HuPSA and MoPSA Atlases Revealed Novel Cell Populations and Lineage Plasticity at Single-Cell Resolution in Prostate Cancer

Siyuan Cheng[1,2], Lin Li[1,2], Yingli Shi[1,2], Xiuping Yu[1,2,3]

1. Department of Biochemistry and Molecular biology, LSU Health Shreveport, Shreveport, LA

2. Feist-Weiller Cancer Center, LSU Health Shreveport, Shreveport, LA

3. Department of Urology, LSU Health Shreveport, Shreveport, LA

Correspondence to S. Cheng and X. Yu

Abstract

Androgen deprivation therapy has improved patient survival. Nevertheless, treatment resistance inevitably emerges due to the complex interplay of tumor heterogeneity and lineage plasticity. We integrated scRNAseq data from billions of cells, including both public cohorts and data generated in our laboratory, and generated the HuPSA (Human Prostate Single cell Atlas) and MoPSA (Mouse Prostate Single cell Atlas) datasets.Through unsupervised clustering and manually annotation, both atlases not only validated previously known prostate adenocarcinoma (AdPCa), neuroendocrine prostate cancer (NEPCa), stromal, and immune cell populations but also unearthed the less described populations including MMP7^{+} normal prostate club cells and two novel lineage plastic cancerous populations, namely progenitor-like and KRT7^{+} cells. Immunostaining experiments confirmed the presence of these populations in both human and mouse tissues, solidifying their significance in PCa biology. To unravel the drivers of these distinct cell populations, we calculated the upstream regulators of the genes enriched in these cells. Furthermore, leveraging the power of state-of-the-art bioinformatics analyses, we scrutinized over one thousand human PCa bulk RNAseq samples. Employing HuPSA-based deconvolution, we reclassified these samples into different molecular subtypes, including the newly discovered KRT7 and progenitor-like groups. The HuPSA and MoPSA provide invaluable blueprints for analyzing and interpreting external PCa single-cell RNAseq datasets. Our data elucidates the roadmap of PCa progression, showcasing the development of heterogeneous populations through lineage plasticity. This understanding holds promise for guiding the development of precise medicine in PCa field.

Introduction

Prostate cancer (PCa) represents a significant public health challenge in the United States, being the most commonly diagnosed cancer among men[1]. Although various treatments including anti-androgen drugs and androgen synthesis inhibitors have been employed, these treatments eventually fail, leading to the

progression of the cancer into a "castration-resistant" stage. In this stage, cancer cells acquire the ability to proliferate independently of androgen stimulation[2]. A subset of castration-resistant prostate cancer (CRPCa) cases retains the characteristics of prostate adenocarcinoma (AdPCa), expressing the markers associated with the luminal epithelium of the prostate, such as CK8, CK18, and HOXB13[3,4]. Another portion of CRPCa cases, approximately 30%, presents an alternative phenotype known as neuroendocrine prostate cancer (NEPCa). NEPCa cells express low or absent levels of the androgen receptor (AR) and its downstream genes, while expressing high levels of neuroendocrine markers such as CHGA and NCAM1 (CD56). [5]. Our previous study found NEPCa cells lose the prostatic lineage features represented by "HOX code"[4]. In addition to the aforementioned phenotypes, there exists a lesser studied phenotype known as double negative prostate cancer (DNPCa), characterized by the absence of both luminal and neuroendocrine features[6,7].

Lineage plasticity is a phenomenon where tumor cells undergo changes in their morphology and function to adapt to stressful conditions or environmental pressures. This dynamic process enables cells to transit to different cellular states[8]. In our previous study, we developed a method to quantify the lineage plasticity by calculating the Pearson correlation between the expression of all 39 HOX genes in samples and the prostatic HOX code derived from human prostate cancer (PCa) cell lines[4]. Additionally, previous studies have shown that SOX2, SRRM4, ONECUT2, BRN2, PEG10, Jak-stat signaling can contribute to prostate cancer lineage plasticity[7,9-13].

Numerous studies utilizing single-cell RNA sequencing (scRNAseq) have been conducted on human and mouse tissues to comprehensively explore the progression path of lineage plasticity and the spectra of population heterogeneity. These investigations have revealed distinct populations in human samples, including AdPCa, NEPCa, basal epithelial, club, stromal, and immune cells. In mouse samples, populations such as AdPCa, NEPCa, Foxi1^{+}, Pou2f3^{+}, and basal epithelial cells have been identified[7,14-28]. Additionally, it is crucial to recognize that the particular populations identified in these studies may exhibit inconsistencies across different studies. Moreover, due to the utilization of relatively small sample sizes in each study, there is a possibility that certain rare populations could be overlooked. It is also important to exercise caution in the interpretation of cell populations because they might be model-specific.

In the current study, we performed the state-of-the-art bioinformatics analysis of combined scRNAseq datasets. Through alignment and integration, we established two integrated scRNAseq datasets: Human Prostate Single cell Atlas (HuPSA) and Mouse Prostate Single cell Atlas (MoPSA). Leveraging these atlases, we were able to identify and characterize previously undiscovered populations within prostatic tissues.

Results

1. Establishment of HuPSA and MoPSA by compiling single cell RNA sequencing data obtained from multiple studies

Using publicly available scRNAseq data, we established a combined dataset and visualized the data on UMAP as Human Prostate Single cell Atlas (HuPSA). This dataset compiled single cell RNA sequencing data of 73 distinct samples from 6 studies, consisting of 368,831 high-quality cells. Based on their original records, these 73 samples span eight histology groups, namely Normal, Normal adjacent, Benign, AdPCa, CSPCa (castration-sensitive PCa), CRPCa, mCRPCa (metastatic CRPCa), and Cribriform PCa.

In parallel, Mouse Prostate Single cell Atlas (MoPSA) was also established. Because the publicly available mouse prostate scRNAseq datasets lack TRAMP model, a widely utilized mouse model for NEPCa research, we included in the MoPSA dataset TRAMP scRNAseq data that were newly generated by our group. The compiled MoPSA dataset consisted of 395,917 cells originated from 138 samples obtained from 13 studies. These 138 samples were derived from mice under different treatment modalities, which were classified into six groups including WT (wild-type), GEM (genetically engineered mice), Control (GEM control mice that express tamoxifen-induced YFP in prostate), Castration, Regeneration (androgen -induced prostate regeneration), and Development (developing prostate).

UMAP analysis of the compiled scRNAseq data revealed distinct cell clusters, including mesenchymal, endothelial, multiple epithelial and immune clusters (Fig. 1). Each cluster was further separated into different populations based on their molecular features, resulting in 26 and 21 populations in HuPSA and MoPSA, respectively (sFig. 1). Sankey diagrams, which visualize the data flow between the subpopulations of cells and the histology of samples, demonstrated the sample types from which the cells of each population were derived (sFig. 2).

Most epithelial clusters (including basal/basal-like, luminal, and NEPCa) were shared between HuPSA amd MoPSA, whereas Progenitor-like cluster only stood out as a cluster in HuPSA and Pou2f3 cluster was only in MoPSA. The epithelial clusters were further separated into different populations and each population was defined based on copy-number variation (CNV), expression of marker genes, and recorded sample histology (sFig. 1, sFig. 2 & sFig. 6).

2. Luminal cells in HuPSA and MoPSA

In HuPSA, the Luminal cluster expressed HOXB13, FOXA1 and NKX3-1. This cluster consisted of seven populations including Normal (contributed by Normal, Normal-adjacent, Benign, and some AdPCa samples), three groups of AdPCa_AR+ as well as AdPCa_AR high (ARhi), AdPCa_AR low (ARlo), and highly_ proliferating groups. The AdPCa_AR+_1 population only slightly deviated from Normal cells on

UMAP (Fig. 2). Its infered copy number variation (CNV) pattern was also similar to Normal population (sFig.6). This population may represent low grade PCa and cells in this population are still at early stage of transformation. On the contrary, AdPCa_AR+_2 population harbored higher level of copy number variations (sFig. 6). The highest level of CNV was present in the AdPCa_AR_high population, suggesting a genome instability in these cells. The AdPCa_AR+_2 and AR-high populations were mostly contributed by advanced PCa samples including CRPC, mCRPC, and CSPC (Fig, 2E). Another population that was mostly derived from advanced PCa samples is AdPCa_AR+_3. It departed from the AdPCa mainbody but its infered CNV pattern did not show much difference compared with the Normal cells, suggesting that genome instability may not be a driving mechanism in these cells. It is interesting to note that AdPCa_AR+_3 expressed Neuropeptide Y (NPY) as well as two smooth muscle markers (TAGLN and ACTA2). This suggests the AdPCa_AR+_3 population may have acquired certain mesenchymal features while still expressing luminal genes (HOXB13, FOXA1 and NKX3-1) (Fig, 2C&D; sFig1). These cancer cells likely have some epigenetic reprogramming.

The AdPCa_AR low population, although low in AR expression, still expressed KLK3 and NKX3-1, two AR target genes, at decent levels, indicating an active AR signaling in these cells.

The proliferating population exhibited high proliferative activity as indicated by the high expression of Ki67 and TOP2A. This population was mostly contributed by CRPC and mCRPC samples. Interestingly, CSPC samples also contributed to this population, suggesting the presence of proliferating cells in CSPC tumors (treatment naiive). The highly proliferating population also exhibited active AR signaling, indicated by the expression of AR and AR target, KLK3.

The Luminal cluster in MoPSA also contained multiple populations, including LE/AdPCa_1, AdPCa_2 & _3, and highly proliferating populations, as well as a unique Spink1 population that was not detected in HuPSA. The Spink1 population was dominantly contributed by WT mouse tissues (sFig. 2) and harbored low chromosomal copy number variation (sFig. 6). Because of its absence in HuPSA and the relatively normal molecular profiles, the Spink1 population will not be further characterized in this study.

All the other Luminal cells in MoPSA expressed HOXB13 and FOXA1, like those in HuPSA. But NKX3-1, a prostate differentiation marker, was only expressed in the normal LE/AdPCa_1 population at high level, suggesting a loss of differentiation state in the other populations. The LE/AdPCa_1 population was mostly originated from wild type mice, but GEM samples contributed to this population as well. The LE/ADPCa_1 populations harbored low chromosomal copy number variation (sFig.6). Therefore, this population may represents normal cells and the cells at early stage of transformation. On the contrary, the other AdPCa populations, including ADPCa_2 & _3, and highly proliferating populations, were

predominantly contributed by GEM (sFig. 2) and harborded slightly higher level of CNV than the LE/AdPCa_1 population (sFig. 6).

Compared with HuPSA, where all the Luminal cells expressed KLK3, an AR target gene, regardless of AR's expression status, in MoPSA, only the normal population express Pbsn, an AR target gene in murine prostate. Its low expression level in all the other luminal populations further coraborated the de-differentiation state of the PCa cells in GEM.

3. Basal/basal-like cells in HuPSA and MoPSA

The Basal/basal-like cluster in HuPSA contained three populations including Basal (KRT5+, KRT14+, TP63+), Club (MMP7+, WFDC2+) and KRT7 (KRT7+, S100A9+). The Club and KRT7 populations, although clustered together with basal epithelial cells, express low to no basal markers, suggesting their basal features but not basal identity. In MoPSA, the Basal and Krt7 populations were present as well, but the Club group was not (Fig. 1).

<u>In HuPSA, KRT7 population exhibited a lineage-switched double-negative PCa phenotype.</u>

The KRT7 population in HuPSA was characterized by the high expression of KRT7 and S100A9. Although KRT7 expression was also detected in other three populations including Club, Basal, and NEPCa (sFig. 1 and Fig. 4A), its expression level was lower in these populations, in addition to other differences in the transcription profiles, e.g. the differential expression of S100A9, thus separating them from KRT7 population. The KRT7 population in HuPSA comprised 1937 cells, 1828 of them (94%) were derived from mCRPCa samples. Also, this population exhibited higher levels of copy number variations compared to Normal cells or early AdPCa (sFig.6). These features suggest that the KRT7 population in HuPSA are cancer cells but not un-transformed basal cells, which express KRT7 as well, albeit at lower level. This is also true for the Krt7 population in MoPSA, where GEM samples were the major cell source, contributing 80% (32,011 of 39,891) of the cells in this population. The murine Krt7 population also harbored higher level of CNV than the normal LE/AdPCa_1 group (sFig.6), supporting that the cells in this population, like those in HuPSA, are not un-transformed basal cells.

The KRT7 population in HuPSA exhibited low to negligible levels of androgen receptor (AR) and AR signaling activity (Fig. 4B). Also, this population expressed low to no prostate lineage and differentiation markers FOXA1, HOXB13, and NKX3-1 (sFig.1 & Fig.4), suggesting the KRT7 cancer cells have switched lineage from luminal epithelia (Fig.4B).

This was further corroborated by the KRT7 protein expression pattern in human prostatic tissues. As shown in Fig. 5A, KRT7 expression was present in the basal epithelia but absent in the luminal epithelial cells of benign prostatic tissues. Its expression was absent in most AdPCa specimens (Fig.5B) but

scattered or patchy KRT7 expression was detected in some rare cells of advanced AdPCa (Figs. 5C and 5D).

Further, we identified the upstream signaling pathways enriched in the KRT7 population. We found that the Wnt signaling pathway, an important stem signaling that controls cell fate determination and cell proliferation[2], was active in this population. Additionally, several transcription factors, including TCF4, PAX6, NFE2L2, and HNF4G, were identified as upstream regulators of this population (sFig.8).

Krt7 population in MoPSA

Krt7 population in MoPSA expressed low to no basal markers Krt5, Krt14 and Trp63, like their counterpart in HuPSA. However, unlike the KRT7 population in HuPSA, the MoPSA Krt7 population expressed Ar and exhibited high levels of Ar signaling activity. Additionally, the murine Krt7 population expressed Hoxb13 and Foxa1, suggesting that the murine Krt7 population still sustains the prostatic luminal epithelial lineage (Fig. 4B).

Although GEM samples were the major contributors of the Krt7 population, wild type prostate samples contributed to this population as well (sFig. 2). Previous studies have shown that Krt7 is expressed in prostate basal epithelial cells. However, our analysis indicates the Krt7 population in MoPSA are none basal. To investigate the possible biological function of Krt7 population in murine prostate, we analyzed the "GSE146811" dataset generated by Sawyers' group, which is also included in MoPSA. This dataset contains prostate samples collected from mice at various time points during androgen-mediated prostate regeneration [29]. Our analysis focused on the GSE146811-derived non-basal epithelial cells in MoPSA. These cells' scRNAseq data were visualized on the UMAP (Fig. 5M) and the sample source on this UMAP was labelled in Fig. 5N, which depicted a clear castration-regeneration path with colors representing intact, castration (CX) from 1 Day to 28 Days, and regeneration (R) from 1 Day to 14 Days. As shown in Figs. 5M and 5N, the cells of Krt7 population were mostly derived from CX 28 day samples and some CX14 day samples. sFig.7 provided a split view of the same plot; it clearly illustrated the position of each sample during prostate regeneration. Interestingly, some cells (mostly derived from CX 14 day samples) were clustered in the AdPCa_2 population on MoPSA, indicating the high similarity (low Pbsn expression) in their tramscriptomes with those of PCa cells.

To validate the Krt7 populational changes during prostate regeneration, we conducted immunostaining to assess the Krt7 expression in prostate tissues since this population is characterized by high expression of Krt7. As depicted in Fig.5O-T, in the prostate tissues derived from intact mice, Krt7 expression was absent in majority of luminal epithelial cells but was present in some rare luminal epithelial cells that were located at the tips of the gland. However, following castration from 7 days to 21 days, the proportion

of Krt7+ luminal cells gradually increased. After testosterone supplement for 1 day to 15 days, the propotion of Krt7+ luminal epithelial cells gradually decreased, reverting back to baseline level like those in intact prostate. This suggests that the positive expression of Krt7 may label the prostatic luminal epithelial cells that have survived androgen deprivation. This was further coraborated by the Krt7 expression pattern in the TRAMP tumors. In the prostatic intraepithelial neoplasia (PIN) tumors developed in intact TRAMP mice, Krt7 was only expressed in a small subset of PIN cells and majority PIN cells did not express this gene (Fig.5U). However, in castrated TRAMP mice, which develop NEPCa after castration[30], Krt7 was expressed in the residual PIN cells but negative in NEPCa cells (Fig. 5V). Since the PIN lesions in castrated TRAMP do not express T-antigen, the Krt7's differential expression pattern in intact vs castrated PIN is not due to T-antigen expression but rather likely attributed to casration-mediated selection of fit cells. Therefore, similar to that in wild type prostate, the Krt7 positive PIN cells are also the cells that have survived castration.

MMP7 labels the club population

HuPSA also revealed a cell population that was located adjacent to basal cells and highly expressed MMP7 gene (Fig. 1). This population was recently discovered in an scRNAseq experiment, which is also included in our HuPSA data. In that study, this population was defined as "club" cells based on the expression of marker genes PIGR, MMP7, and CP[28], which correspond to the club cells in lung[31]. The Club population in HuPSA were characterized by the high expression of MMP7 and WFDC2. MMP7 encodes the smallest secreted matrix metalloproteinase[32]. WFDC2 encodes human epididymis secretory protein 4 (HE4) which is a protease inhibitor[33]. As depicted in Fig.6C, MMP7 exhibited high expression in club cells within the HuPSA dataset. WFDC2 genes were highly expressed not only in club cells but also in basal cells. To validate the presence of the club population, we performed IHC staining targeting Mmp7 and Wfdc2 proteins using human benign prostatic hyperplasia (BPH, non-cancer) samples. Staining of serial sections demonstrated the detection of Mmp7 and Wfdc2 proteins in a sub-population of prostate luminal epithelial cells (Fig.6D). Immunofluorescence staining on the same serial sections further confirmed that club cells are luminal cells, rather than basal epithelial cells, based on the exclusive localization of Mmp7 and the basal marker Krt5, as well as the co-localization of Mmp7 with the luminal marker Krt8 (Fig.6E). The club cells are likely non-cancer cells because of their relatively low CNV (sFig.6).

4. The Progenitor-like cluster in HuPSA and MoPSA

The Progenitor-like cluster was only present in HuPSA. It was predominantly contributed by mCRPCa samples (sFig.2). In line with this, the cells in this cluster harbored high CNVs (sFig.6). The Progenitor-like cluster was located in between the Luminal and NEPCa clusters on UMAP plot (Fig.1). Cells in this

cluster highly expressed SOX2 and FOXA2, two genes that are expressed in both embryonic prostate and NEPCa, but majority of the progenitor-like cells express low or no NEPCa markers including NCAM, INSM1 and CHGA. Additionally, the progenitor-like cells did not express AR or AR targets (KLK3 and NKX3-1), indicating an inactive AR signaling in them. These cells expressed low to no FOXA1 or HOXB13, two prostatic lineage markers (sFig.1), suggesting a poorly differentiated state of these cells.

Pou2f3 cluster was detected in MoPSA but not in HuPSA. This Pou2f3 population was first identified and reported in double/ triple knock-out mouse models[7]. This cluster was characterized by the expression of Pou2f3, a marker of the cells that are in a transition state from AdPCa to NEPCa26. From MoPSA, these cells expressed low level of Sox2, thus the Pou2f3 population is possibly the murine quanterpart of HuPSA's progenitor-like population.

5. Human NEPCa Can Be Further Detailed into Subpopulations, Unlike Mouse NEPCa

The NEPCa cells was extracted from HuPSA and MoPSA and visualized on UMAP for higher resolution. The results indicate that the human NEPCa cluster could be further divided into different sub-populations, whereas the mouse NEPCa population was relatively uniform. Through unsupervised clustering and examination of marker genes, the HuPSA NEPCa cluster was classified into five sub-populations: CHGA, NEUROD1, G2M, TOX, and STAT3 (Fig.3A). All NEPCa sub-populations exhibited high expression levels of NE markers INSM1 and NCAM1 (CD56) (sFig.5A-C) but each subpopulation expressed its unique gene. The CHGA subpopulation expressed high levels of the NE marker CHGA. It represents the canonical neuroendocrine phenotype in prostate cancer, (Fig. 3D). Interestingly, the remaining HuPSA sub-populations expressed relatively low but not negligible levels of CHGA.

The expression of ASCL1, a driver transcription factor of NEPCa[34], was detected in all the NEPCa subpopulations except NEUROD1. The NEUROD1+ sub-population was characterized by the high expression of NEUROD1. It exhibit elevated Wnt signaling activity (Fig.3B). This is consistent with a previous study which has demonstrated that Wnt signaling activates NEUROD1 expression during neurogenesis in mice[35]. By examining the feature plots of NEUROD1 and ASCL1 in HuPSA, we discovered an exclusive expression pattern between NEUROD1 and ASCL1 (Fig.3D).

The G2M sub-population was characterized by the high expression of the signature genes of cell cycle G2M phase such as NUCKS1, TOP2A, CENPF (sTable5). In contrast, the other subpopulations exhibited evenly distribution of the S and G2M phases (sFig.5D). The fact that G2M subpopulation stands out as a group suggests these cells may have prolonged G2M phase and higher level of DNA damage compared with the other subpopulations.

The TOX sub-population was characterized by the high expression of Thymocyte selection-associated HMG box gene (TOX). These cells exhibited elevated activity in the VEGF signaling pathway and high expression of TCF4, a transcription factor of Wnt signaling (Fig.3B-D). The STAT3 sub-population not only expressed high levels of the STAT3 gene but also displayed high activity in JAK-STAT signaling and the transactivation of STAT3 target genes, suggesting this sub-population might be driven by JAK-STAT3 signaling (Fig.3B-D).

6. Non-epithelial clusters

In addition to epithelial clusters, the HuPSA and MoPSA visualized mesenchymal, endothelial, and immune cells as well. The Mesenchymal cluster can be further separated into three populations, Fibroblast, Smooth muscle, and Pericyte, for both species. Compared with MoPSA, a larger proportion of T cells were detected in HuPSA (Fig.1). Based on the Sankey diagrams (sFig.2), the T cells in HuPSA were mainly contributed by normal adjacent and AdPCa samples. The other histology groups (normal, benign, cribriform, CSPC, CRPCa and mCRPCa samples) contained fewer T cells, which was like the mouse tissues in MoPSA.

7. HuPSA-based deconvolution of RNAseq data re-classified human PCa specimens and identified novel KRT7 and progenitor-like samples

Bulk RNA-seq is commonly used in prostate cancer research to identify diagnostic markers and treatment targets. However, bulk RNA-seq data can only reflect gene expression in a mixed cell populations. It faces challenges in molecular subtyping due to two reasons: limited knowledge of prostate cancer subtypes and tumor heterogeneity. The gene expression profiles established in HuPSA for each cell population can be used to overcome these challenges. HuPSA dataset integrated RNAseq data of 368,831 cells at single cells resolution and spaned various types of cell populations. Each population was defined based on its transcriptome profile. By employing covariance-based deconvolution technology against bulk RNA-seq data [36], we can estimate the populational composition for each PCa samples.

In this study, we analyzed publicly avalible bulk RNAseq data that were derived from 877 prostatic specimens after data processing from raw read files and ragid quanlity control. We examined each sample's composition of the following populations: Luminal, KRT7, Basal, Club, Progenitor, NEPCa, Mesenchymal (Fibroblast, SM, Pericyte), and Immune. The luminal populations were combined into a single group because of the relatively small differences in their transcriptomes. The prostatic tissues were re-classified based on their dominant populational composition. As shown in Fig. 7, most specimens can be re-classified as Luminal (n=570), KRT7 (n=9), Progenitor-like (n=7), and NEPCa (n=43). Samples contained more than one dominant cell populations and assigned to three mixture groups: KRT7/AdPCa

mix (n=1), progenitor-like/AdPCa mix (n=1), and NEPCa/AdPCa mix (n=1). Samples with a low percentage of dominant population were categorized as unclassified (n=167). The contaminated samples (n=78) were dominant by mesenchymal or immune cells, but they contained cancer cells as well.

The Luminal samples were further separated into Normal (n=67) and AdPCa groups based on their original histology record because the HuPSA cannot differentiate early stage AdPCa and luminal epithelial transcriptome well. Among the AdPCa group, 293 samples expressed AR mRNA at high (>30) and the other 210 samples expressed low levels of AR.

To facilitate data transparency and for researchers to visualize the expression of their genes of interest in prostatic cells on HuPSA , similar to our previous CTPC App[37], we generated a web App (https://pcatools.shinyapps.io/ProAtlas_dev/) to visualize the gene expression in human samples. Through inputting the official gene symbol, researchers can visualize the mRNA expression level in newly defined cell populations including the rare KRT7 and progenitor-like groups.

Discussion

Advanced technologies have revolutionized our ability to understand biological questions. RNA sequencing, for instance, has provided an unbiased and high-throughput method to detect gene expression. Another significant breakthrough is single cell RNA sequencing, which allows researchers to investigate transcriptome with single-cell resolution. These technological advancements have expanded our research scope, enabling us to address classical questions in cancer research, such as the tumor microenvironment and immune response, as well as newly proposed challenges like phenotypic plasticity, cellular senescence, and nonmutational epigenetic reprogramming[38]. HuPSA and MoPSA, as integrated scRNAseq datasets, have the capability to detect various cell populations and visualize cancer cells with different phenotypes on dimensional reduction maps. By analyzing differentially expressed genes, upstream transcription factor activities, and signaling pathway activities (sFig.9&10), both atlases can provide preliminary insights into potential druggable targets for different phenotypes and shed light on the driver mechanisms of phenotypic plasticity.

In the field of cancer immunology, immune cell populations from HuPSA and MoPSA can be further refined by domain experts as they can define immune sub-populations more precisely. Still, HuPSA and MoPSA can serve as immune landscapes for different PCa subtypes by providing counts of different immune cells within specific samples or groups of samples. Furthermore, the characterization of immune populations using existing bulk RNA-seq data has been an area of interest for several years. Our pipeline, which utilizes HuPSA to deconvolute PCa patient bulk RNA-seq data, can be repurposed to decipher the immune composition within PCa tissue.

In addition to immune studies, HuPSA and MoPSA hold great potential for uncovering cancer-stromal interactions. Cellular communication often relies on ligand-receptor interactions. By utilizing the existing ligand-receptor database "CellChatDB," we predicted ligand-receptor pairs between and within the populations in HuPSA and MoPSA. As depicted in sFig.8, the complex network of cell-cell interactions highlights the necessity to expand our research scope beyond cancer cells and delve into the entire tumor microenvironment. Genes that are not expressed in cancer cells may still contribute to tumor growth, and signaling pathways activated in cancer cells may not have direct effects on the cancer cells themselves.

While integrated scRNAseq datasets like HuPSA and MoPSA provide valuable insight for studying epigenetic reprogramming, cell lineage and cancer progression, these large scale data still have limitations. As mentioned in the results section, scRNAseq, as a transcriptome study, cannot effectively differentiate early-stage cancer (AdPCa_AR+) from the normal progenitor cell population (luminal epithelial cells). On the UMAP plots, normal and early-stage cancer cells are located together, suggesting a relatively similar transcriptome. It is possible that the data integration process in scRNAseq minimized the differences between these populations. However, differential expression analysis on luminal epithelial cells (cells from normal tissues) and early-stage AdPCa cells using non-integrated data could not differentiate these two population well either. This indicates that even in the untransformed raw transcriptome, there is only minimal difference between these populations. Therefore, scRNAseq may not be the optimal tool for characterizing the early stages of carcinogenesis, which are primarily driven by DNA alterations.

Another drawback pertains to the analysis of scRNAseq data itself. The high-throughput nature of scRNAseq data, with its complex multi-dimensional gene expression profiles, can be overwhelming for human interpretation. Transcriptome identities are defined by thousands of gene expression values along numerous axes in a multi-dimensional space. To aid visualization, dimensional reduction techniques such as PCA, t-SNE, and UMAP are applied to reduce the data to lower dimensions. However, it is important for researchers to recognize that the transcriptome identities represented on UMAP plots are not a perfect reflection of reality. In addition, the visual representation can change significantly based on the choice of dimensional reduction parameters. In our opinion, as a famous ancient Chinese poet Su Shi eloquently put it, "It's a range viewed in face and peaks viewed from the side, assuming different shapes viewed from far and wide.", researchers should objectively process the scRNAseq data using appropriate pipelines, but data interpretation may still be subjective.

Methods

1. Single-cell RNA sequencing on TRAMP tumor

Mouse prostate tumor tissues were harvested and dissociated using "Tumor Dissociation Kit, mouse, #130-096-730" from Miltenyi Biotec. The dead cells were removed by "Dead Cell Removal Kit, #130-090-101" from Miltenyi Biotec. The red blood cells were removed by "Red Blood Cell Lysis Solution, # 130-094-183" from Miltenyi Biotec. Then the cDNA libraries were constructed using Chromium Next GEM Single Cell 3' Reagent Kit v3.1. cDNA was fragmented, processed, and size selected. Adaptors were ligated and samples were cleaned up. Dual indexes were attached via PCR. The completed libraries were cleaned up and assessed for quality with the Agilent TapeStation D5000 High Sensitivity Kit. Libraries were quantitated with qPCR (NEBNext Library Quant Kit for Illumina, Bio-rad CFX96 Touch Real-Time PCR). Libraries were normalized to 1.5 pM and pooled. The library pool was denatured and diluted to approximately 300 pM. A library of 1.5 pM PhiX was spiked in as an internal control. Paired-end sequencing was performed on an Illumina NovaSeq 6000.

2. Public scRNAseq collection and alignment

The human and mouse scRNAseq data were collected from Gene Expression Omnibus (GEO). Only 10X genomic based single-cell experiments were considered. [7,14-27,29,39]. All FASTQ files were downloaded using SRA-Toolkit fasterq-dump function and aligned to human (GRCh38 from NCBI) and mouse (GRCm38 from Ensembl) using Alevin-fry[40]. The nf-core/scrnaseq pipeline was used for software environment, paralleling and quantity control. Linux ubuntu was used as the operating system.

3. scRNAseq Data integration and processing

Single-cell read count data were load into R environment using "loadFry" function from Fishpond package[41]. Subsequently "Seurat V5"[42] was used for quality control, data integration, UMAP dimensional reduction, clustering, annotation and data projection. Quality control was performed based on the total RNA count and feature count for each cell. Cells with extremely low or high RNA/feature counts were excluded to ensure the removal of low-quality and doublets. The RPCA model was used for data integration. In the integration process, each study (instead of sample) was treated as one "batch" to prevent over integration. After the UMAP dimensional reduction and clustering, the clusters were then labeled with assign population names based on the marker genes expression and original histology information. The human lung V2 (HLCA) annotated reference dataset from "Azimuth"[43] were referred manually for immune and stromal populations annotation. The R package "SCpubr"[44] was used for data visualization. Dim plots were utilized to visualize the populations of HuPSA and MoPSA on UMAP. Feature plots were employed to visualize gene expression on the UMAP. Geyser plots were used to

visualize and quantify genes expression. Sankey plots were used to depict the "flow" of cells between populations, histology, and study groups. The single cell copy number alterations were calculated and visualized using “InferCNV” package[45]. Pathway activity was calculated by “decoupleR”[46]. Transcription factor activity was calculated using “dorothea”[46]. The “CSCDRNA” package[36] was used for HuPSA based bulk RNAseq deconvolution. The “CellChat” package[47] was used to predict ligand-receptor pairs in HuPSA and MoPSA. All detailed parameters were described in the R scripts deposited to GitHub pages.

4. Bulk RNAseq data collection and processing

The human PCa bulk RNAseq data raw FASTQ files was collected from multiple cohorts. GEO accessions: GSE120795[48], GSE126078[49], GSE156289[50], GSE114740, GSE217260, GSE179321[51], GSE112786[52], GSE104131[53], GSE31528[54], GSE54460[55], GSE22260[56], GSE216490[57]; dbGaP accessions: phs000909.v1.p1[58], phs000915.v2.p2[3]. The TCGA_PRAD read counts and TPM data were acquired using “recount3”[59] R package. The sequencing reads were aligned to human GRCh38 genome (from NCBI) using STAR[60] and Salmon[61]. The nf-core/rnaseq pipeline was used for software environment, paralleling and quantity control. Linux ubuntu was used as the operating system. Samples were considered high quality and retained if they met the following criteria: duplicated reads less than 60%, total reads more than 20 million, and STAR unique mapped percentage higher than 70%. However, the TCGA samples were not filtered as these specific metrics were not available. Eventually 877 high quality samples from normal, BPH, tumor, primary, AdPCa, CRPCa, mCRPCa, NEPCa, organoid, PDX (patient-derived xenograft) histology groups were kept and used for further HuPSA based deconvolution.

5. Mouse castration and Testosterone treatments

Wild-type mice were obtained from Jackson laboratory (Bar Hrbor, ME). Mice at 8 weeks of age were used for this study. Mice were divided into 3 groups: intact group, surgically castrated group and surgically castrated mice given exogenous testosterone replacement at 3 weeks post-castration. Additionally, the castrated mice group were divided into two subgroups based on the time since surgery, consisting of 7 days and 21 days. Meanwhile, the castrated mice receiving exogenous testosterone replacement were further divided into six groups based on the duration of testosterone treatment, which included 1 day, 2days, 3 days, 4 days, 7 days, and 15 days. All isolated prostate tissue was fixed in 10% neutral buffered formalin for embedding and sectioning. All animal procedures were performed in accordance with institutional guidelines and approved by the Animal care and Use committee at the LSU Health Shreveport.

6. Immunohistochemical, immunofluorescence staining and Western blot

De-identified human prostate tissue specimens were obtained from LSU Health-Shreveport Biorepository Core, Overton Brooks VA Medical Center, Ochsner Health System Biorepository and Tissue for Research as described previously[62].

Immunostaining was performed using Vectastain elite ABC peroxidase kit (Vector Laboratories, Burlingame, CA). Primary antibodies include KRT7 (abclonal, A4357), MMP7 (abclonal, A0695), CK5 (BioLegend, 904801), CK8 (BioLegend,904801). The tissue sections were counterstained, mounted, and imaged with a Zeiss microscope (White Plains, NY). Cells were collected in PBS and lysed in passive lysis buffer (Promega, Madison, WI).

Equal amounts of protein were loaded for Western blot analyses. ProSignal Dura ECL Reagent (Genesee Scientific, San Diego, CA) and Chemidoc (Bio-Rad, Hercules, CA) were used to visualize the proteins. Beta-tubulin was used as loading control.

Acknowledgement

We thank the Broad Institute for generating high-quality sequence data supported by NHGRI funds (grant # U54 HG003067) with Eric Lander as PI. We also thank Stand Up To Cancer-Prostate Cancer Foundation Prostate Dream Team Translational Cancer Research Grant for supporting generated high-quality data used in this manuscript. The datasets used in this manuscript were obtained from dbGaP at http://www.ncbi.nlm.nih.gov/gap through dbGaP accession number phs000909 and phs000915. We also thank Weill Cornell Medical College for the samples collection.

Figures

Fig.1

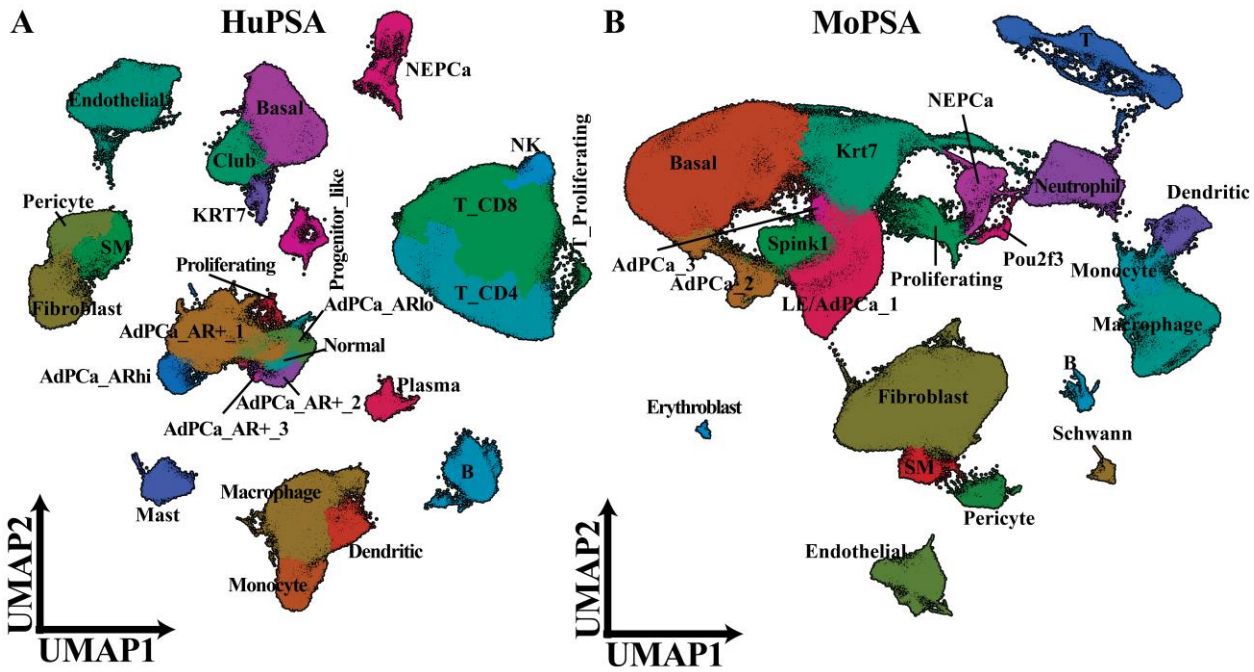


Fig. 1 UMAP dimensional reduction of HuPSA (A) and MoPSA (B). Defined cell populations are represented by different colors and text.

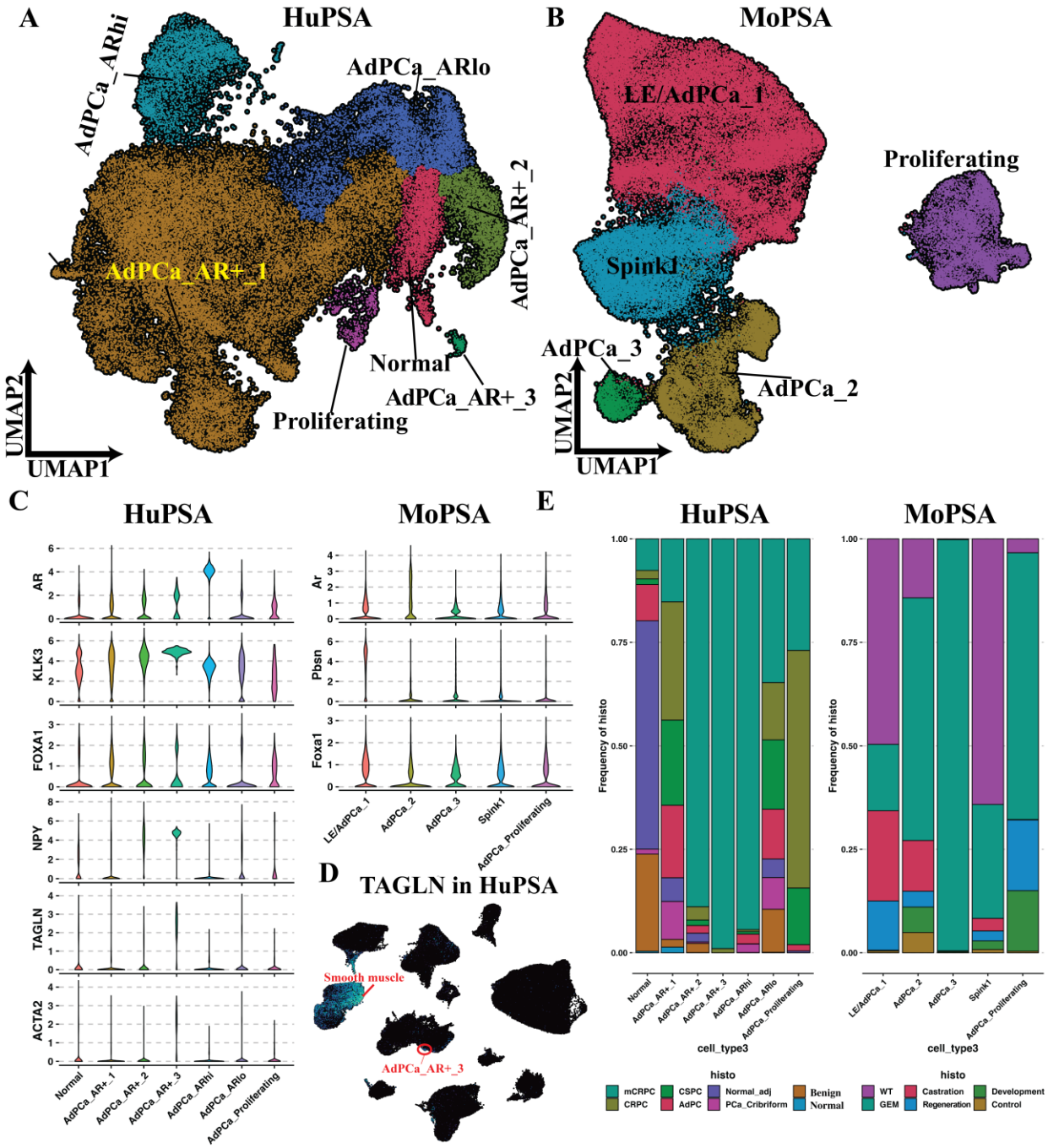


Fig. 2 Characterization of AdPCa populations. UMAP dimensional reduction of AdPCa cells from HuPSA (A) and MoPSA (B). Different cell populations within the AdPCa datasets are represented by distinct colors and text labels. Violin plots were utilized to quantify gene expression in different AdPCa subclusters in HuPSA (left) and MoPSA (right). (D) A Feature plot was employed to visualize the expression of the human TAGLN gene specifically in HuPSA. Cells with high expression of TAGLN, namely smooth muscle and AdPCa_AR+_3, are highlighted by text. (E) Bar plots were used to depict the composition of histology groups within the different AdPCa subclusters in HuPSA (left) and MoPSA (right).

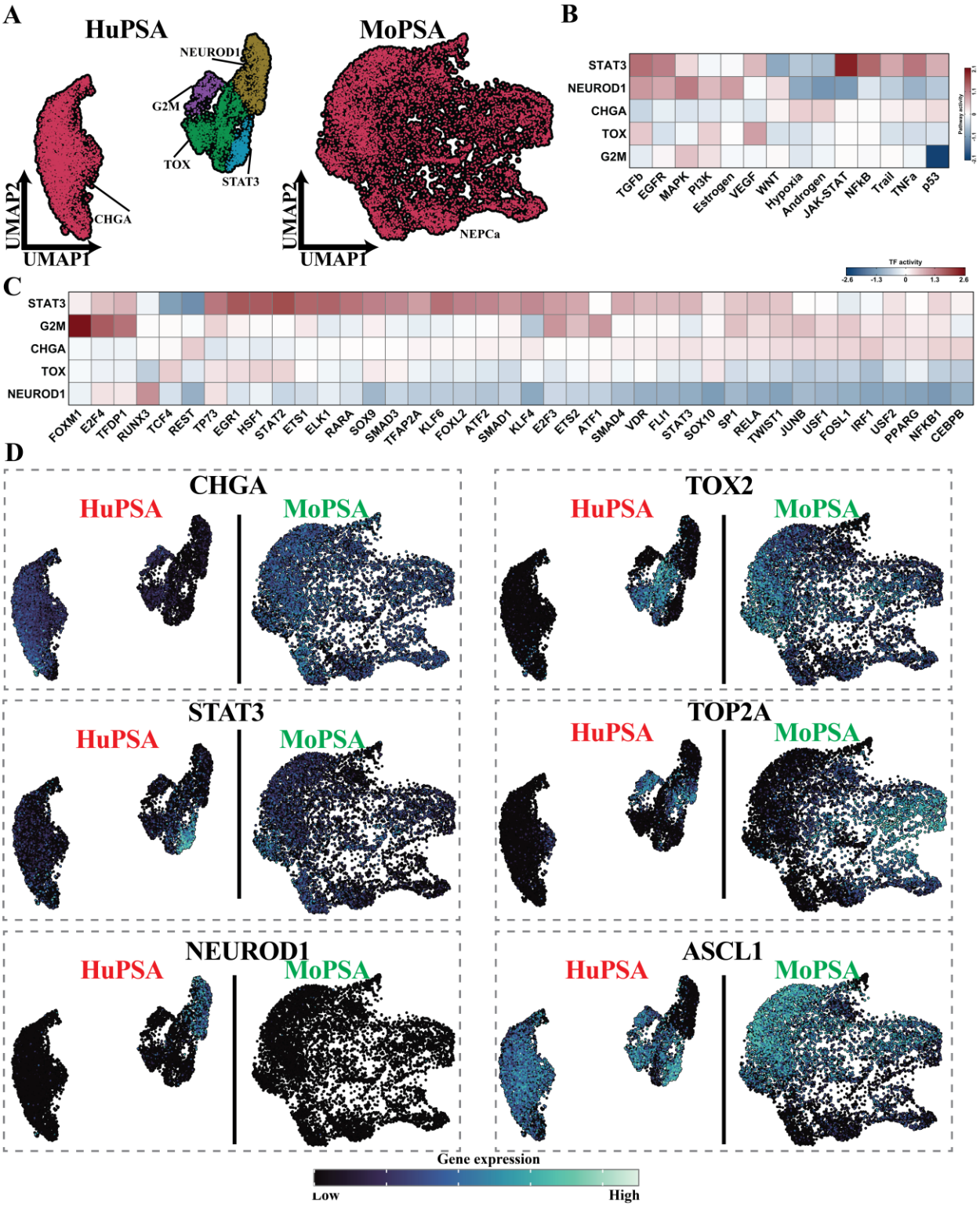


Fig. 3 Characterization of NEPCa populations. The UMAP plots revealed heterogenous NEPCa cells in HuPSA and homogenous NEPCa cells in MoPSA, indicating distinct cell population characteristics. Further analysis incorporating clustering and marker gene expression allowed for the subdivision of HuPSA NEPCa cells into 5 subpopulations (A). Heatmaps were employed to illustrate the predicted activity of signaling pathways (B) and upstream transcription factors (C) for the HuPSA NEPCa subclusters, providing insights into their regulatory networks. Additionally, Feature plots were used to visualize the expression of NEPCa subcluster marker genes in both HuPSA and MoPSA datasets (D).

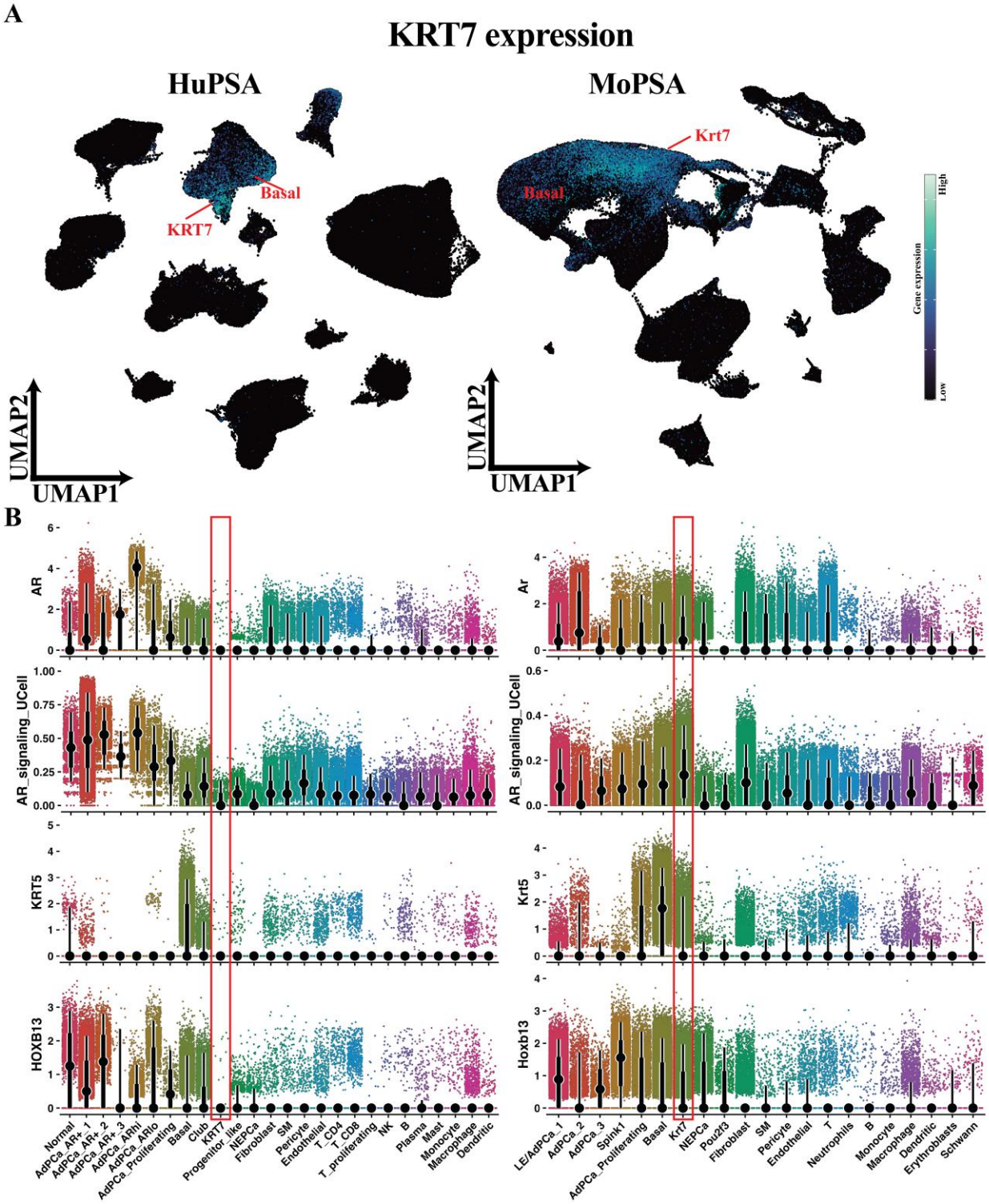


Fig. 4 KRT7 gene expression represents unique populations in HuPSA and MoPSA. Feature plots of KRT7 (Krt7 as mouse gene symbol) gene expression showed the KRT7 expressed in KRT7 cluster in HuPSA (left) and Krt7 cluster in MoPSA (right) and some basal epithelium cells in both datasets (A). Geyser plots (scatter plots) quantified genes expression in HuPSA (left) and MoPSA (right). The KRT7/ Krt7 populations were highlighted by red rectangle.

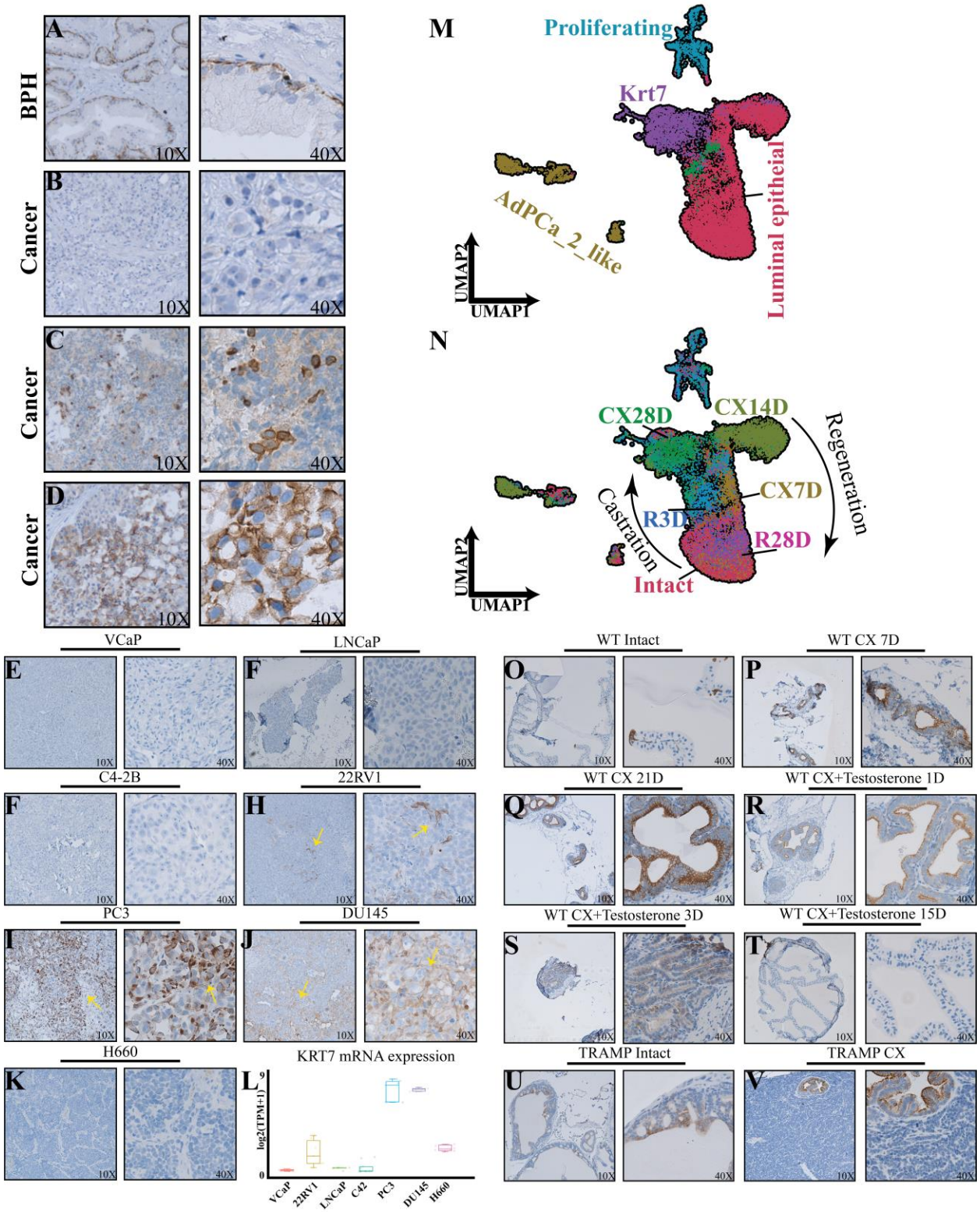


Fig. 5 Characterizing KRT7 populations in human and mouse. In human benign prostatic hyperplasia (BPH) tissue, Krt7 was highly expressed in basal epithelium cells but not in luminal epithelium cells (A). Prostate cancer cells showed negative Krt7 expression (B). In some cases, scattered populations (C) or major populations (D) exhibited high Krt7 expression. Advanced prostate cancer cell line xenograft tissues showed detectable Krt7 expression, with high expression in subpopulations of 22RV1 and PC3 cells, and moderate expression in the majority of DU145 cells (E-K). KRT7 mRNA expression in cultured cells

correlated roughly with the IHC staining, with high levels in PC3 and DU145 cells, and low levels in 22RV1 and H660 cells (F). The Krt7 and AdPCa populations from "GSE210358" study were isolated from MoPSA and visualized by UMAP as a dim plot (M). The same UMAP plot was labeled with treatment information consisting of intact, castration in 1-28 days and regeneration in 1-14days (N). In wild-type (WT) mouse prostate, Krt7 expression was detected in sub populations both basal and luminal epithelium cells. The normal Krt7+ luminal epithelium cells exhibited resistance to castration (CX). Over the treatment time scale, in intact mouse prostate, Krt7 expression was rarely detected in luminal epithelium cells. However, Krt7+ populations were enriched in CX 7 days and CX 21 days samples. After testosterone treatment, along with luminal epithelium regeneration, the percentage of Krt7+ cells decreased. In transgenic adenocarcinoma of the mouse prostate (TRAMP) tissues, intact mouse tissue showed Krt7 expression in a subpopulation of prostatic intraepithelial neoplasia (PIN) cells. After castration, Krt7 was detected in all PIN cells but not in neuroendocrine prostate cancer cells (O-V).

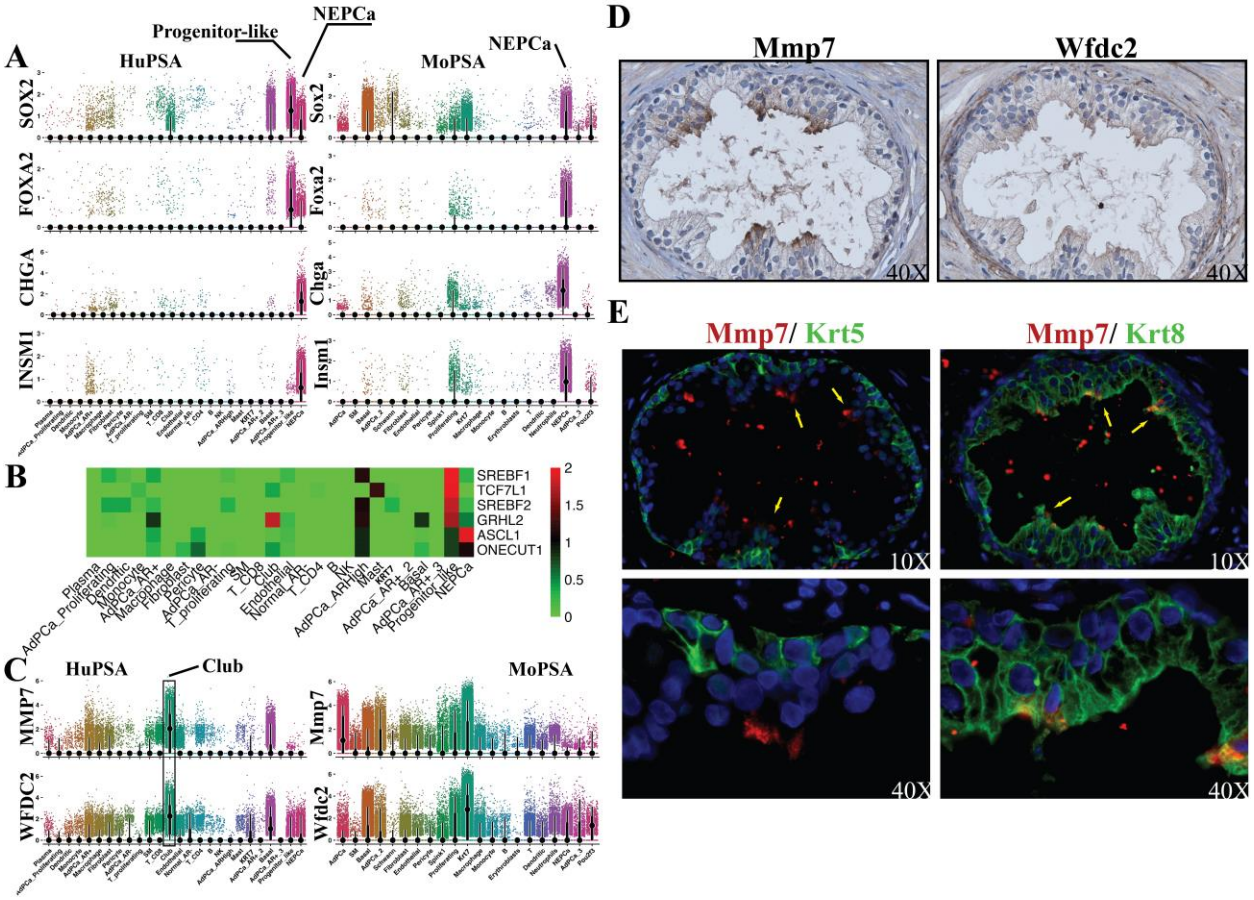


Fig. 6 Characterization of HuPSA unique novel populations: Geyser plots depict the expression of progenitor and neuroendocrine marker genes in HuPSA and MoPSA datasets (A). A heatmap reveals predicted upstream regulators associated with progenitor-like and NEPCa populations (B). Geyser plots demonstrate MMP7 and WFDC2 expression in HuPSA and MoPSA (C). Immunohistochemistry on serial sections confirms the presence of Mmp7 and Wfdc2 in specific subpopulations of luminal epithelium cells in human BPH samples (D), while immunofluorescence staining supports the colocalization of Mmp7 with luminal epithelium markers (E).

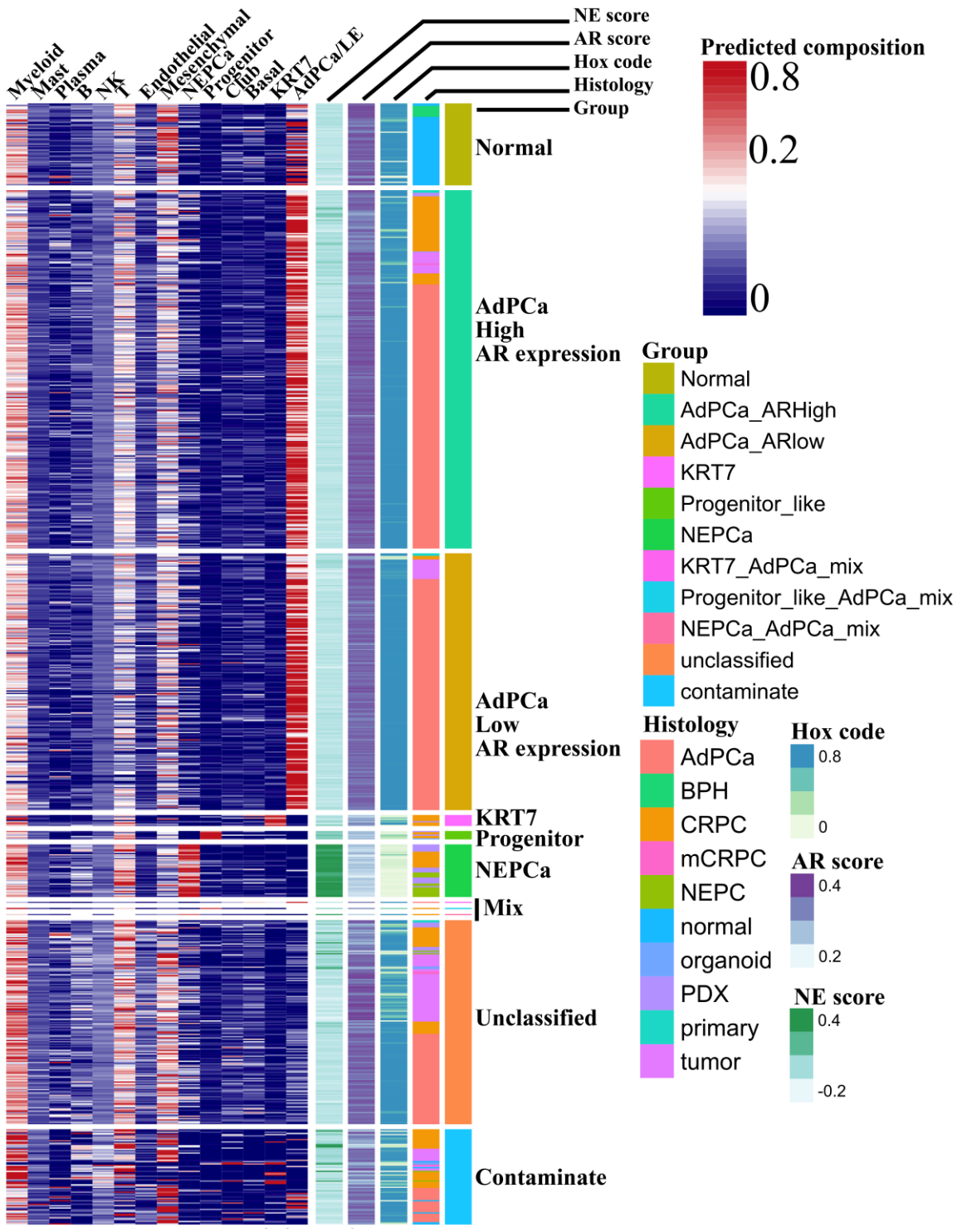
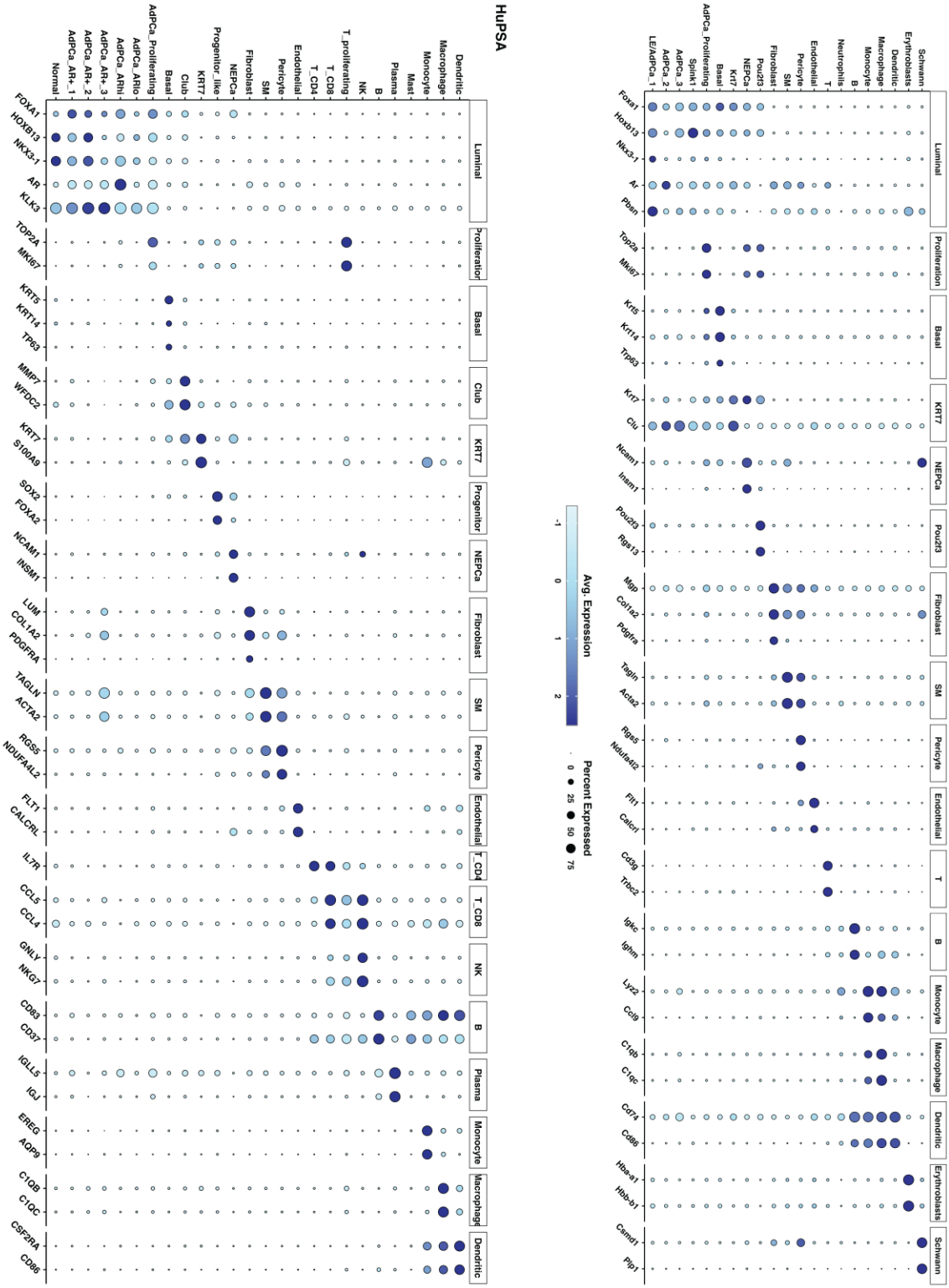
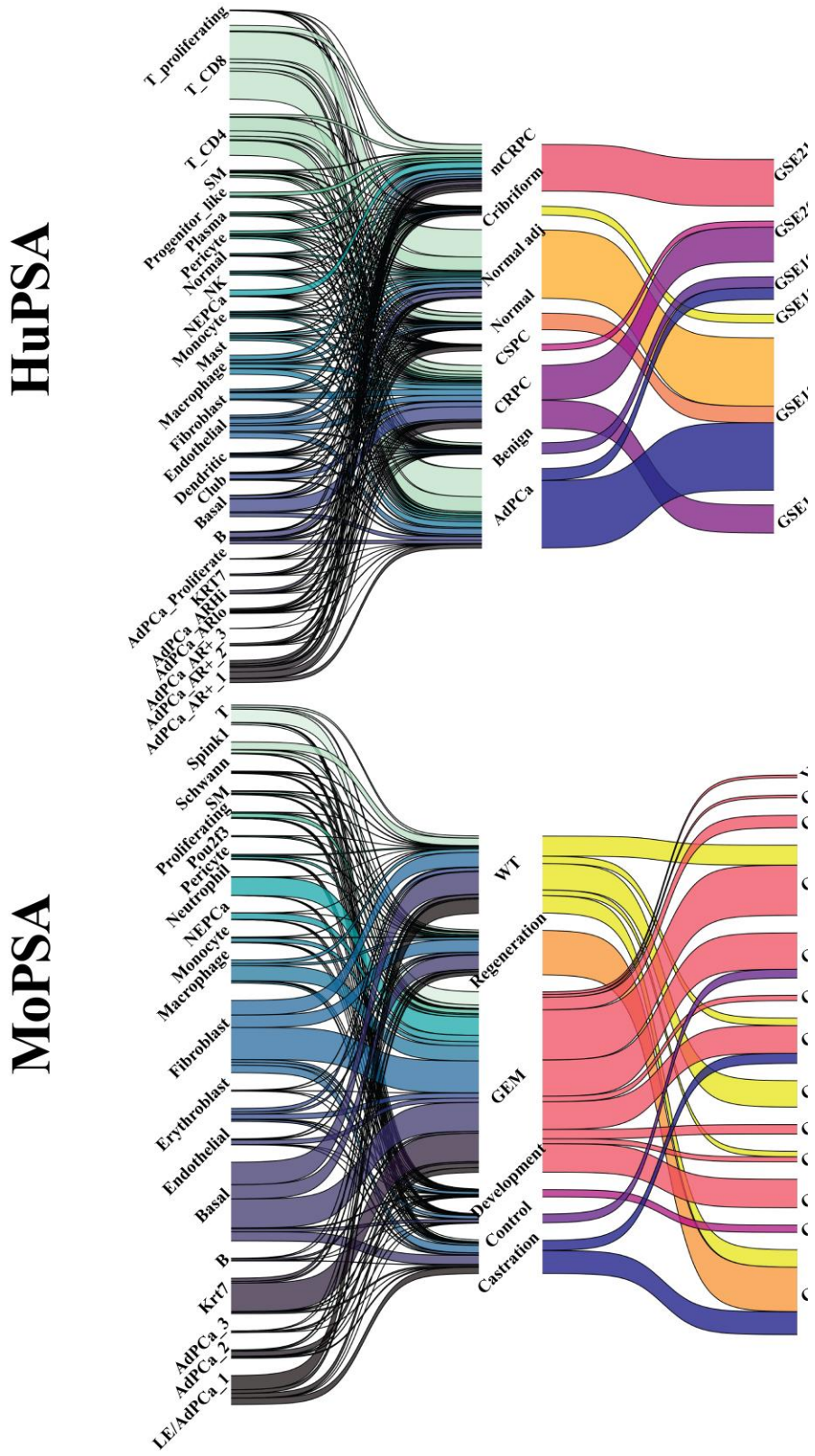


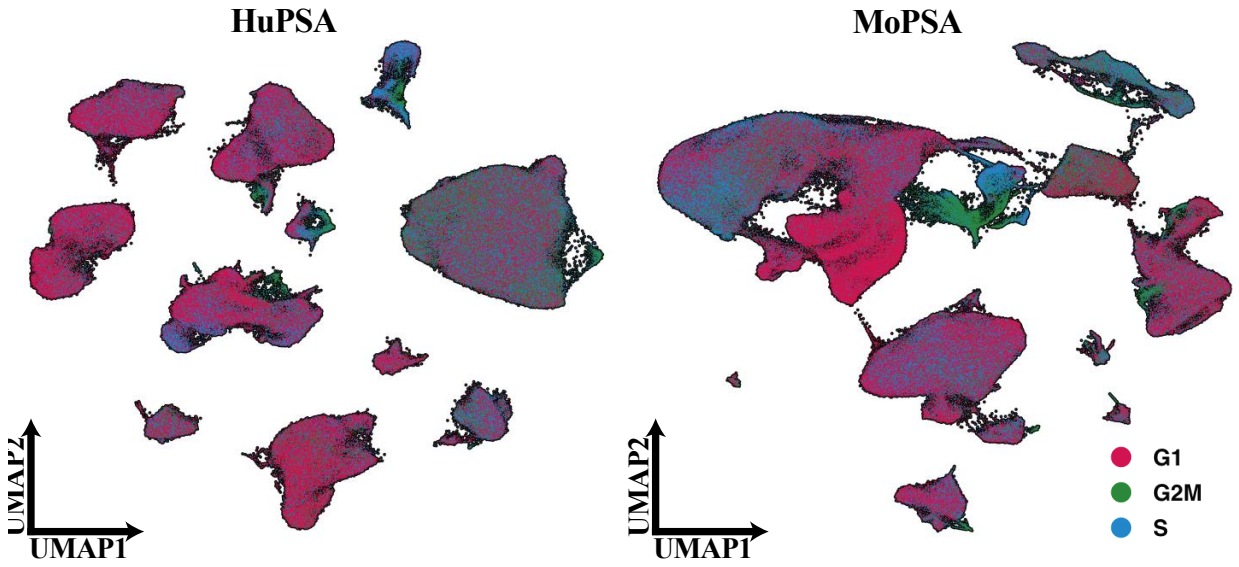
Fig. 7 Heatmap of HuPSA-based human PCa bulk RNAseq deconvolution results. Each column in the heatmap represents a cell population defined by HuPSA, while each row corresponds to a bulk RNA sequencing sample. The color in the heatmap indicates the estimated percentage of each cell population within each sample. Additionally, the samples are annotated with NE score, AR score, (prostatic) HOX code, histology information, and classified group. The samples are categorized into different groups based on the deconvolution results and labeled accordingly. The 'Mix' group contains samples with multiple dominant cancer populations. The 'Unclassified' group comprises samples that did not meet all the criteria for the other groups. The 'Contaminate' group includes samples with a dominant presence of stromal or immune populations.



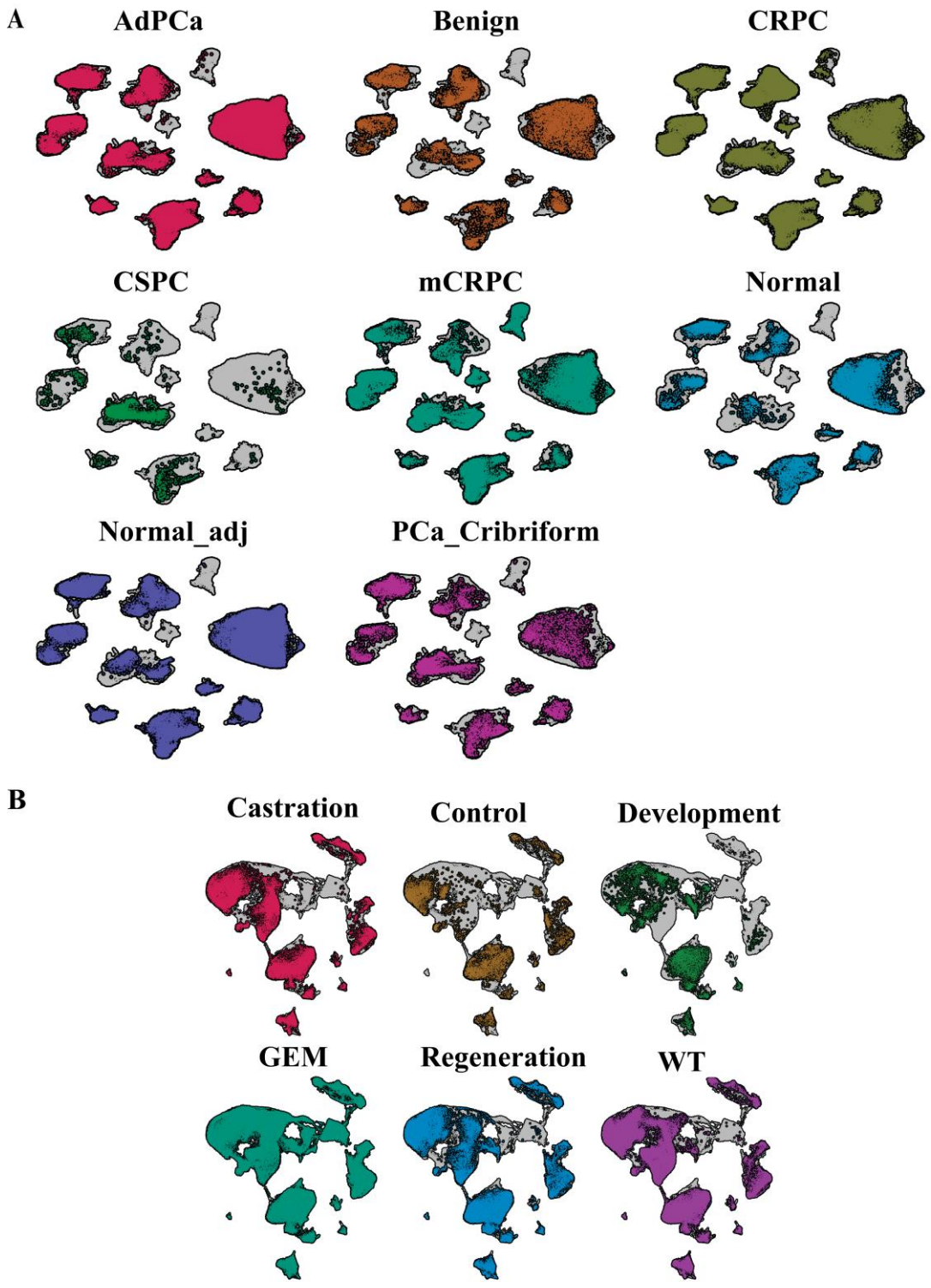
sFig. 1 Dot plots visualizing marker gene expression in different populations in HuPSA (left) and MoPSA (right). Each dot represents a gene, and its color represents the gene expression level, ranging from low (light blue) to high (dark blue). The size of the dot corresponds to the percentage of cells expressing that gene within each population.



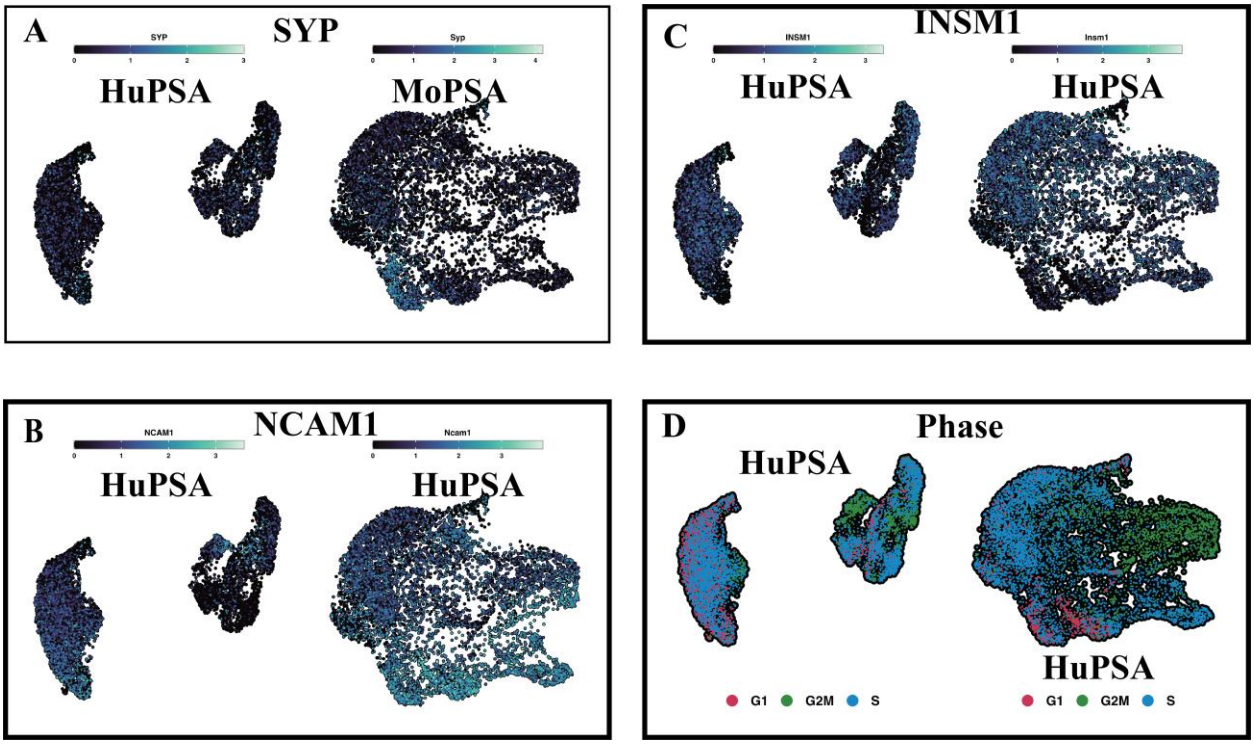
sFig. 2 Sankey plots of HuPSA (A) and MoPSA (B) depicting defined cell populations (left), histology groups (middle), and studies (right).



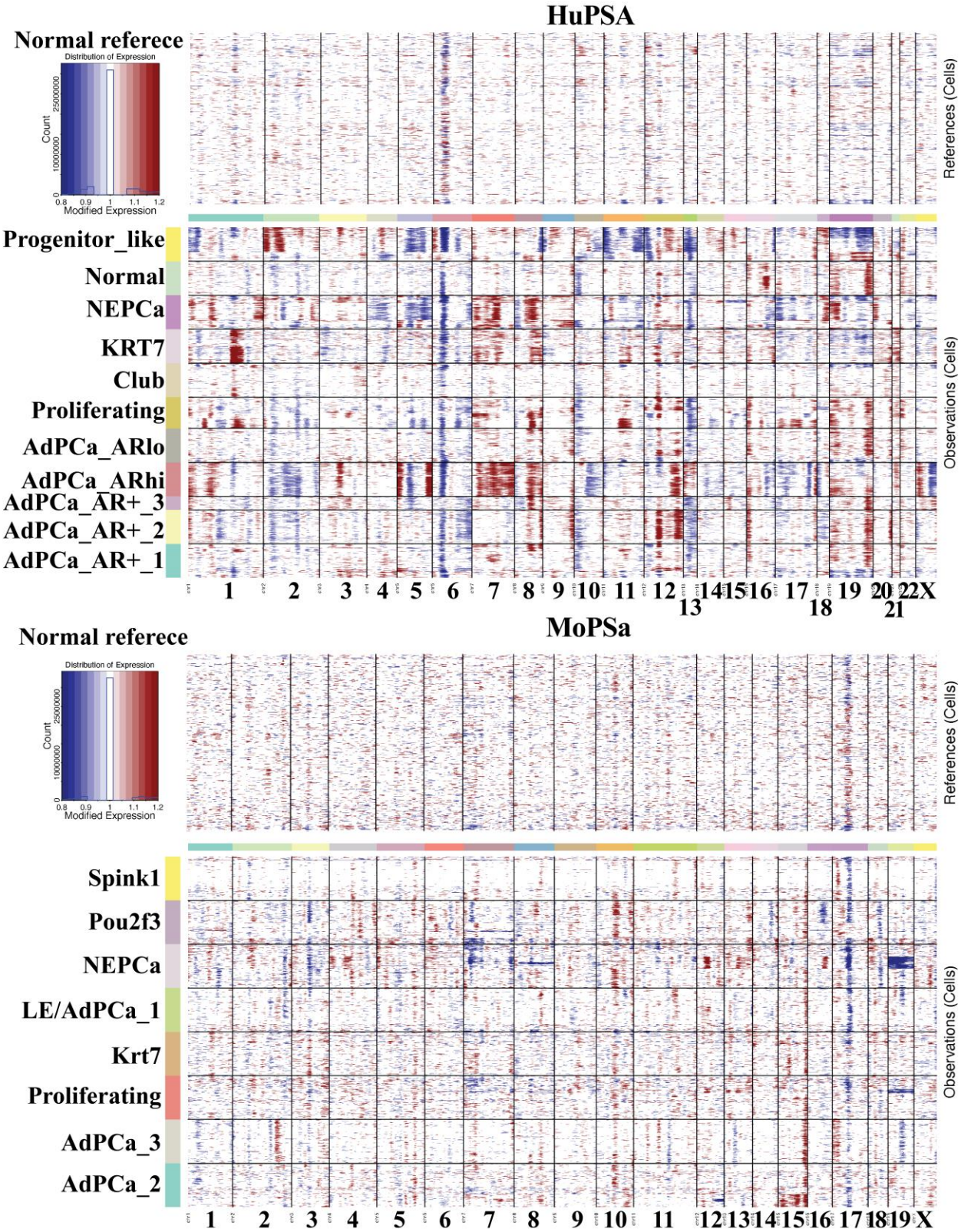
sFig. 3 UMAP of HuPSA and MoPSA showing predicted cell phases. G1 phase in red, G2M phase in green, and S phase in blue.



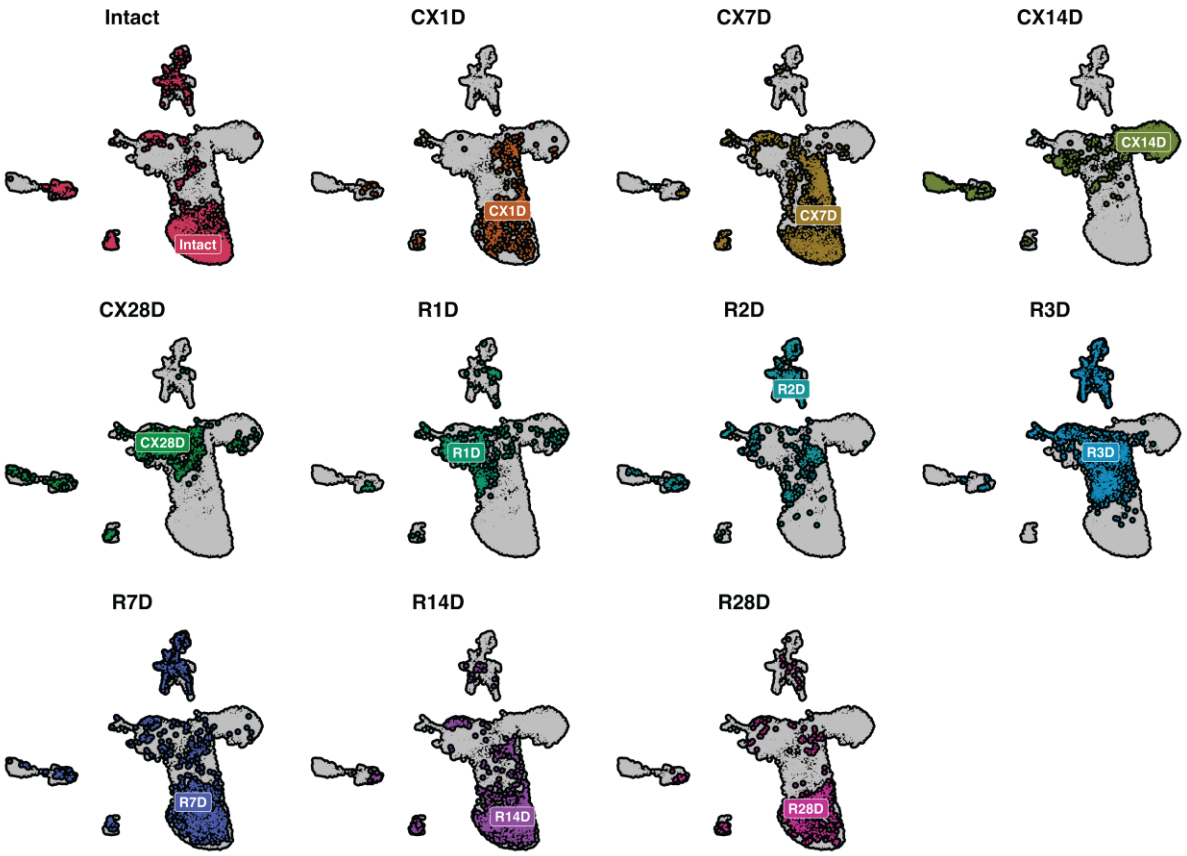
sFig. 4 UMAP of HuPSA (A) and MoPSA (B) spited by histology information. Color represents different histology groups defined by original publications.



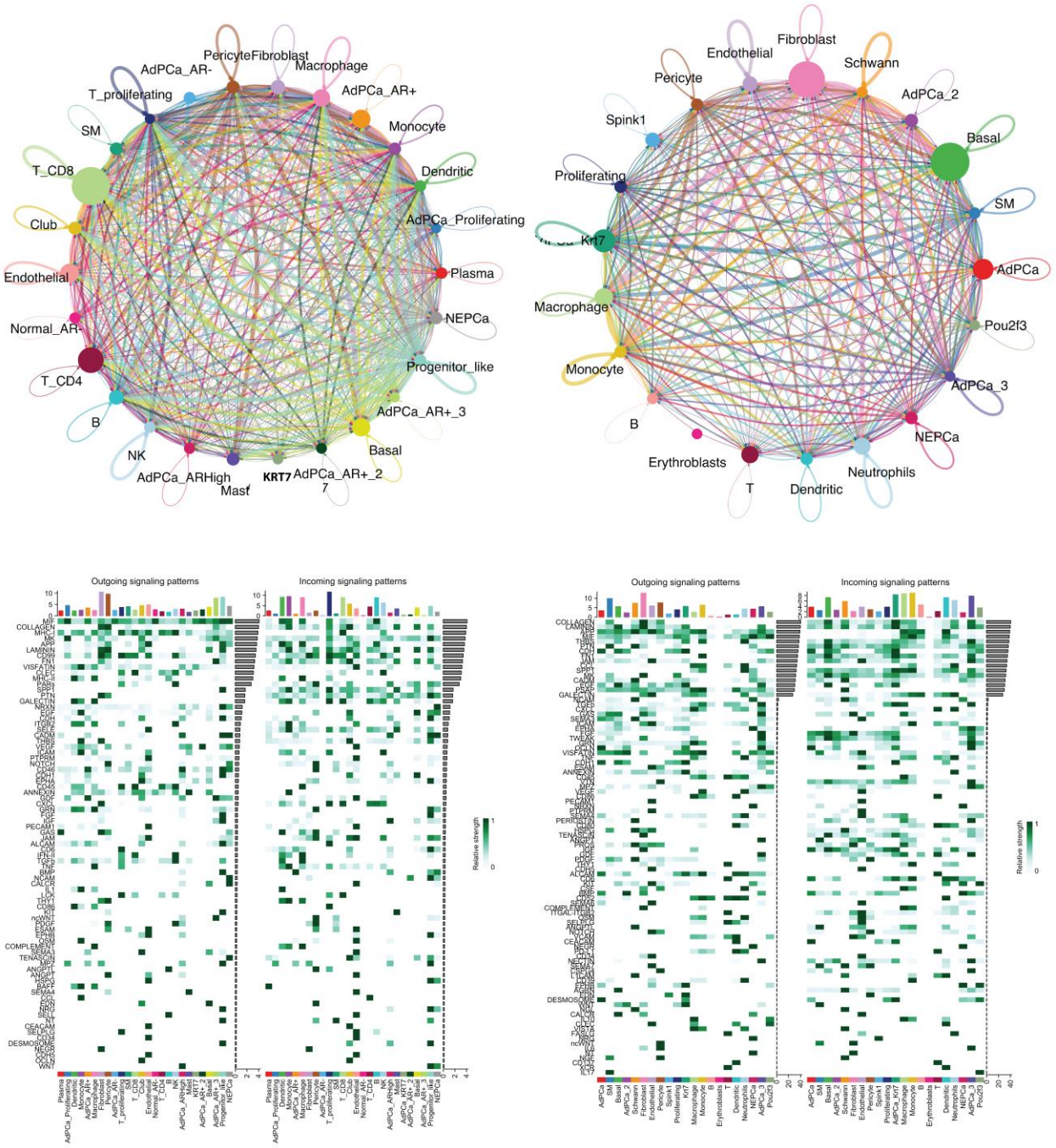
sFig. 5 The HuPSA NEPCa sub-populations expressed NE marker genes. The expression of NE marker genes SYP (A), NCAM1 (B) and INSM1 (C) from HuPSA and MoPSa NEPCa cells were visualized by feature plots. The predicted cell cycle phase of NEPCa cells were visualized by Dim plots (D).



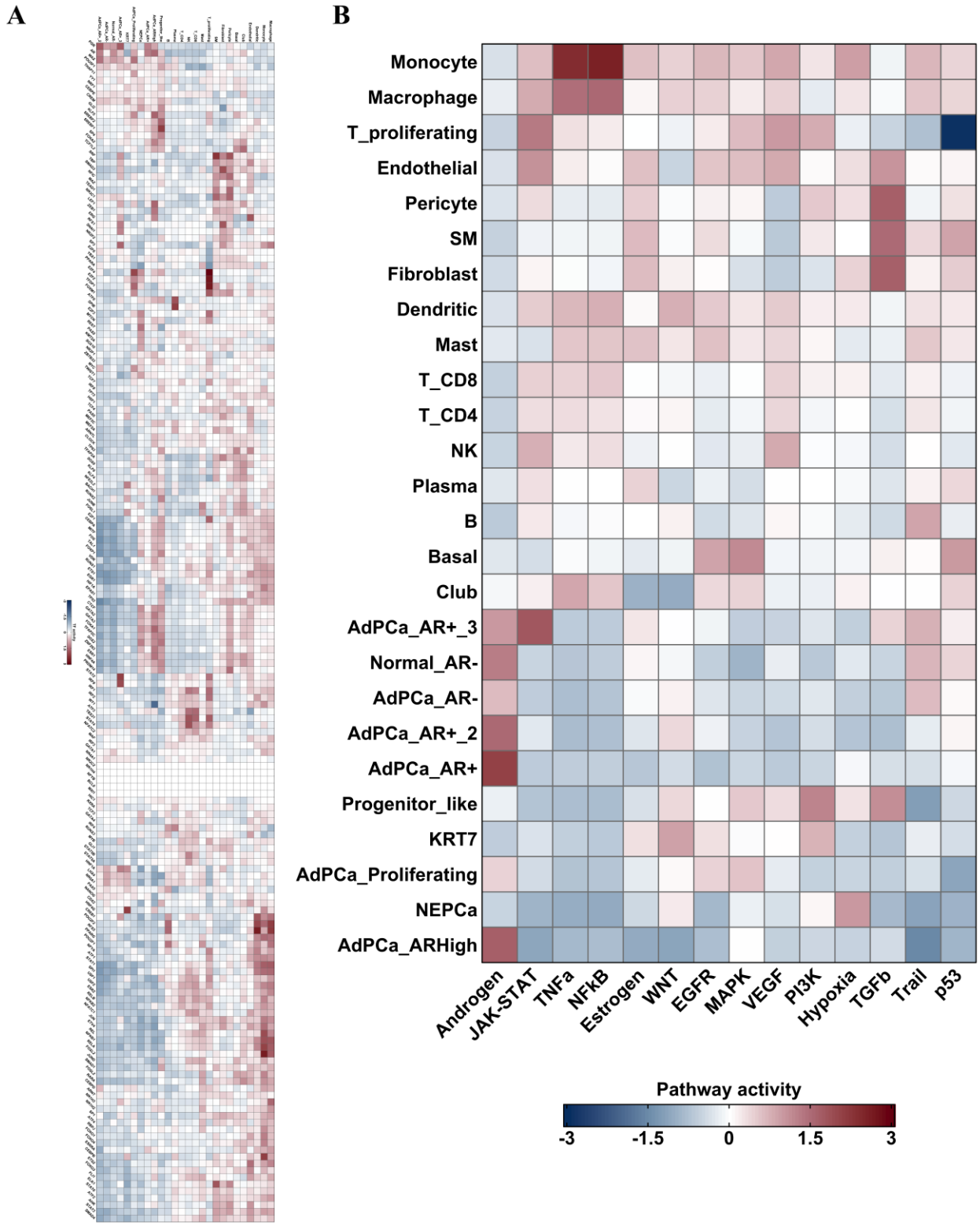
sFig. 6 InferCNV results for both HuPSA (up) and MoPSA (down). Heatmaps showed the predicted copy number alterations of selected cell populations compared to normal references (immune and other stromal cell populations).



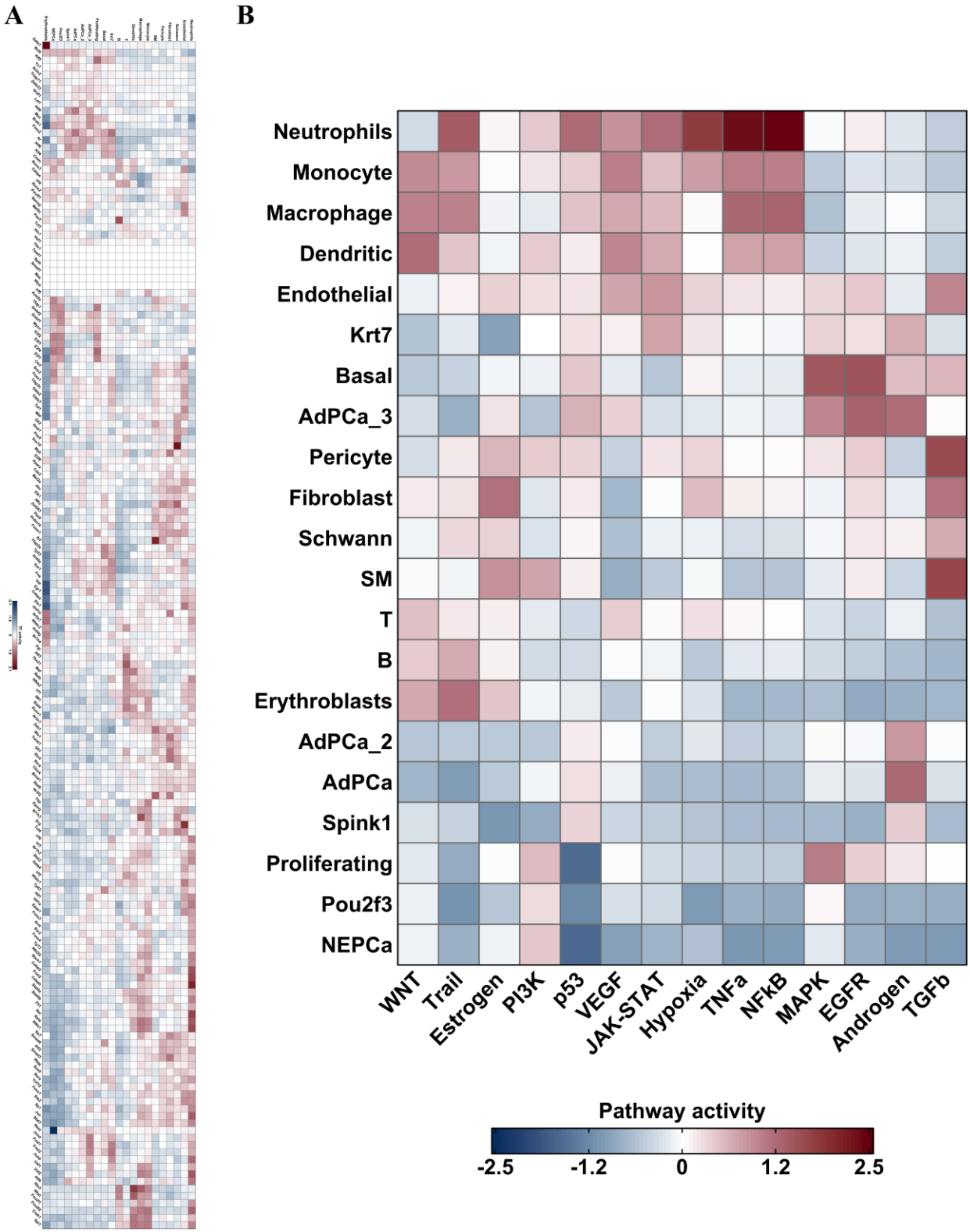
sFig. 7 Dim plot of the MoPSA AdPCa and Krt7 populations from castration and regeneration study "GSE146811" labeled by treatment information. This is the split view or Figure5N.



sFig. 8 Ligand-receptor analysis on HuPSA and MoPSA. Aggregated cell-cell communication networks were visualized by circle plots for both HuPSA (A) and MoPSA (B). The significant outgoing and incoming signalings were visualized by heatmap for HuPSA (A) and MoPSA (B).



sFig. 9 Upstream analysis for HuPSA. Heatmaps were used to visualize the calculated upstream transcription factor activity (A) and signaling pathway analysis for HuPSA.



sFig. 10 Upstream analysis for MoPSA. Heatmaps were used to visualize the calculated upstream transcription factor activity (A) and signaling pathway analysis for MoPSA.

References

1 Society, A. C. *Key Statistics for Prostate Cancer*, <https://www.cancer.org/cancer/prostate-cancer/about/key-statistics.html> (2022).

2 Yeh, Y. *et al.* Wnt/Beta-catenin signaling and prostate cancer therapy resistance. *Prostate Cancer*, 351-378 (2019).

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

4 Cheng, S. *et al.* Neuroendocrine prostate cancer has distinctive, non-prostatic HOX code that is represented by the loss of HOXB13 expression. *Scientific reports* **11**, 1-11 (2021).

5 Beltran, H. *et al.* Aggressive variants of castration-resistant prostate cancer. *Clinical Cancer Research* **20**, 2846-2850 (2014).

6 Bluemn, E. G. *et al.* Androgen receptor pathway-independent prostate cancer is sustained through FGF signaling. *Cancer cell* **32**, 474-489 (2017).

7 Chan, J. M. *et al.* Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science*, eabn0478 (2022).

8 Ferguson, A. M. & Rubin, M. A. Lineage plasticity in prostate cancer: Looking beyond intrinsic alterations. *Cancer letters* **548**, 215901 (2022).

9 Mu, P. *et al.* SOX2 promotes lineage plasticity and antiandrogen resistance in TP53-and RB1-deficient prostate cancer. *Science* **355**, 84-88 (2017).

10 Zhang, X. *et al.* SRRM4 expression and the loss of REST activity may promote the emergence of the neuroendocrine phenotype in castration-resistant prostate cancer. *Clinical Cancer Research* **21**, 4698-4708 (2015).

11 Rotinen, M. *et al.* ONECUT2 is a targetable master regulator of lethal prostate cancer that suppresses the androgen axis. *Nature medicine* **24**, 1887-1898 (2018).

12 Bishop, J. L. *et al.* The master neural transcription factor BRN2 is an androgen receptor–suppressed driver of neuroendocrine differentiation in prostate cancer. *Cancer discovery* **7**, 54-71 (2017).

13 Akamatsu, S. *et al.* The placental gene PEG10 promotes progression of neuroendocrine prostate cancer. *Cell reports* **12**, 922-936 (2015).

14 Dong, B. *et al.* Single-cell analysis supports a luminal-neuroendocrine transdifferentiation in human prostate cancer. *Communications biology* **3**, 778 (2020).

15 Heidegger, I. *et al.* Comprehensive characterization of the prostate tumor microenvironment identifies CXCR4/CXCL12 crosstalk as a novel antiangiogenic therapeutic target in prostate cancer. *Molecular Cancer* **21**, 132 (2022).

16 Hirz, T. *et al.* Dissecting the immune suppressive human prostate tumor microenvironment via integrated single-cell and spatial transcriptomic analyses. *Nature Communications* **14**, 663 (2023).

17 Wong, H. Y. *et al.* Single cell analysis of cribriform prostate cancer reveals cell intrinsic and tumor microenvironmental pathways of aggressive disease. *Nature Communications* **13**, 6036 (2022).

18 Kim, W. K. *et al.* Aberrant androgen action in prostatic progenitor cells induces oncogenesis and tumor development through IGF1 and Wnt axes. *Nature communications* **13**, 1-18 (2022).

19 Terzic, J. *et al.* Hypoxia-inducible factor 1A inhibition overcomes castration resistance of prostate tumors. *EMBO Molecular Medicine*, e17209 (2023).

20 Zhu, Y. *et al.* Targeting the chromatin effector Pygo2 promotes cytotoxic T cell responses and overcomes immunotherapy resistance in prostate cancer. *Science Immunology* **8**, eade4656 (2023).

21 Germanos, A. A. *et al.* Defining cellular population dynamics at single-cell resolution during prostate cancer progression. *Elife* **11**, e79076 (2022).
22 Graham, M. K. *et al.* Single-cell atlas of epithelial and stromal cell heterogeneity by lobe and strain in the mouse prostate. *The Prostate* **83**, 286-303 (2023).
23 Hiroto, A. *et al.* Stromal androgen signaling acts as tumor niches to drive prostatic basal epithelial progenitor-initiated oncogenesis. *Nature Communications* **13**, 6552 (2022).
24 Qiu, X. *et al.* MYC drives aggressive prostate cancer by disrupting transcriptional pause release at androgen receptor targets. *Nature communications* **13**, 2559 (2022).
25 Troiani, M. *et al.* Single-cell transcriptomics identifies Mcl-1 as a target for senolytic therapy in cancer. *Nature communications* **13**, 2177 (2022).
26 Brady, N. J. *et al.* Temporal evolution of cellular heterogeneity during the progression to advanced AR-negative prostate cancer. *Nature Communications* **12**, 3372 (2021).
27 Lee, D.-H. *et al.* Androgen action in cell fate and communication during prostate development at single-cell resolution. *Development* **148**, dev196048 (2021).
28 Song, H. *et al.* Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nature communications* **13**, 141 (2022).
29 Karthaus, W. R. *et al.* Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science* **368**, 497-505 (2020).
30 Grabowska, M. M. *et al.* Mouse models of prostate cancer: picking the best model for the question. *Cancer and Metastasis Reviews* **33**, 377-397 (2014).
31 Henry, G. H. *et al.* A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell reports* **25**, 3530-3542 (2018).
32 Liao, H.-Y., Da, C.-M., Liao, B. & Zhang, H.-H. Roles of matrix metalloproteinase-7 (MMP-7) in cancer. *Clinical biochemistry* **92**, 9-18 (2021).
33 Lamy, P.-J., Plassot, C. & Pujol, J.-L. Serum HE4: an independent prognostic factor in non-small cell lung cancer. *PloS one* **10**, e0128836 (2015).
34 Nouruzi, S. *et al.* ASCL1 activates neuronal stem cell-like lineage programming through remodeling of the chromatin landscape in prostate cancer. *Nature Communications* **13**, 2282 (2022).
35 Kuwabara, T. *et al.* Wnt-mediated activation of NeuroD1 and retro-elements during adult neurogenesis. *Nature neuroscience* **12**, 1097-1105 (2009).
36 Karimnezhad, A. More accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Bioinformatics Advances* **2**, vbac049 (2022).
37 Cheng, S. & Yu, X. CTPC, a combined transcriptome dataset of prostate cancer cell lines. *bioRxiv* (2022).
38 Hanahan, D. Hallmarks of cancer: new dimensions. *Cancer discovery* **12**, 31-46 (2022).
39 Li, D. *et al.* ETV4 mediates dosage-dependent prostate tumor initiation and cooperates with p53 loss to generate prostate cancer. *Science Advances* **9**, eadc9446 (2023).
40 He, D. *et al.* Alevin-fry unlocks rapid, accurate and memory-frugal quantification of single-cell RNA-seq data. *Nature Methods* **19**, 316-322 (2022).
41 Zhu, A., Srivastava, A., Ibrahim, J. G., Patro, R. & Love, M. I. Nonparametric expression analysis using inferential replicate counts. *Nucleic Acids Research* **47**, e105-e105 (2019).
42 Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nature Biotechnology*, 1-12 (2023).
43 Bakken, T. E. *et al.* Comparative cellular analysis of motor cortex in human, marmoset and mouse. *Nature* **598**, 111-119 (2021).
44 Blanco-Carmona, E. Generating publication ready visualizations for single cell transcriptomics using SCpubr. *bioRxiv*, 2022-2002 (2022).

45 Puram, S. V. *et al.* Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611-1624 (2017).
46 Badia-i-Mompel, P. *et al.* decoupleR: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances* **2**, vbac016 (2022).
47 Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nature communications* **12**, 1-20 (2021).
48 Suntsova, M. *et al.* Atlas of RNA sequencing profiles for normal human tissues. *Scientific data* **6**, 36 (2019).
49 Labrecque, M. P. *et al.* Cabozantinib can block growth of neuroendocrine prostate cancer patient-derived xenografts by disrupting tumor vasculature. *Plos one* **16**, e0245602 (2021).
50 Cejas, P. *et al.* Subtype heterogeneity and epigenetic convergence in neuroendocrine prostate cancer. *Nature communications* **12**, 5775 (2021).
51 Luo, G. *et al.* circITGA7 Acts as a miR-370-3p sponge to suppress the proliferation of prostate cancer. *Journal of Oncology* **2021** (2021).
52 Puca, L. *et al.* Patient derived organoids to model rare prostate cancer phenotypes. *Nature communications* **9**, 2404 (2018).
53 Teslow, E. A. *et al.* Exogenous IL-6 induces mRNA splice variant MBD2_v2 to promote stemness in TP53 wild-type, African American PCa cells. *Molecular oncology* **12**, 1138-1152 (2018).
54 Sowalsky, A. G. *et al.* Whole transcriptome sequencing reveals extensive unspliced mRNA in metastatic castration-resistant prostate cancer. *Molecular Cancer Research* **13**, 98-106 (2015).
55 Long, Q. *et al.* Global transcriptome analysis of formalin-fixed prostate cancer specimens identifies biomarkers of disease recurrence. *Cancer research* **74**, 3228-3237 (2014).
56 Kannan, K. *et al.* Recurrent chimeric RNAs enriched in human prostate cancer identified by deep sequencing. *Proceedings of the National Academy of Sciences* **108**, 9172-9177 (2011).
57 Acosta-Vega, N. L. *et al.* Metabolic pathways enriched according to ERG status are associated with biochemical recurrence in Hispanic/Latino patients with prostate cancer. *Cancer Medicine* **12**, 4306-4320 (2023).
58 Beltran, H. *et al.* Molecular characterization of neuroendocrine prostate cancer and identification of new drug targets. *Cancer discovery* **1**, 487-495 (2011).
59 Wilks, C. *et al.* recount3: summaries and queries for large-scale RNA-seq expression and splicing. *Genome biology* **22**, 1-40 (2021).
60 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013).
61 Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nature methods* **14**, 417-419 (2017).
62 Cheng, S. *et al.* The expression of YAP1 is increased in high-grade prostatic adenocarcinoma but is reduced in neuroendocrine prostate cancer. *Prostate cancer and prostatic diseases* **23**, 661-669 (2020).