

RESEARCH

Open Access



# Meta-analyses of mouse and human prostate single-cell transcriptomes reveal widespread epithelial plasticity in tissue regression, regeneration, and cancer

Luis Aparicio<sup>1,2,3,8†</sup>, Laura Crowley<sup>2,4,5,6,8†</sup>, John R. Christin<sup>2,4,5,6,8</sup>, Caroline J. Laplaca<sup>2,4,5,6,8</sup>, Hanina Hibshoosh<sup>7,8</sup>, Raul Rabadan<sup>1,2,3,8\*</sup> and Michael M. Shen<sup>2,4,5,6,8\*</sup>

## Abstract

**Background** Despite extensive analysis, the dynamic changes in prostate epithelial cell states during tissue homeostasis as well as tumor initiation and progression have been poorly characterized. However, recent advances in single-cell RNA-sequencing (scRNA-seq) technology have greatly facilitated studies of cell states and plasticity in tissue maintenance and cancer, including in the prostate.

**Methods** We have performed meta-analyses of new and previously published scRNA-seq datasets for mouse and human prostate tissues to identify and compare cell populations across datasets in a uniform manner. Using random matrix theory to denoise datasets, we have established reference cell type classifications for the normal mouse and human prostate and have used optimal transport to compare the cross-species transcriptomic similarities of epithelial cell populations. In addition, we have integrated analyses of single-cell transcriptomic states with copy number variants to elucidate transcriptional programs in epithelial cells during human prostate cancer progression.

**Results** Our analyses demonstrate transcriptomic similarities between epithelial cell states in the normal prostate, in the regressed prostate after androgen-deprivation, and in primary prostate tumors. During regression in the mouse prostate, all epithelial cells shift their expression profiles toward a proximal periurethral (PrU) state, demonstrating an androgen-dependent plasticity that is restored to normal during androgen restoration and gland regeneration. In the human prostate, we find substantial rewiring of transcriptional programs across epithelial cell types in benign prostate hyperplasia and treatment-naïve prostate cancer. Notably, we detect copy number variants predominantly within luminal acinar cells in prostate tumors, suggesting a bias in their cell type of origin, as well as a larger field of transcriptomic alterations in non-tumor cells. Finally, we observe that luminal acinar tumor cells in treatment-naïve prostate cancer display heterogeneous androgen receptor (AR) signaling activity, including a split between AR-positive and AR-low profiles with similarity to PrU-like states.

<sup>†</sup>Luis Aparicio and Laura Crowley contributed equally to this work.

\*Correspondence:

Raul Rabadan  
rr2579@cumc.columbia.edu

Michael M. Shen  
mshen@columbia.edu

Full list of author information is available at the end of the article



**Conclusions** Taken together, our analyses of cellular heterogeneity and plasticity provide important translational insights into the origin and treatment response of prostate cancer. In particular, the identification of AR-low tumor populations suggests that castration-resistance and predisposition to neuroendocrine differentiation may be pre-existing properties in treatment-naïve primary tumors that are selected for by androgen-deprivation therapies.

**Keywords** Prostate cancer, scRNA-seq, Tumor heterogeneity, Plasticity, Castration, Androgen receptor, Field cancerization

## Background

Despite decades of investigation, regression and regeneration of the prostate gland as well as its oncogenic transformation represent fundamental biological processes that are poorly understood. In particular, androgen signaling represents a key regulatory program that maintains the identity of prostate tissue, yet the roles for androgen regulation in specific cell types remain unclear. In this regard, the advent of scRNA-seq technology has provided new tools to investigate the dynamics of prostate cell identity at the molecular level in both homeostasis and disease.

Although the prostate surrounds the urethra directly underneath the bladder, there are substantial anatomic differences between mammalian species. The mouse prostate is comprised of four distinct lobes, corresponding to the ventral (VP), lateral (LP), dorsal (DP), and anterior prostate (AP) lobes, whereas the human prostate lacks distinct lobular organization but can be subdivided into central, transition, and peripheral zones [1]. These prominent anatomic differences have led in part to long-standing questions about the relationship of cell types and molecular pathways between the mouse and human prostate.

Classically, histological and ultrastructural analyses have described three major epithelial cell types in the prostate: luminal cells, basal cells, and rare neuroendocrine cells [2, 3], with less well-defined stromal cell types. However, recent scRNA-seq analyses have revealed considerable cellular heterogeneity and novel cell types in the mouse prostate epithelium [4–8] and stroma [9, 10]. Although these studies independently reported multiple cellular populations with similar features, there are notable discrepancies in their nomenclature and description [3], perhaps due to methodological differences in sample collection, preparation, computational analyses, and/or annotations. Similar issues also apply for scRNA-seq analyses focused on normal human prostate [4–6, 11], as well as in the context of pan-tissue resources [12, 13]. As a consequence, published scRNA-seq analyses of the mouse and human prostate are not readily comparable, and the precise relationships between cell populations described in different studies are unclear.

To address these issues, we have performed a meta-analysis of independent scRNA-seq datasets from the mouse prostate, aggregating datasets from two different mouse strains published by seven different laboratories and using two distinct bioinformatic approaches for their analysis to generate a comprehensive reference atlas. We have included new datasets to supplement rare cell populations, including a dataset of the proximal prostate (closest to the urethra) to examine the periurethral (PrU) cells residing in this region, and report gene signatures for each well-documented population during homeostasis. We have also analyzed time courses of prostate regression and regeneration, which demonstrate that each epithelial cell type displays similar transcriptomic shifts toward a PrU-like state following castration and returns to normal when androgen is reintroduced, revealing substantial androgen-dependent plasticity.

Similarly, we have performed a meta-analysis of the normal human prostate [4, 5, 7, 13] to generate a consensus atlas of human prostate cell types during homeostasis. Since these studies have used different naming schemes and definitions for cell types, we have generated a nomenclature comparison and proposed a common descriptive naming scheme. In addition, we have compared the transcriptomic profiles of normal human and mouse epithelial cell types and show that PrU cells in the human prostate have transcriptomic profiles consistent with reduced androgen sensitivity.

Finally, we have investigated changes in cell states that occur during progression to prostate adenocarcinoma. We have analyzed the dynamic changes in profiles of each cell type in homeostasis, hyperplasia [7], and adenocarcinoma [6, 14–17]. Notably, we have found that luminal acinar (LumAcinar) cells display the greatest transcriptomic changes during progression to adenocarcinoma. Moreover, we found that copy number variants (CNVs) are only present in LumAcinar and rare neuroendocrine (NE) tumor cells, suggesting a predominant cell of origin for prostate cancer (PCa). However, we find that many LumAcinar cells lacking CNVs as well as other epithelial cell types also display extensive transcriptomic alterations, which is suggestive of a field effect similar to those observed in other tumor types. Most interestingly, we observe that tumor cells found in some treatment-naïve

adenocarcinomas display a transcriptomic shift toward a PrU-like or LumDuctal state that displays decreased AR signaling activity. Taken together, our single-cell analyses demonstrate the cross-species conservation of prostate cell types and underscore the significance of cellular plasticity following androgen deprivation as well as oncogenic transformation.

## Methods

### Mouse prostate tissue

Mouse strains and genotyping, isolation of mouse prostate tissue, dissociation of mouse prostate tissue, and prostate single-cell RNA-sequencing were carried out as previously described [4]. Mice were maintained under specific-pathogen free (SPF) conditions in accordance with USPHS, USDA, and AAALAC requirements. Euthanasia was performed by carbon dioxide inhalation followed by cervical dislocation, as described by the AVMA guidelines for euthanasia. All animal studies were approved by and conducted according to standards set by the Columbia University Irving Medical Center (CUIMC) Institutional Animal Care and Use Committee (IACUC) under protocol AABT5655.

Tissue was isolated from wild type C57BL/6 (C57BL/6NTac, 8–10 weeks old) mice to generate the two new datasets described in this study [18]. For the ventral prostate (VP) lobe dataset (ML003/GSM7024431), the entire extent of the VP lobes was dissected from one male mouse at 8 weeks of age, from the distal tips to the proximal end within the rhabdosphincter. For the proximal and periurethral prostate dataset (ML008/GSM7024432), a proximally enriched region was dissected from 3 male mice, 10 weeks of age. The rhabdosphincters were removed, and prostate tissue was collected from the periurethral junction with the urethra on one end (including minimal surrounding urethra), to 1–2 mm beyond the proximal:distal boundary on the

other end (to include some distal cells). Additionally, a tiny region of proximal seminal vesicle (SV) was dissected from 2 mice to include in the sample after removal of secretions.

### Human prostate tissue

Human prostate tissue specimens were obtained from patients undergoing radical prostatectomy at Columbia University Irving Medical Center. Patients gave informed consent under an Institutional Review Board-approved protocol (AAAN8850). The clinical characteristics of these patients are provided in Table 1. Processing and analysis of tissue was performed as previously described [4].

### Electron microscopy

Prostate tissue was taken from a C57BL/6 J mouse at 8 weeks of age. An approximately 2 mm region of the AP lobe within the rhabdosphincter near the periurethral-proximal boundary was micro-dissected. The sample was fixed, processed, sectioned, and imaged as previously described [4].

### Immunofluorescence imaging

Paraffin embedding, sectioning, immunofluorescence staining, and imaging of tissue sections were performed as previously described [4] on prostate tissue sections from C57BL/6 J mice, 8–10 weeks old. Antibodies used for immunofluorescence staining were anti-mouse/human Krt4 monoclonal antibody (1:50  $\mu$ L dilution, Invitrogen catalog # MA1-35,558, lot TB2524522, clone 6B10), Krt5 polyclonal antibody (1:1000  $\mu$ L dilution, BioLegend cat. 905,901, lot B271562), Krt14 polyclonal antibody (1:500  $\mu$ L dilution, BioLegend cat. 905,301, clone Poly19053, lot B308016), Krt8/18 monoclonal antibody (also called Troma-1, 1:100  $\mu$ L dilution, Developmental Studies Hybridoma Bank, antibody registry ID AB

**Table 1** Human prostate samples and corresponding clinical data

Sample ID	Diagnosis	Gleason grade	Gleason score (highest)	Stage	Treatment	Age	Race	Ethnicity	PSA at sample acquisition
LHu1	Adenocarcinoma	GG3	4 + 3 and 3 + 3	pT2N0	None	51–60: 20% 61–70: 80%	White: 40% Combination not described: 60%	Non-Hispanic: 60% Hispanic: 40%	PSA < 4: 0 patients PSA < 10: 2 patients PSA > 10: 3 patients
LHu2	Adenocarcinoma	GG1	3 + 3	pT2N0	None				
LHu3	Adenocarcinoma (metastatic)	GG5	4 + 5	ypT3N1	Degarelix				
LHu4	Adenocarcinoma	GG2	3 + 4	pT3N0	None				
LHu5	Adenocarcinoma	GG4	4 + 4	pT2N0	Tamsulosin				

531826), MSMB polyclonal antibody (1:100  $\mu$ L dilution, Abclonal, cat. A10092, lot 0204440101), and Chga polyclonal antibody (1:200  $\mu$ L dilution, Abcam, cat. ab15160). Sectioned tissues underwent standard antigen retrieval (citrate-based antigen unmasking solution, Vector Labs, H-3300–250) for all antibodies except anti-Chga, which required high pH antigen retrieval (tris-based antigen unmasking solution, Vector Labs, H-3301–250).

### Datasets analyzed in this study

We collected and analyzed published as well as new scRNA-seq datasets for normal mouse prostate, mouse prostate during tissue regression and regeneration, normal and benign human prostate, and human primary prostate tumors. These datasets and their analysis are described below. Please note that no original code was generated in this study for these analyses.

For normal mouse prostate, we used two separate pipelines in parallel for our analysis: Seurat and Randomly. The Seurat pipeline is summarized in the following section. The Randomly pipeline is similar to what was previously published [4], with detailed description in the following sections. For the analyses used to generate aggregated datasets, we used both pipelines to analyze the whole prostate as well as anterior lobe, dorsal lobe, and lateral lobe datasets (GSM4556594, GSM4556596, GSM4556597, GSM4556599) from Crowley et al. [4, 19], the whole prostate dataset (GSM4338122) from Joseph et al. [7, 20], the whole prostate dataset (GSM4594201, GL64) from Mevel et al. [8, 21], the T00\_intact\_1 anterior prostate dataset (GSM4474186) from Karthaus et al. [6, 22], the anterior prostate lobe (OEX003110) from Guo et al. [5, 23], and the proximal-enriched whole prostate (GSM7024432) from this study [18]. There were minor differences between the analyses as the Randomly pipeline also incorporated the ventral prostate lobe dataset (GSM4556598) from Crowley et al. [4, 19], whereas the Seurat pipeline also utilized the T00\_intact\_2 anterior prostate dataset (GSM4474187) from Karthaus et al. [6, 22], the ventral as well as dorsal and lateral lobe datasets (OEX003110) from Guo et al. [5, 23], and the ventral lobe dataset (GSM7024431) from this study [18]. We also used the Randomly pipeline to perform analyses of the adult mouse urethra dataset (GSM4338169) from Joseph et al. [7, 20], as well as the anterior lobe regression and regeneration datasets (GSM4474191 through GSM4474210) from Karthaus et al. [6, 22].

For analyses of human prostate, we exclusively utilized the Randomly pipeline to analyze the following datasets: normal human prostate from two organ donors (TS\_Prostate.h5ad) from Tabula Sapiens [13, 24], human prostate transition zone from an organ donor (GSM3293878,

GSM4337424) from Henry et al. [11, 25], benign prostate tissue from prostatectomies (GSM4556601) from Crowley et al. [4, 19], benign prostate tissue from prostatectomies (OEP000825) from Guo et al. [5, 23], and tissue from three patients with benign prostatic hyperplasia (GSM4337069 through GSM4337071) from Joseph et al. [9, 26]. For analyses of prostate cancer, we examined datasets from patients HP98, HP99, HP100, and HP103 (DUOS-000115) from Karthaus et al. [6, 27], from patients 1 through 9, 11, 12, and 13 (GSM4203181) from Chen et al. [17, 28], from patients 3, 6, 9, 10, 11, 12, 15, and 18 (GSM5494350, GSM5494356, GSM5494360, GSM5494362, GSM5494363, GSM5494365, GSM5494368, and GSM5494373) from Hirz et al. [15, 29], from patients 1 through 11 (GSM5353214 through GSM5353248) from Song et al. [14, 30], and from patients 1 through 6 and 8 through 13 (HRA000823) from Ge et al. [16, 31].

### Seurat scRNA-seq analysis pipeline

FASTQ files for all datasets used in the Seurat pipeline were either already in house or downloaded from the Short Read Archive (SRA) or the National Omics Data Encyclopedia (NODE) (Additional file 2: Table S1). FASTQ files were then aligned and quantified using Cell Ranger v7.0.0. All scRNA-seq counts were corrected for ambient RNA using the SoupX package. The cleaned counts were then converted into Seurat objects using the Seurat package. Cells with high mitochondrial DNA content, low gene detection, and/or high RNA counts were filtered out to enrich for live single cells. Datasets were normalized, and their variances were stabilized with SCTransform. Cell doublets were then computationally detected and filtered out using DoubletFinder. All individual datasets were then merged into a new Seurat object, their original counts normalized, and their variance stabilized using SCTransform.

To produce an integrated dataset, integration anchors were calculated, and the datasets were then integrated using the reciprocal principal component analysis (RPCA) reduction in Seurat. A new PCA reduction and UMAP reduction were then generated using the integrated dataset. For first-pass cluster calling, neighbors and clusters were determined using Seurat at a resolution of 0.8 and a Louvain algorithm with multilevel refinement. These clusters were then minimally manually adjusted to reflect physical anatomy and marker expression previously validated by immunofluorescence staining. Gene set enrichment analysis was performed using the escape package with the “UCell” method and Hallmark mouse gene sets provided by MSigDB. Visualizations of the gene set enrichment analysis were performed using the dittoSeq package.

Mouse prostate population signatures were generated in Seurat dataset with the `wilcoxauc` function of the `presto` package on the aggregated dataset. Signature genes with the most globally distinguished expression patterns were determined by applying a filter to collect only genes with an  $AUC \geq 0.75$  and an adjusted  $p$  value  $\leq 0.05$ . This was performed for each distinct cell type/cluster, as well as for informative subgroups (such as all distal luminal cells versus proximal luminal cells).

### Randomly scRNA-seq analysis pipeline

Randomly analyses were conducted as previously described [4]. Sequencing data were aligned and quantified using the Cell Ranger Single-Cell Software Suite (v.2.1.1) with either the GRCm38 mouse or the GRCh38 human reference genomes. There are 4 major steps: (1) filtering the raw sequencing data expression matrix, (2) correcting for batch effects using Seurat and processing with Randomly (<http://52.201.223.58:1234/>) [32], (3) clustering data using the Leiden algorithm (<https://scanpy.readthedocs.io/en/stable>), and (4) dimensional reduction for visualizations such as t-SNE and UMAP plots (included in Randomly package). Departures from our previous methods will be summarized below.

### Filtering the expression matrix

Cell-gene matrices were pre-processed by filtering cells with less than 500 genes detected. We also removed cells whose proportion of transcripts derived from mitochondrially encoded genes was greater than 10%. The expression matrices were normalized by  $\log_2(1 + TPM)$ , where  $TPM$  is transcripts per million.

### Random matrix theory application to denoise scRNA-seq

Random matrix theory (RMT) was first introduced by Wishart in 1928, but the mathematical foundations of RMT were developed by the theoretical physicist Dyson in the 1960s when he was describing heavy atomic nuclei energy levels. A key feature of RMT is universality, namely the insensitivity of certain statistical properties to variations of the probability distribution used to generate the random matrix. This property provides a unified and universal way to analyze single-cell data [32] and we

previously used this method to describe new cell populations in prostate [4].

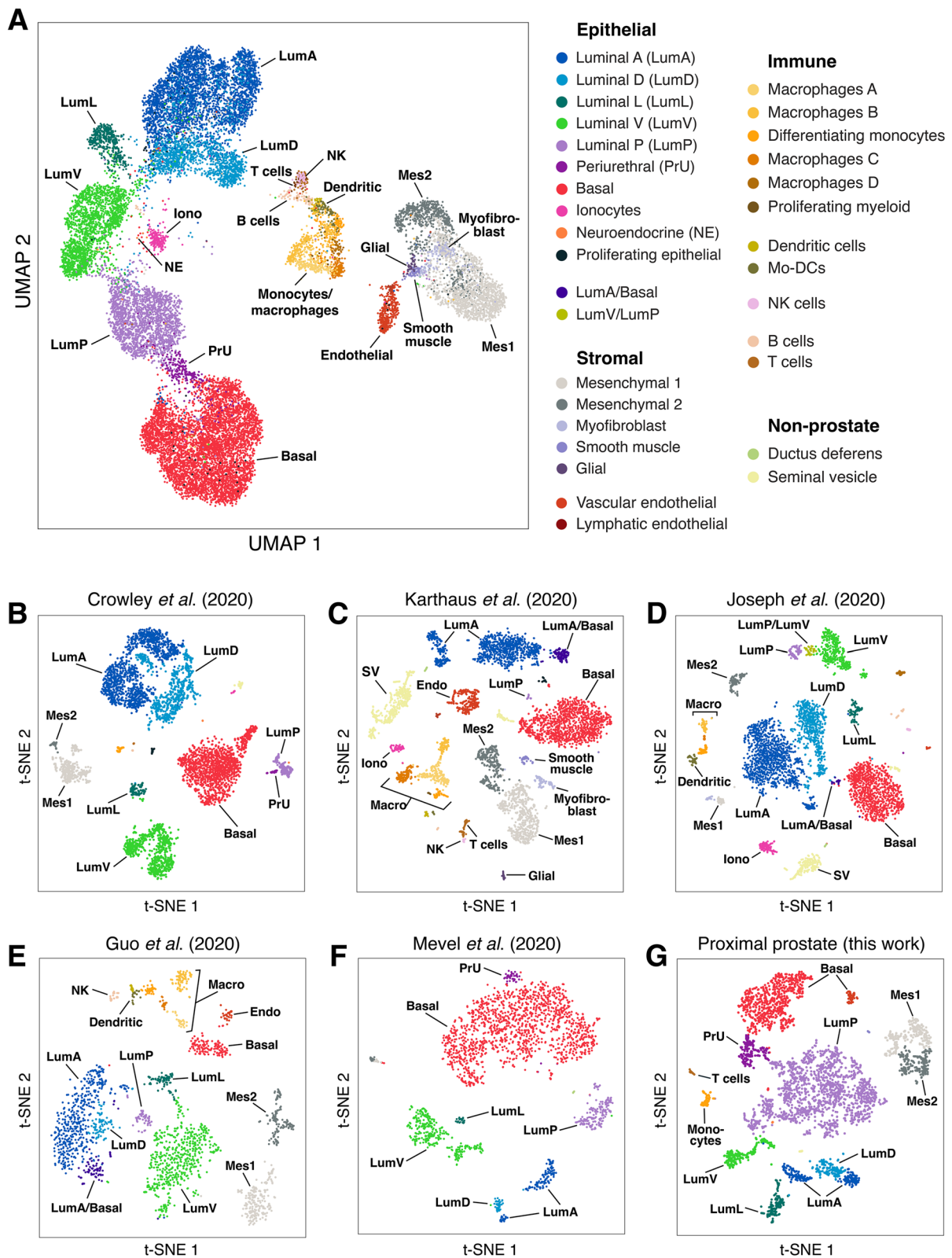
The RMT strategy relies on the fact that single-cell datasets show a threefold structure: a random matrix, a sparsity-induced signal, and a biological signal. Indeed, 95% or more of the single-cell expression matrix is compatible with being a random matrix [32]. This could be understood as if the dataset is showing cells whose expression is randomly sampled from a given distribution in approximately 95% of the matrix inputs. In single-cell datasets, sparsity is also a key feature, as it can generate a fake signal that after removal increases the quality and performance of clustering in prostate scRNA-seq analyses, and led to identification of the novel PrU population [4]. From an operative point of view, the presence of localized eigenvectors related with sparsity implies the existence of an undesired (fake) signal.

### Clustering

Clustering was performed using the Leiden algorithm, as implemented in [33, 34]. This clustering algorithm leverages on the latent space generated by the RMT algorithm. Based on RMT, Randomly determines the dimensions of the latent space and projects the data into this space using the distribution of eigenvalues (Marchenko-Pastur and Tracy-Widom distributions) and eigenvectors (Porter-Thomas distribution). This projection is subsequently clustered in groups using the Leiden algorithm. The determination of the optimal number of clusters relied on the mean silhouette score. Specifically, we conducted a series of clustering analyses across various Leiden resolutions (the clustering parameter) and calculated the mean silhouette score for each scenario. We established a relationship between the mean silhouette score, acting as a function of the Leiden resolution, and the respective number of clusters for each case (see Fig. 1—figure supplement 2 in [4]). We selected the absolute maximum of this curve and took the corresponding number of clusters. In certain instances, sub-clustering specific clusters proved beneficial. The process involved repeating the described procedure for a designated cluster. Sub-clustering was particularly valuable for unraveling immune populations or differentiating between vas deferens and seminal vesicle populations. The robustness

(See figure on next page.)

**Fig. 1** Reference plots of mouse prostate scRNA-seq data demonstrate extensive cell type heterogeneity. **A** Aggregated composite UMAP plot of all mouse prostate cell types. **B–G** t-SNE plots of the individual contributing datasets in full. **B** C57BL/6 whole prostate [4]. **C** FVB anterior prostate lobe [6]. **D** C57BL/6 whole prostate [7]. **E** C57BL/6 whole prostate [5]. **F** C57BL/6 whole prostate [8]. **G** C57BL/6 proximal prostate (this work). Datasets were processed using the *Randomly* pipeline, which revealed 12 epithelial populations, 7 stromal populations, and 11 immune populations found across multiple datasets. Non-prostatic populations as well as populations that may correspond to cell states are only shown for the individual datasets and have been removed from **A**



**Fig. 1** (See legend on previous page.)

of sub-clustering was verified through supervised plotting of known genes associated with the aforementioned populations.

#### Batch effect correction and data integration

Datasets for scRNA-seq samples corresponding to mice and humans were aggregated using the BBKNN method [34] with default parameters. Data for luminal acinar cells from normal human prostate, BPH, and prostate tumors were integrated using SCANORAMA [35], using default parameters for batch correction.

#### Differential expression analysis

The genes highlighted and presented in the dot-plots were chosen using a strategy based on differential expression. These selected genes underwent a *t*-test (one group vs. all others) with a corrected *p* value (Benjamini–Hochberg correction) below 0.01. Additionally, as a secondary threshold for selection, these genes were required to display expression in a minimum of 60% of cells within the target population and in less than 25% of cells for all other populations.

Gene signatures for each human epithelial population were generated using a similar approach. All of the most differentially expressed genes for each population were selected that had corrected *p* values of  $\leq 0.05$ , with the secondary requirement of expression in a minimum of 60% of the cells within the target population and a maximum of 25% of the cells for other populations. The hyperplasia group for human prostate corresponds to cells from samples with hyperplasia (BPH327PrGF\_Via, BPH340PrGF\_Via, and BPH342PrF\_Via from Table S1 of [7]).

#### Representations and visualizations

To visualize the single-cell clusters, we performed dimensional reduction to two dimensions through *t*-distributed Stochastic Neighbor Embedding (*t*-SNE) and Uniform Manifold Approximation and Projection (UMAP) representations. Default parameters were utilized for both techniques: a learning rate of 1000, perplexity of 30, and early exaggeration of 12 for *t*-SNE; for UMAP, we set the number of neighbors to 15 and minimum distance to 0.3. Visualizations using *t*-SNE, such as dot-plots or ridge-plots, were carried out using the visualization functions from the Randomly public package in [32], and the visualization functions within SCANPY in [33, 34].

Two-dimensional visualization of human prostate epithelial cells was performed using PHATE [36], depicting the luminal acinar cells from 2 normal prostates [13], 3 prostates with BPH [7], and 46 prostates with PCa from Memorial Sloan Kettering Cancer Center [6], University of California, San Francisco [14], Massachusetts General

Hospital [15], Peking University Third Hospital [16], and Shanghai Changhai Hospital [17], using default parameters. The data were integrated using SCANORAMA [35].

#### Pseudotime analysis

To infer the potential developmental trajectory and cellular fates of aggregated luminal acinar cells from human prostates, we employed PALANTIR [37] for the detection of single-cell trajectories in pseudotime. To perform trajectory analysis using Palantir, we used normal human prostate LumAcinar cells from the Tabula Sapiens dataset as the root for the analysis. We randomly selected individual normal LumAcinar cells and obtained similar results in several replicate analyses, indicating that the inferred pseudotime trajectories were robust.

#### Cell type score

For analyses of mouse prostate regression and regeneration, we constructed a “cell type score” to quantify the changes in transcriptomic profile for each cell type, based on the genes that are most specific and differentially expressed among the basal, LumA, LumP, Mes1, Mes2, myofibroblast, smooth muscle, and vascular endothelial populations. The cell type score was generated by assessing the mean expression of a specific set of differentially expressed genes that effectively characterize each population. The chosen genes for the cell type score underwent a *t*-test (one group vs. all others) with a corrected *p* value (Benjamini–Hochberg correction) less than 0.01. Additionally, these genes were required to be expressed in at least 60% of cells within the target population and in fewer than 25% of cells for all other populations. To construct the cell type score, the mean expression of differentially expressed genes for each population was compared at each time point during regression. This was followed by division by the mean expression of the genes in the normal tissue before castration, and the resulting values were normalized to a scale of 1–100.

#### Identification of tumor cells

We applied InferCNV [38] to the scRNA-seq datasets to discern malignant epithelial cells exhibiting genomic instability. Epithelial cells classified as non-malignant based on copy number alterations (CNV) via inferCNV may represent authentic benign cells or transformed cells lacking identifiable CNVs through scRNA-seq inference. The analytical approach involved initial examination of each patient sample independently, employing denoising and clustering transcriptomic analyses as detailed above, to identify cell populations akin to those in the human consensus atlas. Subsequently, inferCNV was executed for each patient sample within the same cohort to pinpoint cell populations with CNVs.

To serve as control populations in inferCNV, various non-epithelial cell types were employed, along with external controls sourced from Tabula Sapiens [13]. For the Karthaus cohort [6], where only epithelial cells were identified, Tabula Sapiens external controls were exclusively utilized as a reference population. In contrast, for the remaining cohorts we used mesenchymal populations detected in these samples and populations from Tabula Sapiens as controls.

For each tumor, the CNV matrix obtained from inferCNV was presented in a heatmap, displaying previously identified populations with copy number alterations, using the standard/default scale, and applying the option of hierarchical clustering to visualize the heatmap. Further insights into copy number differences were derived through dimensional reduction of the CNV matrix analysis, employing PCA with its first 20 principal components. This dissimilarity revealed that clusters lacking clear CNVs, which include internal controls such as mesenchymal cells, tended to aggregate together, whereas clusters with distinct CNVs formed separate groupings.

Epithelial populations devoid of CNVs but observed in cancer patients were labeled as “abnormal,” based on changes in their transcriptomic profiles. This categorization applied to LumAcinar cells lacking CNVs as well as non-acinar epithelial populations. We note that the approach utilized for identifying tumor cells via inferCNV has intrinsic limitations, given its basis on inference from scRNA-seq data, and that many alterations such as mutations that might be present in tumor cells would not be captured by the techniques analyzed in this study.

#### Assessing population similarity using optimal transport theory

We employed Optimal Transport [39, 40] to evaluate the transcriptomic similarity between cell types, as done previously [4]. The Wasserstein-1 distance serves as a metric for phenotypic distance among cell populations, defined as a distance function between probability distributions in a measurable metric space. Conceptually, the Wasserstein-1 distance aligns with the earth mover’s distance, wherein probability distributions are envisioned as piles of dirt, and the cost of transforming one pile into another corresponds to the Wasserstein distance. We employed this approach to compare normal tissue with each time point during mouse prostate regression and regeneration, using the aggregated mouse prostate dataset as a reference.

A similar methodology was applied to assess the similarity between normal mouse and human prostate epithelial populations. In this case, optimal transport and

Wasserstein distance were utilized to compare the aggregated mouse dataset with Tabula Sapiens. Initially, we identified orthologous gene pairs, separately normalized mouse and human datasets using  $\log_2(1 + TPM)$ , filtered out genes with an average expression less than 0.1 for human or mouse, and merged the corresponding mouse and human datasets. Employing RMT to eliminate sparsity-induced signals, we selected genes with biological signals in the shared mouse/human dataset. Subsequently, we calculated the Wasserstein distance in the common space between mouse and human, visualizing these distances through a set of nested heatmaps. We also used Wasserstein distance to compare the phenotype of cancer cell states with normal tissue cell types.

## Results

### Epithelial populations of the mouse prostate

To generate a comprehensive aggregated single-cell dataset for the mouse prostate, we gathered publicly available scRNA-seq datasets generated from C57BL/6 and FVB mice and generated new data focusing on the proximal and periurethral regions of the prostate, which have been less studied (Additional file 2: Table S1). We analyzed these datasets using two independent computational approaches to confirm the reproducibility of our interpretations. In the first approach, we de-noised each dataset using random matrix theory (RMT), which improves the ability to separate and detect rare cell populations [32]. We then sequentially clustered in each dataset to identify cell populations, following the same strategy used previously [4]. This analysis of 11 datasets resulted in an aggregated dataset of 21,952 cells arranged in 30 prostate cell clusters (Fig. 1A, Additional file 1: Fig. S1A, Additional file 1: Fig. S2). As a second approach, we used a standard Seurat pipeline to generate an aggregated dataset from 13 datasets of sufficiently high quality, which was composed of 30,433 cells in 18 distinct clusters (Additional file 1: Fig. S3A).

These parallel approaches allowed for the identification and comparison of cell populations across datasets in a uniform manner, independent of differences in reporting and labeling between publications. To define robust cell populations, we required that the population be identified in at least three independent datasets and have nearly complete overlap in globally distinguishing gene expression. For rare cell populations, we only required that the population be present in at least two independent datasets. Of note, the majority of clusters were identical using both the RMT and Seurat approaches. The RMT approach handled sparse data differently, yielding a greater number of small clusters and providing better discrimination between populations with low cell numbers.



We found that epithelial populations were remarkably consistent across datasets and approaches. Interestingly, no distinct subclusters were formed based on mouse strain background, which did not significantly contribute to prostate epithelial heterogeneity. In particular, basal cells formed a single contiguous cluster in individual datasets (Fig. 1B–F; Additional file 1: Fig. S3B–F), as previously reported [4–8], and in our aggregated datasets (Fig. 1A, Additional file 1: Fig. S3A). We did not observe evidence of a distinct basal subcluster with expression of *Zeb1* or other epithelial-mesenchymal transition (EMT) markers [41]. However, in the Seurat pipeline, we observed that a small subset of basal cells adjoins the periurethral (PrU) cluster, is proximally enriched, and expresses slightly more proximal and luminal markers (Additional file 1: Fig. S3A, G).

We identified multiple luminal epithelial clusters, which represent distinct cell types that are separated by prostate region (Fig. 1A–F, Additional file 1: Fig. S3A–F), as previously reported for individual datasets [4–8]. Since the nomenclature for these populations differs between laboratories (summarized in [3]), we follow a descriptive naming system [4] that denotes lobe-specific prostate populations (e.g., LumA for the distal anterior lobe) as well as proximal populations (LumP for proximal prostate). Notably, although the dorsal and lateral lobes have often been combined as a “dorsolateral lobe,” highly distinct dorsal (LumD) and lateral (LumL) populations were always found in each individual dataset as well as the aggregated datasets (Fig. 1A, B, D–F, Additional file 1: Fig. S3A, B, D–F). In contrast, the anterior (LumA) and dorsal (LumD) distal luminal populations consistently displayed the most transcriptomic overlap (Fig. 1A, B, D–F, Additional file 1: Fig. S3A, B, D–F).

Unlike distal luminal cells, which differ by lobe, proximal luminal cells (LumP) formed a single cluster without lobe-specific identity (Fig. 1A–F, Additional file 1: Fig. S3A–F) [4–8]. The vast majority of LumP cells are located in the proximal region of the prostate, though rare distal cells can be observed [4, 5, 8], and functional heterogeneity within the population has been reported [5]. In this regard, in the Seurat pipeline, we observed that a subset of LumP cells is adjacent to distal luminal cells (Additional file 1: Fig. S3A).

Neuroendocrine (NE) cells represent a rare and historically elusive epithelial population that could be detected in both analytical pipelines (Fig. 1, Additional file 1: Fig. S3). Ionocytes are another rare population that was recently described in the prostate [6], and our meta-analysis revealed their presence in additional datasets (Fig. 1A, B, D, Additional file 1: Fig. S3A, B, D) [4, 7]. Though ionocytes have some transcriptional similarities to NE cells, they express *Foxi1* and *Atp6v1g3* but not

specific luminal or basal markers (Additional file 1: Fig. S2I, J). Both cell types were observed in higher proportions in the proximal dataset, and the PrU population is described in detail below.

Using the aggregated datasets, we generated reference gene expression signatures that are specific for each prostate epithelial cell type (Additional file 3: Table S2). In addition, we examined the Gene Set Enrichment Analysis Hallmark signatures and found increased expression of genes involved in protein secretion in seminal vesicle and distal luminal cells, and the lowest levels of Notch signaling genes in NE cells (Additional file 1: Fig. S4). Finally, we observed rare epithelial clusters in individual datasets that may represent cell states. In particular, a subset of LumA cells expresses both LumA and basal markers and may correspond to “intermediate” cells with hybrid luminal and basal features (Additional file 1: Fig. S1B, C).

### Non-epithelial cell populations

Our scRNA-seq meta-analysis also provided consistent insights into non-epithelial cell types in the mouse prostate. The mesenchymal/stromal cells present in these datasets are predominantly fibroblasts and can be divided into several different clusters (Fig. 1, Additional file 1: Fig. S2E, F, Additional file 1: Fig. S3). The Mesenchyme 1 (*Mes1*) population is proximally enriched, lies adjacent to the epithelium, and expresses *Srd5a2* as well as many Wnts and other signaling factors, whereas Mesenchyme 2 (*Mes2*) is enriched more distally, is located slightly farther from the epithelium, and expresses many chemokines and complement components [4, 10]. We also identified distinct myofibroblast and smooth muscle populations that express smooth muscle actin (*Acta2*), and observed that a subset of myofibroblasts expresses *Lgr5* [4, 6, 10, 42]. Although a third fibroblast population has been reported [10], it did not appear as a distinct cell type in our analyses, but rather as a subset of *Mes2* (Additional file 1: Fig. S2E, F). Interestingly, several mesenchymal cell types reported to exist in the prostate (e.g., telocytes) were not detected in any dataset, suggesting that the prostate stromal compartment is incompletely captured in existing scRNA-seq data.

Hematopoietic lineage populations (such as B and T lymphocytes, dendritic cells, and NK cells) were also detected across multiple datasets, with the immune compartment displaying a notable myeloid bias. In particular, macrophages divided into distinct subclusters along a continuous spectrum, which was most evident in the RMT pipeline. Since profiles for M1 and M2 macrophages could not be definitively identified, we have named these populations alphabetically (Fig. 1, Additional file 1: Fig. S2A–D). In addition to the macrophage populations, we detected a population with substantial

overlap in gene expression to macrophages, which appeared to correspond to differentiating monocytes (Additional file 1: Fig. S2A–D).

Finally, we also observed contaminating seminal vesicle cells across multiple datasets. Seminal vesicle epithelial cells could be clustered into a single basal population as well as luminal populations with more proximal markers or more distal markers (Additional file 1: Fig. S2G, H), suggesting potential epithelial heterogeneity within this tissue.

### The periurethral region

We define the periurethral (PrU) region as the most proximal extent of each prostate lobe nearest the junction with the urethra. PrU cells make up most of the epithelium in this region. Because this region is located exclusively within the rhabdosphincter and hence is more difficult to dissect, many prostate scRNA-seq samples have not captured the epithelial populations in this region. However, our meta-analysis detected PrU epithelial cells in several datasets [4, 7] as well as many in our proximal prostate scRNA-seq dataset (Fig. 1G, Additional file 1: Fig. S3G). Uniquely, PrU epithelial cells display hybrid luminal and basal features, similar to urothelial cells in the adjacent urethra [4]. However, PrU cells can be readily distinguished from urethral cells by lineage-tracing with an *Nkx3.1-Cre* driver [4].

To understand the unique morphological features of PrU cells, we imaged the periurethral region by electron microscopy and immunofluorescence staining (Fig. 2). At the ultrastructural level, PrU cells share some features with distal luminal (LumDist) cells, such as organelles involved in protein secretion, and many features with LumP cells, including a high density of mitochondria (Fig. 2A–F). Interestingly, several features of PrU cells also resemble urothelial cells of the urethra, including the nuclear orientation of more basally situated PrU cells, as well as the lumen-facing structures of apically situated cells, which may resemble the rigid, uroplakin-filled surface of urothelial cells. Thus, PrU cells share ultrastructural features of both the prostate and the urethral urothelium and may represent a physical transition between the two tissues.

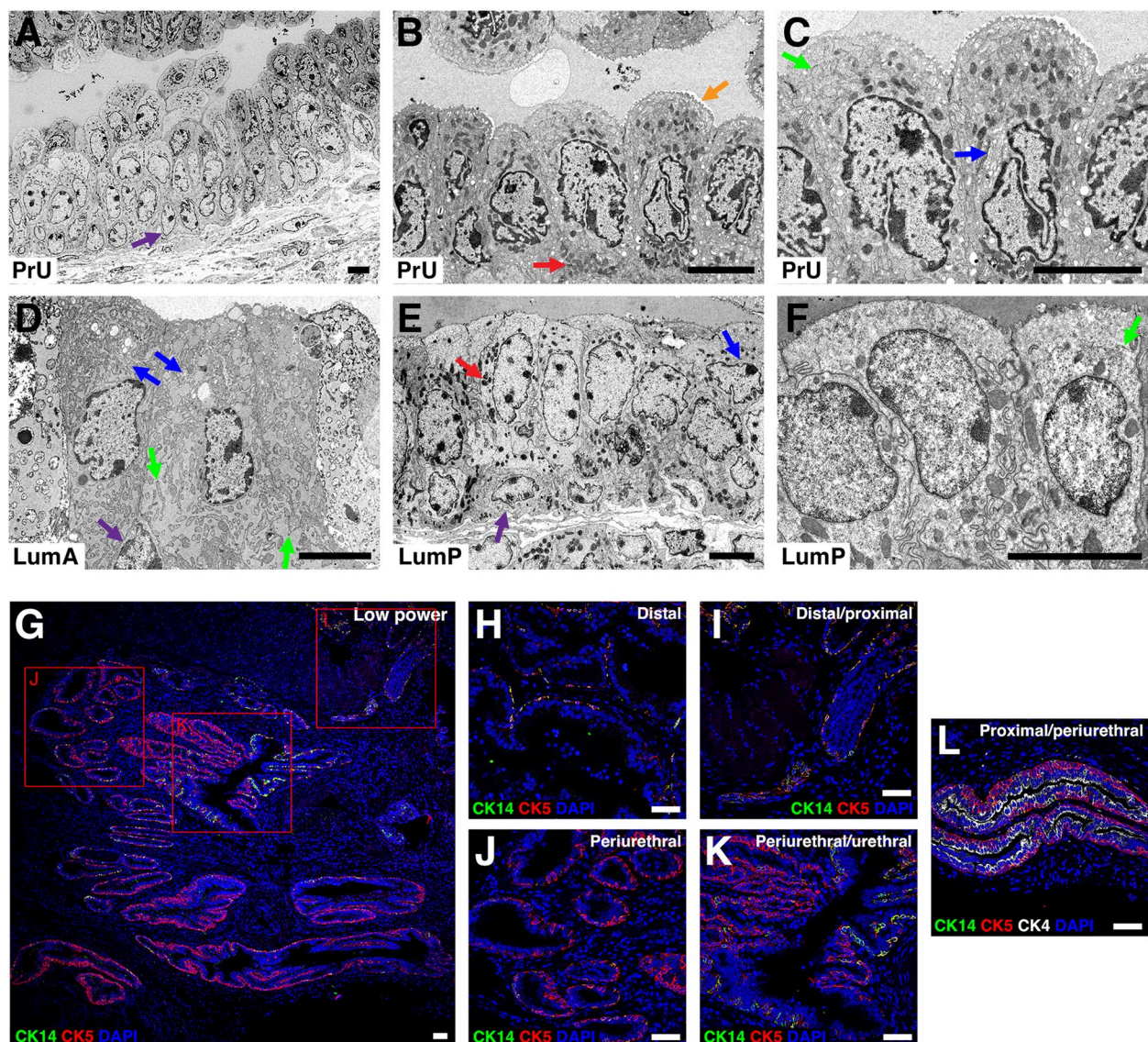
At the level of gene expression, PrU cells uniquely express *Lmo1*, *Anxa8*, *Dapl1*, and *Aqp3* and have higher *Ly6d* and *Sca-1* expression than LumP cells [4] (Additional file 1: Fig. S5B). Although *Krt5* and *Krt14* expression overlaps in the basal layer throughout more distal regions of the prostate, *Krt14* expression becomes intermittent in the PrU region and *Krt5* is maintained, whereas basal cells of the urothelium rarely express *Krt14* (Fig. 2G–L). Based on our re-analysis of a scRNA-seq dataset of the proximal prostate and urethra [7], we

could define two distinct urethral populations, a luminal-intermediate urothelial cell group (which we term Urethral 2) with transcriptomic similarity with LumP cells and a basal-intermediate urothelial cell group (Urethral 1) with similarity with PrU cells (Additional file 1: Fig. S5). Notably, at homeostasis, PrU and LumP cells can be readily distinguished from urothelial cells by key markers, including several uroplakins (Additional file 1: Fig. S5). Thus, PrU cells also represent a transition population in terms of molecular features, such as gene expression.

### The transcriptomic response to androgen deprivation and restoration

The prostate regresses in response to androgen-deprivation and regenerates after androgen restoration, which can be repeated through at least 30 cycles in the mouse [43, 44]. Following castration, the prostate undergoes rapid shrinkage and involution resulting in a stable regressed state, whereas restoration of androgen levels results in prostate regrowth to its former size [45]. To examine the response of individual cell populations to androgen-deprivation and restoration, we examined scRNA-seq data of mouse prostate through time courses of regression and regeneration [5, 6]. For this analysis, we defined a “cell type score” to represent the average of the most specific and differentially expressed genes for each cell type (“Methods”). In response to castration, every cell type except endothelial cells showed a significant decrease in its cell type score (Fig. 3A, B). Interestingly, the rates of transcriptomic change were different for each population, as distal luminal (LumDist) cells, myofibroblasts, and *Mes1* cells rapidly lost almost all of their cell-type specific gene expression, whereas LumP, basal, smooth muscle, and *Mes2* cells only lost approximately half of their cell-type specific gene expression, with *Mes2* cells retaining their gene expression profile the longest (Fig. 3A, B).

Interestingly, our analysis indicated that mouse prostate epithelial cells shift toward a PrU-like expression profile during regression. A detailed examination of gene expression patterns in LumA, basal, and LumP populations showed that each population lost expression of many specific genes but retained its distinctive expression of select distal luminal, basal, or proximal luminal markers during the regression-regeneration cycle (Fig. 3C–E). However, each epithelial population gained expression of multiple PrU markers following castration and lost this expression after androgen restoration; furthermore, the markers retained by LumP cells during regression were those that are co-expressed by PrU cells. The epithelial populations did not shift toward urethral gene expression profiles, as only rare LumP cells expressed any urothelial markers (Additional file 1: Fig.

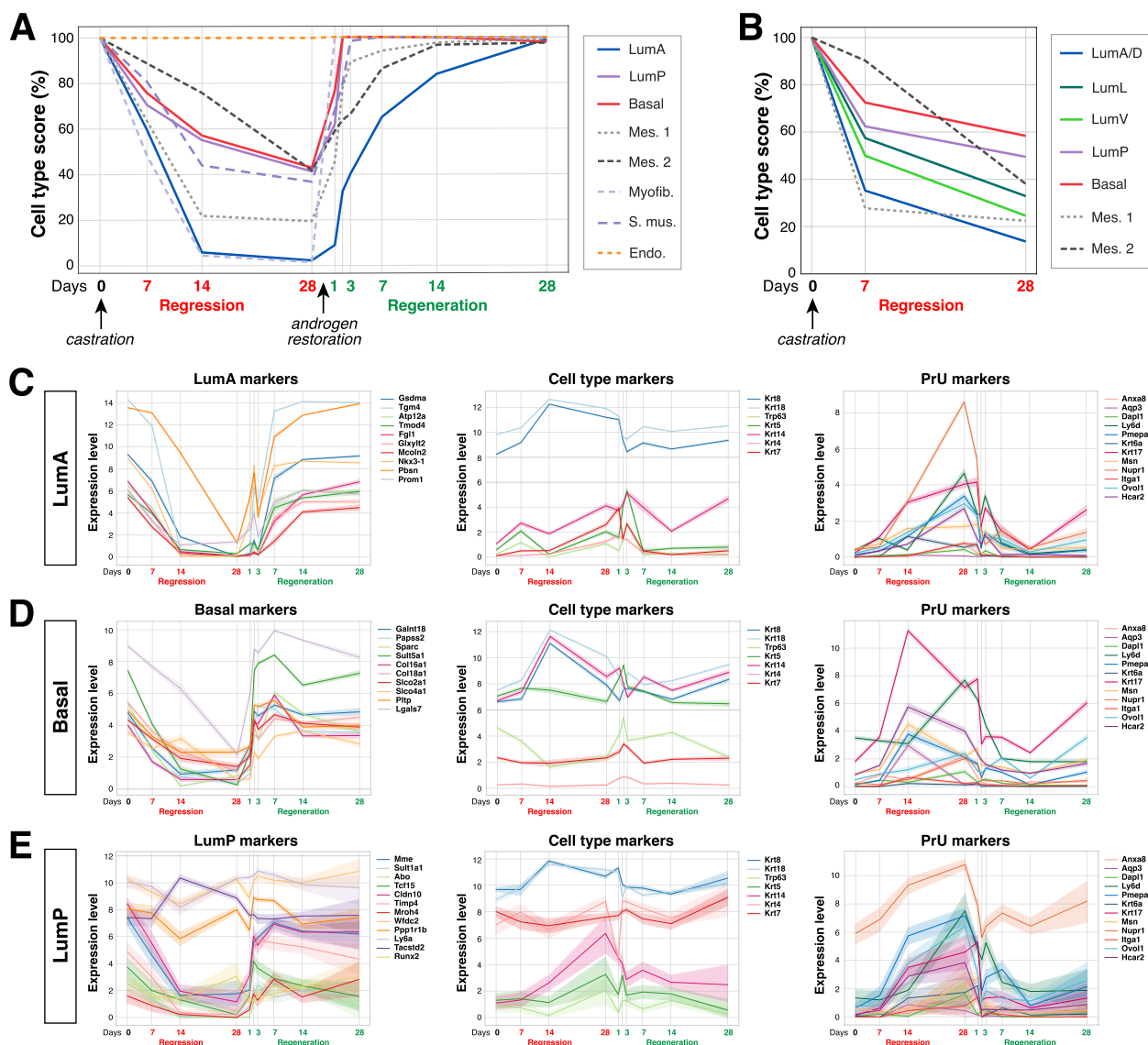


**Fig. 2** Imaging of mouse PrU cells reveals unique and shared features with prostatic and urethral cells. Scanning electron microscopy (EM) images of PrU cells show a focal region of cells where they appear to be multilayered (A), a region that is not multilayered and displays unique features (B), and a higher magnification of this region (C). The features of distal LumA cells (D), proximal LumP and basal cells (E), and LumP cells at higher magnification (F) are shown for comparison. Arrows indicate basal nuclear orientation (purple), mitochondrial density (red), apical membrane structures (orange), rough endoplasmic reticulum (green), and Golgi apparatus (blue). Scale bars in A–F indicate 5 μm. G–L Immunofluorescence staining show changes in basal and proximal keratin expression. G Overview of the periurethral region with neighboring urethral and proximal cells at low power. Insets show co-expression of basal keratins CK5 (red) and CK14 (green) in distal (H) and proximal (I) basal cells, and consistent CK5 but reduced CK14 in periurethral (J) and periurethral and urethral (K) basal cells. Proximal keratin CK4 (white) is maintained through the proximal and periurethral region (L). No superficial-like cells were observed in the periurethral region. Scale bars in G–L indicate 50 μm

S6, Additional file 1: Fig. S7). Notably, while the normal PrU profile includes some genes that are co-expressed by either LumP or the urethral urothelium, the regressed epithelium expresses many PrU-specific genes that are distinct from both (Additional file 1: Fig. S6E). These findings highlight PrU-like transcriptomic profiles and provide a broader context for the previously reported

shift from LumA toward LumP in the anterior prostate following androgen deprivation [6].

A transcriptomic shift was also observed in the prostate stroma during regression, as both the Mes1 and Mes2 fibroblast populations altered gene expression in response to androgen deprivation. Mes1 cells rapidly shifted toward a Mes2 expression profile and lost



**Fig. 3** Time course of prostate regression and regeneration reveals androgen-dependent plasticity. **A, B** Meta-analyses of single-cell RNA-seq datasets for prostate regression and regeneration. **A** As described in [6], for regression time points, wild-type mice were castrated and prostate tissues from 2 biological replicates were collected at 1 day, 7 days, 14 days, and 28 days after castration; for regeneration time points, mice that had been castrated for at least 4 weeks were subcutaneously implanted with dihydrotestosterone (DHT) pellets, with 2 biological replicates collected at 1, 2, 3, 7, 14, and 28 days after pellet implantation. **B** As described in [5], prostate tissues were collected from wild-type mice at 7 and 28 days after castration. “Cell type score” is defined as the percentage of most specific differentially expressed genes for each population, averaged over the whole population (“Methods”). Changes in gene expression that are enriched in urethral but not PrU cells, such as *Areg* and *Ociad2*, in the LumA (**C**), basal (**D**), and LumP (**E**) populations, showing distinguishing genes for each population (left column), genes for general compartmental markers, and genes that are enriched for PrU and not co-expressed in LumP (right column), where the line indicates the average expression for each gene across the population and the bar indicates confidence interval ( $\pm 95\%$ )

expression of several defining factors including Wnts, whereas Mes2 cells changed gene expression more slowly (Fig. 3A, B, Additional file 1: Fig. S6, Additional file 1: Fig. S7). Thus, we conclude that transcriptomic reprogramming following androgen deprivation is not

exclusive to the luminal or distal compartments, but instead represents a tissue-wide alteration of cell states.

### Atlas of the human prostate

Next, we performed a meta-analysis of published scRNA-seq datasets to establish a corresponding reference atlas of the normal human prostate [4, 5, 11, 13] using the criteria described for the RMT pipeline (“Methods”). Despite differences in the relative proportions of cell populations between these datasets, the data were remarkably consistent. We found that the human prostate has a single basal epithelial population, two luminal populations corresponding to luminal acinar (LumAcinar) and luminal ductal (LumDuctal), and a periurethral-like (PrU) population (Fig. 4A–F, Additional file 1: Fig. S8A). The stromal populations were more variable and less well-represented across datasets, but corresponded to at least 1 endothelial population and 3 fibroblast-like populations (Fig. 4A, Additional file 1: Fig. S8A, B). Of the 3 fibroblast-like populations, the first expressed several classic fibroblast markers and did not subdivide readily (we denote these as general fibroblasts), the second corresponded to fibroblasts that express several muscle genes (myofibroblasts), and the third to fibroblast-like cells that express many contractile muscle genes (fibromyocytes) (Fig. 4F) [46]. Based on differential gene expression, we generated signatures for each epithelial and mesenchymal population (Additional file 4: Table S3). Within the immune compartment, we detected relatively fewer cells with variable representation of cell types between patients, so these populations were grouped as either myeloid or lymphoid. Interestingly, the zone of the prostate tissue did not have a clear effect on the transcriptome (Additional file 1: Fig. S8B, C), as previously reported [5].

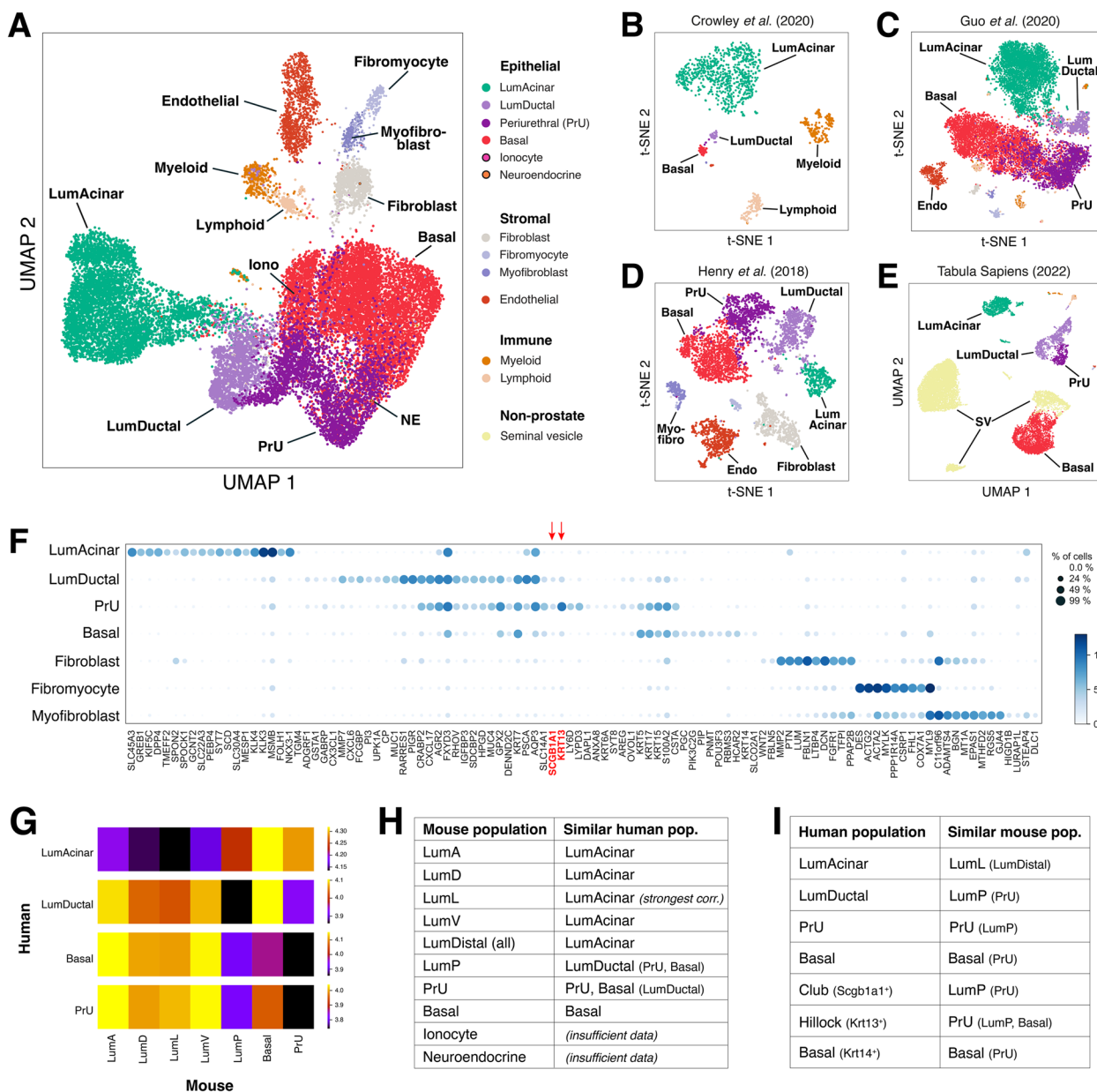
Since the nomenclature of human prostate epithelial populations differs between publications, we compared our previous nomenclature [4] to an alternative system that uses “Club” and “Hillock” lung terminology [11], using the Tabula Sapiens as a source of normal tissue (Fig. 4E, Additional file 1: Fig. S8D, E). Notably, we found that most of the “Club” cells corresponded to LumDuctal and PrU cells (Additional file 1: Fig. S8B, C), as they expressed common luminal genes and more specific markers like *RARRES1*, but did not consistently express the defining marker *SCGB1A1* (Fig. 4F). Similarly, most “Hillock” cells corresponded to PrU cells (Additional file 1: Fig. S8B, C), as they expressed common luminal and basal genes as well as more specific markers such as *KRT7*, *PSCA*, *RARRES1*, *LYPD3*, and *AQP3*; moreover, expression of the Club- and Hillock-defining markers were not specific (Fig. 4). The remaining luminal cells corresponded to LumAcinar cells (Additional file 1: Fig. S8D, E), expressing common luminal cytokeratins as well as more specific markers including *KLK3*, *MSMB*, *FOLH1*, and *TGM4* (Fig. 4F). These transcriptional similarities were separately confirmed by plotting the expression of

each of these genes on the prostate single-nuclei RNA-seq data from the GTEx project portal [12]. Based on these analyses, we find that our descriptive nomenclature of human prostate epithelial populations correlates with lung terminology but appear to align more accurately with distinct cell types in the prostate.

To perform an updated cross-species comparison of cell type identities [4], we calculated the Wasserstein distance between gene expression profiles for each population in the aggregated mouse and human datasets in transcriptomic latent space (“Methods”) (Fig. 4G). While human and mouse basal cells have notably different profiles, human basal and PrU populations most closely resemble mouse PrU, human LumDuctal most closely resembles mouse LumP, and human LumAcinar most closely resembles mouse LumL followed by LumD (Additional file 1: Fig. S8D). To test the robustness of this analysis, we individually removed the top 20 differentially expressed genes from the mouse LumL expression profile and repeated the comparisons, which revealed that the greater similarity of human LumAcinar to LumL was primarily dependent on differential expression of a single gene, *MsmB*; otherwise, the transcriptomes of the different LumDist populations had similar marker overlap with human LumAcinar. Consequently, we suggest that human LumAcinar cells, which are distributed throughout different zones of the human prostate, may correspond more generally to mouse LumDist populations of all lobes. We additionally plotted the signatures of each population on the aggregated data of the other species to see how the differentiating genes versus the whole transcriptome compare across species (Fig. 4H, I). Together, these results suggest a clear correlation across species.

### Distinguishing human prostate cancer progression by AR signaling levels

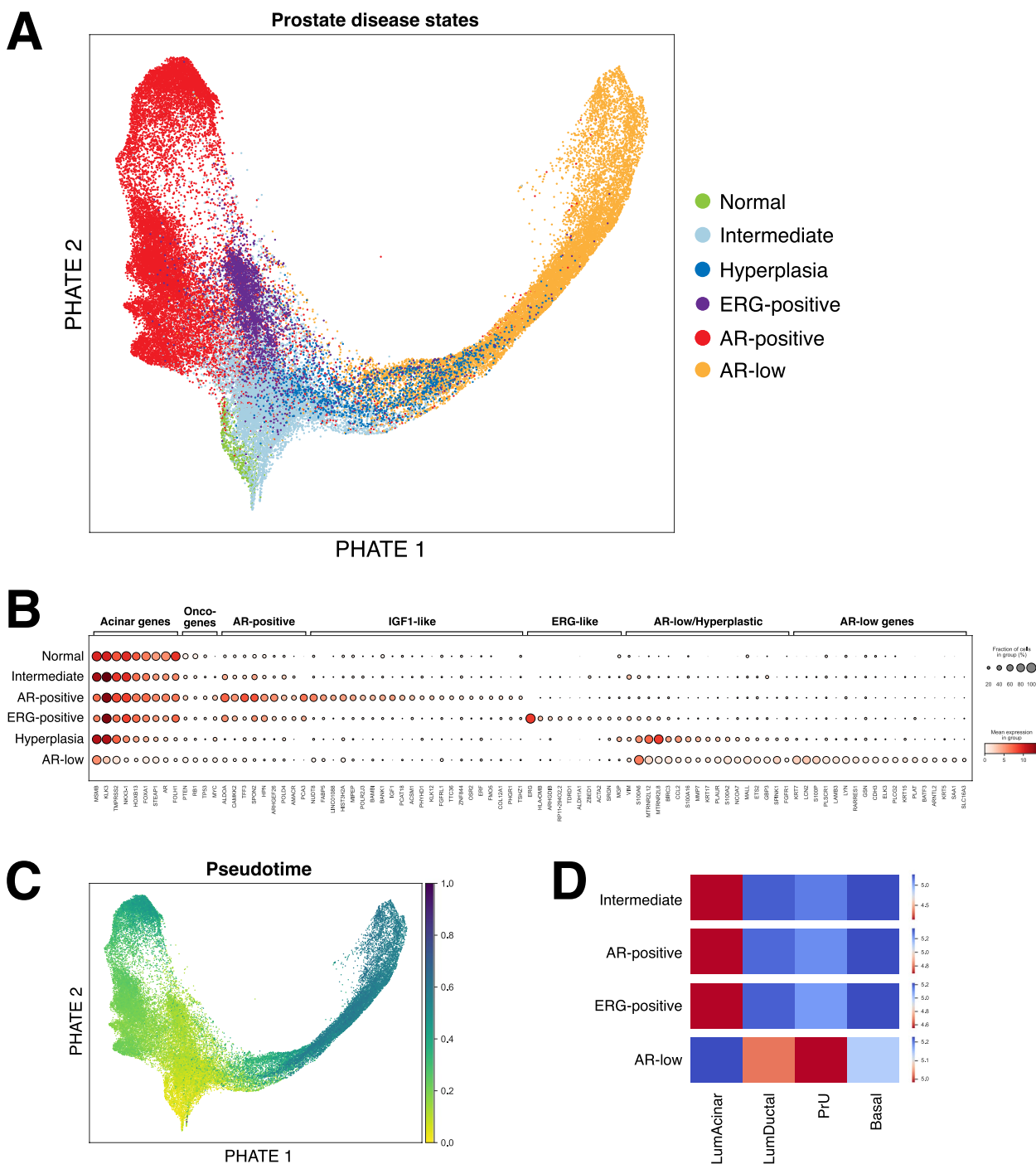
To examine alterations of the human prostate due to disease, we combined the normal prostate scRNA-seq datasets with those from patients with benign prostate hyperplasia (BPH) [7] and treatment-naïve prostate cancer [6, 14–17] using the RMT pipeline; we did not pursue a parallel Seurat pipeline analysis due to lack of availability of FASTQ files for several of these datasets. In these aggregated data of 99,611 cells from 66 datasets, we observed heterogeneous gene expression profiles across treatment-naïve tumors, which was particularly apparent in LumAcinar cells from prostatectomy samples. Therefore, we performed PHATE visualization of the LumAcinar cells from the aggregated data to depict local and global data structures (Fig. 5A, Additional file 1: Fig. S9A, B). This analysis revealed that LumAcinar cells across early prostate disease stages can be sub-clustered into six primary groups with different gene



**Fig. 4** Reference plots for human prostate scRNA-seq data. **A** Aggregated composite UMAP plot for samples of benign human prostate and adjacent benign prostate. **B–E** Plots of individual datasets. **B** UMAP plot corresponding to primarily LumAcinar cells taken from the peripheral zone of 1 patient [4]. **C** Plot containing primarily basal and PrU cells from 1 patient with PCa [5]. **D** Dataset containing primarily basal, PrU, and LumDuctal cells from 1 patient without PCa [11]. **E** Dataset containing mixture of prostate and seminal vesicle, originating from 2 organ donor patients with no history of prostate disease [13]. **F** Dot plot of select top differentially expressed genes (among genes that are expressed in more than 60% of the population and have the highest mean expression in that population) for the epithelial and stromal cell populations from the reference aggregated normal human prostate. The lung club cell marker SCGB1A1 and hillock cell marker KRT13 are highlighted, indicating that these do not clearly correspond to single, distinct prostate cell types. **G** Heatmap comparing the total gene expression profiles of the cell types in the normal human prostate dataset [13] to those of the aggregated normal mouse prostate, using Wasserstein distance as a metric. Darker color indicates greater transcriptomic similarity. Tables listing the most similar mouse and human epithelial populations based on gene expression, generated by overlaying the mouse cell type signatures onto the human populations (**H**) and vice versa (**I**)

expression profiles (Fig. 5B, Additional file 5: Table S4). Notably, these groups divide into two distinct arms that correlate with androgen receptor (AR) signaling levels

and PrU-like gene expression. Pseudotime analysis suggested a model in which the AR-positive and AR-low arms may both arise through a trajectory from normal



**Fig. 5** Meta-analysis of scRNA-seq datasets from human prostate adenocarcinoma reveals disease evolution of luminal acinar cells. **A** PHATE plot of LumAcinar populations from 2 normal prostates [13], 3 prostates with BPH [7], and 46 prostates with PCa from Memorial Sloan Kettering Cancer Center [6], University of California, San Francisco [14], Massachusetts General Hospital [15], Peking University Third Hospital [16], and Shanghai Changhai Hospital [17]. Clustering of the aggregated data reveals that LumAcinar cells show the greatest variation, as LumAcinar cells from the normal and BPH prostates occupy 1 cluster each (normal and hyperplasia, respectively), while cells in the PCa samples subdivide into 4 major subpopulations (intermediate, ERG-positive, AR-positive, and AR-low). The PHATE plot splits these PCa subpopulations along two major branches. **B** Dot plot of cell type, PCa, subpopulation-defining, and other relevant genes indicates that AR signaling is a major differentiating factor across the two branches. **C** Pseudotime analysis of the cells in **A** suggests a normal or intermediate origin for both branches in PCa. This also suggests progression through hyperplasia to PCa in some cells. **D** Heatmap comparing the total gene expression profiles of the LumAcinar subpopulations in **A** to all normal epithelial populations. The AR-low branch has a notable shift toward PrU and LumDuctal marker expression. Wasserstein distance is used as a metric, and darker red indicates greater transcriptomic similarity

(from healthy prostates) and/or intermediate stages (normal-like gene expression from prostatectomy samples), potentially passing through a hyperplastic expression stage before splitting into distinct arms (Fig. 5C).

We performed several analyses to understand this division of gene expression profiles during prostate cancer progression. First, we compared the profiles of the four LumAcinar groups found in PCa samples to profiles of normal human prostate epithelial populations. We found that the cells of the AR-positive and ERG-positive groups as well as the intermediate group resembled normal LumAcinar cells and retained the expression of many differentiated LumAcinar genes (Fig. 5B, D). In contrast, LumAcinar cells of the AR-low group shifted from LumAcinar toward PrU expression patterns (Fig. 5B, D). Additional genes that are lost or gained during the transition from normal LumAcinar to AR-positive or AR-low tumor cells were also noted (Additional file 1: Fig. S9D), including increased expression of PrU and LumDuctal signature genes in the AR-low population (Additional file 1: Fig. S9E). To confirm this analysis, we mapped a signature of the most differentially expressed genes in the AR-low arm, as well as the Hallmark AR response signature and an independent AR response signature [47] onto the aggregated LumAcinar populations (Additional file 1: Fig. S10). Together, these data indicate that LumAcinar cells in primary treatment-naïve prostate cancer can be divided into two primary groups of gene expression patterns based on AR signaling levels. AR-positive and ERG-positive cells display elevated AR signaling relative to normal LumAcinar cells and are associated with classical PCa features. In contrast, AR-low cells have dramatically reduced AR signaling levels and shift toward PrU and some LumDuctal expression profiles, unlike other transformed groups.

In addition, our analyses identified two rare and distinct subsets of LumAcinar cells that display markers of partial neuroendocrine differentiation (Additional file 1: Fig. S11A). One group expresses genes such as ASCL2 and POU2F3 and is located in the AR-low arm (Additional file 1: Fig. S11B). The other group expresses genes including ONECUT2 and INSM1 and is located predominantly within the ERG-positive subset in the AR-positive arm (Additional file 1: Fig. S11C). Interestingly, these two

groups may represent the early emergence of neuroendocrine transdifferentiation from luminal adenocarcinoma cells, corresponding to the Class 1 and 2 pathways, respectively, which were recently defined in analyses of a model of prostate neuroendocrine differentiation [48].

### CNVs are specific for LumAcinar cells

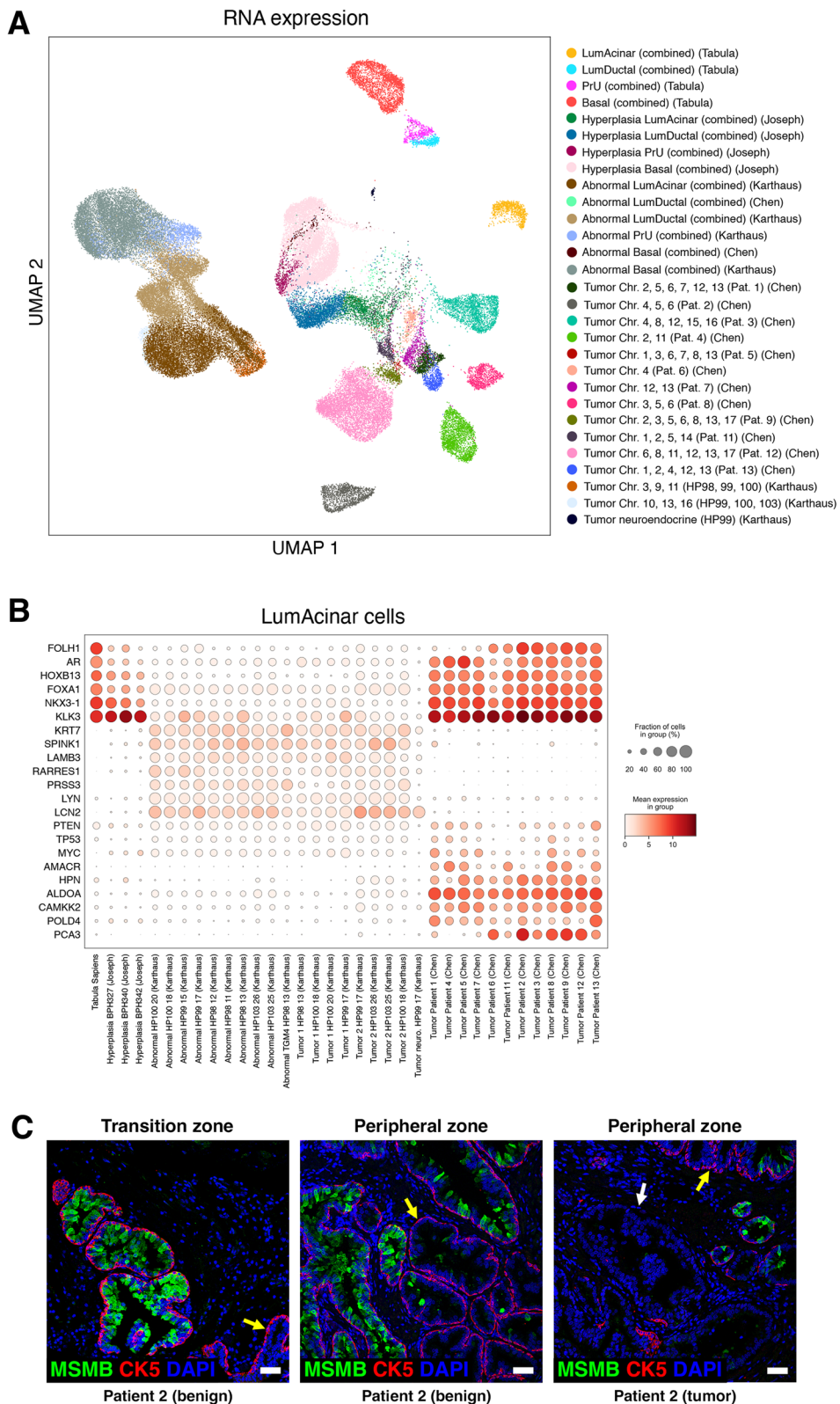
For robust identification of definitive tumor cells in the human prostate cancer scRNA-seq datasets, we used InferCNV to identify copy number variants (CNVs) in the aggregated data (“Methods”). We found that CNVs could only be readily detected and considered to be enriched in a subset of LumAcinar cells from patients with PCa, as well as in a small neuroendocrine (NE) population (Fig. 6A, Additional file 1: Fig. S9B, Additional file 1: Fig. S12, Additional file 1: Figs. S14–S17). Importantly, we could confidently assign LumAcinar identity to the CNV-containing tumor cells despite the transcriptional shifts observed, due to the retained similarity of global transcriptional properties as well as specific genes among these tumor cells and adjacent benign cells (Additional file 1: Fig. S12D).

This analysis also revealed co-occurring CNV profiles in patients from certain datasets; for example, in the Karthaus cohort [6], we named these tumor acinar populations “Tumor 1” (marked by CNVs on chromosomes 3, 9, and 11) and “Tumor 2” (CNVs on chromosomes 10, 13, and 16) to distinguish them from abnormal acinar populations that did not have CNV enrichment (Additional file 1: Fig. S13A–D). Interestingly, the pattern of two groups of common CNVs and expression profiles across the Karthaus cohort were not present in other datasets; for example, the Ge cohort [16] had some overlap of specific inferred CNVs, but not in the overall profiles (Additional file 1: Fig. S16). In comparison, the Chen cohort [17] displayed a more random distribution of CNVs, with no more than 2 CNVs overlapping between any patient tumors (Fig. 6A, Additional file 1: Fig. S15), perhaps consistent with later-stage tumors in these patients. Intriguingly, the patient 1 tumor in the Ge cohort contained multiple different clones, one with AR-positive and the other with AR-low features, indicating that these different clones can coexist (Additional file 1: Fig. S16, Additional file 1: Fig. S17).

(See figure on next page.)

**Fig. 6** Variable CNV and gene expression patterns in LumAcinar cells from human prostate adenocarcinomas. **A** UMAP plot of the epithelial populations from 2 healthy prostates [13], 3 prostates with BPH [7], and 17 prostates with treatment-naïve PCa from Memorial Sloan Kettering Cancer Center [6] and Shanghai Changhai Hospital [17]. **B** Dot plot of key prostate cell type and cancer markers. **C** Immunofluorescence staining of varying levels of MSMB (green) in LumAcinar cells across multiple prostate zones. MSMB is expressed at high levels in normal LumAcinar cells, but is reduced or absent in abnormal acinar areas where basal cells expressing CK5 (red) are intact (yellow arrows), and in tumor-containing areas where basal cells are absent (white arrow)





**Fig. 6** (See legend on previous page.)

Finally, one tumor in the Karthaus cohort contained a small neuroendocrine tumor population with a CNV profile that partially overlapped with one of the acinar tumor populations from the same region of the same tumor (Fig. 6A, B, Additional file 1: Fig. S13E, Additional file 1: Fig. S14). This observation suggests a potential common origin for these two transformed cell types.

#### Transcriptomic changes in LumAcinar cells in proximity to prostate tumors

In addition to different CNV profiles, we observed fundamentally distinct features in the gene expression profiles of tumors in the Chen, Ge, and Karthaus cohorts. The tumors in the Chen cohort retained more classical acinar features, displaying increased expression for many normal LumAcinar, hyperplastic LumAcinar, and AR-responsive genes relative to tumors in the Karthaus cohort. As a result, the expression profiles of tumors in the Chen cohort overlap with those of hyperplastic acinar populations to a greater extent than those in the Karthaus cohort, as shown by signature comparisons, heatmap, and dot plot analyses (Fig. 6A, B, Additional file 1: Fig. S9A, B, Additional file 1: Fig. S14, Additional file 1: Fig. S15). In contrast with the Chen cohort, the Karthaus tumors show significant loss of acinar features across the Tumor 1, Tumor 2, and abnormal acinar populations (as well as non-acinar populations); instead, these transformed and abnormal LumAcinar cells shifted toward PrU and LumDuctal profiles. Intriguingly, the tumor populations in the Karthaus cohort also express lower levels of selected prostate cancer-relevant markers relative to the Chen tumors, including *POLD4*, *AMACR*, and *CAMKK2* (Fig. 6B). In comparison, the Ge cohort displayed a mixture of transcriptomic features, even within the same patient (Additional file 1: Fig. S9A, B, Additional file 1: Fig. S16, Additional file 1: Fig. S17). These findings underscore the extent of transcriptomic variability among the treatment-naïve prostate tumors.

Furthermore, although we only inferred increased CNVs in LumAcinar and neuroendocrine cells in prostate tumor samples (Additional file 1: Figs. S14–S16), we observed that transcriptomic changes in these samples were detectable in all epithelial cell types. Compared to cells in the healthy normal prostate samples, the LumAcinar, LumDuctal, PrU, and basal cells in these samples retained expression of most of the key marker genes, but showed substantial transcriptomic reprogramming, including loss of several components of major developmental signaling pathways (Additional file 1: Fig. S12D–G). Notably, these gene expression changes were not limited to definitive tumor regions or transformed cell types, as we observed that a large number of LumAcinar cells outside of the definitive tumor displayed altered

transcriptomes (Additional file 1: Fig. S12A, C–G). Therefore, we performed immunofluorescence staining to validate these changes in gene expression adjacent to tumor lesions, using MSMB as a representative LumAcinar marker. We observed an apparent shift in LumAcinar gene expression close to the tumor, where the basal cell layer was disrupted, as well as at a distance from the tumor where the basal cell layer is completely intact (Fig. 6C, Additional file 1: Fig. S13G).

Overall, LumAcinar cells showed the greatest amount of reprogramming, including some changes that were specific to BPH and not found in cancer (Additional file 1: Fig. S13A), as well as in cancer (Additional file 1: Fig. S13B–D). Interestingly, one subset of tumor LumAcinar cells had greater transcriptomic similarity to BPH LumAcinar cells (Additional file 1: Fig. S13F), whereas a different set more closely resembled abnormal LumAcinar cells (Additional file 1: Fig. S13B). Notably, these similarities applied even when the LumAcinar cells were analyzed from different patients, perhaps suggesting that the abnormal or BPH LumAcinar cells might occasionally serve as a precursor state for prostate adenocarcinoma, consistent with our pseudotime analysis (Fig. 5C). In addition, we found that this transcriptomic rewiring of expression patterns could extend far from the tumor focus, as suggested by analysis of MSMB expression (Additional file 1: Fig. S13G). Taken together, these findings support the identification of “fields” of transcriptionally altered LumAcinar cells that lack CNVs and are not themselves transformed. Interestingly, this transcriptomic shift was not restricted to the peripheral zone, and transcriptomic shifts were also detected among other epithelial cell types (Additional file 1: Fig. S12D–G).

#### Discussion

Our analyses have revealed several notable features of the normal mouse prostate. First, the remarkable consistency across datasets has allowed us to define 9 mouse prostate epithelial cell types and suggest an additional intermediate cell state. Second, lobe identity along the dorsal–ventral axis corresponds to the identity of a cognate distal luminal population, whereas each lobe is further divided along its proximal–distal axis into three distinct parts, corresponding to the periurethral PrU region, the proximal LumP region, and a lobe-specific distal region. Third, only luminal cell types are reliably distinct in each region along the proximal–distal axis, and thus luminal cells specifically reflect spatial identity. Finally, PrU cells have a hybrid luminal-basal identity, share gene expression with both the prostate and the urethra, and have one PrU subset displaying greater luminal features and the other more basal, which resembles the organization of the urothelium. Consequently, PrU cells have properties

of a transition population at the junction of the urethra and prostate.

In addition, our analyses of prostate regression and regeneration highlight the plasticity of both epithelial and stromal cell types, as nearly all significantly change gene expression profiles in an androgen-dependent manner. In particular, all luminal and basal epithelial populations shift after androgen deprivation toward a transcriptomic profile that resembles a PrU-like state; these transcriptomic shifts have also been investigated by detailed single-cell analyses and lineage-tracing in a recent publication [49]. We speculate that this PrU-like state may mirror that of epithelial cells within the prostatic urogenital sinus and prostate epithelial buds during early organogenesis, when androgen levels are relatively low. At these early stages, both luminal and basal progenitors retain bipotent progenitor properties and may display hybrid luminal-basal features [50, 51]. Notably, PrU cells express the highest levels of *Sca-1*, *Ly6d*, and other markers that have been associated with progenitor-like properties [4], and also have luminal-basal features. Furthermore, we have observed that treatment-naïve primary tumors often contain cells with an AR-low state resembling PrU, suggesting that PrU-like states may also recur in castration-resistant prostate cancer, when lineage plasticity leads to an increase of tumor cells with hybrid luminal-basal states [52].

Our studies have also addressed the similarity of rodent and human prostate cell populations, which has represented a long-standing question. Early histological studies suggested that the ventral lobe most closely resembled the human prostate [53], whereas later analyses claimed that the rat dorsal lobe most closely resembled the human prostate [54, 55]. Our meta-analysis identifies discrete LumAcinar, LumDuctal, basal, and PrU populations in the human prostate that have transcriptomic similarities to mouse LumDist, LumP, basal, and PrU, respectively. While the mouse LumL cells of the lateral lobe have the greatest transcriptomic similarity to human LumAcinar cells, this relationship is driven by a small number of genes, and ultimately mouse LumDist cells of all lobes generally resemble human secretory LumAcinar cells. However, the similarity of luminal cells from different zones of the human prostate to mouse luminal populations remains to be elucidated. Intriguingly, this analysis also revealed that the gene expression profiles of basal cells are very different between species, which is consistent with their unique histological and ultrastructural features such as differing basal:luminal ratios [56].

Our reference atlas for the human prostate has also addressed the nomenclature for human epithelial cell types. In particular, we found that the LumDuctal and PrU populations resemble the “Club” and “Hillock”

populations that were previously named due to their transcriptomic similarity to cell populations described in the lung [11]. Although this is an intriguing observation, the prostate LumDuctal/Club and PrU/Hillock populations do not uniformly and specifically express the defining markers for the corresponding lung populations (*SCGB1A1* and *KRT13*, respectively). Moreover, since these prostate cell types may not have similar functions or localizations as those in the lung, we favor the use of a simpler, descriptive nomenclature and find Club- and Hillock-like cells to be subsets within the LumDuctal and PrU populations, respectively.

Our analysis of human prostate cancer has also led to several interesting findings. Notably, we only observed significant CNV alterations in LumAcinar cells and rare NE cells across independent cohorts of treatment-naïve prostate cancer [6, 14–17], as previously noted for one of these studies [15]. This finding implies that a major cell type of origin for prostate adenocarcinoma is either a normal LumAcinar cell, or a progenitor that generates LumAcinar cells. Furthermore, primary NE prostate cancer may arise de novo from NE cells themselves, or from a progenitor that can give rise to NE cells, consistent with our identification of a CNV profile shared between acinar tumor cells and NE tumor cells in a patient sample. In this regard, our observation of two distinct LumAcinar cell states with neuroendocrine features in primary treatment-naïve prostate tumors may correspond to early steps in the transdifferentiation of luminal adenocarcinoma cells to neuroendocrine fates [48, 57, 58]. Intriguingly, both cell states as well as the NE tumor clone could be detected in the absence of androgen-deprivation therapies.

In addition, we have found that tumor-adjacent benign tissue contains cells with transcriptomic alterations that are broadly present across cell types and different regional samples. Notably, although CNVs were only observed in LumAcinar and NE cells, these transcriptomic alterations were found across epithelial cell types and were validated in LumAcinar cells by immunofluorescence staining. This widespread transcriptomic reprogramming is highly suggestive of field cancerization or “field effect” in which benign tissue contains genetic or transcriptomic alterations resembling adjacent tumor tissue. Such field cancerization has been documented in many other tumor types [59] and has been suggested in prostate cancer [60], but is not well understood.

Our findings demonstrating single-cell heterogeneity of AR signaling in treatment-naïve prostate adenocarcinomas provide deeper insights into previous studies that have classified primary tumors into subclasses with high and low AR activity [47, 61]. In particular, analyses of nearly 20,000 patient tumors analyzed by the Decipher

clinical assay revealed heterogeneity of *AR* gene expression and a signature of canonical *AR* target genes, splitting tumors into *AR*-positive and *AR*-low subsets, with the *AR*-low tumors displaying worse treatment response and increased expression of neuroendocrine markers [47]. These results are consistent with an independent retrospective study of over 600,000 patients, showing poorer outcomes and higher expression of neuroendocrine markers by high-grade tumors expressing low levels of prostate-specific antigen (PSA), an *AR*-regulated gene [61]. Interestingly, we have identified at least one cell state with neuroendocrine features that is associated with the *AR*-low population.

Our current study indicates that this heterogeneity in *AR* signaling exists at the single-cell level within patient tumors, and that the previous classifications of *AR*-positive and *AR*-low tumors should be further refined to reflect the heterogeneous composition of patient tumors and the possibility of tumor evolution altering the balance of *AR*-positive and *AR*-low states. Given that the *AR*-low population transcriptionally resembles a PrU-like state, these *AR*-low tumor cells may display greater castration-resistance. Notably, in the mouse prostate, a transition to PrU-like expression profiles is observed in the context of regression following castration (Fig. 3), with distal LumA cells displaying the most pronounced shift (Additional file 1: Fig. S7). Furthermore, PrU cells have the greatest progenitor potential among the epithelial populations in functional assays [4], a feature that may also contribute to castration-resistance.

## Conclusions

Our study has generated aggregated reference atlases for the human and mouse prostate and has shown the remarkable consistency of prostate tissue cell types between species. In addition, our findings have revealed profound cellular heterogeneity and plasticity that has significant implications for the origin and phenotypes of prostate diseases. Notably, the transcriptomic shifts observed following castration suggest the importance of cells with hybrid luminal-basal features resembling the PrU cell type in normal tissue homeostasis as well as in cancer. However, we also found that widespread transcriptomic plasticity is not necessarily dependent on loss of androgen signaling, as we observed transcriptomic reprogramming reminiscent of field cancerization in multiple patient samples with hormonally intact tumors. Consistent with this view, we identified substantial *AR*-low populations mirroring a PrU transcriptome within treatment-naïve primary prostate tumors. Taken together, these findings raise the possibility that castration-resistance and predisposition to neuroendocrine differentiation are pre-existing properties

that may be selected in part by androgen-deprivation therapies. Consequently, the detection of such *AR*-low tumor cells in treatment-naïve tumors may represent an important step in designing precision therapies for primary prostate cancer.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13073-025-01432-w>.

Supplementary Material 1. Figures S1-S17.

Supplementary Material 2. Table S1. Key resources table.

Supplementary Material 3. Table S2. Mouse prostate epithelial expression signatures.

Supplementary Material 4. Table S3. Human prostate expression signatures.

Supplementary Material 5. Table S4. Signatures of LumAcinar groups in primary human prostate tumors.

## Acknowledgements

We would like to thank Cory Abate-Shen, Brett Carver, and Charles Sawyers for helpful discussions on this work, and thank Erin Bush and Peter Sims for assistance with single-cell sequencing. This work utilized the Columbia Genomics and High Throughput Screening Shared Resource as well as the Molecular Pathology Shared Resource of the Herbert Irving Comprehensive Cancer Center, which is supported in part by the Cancer Center Support Grant P30CA013696. For assistance with electron microscopy, we thank Alice Liang, Chris Petzold, and Joseph Sall at the New York University Langone Health DART Microscopy Lab, which is partially funded by NYU Cancer Center Support Grant P30CA016087 and by S10OD019974. We also want to thank Teresa Rosa for her assistance in organizing the project.

## Authors' contributions

Conceptualization: L.A., L.C., R.R., and M.M.S.; methodology: L.A., L.C.; formal analysis: L.A., L.C., J.R.C.; investigation and experiments: L.C.; resources: C.J.L., H.H.; data curation: L.A., L.C., and J.R.C.; writing (original draft): L.C.; writing (review and editing): L.C., M.M.S., L.A., H.H., and R.R.; supervision: R.R. and M.M.S.; funding acquisition: L.C., J.R.C., R.R., and M.M.S. All authors read and approved the final manuscript.

## Funding

These studies were supported by NIH grants R01CA238005 (M.M.S.), R01CA251527 (M.M.S.), P01CA265768 (M.M.S.), U01CA261822 (R.R. and M.M.S.), R35CA253126 (R.R.), U01CA243073 (R.R.), by SU2C Convergence 3.14 (R.R.), by the Prostate Cancer Foundation (M.M.S.), and by fellowships from the NIH (F32CA261152; J.C.) and the National Science Foundation DGE 16-44869 (L.C.).

## Availability of data and materials

Single-cell RNA-sequencing data generated in this study are available in the Gene Expression Omnibus (GEO) under the accession number GSE224452 at <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi> [18]. Other datasets analyzed in this study are described and cited in the corresponding section of the "Methods" and are also listed in Additional file 2: Table S1.

## Declarations

### Ethics approval and consent to participate

Animal studies were approved by and were performed according to ethical standards set by the Columbia University Irving Medical Center (CUIMC) Institutional Animal Care and Use Committee (IACUC), under protocol AABT5655. For human studies, patients provided written informed consent to participate in this study under protocol AAAN8850 approved by the Institutional Review Board at Columbia University Irving Medical Center. Prostate tissues were

provided by the Tissue Bank of the Herbert Irving Comprehensive Cancer Center. This research conformed to the principles of the Helsinki Declaration.

#### Consent for publication

Not applicable.

#### Competing interests

R.R. is a founder of Genotwin and a member of the Scientific Advisory Board of Diotech Pharmacogenetics and Flahy. M.M.S. is a consultant for K36 Therapeutics. None of these activities is related to the work described in this manuscript. The other authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Program for Mathematical Genomics, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA. <sup>2</sup>Department of Systems Biology, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA. <sup>3</sup>Department of Biomedical Informatics, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA. <sup>4</sup>Department of Medicine, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA. <sup>5</sup>Department of Genetics and Development, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA. <sup>6</sup>Department of Urology, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA. <sup>7</sup>Department of Pathology and Cell Biology, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA. <sup>8</sup>Herbert Irving Comprehensive Cancer Center, Columbia University Vagelos College of Physicians and Surgeons, New York, NY 10032, USA.

Received: 15 April 2024 Accepted: 8 January 2025

Published online: 17 January 2025

#### References

- Abate-Shen C, Shen MM. Molecular genetics of prostate cancer. *Genes Dev.* 2000;14:2410–34.
- Toivanen R, Shen MM. Prostate organogenesis: tissue induction, hormonal regulation and cell type specification. *Development.* 2017;144:1382–98.
- Crowley L, Shen MM. Heterogeneity and complexity of the prostate epithelium: new findings from single-cell RNA sequencing studies. *Cancer Lett.* 2022;525:108–14.
- Crowley L, Cambuli F, Aparicio L, Shibata M, Robinson BD, Xuan S, et al. A single-cell atlas of the mouse and human prostate reveals heterogeneity and conservation of epithelial progenitors. *eLife.* 2020;9:e59465.
- Guo W, Li L, He J, Liu Z, Han M, Li F, et al. Single-cell transcriptomics identifies a distinct luminal progenitor cell type in distal prostate invagination tips. *Nat Genet.* 2020;52:908–18.
- Karthus WR, Hofree M, Choi D, Linton EL, Turkekel M, Bejnood A, et al. Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science.* 2020;368:497–505.
- Joseph DB, Henry GH, Malewska A, Iqbal NS, Ruetten HM, Turco AE, et al. Urethral luminal epithelia are castration-insensitive cells of the proximal prostate. *Prostate.* 2020;80:872–84.
- Mevel R, Steiner I, Mason S, Galbraith LC, Patel R, Fadlullah MZ, et al. RUNX1 marks a luminal castration-resistant lineage established at the onset of prostate development. *eLife.* 2020;9:e60225.
- Joseph DB, Henry GH, Malewska A, Reese JC, Mauack RJ, Gahan JC, et al. Single-cell analysis of mouse and human prostate reveals novel fibroblasts with specialized distribution and microenvironment interactions. *J Pathol.* 2021;255:141–54.
- Kwon OJ, Zhang Y, Li Y, Wei X, Zhang L, Chen R, et al. Functional heterogeneity of mouse prostate stromal cells revealed by single-cell RNA-seq. *iScience.* 2019;13:328–338.
- Henry GH, Malewska A, Joseph DB, Malladi VS, Lee J, Torrealba J, et al. A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell Rep.* 2018;25:3530–42.
- Eraslan G, Drokhylyansky E, Anand S, Fiskin E, Subramanian A, Slyper M, et al. Single-nucleus cross-tissue molecular reference maps toward understanding disease gene function. *Science.* 2022;376: eabl4290.
- Tabula Sapiens C, Jones RC, Karkania J, Krasnow MA, Pisco AO, Quake SR, et al. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Science.* 2022;376: eabl4896.
- Song H, Weinstein HNW, Allegakoen P, Wadsworth MH 2nd, Xie J, Yang H, et al. Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nat Commun.* 2022;13:141.
- Hirz T, Mei S, Sarkar H, Kfoury Y, Wu S, Verhoeven BM, et al. Dissecting the immune suppressive human prostate tumor microenvironment via integrated single-cell and spatial transcriptomic analyses. *Nat Commun.* 2023;14:663.
- Ge G, Han Y, Zhang J, Li X, Liu X, Gong Y, et al. Single-cell RNA-seq reveals a developmental hierarchy super-imposed over subclonal evolution in the cellular ecosystem of prostate cancer. *Adv Sci (Weinh).* 2022;9: e2105530.
- Chen S, Zhu G, Yang Y, Wang F, Xiao YT, Zhang N, et al. Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. *Nat Cell Biol.* 2021;23:87–98.
- Aparicio L, Crowley L, Christin JR, Laplaca CJ, Hibshoosh H, Rabadan R, et al. Single-cell RNA-seq analysis of additional mouse prostates. *Gene Expression Omnibus*; 2024. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE224452>.
- Shen MM, Aparicio L, Cambuli F, Crowley L, Shibata M. Single-cell RNA-seq analysis of mouse and human prostate. *Gene Expression Omnibus*; 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE150692>.
- Strand DW, Henry GH. Single-cell RNA-sequencing of adult mouse prostates. *Gene Expression Omnibus*; 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145861>.
- Mevel R, Lacaud G. Runx1 marks a luminal castration resistant lineage established at the onset of prostate development. *Gene Expression Omnibus*; 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE151944>.
- Karthus WR, Hofree M, Choi D, Linton EL, Turkekel M, Bejnood A, et al. Differentiated prostate luminal cells acquire enhanced regenerative potential after androgen ablation. *Gene Expression Omnibus*; 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE146811>.
- Guo W. Single-cell transcriptomics identifies a distinct luminal progenitor cell type in the distal prostate invagination tips. *National Omics Data Encyclopedia*; 2020. <https://www.biosino.org/node-cas/login?service=https://www.biosino.org/node/auth/login>.
- Tabula Sapiens C. The Tabula Sapiens: a multiple-organ, single-cell transcriptomic atlas of humans. *Tabula Sapiens*; 2022. <https://tabula-sapiens.sf.czbiohub.org>.
- Henry GH. Viable human prostate transition zone. *Gene Expression Omnibus*; 2018. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSM3293878>.
- Strand DW, Henry GH. Single-cell RNA-sequencing of adult human prostates from BPH patients. 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE145838>.
- Karthus WR, Hofree M, Choi D, Linton EL, Turkekel M, Bejnood A, et al. Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Data use oversight system*; 2020. <https://duos.broadinstitute.org>.
- Ren S. Single cell analysis reveals onset of multiple progression associated transcriptomic remodellings in prostate cancer. *Gene Expression Omnibus*; 2020. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141445>.
- Mei S. Dissecting the immune suppressive human prostate tumor microenvironment via integrated single-cell and spatial transcriptomic analyses. *Gene Expression Omnibus*; 2022. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181294>.
- Song H, Weinstein HN, Allegakoen P, Wadsworth II MH, Xie J, Yang H, et al. Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Gene Expression Omnibus*; 2021. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE176031>.
- Ci W. PRAD single-cell sequencing. *Genome sequence archive for human*; 2022. <https://ngdc.cncb.ac.cn/gsa-human/browse/HRA000823>.
- Aparicio L, Bordyuh M, Blumberg AJ, Rabadan R. A random matrix theory approach to denoise single-cell data. *Patterns.* 2020;1: 100035.
- Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* 2018;19:15.

34. Polanski K, Young MD, Miao Z, Meyer KB, Teichmann SA, Park JE. BBKNN: fast batch alignment of single cell transcriptomes. *Bioinformatics*. 2020;36:964–5.
35. Hie B, Bryson B, Berger B. Efficient integration of heterogeneous single-cell transcriptomes using Scanorama. *Nat Biotechnol*. 2019;37:685–91.
36. Moon KR, van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, et al. Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol*. 2019;37:1482–92.
37. Setty M, Kiseliovas V, Levine J, Gayoso A, Mazutis L, Pe'er D. Characterization of cell fate probabilities in single-cell data with Palantir. *Nat Biotechnol*. 2019;37:451–60.
38. Tickle T, Tirosh I, Georgescu C, Brown M, Haas B. inferCNV of the Trinity CTAT Project. Klarman Cell Observatory. Cambridge: Broad Institute of MIT and Harvard; 2019.
39. Kolouri S, Park S, Thorpe M, Slepcev D, Rohde GK. Optimal mass transport: signal processing and machine-learning applications. *IEEE Signal Process Mag*. 2017;34:43–59.
40. Villani C. Topics in optimal transportation. *Am Math Soc*. 2003;58:58.
41. Wang X, Xu H, Cheng C, Ji Z, Zhao H, Sheng Y, et al. Identification of a Zeb1 expressing basal stem cell subpopulation in the prostate. *Nat Commun*. 2020;11:706.
42. Wei X, Zhang L, Zhang Y, Cooper C, Brewer C, Tsai CF, et al. Ablating Lgr5-expressing prostatic stromal cells activates the ERK-mediated mechanosensory signaling and disrupts prostate tissue homeostasis. *Cell Rep*. 2022;40: 111313.
43. Isaacs JT: Control of cell proliferation and cell death in the normal and neoplastic prostate: a stem cell model. In *Benign prostatic hyperplasia*. Edited by Rodgers CH, Coffey DS, Cunha G, Grayshack JT, Henman R, Horton R. Washington, DC: Department of Health and Human Services; 1985:85–94.
44. Tsujimura A, Koikawa Y, Salm S, Takao T, Coetzee S, Moscatelli D, et al. Proximal location of mouse prostate epithelial stem cells: a model of prostatic homeostasis. *J Cell Biol*. 2002;157:1257–65.
45. Shen MM, Abate-Shen C. Molecular genetics of prostate cancer: new prospects for old challenges. *Genes Dev*. 2010;24:1967–2000.
46. Travaglini KJ, Nabhan AN, Penland L, Sinha R, Gillich A, Sit RV, et al. A molecular cell atlas of the human lung from single-cell RNA sequencing. *Nature*. 2020;587:619–25.
47. Spratt DE, Alshalalfa M, Fishbane N, Weiner AB, Mehra R, Mahal BA, et al. Transcriptomic heterogeneity of androgen receptor activity defines a de novo low AR-active subclass in treatment naive primary prostate cancer. *Clin Cancer Res*. 2019;25:6721–30.
48. Chen CC, Tran W, Song K, Sugimoto T, Obusan MB, Wang L, et al. Temporal evolution reveals bifurcated lineages in aggressive neuroendocrine small cell prostate cancer trans-differentiation. *Cancer Cell*. 2023;41(2066–2082): e2069.
49. Kirk JS, Wang J, Long M, Rosario S, Tracz A, Ji Y, et al. Integrated single-cell analysis defines the epigenetic basis of castration-resistant prostate luminal cells. *Cell Stem Cell*. 2024;31:1203–21.
50. Ousset M, Van Keymeulen A, Bouvencourt G, Sharma N, Achouri Y, Simons BD, et al. Multipotent and unipotent progenitors contribute to prostate postnatal development. *Nat Cell Biol*. 2012;14:1131–8.
51. Shibata M, Epsi NJ, Xuan S, Mitrofanova A, Shen MM. Bipotent progenitors do not require androgen receptor for luminal specification during prostate organogenesis. *Stem Cell Reports*. 2020;15:1026–36.
52. Chan JM, Zaidi S, Love JR, Zhao JL, Setty M, Wadosky KM, et al. Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science*. 2022;377:1180–91.
53. Price D. Comparative aspects of development and structure in the prostate. *Natl Cancer Inst Monogr*. 1963;12:1–27.
54. Aumuller G, Seitz J, Lilja H, Abrahamsson PA, von der Kammer H, Scheit KH. Species- and organ-specificity of secretory proteins derived from human prostate and seminal vesicles. *Prostate*. 1990;17:31–40.
55. Berquin IM, Min Y, Wu R, Wu H, Chen YQ. Expression signature of the mouse prostate. *J Biol Chem*. 2005;280:36442–51.
56. El-Alfy M, Pelletier G, Hermo LS, Labrie F. Unique features of the basal cells of human prostate epithelium. *Microsc Res Tech*. 2000;51:436–46.
57. Beltran H, Prandi D, Mosquera JM, Benelli M, Puca L, Cyrta J, et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat Med*. 2016;22:298–305.
58. Zou M, Toivanen R, Mitrofanova A, Floch N, Hayati S, Sun Y, et al. Transdifferentiation as a mechanism of treatment resistance in a mouse model of castration-resistant prostate cancer. *Cancer Discov*. 2017;7:736–49.
59. Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization. *Nat Rev Cancer*. 2018;18:19–32.
60. Nonn L, Ananthanarayanan V, Gann PH. Evidence for field cancerization of the prostate. *Prostate*. 2009;69:1470–9.
61. Mahal BA, Yang DD, Wang NQ, Alshalalfa M, Davicioni E, Choerung V, et al. Clinical and genomic characterization of low-prostate-specific antigen, high-grade prostate cancer. *Eur Urol*. 2018;74:146–54.

## Publisher's Note

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