* setting.

In conclusion, we successfully merged the RNA sequencing data from several prostate cancer studies, covering different disease stages. Based on that, we delineate the roadmap to prostate cancer progression in an unprecedented, qualitative, and quantitative manner. Furthermore, we also show how

individual tumor cells can be tracked along the progression trajectory in response to pharmacological perturbations. Because transcriptome data of advanced metastatic disease will become more readily available for other tumor types, the current study may serve as a blueprint for their analysis and exploitation.

## Methods

### DATA AVAILABILITY

The data/analyses presented in the current publication include the use of protected study data downloaded from the dbGaP web site, under phs000915.v2.p2<sup>11</sup>, phs000673.v4<sup>7</sup>, phs000909.v1<sup>72</sup>, phs000424.v8<sup>18</sup>, phs000178.v11<sup>26</sup>. Access to dbGaP was granted for Project #24196, entitled “Unravelling determinants of acquired resistance to hormone therapy in cancers”. Publicly available data were retrieved from GEO: GSE120795<sup>23</sup>, GSE120741<sup>19</sup>, GSE118435<sup>22</sup>, GSE126078<sup>21</sup> and SRA: PRJNA477449<sup>73</sup>, PRJEB21092<sup>74</sup>. Datasets generated in this study were made available at EMBL-EBI. Bulk RNA-Seq data used in the study were deposited with the accession number E-MTAB-9930. Single-cell RNA-Seq data for LuCaP PDX models and LNCaP cells were deposited with the accession number E-MTAB-9903. Processed data in form of h5 files can also be downloaded directly from <https://www.pcaprofiler.com>. All the software used for the analyses is described and referenced in the respective Method Details subsections.

### EXPERIMENTAL MODEL AND SUBJECT DETAILS

#### Plasmids

The pHAGE-puro (Plasmid #118692) and the pHAGE-EZH2 (Plasmid #116738) were purchased from Addgene.

#### Cell Lines

PC3, DU-145, 22rV1, MDA-PCa-2b, LAPC-4, LNCaP, VCaP, and HEK 293T cell lines were purchased from ATCC (American Tissue Culture Collection) (Manassas, USA). The LAPC-4 cell line was a gift from Prof. Helmut Klocker while the LNCaP-abl cell line was a gift from Prof. Myles Brown (DFCI, Boston).

#### Immunohistochemically staining

EZH2 protein expression was analyzed on a previously described tissue microarray including, matched primary and CRPC samples<sup>36</sup>. All prostate cancer samples were obtained under approval by the Ethics Committee of Northwestern and Central Switzerland (EKNZ, No EK/1311 and 2015/228). Tumor-free prostate core needle biopsies were used to analyze benign prostate (n = 3 patients). Prostate cancer biopsies included in the TMA were taken during routine clinical treatment. Samples were selected based on the following inclusion criteria: (a) histologically-diagnosed PCa, (b) tumor-containing biopsies

available at HN and CR state, and (c) sufficient quality and amount of material, as evaluated by experienced pathologists (LB and KM). Castration-resistance was defined as either biochemical progression (ie, serum PSA progression according to *Prostate Cancer Clinical Trials Working Group* criteria or clinical progression. A TMA comprising 112 matched HN/CR tissue specimens, and including 107 transurethral resections and five distant metastases derived from 55 PCa patients was constructed as previously described (Federer-Gsponer JR et al. Prostate 2020). Due to tissue loss, a common problem associated with TMA technology, 33 high-quality matched tissue samples of primary and CRPC remained after sectioning.

For EZH2 IHC, slides were analyzed with the Bond-III automated staining system (Leica) using manufactured reagents for the entire procedure. For antigen retrieval, slides were incubated for 60 min in Citrate buffer at pH6 at 98°C. Thereafter, slides were incubated with a rabbit monoclonal antibody against EZH2 (D2C9, CST5246 from Cell Signaling) at the dilution of 1:500. Detections were performed using the detection refine DAB kit (Leica). Immunohistochemical staining was evaluated as a percentage of tumor cells with nuclear positivity for EZH2 using Aperio ImageScope (Leica).

For the assessment of CD11c protein expression in primary prostate cancer and the correlation with PSA-recurrence patients were selected from two previously characterized tissue microarrays cohorts constructed in Zurich and Bern<sup>75-77</sup>. Due to tissue loss, a common problem associated with TMA technology, a total of 482 high-quality tissue samples of primary tumor remained after sectioning (n = 272 from Zurich and n = 210 from Bern). In each case, the local scientific ethics committees approved (StV-Nr. 25/2007 and StV-Nr. 25–2008) and informed consent was obtained from all patients. Recurrence-free survival curves were calculated using the Kaplan-Meier method. Patients were censored at the time of their last tumor-free clinical follow-up visit. Time to PSA recurrence was selected as the clinical endpoint. Only patients undergoing radical prostatectomy were used for survival analysis.

For CD11c IHC, slides were analyzed with the Bond-III automated staining system (Leica) using manufactured reagents for the entire procedure. For antigen retrieval, slides were incubated for 20 min in Citrate buffer at pH6 at 98°C. Thereafter, slides were incubated with a rabbit anti-CD11c antibody targeting the C-terminus (ab52632) at the dilution of 1:1000 for one hour at room temperature. Detections were performed using the detection refine DAB kit (Leica). Immunohistochemical staining was evaluated with the automated Aperio ImageScope (Leica) image quantification system using a two-tiered score, i.e. tumor spots with at least three percent of CD11c-positive cells were classified as CD11c high, while the remaining cases were classified as CD11c low. In the case of the Bern cohort, multiple spots per tumor were available and the percentage of CD11c-positive tumor cells was established based on the average of two spots displaying the highest Gleason pattern.

## Cell Culture

PC3, DU-145, 22rV1, LAPC-4, and LNCaP cell lines were cultured in RPMI 1640 (21875-034 Life Technologies) supplemented with 10% Fetal Bovine Serum (FBS-11A Capricorn Scientific) and 1%

Penicillin/Streptomycin (15140-122 Life Technologies) with 5% CO<sub>2</sub> at 37°C. LAPC4 was also supplemented with 1 nM DHT.

The LNCaP-abl cell line was cultured in Phenol-Red Free RPMI 1640 (11835063 Life Technologies) containing 10% charcoal-stripped serum (CSS Fetal Bovine Serum, Charcoal Stripped A3382101 Life Technologies) and 1% Penicillin/Streptomycin with 5% CO<sub>2</sub> at 37°C.

VCaP and 293THEK cell lines were cultured in DMEM (61965059 Life Technologies) supplemented with 10% FBS and 1% µg/ml Penicillin/Streptomycin with 5% CO<sub>2</sub> at 37°C.

MDA-PCa-2b cell line was cultured in ATCC-formulated F-12K medium (30-2004) supplemented with 20% FBS, 25 ng/ml Cholera Toxin (C8252 Sigma), 10 ng/ml Epidermal Growth Factor (AF-100-15 Peprotech), 0.005 mM Phosphoethanolamine (P1348 Sigma), 100 pg/ml Hydrocortisone (H0135 Sigma), 45 nM Selenium Acid (211176 Sigma), 0.005 mg/ml Human Recombinant Insulin (I1884 Sigma) and 1% Penicillin/Streptomycin with 5% CO<sub>2</sub> at 37°C.

### **Transfection and Infection**

The HEK 293T cells were transfected with pHAGE (Empty) and PHAGE EZH2 vectors as previously described<sup>78</sup>. Forty-eight hours after the transfection, the viral supernatants were collected and filtered through a 0.45 µm filter. The LNCaP cell line was incubated with viral supernatant and 8µg/ml Polybrene (H9268 Sigma) for 72 h and then selected with 2 µg/ml puromycin (P8833 Sigma) for two weeks. Western blotting was used to verify the protein overexpression of EZH2.

### **Ex vivo culture of PDX**

PDX tumor tissue was cut into small pieces (1–0.5 mm) with a scalpel blade and then digested in Collagenase Type I media solution (200U/ml Cat#SCR103, Millipore) at 37 °C for 45-60 min. After enzymatic dissociation, the cell suspension was passed through a 100 µm cell strainer (11814389001 Roche) to eliminate macroscopic tissue pieces and centrifuged. The cell pellet was then resuspended in 2-volume pf RBC lysis buffer (11814389001 Roche), incubated for 3 min at RT, and, after centrifugation, resuspended in media. Cell suspension of PDX cells was then propagated in 3D vitro culture (<sup>44</sup> and <sup>79</sup>). The cells were embedded in 50% phenol red-free Matrigel (356231 Corning) and plated as a drop in a 96-well-plate (10000 cells/ well) and maintained in the medium 7-10 days.

### **DHT Dose Response Assay**

3D culture of PDXs (see session Ex vivo culture of PDX) or 2D culture of LNCaP cell line (5000 cells/well in 10% CSS medium) were seeded in triplicate in a 96-well plate and subsequently treated with serial dilutions of DHT (concentration range of 0.01nM-30µM). Proliferation was assessed after 7-10 days by Cell-Titre-Glo assay (G9241 Promega) for 3D culture or MTT (Methylthiazolyldiphenyl-tetrazolium bromide) assay (M5655 Sigma) for the 2D culture. For each time point, absorbance (OD, 590 nm) was measured in a microplate reader (Cytation 3 Imaging Reader Biotek).

## Colony Formation Assay in DHT-free medium

VCAP ( $5 \times 10^5$  cell/well), LAPC4 ( $2.5 \times 10^5$  cell/wells), LNCaP ( $2.5 \times 10^5$  cell/wells), or LNCaP over-expressing EZH2 were seeded in triplicate in 6-well plates in a standard medium. After 24-48h, when the cells attached to the plate and formed a confluent layer, the medium was replaced with 10% CSS medium (DHT free medium) with/without 1  $\mu$ M GSK126 and keep in culture until the formation of the colonies (4-6 weeks). The medium/treatment was weekly replaced. At the end time point, the cells were gently washed with PBS, fixed with 0.01% crystal violet and 20% of EtOH for 30 min, and then wash out with water. The images' colonies were acquired using the Fusion Solo IV LBR system and the quantification of colonies was performed by ImageJ software.

## Animal Experiments

All animal experiments were carried out according to the Swiss Veterinary Authority (TI-42-2018 and TI-10-2010). All *in vivo* studies were used 6-8 weeks old male NRG (NOD-*Rag1*<sup>null</sup> *IL2rg*<sup>null</sup>, NOD rag gamma) mice. Patient-derived xenografts (PDX) LuCaP-147, -145.2,-78, -35 -23.1 were provided by Dr. Eva Corey<sup>41</sup>. Dr. Marianna Kruithof-de Julio provided PNPCa. PDXs tumors were maintained by subcutaneous implantation of matrigel-embedded tumor fragment (1-2-mm average diameter tumor or take rate varied from 1 to 6 months. For the experiment in castration LNCaP or LuCaP-147 cells (obtained from tumor dissociation see details in a session of ex vivo PDX culture) were suspended in PBS and 50% Matrigel and subcutaneously injected into the dorsal flanks of the mice ( $2 \times 10^6$  cells/mouse). Tumor growth was recorded using a digital caliper, and tumor volumes were calculated using the formula  $(L \times W^2) / 2$ , where  $L$ =length and  $W$ =width of the tumor. Tumor volume was measured 2 times per week. When the tumor reached the dimension of 50-100 mm<sup>2</sup>, mice were surgically castrated. For the GSK126 treatment, the mice were treatment one week after castration by daily i.p. at the dose of 100mg/Kg for 3 weeks. At the end of the experiment, mice were euthanized, tumors explanted, and used for the molecular assessment.

## Antibodies and Western Blot Analysis

Primary antibodies used: anti-GAPDH (sc-47724 Santa Cruz), anti-AR (sc-7305 Santa Cruz), anti-DNMT3A (sc-365769 Santa Cruz), anti- EZH2 (612667 BD Transduction Laboratory), anti-DNMT(15032S Cell Signaling Technologies), anti-EED (85322S Cell Signaling Technologies), anti-SUZ12 (3737S Cell Signaling Technologies), anti-Aurora A (14475T Cell Signaling Technologies), antiH3K27me3 (9733S Cell Signaling Technologies), anti- PLK1 B290751 Biologend, anti-H3K9me2 (ab1220 Abcam), antiH3K27ac (ab4729 Abcam) and H3K4me3 (ab6000 Abcam).

Tumor Tissues (25-30 mg) or cellular pellet were lysates with RIPA Buffer supplemented with cocktail phosphatase inhibitors (4906845001 Roche) and proteases inhibitors (5892953001 Roche). Protein concentration was determined by BCA reagent (A52255 Thermo Fisher Scientific), 30-50  $\mu$ g of whole protein lysate were separated on 8-12% SDS–polyacrylamide gels and transferred onto PVDF membrane (88518 Thermo Fisher Scientific). The membranes were blocked with 5% milk in Tris Buffered Saline with

Tween 20 (TBST) for 30 min at RT, incubated overnight at 4°C with primary antibodies, and incubated for 1h at RT with secondary antibodies (anti-rabbit IgG HRP W401B and anti-mouse IgG HRP W402B Promega). The protein bands were visualized using the western bright quantum reagent (K-12042-D20 Advansta) and quantified using the Fusion Solo IV LBR system.

### **RNA Extraction for RNA-seq analysis**

According to the manufacturer's guidelines, the RNA extraction was performed from PDXs frozen fragment (25-30 mg) of cellular pellet using RNeasy kit (74106 Qiagen). The RNAs were processed using the NEB Next Ultra II Directional Library prep Kit for Illumina (E7765 NEB) and sequenced on the Illumina NextSeq500 with single-end, 75 base pair long reads.

### **Single-cell isolation for scRNA sequencing**

To perform scRNA-seq PDX tumor tissue, they were dissociated into single cells as described above (see session Ex vivo culture of PDX). After resuspension in PBS, single-cell suspensions were loaded into a 10x Chromium Controller (10x Genomics, Pleasanton, CA, USA), aiming for 10000-5000 cells, with the Chromium Next GEM Single Cell 3' v3.1 reagent kit (PN-1000121 10x Genomics), according to the manufactured instructions.

### **RNA-Seq Data processing**

#### **Sequencing of Xenografts, 2D and 3D cultures**

We retrieved bulk RNA-Seq data for cellular models of prostate cancer from various available datasets and extended these by performing bulk RNA-Seq of several prostate-cancer Xenografts models (i.e. PNPcA; LuCaP-78; LuCaP-23; LuCaP-35; LuCaP-145; LNCAP), and their derived 3D cultures. Additional sequencing was performed for 2D cultures of LNCaP, LNCaP-all, LAPC4, and VCaP cells. (See data availability section)

#### **Prostate Cancer Transcriptome Atlas**

To build an integrated resource of transcriptional features representing all stages of prostate cancer progression, we collected raw sequencing data from a large panel of independent datasets. We gathered raw data for 1223 clinical samples (1104 excluding technical replicates, 1044 excluding multiple metastatic sites derived from the same individual). The resulting integrated cohort is representative of various stages of disease progression, namely, normal prostate specimens (n=174), primary tumors (n=714), castration-resistant prostate cancers (n=316), and castration-resistant prostate cancers showing features of neuroendocrine trans-differentiation (n=19). Raw sequencing files were retrieved from following sources: 1) Gene Tissue Expression Database (GTEx); 2) The Cancer Genome Atlas (TCGA); 3) Atlas of RNA sequencing profiles of normal human tissues (GSE120795); 4) Integrative epigenetic taxonomy of primary prostate cancer (GSE120741); 5) Prognostic markers in locally advanced lymph node-negative prostate cancer (PRJNA477449); 6) The Long Noncoding RNA Landscape of

Neuroendocrine Prostate Cancer and its Clinical Implications (PRJEB21092); 7) Integrative Clinical Sequencing Analysis of Metastatic Castration Resistant Prostate Cancer Reveals a High Frequency of Clinical Actionability (PRJNA283922; dbGaP: phs000915); 8) CSER - Exploring Precision Cancer Medicine for Sarcoma and Rare Cancers (PRJNA223419; dbGaP: phs000673); 9) Molecular Basis of Neuroendocrine Prostate Cancer (PRJNA282856; dbGaP: phs000909); 10) Heterogeneity of Androgen Receptor Splice Variant-7 (AR-V7) Protein Expression and Response to Therapy in Castration Resistant Prostate Cancer (CRPC) (GSE118435); 11) Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer (PRJNA520923; GEO: GSE126078). Depending on the specific dataset considered, *fastq* files were downloaded either by using *gdc-client* (TCGA) or *sra-toolkit* (SRA, dbGaP). Detailed information along with all available clinical annotations are provided in Supp.Table1.

### **RNA-seq data processing of clinical samples**

The overall quality of sequencing reads was evaluated using *FastQC* (Andrews S., 2010). Sequence alignments to the reference human genome (GRCh38) were performed using STAR (v.2.6.1c) in two-pass mode. Gene-expression was quantified at the gene level by using the comprehensive annotations made available by *Gencode* (v29 GTF-File). Strand-specific information was not maintained to avoid technical differences between stranded and unstranded libraries. Samples were adjusted for library size and normalized with the variance stabilizing transformation (*vst*) in the R statistical environment using DESeq2 (v1.28.1) pipeline. When performing differential expression analysis between groups we applied the embedded *IndependentFiltering* procedure to exclude genes that were not expressed at appreciable levels in most of the samples considered. If not otherwise specified, all gene set enrichment analyses were performed using the *limma* package (*Camera*, use. ranks set to TRUE)<sup>70</sup>. Gene-Sets collections were retrieved either from the Molecular Signature Database (MSigDB, or from previous publications (AR/NE-Score)<sup>80</sup>.

### **Batch effects correction and Principal Component Analysis**

In the processes of integrating different datasets from a variety of sources, we verified that batch effects did not overwhelm the biological signal. Batch effects may derive not only from differences across datasets, but also may be consequent of a different sequencing technique (PolyA+; TotalRNA; Hybrid Capture Sequencing) or originate from other unknown sources. Principal component analysis (PCA), by identifying the transcriptional features endowed with the highest variance across samples, is a very useful tool to detect relevant batch effects. When the latter are overwhelming, they are likely to appear among the top principal components and cluster together samples sharing the same batch effect-related features. A PCA analysis performed on the complete set of 1223 samples (Figure S1B) showed that the largest source of batch effects was associated with the Hybrid Capture Sequencing technique (HCS), while no relevant differences could be clearly associated with the dataset of origin. Only two of the CRPC datasets (phs000915, phs000673) contained samples sequenced using HCS, and for several of these, matched technical replicates sequenced using PolyA+ technology were also available. This allowed us to

assess and remove technology-associated bias in gene expression (ComBat, PolyA+ samples set as reference batch). We further reduced the possibility of confounding biological with technical variation by generating a training-subset of our data, consisting of 883 PolyA+ samples (52 Normal prostate, 620 Primary tumors, 193 CRPCs, 19 NEPCs) and determined the top 2000 genes showing the highest amount of variation within the PolyA+ training set only. This way, for PCA representation we avoid the selection of genes that are possibly affected by the sequencing technique, despite the correction we had already performed on the data.

Hence, we used the same 2000 genes to generate a PCA plot computed on the extended set of samples. The results depicted in the PCA plot shown in Figure1A clearly show that the positioning of tumors at the same stages of cancer progression overlap with each other irrespectively of the dataset of origin and the sequencing technology. This indicates that the different positioning of normal prostate, primary tumors, CRPCs, and NEPCs is due to a real biological signal and not consequent to an unwanted dataset-specific batch effect.

### **Integration and validation of additional bulk RNA-Seq samples and pseudo-time inference**

We developed a method to include new prostate tumor samples in our current analysis by starting from raw counts, which allows the computation of pseudo-time and Principal components without modifying the original data and plots. Ideally, RNA-Seq should be quantified using the sample genome (hg38) and references used for the current study (Gencode V29). Predictions can be performed sequentially, one sample at a time. For each new sample of interest, raw counts will be merged with the ones composing our full set. The obtained numeric matrix (the original matrix + 1 extra sample of interest) undergoes the same normalization and processing steps up to the computation of the PCA. Here, coordinates may slightly differ from the original ones, due to the adding of a new sample which might exert a small effect on the global re-normalization of all samples. To address this behavior, we apply a machine learning-based approach that generates at runtime three elastic net models, one for each of the top 3 principal components, and train them to predict the error between the original coordinates and ones that are recomputed following the addition of the extra sample of interest. Hence, we apply these models to adjust the computed PC1, PC2 and PC3 coordinates of the extra sample which can now be added to the PCA plot and pseudo-time can be determined using slingshot.

### **Trajectory analysis**

Trajectory and pseudo-time inference are frequently used in single-cell RNA sequencing data analysis to model developmental trajectories through smooth curves following dimensionality reduction and clustering. Here we applied one of these tools, slingshot (v1.6.0), to infer progression-associated trajectory and pseudo-time from our integrated set of bulk-RNA sequencing samples. We selected slingshot because of its capability to also determine branches along the trajectory if any. PCA positioning (PC1-PC2) of the individual samples was used as input for slingshot, along with the information that the computed trajectory had to start from the Normal tissue cluster. The analysis was performed using 1106 samples, discarding all technical replicates, in order not to overweight some samples and influence the



computation of the trajectory. Metastatic lesions from the same individual but localized in different organs were admitted for this analysis. Subsequently, we could associate a pseudo-time for each sample, ranging from 0 to 250 (Figure 1B).

### **Correlation of genes and pathways to pseudo-time**

Having defined a unique pseudo-time value for each sample, we computed the correlation between pseudo-time and mRNA expression for each gene. For this purpose, we used Pearson's correlation over Spearman's because we aimed at identifying the strength of the linear relationship between gene expression and pseudo-time. However, to be more robust to outliers, we opted for 10 times repeated leave one third out procedure. Precisely, we randomly selected 10 subsets composed of 66% of the samples and computed correlation coefficients between pseudo-time and expression of each gene in all subsets. Finally, we averaged these values and ranked them according to their correlation coefficient to pseudo-time. Subsequently, using this ranking we applied Camera to perform gene-set enrichment analysis procedure (use.ranks = TRUE) and determined which gene-set were mostly directly or inversely associated with pseudo-time (Figure S1F).

### **Correlation of mRNA expression and protein abundances**

Proteomics data were retrieved from the Proteomics Identifier Database (PRIDE: projects PXD009868, PXD003430, PXD003452, PXD003515, PXD004132, PXD003615, PXD003636). The dataset includes 28 gland confined prostate tumors and 8 adjacent non-malignant prostate tissue obtained from radical prostatectomy procedures, plus 22 bone metastatic prostate tumors obtained from patients operated to relieve spinal cord compression. To compute the correlation between mRNA expression and protein abundance we first computed, for each gene, the average Fold-change (log2) between CRPC and PRIMARY tumors based on mRNA expression. Then the same was applied to the proteomics data to obtain for each protein a log fold change representing differential abundance between CRPCs and primary tumors. For protein/mRNA correlation purposes, we discarded all genes that had not been evaluated in the proteomic data. Finally, we used Pearson's method to evaluate the strength of correlation and the associated statistical significance.

### **Retrieval of genetic information and correlation with progression**

Matched genetic information respective to mutations and copy number status could be retrieved for 763 samples through cBioportal. Samples for which this information was available are indicated in Supp.Table1. To determine associations between mutations and tumor progression, for each gene we compared the pseudo-time of mutant vs wild-type samples, by performing statistical testing using the Wilcoxon-sum rank test. Mutations were ordered according to their False Discovery Rate adjusted P-values and analyses were performed separately in PRIMARY and CRPC+NEPC tumors, to determine the relative contribution of mutations at various stages of disease progression. We only screened for genes being mutated in more than 5 individuals (Figure S1L). To determine associations between copy-number alterations and tumor progression, we associated for each gene a value of either -2 (homozygous

deletion), -1 (heterozygous deletion), 0 (Wild-Type), 1 (Gain), 2(Amplification), and subsequently computed Pearson's correlation between these values and pseudo-time. We restricted this last analysis to genes being frequently deleted or amplified in prostate tumors, namely, MYC, AR, RB1, PTEN, and TP53 (Figure 1E). The above-described analyses were performed discarding technical replicates. Metastatic lesions from the same individual but localized in different organs were admitted for this analysis.

### **Quantification of immune infiltrates and correlation with progression**

Quantification of immune infiltrates for all samples in our cohort was inferred from transcriptomic data using *CibersortX*<sup>81</sup> by using the default signature matrix "LM22" to deconvolve 22 immune cell subsets from bulk RNA-Seq (Absolute quantification mode). The abundance of inferred immune populations was correlated to pseudo-time using the same strategy applied to correlate gene-expression and pseudo-time. We opted for 10 times repeated leave one-third out procedure. Precisely, we randomly selected 10 subsets composed of 66% of the samples and computed correlation coefficients between pseudo-time and each immune population in all subsets. Finally, we averaged these values and ranked them according to their correlation coefficient to pseudo-time. Pearson's correlation-associated P-Values were corrected for multiple testing using the False Discovery Rate (FDR).

### **Macrophage Polarization Index**

The Macrophage Polarization Index, indicating polarization towards M1 or M2 phenotypes was computed for all bulk-RNA samples in our cohort using MacSpectrum<sup>71</sup>.

## **SINGLE-CELL RNA-SEQ DATA PROCESSING**

### **Quantification of gene expression**

*Fastq* files were generated by demultiplexing raw data using *cellranger mkfastq* (v3.1.0) To make single-cell gene-expression quantification more comparable to those of bulk RNA-Seq, we generated a custom genome with *cellranger more*, using the very same reference (GRCh38.p12) and annotations (*encode* v29).used for *STAR* when performing bulk RNA-Sequencing analysis. To discriminate between human and murine cells that may infiltrate the tumors in the in vivo setting, we created a Mouse-Human reference, by creating a hybrid genome (GRCh38.p12+GRCm38.p6) and hybrid gene-annotations (*gencode* v29 and M25, for human and mouse genes respectively). To avoid conflicts, mouse genomic coordinates were preceded by a prefix (i.e. mm\_chr1, mm\_chr2, etc.). Subsequently, *cellranger count* was used to quantifying gene-expression in form of an h5 filtered matrix where Ensembl gene IDs are used as identifiers.

### **Data filtering and clustering**

Expression quantification files were imported in R statistical environment using Seurat (v3.1.5) package. We discarded individual cells from our data matrix by using two filtering procedures: first, we aimed at detecting transcriptional outliers, second, we looked for putative doublets, which we also discarded.

Briefly, we computed per-cell quality control metrics using *scatter* (v1.16.1). The total amount of mitochondrial and ribosomal gene expression was quantified for both human and mouse cells. The number of genes being detected per cell, the total amount of reads per cell, and the mitochondrial and ribosomal fraction of the transcriptome were used to determine the skewness-adjusted multivariate outlyingness for each cell (*robustbase* v0.93-6). Outliers were detected by median absolute deviation (MAD) and removed at both tails. Counts were then normalized (*Seurat::NormalizeData*, method = LogNormalize, scale.factor = 1000) and the top 2000 most variable features were selected (*Seurat::FindVariableFeatures*, method = vst). Data were then scaled (*Seurat::ScaleData*) and principal component analysis was performed up to the top 50 components (*Seurat::RunPCA*). Subsequently, we identified and eliminated putative doublets using *DoubletFinder* (v2.0.3). Having identified outliers and doublets, we removed them from the original count data and went through the pre-processing step again (i.e. normalization, scaling, and pca-reduction). We proceeded to the determination of the k-nearest neighbors of each cell and the construction of a Shared Nearest Neighbor (SNN) Graph (*Suerat::FindNeighbors*), then we identified clusters using the shared nearest neighbor (SNN) modularity optimization based clustering algorithm (*Seurat:: FindClusters*, resolution = 0.5). Finally, we performed Umap dimensionality reduction on the first 10 Principal Components, annotated the previously identified clusters, and generated plots accordingly.

## Identification of Cell-Cycle Phase and Cell-Type

We retrieved the list of cell cycle markers<sup>82</sup> and subdivided it into markers of G2/M phase or S phase, according to Seurat's annotations. We then used this information to infer the cell cycle phase in our samples (*Seurat::CellCycleScoring*). Murine cells could be clearly distinguished from human cancer cells, because of the intrinsic differences that could be easily spotted thanks to the alignment and quantification performed using a hybrid human-mouse genome. Murine cell types were identified using SingleR (v1.2.4)<sup>66</sup>, using ImmGen repository<sup>67</sup>.

## Dealing with Drop-out events

Drop-out events are very frequent in the single-cell experiment performed using chromium 10x technology. To address these issues, we applied Markov Affinity-based Graph Imputation of Cells (RMagic v2.0.3)<sup>68</sup>.

## Differential expression analysis and gene-set enrichment

Differential expression was performed between different cell-clusters and between clusters subjected to different treatment conditions (*Seurat::Findmarkers*) using a hurdle model tailored to scRNA-seq data (MAST method). Genes were subsequently ranked for log<sub>2</sub> Fold-Change and the Camera algorithm (pre-ranked) was used to determine gene-set enrichments for each comparison. Cell-specific gene-set enrichments were determined using single-sample GSEA, computed using gene-expression values of each cell following RMagic imputation.

## Macrophage Polarization Index of macrophages

The Macrophage Polarization Index, indicating polarization towards M1 or M2 phenotypes was computed for all cells being identified as macrophages from SingleR analysis (<https://macspectrum.uconn.edu>).

## Macrophage Reclustering

We could identify a sustained number of murine macrophages infiltrating all xenograft models, except for PNPc cells. We isolated them and performed a cell-type-specific analysis by repeating all previously described processing steps (i.e. normalization, scaling, and pca-reduction). Dropout events were addressed using RMagic, and cell-specific enrichments were computed using a single sample GSEA.

## Integration of scRNA-Seq with bulk-RNA samples, PCA, and pseudo-time inference

Single-cell experiments can be easily integrated with bulk-RNA experiments by simply summing up together gene-counts for all individual cells into one meta-element. This has proven to be extremely comparable in terms of pseudo-time inference and PCA positioning, as scRNA-Seq and bulk RNA-Seq experiments performed on the same cells are overimposable to each other. The same applies for the integration of single-cell derived clusters, provided that the number of cells composing each cluster is not so critically low that the number of drop-out events results in a matrix composed of too many missing genes. If this is the case, or if just a single cell is to be integrated into the analysis, we suggest running RMagic to deal with the drop-out events, and then simply proceed as previously described.

## QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification methods and statistical analysis methods for \*\* were mainly described and referenced in the respective Method Details subsection. If not otherwise specified, all statistical tests were corrected for multiple comparisons using the false discovery rate (FDR) correction method.

## ADDITIONAL RESOURCES

### PCAProfiler

We provide a resource for the research community endowed with a web-based interface to facilitate the mining of the prostate cancer transcriptome atlas, called the PCaProfiler (<https://www.pcaprofiler.com>). Using this resource, scientists can easily interrogate the atlas, recapitulates the findings shown in this study, and extend these by exploiting correlations between genes of interest and prostate cancer progression. PCaProfiler will allow integration and pseudo-time inference of new cancer transcriptomes that the user can directly upload, compute and visualize on the server. All results can be downloaded and re-uploaded to PCA-Profiler when needed. Pre-loaded are PCA-positioning and Pseudo-time inferences of cell-line, xenografts, and organoid models, as well as single-cell clusters and additional transcriptional

datasets not included in the current study (i.e. PRJEB25542, ESCAPE Trial). PCAPProfiler will be updated frequently with new data as new samples are being released or under specific requests.

## Declarations

### Acknowledgments

The results shown here are in whole or part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. J.P.T is funded by a Swiss National Science Foundation Professorship (PP00P3\_150645 & PP00P3\_179072) grant, and grants by the Swiss Cancer League, the Lega Ticinese contro il cancro, and the Fidinam and Barletta Foundation. Thank the CINECA Server Bologna for enabling this study.

### Author Contributions

M.B. and J.P.T. originally developed the concept, further elaborated on it, and designed the experiments together with D.B. and A.V.; D.B., M.C., A.R., and V.C. performed experiments and analyzed the experimental data. The TMAs for this study have been provided by L.B and IHCs have been performed by S.M. and analyzed by J.P.T and D.B. The PNPcCa PDX model has been provided by M.K.J. Prostate cancer tissues and clinical input has been provided by R.P.M, M.F., F.S., F.J., S.G, H.M., P.S., M.S., G.T.

### Declaration of Interests

MAR is listed as a co-inventor on the US and International patents in the diagnostic and therapeutic fields of ETS gene fusion prostate cancers (Harvard and the University of Michigan) and SPOP mutations (Weill Cornell Medicine). JPT has received funding for the venue of scientific conferences from Astellas, MSD, and Janssen/Cilag. SG (last 3 years): Honoraria—Janssen Cilag; Consulting or Advisory role (including IDMC) —Astellas Pharma, Amgen, Roche, Pfizer, AAA International, Janssen, Innocrin Pharma Inst, Sanofi, Bayer, Orion Pharma GmbH, Clovis Oncology, Menarini Silicon Biosystems, Tolero Pharmaceuticals, and MSD; patents, royalties, other intellectual property—Method for biomarker WO2009138392; Travel grant—ProteoMediX; Other relationship—Aranda.

The remaining authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.P.T. (JeanP\_Theurillat@ior.ios.ch)

## References

1. Vogelstein, B. *et al.* Cancer genome landscapes. *Science***339**, 1546–1558, doi:10.1126/science.1235122 (2013).
2. Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell***153**, 17–37, doi:10.1016/j.cell.2013.03.002 (2013).

3. Beltran, H. *et al.* Whole-Exome Sequencing of Metastatic Cancer and Biomarkers of Treatment Response. *JAMA Oncol***1**, 466–474, doi:10.1001/jamaoncol.2015.1313 (2015).
4. Yaeger, R. *et al.* Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. *Cancer cell***33**, 125–136 e123, doi:10.1016/j.ccell.2017.12.004 (2018).
5. Janjigian, Y. Y. *et al.* Genetic Predictors of Response to Systemic Therapy in Esophagogastric Cancer. *Cancer discovery***8**, 49–58, doi:10.1158/2159-8290.CD-17-0787 (2018).
6. Zehir, A. *et al.* Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nature medicine***23**, 703–713, doi:10.1038/nm.4333 (2017).
7. Robinson, D. R. *et al.* Integrative clinical genomics of metastatic cancer. *Nature***548**, 297–303, doi:10.1038/nature23306 (2017).
8. Morris, L. G. T. *et al.* The Molecular Landscape of Recurrent and Metastatic Head and Neck Cancers: Insights From a Precision Oncology Sequencing Platform. *JAMA Oncol***3**, 244–255, doi:10.1001/jamaoncol.2016.1790 (2017).
9. Lefebvre, C. *et al.* Mutational Profile of Metastatic Breast Cancers: A Retrospective Analysis. *PLoS Med***13**, e1002201, doi:10.1371/journal.pmed.1002201 (2016).
10. Armenia, J. *et al.* The long tail of oncogenic drivers in prostate cancer. *Nature genetics***50**, 645–651, doi:10.1038/s41588-018-0078-z (2018).
11. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell***161**, 1215–1228, doi:10.1016/j.cell.2015.05.001 (2015).
12. Demircioglu, D. *et al.* A Pan-cancer Transcriptome Analysis Reveals Pervasive Regulation through Alternative Promoters. *Cell***178**, 1465–1477 e1417, doi:10.1016/j.cell.2019.08.018 (2019).
13. Ma, X. *et al.* Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature***555**, 371–376, doi:10.1038/nature25795 (2018).
14. Dhanasekaran, S. M. *et al.* Delineation of prognostic biomarkers in prostate cancer. *Nature***412**, 822–826, doi:10.1038/35090585 (2001).
15. Varambally, S. *et al.* The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature***419**, 624–629, doi:10.1038/nature01075 (2002).
16. Kumar, A. *et al.* Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. *Nature medicine***22**, 369–378, doi:10.1038/nm.4053 (2016).
17. Beltran, H. *et al.* Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nature medicine***22**, 298–305, doi:10.1038/nm.4045 (2016).
18. Consortium, G. T. The Genotype-Tissue Expression (GTEx) project. *Nature genetics***45**, 580–585, doi:10.1038/ng.2653 (2013).
19. Stelloo, S. *et al.* Integrative epigenetic taxonomy of primary prostate cancer. *Nature communications***9**, 4900, doi:10.1038/s41467-018-07270-2 (2018).

20. Lapuk, A. V. *et al.* From sequence to molecular pathology, and a mechanism driving the neuroendocrine phenotype in prostate cancer. *J Pathol***227**, 286–297, doi:10.1002/path.4047 (2012).
21. Labrecque, M. P. *et al.* Molecular profiling stratifies diverse phenotypes of treatment-refractory metastatic castration-resistant prostate cancer. *J Clin Invest***129**, 4492–4505, doi:10.1172/JCI128212 (2019).
22. Sharp, A. *et al.* Androgen receptor splice variant-7 expression emerges with castration resistance in prostate cancer. *J Clin Invest***129**, 192–208, doi:10.1172/JCI122819 (2019).
23. Suntsova, M. *et al.* Atlas of RNA sequencing profiles for normal human tissues. *Sci Data***6**, 36, doi:10.1038/s41597-019-0043-4 (2019).
24. Abida, W. *et al.* Genomic correlates of clinical outcome in advanced prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America***116**, 11428–11436, doi:10.1073/pnas.1902651116 (2019).
25. Barbieri, C. E. *et al.* Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nature genetics***44**, 685–689, doi:10.1038/ng.2279 (2012).
26. Cancer Genome Atlas Research, N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell***163**, 1011–1025, doi:10.1016/j.cell.2015.10.025 (2015).
27. Shoag, J. *et al.* SPOP mutation drives prostate neoplasia without stabilizing oncogenic transcription factor ERG. *J Clin Invest***128**, 381–386, doi:10.1172/JCI96551 (2018).
28. Bernasocchi, T. *et al.* Dual functions of SPOP and ERG dictate androgen therapy responses in prostate cancer. *Nature communications***12**, 734, doi:10.1038/s41467-020-20820-x (2021).
29. Gaytan de Ayala Alonso, A. *et al.* A genetic screen identifies novel polycomb group genes in *Drosophila*. *Genetics***176**, 2099–2108, doi:10.1534/genetics.107.075739 (2007).
30. Xu, K. *et al.* EZH2 oncogenic activity in castration-resistant prostate cancer cells is Polycomb-independent. *Science***338**, 1465–1469, doi:10.1126/science.1227604 (2012).
31. Yu, J. *et al.* An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer cell***17**, 443–454, doi:10.1016/j.ccr.2010.03.018 (2010).
32. Wang, Q. *et al.* Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell***138**, 245–256, doi:10.1016/j.cell.2009.04.056 (2009).
33. Pomerantz, M. M. *et al.* The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nature genetics***47**, 1346–1351, doi:10.1038/ng.3419 (2015).
34. Pomerantz, M. M. *et al.* Prostate cancer reactivates developmental epigenomic programs during metastatic progression. *Nature genetics***52**, 790–799, doi:10.1038/s41588-020-0664-8 (2020).
35. Iglesias-Gato, D. *et al.* The Proteome of Prostate Cancer Bone Metastasis Reveals Heterogeneity with Prognostic Implications. *Clinical cancer research: an official journal of the American Association for Cancer Research***24**, 5433–5444, doi:10.1158/1078-0432.CCR-18-1229 (2018).
36. Federer-Gsponer, J. R. *et al.* Patterns of stemness-associated markers in the development of castration-resistant prostate cancer. *The Prostate***80**, 1108–1117, doi:10.1002/pros.24039 (2020).

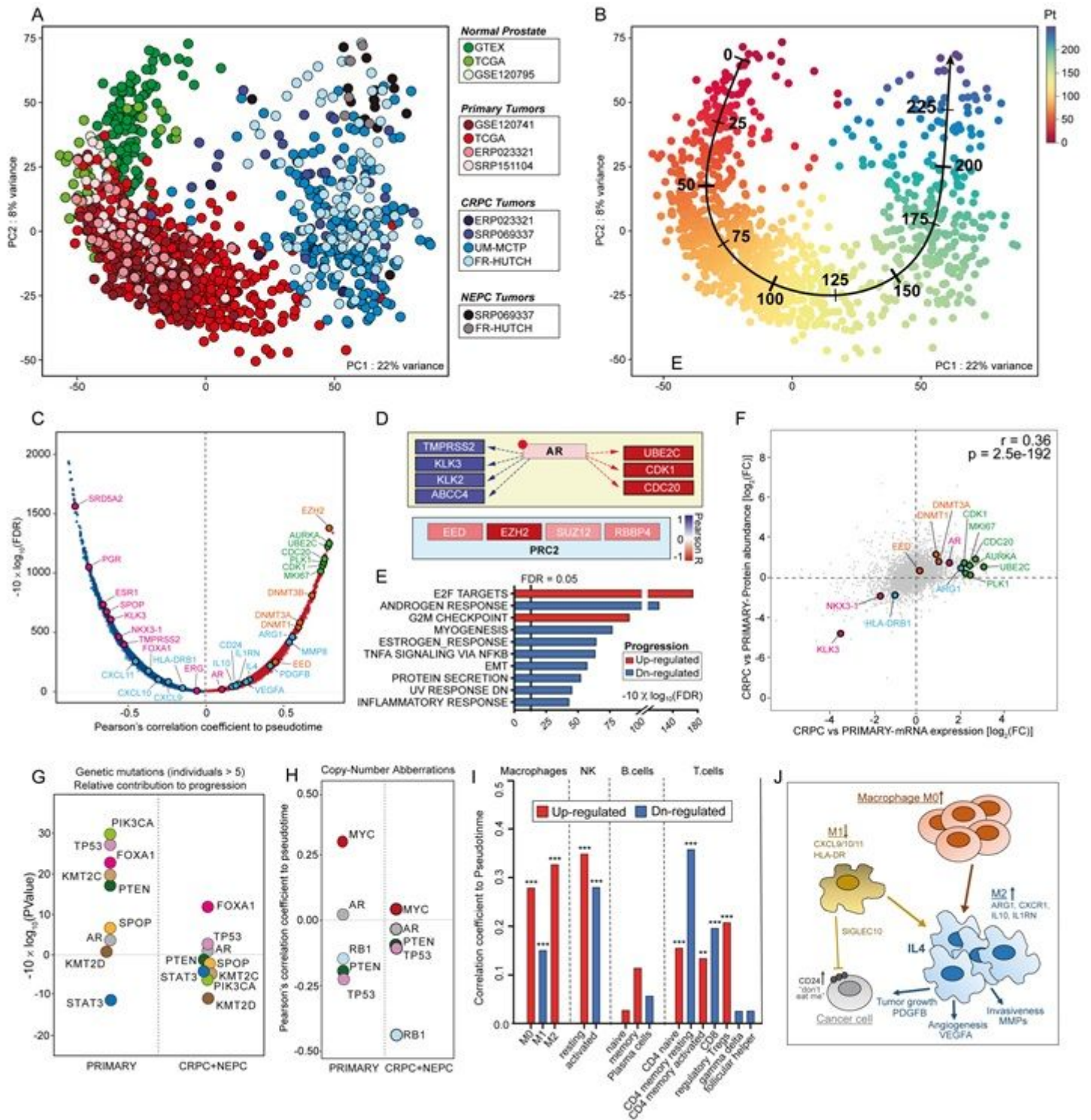
37. Di Mitri, D. *et al.* Re-education of Tumor-Associated Macrophages by CXCR2 Blockade Drives Senescence and Tumor Inhibition in Advanced Prostate Cancer. *Cell reports***28**, 2156–2168 e2155, doi:10.1016/j.celrep.2019.07.068 (2019).
38. Kowal, J., Kornete, M. & Joyce, J. A. Re-education of macrophages as a therapeutic strategy in cancer. *Immunotherapy***11**, 677–689, doi:10.2217/imt-2018-0156 (2019).
39. Barkal, A. A. *et al.* CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. *Nature***572**, 392–396, doi:10.1038/s41586-019-1456-0 (2019).
40. Karkampouna, S. *et al.* Patient-derived xenografts and organoids model therapy response in prostate cancer. *Nature communications***12**, 1117, doi:10.1038/s41467-021-21300-6 (2021).
41. Nguyen, H. M. *et al.* LuCaP Prostate Cancer Patient-Derived Xenografts Reflect the Molecular Heterogeneity of Advanced Disease and Serve as Models for Evaluating Cancer Therapeutics. *The Prostate***77**, 654–671, doi:10.1002/pros.23313 (2017).
42. Pauli, C. *et al.* Personalized In Vitro and In Vivo Cancer Models to Guide Precision Medicine. *Cancer discovery***7**, 462–477, doi:10.1158/2159-8290.CD-16-1154 (2017).
43. Akamatsu, S. *et al.* The Placental Gene PEG10 Promotes Progression of Neuroendocrine Prostate Cancer. *Cell reports***12**, 922–936, doi:10.1016/j.celrep.2015.07.012 (2015).
44. Beshiri, M. L. *et al.* A PDX/Organoid Biobank of Advanced Prostate Cancers Captures Genomic and Phenotypic Heterogeneity for Disease Modeling and Therapeutic Screening. *Clinical cancer research: an official journal of the American Association for Cancer Research***24**, 4332–4345, doi:10.1158/1078-0432.CCR-18-0409 (2018).
45. Gao, D. *et al.* Organoid cultures derived from patients with advanced prostate cancer. *Cell***159**, 176–187, doi:10.1016/j.cell.2014.08.016 (2014).
46. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell***177**, 1888–1902 e1821, doi:10.1016/j.cell.2019.05.031 (2019).
47. Berger, A. *et al.* N-Myc-mediated epigenetic reprogramming drives lineage plasticity in advanced prostate cancer. *J Clin Invest***129**, 3924–3940, doi:10.1172/JCI127961 (2019).
48. Mu, P. *et al.* SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science***355**, 84–88, doi:10.1126/science.aah4307 (2017).
49. Hall, M. E. *et al.* Metastatic Hormone-sensitive Prostate Cancer: Current Perspective on the Evolving Therapeutic Landscape. *Onco Targets Ther***13**, 3571–3581, doi:10.2147/OTT.S228355 (2020).
50. Kyriakopoulos, C. E. *et al.* Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer: Long-Term Survival Analysis of the Randomized Phase III E3805 CHARTED Trial. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology***36**, 1080–1087, doi:10.1200/JCO.2017.75.3657 (2018).
51. de Bono, J. S. *et al.* Prednisone plus cabazitaxel or mitoxantrone for metastatic castration-resistant prostate cancer progressing after docetaxel treatment: a randomised open-label trial. *Lancet***376**, 1147–1154, doi:10.1016/S0140-6736(10)61389-X (2010).



52. Petrylak, D. P. *et al.* Docetaxel and estramustine compared with mitoxantrone and prednisone for advanced refractory prostate cancer. *N Engl J Med***351**, 1513–1520, doi:10.1056/NEJMoa041318 (2004).
53. Gandaglia, G., Fossati, N., Suardi, N., Montorsi, F. & Briganti, A. STAMPEDE trial and patients with non-metastatic prostate cancer. *Lancet***388**, 234–235, doi:10.1016/S0140-6736(16)31038-8 (2016).
54. Clarke, N. W. *et al.* Addition of docetaxel to hormonal therapy in low- and high-burden metastatic hormone sensitive prostate cancer: long-term survival results from the STAMPEDE trial. *Ann Oncol***30**, 1992–2003, doi:10.1093/annonc/mdz396 (2019).
55. Sweeney, C. J. *et al.* Chemohormonal Therapy in Metastatic Hormone-Sensitive Prostate Cancer. *N Engl J Med***373**, 737–746, doi:10.1056/NEJMoa1503747 (2015).
56. Tannock, I. F. *et al.* Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. *N Engl J Med***351**, 1502–1512, doi:10.1056/NEJMoa040720 (2004).
57. Yu, J. *et al.* A polycomb repression signature in metastatic prostate cancer predicts cancer outcome. *Cancer research***67**, 10657–10663, doi:10.1158/0008-5472.CAN-07-2498 (2007).
58. Xiao, L. *et al.* Epigenetic Reprogramming with Antisense Oligonucleotides Enhances the Effectiveness of Androgen Receptor Inhibition in Castration-Resistant Prostate Cancer. *Cancer research***78**, 5731–5740, doi:10.1158/0008-5472.CAN-18-0941 (2018).
59. Karthaus, W. R. *et al.* Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science***368**, 497–505, doi:10.1126/science.aay0267 (2020).
60. Laudato, S., Aparicio, A. & Giancotti, F. G. Clonal Evolution and Epithelial Plasticity in the Emergence of AR-Independent Prostate Carcinoma. *Trends Cancer***5**, 440–455, doi:10.1016/j.trecan.2019.05.008 (2019).
61. Linja, M. J. & Visakorpi, T. Alterations of androgen receptor in prostate cancer. *J Steroid Biochem Mol Biol***92**, 255–264, doi:10.1016/j.jsbmb.2004.10.012 (2004).
62. Koivisto, P. *et al.* Androgen receptor gene amplification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. *Cancer research***57**, 314–319 (1997).
63. Antonarakis, E. S. *et al.* AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N Engl J Med***371**, 1028–1038, doi:10.1056/NEJMoa1315815 (2014).
64. Gaddipati, J. P. *et al.* Frequent detection of codon 877 mutation in the androgen receptor gene in advanced prostate cancers. *Cancer research***54**, 2861–2864 (1994).
65. Calcinotto, A. *et al.* IL-23 secreted by myeloid cells drives castration-resistant prostate cancer. *Nature***559**, 363–369, doi:10.1038/s41586-018-0266-0 (2018).
66. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol***20**, 163–172, doi:10.1038/s41590-018-0276-y (2019).
67. Shay, T. & Kang, J. Immunological Genome Project and systems immunology. *Trends Immunol***34**, 602–609, doi:10.1016/j.it.2013.03.004 (2013).

68. van Dijk, D. *et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell***174**, 716–729 e727, doi:10.1016/j.cell.2018.05.061 (2018).
69. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Bio***16**, 278, doi:10.1186/s13059-015-0844-5 (2015).
70. Wu, D. & Smyth, G. K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic acids research***40**, e133, doi:10.1093/nar/gks461 (2012).
71. Li, C. *et al.* Single cell transcriptomics based-MacSpectrum reveals novel macrophage activation signatures in diseases. *JCI Insight***5**, doi:10.1172/jci.insight.126453 (2019).
72. Beltran, H. *et al.* Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer discovery***1**, 487–495, doi:10.1158/2159-8290.CD-11-0130 (2011).
73. Oberhuber, M. *et al.* STAT3-dependent analysis reveals PDK4 as independent predictor of recurrence in prostate cancer. *Mol Syst Bio***16**, e9247, doi:10.15252/msb.20199247 (2020).
74. Ramnarine, V. R. *et al.* The long noncoding RNA landscape of neuroendocrine prostate cancer and its clinical implications. *Gigascience***7**, doi:10.1093/gigascience/giy050 (2018).
75. Groner, A. C. *et al.* TRIM24 Is an Oncogenic Transcriptional Activator in Prostate Cancer. *Cancer cell***29**, 846–858, doi:10.1016/j.ccell.2016.04.012 (2016).
76. Cyrt, J. *et al.* Role of specialized composition of SWI/SNF complexes in prostate cancer lineage plasticity. *Nature communications***11**, 5549, doi:10.1038/s41467-020-19328-1 (2020).
77. Spahn, M. *et al.* Expression of microRNA-221 is progressively reduced in aggressive prostate cancer and metastasis and predicts clinical recurrence. *International journal of cancer. Journal international du cancer***127**, 394–403, doi:10.1002/ijc.24715 (2010).
78. Kuroda, H., Kutner, R. H., Bazan, N. G. & Reiser, J. Simplified lentivirus vector production in protein-free media using polyethylenimine-mediated transfection. *J Virol Methods***157**, 113–121, doi:10.1016/j.jviromet.2008.11.021 (2009).
79. Drost, J. *et al.* Organoid culture systems for prostate epithelial and cancer tissue. *Nature protocols***11**, 347–358, doi:10.1038/nprot.2016.006 (2016).
80. Bluemn, E. G. *et al.* Androgen Receptor Pathway-Independent Prostate Cancer Is Sustained through FGF Signaling. *Cancer cell***32**, 474–489 e476, doi:10.1016/j.ccell.2017.09.003 (2017).
81. Steen, C. B., Liu, C. L., Alizadeh, A. A. & Newman, A. M. Profiling Cell Type Abundance and Expression in Bulk Tissues with CIBERSORTx. *Methods in molecular biology***2117**, 135–157, doi:10.1007/978-1-0716-0301-7\_7 (2020).
82. Kowalczyk, M. S. *et al.* Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res***25**, 1860–1872, doi:10.1101/gr.192237.115 (2015).

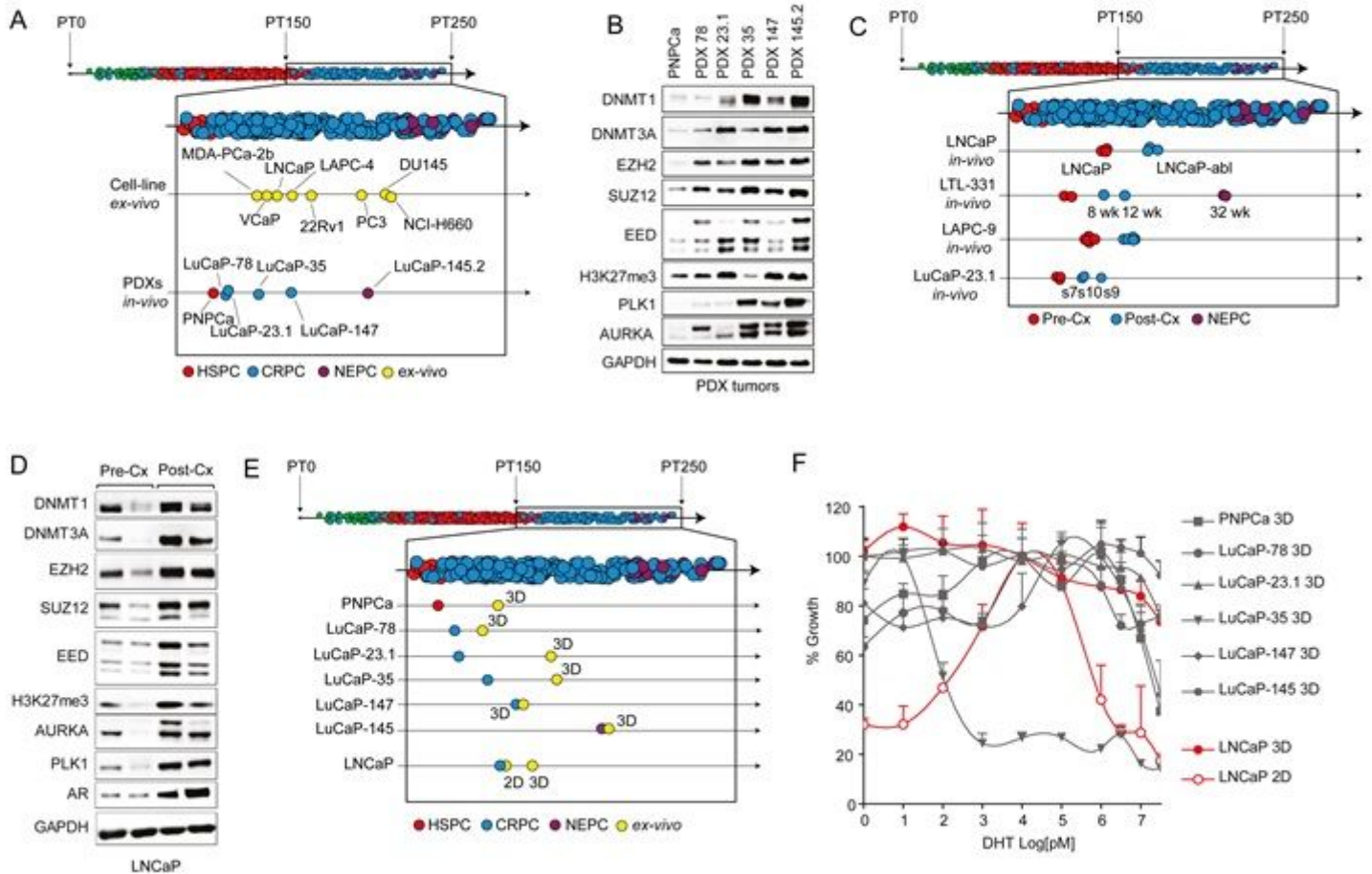
## Figures



**Figure 1**

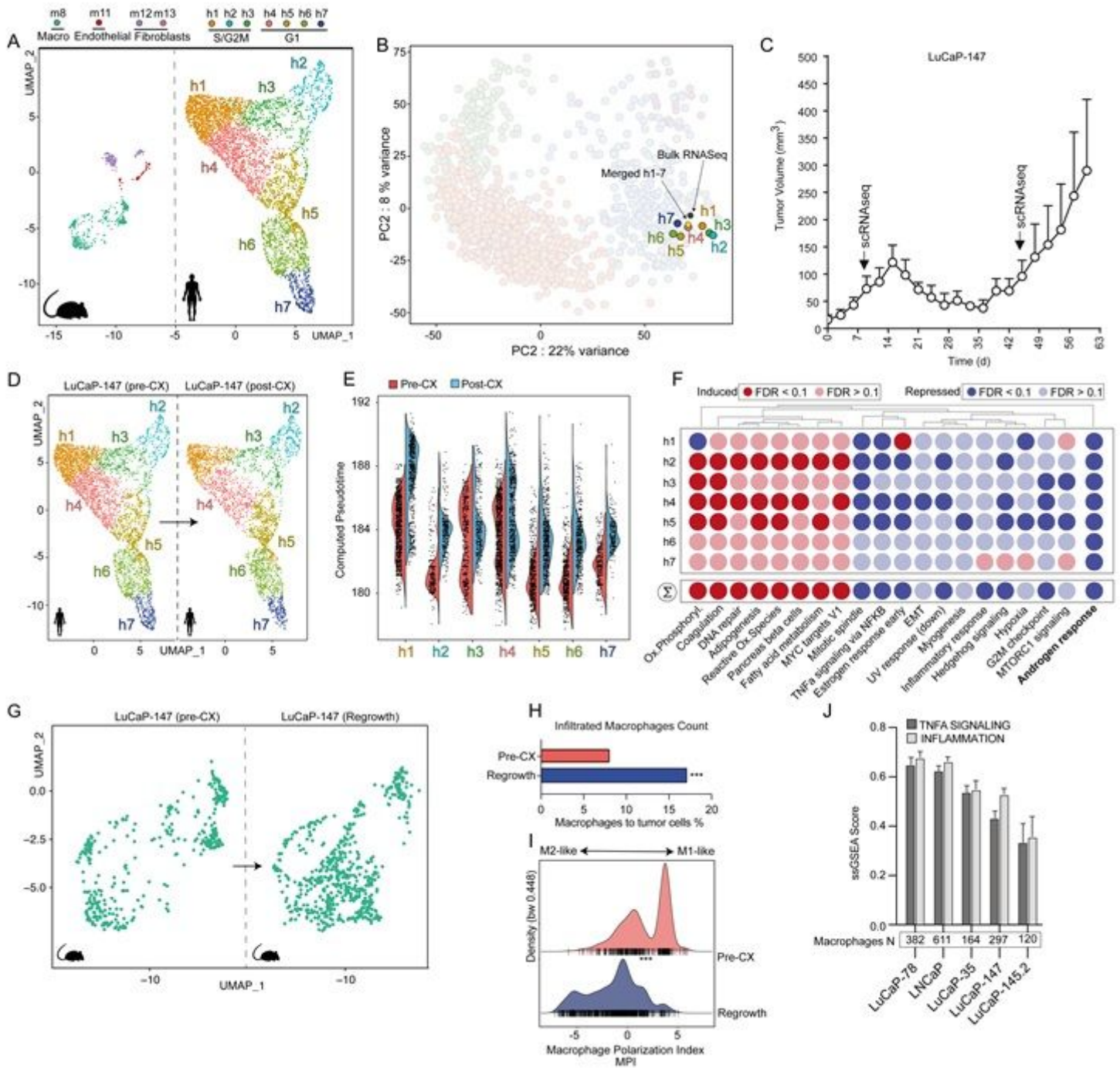
Trajectory to Prostate Cancer Progression (A) Principal component analysis (PCA) of pan-prostate cancer transcriptomes obtained from the indicated studies of normal, primary, castration-resistant (CRPC), and neuroendocrine prostate cancer (NEPC). (B) Unbiased trajectory analysis identifies the path to disease progression. Quantification of the path is indicated by inferred pseudo-time. (C) Plot representing the correlation between mRNAs and pseudo-time inferred along the trajectory. Polycomb repressive complex

related genes highlighted in orange, cell cycle-related genes in green, immune response in blue, and AR signaling in red. X-axis: Pearson's correlation coefficient between mRNAs and pseudo-time; Y-axis: The associated significance adjusted for False Discovery Rate (FDR) and expressed in the form of  $-10 \times \log_{10}(\text{FDR})$ . (D) Schematic representation of gene expression changes in AR-regulated target genes related to cell differentiation and proliferation and PRC2 components along the trajectory. Correlation coefficients between mRNA expression and pseudo-time are depicted in a three-color scale (blue: -1; white: 0; red: +1). (E) Gene set enrichment analysis performed on genes ranked for their Pearson's coefficient as determined by the correlation between mRNA expression and pseudo-time inferred from the trajectory. Increasing pseudo-time results in an increase of cell cycle-related genes and concomitant down-regulation of androgen-responsive genes. (F) Scatterplot revealing correlation between mRNAs and protein abundances, expressed in the form of fold-change (log-scale) between CRPCs and Primary tumors. (G) Pearson's coefficients, as determined from the correlation between somatic mutations (0:wild-type; 1:non-synonymous mutation) and inferred pseudo-time along the trajectory. To dissect the relative impact on disease progression at different stages, coefficients were computed separately in primary and CRPC/NEPC samples. The analysis was performed only for genes mutated at least in 6 individuals. Significance level (p-values): \* < 0.05, \*\* < 0.01, \*\*\* < 0.001. (H) Computed Pearson's correlation between samples' numeric copy number status (-2: homozygous deletion; -1: heterozygous deletion; 0: wild-type; 1:gain; 2:amplification) and inferred pseudo-time, stratified for primary and metastatic tumors (CRPC, NEPC). (I) Histograms depicting the correlation between the inferred abundance of the indicated immune cell populations (as determined by Cibersortx) and pseudo-time. P-values associated with Pearson's correlation coefficients were adjusted for multiple testing using the false discovery rate (FDR). (J) Schematic representation of expression changes in genes linked to the tumor environment. Transcripts specific to M1-macrophages decrease along the trajectory while those of the M2 counterpart increase.



**Figure 2**

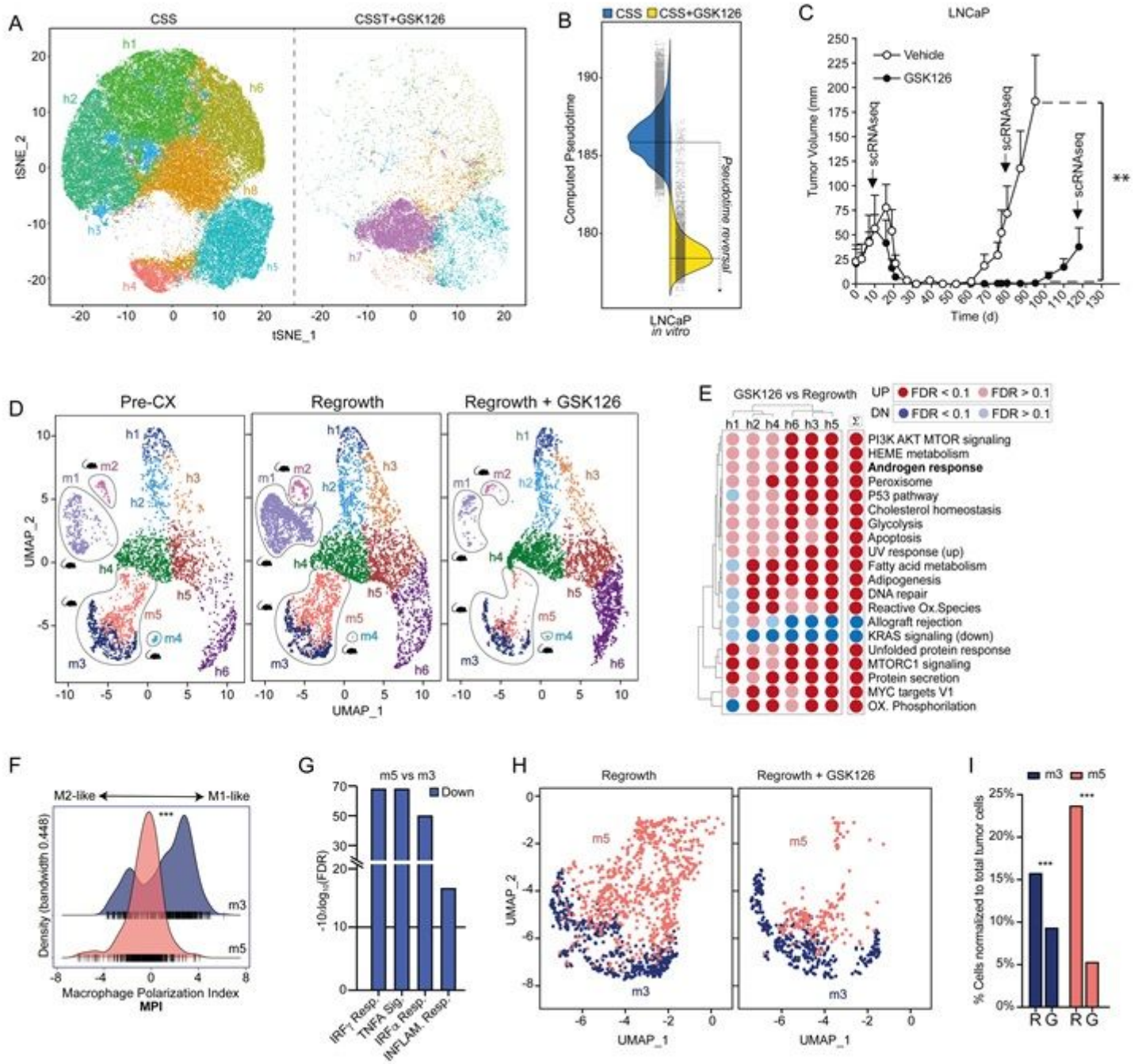
Mapping of Human Prostate Cancer Models to the Trajectory (A) Projection of the indicated human cell lines *ex vivo* and patient-derived xenograft (PDX) models *in vivo* to the trajectory (PT = pseudo-time). (B) Immunoblotting analysis of the indicated proteins across PDX models indicates an upregulation of polycomb repressive complex 2 members and G2M cell cycle checkpoint genes. (C) Androgen-dependent xenograft models progress along the trajectory when developing castration resistance. (D) Corresponding immunoblot analysis of LNCaP xenograft models shows upregulation of AR, polycomb repressive complex 2 members, and G2M cell cycle checkpoint genes upon recurrence after castration (Cx). (E) 3D *ex vivo* cultures in Matrigel of the indicated PDX and the xenografted LNCaP cells show higher PT than their *in vivo* counterparts. (F) Corresponding dihydrotestosterone (DHT) dose-response curves of the indicated models in 3D using Matrigel versus standard 2D culture. In the 3D conditions, DHT-dependency is largely abolished. See also Supplementary Figure 2.



**Figure 3**

Single-Cell Resolution to the Trajectory (A) Dimensionality reduction of single-cell distribution of LuCaP-147 PDX model in vivo using Uniform Manifold Approximation and Projection (UMAP) and subsequent identification of cell clusters performed using Seurat 46 workflow. Human (right) and mouse cells (left) are separated from each other. A total amount of 7 and 4 clusters could be identified for human and mouse cells, respectively. For the latter, we indicated the cells of origin corresponding to the various clusters on top. Inference of cell-types was performed with SingleR 66 through the exploitation of the ImmGen repository 67. For human cell clusters, we indicated the inferred cell-cycle phase as predicted

using Seurat. (B) Projection of single-cell clusters on the PCA plot. The position of merged single-cell data corresponds to the one from bulk RNA sequencing data. Please refer to Star Methods for detailed information on scRNA-Seq data integration with bulk RNA-Seq. (C) LuCaP-147 xenografts regress and re-grow within 4 weeks after castration. (D) Comparison of tumor single-cell clusters before (left) and after castration (right). (E) Violin plot shows an increase in pseudo-time of individual cells within the cell clusters after castration. To deal with drop-out events, the pseudo-time inference was performed for each cell following imputation of missing genes using RMagic 68. (F) Gene-sets perturbed in LuCaP-147 xenografts' single-cell clusters at regrowth (post castration) compared to pre castration. Most hallmark gene sets are up-or down-regulated similarly. A marked downregulation of AR responsive genes is noted. Differential expression for each cluster denoting the transcriptional changes occurring after castration was determined using the MAST algorithm 69. Subsequently, we determined the gene set enrichments using Camera (pre-ranked) 70. (G) Dimensionality reduction (UMAP) of murine macrophages before (left) and post-castration (right) highlights a notable increase in macrophage count at regrowth. (H-I) After castration, the percentage of infiltrated macrophage to tumor cell ratio not only increases (H) but also displays more M2-like transcriptional features according to the Macrophage Polarization Index (I), as determined by using MacSpectrum 71. Significance levels (p-values) were determined using Wilcoxon rank-sum test: \* < 0.05, \*\* < 0.01, \*\*\* < 0.001. (J) Single samples gene-set enrichment analysis of inflammation-related pathways performed following reclustering of murine macrophages extracted from the corresponding single-cell RNA-Seq experiments. Missing gene-expression values (dropout events) for each cell were imputed using RMagic. With increasing pseudo-time along the trajectory, macrophages of xenograft models display less active TNFA and inflammatory signaling. PNPc xenografts were excluded from the analysis because of the limited number of infiltrated macrophages. See also Supplementary Figure 3.



**Figure 4**

EZH2 Inhibition Cooperates with Castration (A) Dimensionality reduction (TSNE) of single-cell RNASeq performed on LNCaP cells cultured *in vitro* with charcoal-stripped serum (CSS) in the presence (right) or absence (left) of the EZH2 inhibitor GSK126. Identification of cell clusters (h1-h8) was performed using the Seurat workflow. EZH2 inhibition has a dramatic impact on LNCaP cells, as most of the clusters disappear, while the remaining cells undergo such deep transcriptional modifications that give rise to a novel cluster (h7). (B) Pseudo-time of individual LNCaP cultured in charcoal-stripped serum (CSS) is significantly reduced upon GSK126 treatment. Pseudo-time was computed for each cell, following



imputation of missing genes (drop-outs) using RMagic. (C) GSK126 treatment for 3 weeks upon castration significantly delays the regrowth of LNCaP xenografts after castration. (D) Dimensionality Reduction (UMAP) of LNCaP xenografts performed on scRNA-Seq experiments derived from mice before castration (left), after castration (center), and after concomitant castration and EZH2 inhibition with GSK126. Murine cells can be subdivided into 5 clusters corresponding to different cell populations according to SingleR (m1:fibroblasts; m2:endothelial cells; m3,m5:macrophages; m4:monocytes). Human malignant cells can be separated into 6 clusters. An increase in the relative number of cells in cluster h6 and a concomitant reduction of murine macrophages following EZH2 inhibition is observed. (E) Upon GSK126 pre-treatment, for each cluster, we determined differentially expressed genes (MAST algorithm) and performed gene set enrichment using Camera (pre-ranked). Results highlight a global increase in androgen-responsive genes. (F) The density plot of macrophage polarization index (MPI) reveals that macrophage cluster m5 (which decreases upon GSK126 administration) shows M2-like transcriptional features, while cluster m3 corresponds to an increased M1-like polarization as determined by MacSpectrum. (G) Differential expression (MAST algorithm) shows that M1-like inflammatory signaling pathways are downregulated in m5 compared to the m3 cluster. (H) Snapshot of macrophage clusters extracted from panel D shows that the relative cell count of the m5 cluster specifically decreases upon GSK126 treatment. (I) Corresponding relative changes in the abundance of m3- and m5-cluster macrophages upon GSK126 treatment. Significance level (p-values): \* < 0.05, \*\* < 0.01, \*\*\* < 0.001. See also Supplementary Figure 4.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Suppl.Table1.xlsx](#)
- [SupplementaryFigures.pdf](#)