



OPEN

Transcriptional mediators of treatment resistance in lethal prostate cancer

Meng Xiao He^{1,2,3}, Michael S. Cuoco³, Jett Crowdis^{2,3}, Alice Bosma-Moody^{2,3,4}, Zhenwei Zhang^{2,14}, Kevin Bi^{2,3}, Abhay Kanodia², Mei-Ju Su², Sheng-Yu Ku², Maria Mica Garcia², Amalia R. Sweet⁵, Christopher Rodman³, Laura DelloStritto^{2,6}, Rebecca Silver², John Steinharter², Parin Shah², Benjamin Izar^{7,8}, Nathan C. Walk², Kelly P. Burke^{2,9,10}, Ziad Bakouny², Alok K. Tewari², David Liu^{2,3}, Sabrina Y. Camp^{2,3}, Natalie I. Vokes^{2,3,15,16}, Keyan Salari^{3,11}, Jihye Park^{2,3}, Sébastien Vigneau^{2,3,6}, Lawrence Fong², Joshua W. Russo², Xin Yuan⁵, Steven P. Balk⁵, Himisha Beltran², Orit Rozenblatt-Rosen³, Aviv Regev^{3,13,17}, Asaf Rotem^{2,3,6,18}, Mary-Ellen Taplin^{2,19} and Eliezer M. Van Allen^{2,3,6,19} ✉

Metastatic castration-resistant prostate cancer is typically lethal, exhibiting intrinsic or acquired resistance to second-generation androgen-targeting therapies and minimal response to immune checkpoint inhibitors¹. Cellular programs driving resistance in both cancer and immune cells remain poorly understood. We present single-cell transcriptomes from 14 patients with advanced prostate cancer, spanning all common metastatic sites. Irrespective of treatment exposure, adenocarcinoma cells pervasively coexpressed multiple androgen receptor isoforms, including truncated isoforms hypothesized to mediate resistance to androgen-targeting therapies^{2,3}. Resistance to enzalutamide was associated with cancer cell-intrinsic epithelial-mesenchymal transition and transforming growth factor- β signaling. Small cell carcinoma cells exhibited divergent expression programs driven by transcriptional regulators promoting lineage plasticity and HOXB5, HOXB6 and NR1D2 (refs. 4–6). Additionally, a subset of patients had high expression of dysfunction markers on cytotoxic CD8⁺ T cells undergoing clonal expansion following enzalutamide treatment. Collectively, the transcriptional characterization of cancer and immune cells from human metastatic castration-resistant prostate cancer provides a basis for the development of therapeutic approaches complementing androgen signaling inhibition.

Prostate cancer cells depend on the activity of the transcription factor androgen receptor (AR), which represents a major therapeutic target both in primary and advanced disease. Metastatic castration-resistant prostate cancer (mCRPC) is treated with

second-generation androgen-targeting therapies, including enzalutamide and abiraterone¹. Development of resistance to these therapies is nearly universal and neuroendocrine prostate cancer (NEPC) is often intrinsically resistant^{1,7}. Despite successes in other solid tumors, immune checkpoint inhibitors have provided minimal benefit except in rare subsets of patients¹. Both cancer cell-intrinsic and extrinsic programs contributing to resistance remain areas of active investigation. Large-scale genomic analyses have associated alterations in *RBI1*, *AR*, *TP53*, Wnt/ β -catenin pathway, PI3K pathway and homologous recombination repair genes with worse outcomes on second-generation androgen-targeting therapies^{8–11}. Moreover, substantial attention has been focused on a diverse set of AR isoforms, particularly those lacking a ligand-binding domain and hypothesized to be constitutively active^{2,3}. As bulk sequencing of biopsies of varying tumor purity collected from diverse metastatic sites, which harbor different nonmalignant cells, may obscure cell type-specific expression programs, we profiled human mCRPC using single-cell transcriptomics.

With informed consent, we collected fresh biopsies from the three common mCRPC metastatic sites (bone, lymph node and liver) for whole exome sequencing, bulk RNA sequencing (RNA-seq) and single-cell RNA-seq (scRNA-seq) using the Smart-seq2 protocol (Methods)¹². Patients had varied treatment histories, with biopsies approximately evenly represented before and after enzalutamide exposure (Extended Data Fig. 1 and Supplementary Table 1). In addition to adenocarcinomas, one tumor exhibited a small cell carcinoma histology. After quality control, our cohort consisted of 2,170 cells from 14 patients and 15 biopsies (Methods; Supplementary

¹Harvard Graduate Program in Biophysics, Boston, MA, USA. ²Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. ³Broad Institute of Harvard and MIT, Cambridge, MA, USA. ⁴Harvard Medical School, Boston, MA, USA. ⁵Department of Medicine, Division of Hematology/Oncology, Beth Israel Deaconess Medical Center, Boston, MA, USA. ⁶Center for Cancer Genomics, Dana-Farber Cancer Institute, Boston, MA, USA. ⁷Columbia Center for Translational Immunology, New York, NY, USA. ⁸Department of Medicine, Division of Hematology/Oncology, Columbia University Medical Center, New York, NY, USA. ⁹Department of Immunology, Blavatnik Institute, Harvard Medical School, Boston, MA, USA. ¹⁰Evergrande Center for Immunologic Diseases, Harvard Medical School and Brigham and Women's Hospital, Boston, MA, USA. ¹¹Department of Urology, Massachusetts General Hospital, Boston, MA, USA. ¹²Division of Hematology and Oncology, University of California, San Francisco, San Francisco, CA, USA. ¹³Department of Biology, Howard Hughes Medical Institute and Koch Institute for Integrative Cancer Research, MIT, Cambridge, MA, USA. ¹⁴Present address: Department of Pathology, University of Massachusetts Memorial Medical Center, Worcester, MA, USA. ¹⁵Present address: Department of Thoracic/Head and Neck Oncology, MD Anderson Cancer Center, Houston, TX, USA. ¹⁶Present address: Department of Genomic Medicine, MD Anderson Cancer Center, Houston, TX, USA. ¹⁷Present address: Genentech, South San Francisco, CA, USA. ¹⁸Present address: AstraZeneca, Waltham, MA, USA. ¹⁹These authors jointly supervised this work: Mary-Ellen Taplin, Eliezer M. Van Allen. ✉e-mail: eliezerm_vanallen@dfci.harvard.edu

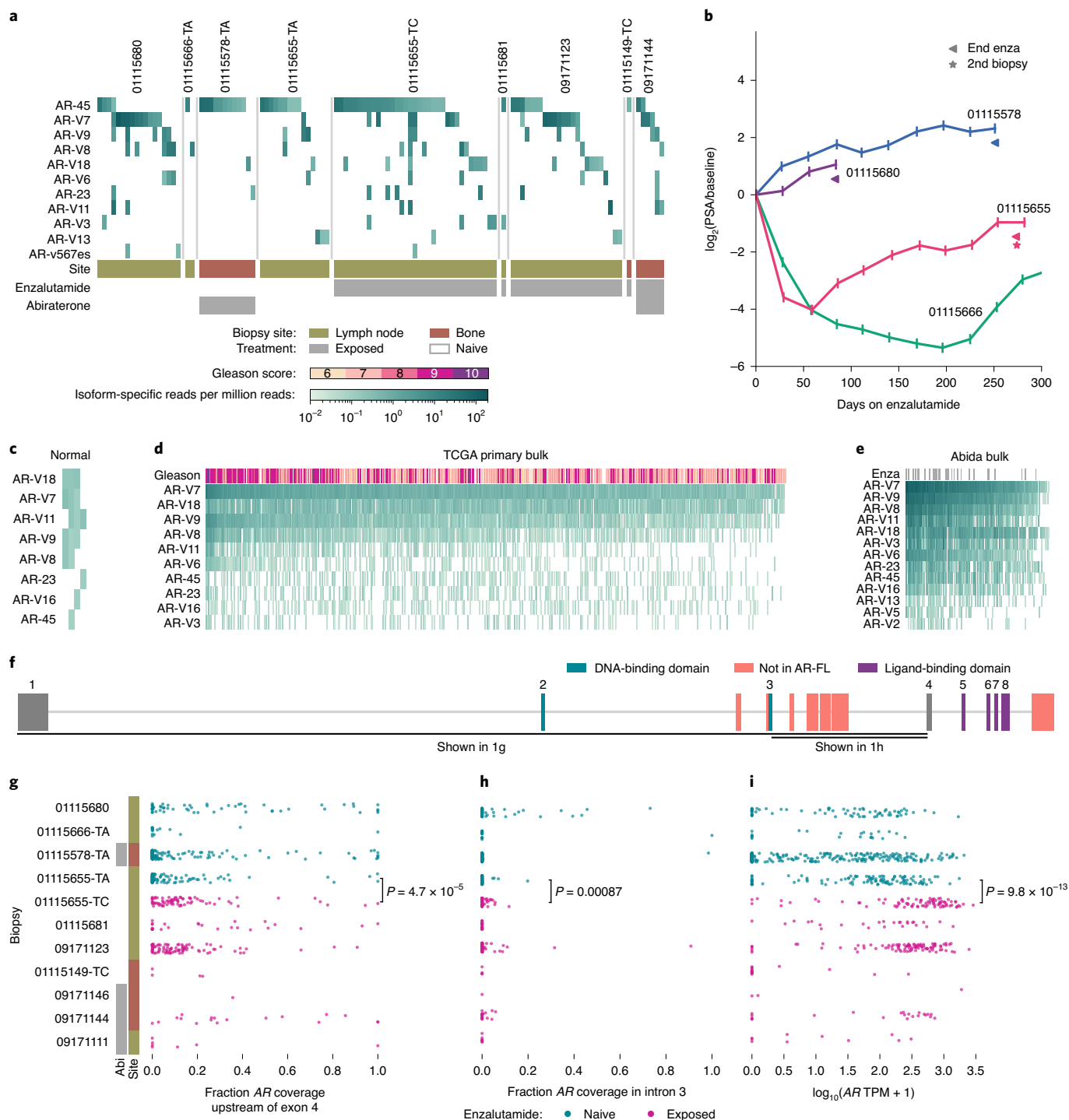


Fig. 1 | Complex AR isoform coexpression within individual cells and tumors is common across disease stages and resistance states. **a**, Heat map displaying frequency of isoform-specific reads mapping to AR splice variants (Supplementary Fig. 1). Each column represents AR variants detected in a single cell, with only cells and isoforms that had at least one isoform-specific read shown. Short-read-based approaches cannot specifically identify full-length AR (Methods). **b**, PSA histories for the four patients for whom immediately pre-enzalutamide biopsies are shown in **a**. Patient 01115666 remained on enzalutamide beyond the 300 d shown. **c–e**, Heat maps displaying frequency of isoform-specific reads in bulk RNA-seq of healthy prostate ($n=4$)¹⁵, TCGA primary prostate adenocarcinoma ($n=496$)¹⁴ and advanced prostate adenocarcinoma ($n=127$, poly-A-sequenced only, enzalutamide and abiraterone-naive or enzalutamide-exposed)⁸. Gleason score and enzalutamide exposure status are shown for TCGA and Abida et al. cohorts, respectively. For each cohort, AR variants detected in at least 20% of samples are shown. **f**, Schematic representation of AR locus. Rectangles indicate exons. Exons corresponding to the full-length AR transcript are numbered, with exons comprising different functional domains colored. Select alternative exons included in AR splice variants are indicated. **g**, Fraction of total AR coverage upstream of exon 4 (including the DNA-binding domain but excluding the ligand-binding domain) in single cells. **h**, Fraction of total AR coverage in intron 3 (including multiple cryptic/alternative exons included in truncated splice variants) in single cells. **i**, Total AR expression in single cells. **g–i**, P value compares cells before ($n=112$) and after ($n=83$) enzalutamide treatment for patient 01115655 (two-sided Mann-Whitney U -test).

Table 3). We labeled clusters for dominant cell type based on marker gene expression (Extended Data Fig. 2; Methods).

To investigate the role of AR isoforms in resistance, we developed methodology to specifically identify literature-described isoforms (Supplementary Figs. 1 and 2a,b; Methods). Isoform-specific reads corresponding to multiple AR variants were detected in the same cell across tumors and treatment exposures. AR-V7, AR-V8 and AR-V9, which lack the ligand-binding domain, were present both in cells exposed to and naive to second-generation androgen-targeting compounds (Fig. 1a). All three were detected in enzalutamide- and abiraterone-naive cells from patient 01115655, who nonetheless experienced a large decrease in prostate-specific antigen (PSA) levels following enzalutamide initiation (Fig. 1b). In post-enzalutamide cells from the same patient, we observed no evidence of a selective sweep driven by any dominant single AR isoform, and almost all post-treatment isoforms were detectable in some cells before enzalutamide. In cancer cells from patient 01115578, who had no clinical response to enzalutamide, we did not robustly detect any isoform, except AR-45 (hypothesized to be dominant negative)^{2,13}. Across the cohort, AR-45 and AR-V7 were coexpressed in significantly fewer cells than expected by chance, but we did not observe a replacement of AR-45-expressing cells by AR-V7-expressing cells in post-enzalutamide biopsies (Supplementary Fig. 2c). More generally, we did not consistently detect any isoform specific to post-enzalutamide biopsies.

Applying our approach to bulk RNA-seq of normal prostate, TCGA primary prostate adenocarcinoma and mCRPC adenocarcinomas, we consistently detected coexpression of multiple isoforms within the same tumor across disease stages (Fig. 1c–e)^{8,14,15}. AR-V7, AR-V8 and AR-V9 were detected even in normal prostate and in mCRPC, no AR isoform was significantly associated with duration on treatment (Extended Data Fig. 3a). While we observed robust increases in AR-V7 after enzalutamide exposure, increases were observed for most of the detected isoforms (Extended Data Fig. 3b). After normalizing the number of isoform-specific reads to AR read count (as opposed to total read count), we observed no significant differences between exposed and naive tumors (Extended Data Fig. 3c). These observations suggest that the increased abundance after enzalutamide exposure of many isoforms (including AR-V7) may largely be a consequence of increased total AR expression and that the relative abundances of the different isoforms may be largely unchanged. Similarly, higher Gleason scores were more common in primary tumors with increased abundance of many AR isoforms but not of a specific variant (Fig. 1d). Collectively, these results suggest no single AR isoform is a major driver of resistance to therapy.

Substantial numbers of reads from the AR genomic interval failed to map to any curated isoform, suggesting described isoforms form an incomplete census (Supplementary Fig. 2d). Short-read sequencing precludes direct isoform reconstruction, so we defined two measures of AR splicing that may be functionally informative. Intron 3 contains many of the terminal cryptic/alternative exons included in truncated isoforms, so we quantified the proportion of total AR coverage from intron 3 or from a larger interval that includes upstream exons (Fig. 1f)². Increases in these measures may suggest expression of a greater proportion of transcripts that encode ligand-independent proteins.

We detected an increase in both measures after enzalutamide treatment for patient 01115655, but the more dramatic effect was an increase in total AR expression and copy number (Fig. 1g–i). Across samples, wide variation in cells from the same tumor was pervasive. Some enzalutamide-exposed biopsies (09171123 and 0917144) had high proportions of cells with high values for both measures, but so did some number of cells from all enzalutamide-naive biopsies, including from both patients with subsequent PSA response. More generally, the observed isoform coexpression patterns muddy interpretation of their functional effects, especially given poorly understood heterodimerization^{2,13}. Combined with the lack of systematic differences before and after therapy, our observations suggest that assessments of single AR variants may be insufficient for a causal understanding of a tumor's sensitivity to androgen-targeting therapies even if their detection can serve as a proxy of total AR.

Next, we sought to identify broad expression changes following enzalutamide exposure. We scored cancer cells for expression of the MSigDB hallmark gene sets and curated literature-derived gene sets corresponding to proposed resistance mechanisms using a subsampling procedure to filter out patterns driven by individual tumors (Fig. 2a; Methods)^{16–27}. Enzalutamide-exposed cells upregulated gene sets for epithelial–mesenchymal transition (EMT) and transforming growth factor (TGF)- β signaling (Fig. 2a–c). We observed consistent increases when restricting only to the patient with paired biopsies (Supplementary Fig. 4a,b). In a cohort of mCRPC bulk RNA-seq, we found a similar effect for TGF- β signaling in enzalutamide-exposed lymph node biopsies, although the number of exposed biopsies was small and the effect was not statistically significant (Fig. 2d)⁸. EMT scores were confounded with tumor purity in these bulk data, with large differences likely driven by presence of benign cells (Supplementary Fig. 4c)²⁸.

Concordantly, in VCaP-16, a cell line generated from long-term culture in enzalutamide until emergence of resistance, we observed increased SMAD2 phosphorylation following equivalent TGF- β

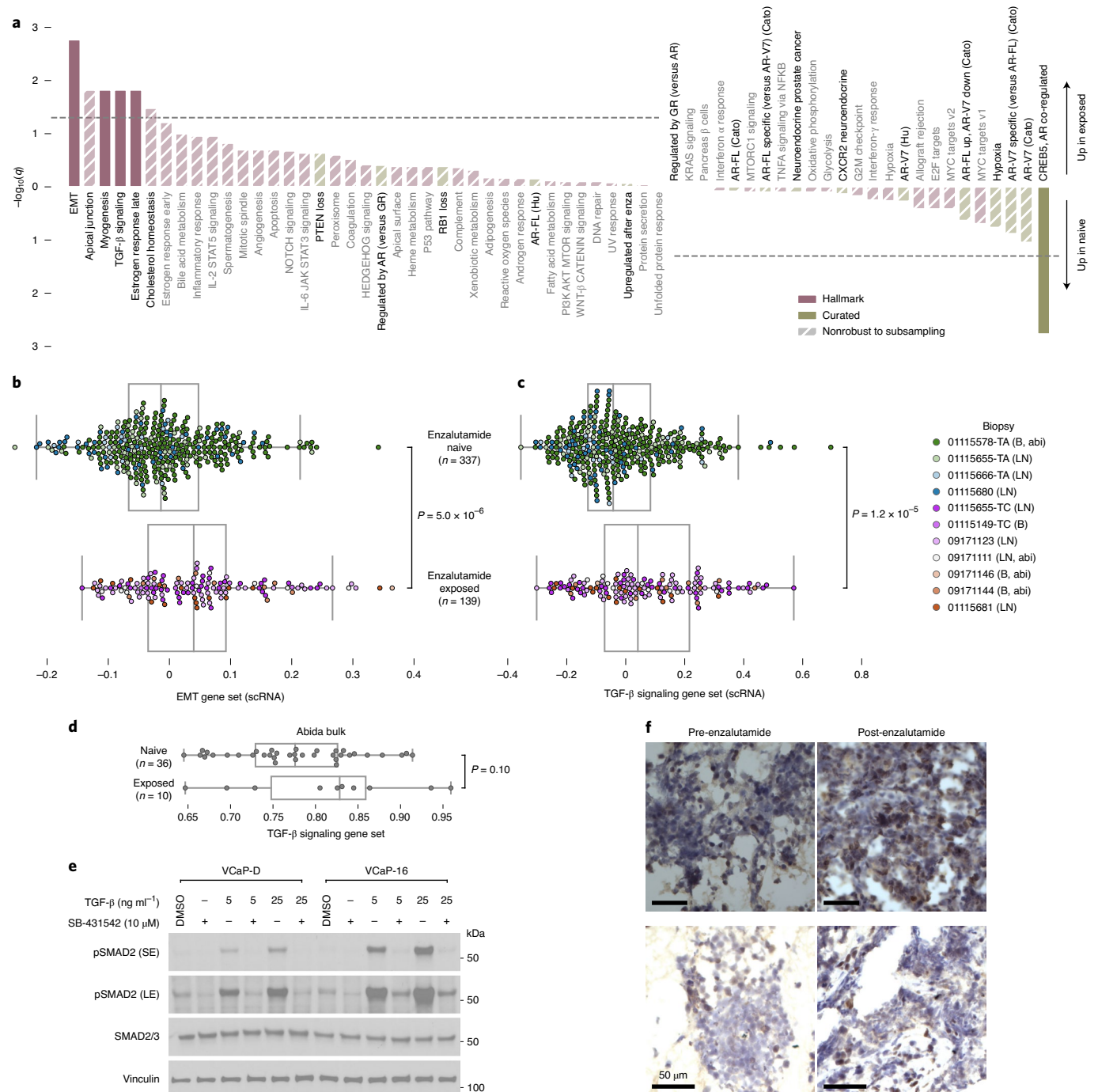
Fig. 2 | Enzalutamide-exposed adenocarcinoma cells upregulate expression programs associated with EMT and TGF- β signaling. **a**, Comparison of gene set expression scores in enzalutamide-exposed versus naive adenocarcinoma cells. Gene sets included include the MSigDB hallmark collection¹⁶ and literature-curated gene sets hypothesized to be related to enzalutamide resistance (Supplementary Table 5). *q* values from Benjamini–Hochberg adjustment of *P* values from two-sided Mann–Whitney *U*-tests (Supplementary Table 6). The broken lines indicate *q* = 0.05. Only G1 cells were included in analyses. Results shown are the median of a subsampling procedure designed to even out representation of cells from different biopsies. Each subsample included 67 exposed and 76 naive cells. Hashed bars correspond to results that were nonrobust in an additional leave-one-sample-out sampling step, suggesting that the effect is driven by patient-specific mechanisms. For details of subsampling used during statistical testing, see Methods and Supplementary Fig. 3. enza, enzalutamide; IL, interleukin; GR, glucocorticoid receptor. **b, c**, MSigDB hallmark EMT (**b**) and TGF- β signaling gene set expression scores (**c**) for individual cells (G1 only) collected before and after enzalutamide treatment. Each dot represents a single cell and is colored corresponding to biopsy. Biopsy clinical attributes are indicated in parentheses in legend (B, bone biopsy; LN, lymph node biopsy; abi, previous abiraterone exposure). *P* values from two-sided Mann–Whitney *U*-test, including all G1 cells. **d**, MSigDB hallmark TGF- β signaling gene set expression scores for bulk RNA-seq of mCRPC lymph node biopsies⁸ collected before and after enzalutamide treatment. Each dot represents a single tumor. *P* value from one-sided Mann–Whitney *U*-test. **e**, Western blot of SMAD2/3 and phospho-SMAD2 levels in enzalutamide-sensitive VCaP-D and enzalutamide-resistant VCaP-16. Cells were grown in basal maintenance medium with 0.5% FBS for 48 h, then treated for 2 h with the TGF- β receptor inhibitor SB-431542, recombinant TGF β -1 protein or a combination of both. Cells treated with SB-431542 were pretreated for 24 h with SB-431542 before addition of recombinant TGF β -1 protein. SE, short exposure; LE, long exposure. For uncropped images and both sets of vinculin loading controls, see Source Data. DMSO, dimethylsulfoxide. **f**, Immunohistochemical staining of phospho-SMAD2 in longitudinal biopsies from two patients immediately before and after enzalutamide treatment. Each row corresponds to one patient. All biopsies are from bone metastases. Scale bars, 50 μ m. Box plots are represented by center line, median; box limits, upper and lower quartiles; whiskers extend at most 1.5 \times interquartile range past upper and lower quartiles.

stimulation compared to enzalutamide-sensitive VCaP-D cells, indicating they are primed to amplify downstream signaling by TGF- β (Fig. 2e; Methods). Short-term exposure of wild-type VCaP cells to enzalutamide had minimal effect (Supplementary Fig. 4d). Furthermore, in two pairs of patient biopsies collected immediately before enzalutamide initiation and after progression, nuclear localized phosphorylated SMAD2 increased in the post-treatment biopsies (Fig. 2f). While other hypothesized resistance programs were not robustly upregulated in our cohort, our observations support clinical studies of TGF- β inhibition in the context of enzalutamide resistance (Fig. 2a).

One biopsy was of a small cell carcinoma, a rare, aggressive form of prostate cancer that does not respond to androgen-targeting

therapies⁷. Cancer cells from this biopsy had no detectable AR, downregulated an AR-regulated gene set and upregulated a NEPC gene set (Fig. 3a,b and Supplementary Fig. 5)^{17,18}. With only one NEPC biopsy, we analyzed expression levels of gene sets inferred in our scRNA-seq to be regulated by specific transcription factors (regulons), an approach shown to identify more generalizable expression programs (Methods)²⁹.

Comparing the single-cell results and a bulk-sequenced cohort that we scored for expression of the regulons inferred in scRNA-seq, we observed concordant patterns of differential regulon activity between adenocarcinoma and small cell carcinoma (Fig. 3c; Methods)⁸. Adenocarcinomas exhibited higher activity of HOXB13, which mediates AR regulatory activity, and of several ETS family



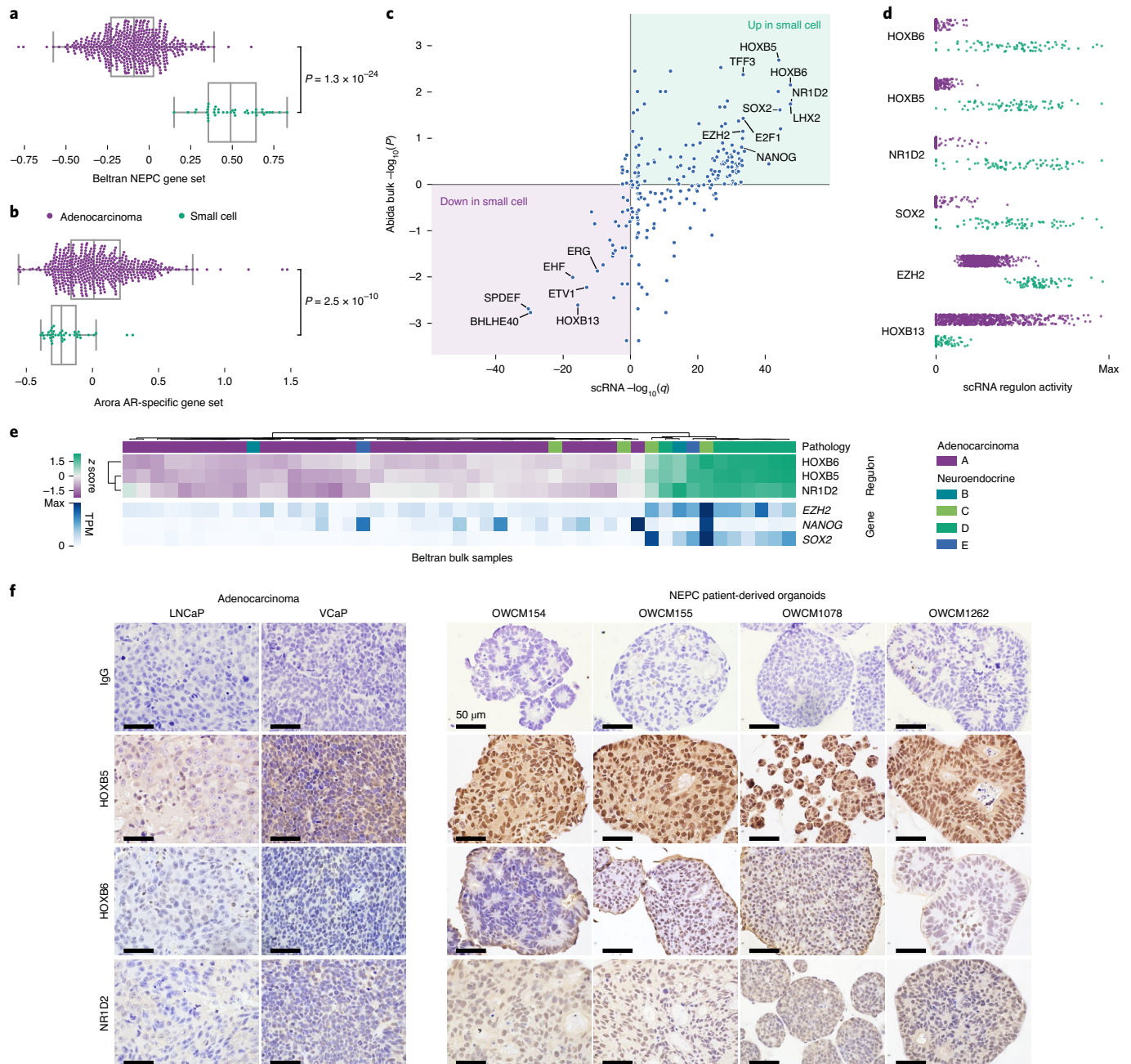


Fig. 3 | Cancer cells from small cell carcinoma employ a highly divergent regulatory program compared to adenocarcinoma cells. a, b, Gene set expression scores of single G1 cells using an expression signature of NEPC¹⁸ (**a**) and a set of genes under regulation by AR¹⁷ (**b**). Box plots include center line, median; box limits, upper and lower quartiles; whiskers extend at most 1.5 \times interquartile range past upper and lower quartiles; P values are from two-sided Mann-Whitney U -test. **c,** Inferred activity of regulons of different transcriptional regulators. x axis, q values from comparison of inferred regulon activity in cancer cells from small cell carcinoma ($n = 76$) versus cancer cells from adenocarcinomas ($n = 188$, sampled as described in Methods) (negative values indicate regulon is less active in small cell carcinoma; two-sided Mann-Whitney U -test, median outcome of sampling iterations (Methods) with Bonferroni FWER correction). y axis: P values (two-sided Mann-Whitney U -test, signed as previous) from comparison of expression scores of scRNA-derived regulons in bulk RNA-seq of small cell carcinomas ($n = 8$) versus adenocarcinomas ($n = 18$) from a published cohort⁸. **d,** Regulon activity in single cells for select transcriptional regulators. **e,** Hierarchical clustering of bulk RNA-seq of a published cohort of prostate cancers of known histopathology¹⁵ based on expression of HOXB5, HOXB6 and NR1D2 regulons inferred from scRNA-seq. B–E correspond to different NEPC subtype labels from original publication. Expression levels of *EZH2*, *NANOG* and *SOX2* are shown for reference but were not used in clustering ($n = 34$ adenocarcinoma, 15 NEPC). **f,** Immunohistochemical staining of HOXB5, HOXB6 and NR1D2 protein levels in two prostate adenocarcinoma xenografts (LNCaP and VCaP) and four NEPC patient-derived organoids. Scale bar, 50 μ m.

transcription factors³⁰. Small cell carcinoma had higher activity of *NANOG*, *SOX2* and *EZH2*, which promote lineage plasticity and resistance to androgen-targeting therapies^{4–7}. Additionally, we observed increased activity of three transcriptional regulators

whose role in NEPC has not been reported: *HOXB5* and *HOXB6*, two homeobox containing transcription factors and *NR1D2*, a nuclear hormone receptor family member (Fig. 3c,d). In a second, independent, bulk-sequenced clinical cohort, inferred regulon

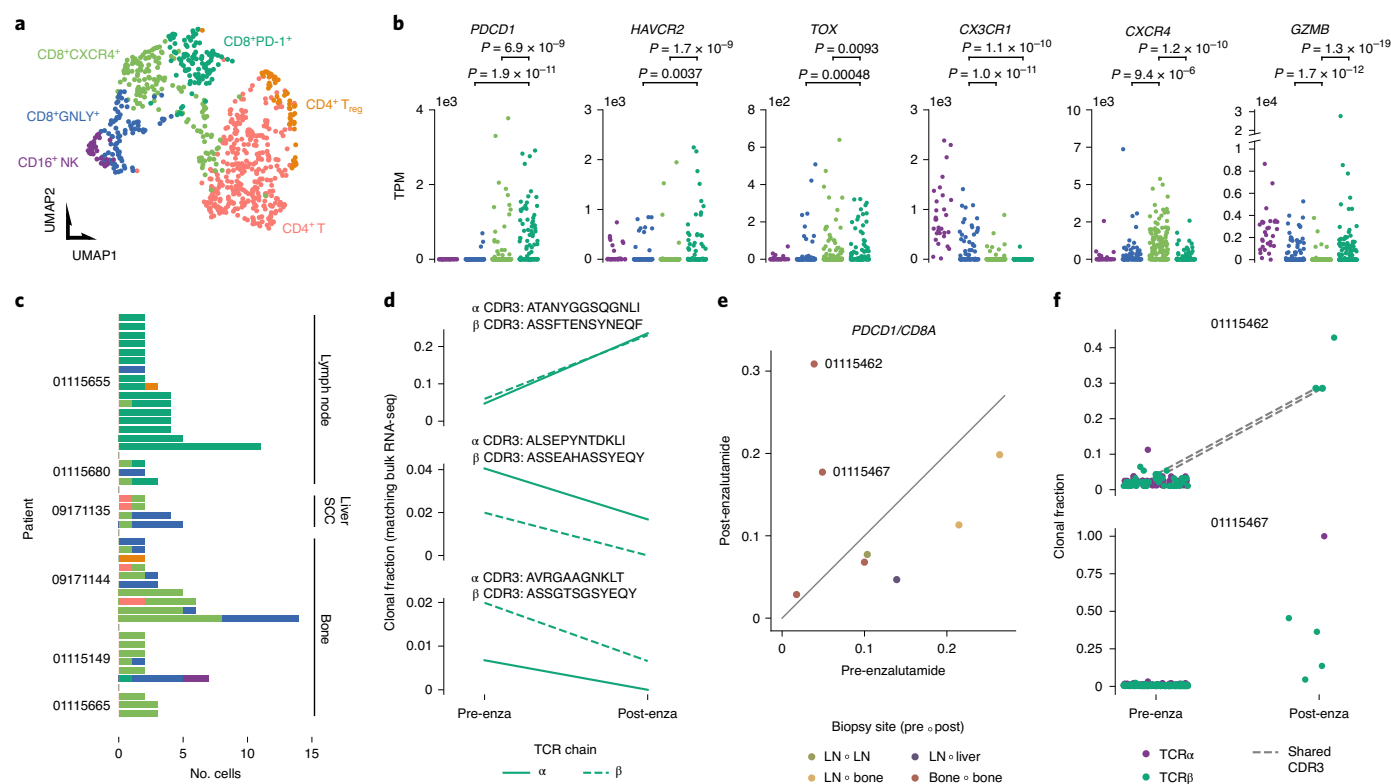


Fig. 4 | Cytotoxic lymphocyte populations and clonal expansions in metastatic niches. **a**, Subclustering of NK and T cells. Each dot represents a single cell projected onto uniform manifold approximation and projection (UMAP) space colored corresponding to clustering via the Louvain algorithm. Clusters are manually labeled with dominant phenotype/cell type from patterns of marker gene expression. Cluster colors are used throughout subpanels. **b**, Expression of select marker, effector and co-inhibitory receptor genes within cytotoxic clusters, CD16⁺ NK ($n = 30$), CD8⁺GNLY⁺ ($n = 84$), CD8⁺CXCR4⁺ ($n = 157$) and CD8⁺PD-1⁺ ($n = 106$). Nominal P values from a two-sided Mann–Whitney U -test are shown. TPM, transcripts per million. **c**, T cell clonotypes from scRNA-seq TCR reconstruction. Each bar represents cells sharing a reconstructed productive TCR CDR3 sequence. Bars are grouped by patient. SCC, small cell carcinoma. **d**, Changes in clonal fractions of cytotoxic T cell clonotypes in patient 01115655 following enzalutamide treatment. Each subplot corresponds to a single clonotype with TCRα and β CDR3 amino acid sequences paired from scRNA-seq. Clonal fractions for the same CDR3 sequences (matching at both nucleotide and amino acid level) inferred from TCR reconstruction in bulk RNA-seq are plotted. All detected single cells of the displayed clonotypes came from the CD8⁺PD-1⁺ T cell cluster. **e**, *PDCD1* to *CD8A* expression ratio in bulk RNA-seq of paired biopsies from the same patient before and after enzalutamide treatment. Paired biopsies did not always derive from the same site; color of dots indicate biopsy site types. For patient 01115462, the pre-enzalutamide biopsy was collected from the sacrum and the post-enzalutamide biopsy was collected from a left ischial lesion. **f**, Expanded T cell clonotypes following enzalutamide in patients 01115462 and 01115467. Each dot represents a single CDR3 sequence detected in bulk RNA-seq. CDR3 sequences identified both before and after enzalutamide are connected by a broken gray line.

activity of these transcription factors partitioned tumors by histopathology classification (Fig. 3e)¹⁸. Furthermore, all three were consistently expressed in NEPC patient-derived organoids but not in adenocarcinoma models (Fig. 3f). In aggregate, we recovered generalizable patterns of cancer cell-intrinsic expression in NEPC, recovering known regulators mediating resistance and implicating new transcription factors.

To complement therapeutic strategies targeting cancer cell-intrinsic mechanisms, clinical trials have tested immune checkpoint inhibitors in prostate cancer, but responses in mCRPC have been limited¹. We characterized infiltrating cytotoxic cells from various metastatic niches to improve understanding of mCRPC immunobiology. We subclustered T and natural killer (NK) cells into two CD4⁺ T cell populations, three largely CD8⁺ T cell populations and a population of strongly CD16⁺ and largely CD3⁻ cells dominated by NK cells (Fig. 4a and Extended Data Fig. 4a). Aside from a population of *CXCR4*-expressing CD8⁺ T cells chiefly derived from bone biopsies, cytotoxic clusters exhibited widespread *GZMB* expression (Fig. 4b and Supplementary Fig. 6a). One CD8⁺ T cell population largely derived from lymph node biopsies was marked by expression of *PDCD1* (encoding PD-1), *HAVCR2*, *TOX*, *TIGIT*, *ICOS*, *FASLG*

and *LAG3*, suggestive of a dysfunctional effector phenotype (Fig. 4b and Extended Data Fig. 4b,c). This population also exhibited elevated expression of both *ENTPD1* and *ITGAE*, whose coexpression identifies infiltrating cytotoxic cells reactive to cancer cells in other human cancers³¹. We did not observe a distinct cluster of *TCF7* and *SLAMF6* dual-expressing progenitor cells reported to mediate response to anti-PD-1 therapy in melanoma, but it remains unclear whether this explains poor response to immune checkpoint inhibitors in mCRPC (Extended Data Fig. 4e,f)³².

To identify potential immune system responses to metastasis, we examined T cell clonal expansion by reconstructing T cell receptor (TCR) complementarity-determining region 3 (CDR3) sequences in individual cells and matching bulk RNA-seq (Methods). Groups of T cells forming part of an expanded clonotype group with a shared CDR3 sequence were detected in six patients. Clonotype groups detected in lymph node metastases largely comprised cells from the CD8⁺ T cell cluster with elevated co-inhibitory receptor expression, whereas clonotype groups detected in bone metastases largely comprised cells from the *CXCR4*-expressing CD8⁺ T cell cluster (Fig. 4c). In one enzalutamide-exposed bone biopsy (09171144), a large clonotype group included cells with high

CX3CR1 expression. CD8⁺ T cells with high effector molecule, low exhaustion marker and *CX3CR1* expression have recently been characterized as highly cytolytic but poorly proliferative and unresponsive to PD-L1 blockade^{33–35}.

In patient 01115655, we collected cells from before and after enzalutamide treatment and noted marked changes in the infiltrating T cell populations. Before treatment, cytotoxic cells formed a minority of infiltrating T cells, but following treatment, most infiltrating T cells were dysfunctional *PDCD1*-expressing CD8⁺ T cells. For the largest clonotype group in this patient, we detected the corresponding TCR α and TCR β CDR3 sequences in bulk RNA-seq from both time points. The clonal fraction estimated from bulk RNA-seq increased sharply from approximately 5% before treatment to 25% after treatment, making it the largest detected clone (Fig. 4d). All cells of this clonotype group detected in scRNA-seq were part of the *PDCD1*-expressing dysfunctional cluster. To assess whether this phenomenon occurs in other patients, we examined a longitudinal mCRPC cohort (NCT01942837) with bulk RNA-seq, comparing biopsies collected before and after enzalutamide treatment. As a measure of the abundance of CD8⁺ T cells resembling those clonally expanded in 01115655, we examined the ratio of *PDCD1* to *CD8A* expression (Fig. 4e). Two patients had marked increases in this ratio following enzalutamide treatment and in both, we observed reduced T cell diversity after enzalutamide. Owing to increased tumor purity in enzalutamide-exposed biopsies, we generally detected fewer clonotypes after treatment; however, in patient 01115462, we detected large increases in clonal fraction of two CDR3 sequences (Fig. 4f).

The observed clonal expansions paired with relatively increased *PDCD1* expression suggests that in some tumors, CD8⁺ T cells do mount an antitumor response in the context of enzalutamide exposure. This may explain why some patients respond to immune checkpoint inhibition following enzalutamide resistance³⁶. While atezolizumab, an anti-PD-L1 therapy, was recently reported to provide no benefit in conjunction with enzalutamide, the observation of expanded T cell clones after enzalutamide may indicate that tumor killing by enzalutamide is necessary to drive the tumor immunity cycle³⁷. At the same time, high *CX3CR1* expression on some expanded clones may suggest poor response in other patients. Functional assays of infiltrating cytotoxic cells in heavily treated mCRPC are needed to decipher the effects of previous therapy and metastatic site on tumor reactivity and potential response to immunomodulatory therapies.

To explore the significance of cell populations other than cancer cells and T cells, we applied approaches leveraging genome-wide association studies (GWAS) of inherited prostate cancer risk. We began by examining cell type specificity of 109 genes implicated in an analysis of a GWAS of over 100,000 men (Methods)^{38,39}. Genes with cell type-specific expression were dominated by cancer cell-specific genes (Extended Data Fig. 5a). Using linkage disequilibrium score regression applied to specifically expressed genes (LDSC-SEG) to directly integrate GWAS summary statistics with scRNA-seq expression patterns, we likewise found germline heritability enriched only in genomic intervals near genes specifically expressed in cancer cells (Extended Data Fig. 5b; Methods)⁴⁰. These results suggest that any inherited risk for prostate cancer still operant during advanced disease is primarily mediated through cancer cell-intrinsic mechanisms.

In summary, we used single-cell transcriptomics and validation in external cohorts to characterize mCRPC expression programs in the context of therapy. In adenocarcinomas, our findings suggest therapeutic targeting of specific AR isoforms may be insufficient but support clinical inhibition of TGF- β and further characterization of cytotoxic cells. *HOXB5*, *HOXB6* and *NR1D2* warrant functional study and potential therapeutic development in NEPC. Longitudinal studies integrating multi-omics and spatially resolved

expression in larger cohorts may inform additional mediators of clinical resistance.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-021-01244-6>.

Received: 20 June 2020; Accepted: 13 January 2021;

Published online: 4 March 2021

References

- Teo, M. Y., Rathkopf, D. E. & Kantoff, P. Treatment of advanced prostate cancer. *Annu Rev. Med.* **70**, 479–499 (2019).
- Lu, J., der Steen, T. V. & Tindall, D. J. Are androgen receptor variants a substitute for the full-length receptor? *Nat. Rev. Urol.* **12**, 137–144 (2015).
- Antonarakis, E. S. et al. AR-V7 and resistance to enzalutamide and abiraterone in prostate cancer. *N. Engl. J. Med.* **371**, 1028–1038 (2014).
- Borges, G. T. et al. Conversion of prostate adenocarcinoma to small cell carcinoma-like by reprogramming. *J. Cell Physiol.* **231**, 2040–2047 (2016).
- Mu, P. et al. SOX2 promotes lineage plasticity and antiandrogen resistance in TP53- and RB1-deficient prostate cancer. *Science* **355**, 84–88 (2017).
- Ku, S. Y. et al. Rb1 and Trp53 cooperate to suppress prostate cancer lineage plasticity, metastasis, and antiandrogen resistance. *Science* **355**, 78–83 (2017).
- Puca, L., Vlachostergios, P. J. & Beltran, H. Neuroendocrine differentiation in prostate cancer: emerging biology, models, and therapies. *Cold Spring Harb. Perspect. Med.* **9**, a030593 (2019).
- Abida, W. et al. Genomic correlates of clinical outcome in advanced prostate cancer. *Proc. Natl Acad. Sci. USA* **116**, 11428–11436 (2019).
- Annala, M. et al. Circulating tumor DNA genomics correlate with resistance to abiraterone and enzalutamide in prostate cancer. *Cancer Discov.* **8**, 444–457 (2018).
- Chen, W. S. et al. Genomic drivers of poor prognosis and enzalutamide resistance in metastatic castration-resistant prostate cancer. *Eur. Urol.* **76**, 562–571 (2019).
- Wyatt, A. W. et al. Genomic alterations in cell-free DNA and enzalutamide resistance in castration-resistant prostate cancer. *JAMA Oncol.* **2**, 1598–1606 (2016).
- Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096 (2013).
- Ahrens-Fath, I., Politz, O., Geserick, C. & Haendler, B. Androgen receptor function is modulated by the tissue-specific AR45 variant. *FEBS J.* **272**, 74–84 (2005).
- Abeshouse, A. et al. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
- Fagerberg, L. et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell Proteom.* **13**, 397–406 (2014).
- Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
- Arora, V. K. et al. Glucocorticoid receptor confers resistance to antiandrogens by bypassing androgen receptor blockade. *Cell* **155**, 1309–1322 (2013).
- Beltran, H. et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22**, 298–305 (2016).
- Cato, L. et al. ARv7 represses tumor-suppressor genes in castration-resistant prostate cancer. *Cancer Cell* **35**, 401–413 (2019).
- Hu, R. et al. Distinct transcriptional programs mediated by the ligand-dependent full-length androgen receptor and its splice variants in castration-resistant prostate cancer. *Cancer Res.* **72**, 3457–3462 (2012).
- Yuan, F. et al. Molecular determinants for enzalutamide-induced transcription in prostate cancer. *Nucleic Acids Res.* **47**, 10104–10114 (2019).
- Hwang, J. H. et al. CREB5 promotes resistance to androgen-receptor antagonists and androgen deprivation in prostate cancer. *Cell Rep.* **29**, 2355–2370.e6 (2019).
- Li, Y. et al. Targeting cellular heterogeneity with CXCR2 blockade for the treatment of therapy-resistant prostate cancer. *Sci. Transl. Med.* <https://doi.org/10.1126/scitranslmed.aax0428> (2019).
- Ragnum, H. B. et al. The tumour hypoxia marker pimonidazole reflects a transcriptional programme associated with aggressive prostate cancer. *Br. J. Cancer* **112**, 382–390 (2015).
- Saal, L. H. et al. Poor prognosis in carcinoma is associated with a gene expression signature of aberrant PTEN tumor suppressor pathway activity. *Proc. Natl Acad. Sci.* **104**, 7564–7569 (2007).

26. Ertel, A. et al. RB-pathway disruption in breast cancer. *Cell Cycle* **9**, 4153–4163 (2010).
27. DeTomaso, D. et al. Functional interpretation of single cell similarity maps. *Nat. Commun.* **10**, 1–11 (2019).
28. Rodrigues, D. N. et al. Immunogenomic analyses associate immunological alterations with mismatch repair defects in prostate cancer. *J. Clin. Invest.* **128**, 4441–4453 (2018).
29. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
30. Norris, J. D. et al. The homeodomain protein HOXB13 regulates the cellular response to androgens. *Mol. Cell* **36**, 405–416 (2009).
31. Duhon, T. et al. Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. *Nat. Commun.* **9**, 1–13 (2018).
32. Miller, B. C. et al. Subsets of exhausted CD8⁺ T cells differentially mediate tumor control and respond to checkpoint blockade. *Nat. Immunol.* **20**, 326–336 (2019).
33. Hudson, W. H. et al. Proliferating transitory T cells with an effector-like transcriptional signature emerge from PD-1⁺ stem-like CD8⁺ T cells during chronic infection. *Immunity* **51**, 1043–1058 (2019).
34. Zander, R. et al. CD4⁺ T cell help is required for the formation of a cytolytic CD8⁺ T cell subset that protects against chronic infection and cancer. *Immunity* **51**, 1028–1042 (2019).
35. Yamauchi, T., Hoki, T., Odunsi, K. & Ito, F. Identification of dysfunctional CD8⁺ T-cell subsets rescued by PD-L1 blockade in the tumor microenvironment. *J. Immunol.* **200**, 58.3 (2018).
36. Graff, J. N. et al. Early evidence of anti-PD-1 activity in enzalutamide-resistant prostate cancer. *Oncotarget* **7**, 52810–52817 (2016).
37. Sweeney, C. J. et al. Abstract CT014: IMbassador250: A phase III trial comparing atezolizumab with enzalutamide vs enzalutamide alone in patients with metastatic castration-resistant prostate cancer (mCRPC). *Cancer Res.* **80**, CT014–CT014 (2020).
38. Mancuso, N. et al. Large-scale transcriptome-wide association study identifies new prostate cancer risk regions. *Nat. Commun.* **9**, 1–11 (2018).
39. Schumacher, F. R. et al. Association analyses of more than 140,000 men identify 63 new prostate cancer susceptibility loci. *Nat. Genet.* **50**, 928 (2018).
40. Finucane, H. K. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* **50**, 621–629 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021

Methods

Statistical computing. Except where otherwise specified, analysis and data visualization were performed with Python 3.7, SciPy v.1.3.2, Matplotlib v.3.1.1, seaborn v.0.9.0 and R v.3.5.1.

Reference versions. We used human genome reference b37 and the GENCODE release 30 gene annotation lifted over to GRCh37.

Whole exome analyses. For biopsies with paired tumor and normal samples available, we performed whole exome sequencing with a customized version of a previously described protocol⁴¹. After DNA shearing, hybridization and exome capture were performed using Illumina's Rapid Capture Exome kit (except for the normal sample for 01115149 and the tumor sample for biopsy 01115149-TA, which used the Agilent SureSelect Human All Exon 44Mb v.2.0 bait set). Libraries were sequenced with 76 bp paired-end reads on an Illumina instrument.

Reads were aligned using BWA v.0.5.9 (ref. ⁴²). Somatic mutations were called using a customized version of the Getz Lab CGA WES Characterization pipeline (https://portal.firecloud.org/#methods/getzlab/CGA_WES_Characterization_Pipeline_v0.1_Dec2018/2) developed at the Broad Institute. Copy number alterations, purity, ploidy and whole genome doubling status were called using FACETS v.0.5.14 (ref. ⁴³; Supplementary Table 2). Copy number alterations were evaluated with respect to whole genome doubling status. Summary visualization of mutation and copy number status integrated with clinical variables was performed with a beta version of CoMut⁴⁴.

For biopsies where whole exome sequencing failed, somatic mutation calls, copy number alterations and ETS fusion status were taken where available from OncoPanel, a clinical panel sequencing test available at Dana-Farber Cancer Institute.

Sample collection and dissociation for single-cell RNA-seq. Samples were collected with informed consent and ethics approval by the Dana-Farber/Harvard Cancer Center Institutional Review Board under protocol nos. 09-171, 11-104, 13-301 and 01-045. Tumor samples were collected and transported in Dulbecco's modified Eagle medium, high glucose, pyruvate (DMEM; Thermo Fisher Scientific, 11995073) on ice. Single-cell suspensions for scRNA-seq were obtained from tumor core needle biopsies through mechanical and enzymatic dissociation. Samples were first cut into pieces <1 mm³ using a scalpel. For bone biopsies, soft tissue was also scraped from the hard bone surface using a scalpel blade. Samples were then dissociated using one of two protocols, chiefly to optimize for yield of viable cells from different metastatic sites. Cells obtained from the two protocols were comparable and findings were consistent in subanalyses of cells processed with the same protocol (Supplementary Fig. 7).

For biopsies 01115655-TC, 01115666-TA, 01115680, 01115681, 09171111, 09171135, 09171136 and 09171139, resulting tissue fragments were incubated in 3 ml Accumax (Innovative Cell Technologies, AM105) for 10 min at room temperature on a rocking shaker (ACC protocol). Cell suspensions were then filtered with a 100- μ m cell strainer (Thermo Fisher Scientific, 08-771-19) and spun at 580g for 5 min at 4°C. In cases where cell pellets appeared bloody, red blood cells were lysed with ACK Lysing Buffer (Thermo Fisher Scientific, A1049201) on ice for 1 min, followed by quenching with PBS and an additional centrifugation. The final cell pellet was resuspended in PBS (Thermo Fisher Scientific, MT21040CV) with 2% FBS (Gemini Bio-Products, 100-106).

For biopsies 01115655-TA, 01115665, 01115149-TC, 01115578-TA, 09171123, 09171144 and 09171146, tissue fragments were incubated in 2–3 ml Medium 199, Earle's Salts (M199, Thermo Fisher Scientific, 11150059) with 1 mg ml⁻¹ Collagenase 4 (Thermo Fisher Scientific, NC9836075) and 10–20 μ g ml⁻¹ DNase I (StemCell Technologies, 7900) for 5–10 min in a 37°C water bath with intermittent mixing, followed by additional mixing and pipetting (CD protocol). Cell suspensions were then filtered with a 100- μ m cell strainer, spun at 580g for 5 min at 4°C and the resulting pellet resuspended in PBS with 2% FBS. The blood clot from biopsy 09171144 was processed in a similar manner, with the exception that red blood cells were lysed with ACK Lysing Buffer on ice at 5-min increments for a total of 15 min. For the bone marrow aspirate from biopsy 09171144, mechanical and enzymatic dissociation were not performed and red blood cells were lysed with ACK Lysing Buffer on ice at 5-min increments for a total of 10 min.

Single-cell sorting. Dissociated single cells were enriched for cancer cells, CD45⁺ immune cells and, for 09171144, other populations. Single-cell suspensions in PBS with 2% FBS were stained by incubating for 15 min at room temperature protected from light with anti-human PTPRC (CD45) monoclonal antibody conjugated to FITC (1:200 dilution, VWR ABNOMAB12230), anti-human EPCAM antibody conjugated to PE (1:50 dilution, Miltenyi Biotec, 130-091-253) and either Calcein-AM (1:200 dilution, Thermo Fisher Scientific, C3100MP; biopsies 01115655-TA and 01115665), 7-aminoactinomycin D (7-AAD) (1:200 dilution, Thermo Fisher Scientific, A1310; all other biopsies except sample 01115149-TC) or both (sample 01115149-TC). We first sorted cells with biological dimensions (high FSC-A and high SSC-A), selected single cells and excluded doublets or

triplets (low FSC-W). Next, we sorted live cells (low 7AAD/ high Calcein-AM) that were CD45⁺ (high FITC, enriching for immune cells), EPCAM⁺ (high PE, enriching for cancer cells) or double negative (low FITC/low PE, only in biopsy 09171144) (Supplementary Fig. 8 shows example gating). When possible, we collected at least one full plate of each gating, but as we did not collect all live cells, the representation of immune versus cancer cells in our dataset do not reflect composition within tumors. Cell sorting was performed using a BD Biosciences FACSAria cell sorter (Ilu or UV) with FACSDiva software. Individual cells were sorted into the wells of 96-well plates with 10 μ l TCL buffer (QIAGEN, 1070498) with 1% β -mercaptoethanol (Sigma, 63689) per well. Plates were then sealed, vortexed for 10 s, spun at 3,700 r.p.m. for 2 min at 4°C and frozen on dry ice.

Transcriptome sequencing, alignment and quantification. Library preparation for bulk RNA-seq was performed using the Illumina TruSeq Stranded mRNA Sample Preparation kit (except for biopsy 01115149-TA, which was prepared using the unstranded Illumina TruSeq RNA Sample Preparation protocol (Revision A, 2010)). Libraries were sequenced with 101 bp paired-end reads (except biopsy 01115149-TA, which was sequenced with 76 bp paired-end reads) on an Illumina instrument.

For scRNA-seq, RNA was captured from single-cell lysates with 2.2 \times RNAClean SPRI beads (Beckman Coulter Genomics) without the final elution. After air drying and secondary structure denaturation at 72°C for 3 min, library construction was performed using a slightly customized Smart-seq2 protocol¹² with 21 cycles of PCR for preamplification. Complementary DNA (cDNA) was purified with 0.8 \times Ampure SPRI beads (Beckman Coulter Genomics) and eluted in 21 μ l TE buffer. During tagmentation and PCR amplification, we used 0.2 ng of cDNA per cell and one-eighth of the Illumina NexteraXT (Illumina FC-131-1096) reaction volume. Individual cells were sequenced to a mean depth of ~1.5 million 38 bp paired-end reads on an Illumina NextSeq 500 instrument with 75 cycle high output kits (Illumina TG-160-2005).

After adaptor trimming, reads were aligned using STAR aligner v.2.7.2b⁴⁵ with parameters: `--outFilterMultimapNmax 20 --outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 --alignIntronMin 20 --alignMatesGapMax 1250000 --alignIntronMax 1250000 --chimSegmentMin 12 --chimJunctionOverhangMin 12 --alignSjstitchMismatchNmax 5 -1 5 5 --chimMultimapScoreRange 3 --chimScoreJunctionNonGTAG -4 --chimMultimapNmax 20 --chimNonchimScoreDropMin 10 --peOverlapNbasesMin 12 --peOverlapMMP 0.1 --chimOutJunctionFormat 1`. sjdbOverhang was set to 1 less than the untrimmed read length. We used multi-sample two-pass mapping for all samples from each patient, first mapping all samples (bulk and single-cell transcriptomes), merging the SJ.out.tab files, then running the second pass with the jointly called splice junctions. For samples with germline and/or somatic variant information available from bulk exome sequencing, we supplied a VCF specifying single-nucleotide variants for STAR's variant-aware mapping via the `--varVCFfile` parameter. STAR BAMs were passed into Salmon v.0.14.1 (ref. ⁴⁶) to generate gene-level TPM quantifications with parameters: `--incompatPrior 0.0 --seqBias --gcBias --reduceGCMemory --posBias`. STAR chimeric junctions were supplied to STAR-Fusion⁴⁷ v.1.7.0 in kickstart mode to call ETS family fusions.

Single-cell quality control and clustering. After removing low quality cells (<500 or >10,000 detected genes, <50,000 reads or >25% expression from mitochondrial genes), we used Seurat v.3.1.0 (ref. ⁴⁸) to perform first-pass clustering using the TPM matrix rescaled to exclude mitochondrial genes. We manually identified and removed a small number of cells with anomalous expression patterns (chiefly coexpression of high levels of hemoglobin with marker genes for nonerythroid cells). Additionally, some cells that did not cluster with erythroid cells (easily identified with dominant hemoglobin expression) nonetheless had low levels of hemoglobin detected, suggestive of contamination from ambient RNA released from lysed erythroid cells. To account for this, we identified genes whose expression was correlated (Pearson correlation > 0.2) with total hemoglobin expression levels in nonerythroid cells with detectable hemoglobin. This consisted of a small set of genes with known function in erythroid cell development and function: *AHSP*, *GATA1*, *CA1*, *EPB42*, *KLF1*, *SLC4A1*, *CA2*, *GYP A*, *TFR2*, *RHAG*, *FAXDC2*, *RHD*, *ALAS2*, *SPTA1* and *BLVRB*. To mitigate batch effects driven by different degrees of contaminating ambient erythroid transcripts, we removed these genes, along with the genes encoding hemoglobin subunits, from the expression matrix for all nonerythroid cells. Following this, we did not observe strong batch effects, as evidenced by co-mingling of nonmalignant cells from different biopsies (Extended Data Fig. 1d).

We repeated the clustering and conducted all downstream analyses with the filtered expression matrix. After joint clustering of all cells (Extended Data Fig. 1c), we performed subclustering on three cell subsets: (1) NK and T cells, (2) B-lineage cells and (3) myeloid cells. We manually labeled clusters by dominant cell identity, as assessed by marker gene expression patterns (Extended Data Fig. 2 and Supplementary Table 4). Briefly, cancer cell clusters ($n = 836$, from 12 biopsies) were identified by expression of *AR* and *KLK3* (adenocarcinoma markers) or

CHGA (NEPC marker); T cell populations by *CD3D* and *CD3G*; regulatory T cells by *CD4*, *FOXP3* and *CTLA4*; NK cells by absence of *CD3D* and *CD3G* and expression of *FCGR3A*, *FCGR3B* and *GZMB*; erythroid cells by *HBA* and *HBB*; neutrophils by *ELANE*, *CEACAM8*, *AZU1* and *DEFA1*; macrophages by *APOE*, *CIQA* and *CIQB*; monocytes by *ITGAX*, *CD14*, *FCGR3A* and *FCGR3B*; B cells by *CD19* and *MS4A1*; plasmablasts by *CD19* and absence of *MS4A1*; and plasma cells by *SDC1* and high expression of immunoglobulin genes. Additionally, we confirmed the identity of cancer cell clusters by matching transcriptome-inferred copy number alteration profiles generated from inferCNV v.0.99.7 (<https://github.com/broadinstitute/inferCNV>) with those obtained from corresponding bulk whole exome sequencing.

AR isoform-specific reads. Our scRNA-seq dataset had relatively even sequencing coverage along transcripts as is characteristic of Smart-seq2. To identify reads that uniquely map to an AR isoform, we generated a FASTA transcriptome annotation of spliced sequences from isoforms described in literature (Supplementary Fig. 1)^{249–55}. This allowed us to use a mapping-based approach to quantify reads mapping to specific AR isoform transcripts. This approach is more comprehensive than approaches considering only junction reads and it yields estimates of AR-V7 abundance that are highly correlated with published estimates in existing cohorts (Supplementary Fig. 2a,b). We extracted all reads initially mapped by STAR (using the GENCODE transcriptome annotation) to the AR genomic interval (and flanking intergenic region) X:66753830–67011796 and then remapped them to our AR isoform transcriptome, disallowing clipping, multimapping or chimeric reads and requiring end-to-end mapping (STAR parameters: `--outFilterMultimapNmax 1 --alignEndsType EndToEnd --alignSoftClipAtReferenceEnds No --outFilterMismatchNmax 999 --outFilterMismatchNoverReadLmax 0.04 --peOverlapBasesMin 10 --scoreGap -1000000`). As our AR isoform annotation corresponded to transcript sequences after splicing, we set `scoreGap` to prohibitively penalize gaps introducing additional inferred splice events. For patients with germline or somatic AR single-nucleotide variants detected either via whole exome sequencing or OncoPanel, we transformed single-nucleotide variants to transcript coordinates and supplied this as a VCF to STAR using the `--varVCFfile` parameter for STAR's variant-aware mapping. We reported all reads that mapped uniquely to an isoform with at most one mismatch in Fig. 1a.

Alignment to specific isoforms requires detection of identifying transcript features (for example a unique exon–exon boundary, a portion of an exon not in other isoforms or a unique combination of exons). Illumina reads are not long enough to cover full transcripts in a single read, so not all isoforms are equivalently detectable. For example, reads cannot be attributed to AR-FL, as inclusion of all its component exons is impossible in Illumina reads and it has no unique splice junctions or exons. Thus, we avoid comparing frequencies of isoform-specific reads from different isoforms in the same sample/cell. We can, however, compare levels of the same isoform across samples/cells.

Validation in the three bulk cohorts proceeded in the same fashion, except that for the 4 healthy prostate samples, the initial STAR genomic alignment step was run with `--twopassMode Basic`. Reads were downloaded as FASTQs for healthy prostate (ArrayExpress accession E-MTAB-1733) and TCGA (ISB-CGC, <https://isb-cgc.appspot.com/>). CRAMs for the Abida et al. cohort⁸ were downloaded from Amazon S3 (dbGaP accession code, [phs000915.v2.p2](https://www.ncbi.nlm.nih.gov/bioproject/5000000)) and converted to FASTQs with samtools fastq. Survival analyses incorporating AR variant expression were performed using the Python package lifelines v.0.23.9.

To test for AR-45 and AR-V7 mutual exclusivity, we permuted assignments of AR region reads to single cells for all cells with any isoform-specific reads, while keeping fixed the AR read depth for each cell. This represents a null distribution under which all cells from a biopsy have identical AR isoform expression patterns while preserving the per-cell AR expression levels observed in our data. We performed 10,000 random permutations and compared the number of cells in which there was AR-45 and AR-V7 coexpression (at least one isoform-specific read for each) under the null to that observed in our data.

Gene set scoring. For both bulk samples and single cells, we scored the activity of gene sets with VISION v.2.0.0 (ref. 27). Curated gene sets are shown in Supplementary Table 5. The hallmark gene sets were obtained from v.7.0 of the Molecular Signatures Database (MSigDB)¹⁶.

Analyzing cancer cell scRNA-seq data collected from patient biopsies poses statistical challenges. Cancer cells generally show patient-specific expression patterns that cause two cancer cells from a given patient to resemble each other more than any two cancer cells from different patients. This is driven substantially by patient-specific patterns of large somatic copy number alterations, which cause changes in expression of the affected genes to a degree that the copy number events can be inferred from scRNA-seq and used to cluster patient cancer cells. We aimed to uncover common, generalizable changes in patterns of expression related to enzalutamide exposure and sought to downweigh phenomena unique to individual patients. Patient-specific mechanisms can be detected with high power when cancer cells from a biopsy are particularly unusual or when they represent a larger than typical fraction of the total cohort of cells. To counter these effects, we

developed a subsampling procedure detailed next and visualized in Supplementary Fig. 3. By considering subsets of the data more balanced for representation from different biopsies, we traded reduced power for more robustness.

When comparing VISION scores in cells from biopsies exposed and naive to treatment with enzalutamide, we included only cells inferred to be in G1 by Seurat to reduce discovery of signals introduced by different proportions of cycling cells between tumors⁴⁸. We restricted our initial analyses to biopsies with at least ten G1 cancer cells. From either class (enzalutamide-naive versus exposed), we sampled up to 20 cells per biopsy to prevent results from being dominated by tumors with many recovered cells. Additionally, across repeated sampling iterations, we took a leave-one-out sampling procedure across biopsies. We omitted each biopsy in turn, instead sampling cells from other biopsies within its class, keeping the total number of cells the same. We performed 501 iterations of sampling for each biopsy being excluded. For each gene set being scored with VISION, we used the sampling with the median effect size as the summary of all iterations. When measuring effect size, we consistently compared one class versus the other (that is always exposed relative to naive) to ensure consistency in comparisons of direction of effect. We used the corresponding two-sided Mann–Whitney *U*-test *P* value as the nominal *P* value for the given gene set.

We additionally took the following steps to filter results that seemed to be driven by a single biopsy: for any given biopsy, we compared samplings when cells from the biopsy were held out versus when cells from the biopsy were included. If the proportion of nominally significant results ($P < 0.05$, same direction of effect as the overall median outcome for the given signature) when the biopsy was excluded was $< 80\%$ of the proportion of nominally significant results when the biopsy was included, we considered any overall finding of differential gene set expression as nonrobust. This step was performed after computing Benjamini–Hochberg false discovery rate (FDR)⁵⁶. Note that *P* values shown in Fig. 2b,c are based on all G1 cells and confirmed the findings from this sampling approach.

Regulon activity. As only a single biopsy had a small cell carcinoma histology, we opted against gene-level differential expression analysis to avoid discoveries unique to one patient. Instead, we inferred transcriptional regulatory factor activity using all cancer cells from our cohort using SCENIC v.1.1.2.2²⁹. This approach integrates information from patterns of gene expression covariation within our single-cell data with transcription factor binding site information to infer the sets of genes (regulons) upregulated by particular transcription factors and the activity of regulons within individual cells²⁹. By focusing on inferred transcription factor activity as opposed to comparing individual genes, this approach has been previously shown to identify more generalizable patterns in expression programs in cancer²⁹. In Fig. 3, for single cells, we used SCENIC AUC directly as a measure of regulon activity. For Fig. 3c,e, to infer regulon activity in bulk samples, we extracted gene sets corresponding to regulons from SCENIC and scored bulk samples for activity of the genes sets using VISION^{18,27}.

We compared inferred regulon activities between small cell carcinoma and adenocarcinoma cells within our cohort using a similar subsampling approach as detailed previously for comparisons of gene set scores, except that in comparing SCENIC AUC scores, we did not restrict to only G1 cells, as the regulons had been inferred with all cancer cells together. As there was one small cell carcinoma biopsy, cells from that biopsy were never selected for omission across samplings.

Signature scoring in bulk RNA-seq from Abida et al. cohort. Clinical annotations and expression quantifications were obtained from the published supplementary materials and from the authors directly⁸. We converted gene expression values from fragments per kilobase of transcript per million mapped reads (FPKM) (as supplied by the authors) to TPM for consistency with the rest of our study. As this cohort included samples sequenced at different centers and from different metastatic sites, we further restricted our analyses to avoid batch effects. For Fig. 2d, we analyzed only samples sequenced via transcriptome capture at the University of Michigan, as this was the largest identifiably uniformly sequenced subset. We could not analyze bone biopsies in this cohort due to paucity of post-enzalutamide samples. For Fig. 3c, as the largest number of small cell carcinoma samples were sequenced at Cornell, we included only small cell carcinoma and adenocarcinoma cases from Cornell in our analyses.

Cell culture and western blots. The enzalutamide-resistant cell line VCaP-16 was maintained in DMEM with 10% FBS and 16 μ M enzalutamide (SelleckChem, S1250), while parental VCaP-D was maintained in DMEM with 10% FBS. VCaP-16 and VCaP-D cells were plated into six-well plates in maintenance medium at a density of 50% and allowed to attach for 48 h. VCaP-16 cells are available from S.P.B. upon request. Cells in assay wells were then washed twice with PBS and switched into basal maintenance medium with 0.5% FBS for 48 h (VCaP-16 basal medium still contains 16 μ M enzalutamide). Select samples were pretreated with the TGF- β receptor inhibitor SB-431542 (SelleckChem, S1067) for the last 24 of the 48 h. Following the 48-h incubation period, cells were treated for 2 h with fresh 0.5% FBS maintenance medium containing SB-431542, recombinant TGF- β -1 protein (R&D Systems, 240-B-002), or a combination of both. After 2 h, medium was aspirated from all samples and each well was rinsed briefly in PBS.

Cells were lysed in RIPA buffer with protease and phosphatase inhibitor cocktail (Thermo Fisher Scientific, P178440) on a vortex shaker for 30 mins at 4 °C. Lysates were spun at 16,000g for 15 mins at 4 °C and supernatant containing the soluble protein fraction was transferred to a new tube. Protein lysates were added to loading buffer (2× Laemmli Sample Buffer, Bio-Rad, 1610737) and the samples were boiled at 90 °C for 7 mins. Samples were electrophoresed at 200 V using a 4–20% denaturing PAGE gel (Bio-Rad). Relative protein levels of phospho-SMAD2 (Ser465/467), SMAD2/3 and vinculin were determined via immunoblotting using the following antibodies/dilutions: phospho-SMAD2 (Ser465/467) (138D4) antibody (Cell Signaling Technologies, 3108T) at 1:1,000 dilution, SMAD2/3 antibody (Cell Signaling Technologies, 3102) at 1:1,000 dilution and vinculin antibody (7F9) (Santa Cruz Biotechnology, sc-73614) at 1:10,000 dilution. The experiment and blotting was performed once.

To assess response to short-term enzalutamide exposure, wild-type VCaP cells were seeded in DMEM with 10% FBS in 24-well plates and allowed to attach for 24 h. The cells were treated with 10 μM enzalutamide for 24, 48 or 72 h or DMSO for 72 h. Recombinant TGF-β-1 protein was added 1 h before collecting cells for lysis and western blot. Antibodies are as previously stated, with the addition of anti-PSA antibody (rabbit polyclonal, Bidesign, K9211DR) at 1:10,000 dilution and β-actin HRP-conjugate antibody (Abcam, ab20272) at 1:20,000 dilution.

pSMAD2 immunohistochemistry. Samples frozen in OCT were cryosectioned at 5-μm thickness, fixed in 10% neutral buffered formalin for 12 h and washed in 1× PBS. For each biopsy, slides were cut from multiple tissue blocks ($n = 2–6$, depending on available tissue) and evaluated for tumor content, with representative images shown from regions with sufficient tumor cellularity. Immunohistochemical staining of phospho-SMAD2 was performed on the automated Leica Bond RX system, using the Bond Polymer Refine kit with anti-phospho-SMAD2 (Ser465/467) antibody (E8F3R, rabbit monoclonal, Cell Signaling Technology) at a dilution of 1:100 overnight at 4 °C.

HOXB5, HOXB6, NR1D2 immunohistochemistry. Patient-derived organoids of NEPC were developed from metastatic biopsies as previously described³⁷. NEPC organoids (OWCM154, OWCM155, OWCM1078 and OWCM1262) and prostate adenocarcinoma (LNCaP and VCaP) xenograft models were formalin-fixed, paraffin embedded and serial sectioned at 4 μm thickness. Formalin-fixed, paraffin embedded slides were immersed in xylene and ethanol solutions for deparaffinization and dehydration. Antigens were retrieved in boiled 10 mM sodium citrate buffer (Sigma-Aldrich). Slides were blocked using horse serum and incubated with primary antibodies at 4 °C overnight. Slides were washed and developed using the ABC kit (Vector Laboratories), followed by hematoxylin counterstaining. Each stain was performed twice independently on the same batch of freshly sectioned slides and at least three images were taken from each in low and high magnification, with representative images shown in Fig. 3f. Dilutions for primary antibodies were anti-NR1D2 (1:50, NOVUS, NBP2-56141), anti-HOXB5 (1:10, Sigma, HPA043851) and anti-HOXB6 (1:200, Abcam, ab219499).

TCR reconstruction. We performed TCR reconstruction and clonotype inference from scRNA-seq with TraCeR v.0.6.0 (ref. ³⁸). We performed TCR reconstruction and estimation of clonal fraction from bulk RNA-seq using MiXCR v.3.0.12 (ref. ³⁹). TCRs were inferred as detected in both bulk and scRNA-seq if the CDR3 nucleic acid (and therefore amino acid) sequence matched.

Expression of prostate cancer risk genes. For Extended Data Fig. 5a, we examined expression of a published set of prostate cancer risk genes³⁸. For each cluster of cells, we computed the fraction of cells expressing each gene at TPM ≥ 1. For nonmalignant cells, we combined cells from different biopsies in each cluster. For cancer cells, we considered each biopsy separately as its own cluster. We further grouped cell clusters into ‘superclusters’ of related cell types (Supplementary Table 4 and Extended Data Fig. 5). We deemed a gene to be cell type-specific if, in the cluster where it was most frequently expressed, it was expressed in at least a quarter of the cells and if the proportion expressing was at least 2.5-times that of every cluster in any other supercluster. This allows the identification of, for example, *PIK3C2B*, which is highly expressed in both B cells and plasmablasts but only sparsely in unrelated cell types.

Cluster-specific genes and LDSC-SEG. LDSC-SEG considers genome-wide signals of disease association, including mostly weak, nonsignificant signals and determines whether they are significantly concentrated in subsets of the genome. More specifically, LDSC-SEG examines heritability in regions of the genome surrounding genes specifically expressed in different cell populations to nominate cell types relevant to disease biology⁴⁰. We grouped cell clusters into ‘superclusters’ of related cell types as before and performed differential expression to identify markers for each cell cluster, omitting cells in the same supercluster. To mitigate uneven representation of cell types, when comparing against any cluster, we subsampled the same number of cells from each other supercluster and used as even a representation as possible of the contained clusters. In determining cancer cell markers, we used as even a representation as possible of cells from each biopsy while sampling 200 cancer cells total per iteration. For each cluster, we repeated

sampling 500 times. In each sampling, we performed a one-sided Mann–Whitney *U*-test for differential expression on all genes with at least one TPM expression in at least 10% of the cluster’s cells. We then selected the top 10% most-upregulated genes (lowest median *P* value across samplings) as cluster specifically expressed genes. We used a 100 kb interval around genes for heritability partitioning with LDSC-SEG v.1.0.1, additionally including an annotation corresponding to all genes and the baseline v.1.1 model⁴⁰.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

scRNA-seq expression and clustering data generated in this study are available at https://singlecell.broadinstitute.org/single_cell/study/SCP1244/transcriptional-mediators-of-treatment-resistance-in-lethal-prostate-cancer. Raw sequence data generated in this study are being deposited in dbGaP (accession [phs001988.v1.p1](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=phs001988.v1.p1)). FASTQs for healthy prostate RNA-seq are from ArrayExpress under accession [E-MTAB-1733](https://www.ebi.ac.uk/ena/arrayexpress/experiments/E-MTAB-1733). TCGA data are available at dbGaP accession [phs000178.v1.p8](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=phs000178.v1.p8). Abida et al. data are available at dbGaP accession [phs000915.v2.p2](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=phs000915.v2.p2). The hallmark gene sets from MSigDB can be accessed at <https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp#H>. Source data are provided with this paper.

References

- Fisher, S. et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* **12**, R1 (2011).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Shen, R. & Sheshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucl. Acids Res.* **44**, e131 (2016).
- Crowdis, J., He, M. X., Reardon, B. & Van Allen, E. M. CoMut: visualizing integrated molecular information with comutation plots. *Bioinformatics* **36**, 4348–4349 (2020).
- Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon: fast and bias-aware quantification of transcript expression using dual-phase inference. *Nat. Methods* **14**, 417–419 (2017).
- Haas, B. J. et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biol.* **20**, 213 (2019).
- Butler, A., Hoffman, P., Smibert, P., Papalexis, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
- Hu, R. et al. Ligand-independent androgen receptor variants derived from splicing of cryptic exons signify hormone-refractory prostate cancer. *Cancer Res.* **69**, 16–22 (2009).
- Watson, P. A. et al. Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. *Proc. Natl Acad. Sci. USA* **107**, 16759–16765 (2010).
- Lu, C. & Luo, J. Decoding the androgen receptor splice variants. *Transl. Androl. Urol.* **2**, 178–186 (2013).
- Hu, D. G. et al. Identification of androgen receptor splice variant transcripts in breast cancer cell lines and human tissues. *Horm. Cancer* **5**, 61–71 (2014).
- Hu, R., Isaacs, W. B. & Luo, J. A snapshot of the expression signature of androgen receptor splicing variants and their distinctive transcriptional activities. *Prostate* **71**, 1656–1667 (2011).
- Sun, S. et al. Castration resistance in human prostate cancer is conferred by a frequently occurring androgen receptor splice variant. *J. Clin. Invest.* **120**, 2715–2730 (2010).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57**, 289–300 (1995).
- Puca, L. et al. Patient derived organoids to model rare prostate cancer phenotypes. *Nat. Commun.* **9**, 2404 (2018).
- Stubbington, M. J. T. et al. T cell fate and clonality inference from single-cell transcriptomes. *Nat. Methods* **13**, 329–332 (2016).
- Bolotin, D. A. et al. MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12**, 380–381 (2015).
- Chakravarty, D. et al. OncoKB: a precision oncology knowledge base. *JCO Precis. Oncol.* <https://doi.org/10.1200/PO.17.00011> (2017).

Acknowledgements

We thank the patients who participated in this study. We are grateful to Y. Huang and the Molecular Pathology Core Laboratory, Department of Oncologic Pathology, Dana-Farber Cancer Institute, for assistance with immunohistochemistry in patient samples. We

thank Irene I. Lee and M. Brown for helpful discussions on TGF- β signaling. Extended Data Fig. 1b was created with BioRender.com. This work was supported by the National Science Foundation (GRFP DGE1144152 (M.X.H.)), the National Institutes of Health (T32 GM008313 (M.X.H.), T32 CA009172 (K.P.B. and A.K.T.), K08 CA234458 (D.L.), K08 CA222663 (B.I.), R37 CA241486-01A1 (H.B.), R01 CA227388 (E.M.V.A.), R01 CA223484 (L.F.) and U01 CA233100 (L.F. and E.M.V.A.)), a Burroughs Wellcome Fund Career Award for Medical Scientists (B.I.), the Conquer Cancer Foundation (D.L. and N.I.V.), the Society for Immunotherapy of Cancer (D.L. and N.I.V.), a Damon-Runyon Physician Scientist Award (N.I.V.), the Prostate Cancer Foundation (L.F.; J.W.R. (18YOUN24); S.-Y.K. (18YOUN10); and M.-E.T. (16CHAS03)), a Mark Foundation Emerging Leader Award (E.M.V.A.), a Dana-Farber Medical Oncology Translational grant (A. Rotem and E.M.V.A.), the Ambrose Monell Foundation (E.M.V.A.), Dana-Farber Cancer Institute Pan Mass Challenge Funds (Team Fat Boys/Slim Sisters) (M.-E.T.) and a PCF-V Foundation Challenge Award (E.M.V.A.). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The results shown here are in part based upon data generated by the TCGA Research Network (<https://www.cancer.gov/tcga>).

Author contributions

M.X.H., A. Rotem, A. Regev, M.-E.T. and E.M.V.A. conceived and designed the overall study. M.S.C., A.K., M.-J.S. and C.R. processed the samples and prepared libraries for scRNA-seq. L.D., J.P., A. Rotem and O.R.-R. supervised and administered the sample collection and preparation. P.S., B.I. and A. Rotem developed the tissue dissociation methodology. Z.Z., L.D., Z.B., R.S., J.S. and A.K.T. organized patient participation and clinical data. M.X.H. and J.C. analyzed bulk exome sequencing data. S.Y.C. contributed to germline variant calling from exome sequencing. M.X.H., J.C., A.B.-M. and K.S. contributed to AR splicing analyses. M.X.H., D.L. and N.I.K. contributed to genomic analyses of enzalutamide resistance. A.R.S., J.W.R. and X.Y. performed preclinical TGF- β experiments with supervision by S.P.B. N.C.W. contributed to analysis of pSMAD2 immunohistochemistry. M.X.H. and K.B. performed single-cell regulon analyses. S.-Y.K. and M.M.G. performed immunohistochemistry experiments involving NR1D2 and HOXB6 with supervision by H.B. M.X.H., K.B. and K.P.B. analyzed immune cells. M.X.H. and J.C. analyzed clonotypes in bulk transcriptome sequencing. M.X.H. performed heritability enrichment analysis. M.X.H., L.F., M.-E.T. and E.M.V.A. interpreted the data. M.X.H., S.V., J.W.R. and S.-Y.K. wrote the manuscript. All authors reviewed and approved the final manuscript.

Competing interests

M.X.H. has been a consultant to Amplify Medicines and Ikena Oncology. Z.B. reports research support from Bristol-Meyers Squibb (BMS) and Genentech/imCORE unrelated to the current study. B.I. is a consultant for Merck and Volastra Therapeutics. D.L. reports funding by a postdoctoral fellowship from the Society for Immunotherapy of Cancer, which is funded in part by an educational grant from BMS. BMS has had no input into the conception, conduct or reporting of the submitted work. N.I.V. has served on an advisory board for Sanofi/Genzyme and is supported by a grant from the Society of Immunotherapy of Cancer that is funded in part by Genentech. L.F. reports receiving commercial research grants from AbbVie, Bavarian Nordic, BMS, Dendreon, Janssen, Merck and Roche/Genentech. S.P.B. served as an advisor for Sanofi. H.B. reports advisory/consulting from Janssen, Amgen, Astra Zeneca, Pfizer, Astellas, Sanofi Genzyme and research funding from Janssen, AbbVie Stemcentryx, Eli Lilly, Millenium and Astellas. A. Regev is a founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas Therapeutics and, until 31 August 2020, was a SAB member of Syros Pharmaceuticals, Neogene Therapeutics, Asimov and Thermo Fisher Scientific; from 1 August 2020, A. Regev is an employee of Genentech. A. Rotem is an employee of Astra Zeneca and an equity holder in NucleAI and Celsius Therapeutics. M.-E.T. reports advisory relationships with Celgene, Janssen, GSK/Parexel, Bayer, Astra Zeneca, Riovant, AbbVie, Arcus, Astellas, Pfizer, Constellation, Summus Global, MH Life Sciences/Intellisphere, Targeted Oncology and Aptitude Health. E.M.V.A. reports advisory relationships and consulting with Tango Therapeutics, Genome Medical, Invitae, Illumina, Enara Bio, Manifold Bio and Janssen; research support from Novartis and BMS; equity in Tango Therapeutics, Genome Medical, Syapse, Manifold Bio and Enara Bio; and travel reimbursement from Roche and Genentech, outside the submitted work.

Additional information

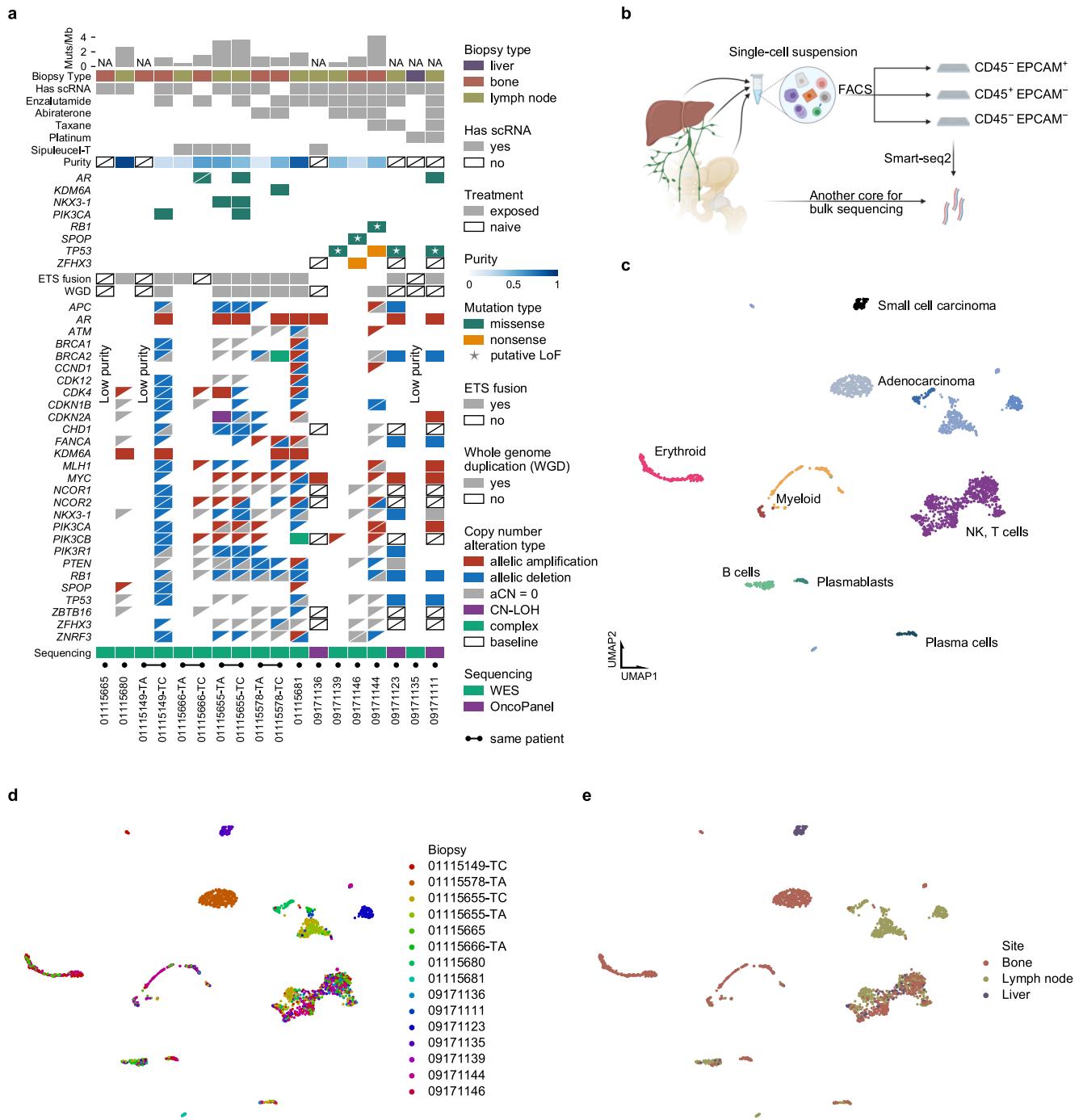
Extended data is available for this paper at <https://doi.org/10.1038/s41591-021-01244-6>.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-021-01244-6>.

Correspondence and requests for materials should be addressed to E.M.V.A.

Peer review information Javier Carmona was the primary editor on this article, and managed its editorial process and peer review in collaboration with the rest of the editorial team.

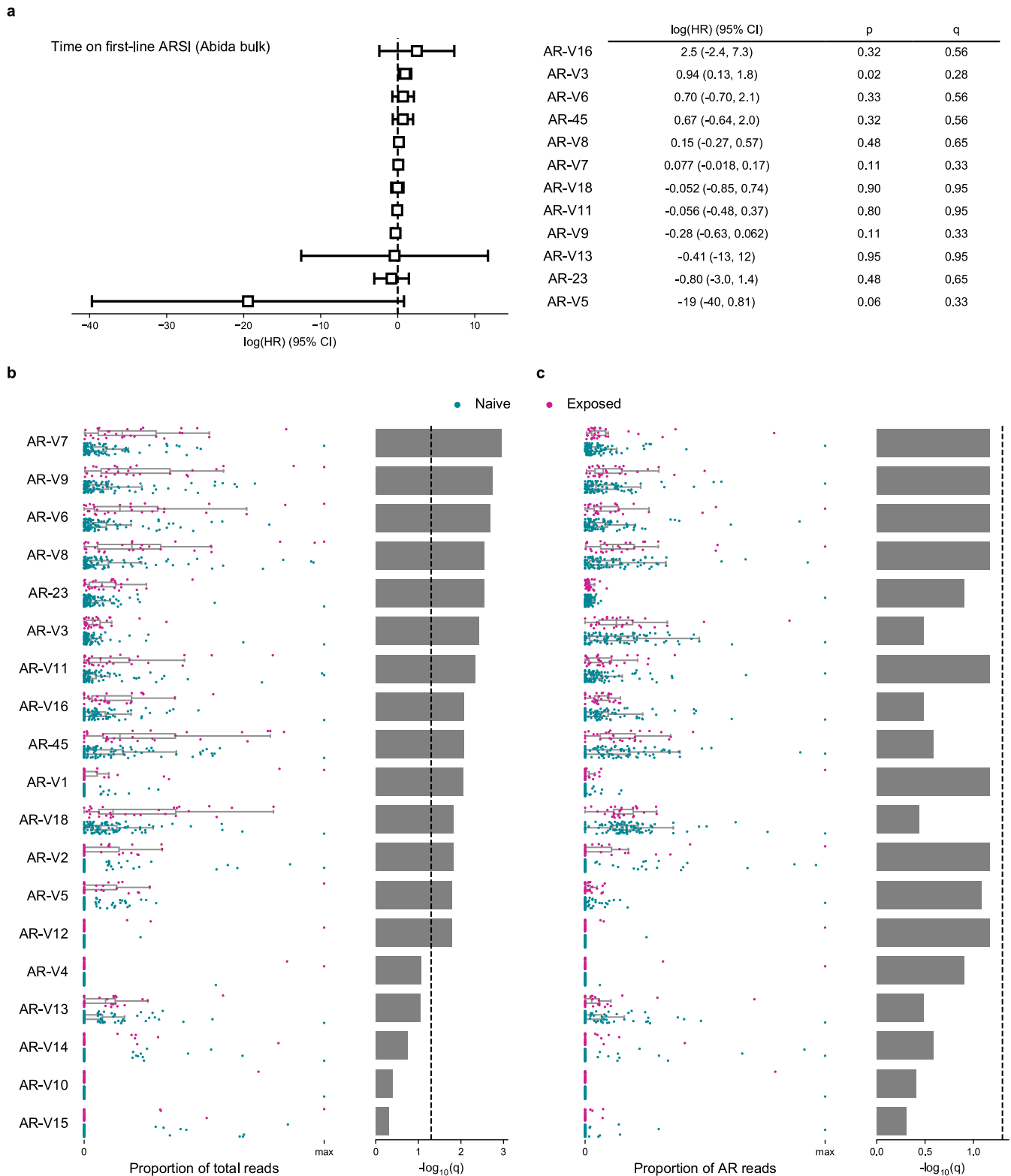
Reprints and permissions information is available at www.nature.com/reprints.



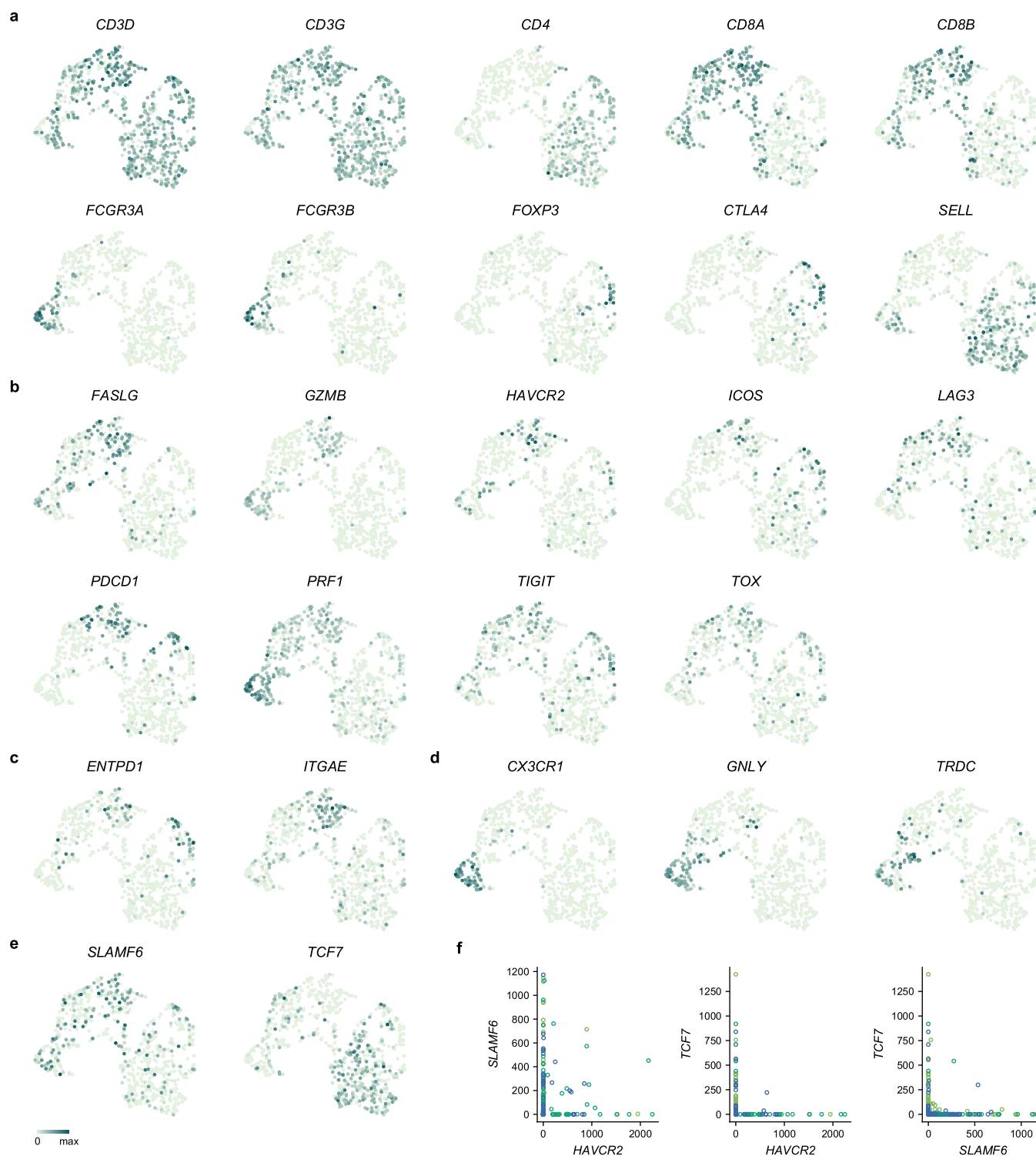
Extended Data Fig. 1 | Cellular atlas of mCRPC. a, Summary of clinical and select genomics features of patients and biopsies forming the study cohort. Each column represents a single biopsy. Where available, multiple biopsies from the same patient are displayed in adjacent columns. Patients are identified by numerical prefix, while suffixes after a dash, when present, identify biopsies from the same patient. Copy number calls based on whole exome sequencing (WES), but not OncoPanel, are allelic, with calls for the two alleles indicated by two triangles. In cases with whole genome doubling, it is possible for one allele to be amplified and one or both copies of the other allele to be lost. *AR* and *KDM6A* are on the X chromosome and so have only a single copy in these patients; they are represented with solid boxes for copy number status. Boxes with diagonal slashes indicate missing data, for example for genes not included in OncoPanel or for low tumor purity samples for which FACETS does not produce a purity estimate. Putative loss of function (LoF) missense mutations were annotated as LoF or likely LoF in OncoKB or mutated the same amino acid as a LoF mutation⁶⁰. **b**, Study design overview. Dissociated single cells were sorted to enrich for tumor (CD45⁻ EPCAM⁺), immune (CD45⁺ EPCAM⁻), or other populations (CD45⁻ EPCAM⁻). **c**, Projection of single-cell expression onto the first two dimensions of UMAP space. Each dot represents a single cell, and colors correspond to clusters identified by the Louvain algorithm. Clusters are manually labelled with dominant cell type(s) inferred from cluster-specific expression of marker genes. Cells colored corresponding to **(d)** biopsy of origin or **(e)** metastatic site. Non-malignant cells from different patients jointly cluster by cell type, while cancer cells from different patients largely form distinct clusters.



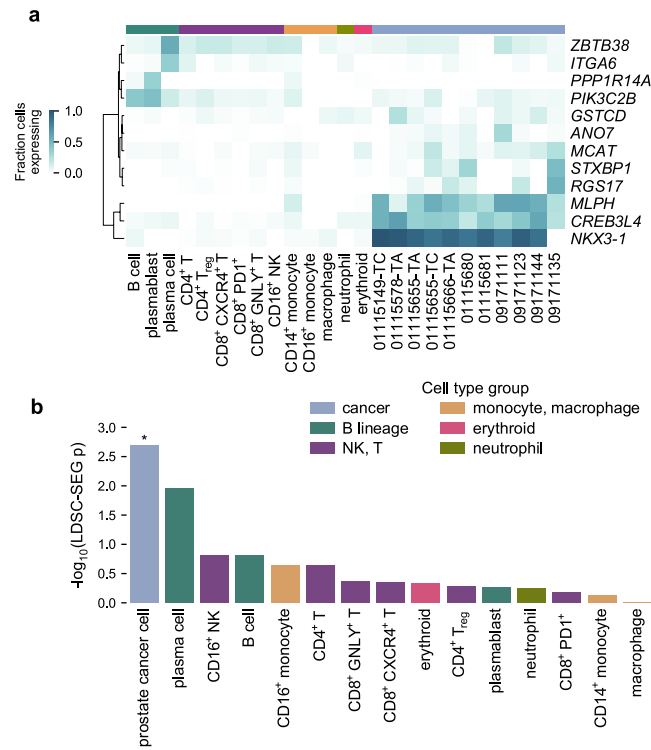
Extended Data Fig. 2 | Marker gene expression used for cluster labeling. Expression of select cell type marker genes for (a) prostate cancer cells (*AR* is expressed in adenocarcinoma, and *CHGA* is expressed in small cell carcinoma) (b) erythroid cells (c) T and NK cells (see also Extended Data Fig. 4) (d) neutrophils (e) macrophages (f) monocyte subsets, and (g) B lineage cells (see Methods for details on using combinations of markers to distinguish B cells, plasmablasts, and plasma cells). UMAP projections as in Extended Data Fig. 1c.



Extended Data Fig. 3 | Association of AR variant expression with clinical features in advanced prostate cancer. Analysis of isoform-specific read frequency in poly-A-sequenced samples with an adenocarcinoma histology from Abida et al. cohort⁸. **a**, Summary of multivariate Cox proportional hazards analysis of time on first-line androgen receptor signaling inhibitor (ARSI; abiraterone or enzalutamide) using AR variant frequency (isoform-specific reads / million total reads) in pre-treatment ARSI-naïve samples ($n = 46$). AR variants detected in at least 20% of samples were included. 95% confidence intervals and p -values are unadjusted for multiple hypotheses. q values correspond to Benjamini-Hochberg FDR. Comparison of expression levels of AR variants measured as a proportion of **(b)** all reads or **(c)** AR reads in ARSI-naïve ($n = 98$) vs. enzalutamide-exposed patients ($n = 29$). In stripplots, each dot corresponds to a single tumor. Bar plots show q values from Benjamini-Hochberg adjustment of p values from two-sided Mann-Whitney U tests. Dashed lines indicate $q = 0.05$. Boxplots: center line: median; box limits: upper and lower quartiles; whiskers extend at most 1.5x interquartile range past upper and lower quartiles.



Extended Data Fig. 4 | Marker gene expression in NK and T cells. Expression of (a) cell type markers, (b) dysfunction and activation markers, (c) markers of tumor-reactive cytotoxic cells³¹, (d) genes expressed in a *GNLY*-expressing cytotoxic subset, and (e) genes reported to mark a progenitor population necessary for response after anti-PD-1 therapy in melanoma³². Cells are projected onto UMAP space as in Fig. 4a. f, Scatterplots showing pairwise co-expression of *HAVCR2*, *SLAMF6*, and *TCF7* in CD8⁺ T cells. Points are colored according to cluster membership as in Fig. 4a.



Extended Data Fig. 5 | Germline heritability for prostate cancer is enriched near genes specifically expressed in prostate cancer cells. **a**, Expression of a subset of a published set of prostate cancer risk genes that show cell type specificity (Methods)³⁸ **(b)** LDSC-SEG enrichment of heritability for prostate cancer near genes specifically expressed in each cell type (compared to cell types in other cell type groups) (Methods). *: Benjamini-Hochberg FDR < 0.05.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

FACSDiva 8.0.2

Data analysis

General statistical computing & visualization:

SciPy v1.3.2
Python 3.7
Matplotlib v3.1.1
seaborn v0.9.0
R v3.5.1

Whole exome analyses:

BWA v0.5.9
Getz Lab CGA WES Characterization pipeline (https://portal.firecloud.org/#methods/getzlab/CGA_WES_Characterization_Pipeline_v0.1_Dec2018/2), which includes ContEst, MuTect v1.1.6, Strelka v1.0.11, DeTiN v1.8.5, Orientation Bias Filter v17, MAFPoNFilter, and Oncotator v1.9.9.0
DeepVariant v0.8.0
vcf2maf v1.6.17
FACETS v0.5.14

RNA analyses:

cutadapt v2.2
STAR v2.7.2b
Salmon v0.14.1
STAR-Fusion v1.7.0
inferCNV v0.99.7
VISION v2.0.0

SCENIC v1.1.2.2
TraCeR v0.6.0
MiXCR v3.0.12

Other:
Seurat v3.1.0
lifelines v0.23.9
LDSC-SEG v1.0.1

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

scRNA-seq expression and clustering data generated in this study are available at https://singlecell.broadinstitute.org/single_cell/study/SCP1244/transcriptional-mediators-of-treatment-resistance-in-lethal-prostate-cancer. Raw sequence data generated in this study are being deposited in dbGaP (accession phs001988.v1.p1). FASTQs for healthy prostate RNA-seq are from ArrayExpress accession E-MTAB-1733. TCGA data are available at dbGaP accession phs000178.v11.p8. Abida et al. data are available at dbGaP accession phs000915.v2.p2. The Hallmark gene sets from the Molecular Signatures Database (MSigDB) can be accessed at <https://www.gsea-msigdb.org/gsea/msigdb/collections.jsp#H>.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	2,170 post-QC cells from 14 patients and 15 biopsies. No sample-size calculation was performed a priori, as this is a descriptive study. We accrued biopsies from all available patients who consented to the study, and patient numbers are consistent with cancer scRNA-seq studies in literature.
Data exclusions	To arrive at 2,170 post-QC cells, sequenced cells were excluded from downstream analyses based on gene count (<500 or >10,000), read depth (< 50,000 reads), and expression patterns suggestive of ambient RNA contamination (see Methods for details). QC metric thresholds were not fixed before data analysis, but the QC procedure is consistent with current practices in field.
Replication	scRNA-seq data to replicate our findings are not directly available. While the comparisons are limited due to the mismatch in data type, we attempted to identify patterns of expression differences identified in our scRNA-seq data in bulk RNA-seq cohorts. Additionally, we used cell lines, patient biopsy tissues, and organoid models to test expression programs identified from sequencing analyses. VCaP-16 and VCaP-D TGF- β stimulation was tested at multiple concentrations and showed a clear dose-response relationship in both. pSMAD2 IHC was performed in pairs of pre/post-enzalutamide tissues from two separate patients. For each biopsy, slides were cut from multiple tissue blocks (n = 2-6, depending on available tissue) and evaluated for tumor content, with representative images shown from regions with sufficient tumor cellularity. HOXB5, HOXB6, NR1D2 immunohistochemistry was performed twice independently on the same batch of fresh sectioned slides and at least three images were taken from each in low and high magnification.
Randomization	This study did not involve allocation of patients/samples to interventions/experimental perturbations, so randomization did not apply.
Blinding	This study did not include experimental perturbations of biopsies. Comparisons of nuclear pSMAD2 staining in Fig. 2f were performed by a pathologist who was blinded to the timepoint label for each biopsy. There was no additional blinding, as this study did not involve prospective allocation of patients into arms/groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

Methods

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

anti-human PTPRC (CD45) monoclonal antibody conjugated to FITC (VWR #ABNOMAB12230)
 anti-human EPCAM antibody conjugated to PE (Miltenyi Biotec #130-091-253)
 anti-phospho-SMAD2 (Ser465/467) (138D4) antibody (Cell Signaling Technologies #3108T)
 anti-SMAD2/3 antibody (Cell Signaling Technologies #3102)
 anti-vinculin antibody (7F9) (Santa Cruz Biotechnology, #sc-73614)
 anti- β -actin HRP-conjugate antibody (Abcam #ab20272)
 anti-PSA antibody (rabbit polyclonal, Biodesign, Kennebunk, ME, #K9211DR)
 anti-phospho-SMAD2 (Ser465/467) antibody (E8F3R, rabbit monoclonal, Cell Signaling Technology #18338)
 anti-NR1D2 (NOVUS #NBP2-56141)
 anti-HOXB5 (Sigma #HPA043851)
 anti-HOXB6 (Abcam #ab219499)

Validation

For FACS, anti-human PTPRC (CD45) and anti-human EPCAM antibodies were validated by the manufacturers as specific for human and for use in flow cytometry.
 Per manufacturer, anti-phospho-SMAD2 (Ser465/467) (138D4) detects endogenous levels of Smad2 only when dually phosphorylated at serines 465 and 467, and may detect Smad3 phosphorylated at the equivalent sites. This antibody does not cross-react with other Smad-related proteins. Species reactivity: human, mouse, rat, and mink. Validated for use in Western blot.
 Per manufacturer, anti-SMAD2/3 antibody is recommended for use in Western blot and immunoprecipitation, with reactivity to human, mouse, rat, and monkey total Smad2/3 protein.
 Per manufacturer, anti-vinculin antibody is recommended for use in Western blot, immunofluorescence, immunoprecipitation, and immunohistochemistry, with reactivity to vinculin of human, mouse, rat and avian origin.
 Per manufacturer, anti- β -actin antibody is validated for use in Western blot and has specificity to mouse, rat, rabbit, chicken, cow, dog, human, pig, *Drosophila melanogaster*, African green monkey, and Chinese hamster.
 anti-PSA antibody: specificity and sensitivity is supported by repeated results showing the detection of a protein band at the expected size. Protein detection was increased by androgen stimulation in prostate cancer cells and reduced by androgen receptor downregulation with siRNA or degrader or AR antagonist (e.g. enzalutamide).
 anti-phospho-SMAD2 (Ser465/467) antibody (E8F3R): per manufacturer, antibody has specificity to human, mouse, and rat and is suitable for use in Western blot, immunoprecipitation, immunocytochemistry, flow cytometry, and ChIP. We performed titration experiments using 22Rv1 cells with and without TGF- β stimulation to determine antibody concentration.
 anti-NR1D2, anti-HOXB5, anti-HOXB6: Antibodies for immunohistochemistry experiments were validated by the Human Protein Atlas (HPA) project according to manufacturer's website. We confirmed by serial dilution to determine specificity and expression location using manufacturer datasheet as reference. In addition, these antibodies are specific to human antigens but not mouse. We validated in patient derived xenografts showing no expression in mouse tissues.

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

VCaP-16 cells were derived from long term culture of VCaP cells in enzalutamide (see Methods). VCaP cells were purchased from ATCC.

Authentication

VCaP cells were authenticated via DNA typing.

Mycoplasma contamination

VCaP cells tested negative for Mycoplasma via a DNA-based PCR test.

Commonly misidentified lines
(See [ICLAC](#) register)

None

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

Patients were men who had metastatic castration resistant prostate cancer. No genomic features were used for participant selection. 13/14 patients had prostate adenocarcinoma, and 1/14 had small cell carcinoma. Biopsies were collected from metastases to bone (8/18), lymph node (9/18), and liver (1/18). All patients had received androgen deprivation therapy. Patients ranged in age from 57-79.

Recruitment

Patients were recruited as part of the research program at Dana-Farber Cancer Institute. The consenting patient population at this site may not necessarily be representative (e.g. potentially more urban) of all advanced prostate cancer patients; however, the genomic and clinical characteristics of our cohort are consistent with those of participants from large-scale bulk sequencing studies of mCRPC conducted in the US.

Ethics oversight

This study was approved by the Dana-Farber/Harvard Cancer Center Institutional Review Board under protocols # 09-171, 11-104, 13-301, and 01-045.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Tumor samples were collected and transported in Dulbecco's Modified Eagle Medium, on ice. Single-cell suspensions for single-cell RNA-seq were obtained from tumor core needle biopsies through mechanical and enzymatic dissociation. Samples were first cut into pieces smaller than 1 mm³ using a scalpel. For bone biopsies, soft tissue was also scraped from the hard bone surface using a scalpel blade. Samples were then dissociated using one of two protocols, chiefly to optimize for yield of viable cells from different metastatic sites (Methods). Single cell suspensions in PBS with 2% FBS were stained by incubating for 15 minutes at room temperature protected from light with anti-human PTPRC (CD45) monoclonal antibody conjugated to FITC (1:200 dilution), anti-human EPCAM antibody conjugated to PE (1:50 dilution), and either Calcein-AM (1:200 dilution), 7-Aminoactinomycin D (7-AAD) (1:200 dilution), or both.

Instrument

BD Biosciences FACSAria cell sorter (Ilu or UV)

Software

FACSDiva

Cell population abundance

Post-sort fractions included 45.5% viable cells (low 7AAD), 0.1% immune cells (high CD45-FITC), and 0.3% epithelial/tumor cells (high EPCAM-PE) as demonstrated by sample 09171136.

Gating strategy

We first sorted cells with biological dimensions (high FSC-A and high SSC-A), selected single cells, and excluded doublets or triplets (low FSC-W). Next, we selected live cells (low 7AAD/ high Calcein-AM) that were CD45+ (high FITC, enriching for immune cells), EPCAM+ (high PE, enriching for tumour cells), or double negative (low FITC/low PE).

- Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.