

# Single cell and spatial transcriptomics highlight the interaction of club-like cells with immunosuppressive myeloid cells in prostate cancer

Received: 3 June 2024

Accepted: 8 November 2024

Published online: 16 November 2024

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Prostate cancer treatment resistance is a significant challenge facing the field. Genomic and transcriptomic profiling have partially elucidated the mechanisms through which cancer cells escape treatment, but their relation toward the tumor microenvironment (TME) remains elusive. Here we present a comprehensive transcriptomic landscape of the prostate TME at multiple points in the standard treatment timeline employing single-cell RNA-sequencing and spatial transcriptomics data from 120 patients. We identify club-like cells as a key epithelial cell subtype that acts as an interface between the prostate and the immune system. Tissue areas enriched with club-like cells have depleted androgen signaling and upregulated expression of luminal progenitor cell markers. Club-like cells display a senescence-associated secretory phenotype and their presence is linked to increased polymorphonuclear myeloid-derived suppressor cell (PMN-MDSC) activity. Our results indicate that club-like cells are associated with myeloid inflammation previously linked to androgen deprivation therapy resistance, providing a rationale for their therapeutic targeting.

Approximately one in eight men develop prostate cancer (PCa) during their lifetime. A favorable prognosis for localized tumors is contrasted by the high mortality rate and challenging treatment landscape of metastatic disease<sup>1</sup>. Genomics and transcriptomics data from clinical samples have revealed the various aberrations underlying primary<sup>2,3</sup> and treatment-resistant tumors<sup>4,5</sup>. Single-cell RNA-sequencing (scRNA-seq) studies have elucidated the heterogeneity inside these tumors, demonstrating that bulk-level investigations fail to address their whole complexity<sup>6–9</sup>. While powerful, scRNA-seq is hampered by a lack of spatial information making it a suboptimal tool for observing interactions in the tumor microenvironment (TME). A handful of studies have investigated the prostate TME in a spatial context, reporting insights such as transcriptomic differences between Gleason grades<sup>10</sup>, malignant cells in histologically benign areas<sup>11</sup>, and immunosuppressive regulatory T-cell activity in

coordination with myeloid cells<sup>12</sup>, but have included only few patients. The tumor cells' interaction with the TME is a question of high interest, as mounting evidence suggests that the TME actively participates in the progression, treatment response, and metastasis of various tumors<sup>13–15</sup>.

The scRNA-seq-based characterization of PCa has led to an appreciation of epithelial cell type heterogeneity, identifying four specialized subtypes: luminal, basal, hillock, and club cells<sup>16,17</sup>. Basal and luminal epithelium are well-described parts of the normal prostate glandular structure. The luminal epithelium is regarded as the cellular origin of prostate adenocarcinomas<sup>18,19</sup>. Hillock cells are a source of discrepancy as they have been claimed to be found<sup>9,12,16,20</sup> or to be undiscernible<sup>17</sup>, depending on the study. In some cases, hillock and club cells have been described together as *intermediate* epithelial cells, reflecting the similarity of these cell populations<sup>7,21</sup>.

✉ e-mail: [alfonso.urbanucci@tuni.fi](mailto:alfonso.urbanucci@tuni.fi); [matti.nykter@tuni.fi](mailto:matti.nykter@tuni.fi)

The role of club cells in the prostate is poorly understood. Club cells were originally described in the lung and characterized for their expression of uteroglobin *SCGB1A1*<sup>16,22</sup>. Initial observations noted that club cells reside in the urethral epithelium and proximal prostate ducts<sup>20,23</sup>. In contrast, more recent work has highlighted their presence in the peripheral zone and inflamed regions of the prostate<sup>24</sup>. Intriguingly, human club cells have a transcriptomic profile similar to the castration resistant luminal progenitor phenotype in mice<sup>25–27</sup>, raising questions regarding their function in response to androgen deprivation therapy (ADT) in humans.

Here we investigate the cellular composition of the prostate TME and its response to androgen deprivation. We generate spatial transcriptomics (ST) data from benign prostatic hyperplasia, treatment-naïve primary PCa, neoadjuvant-treated PCa, and castration-resistant PCa, representing a cross-section of cancer and benign prostate disease states together with TME from tissue adjacent to diseased areas. We augment these ST data with publicly available scRNA-seq data, analyzing high-resolution transcriptomics data from 120 patient samples in total. Our study provides an exhaustive map of the prostate TME, ranging from benign tissue to treatment-naïve tumors and advanced, treatment-resistant disease.

## Results

We used the Visium Spatial Transcriptomics assay to generate data from 80 fresh frozen tissue sections from 56 prostatectomy samples. Processed samples contained 1835 tissue-covered spots, 3275 median genes per spot, and 9549 median UMIs per spot on average (Supplementary Data S1). These samples were divided into two groups, constituting a discovery (48 sections from 48 patients) and a validation cohort (32 sections from 8 patients, Fig. 1a). We used the discovery cohort, consisting of treatment-naïve benign prostatic hyperplasia (BPH) and prostate cancer (TRNA), neoadjuvant-treated prostate cancer (NEADT), and castration-resistant prostate cancer tumors (CRPC) to develop a computational pipeline for unsupervised exploration of the prostate TME. We used the validation cohort of treatment-naïve primary tumors to cross-compare results and test hypotheses arising from the discovery cohort.

### Sample-independent deconvolution of spatial transcriptomics data with a single cell-derived cell state reference

The ST assay captures polyadenylated mRNAs onto 55  $\mu\text{m}$  sized uniquely barcoded spots. The number of individual cells and cell types underlying this measured expression varies, presenting a major challenge in the analysis of these data<sup>28,29</sup>. A commonly adopted analysis strategy uses scRNA-seq-derived transcriptomics profiles to infer the most probable cell type compositions in each ST spot<sup>30</sup>. These inferred cell type counts can further be used to divide tissue structures into distinct, biologically meaningful regions<sup>31</sup>.

We re-analyzed previously published scRNA-seq datasets from normal prostate tissue<sup>12,32</sup>, low- and high-grade primary tumors<sup>7,9,12,17</sup>, and both locally recurrent and metastatic castration-resistant prostate cancer tumors<sup>6,8</sup> to assemble a highly diverse cell-state reference likely to capture a similar heterogeneous cell type distribution as the discovery cohort. Using unsupervised clustering and non-negative matrix factorization (NMF) to define tissue compositions across samples (Supplementary Methods), we identified 26 cell states in 223,881 cells across 98 samples and 64 patients (Supplementary Data S2). We used this cell state reference to model the most probable cell state composition in spatial locations of our discovery cohort<sup>31</sup>.

Notable redundancy was still present in the inferred cell state counts, as thirteen cell states with the lowest inferred abundance accounted for just 6.7% of the total inferred cell counts (Supplementary Data S3). To address this, we categorized individual spots into eight categories based on their majority cell state contributors (Methods).

This resulted in commonly co-localized cell populations, hereon called single-cell mapping-derived regions (SCM-regions, Fig. 1b).

### Single-cell mapping-derived regions capture well-established biology

We identified eight SCM regions across the discovery cohort. For each sample, we compared the expression between regions to find differentially expressed genes (One-vs-rest two-sided Wilcoxon rank-sum test, Supplementary Data S4). We then determined genes that were enriched among the overexpressed genes in each region ( $p_{\text{adj}} < 0.05$ , one-sided Fisher's exact test), terming these *region-specific markers* (Supplementary Data S5).

We annotated the SCM regions according to their region-specific markers (Fig. 1c). The *Tumor* region overexpressed known PCa-specific markers such as *AMACR*<sup>33</sup>, *PCA3*<sup>34</sup>, and *PCAT14*<sup>35</sup>. Canonical luminal cell markers *MSMB*, *ACPP*, and *CD38* were markers for the *Luminal* region, while *KRT5*, *KRT15*, and *TP63* were overexpressed in the *Basal* region. The *Immune* region overexpressed chemokine receptors (*CXCR3*, *CXCR4*), T cell- (*TRBC1*, *TRBC2*, *CD3D*), and B cell-specific (*CD79A*, *CD22*) genes, as well as myeloid-lineage marker genes (*LYZ*, *HLA-DRA*, *CD74*, *CIQA*, *CIQB*). Stromal regions *Endothelium* (*VWF*, *EPAS1*, *EMPI*), *Fibroblast* (*DCN*, *FBLN1*, *LUM*), and *Muscle* (*ACTA2*, *TAGLN*, *MYL9*) were similarly annotated. One SCM region overexpressed recently proposed markers for club cells<sup>17,24</sup> *MMP7*, *PIGR*, *CP*, and *LTF*, and was consequently named the *Club* region.

Tumor, Luminal, Basal, Club, Immune, Fibroblast, and Muscle regions were reproduced in the validation cohort following an identical computational workflow. We identified an additional region undetected in the discovery cohort, expressing *FOS*, *JUN*, and *EGRI* which we consequently named the *Stressed luminal* region. Both *Endothelium* and the *Stressed luminal* regions were the smallest in their cohorts representing 1.8% and 2.2% of all spots, respectively, and may represent differences due to the nature of the composition of the two cohorts. Each spot's neighborhood was enriched for spots of the same region in both cohorts (Supplementary Fig. S1).

SCM regions agreed with the pathologists' assessment of sample histology (Fig. 2a). To quantitatively assess SCM region accuracy, we categorized samples in the validation cohort into *Low*, *Mid*, and *High cancer %* groups according to the fraction of spots classified as cancer by the pathologists (Supplementary Fig. S1). *Low cancer %* samples had a lower fraction of Tumor region spots than *Mid cancer %* ( $p = 1.2 \times 10^{-4}$ , two-sided Wilcoxon rank-sum test) and *High cancer %* samples ( $p = 1.1 \times 10^{-4}$ ). *High cancer %* samples had lower Basal region fraction than *Mid cancer %* ( $p = 3.5 \times 10^{-3}$ ) or *Low cancer %* ( $p = 1.4 \times 10^{-4}$ ) samples. The fraction of Luminal ( $p = 1.6 \times 10^{-2}$ , Kruskal-Wallis test) and Muscle ( $p = 2.9 \times 10^{-4}$ ) regions also varied across sampling locations while the fraction of Club, Immune, Stressed, and Fibroblast did not.

In the validation cohort, spots annotated as cancer with ISUP grade group grading were enriched for the Tumor region ( $p < 2.2 \times 10^{-16}$ , two-sided Fisher's exact test). 67% of the spots deemed normal by the pathologists' were Luminal, Basal, or Club region spots (Fig. 2b). Of the spots annotated as lymphocytes by the pathologists, 90% were labeled as the Immune region. Similar concordance of histopathology annotation and SCM regions was observed in the discovery cohort (Supplementary Fig. S2).

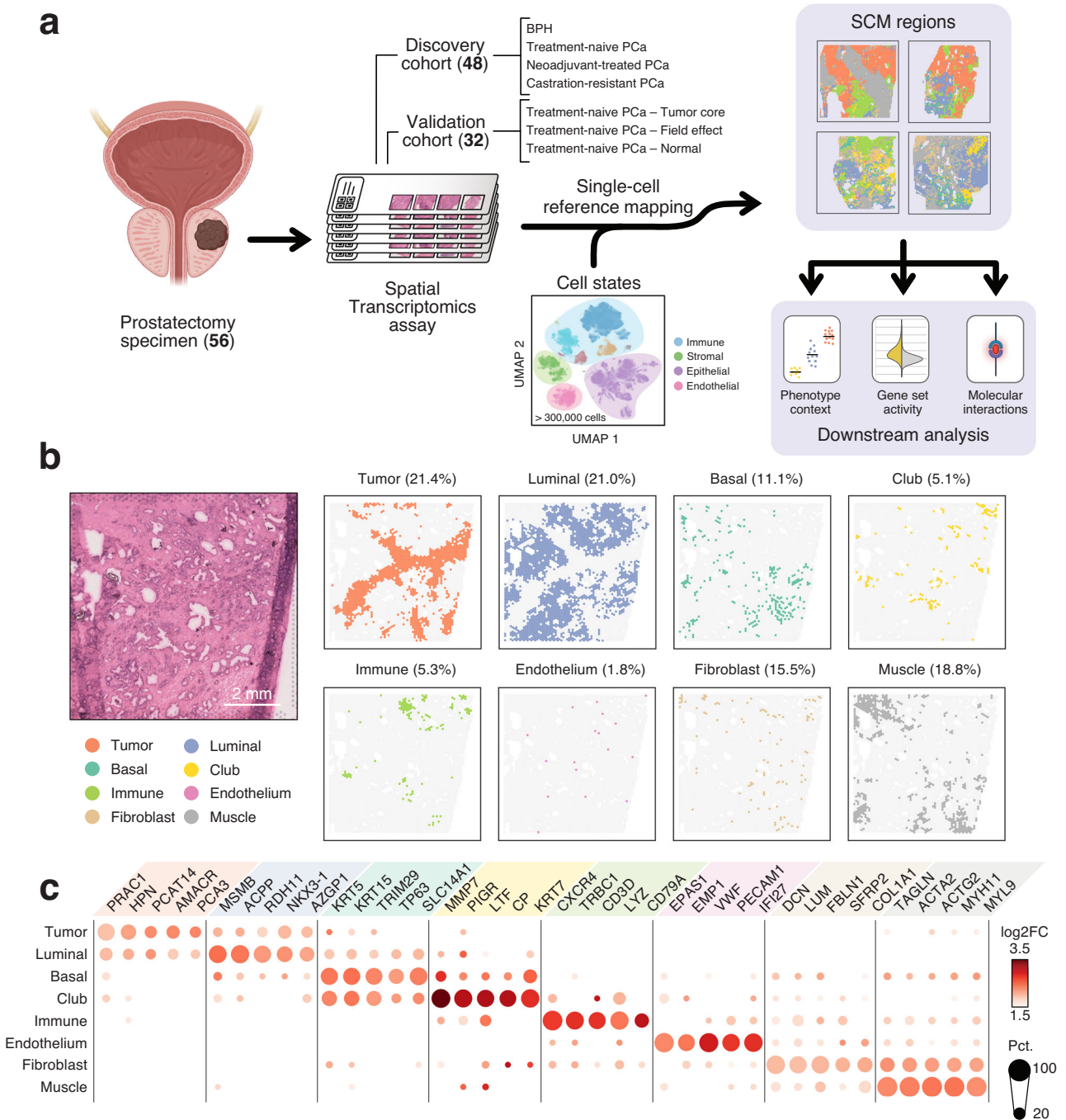
### Androgen deprivation promotes basal and club-like epithelial phenotypes

To investigate the effect of ADT on gene expression, we analyzed the epithelial SCM regions in pre- and post-treatment samples. We observed decreased expression of androgen receptor-regulated (AR-regulated) genes in the Tumor and Luminal regions following treatment (Fig. 2c). These genes were expressed at low levels in the Basal and Club regions in pre- and post-treatment samples. Conversely,

canonical basal cell markers *KRT5*, *KRT15*, and *TP63* were highly expressed in the Basal and Club regions both pre- and post-treatment.

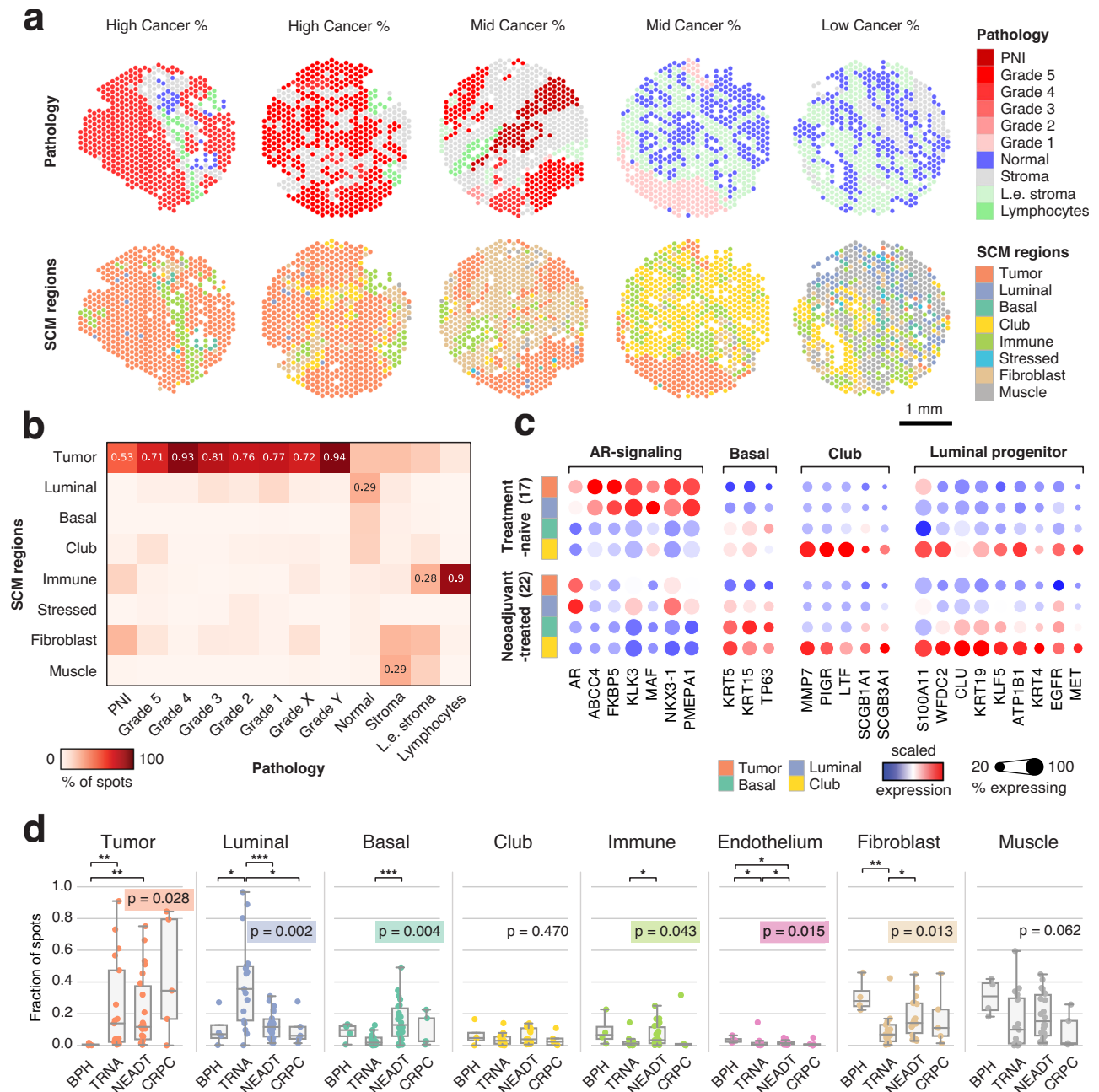
To test how treatment affected gene expression in the Club region specifically, we calculated differentially expressed genes between the Club regions of BPH, TRNA, NEADT, and CRPC sample categories (Two-sided Wilcoxon rank-sum test, Supplementary Data

S6). The expression of AR-regulated genes *ABCC4*, *KLK3*, *MAF*, *NKX3-1*, and *PMEPA1* was lower in NEADT than TRNA samples ( $\log_2 \text{fold-change} \leq -1$ ,  $p_{\text{adj}} < 0.05$ , two-sided Wilcoxon rank-sum test). The expression of canonical club cell marker *LTF* was likewise down-regulated, while the expression of *MMP7*, *PIGR*, *SCGB1A1*, and *SCGB3A1* was unaffected. Mouse luminal progenitor marker genes<sup>36</sup>



**Fig. 1 | Integration of scRNA-seq and ST data reveals the organizational patterns of the prostate TME. a** Sample collection and analysis pipeline overview. Brackets indicate the number of samples in each category. The cell state reference was assembled from previously published single-cell RNA-sequencing data<sup>6-9,12,17,32</sup>. Created in BioRender. Nykter, M. (2023) <https://BioRender.com/n30i289> **b** An untreated primary tumor sample with eight SCM regions shown separately. Percentages represent the share of spots across the discovery cohort. For details on how the SCM regions were calculated see Methods. Scale bar is 2 mm. **c** Expression of cell type gene markers across the discovery cohort. Each region was tested for

differentially expressed genes individually in all samples. Dot size represents the percentage of samples in which the gene was overexpressed (Wilcoxon rank-sum test  $p_{\text{adj}} < 0.05$  &  $\log_2 \text{fold change} \geq 1$ ). Dot color represents the average log-fold change. Region-specific marker, and their enrichment test  $p_{\text{adj}}$  (one-sided Fisher's exact test), and region-specific marker status are indicated in Supplementary Data S5. The number of region-specific markers for each region: Tumor (569), Luminal (1,776), Basal (46), Club (452), Immune (594), Endothelium (139), Fibroblast (294), Muscle (280). SCM regions single-cell mapping-based regions.



**Fig. 2 | SCM regions match histopathological features and reflect treatment-dependent shifts in gene expression.** **a** Histopathology classes and SCM regions of five representative samples from the validation cohort. Grade X and Grade Y are cancer spots with indecisive grading (see Methods). PNI: perineural invasion; L.e. stroma: lymphocyte-enriched stroma. Scale bar is 1 mm. **b** Percentage of SCM regions in each histopathology category in the validation cohort. Values are shown for each region with the highest percentage. **c** Scaled gene expression in SCM regions grouped by ADT exposure. **d** Sample-specific SCM region fractions divided

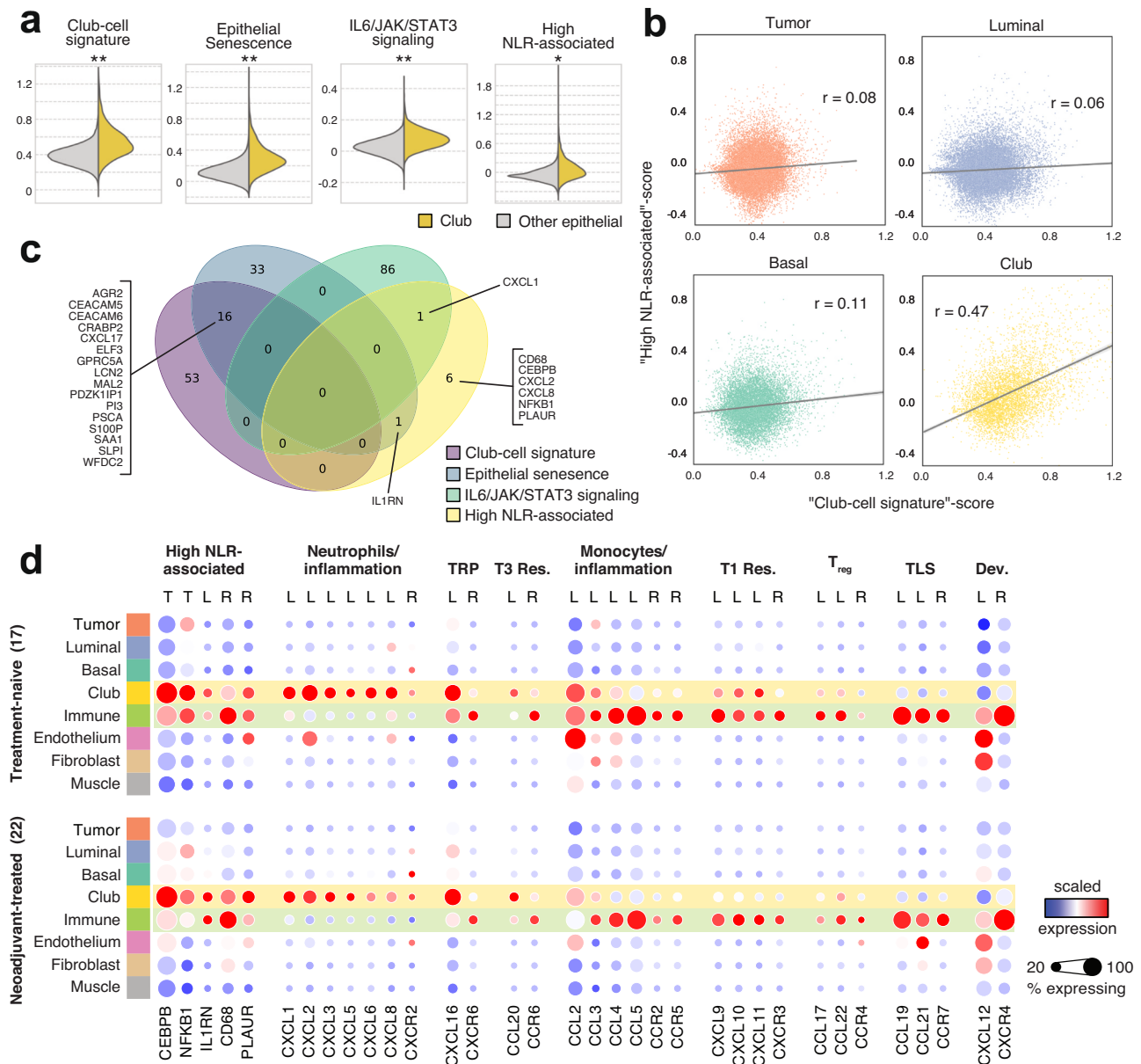
according to sample category: BPH: Benign prostatic hyperplasia ( $n = 4$ ); TRNA: Treatment-naïve prostate cancer ( $n = 17$ ); NEADT: Neoadjuvant-treated prostate cancer ( $n = 22$ ); CRPC: Castration-resistant prostate cancer ( $n = 5$ ). A Kruskal-Wallis test p-value is displayed for each category, while asterisks indicate two-sided Wilcoxon rank-sum test significance level: \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.001$ . Boxes span the interquartile range (IQR) and whiskers extend to points that lie within 1.5 IQRs of the lower and upper quartile. Center line is drawn at the median.

*S100A11*, *WFDC2*, *KRT19*, *KLF5*, *ATP1B1*, *KRT4*, and *MET* were region-specific markers for the Club region, and their expression was similarly unperturbed by treatment.

Compared to BPH, NEADT, and CRPC, the most significantly upregulated DEGs in the TRNA Club region included AR-regulated luminal cell markers such as *KLK3*, *KLK2*, *PMEPA1*, *MSMB*, and *ACPP* (Supplementary Fig. S3). Club-like cells in tumor tissue have been reported to have increased AR-signaling activity compared to healthy prostate tissue<sup>17</sup>. The upregulation of activating transcription factors

*FOS* and *JUN* was also detected in TRNA samples when compared to BPH samples, the expression of which has been previously linked to epithelial cell stress response in tumor progression<sup>17</sup>.

The percentage of Tumor, Luminal, Basal, Immune, Endothelium, and Fibroblast spots varied across the treatment categories (Fig. 2d). TRNA ( $p = 7.2 \times 10^{-3}$ , two-sided Wilcoxon rank-sum test) and NEADT ( $p = 4.5 \times 10^{-3}$ ) samples had a higher percentage of tumor spots than BPH samples. NEADT samples contained a lower percentage of Luminal region than TRNA samples ( $p = 9.2 \times 10^{-4}$ ), while a higher proportion



**Fig. 3 | The Club region has increased expression of genes linked to a senescence-associated secretory phenotype.** **a** Spotwise gene set activity scores in the Club region ( $n = 5595$ ) compared to other epithelial regions ( $n = 59,198$ ). Asterisks indicate two-sided independent samples t-test  $p$ -value ( $p < 0.05$ ) and effect size: \*Other 70th percentile < Club region mean; \*\*Other 80th percentile < Club region mean; \*\*\*Other 90th percentile < Club region mean **b** Gene set score correlation for club-cell signature and high neutrophil-to-lymphocyte ratio

associated (NLR) signature in epithelial regions. **c** A Venn diagram of gene sets shown in **(a)**. **d** Scaled gene expression in SCM regions divided by sample category. Genes are grouped according to immunological function. TRP: tissue-resident program, T3 Res: Type 3 immune response, T1 Res: Type 1 immune response,  $T_{reg}$ : regulatory T cells, TLS: Tertiary lymphoid structure, Dev.: Developmental program, T: transcription factor, L: chemokine ligand, R: chemokine receptor.

of Basal region ( $p = 4.4 \times 10^{-4}$ ). These results indicate an increase in the fraction of basal phenotype epithelial cells in response to ADT.

**The Club region is tied to inflammation and senescence-associated secretory phenotype**

To elucidate the role of club-like cells in the prostate TME, we calculated spot-level gene set activity scores on prostate-specific<sup>16,37,38</sup>, pan-cancer associated<sup>39,40</sup>, hallmark signaling<sup>41,42</sup>, and immune activity-related gene sets<sup>12,43-46</sup> and compared the scores across epithelial cell regions (Supplementary Data S7). *Club cell*<sup>6</sup>, *epithelial senescence (EpiSen)*<sup>40</sup>, *IL6/JAK/STAT3 signaling*, and *high neutrophil-to-lymphocyte ratio associated (high NLR-associated)* gene signatures scored higher in the Club region ( $p < 0.05$ , two-sided independent samples t-test and quantile thresholds, Fig. 3a). Genes related to androgen response

scored lower in the Club region, while the mouse *luminal progenitor*<sup>36</sup>, *lung club cell*<sup>47</sup>, *stem cell-like CRPC*<sup>38</sup>, and *PROSGenesis*<sup>37</sup> signatures scored higher (Supplementary Fig. S4). Multiple inflammation-related hallmark gene signatures scored higher in the Club region, including *inflammatory response*, *interferon gamma signaling*, *interferon alpha signaling*, and *KRAS signaling UP* (Supplementary Fig. S4).

The expression of *high NLR-associated* genes in CRPC primary tumors correlates with the NLR measured from peripheral blood<sup>45</sup>. High blood NLR is associated with shorter survival and treatment resistance<sup>45,48,49</sup>. We detected a four-fold higher correlation between the club cell and high NLR-associated signature scores in the Club region compared to the other epithelial regions (Fig. 3b), despite no overlap between the two signatures (Fig. 3c). We consolidated gene set activity scores with an enrichment analysis between region-specific

gene markers and these gene sets (Supplementary Fig. S5, Supplementary Data S8). Genes of the *high NLR-associated* signature were overrepresented among Club region markers ( $p_{\text{adj}} = 1.5 \times 10^{-3}$ , one-sided Fisher's exact test). One of the genes in *high NLR-associated* signature, *NFKB1*, codes for nuclear factor  $\kappa\text{B}$  (*NF- $\kappa\text{B}$* ). *TNF- $\alpha$ -signaling via NF- $\kappa\text{B}$*  was among the most significantly enriched gene sets for the Club region ( $p_{\text{adj}} = 2.7 \times 10^{-26}$ ). The p53 pathway ( $p_{\text{adj}} = 2.3 \times 10^{-7}$ ) hallmark gene set was uniquely enriched in the Club region, in line with its putative senescent nature<sup>50</sup>.

Club region-specific markers were enriched for genes that encode for proteins whose secretion is increased in the senescence-associated secretory phenotype<sup>51</sup> (SASP) suggesting that club-like cells engage in similar secretory activity ( $p_{\text{adj}} = 7.2 \times 10^{-4}$ ). SASP is typified by the secretion of chemokines, small molecules responsible for inducing chemotaxis of immune cells that have a significant role in TME organization<sup>52,53</sup>. We examined the expression of key chemokines across all SCM regions in a pre- and post-treatment setting (Fig. 3d). Neutrophil chemotaxis-inducing chemokines *CXCL1* ( $p_{\text{adj}} = 1.5 \times 10^{-13}$ , one-sided Fisher's exact test), *CXCL2* ( $p_{\text{adj}} = 7.4 \times 10^{-13}$ ), and *CXCL8* ( $p_{\text{adj}} = 1.7 \times 10^{-5}$ ) were region-specific markers for the Club region, along with *CXCL16* ( $p_{\text{adj}} = 1.5 \times 10^{-13}$ ) and *CCL20* ( $p_{\text{adj}} = 2.0 \times 10^{-4}$ ). Canonical receptors of these chemokines<sup>53</sup> were region-specific markers in the Immune region: *CXCR6* ( $p_{\text{adj}} = 2.7 \times 10^{-2}$ ) for *CXCL16* and *CCR6* ( $p_{\text{adj}} = 4.1 \times 10^{-4}$ ) for *CCL20*. While the sensitivity of ST was not sufficient to detect significant levels of neutrophil chemotactic receptor *CXCR2* expression, the elevated expression of its canonical ligands *CXCL1*, *CXCL2*, and *CXCL8* suggests that club-like cells partake in their recruitment to the TME.

#### Areas proximal to the Club region have increased polymorphonuclear myeloid-derived suppressor cell activity

Overrepresentation analysis of Gene Ontology: Biological Process (GO:BP) terms revealed enrichment of myeloid cell chemotaxis-associated gene sets in the Club region (Fig. 4a). Lymphoid and myeloid immune cell activation-related processes were enriched in the Immune region. Two gene sets indicating prostate-specific myeloid-derived suppressor cell (MDSC) activity derived from mice<sup>43</sup> ( $p = 4.3 \times 10^{-7}$ ) and human<sup>12</sup> ( $p = 2.0 \times 10^{-6}$ ) were also enriched in the Immune region (Supplementary Data S8). A CRISPR-Cas9 validated signature of polymorphonuclear MDSC (PMN-MDSC) immunosuppression<sup>46</sup> was enriched in the Immune region ( $p = 1.9 \times 10^{-2}$ ), indicating an accumulation of immunosuppressive PMN-MDSCs in the prostate TME.

The fraction of Club region spots in a sample correlated positively with its PMN-MDSC activity score calculated from non-Club region spots in both cohorts (Fig. 4b). Tumor region spots had a higher PMN-MDSC activity score when proximal to Club region spots (Fig. 4c). Similar results were acquired for three other MDSC signatures (Supplementary Fig. S6). One MDSC activity signature scored higher in the recurrent samples ( $n = 20$ ) than in non-recurrent samples of the validation cohort ( $n = 12$ ) ( $p = 1.2 \times 10^{-4}$ , two-sided Wilcoxon rank-sum test). PMN-MDSC activity scores didn't differ between the TRNA and NEADT groups ( $p \geq 0.05$ , two-sided independent samples t-test, Fig. 4d).

To validate the observed proximity between club-like cell prevalence and PMN-MDSC infiltration, we performed multiplex immunohistochemistry staining of 16 primary untreated prostate tumor tissue sections using antibodies against previously validated markers for these cell types (Methods). From the resulting staining images, we selected 101 approximately 3 mm<sup>2</sup> regions of interest that were either club-like negative (absent *LTF* staining,  $n = 54$ ) or club-like positive (moderate to strong *LTF* staining,  $n = 47$ ) (Fig. 4e, Fig. 4f, Supplementary Fig. S7). We then trained a random trees cell classifier for seven cell categories, including PanCK<sup>+</sup>LTF<sup>+</sup>/PIGR<sup>+</sup> club-like cells<sup>24</sup> and CD45<sup>+</sup>CD66b<sup>+</sup>CD11b<sup>+</sup>CXCR2<sup>+</sup> PMN-MDSCs<sup>45</sup> (Fig. 4g, Methods).

Across all ROIs, the number of club-like-classified cells correlated positively with the number of PMN-MDSC-classified cells (Fig. 4h). A higher proportion of PMN-MDSCs was present in club-like positive ROIs than in club-like negative ROIs ( $p = 1.7 \times 10^{-8}$ , two-sided Wilcoxon rank-sum test) (Fig. 4i). No difference in the total number of detected cells between club-like positive and negative ROIs was observed ( $p = 0.23$ , two-sided Wilcoxon rank-sum test), while a greater number of club-like cells was present in the club-like positive ROIs ( $p = 3.0 \times 10^{-15}$ ) (Supplementary Fig. S7). Taken together these results demonstrate that the presence of club-like cells is strongly associated with PMN-MDSC infiltration and immunosuppressive activity in the prostate TME.

#### Club-derived ligand-receptor signaling activity is specific to the interacting region

To elucidate the specific molecular interactions occurring between the Club and other regions, we performed ligand-receptor binding analysis in the Club region interfaces (Supplementary Fig. S8, Supplementary Data S9). Briefly, we surveyed spots of each region that were proximal to Club region spots and tested for interactions between these two sets (Methods). We filtered the interactions to those that had a ligand (from Club to other regions) or receptor (from other regions to Club) overexpressed in the Club region. We then tested for region interface-specific enrichment by comparing the number of interfaces where a specific ligand-receptor pair was active.

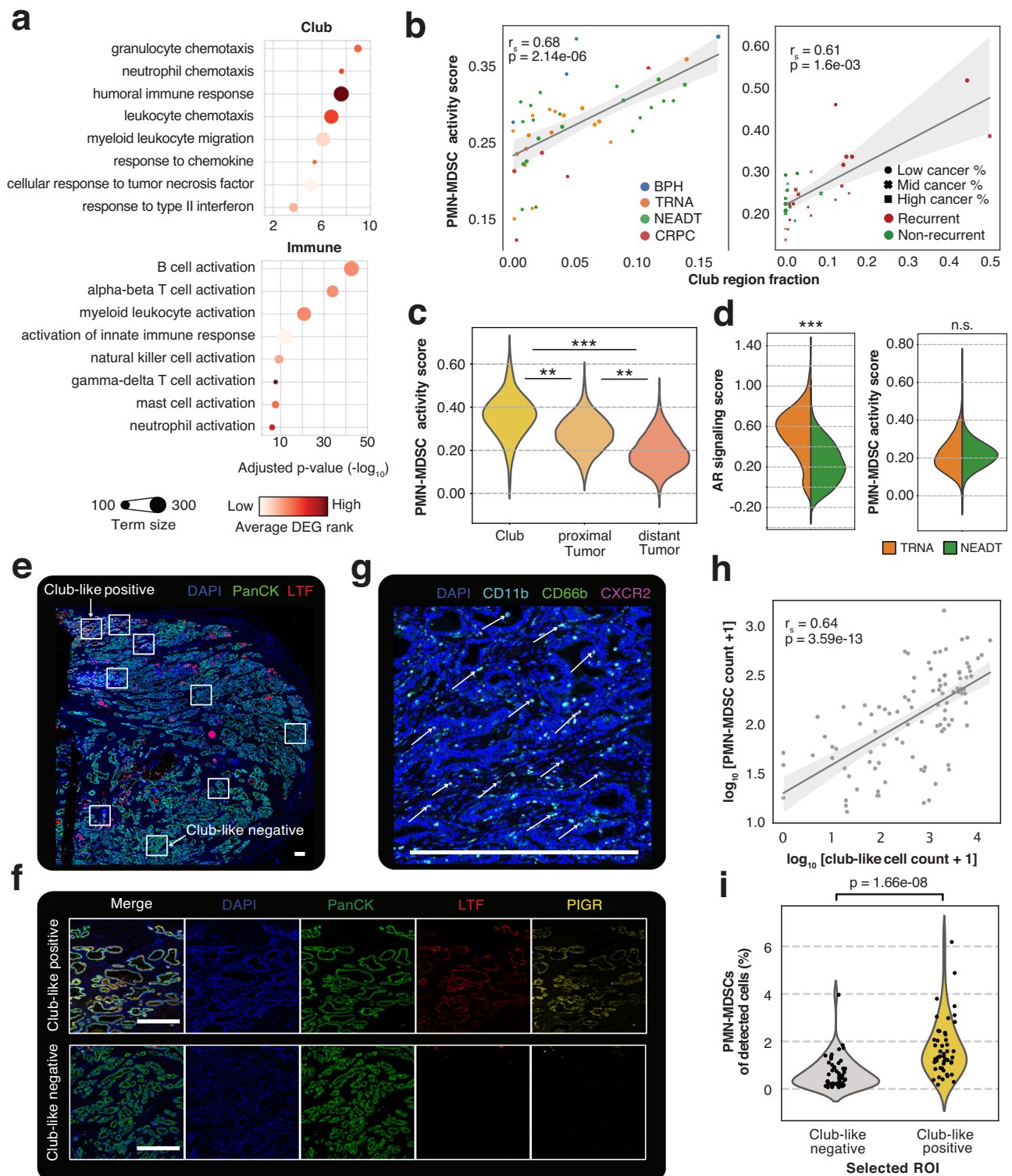
This analysis identified 142 unique ligand-receptor interactions occurring at the Club region interface, 59 (41.5%) of which were enriched at specific interfaces ( $p_{\text{adj}} < 0.05$ , one-sided Fisher's exact test). *Atypical chemokine receptor 1* (*ACKR1*) was the most common target for neutrophil chemotaxis-inducing chemokines *CXCL1*, *CXCL2*, *CXCL5*, *CXCL6*, and *CXCL8* (Supplementary Fig. S8). The *CXCL2* to *ACKR1* interaction was enriched at the Club-Fibroblast interface ( $p_{\text{adj}} = 2.3 \times 10^{-2}$ ). *Dipeptidyl peptidase-4* (*DPP4*) was a unique target of *CXCL2* and enriched at the Club-Luminal interface ( $p_{\text{adj}} = 5.7 \times 10^{-8}$ , Supplementary Fig. S8). *CCL20* and *CCR6* formed an interaction that was enriched in the Club-Immune interface ( $p_{\text{adj}} = 2.2 \times 10^{-3}$ ).

The Club region overexpressed receptors with known oncogenic properties (*MET*, *EPHA2*)<sup>54,55</sup>, inflammatory activities (*CEACAM1*, *PTGS2*)<sup>56,57</sup>, and adhesive function (*ITGA3*, *ITGA6*) (Supplementary Fig. S8). The Fibroblast-Club interface was enriched for *MET* interaction with *decorin* (*DCN*,  $p_{\text{adj}} = 1.0 \times 10^{-6}$ ) and *hepatocyte growth factor* (*HGF*,  $p_{\text{adj}} = 5.6 \times 10^{-5}$ ). *EPHA2* was the target of multiple *ephrin*-family ligands, of which *EFNA1* and *EFNA5* signaling interactions were enriched at the Luminal ( $p_{\text{adj}} = 1.0 \times 10^{-6}$ ) and Muscle ( $p_{\text{adj}} = 2.4 \times 10^{-2}$ ) interfaces, respectively. *TP53* and *PTGS2* interaction was enriched at the Tumor-Club interface ( $p_{\text{adj}} = 3.0 \times 10^{-2}$ ). An interaction between Immune region-expressed *CCL5* and Club region-expressed *SDCI* was the single most enriched interaction between these two regions ( $p_{\text{adj}} = 5.9 \times 10^{-3}$ ). These results elucidate the region-specific signaling relied upon by the club-like cells.

#### Club-like senescence is associated with immunosuppressive PMN-MDSC activity in primary and metastatic tumors

We further examined the connection between the Club region SASP and PMN-MDSC activity. We detected a significant overlap between the *EpiSen* signature and the Club region markers ( $p = 3.2 \times 10^{-22}$ , one-sided Fisher's exact test), as well as the *EpiSen* signature and PMN-MDSC activity signature ( $p = 2.1 \times 10^{-5}$ , Fig. 5a). We termed the 26-gene overlap between *EpiSen* and Club region markers *club-like senescence*.

Of the genes overlapping between these three signatures, we found eleven genes to be expressed in the neutrophil and monocyte populations of metastatic CRPC (mCRPC) scRNA-seq data<sup>58</sup>, suggesting that similar senescence-driven processes could be taking place in the metastatic tumor environment (Fig. 5b). To investigate, we generated ST data from four metastatic prostate cancer tumors collected



from a pelvic lymph node, liver, pericardial region, and the subdural region, each from a different patient (Methods, Supplementary Data S10). We determined expression-based clusters for each sample separately to find spots with similar expression and underlying cell populations (Fig. 5c, Supplementary Fig. S9). We then performed enrichment analysis and calculated gene set activity scores to determine whether any of these clusters corresponded to the club-like cell population identified in primary tumors.

Clusters *Met A1* (liver metastasis) and *Met B5* (subdural metastasis) displayed elevated scores for the Club region markers and the club-like

senescence signature (Fig. 5d). These tissue sites overexpressed genes ( $\log_2$  fold-change  $\geq 1$ ,  $p_{\text{adj}} < 0.05$ , two-sided Wilcoxon rank-sum test, Supplementary Data S10) that were enriched for Club region markers (*Met A1*  $p_{\text{adj}} = 3.2 \times 10^{-7}$  and *Met B5*  $p_{\text{adj}} = 1.3 \times 10^{-13}$ , one-sided Fisher's exact test) and PMN-MDSC activity markers (*Met A1*  $p_{\text{adj}} = 3.2 \times 10^{-2}$ , *Met B5*  $p_{\text{adj}} = 2.3 \times 10^{-2}$ , Fig. 5e). Genes overexpressed in the *Met B1* and *Met B2* clusters were similarly enriched for the Club region markers ( $p_{\text{adj}} = 3.4 \times 10^{-7}$  and  $p_{\text{adj}} = 1.5 \times 10^{-4}$ , respectively). *Met B1* was enriched for all five of the tested signatures, including *EpiSen*, club-like senescence, and high NLR-associated genes. Canonical club cell marker *MMP7* was

**Fig. 4 | Club-like cell enriched regions are associated with increased myeloid-derived suppressor cell infiltration.** **a** GO:BP term enrichments for genes with upregulated expression in the Club (top) and Immune (bottom) regions. One-sided Fisher's exact test was used. **b** Correlation between Club region fraction and PMN-MDSC activity scores in pseudo-bulk for the discovery (left) and the validation (right) cohorts. Two-sided Spearman correlation coefficient and p-value are shown. Error bands in light grey span the 95% confidence interval calculated using a bootstrap. BPH: benign prostatic hyperplasia ( $n = 4$ ), TRNA: treatment-naïve prostate cancer ( $n = 17$ ); NEADT: neoadjuvant-treated prostate cancer ( $n = 22$ ); CRPC: Castration-resistant prostate cancer ( $n = 5$ ). Low cancer%  $n = 11$ , Mid cancer %  $n = 11$ , High cancer %  $n = 10$ . **c** PMN-MDSC activity score in treatment-naïve spots grouped according to their proximity to the Club region. Asterisks indicate two-sided independent samples t-test significance levels ( $p < 0.05$ ) and effect size: \*70th percentile <mean; \*\*80th percentile <mean; \*\*\*90th percentile <mean. Club region

( $n = 278$ ), proximal Tumor ( $n = 687$ ), distant Tumor ( $n = 8751$ ). **d** Violin plots of gene set activity scores across all regions in treatment-naïve ( $n = 36,198$ ) and neoadjuvant-treated ( $n = 59,259$ ) samples. Asterisks are the same as in (c). **e** A representative immunostained prostate tissue section (total  $n = 16$ ) with selected regions of interest (ROIs). **f** Representative images of club-like positive ( $n = 47$ ) and club-like negative ( $n = 54$ ) ROIs corresponding to those annotated in (e). **g** A representative image of a club-like positive ROI ( $n = 47$ ) with a three-colour stain showing CD66b<sup>+</sup>CD11b<sup>+</sup>CXCR2<sup>+</sup> PMN-MDSCs. Arrows point to example cases. Scale bars in white are 500  $\mu\text{m}$  in each panel. **h** Scatterplot of log-transformed club-like cell and PMN-MDSC cell counts. Each dot represents a ROI ( $n = 101$ ). Correlation coefficient, p-value, and error bands were calculated the same as in (b). **i** Violinplot of PMN-MDSC percentage of all detected cells in club-like negative ( $n = 54$ ) and club-like positive ( $n = 47$ ) ROIs. A Wilcoxon rank-sum test p-value is shown.

overexpressed in *Met B1* ( $\log_2$  foldchange = 1.06,  $p_{\text{adj}} = 4.5 \times 10^{-24}$ ) and *Met B5* ( $p_{\text{adj}} = 8.8 \times 10^{-21}$ ). *PIGR* was similarly overexpressed in *Met B5* ( $p_{\text{adj}} = 1.2 \times 10^{-2}$ ). Other Club region markers overexpressed in *Met B5* included *CLDNI* ( $p_{\text{adj}} = 7.8 \times 10^{-13}$ ), *SCNNIA* ( $p_{\text{adj}} = 2.9 \times 10^{-6}$ ), and *TACSTD2* ( $p_{\text{adj}} = 8.5 \times 10^{-61}$ ), all of which are exclusively expressed in epithelial cells. These results suggest the presence of club-like cells in metastatic prostate cancer tumors.

*Club-like senescence* and *PMN-MDSC activity* signature GSVA scores correlated across untreated primary and mCRPC tumors in the TCGA-PRAD and SU2C-PCF cohorts (Fig. 5g, h). The score of *club-like senescence* signature also correlated with two other MDSC signatures, and the *high NLR-associated* signature (Supplementary Fig. S10). *Club-like senescence* score correlated negatively with AR-signaling in both TCGA-PRAD ( $r_s = -0.30$ ,  $p = 1.2 \times 10^{-12}$ , two-sided Spearman correlation test) and SU2C-PCF ( $r_s = -0.28$ ,  $p = 5.0 \times 10^{-6}$ ) cohorts.

Finally, we generated pseudo-bulk expression profiles of the discovery cohort ST samples to investigate how treatment affected the expression of senescence-related genes. Hierarchical clustering of the expression of *AR-signaling* and *club-like senescence* genes revealed nested sample groups within the cohort (Fig. 5f). Altogether 16 NEADT samples were characterized by decreased *AR-signaling* and increased *club-like senescence* activity. Differential expression analysis between the TRNA and NEADT groups revealed increased expression for genes in the *club-like senescence* and the *high NLR-associated* signatures ( $\log_2$  fold-change  $\geq 1$ ,  $p_{\text{adj}} < 0.05$ , two-sided Wald test, Fig. 5i, Supplementary Data S11). None of the genes in these signatures were downregulated post-treatment ( $p_{\text{adj}} \geq 0.05$ , two-sided Wald test).

## Discussion

We used spatial transcriptomics to explore the microenvironment of benign, untreated, neoadjuvant-treated, and castration-resistant prostate tumors. We used a single-cell expression reference to annotate tissue regions across 146,780 spatially defined data points, which not only recapitulated the histology annotated by pathologists but also allowed for joint analysis across a large set of ST data. Our analysis pipeline is inspired by NMF-based approaches that have captured recurrent gene expression programs across different tumor types<sup>39,40,59</sup> emphasizing gene set variation within individual samples before cross-sample integration. The advantage of our analysis is that we avoid batch effect issues related to commonly used clustering workflows<sup>60,61</sup>, resulting in an unbiased cell-state reference.

Here we report that club-like cells have upregulated expression of genes encoding for molecules secreted in SASP<sup>51</sup>, including TME-altering proteases *cathepsin  $\beta$*  (*CTSB*), *stromelysin-2* (*MMP10*), and *urokinase* (*PLAU*), *insulin-like growth factor-binding protein 3* (*IGFBP3*), and *intercellular adhesion molecule 1* (*ICAMI*). Furthermore, club-like cells display elevated expression of similarly SASP-associated myeloid cell chemotaxis-inducing chemokines<sup>31,33</sup> *CXCL1*, *CXCL2*, *CXCL3*, *CXCL8*, and *CCL20*, as well as the transmembrane chemokine *CXCL16*. In a

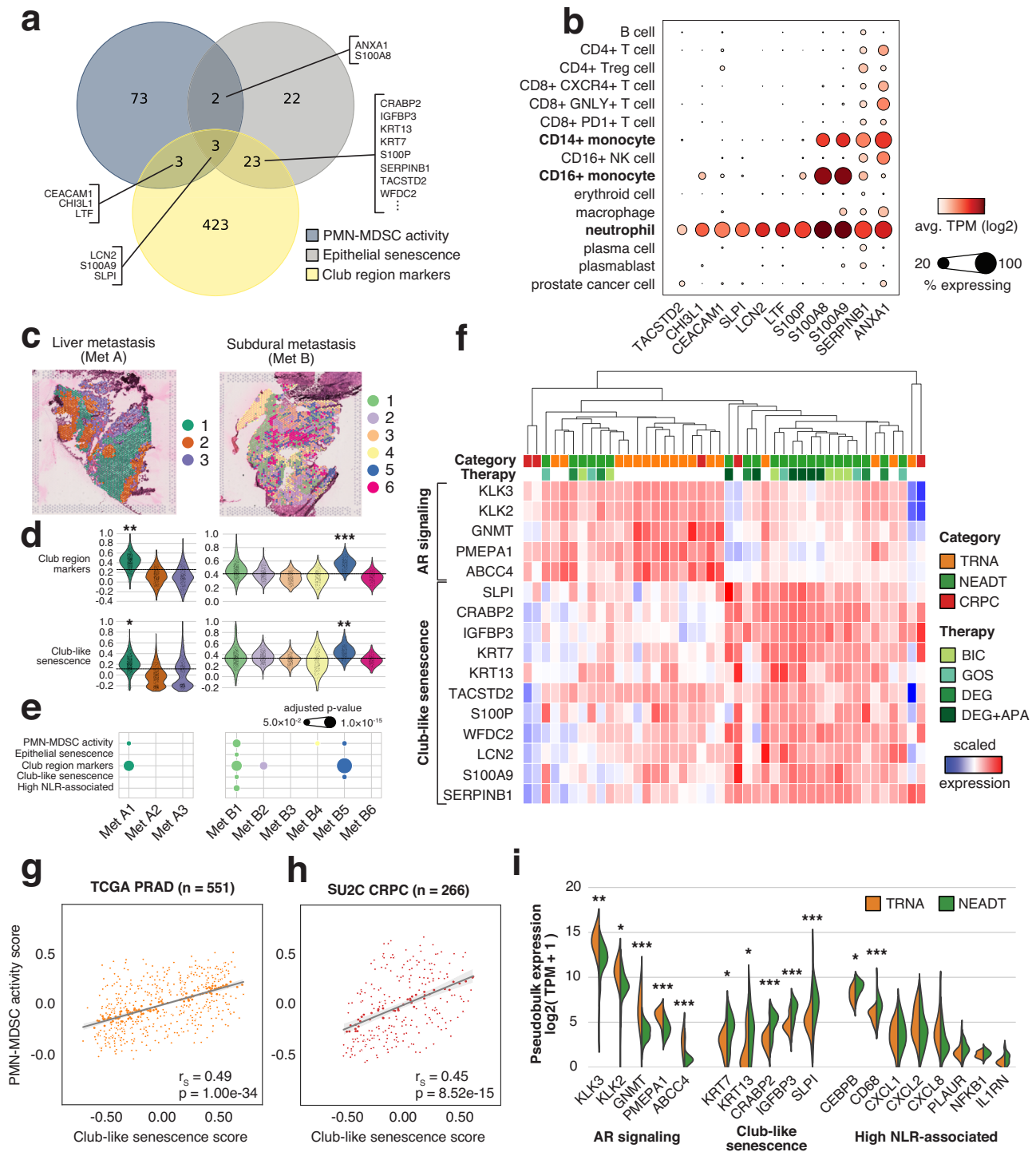
prostate-specific context, the expression of *CXCL1*, *CXCL2*, and *CXCL8* is independently predictive of poor overall survival in metastatic CRPC patients<sup>45</sup>. High *CCL20* expression in the prostate is associated with more aggressive disease<sup>62</sup> and its blockade slows tumor progression in mice<sup>12</sup>. In vitro studies have shown that *CXCL16* expression induces tumor cell and mesenchymal stem cell migration, promoting cancer metastasis to the bone<sup>63,64</sup>. We observed upregulated *CXCL12* expression in the Fibroblast region, and high *CXCR4* and *CXCR6* expression in the Immune region, all of which have been indicated as key molecules in possible metastasis mechanisms acting through *CXCL16*<sup>64</sup>.

Increased cytokine signaling activity in the Club region is in line with reports that club-like cells are enriched in proliferative inflammatory atrophy (PIA)<sup>24,65,66</sup>. PIA has been regarded as a potential precursor for prostate cancer<sup>65</sup>, with DNA damage resulting from cytokine-induced oxidative stress reported as one possible mechanism<sup>14,67</sup>. We found increased expression of PIA gene markers<sup>66</sup> *KRT5*, *KRT8*, and *MET*, and PIA-associated club-like cell gene markers<sup>17,24</sup> *CP*, *MMP7*, *LTF*, and *PIGR* in the Club region. Inflammation-related molecular signaling routes *IL6/JAK/STAT3*, *IFN $\gamma$  response*, and *TNF $\alpha$  signaling* via *NF $\kappa$ B* also had increased activity in the Club region. Similar inflammatory signaling activity has been reported in club-like cells of histologically benign glands proximal to invasive cribriform carcinoma or intraductal carcinoma-enriched regions<sup>9</sup>. Inflammation-associated low-CD38 expressing luminal cells have previously been associated with biochemical recurrence<sup>68</sup>, whereas high IFN $\gamma$  signaling is associated with worse outcomes in high-risk localized disease<sup>69</sup>. These results indicate that the club-like cells are interchangeably linked with inflammation in the prostate TME.

Our findings corroborate a link between club-like cells and the luminal progenitor phenotype reported in mice<sup>27</sup>. Luminal progenitor cells are an inherently castration-resistant cell population in the mouse prostate<sup>25,70</sup> that have also been proposed as a possible cellular origin of prostate cancer<sup>71</sup>. These cells are mainly present in the proximal region of the prostatic duct, with scattered cells found in the distal lobes<sup>20,23,72</sup>. The previously described transcriptomic similarity of the mouse progenitor and human club-like cells is supported by our data<sup>36</sup>. Our analysis also shows that the Club region is unperturbed by ADT, indicating that human club-like cells are similarly resistant to castration. High expression of transcription factors *SOX9*, *KLF5*, and *ELF3* in the Club region is also indicative of the stem-like characteristics of these cells<sup>73</sup>.

Finally, we postulate a connection between club-like cells and clinically significant immunosuppression. Chronic inflammation in the prostate facilitates cancer progression and results in the accumulation of pathologically activated neutrophils or PMN-MDSCs<sup>14,46,74</sup>. Immunosuppressive PMN-MDSCs are a driver of CRPC that can reactivate androgen signaling through IL-23 secretion<sup>43</sup>. In general, tumor sites contain expanded PMN-MDSC populations compared to healthy tissue<sup>74,75</sup>, and the inhibition of their chemotactic receptor *CXCR2* can





**Fig. 5 | Club-like senescence is associated with immunosuppressive PMN-MDSC activity in primary and metastatic tumors.** **a** Venn diagram showing overlaps between PMN-MDSC activity, epithelial senescence, and club-region upregulated gene sets. **b** Dot plot of normalized gene expression in metastatic castration-resistant prostate cancer samples. Data and clusters as reported in He et al. 2021<sup>58</sup>. **c** Expression-based clustering overlaid on two prostate cancer metastasis ST samples. **d** Cluster-specific gene set scores in MET A ( $n = 2190$ ) and MET B ( $n = 2346$ ) ST spots. The dashed line marks the overall score median. Asterisks indicate two-sided independent samples t-test p-value ( $p < 0.05$ ) and effect size (\*cluster 30<sup>th</sup> percentile > overall median; \*\*cluster 20<sup>th</sup> percentile > overall median; \*cluster 10<sup>th</sup> percentile > overall median) **e** Log-transformed overrepresentation  $p_{adj}$  among each cluster overexpressed genes. One-sided Fisher’s exact test was used. **f** Scaled

gene expression of pseudo-bulk spatial transcriptomics samples. **g, h** Score correlation for the Club-like senescence and PMN-MDSC activity signatures in TCGA PRAD ( $n = 551$ ) and SU2C ( $n = 266$ ) cohorts. Two-sided Spearman correlation coefficient and p-value are shown. Error bands in light grey span the 95% confidence interval calculated using a bootstrap. **i** Violin plot of normalized gene expression in pseudo-bulk spatial transcriptomics samples. Asterisks indicate differential gene expression test significance levels between treatment-naïve ( $n = 17$ ) and neoadjuvant-treated ( $n = 22$ ) prostate cancer samples (\* $p_{adj} < 0.05$ , \*\* $p_{adj} < 0.01$ , \*\*\* $p_{adj} < 0.001$ , two-sided Wald test). PMN-MDSC: polymorphonuclear myeloid-derived suppressor cell; TRNA: treatment-naïve prostate cancer ( $n = 17$ ), NEADT: neoadjuvant-treated prostate cancer ( $n = 22$ ); CRPC castration-resistant prostate cancer ( $n = 5$ ), BIC bicalutamide, GOS goserelin, DEG degarelix, APA apalutamide.

reverse castration resistance<sup>45</sup>. We show that club-like cells have elevated expression of canonical neutrophil chemokines and that their presence is associated with increased PMN-MDSC activity in primary tumors. Furthermore, we identify club-like cell populations in metastatic prostate cancer tumors, expanding their relevance outside the primary tumor setting.

Taken together, our analysis highlights club-like cells as a key TME constituent that is associated with PMN-MDSC infiltration. The combination of their ability to persist throughout ADT and to induce PMN-MDSC chemotaxis implicates them as a contributor to treatment resistance. Our findings warrant further research into how club-like cells contribute to prostate cancer initiation, progression, and resistance to therapy.

## Methods

### Sample collection

This study was carried out in compliance with all relevant ethical regulations and guidelines. The collection of patient samples was approved by the local ethical committees in Tampere University Hospital, Universitaire Ziekenhuizen KU Leuven, St. Olav's Hospital in Trondheim, and the Johns Hopkins Hospital. For the discovery cohort, a total of 48 prostatectomy samples were collected at Tampere University Hospital (37) and UZ Leuven (11). Freshly frozen prostate tissue samples from 37 patients of different disease stages, including benign prostatic hyperplasia ( $n = 4$ ), treatment-naïve ( $n = 17$ ) and neoadjuvant-treated primary PCa ( $n = 11$ ), locally recurrent castration-resistant prostate cancer (CRPC,  $n = 5$ ) were obtained from Tampere University Hospital (Tampere, Finland) and utilized for the analysis. The neoadjuvant-treated samples were collected as part of clinical trial [NCT00293696](#) evaluating the molecular effects of anti-androgen and chemical castration treatment<sup>76</sup>. The use of clinical material was approved by the Ethics Committee of the Tampere University Hospital and the National Authority for Medicolegal Affairs. For prospective sample collection, written informed consent was obtained from all the subjects. 11 freshly frozen neoadjuvant-treated primary PCa samples from radical prostatectomies samples of 11 patients were collected at the Universitaire Ziekenhuizen KU Leuven as part of clinical trial [NCT03080116](#)<sup>77</sup>. The mean age of discovery cohort cancer patients at surgery was 63 years (range: 33–79), the mean ISUP grade group was 2.7 (range: 1–5), and the mean pre-operation blood PSA level 12.7 ng/ml (range: 3.2–111) (Supplementary Data S1).

Of the metastatic prostate cancer samples used in this study, three were acquired as part of the Johns Hopkins Medicine Institutional Review Board-approved (NA\_00003925) Project to ELminate lethal CANcer (PELICAN) from patients who provided written informed consent. One sample was acquired under Tampere University Hospital Ethics Committee approval R19074 from a patient who had provided written informed consent.

For the validation cohort, 8 freshly frozen prostate tissue specimens were collected from patients with untreated primary PCa who had given informed written consent before undergoing radical prostatectomy at St. Olav's Hospital in Trondheim between 2008 and 2016. This research received approval from the regional ethical committee of Central Norway (identifier 2017/576) and adhered to both national and EU ethical regulations. The mean age at surgery was 59 years (range: 53–73), the mean ISUP grade was 3.4 (range: 2–5) and the mean pre-operation blood PSA level was 20.1 ng/ml (range: 9.6–45.9). Three patients remained relapse-free for >10 years following surgery, while five patients experienced relapse with confirmed metastasis within three years. For each patient, two samples with cancerous tissue (Cancer), one sample with noncancerous morphology close to the cancerous region (Field effect), and one sample with noncancerous morphology distant from the cancerous area (Normal) were sectioned. In total, 32 samples (no relapse  $n = 12$ , relapse  $n = 20$ ) were collected.

### Spatial transcriptomics assay

For the discovery cohort and the metastatic tumor samples, tissue sections were profiled for spatial transcriptomics using the Visium Spatial Gene Expression Reagent Kit protocol from 10x Genomics (CG000239, Rev F, 10x Genomics). The tissues were cryosectioned with Cryostat SLEE MEV+ at 10  $\mu\text{m}$  thickness to the Visium spatial gene expression slide, fixed in ice-cold 100% methanol for 30 min at  $-20\text{ }^{\circ}\text{C}$ , and hematoxylin and eosin (H&E) staining was done manually based on the Methanol Fixation, H&E Staining & Imaging for Visium Spatial Protocols (CG000160, Rev D, 10x Genomics). The capture areas were imaged individually using Hamamatsu NanoZoomer S60 digital slide scanner. Sequencing library preparation was performed according to the Visium Spatial Gene Expression user guide (CG000239 Rev F, 10x Genomics) using a 20-minute tissue permeabilization time. Sequencing was done on the Illumina NovaSeq 6000 sequencer at Novogene Company Limited, (Cambridge, UK) or CeGaT GmbH (Tübingen, Germany) sequencing facilities, aiming at a minimum of 50,000 read pairs per tissue-covered spot (55  $\mu\text{m}$ ; 1–10 cells) as recommended by the manufacturer.

For the validation cohort samples, 10  $\mu\text{m}$ -thick tissue sections were cut at  $-20\text{ }^{\circ}\text{C}$  using a Cryostar NX70 (Thermo Fisher), placed onto the Visium slides, and processed following the manufacturer's protocol. Before extracting RNA, tissue sections were treated with 100% methanol, stained with H&E, and scanned digitally at 20x magnification. RNA was isolated by treating the tissue sections with a permeabilization agent for 12 minutes. The most effective extraction time had been established earlier using the Visium Spatial Tissue Optimization Slide & Reagent kit (10X Genomics). Following this, a second strand mix was introduced to generate a complementary strand, after which cDNA was amplified using real-time qPCR. The resulting amplified cDNA library was quantified using the QuantStudio™ 5 Real-Time PCR System (Thermo Fisher) via qPCR, and the cDNA libraries were preserved at  $-20\text{ }^{\circ}\text{C}$  until needed for subsequent experiments.

### Histopathology evaluation

For the discovery cohort samples that were generated at Tampere, a board-certified uropathologist (T.M.) assigned individual spots into one of 9 categories (Gleason 5, Gleason 4,

Gleason 4 cribriform, Gleason 3, Atrophy, Prostatic Intraepithelial Neoplasia (PIN), Inflammation, Benign, and Stroma) using Loupe Browser version 6.0 (10x Genomics). Quality control of selected samples and qualitative assessment of histological concordance with SCM regions was performed by an additional pathologist (C.T.A.P.).

For the validation cohort, two experienced uropathologists (T.V. and Ø.S.) independently evaluated the H&E-stained sections from all 32 spatial transcriptomics samples in QuPath (version 0.2.3). They annotated cancer areas according to the International Society for Urological Pathology (ISUP) Grade Group system and also identified aggregates of lymphocytes. For the downstream data analysis, a consensus pathology annotation was reached in agreement with both pathologists. Gleason scores were transformed into ISUP grade groups (ISUP1–5). Cancer areas with uncertain grading were annotated as ISUPX (indecisive between ISUP3 and ISUP5) and ISUPY (indecisive between ISUP1 and ISUP4). Other section annotations included non-cancerous glands, stroma, and stroma with higher levels of lymphocytes (referred to as 'lymphocyte-enriched stroma'). A detailed description of how histology annotations in QuPath were converted into spot-wise histology classes is given elsewhere<sup>78</sup>.

### Spatial transcriptomics data preprocessing

Libraries were sequenced with NovaSeq 6000 PE150. Images of each capture area were transformed from *ndpi* to *tiff*-format at the 20x zoom level using *ndpi2tiff v1.8*<sup>79</sup>. Tissue alignment masks were created using Loupe Browser v4.2.0. Read alignment to GRCh38 and transcript counting were performed using *spaceranger v1.1.0*, where the *tiff*

image and the tissue alignment masks were also used as input. The results were read into Python 3.7 using *scanpy.read\_visium* (scanpy v1.9.1). For each sample individually, genes present in less than 5 spots and spots with less than 500 UMIs were discarded using *scanpy.pp.filter\_genes* and *scanpy.pp.filter\_cells*, respectively. Transcript counts were normalized using the single-cell integration benchmark (scib, v1.1.1) implementation of the scran method by running *scib.preprocessing.normalize*<sup>80,81</sup>.

### Single-cell reference mapping

The scRNA-seq cell-state reference created from publicly available data<sup>6–9,12,17,32</sup> (Supplementary Methods) was transformed into a mappable reference using *cell2location.models.RegressionModel* (cell2location v0.1.3) with the dataset of origin used as *batch\_key*, cell-state annotation as *labels\_key*, and original sample used as *categorical\_covariate\_key*. The regression was performed on unnormalized gene counts. The model was trained for 250 epochs on an NVIDIA Tesla V100 16GB graphics processing unit (GPU). *RegressionModel.export\_posterior* with parameters *num\_samples: 1000*, *batch\_size: 2500* on the same GPU was used to export the posterior distribution. From these distributions, the mean expression of each gene in each cell state (*means\_per\_cluster\_mu\_fg*) was used for mapping. A single cell2location model was created using *cell2location.models.Cell2location* by concatenating the untransformed (raw) ST data gene counts and setting the parameter *batch\_key* as the sample identifier. The model was trained on the same NVIDIA Tesla V100 GPU with the following parameters: *N\_cells\_per\_location = 21*, *detection\_alpha = 20*, *max\_epochs = 30000*, *batch\_size = 34000*, *train\_size = 1*. The inferred cell abundances for each spatial location across all samples were exported using *export\_posterior* with *num\_samples = 1000*.

### Defining single-cell mapping-based regions

The 5th quantile of each cell state's abundance distribution (*q05\_cell\_abundance\_w\_sf*) was used as the inferred cell count as per the developer's instructions (Supplementary Fig. S11). These cell abundances were treated as an alternative to gene counts, where the feature space was defined as 26 cell states. We determined that not all cell states were relevant in terms of tissue organization, and set out to define the smallest number of regions that were 1) ubiquitous, i.e. present in all sample classes, and 2) biologically meaningful.

Non-negative matrix factorization (NMF) as implemented in *scikit-learn* v1.1.3 was used to reduce the feature space. Iterative NMF with an increasing number of components (5 to 12) was run, stopping at the highest number of components where each factor had a unique cell state as the highest-contributing feature. The iteration with the highest number of components where no single cell state was the highest contributor in multiple components was chosen as the optimal iteration (*n\_components = 8*, Supplementary Fig. S12). Each ST data point was then categorized according to its highest contributing component, resulting in eight regions across the whole ST dataset (Supplementary Fig. S13).

### Region-specific gene markers

For each sample individually, *scanpy's rank\_genes\_groups* function with *method = 'wilcoxon'* was used to calculate differentially expressed genes between the SCM regions, where the expression of each gene in each region was compared to the baseline expression of that gene in other regions of the sample. Regions with fewer than 10 members and genes of ribosomal or mitochondrial origin were excluded from the analysis. For each region, genes with log fold change  $\geq 1$  and  $p_{\text{adj}} < 0.05$  were considered differentially expressed. A one-sided Fisher's exact test was then used to test whether a gene was enriched as a marker for a specific region. The test was carried out by comparing the number of instances where the gene was differentially expressed in a region of interest against the number of instances where the gene was

differentially expressed in any other region. Genes with  $p_{\text{adj}} < 0.05$  were considered as region-specific markers. A single gene could be considered a region-specific marker for multiple regions.

### Neighborhood enrichment analysis

Sample-specific region neighborhood enrichments were calculated using *squidpy* v1.2.3. First, a neighborhood graph was built for each sample individually using *squidpy.gr.spatial\_neighbors* with parameters *coord\_type = "grid"*, *n\_neigh = 6*, and *n\_rings = 3*. *squidpy.gr.nhood\_enrichment* was then used with default parameters to calculate the enrichment scores. This score is based on a permutation test where the association between regions is estimated by comparing the true region layout to a randomly sampled configuration. Enrichment scores of missing regions were set to 0 in each sample. The mean of all scores was calculated across all samples to get dataset-wide enrichment scores.

### Spot-level gene set activity scoring and gene set enrichment analysis

Scanpy implementation of Seurat's scoring method in *scanpy.tl.score\_genes* was used to score 154 gene sets on the normalized data<sup>82,83</sup>. Parameter values *ctrl\_size: 50*, *n\_bins: 25* were used. No genes were excluded from the control gene pool. Only genes that passed the quality control were included in calculating the score for each sample. A one-sided Fisher's exact test was used to perform gene set enrichment analysis of custom gene sets on the region-specific gene markers. Gene sets with  $p_{\text{adj}} < 0.05$  were considered enriched. g:Profiler was used to perform enrichment analysis of region-specific gene markers on the Gene Ontology: Biological Process database<sup>84,85</sup>.

### Region fraction and gene signature score correlation analysis

Each sample's un-normalized transcript counts were concatenated while leaving out spots that were annotated as the region of interest. Genes expressed in fewer than 10 samples were discarded, after which the data was normalized using *scanpy.pp.normalize\_total* and log-transformed with *scanpy.pp.log1p*. Gene set activity scores were calculated identically to the spot-level scores. A Spearman correlation coefficient and a corresponding p-value were calculated for the Club region fraction and gene set scores in each sample.

### Multiplex immunohistochemistry staining

Formalin-fixed paraffin-embedded prostate tumor tissue blocks from 16 patients were obtained from Tampere University Hospital (Tampere, Finland). The use of clinical material was approved by the Ethics Committee of the Tampere University Hospital and the National Authority for Medicolegal Affairs. For prospective sample collection, written informed consent was obtained from all the subjects. The tumor samples were fixed in 4% phosphate-buffered formaldehyde and processed into paraffin blocks. 5  $\mu\text{m}$  thick sections were stained with in-house multiplex-IHC protocol based on Multiple Iterative Labeling by Antibody Neodeposition (MILAN)<sup>86</sup>. Tissue sections were treated with heat-induced epitope retrieval (HIER) using Tris-HCl buffer (pH 9.0) prior to antibody labeling. Antigens were stained with anti-PIGR 1:500 (Sigma-Aldrich, HPA012012), anti-LTF 1:500 (Sigma-Aldrich, HPA059976), anti-Pan Keratin (AE1/AE3/PCK26) (Roche, 760-2135), anti-CP 1:150 (HPA001834, Sigma-Aldrich), anti-CD66b 1:50 (Novus Biologicals, NB100-77808), anti-CXCR2 1:2000 (Abcam, ab245982), anti-CD45 1:100 (Cell Signaling Technology, #13917) and anti-CD11b 1:100 (Cell Signaling Technology, #49420). Detection was done with fluorescently labeled secondary antibodies Goat anti-Rabbit IgG (H + L) Highly Cross-Adsorbed Secondary Antibody, Alexa Fluor Plus 647 (Invitrogen, A32733), Goat anti-Mouse IgG (H + L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 750 (Invitrogen, A21037). Nuclei were stained with DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride) (Invitrogen, D1306) and slides were mounted using

Fluorescence Mounting Medium (Agilent, S3023). The staining result was scanned using NanoZoomer S60 (Hamamatsu) whole-slide scanner. The fluorescent multiplex-IHC staining was followed by an HE-staining on the same tissue section. HE-staining was done using Leica ST5010 Autostainer XL and the slides were mounted using DPX mountant (Sigma-Aldrich, 44581).

### Multiplex immunohistochemistry image processing

The DAPI staining images were overlaid (registered) utilizing in-house scripts and Python package VALIS (version 1.0.0rc13)<sup>87</sup>. DAPI staining image of LTF-panCK was used as a fixed reference for each staining round. The transformation matrices obtained from these were applied to each matching staining image to obtain the same orientation. Regions of interest (ROI) were obtained by selecting representative areas from the tissue and re-selecting those from subsampled (long side 2048px) images using Qupath (version 0.4.4) TMA dearrayer (width 120px)<sup>88</sup>. These obtained coordinates were transformed to original images to crop full-size ROIs.

Images were processed into hyperstacks using Fiji-ImageJ version v1.53<sup>89</sup>. Image analysis of registered images was performed in Qupath version 0.4.4. Individual cells were segmented from images using Stardist (model dsb2018\_heavy\_augment)<sup>90</sup>. Cell classifier of seven cell categories (Club-like cells (panCK+, LTF+ and/or PIGR+), CP+ Club-like cells (panCK+, CP+), panCK+ cells (panCK+), Granulocytes (CD45+, CD11b+, CD66b+), MDSCs (CD45+, CD11b+, CD66b+, CXCR2+), Other cells (negative to all except DAPI) and Other immune cells (CD45+) (Rtrees) was created based on 26 training images selected from across samples in the staining cohort and applied to a total of 101 ROIs. Classifier results were exported as cell annotation counts and further analyzed in Python.

### Region interface annotation and ligand-receptor analysis

Spatial neighborhood graphs were constructed for each sample as in the neighborhood enrichment analysis. With the Club region selected as the region of interest, the region identities of neighboring spots were surveyed to find another spot of the same annotation. This was done to exclude sporadic individual spots without neighbors of the same region. All non-club spots within 3 rings (300  $\mu$ m) of two or more adjacent Club region spots were then annotated as 'proximal', and all spots further than 3 rings away as 'distant'. The previously calculated gene set scores in these spot categories were then compared across all samples. 6 samples with fewer than 10 Club region-annotated spots were excluded from the analysis. For region-specific interactions (Fig. 4d), the clause that a spot was to have a neighboring spot of the same region identity was included for the interacting region. Additionally, each spot was required to have at least 2 spots from the interacting cluster within three spots distance to be considered as interacting. At least 10 spots from both regions were required in a sample for it to be used in the ligand-receptor interaction analysis. Interfaces between Club and Endothelium were not considered for downstream analysis due to low prevalence (6).

*squidpy*'s implementation of the *CellphoneDB*<sup>91</sup> method was used to calculate ligand-receptor interactions at the sample level. *squidpy.gr.ligrec* was run between interface-labeled spots using parameter settings *complex\_policy='all'*, *threshold=0.01*, and *n\_perms=1000*. Ligand-receptor pairs with  $p_{adj} < 0.05$  were considered active and were used in the overrepresentation analysis. Interactions with fewer than 3 literature references in the *OmniPath*<sup>92</sup> database were discarded.

### Analysis of metastatic CRPC scRNA-seq data

Single-cell gene expression data of mCRPC samples were downloaded from a public repository (<https://singlecell.broadinstitute.org/single-cell/study/SCP1244/transcriptional-mediators-of-treatment-resistance-in-lethal-prostate-cancer>). TPM normalized counts were

matched with the cell identity annotation as defined in the original analysis. Gene expression was visualized using *scanpy.pl.dotplot*.

### Spatial transcriptomics data analysis of metastatic prostate cancer tumors

Three of the four metastatic tissue samples studied were collected as part of the PELICAN integrated clinical-molecular autopsy study of lethal prostate cancer (PELICAN)<sup>93</sup>. Subjects A14 (Met A), A3 (Met C), and A16 (Met D) included in this study provided written informed consent to participate in the Johns Hopkins Medicine IRB-approved study between 1995 and 2005. The mean age of the study subjects at the time of diagnosis of prostate cancer was 61 years.

Sample GP12 (Met B) was collected as a part of the Geoprostate study, where patients newly diagnosed with PrCa electing radical prostatectomy (RP) with 20% or greater preoperative risk of pelvic lymph node metastasis were eligible to participate under Tampere University Hospital Ethics Committee approval R19074 and provided written informed consent to participate in the study<sup>94</sup>.

ST libraries were generated identically to the discovery cohort primary tumor samples generated in Tampere. An identical data pre-processing procedure was followed. Expression clusters were generated using the standard clustering workflow outlined in the *scanpy* (v1.9.1) manual. The data was scaled using *scanpy.pp.scale()*, followed by *scanpy.tl.pca()*, *scanpy.pp.neighbors()*, and *scanpy.tl.leiden()* with a parameter setting *resolution=0.5*. All the other parameters were used in their standard setting. Differential gene expression analysis was performed by running *scanpy.tl.rank\_genes\_groups()* on the *leiden* clusters and with *method=wilcoxon*. Overexpressed genes in each cluster were defined as having  $p_{adj} < 0.05$  (two-sided Wilcoxon rank-sum test) and  $\log_2 \text{fold-change} \geq 1$ . These genes were used as input in the enrichment analysis. Gene set activity scores were calculated for each data point with *sc.tl.score\_genes()* on the normalized but unscaled expression counts.

### Analysis of bulk transcriptomics data

TCGA-PRAD expression data was downloaded from a public repository (<https://tcga.xenahubs.net>). FPKM normalized polyA mRNA expression data of the SU2C-PCF mCRPC cohort was downloaded from a public repository ([https://www.cbioportal.org/study/summary?id=prad\\_su2c\\_2019](https://www.cbioportal.org/study/summary?id=prad_su2c_2019))<sup>95</sup>. *gseapy.gsva* (v1.1.0) was used to calculate GSVAs for each sample<sup>96</sup>. Possible gene overlap was removed when calculating scores for the correlation analysis of two signatures.

### Spatial transcriptomics pseudo-bulk GSVAs and differential expression analysis

Each ST sample's un-normalized transcript counts were concatenated to form pseudo-bulk expression matrices. Gene counts were TPM normalized and *gseapy.gsva* (v1.1.0) was used to calculate GSVAs for each sample. Possible gene overlap was removed when calculating scores for the correlation analysis of two signatures. Differential expression analysis between TRNA and NEADT samples was performed using *pydeseq2* (v0.4.4)<sup>97</sup> with TRNA used as the reference level.

### Statistics and reproducibility

*scipy.stats* v1.9.3 and *statsmodels.stats.multitest* v0.13.5 in python 3.8 was used to perform statistical testing and multiple testing corrections, respectively. The Benjamini-Hochberg method was used to adjust p-values for multiple testing. No statistical method was used to predetermine sample size. No data were excluded from the analysis, except for the spatial transcriptomics data points that did not meet the criteria for minimum number of UMIs or genes. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

## Reporting summary

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

## Data availability

The processed spatial transcriptomics data presented in this study, excluding the validation cohort, have been deposited in the Gene Expression Omnibus (GEO) archive under accession identifier [GSE278936](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE278936). The raw sequencing data are available from the authors, but restrictions apply to the availability of these data. Data can be shared with qualified researchers in accordance with the conditions of ethical approvals and informed consent to use these data in research of prostatic diseases. All handling of these data must be in compliance with GDPR and other relevant data protection regulations upon completion of material transfer agreement with respective data controllers' information. Data access requests will be processed at the earliest convenience. Data access will be granted for one year. Validation cohort spatial transcriptomics data is available in the European Genome-Phenome Archive (EGA) under accession identifier [[EGAD50000000603](https://ega-archive.org/studies/EGAD50000000603)]. The data is available under restricted access to ensure compliance with ethical and legal standards, including GDPR and approval from the Regional Committee for Medical Research in Norway. Access will be granted to researchers who meet these requirements, and they must sign a Data Access Agreement (DAA). To obtain access, contact the Data Access Committee (DAC) at NTNU, who will facilitate the process and provide access through the FEQA Norway node or HUNT Cloud once the DAA is completed. The DAA will be processed at the earliest convenience. The single cell RNA-sequencing datasets used in this study<sup>6–9,12,17,32</sup> are available on GEO under accession numbers [GSE137829](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE137829), [GSE141445](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141445), [GSE176031](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE176031), [GSE185344](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE185344), and [GSE181294](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE181294), the Sequence Read Archive (SRA) under accession number [PRJNA699369](https://www.ncbi.nlm.nih.gov/sra/PRJNA699369) and on [https://singlecell.broadinstitute.org/single\\_cell/study/SCP1244/transcriptional-mediators-of-treatment-resistance-in-lethal-prostate-cancer](https://singlecell.broadinstitute.org/single_cell/study/SCP1244/transcriptional-mediators-of-treatment-resistance-in-lethal-prostate-cancer)<sup>58</sup>. The bulk RNA-sequencing data used in this study is available for download on <https://tcga.xenahubs.net> (TCGA)<sup>9</sup> and on [https://www.cbioportal.org/study/summary?id=prad\\_su2c\\_2019](https://www.cbioportal.org/study/summary?id=prad_su2c_2019)<sup>93</sup>. The remaining data are available within the Article, Supplementary Information or Source Data file. Source data are provided with this paper.

## Code availability

Code used for data analysis in this manuscript is available on <https://github.com/akiviahho/ST-prostate>.

## References

- Attard, G. et al. Prostate cancer. *Lancet* **387**, 70–82 (2016).
- Taylor, B. S. et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
- Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
- Gundem, G. et al. The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).
- Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
- Dong, B. et al. Single-cell analysis supports a luminal-neuroendocrine transdifferentiation in human prostate cancer. *Commun. Biol.* **3**, 778 (2020).
- Chen, S. et al. Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. *Nat. Cell Biol.* **23**, 87–98 (2021).
- Cheng, Q. et al. Pre-existing castration-resistant prostate cancer-like cells in primary prostate cancer promote resistance to hormonal therapy. *Eur. Urol.* **81**, 446–455 (2022).
- Wong, H. Y. et al. Single cell analysis of cribriform prostate cancer reveals cell intrinsic and tumor microenvironmental pathways of aggressive disease. *Nat. Commun.* **13**, 1–21 (2022).
- Berglund, E. et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun.* **9**, 2419 (2018).
- Erickson, A. et al. Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature* **608**, 360–367 (2022).
- Hirz, T. et al. Dissecting the immune suppressive human prostate tumor microenvironment via integrated single-cell and spatial transcriptomic analyses. *Nat. Commun.* **14**, 1–20 (2023).
- Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19**, 1423–1437 (2013).
- Sfanos, K. S., Yegnasubramanian, S., Nelson, W. G. & De Marzo, A. M. The inflammatory microenvironment and microbiome in prostate cancer development. *Nat. Rev. Urol.* **15**, 11–24 (2017).
- Yuan, S., Almagro, J. & Fuchs, E. Beyond genetics: driving cancer with the tumour microenvironment behind the wheel. *Nat. Rev. Cancer* **24**, 274–286 (2024).
- Henry, G. H. et al. A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell Rep.* **25**, 3530–3542.e5 (2018).
- Song, H. et al. Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. *Nat. Commun.* **13**, 141 (2022).
- Wang, X. et al. A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature* **461**, 495–500 (2009).
- Wang, Z. A., Toivanen, R., Bergren, S. K., Chambon, P. & Shen, M. M. Luminal cells are favored as the cell of origin for prostate cancer. *Cell Rep.* **8**, 1339–1346 (2014).
- Joseph, D. B. et al. Urethral luminal epithelia are castration-insensitive cells of the proximal prostate. *Prostate* **80**, 872–884 (2020).
- Germanos, A. A. et al. Defining cellular population dynamics at single-cell resolution during prostate cancer progression. *Elife* **11**, e79076 (2022).
- Manyak, M. J., Kikukawa, T. & Mukherjee, A. B. Expression of a uteroglobin-like protein in human prostate. *J. Urol.* **140**, 176–182 (1988).
- Crowley, L. et al. A single-cell atlas of the mouse and human prostate reveals heterogeneity and conservation of epithelial progenitors. *Elife* **9**, (2020).
- Huang, F. W. et al. Club-like cells in proliferative inflammatory atrophy of the prostate. *J. Pathol.* **261**, 85–95 (2023).
- Karthaus, W. R. et al. Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science* **368**, 497–505 (2020).
- Chan, J. M. et al. Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science* **377**, 1180–1191 (2022).
- Baures, M. et al. Prostate luminal progenitor cells: from mouse to human, from health to disease. *Nat. Rev. Urol.* **19**, 201–218 (2022).
- Lu, S., Fürth, D. & Gillis, J. Integrative analysis methods for spatial transcriptomics. *Nat. methods* **18**, 1282–1283 (2021).
- Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
- Li, B. et al. Benchmarking spatial and single-cell transcriptomics integration methods for transcript distribution prediction and cell type deconvolution. *Nat. Methods* **19**, 662–670 (2022).
- Kleshchevnikov, V. et al. Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
- Chen, Y. et al. Single-cell transcriptomics reveals cell type diversity of human prostate. *J. Genet. Genomics* **49**, 1002–1015 (2022).
- Rubin, M. A. et al.  $\alpha$ -methylacyl coenzyme A racemase as a tissue biomarker for prostate cancer. *JAMA* **287**, 1662–1670 (2002).
- Hessels, D. et al. DD3PCA3-based molecular urine analysis for the diagnosis of prostate cancer. *Eur. Urol.* **44**, 8–16 (2003).
- Shukla, S. et al. Identification and Validation of PCAT14 as Prognostic Biomarker in Prostate Cancer. *Neoplasia* **18**, 489–499 (2016).

36. Baures, M. et al. Transcriptomic signature and growth factor regulation of castration-tolerant prostate luminal progenitor cells. *Cancers* **14**, 3775 (2022).
37. Taavitsainen, S. et al. Single-cell ATAC and RNA sequencing reveal pre-existing and persistent cells associated with prostate cancer relapse. *Nat. Commun.* **12**, 5307 (2021).
38. Tang, F. et al. Chromatin profiles classify castration-resistant prostate cancers suggesting therapeutic targets. *Science* **376**, eabe1505 (2022).
39. Barkley, D. et al. Cancer cell states recur across tumor types and form specific interactions with the tumor microenvironment. *Nat. Genet.* **54**, 1192–1201 (2022).
40. Gavish, A. et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598–606 (2023).
41. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
42. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
43. Calcinotto, A. et al. IL-23 secreted by myeloid cells drives castration-resistant prostate cancer. *Nature* **559**, 363–369 (2018).
44. Alshetaiwi, H. et al. Defining the emergence of myeloid-derived suppressor cells in breast cancer using single-cell transcriptomics. *Sci. Immunol.* **5**, eaay6017 (2020).
45. Guo, C. et al. Targeting myeloid chemotaxis to reverse prostate cancer therapy resistance. *Nature* <https://doi.org/10.1038/s41586-023-06696-z> (2023).
46. Wang, C. et al. CD300ld on neutrophils is required for tumour-driven immune suppression. *Nature* **621**, 830–839 (2023).
47. Han, G. et al. An atlas of epithelial cell states and plasticity in lung adenocarcinoma. *Nature* <https://doi.org/10.1038/s41586-024-07113-9> (2024).
48. Templeton, A. J. et al. Simple prognostic score for metastatic castration-resistant prostate cancer with incorporation of neutrophil-to-lymphocyte ratio. *Cancer* **120**, 3346–3352 (2014).
49. Valero, C. et al. Pretreatment neutrophil-to-lymphocyte ratio and mutational burden as biomarkers of tumor response to immune checkpoint inhibitors. *Nat. Commun.* **12**, 729 (2021).
50. Kasthuber, E. R. & Lowe, S. W. Putting p53 in Context. *Cell* **170**, 1062–1078 (2017).
51. Coppé, J.-P., Desprez, P.-Y., Krtolica, A. & Campisi, J. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu. Rev. Pathol.* **5**, 99–118 (2010).
52. Hanahan, D. Hallmarks of cancer: new dimensions. *Cancer Discov.* **12**, 31–46 (2022).
53. Mempel, T. R., Lill, J. K. & Altenburger, L. M. How chemokines organize the tumour microenvironment. *Nat. Rev. Cancer* **24**, 28–50 (2024).
54. Zhang, Y. et al. Function of the c-Met receptor tyrosine kinase in carcinogenesis and associated therapeutic opportunities. *Mol. Cancer* **17**, 45 (2018).
55. Xiao, T. et al. Targeting EphA2 in cancer. *J. Hematol. Oncol.* **13**, 114 (2020).
56. Liu, B., Qu, L. & Yan, S. Cyclooxygenase-2 promotes tumor growth and suppresses tumor immunity. *Cancer Cell Int.* **15**, 106 (2015).
57. Dankner, M., Gray-Owen, S. D., Huang, Y.-H., Blumberg, R. S. & Beauchemin, N. CEACAM1 as a multi-purpose target for cancer immunotherapy. *Oncoimmunology* **6**, e1328336 (2017).
58. He, M. X. et al. Transcriptional mediators of treatment resistance in lethal prostate cancer. *Nat. Med.* **27**, 426–433 (2021).
59. Kinker, G. S. et al. Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet.* **52**, 1208–1218 (2020).
60. Stein-O'Brien, G. L. et al. Enter the matrix: factorization uncovers knowledge from omics. *Trends Genet.* **34**, 790–805 (2018).
61. Kotliar, D. et al. Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife* **8**, e43803 (2019).
62. Kfoury, Y. et al. Human prostate cancer bone metastases have an actionable immunosuppressive microenvironment. *Cancer Cell* **39**, 1464–1478.e8 (2021).
63. Lu, Y. et al. CXCL16 functions as a novel chemotactic factor for prostate cancer cells in vitro. *Mol. Cancer Res.* **6**, 546–554 (2008).
64. Jung, Y. et al. Recruitment of mesenchymal stem cells into prostate tumours promotes metastasis. *Nat. Commun.* **4**, 1–11 (2013).
65. De Marzo, A. M., Marchi, V. L., Epstein, J. I. & Nelson, W. G. Proliferative inflammatory atrophy of the prostate: implications for prostatic carcinogenesis. *Am. J. Pathol.* **155**, 1985–1992 (1999).
66. van Leenders, G. J. L. H. et al. Intermediate cells in human prostate epithelium are enriched in proliferative inflammatory atrophy. *Am. J. Pathol.* **162**, 1529–1537 (2003).
67. Mani, R. S. et al. Inflammation-induced oxidative stress mediates gene fusion formation in prostate cancer. *Cell Rep.* **17**, 2620–2631 (2016).
68. Liu, X. et al. Low CD38 identifies progenitor-like inflammation-associated luminal cells that can initiate human prostate cancer and predict poor outcome. *Cell Rep.* **17**, 2596–2606 (2016).
69. Attard, G. et al. Clinical testing of transcriptome-wide expression profiles in high-risk localized and metastatic prostate cancer starting androgen deprivation therapy: an ancillary study of the STAMPEDE abiraterone Phase 3 trial. *Res Sq* <https://doi.org/10.21203/rs.3.rs-2488586/v1> (2023).
70. Karthaus, W. R. et al. Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* **159**, 163–175 (2014).
71. Guo, W. et al. Single-cell transcriptomics identifies a distinct luminal progenitor cell type in distal prostate invagination tips. *Nat. Genet.* **52**, 908–918 (2020).
72. Kwon, O.-J., Zhang, L. & Xin, L. Stem cell antigen-1 identifies a distinct androgen-independent murine prostatic luminal cell lineage with bipotent potential. *Stem Cells* **34**, 191–202 (2016).
73. Bian, X. et al. Integration analysis of single-cell multi-omics reveals prostate cancer heterogeneity. *Adv. Sci.* <https://doi.org/10.1002/adv.202305724> (2024).
74. Veglia, F., Sanseviero, E. & Gabrilovich, D. I. Myeloid-derived suppressor cells in the era of increasing myeloid cell diversity. *Nat. Rev. Immunol.* **21**, 485–498 (2021).
75. Gabrilovich, D. I. & Nagaraj, S. Myeloid-derived suppressor cells as regulators of the immune system. *Nat. Rev. Immunol.* **9**, 162–174 (2009).
76. Lehmusvaara, S. et al. Chemical castration and anti-androgens induce differential gene expression in prostate cancer. *J. Pathol.* **227**, 336–345 (2012).
77. Devos, G. et al. ARNEO: A randomized phase II trial of neoadjuvant degarelix with or without apalutamide prior to radical prostatectomy for high-risk prostate cancer. *Eur. Urol.* **83**, 508–518 (2023).
78. Andersen, M. K. et al. Spatial transcriptomics reveals strong association between SFRP4 and extracellular matrix remodeling in prostate cancer. *Commun. Biol.* **7**, 1462 (2024).
79. Deroulers, C. et al. Analyzing huge pathology images with open source software. *Diagn. Pathol.* **8**, 92 (2013).
80. Lun, A. T. L., Bach, K. & Marionni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).
81. Luecken, M. D. et al. Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2021).
82. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
83. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

84. Ashburner, M. et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
85. Kolberg, L. et al. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.* **51**, W207–W212 (2023).
86. Cattoretti, G., Bosisio, F. M., Marcellis L. & Bolognesi M. M. Multiple Iterative Labeling by Antibody Neodeposition (MILAN) PROTOCOL (Version 5) available at Protocol Exchange (2019).
87. Gatenbee, C. D. et al. Virtual alignment of pathology image series for multi-gigapixel whole slide images. *Nat. Commun.* **14**, 4502 (2023).
88. Bankhead, P. et al. QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).
89. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
90. Schmidt, U., Weigert, M., Broaddus, C. & Myers, G. Cell detection with star-convex polygons. arXiv [cs.CV] (2018).
91. Efremova, M., Vento-Tormo, M., Teichmann, S. A. & Vento-Tormo, R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nat. Protoc.* **15**, 1484–1506 (2020).
92. Türeli, D. et al. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol. Syst. Biol.* **17**, e9923 (2021).
93. Jasu, J. et al. Combined longitudinal clinical and autopsy phenomic assessment in lethal metastatic prostate cancer: recommendations for advancing precision medicine. *Eur. Urol. Open Sci.* **30**, 47–62 (2021).
94. Nurminen, A. et al. Cancer origin tracing and timing in two high-risk prostate cancers using multisample whole genome analysis: prospects for personalized medicine. *Genome Med.* **15**, 82 (2023).
95. Abida, W. et al. Genomic correlates of clinical outcome in advanced prostate cancer. *Proc. Natl Acad. Sci.* **116**, 11428–11436 (2019).
96. Fang, Z., Liu, X. & Peltz, G. GSEApY: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023).
97. Muzellec, B., Teleńczuk, M., Cabeli, V. & Andreux, M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics* **39**, btad547 (2023).

## Acknowledgements

The authors acknowledge the Biocenter Finland (BF) and Tampere Genomics Facility for their service. The authors wish to acknowledge CSC–IT Center for Science, Finland, for computational resources. The results published here are in part based upon data generated by The Cancer Genome Atlas project (dbGaP Study Accession: phs000178.v9.p8) established by the NCI and NHGRI. Information about TCGA and the investigators and institutions who constitute the TCGA research network can be found at <http://cancergenome.nih.gov>. Tissue samples of the validation were collected and stored by Biobank1, St. Olav's Hospital. Tissue sectioning, staining, and scanning were performed by or in collaboration with the Histology lab at the Cellular & Molecular Imaging Core Facility at NTNU, while RNA isolation and sequencing were carried out at the Genomics Core Facility at NTNU. We thank Sari Toivola, Päivi Martikainen, Hanna Selin, and Maria Annala for their assistance in laboratory procedures. The authors wish to thank all participating patients and their families. A.K. has received funding from the Tampere University Doctoral School, the Finnish Cultural Foundation, and the Cancer Foundation Finland. C.T.A.P. has received funding from the Jean Shanks Foundation and the Pathological Society Clinical PhD Fellowship. S.T. has been funded by the Tampere University Doctoral School and the Relander Foundation. E.A. has received funding from the Finnish Cultural Foundation. T.M. has received funding from the Finnish Cancer Institute, Helsinki University Hospital State Funding

(VTR), the Sigrid Juselius Foundation, and the Cancer Foundation Finland. M.B.T., M.K.A., S.Krossa, M.B.R., M.W., E.M., G.F.G., and A.S. have collectively benefited from funding provided by the European Research Council (ERC) under the Horizon 2020 program, the Liaison Committee between the Central Norway Regional Health Authority (RHA) and NTNU, the Norwegian Cancer Society, and the Terje Eugen Johnsen funds. K.J.R. has been funded by the Cancer Foundation Finland, the Sigrid Juselius Foundation, the Emil Aaltonen Foundation, the Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital, the Väre Research Foundation, and the Aamu Pediatric Cancer Foundation. A.U. has received support from the Academy of Finland (project no. 349314), the Cancer Foundation Finland, the Norwegian Cancer Society (project no. 198016-2018 & project no. 333723-2023), and the Tampere Institute for Advanced Study. M.N. has been funded by the Academy of Finland Center of Excellence program (project no. 312043), the Cancer Foundation Finland, the Sigrid Juselius Foundation, and the Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital.

## Author contributions

K.J.R., A.U., and M.N. supervised this work. A.K., K.J.R. A.U., and M.N. conceived and designed the analysis. A.K. designed the figure panels and wrote the original manuscript draft. S.K.E., H.M.L.K., M.K.A., M.H., A.M.T., E.M., K.V., and S. Kint performed experimental studies. A.K., A.M.T., I.S., S.T., O.H., M.I., M.W., A.S., S. Krossa, T.H., E.A., J.K., and M.B.R. performed computational analyses and contributed data analysis tools. C.T.A.P., T.T., T. Viset, Ø.S., and T.M. performed histopathology annotation. X.S., A.G., W.D., T.L.J.T., T.M., G.S.B., S.J., J.V.S., T. Visakorpi, G.A., F.C., T. Voet, K.J.R., and M.B.T. contributed data and provided resources for the study. A.K., S.K.E., H.M.L.K., M.K.A., A.G., C.T.A.P., M.M., I.H., G.F.G., M.B.R., T.M., A.E., L.L., G.S.B., I.G.M., S.J., T.M., G.A., F.C., T. Visakorpi, K.J.R., M.B.T., A.U., and M.N. reviewed and edited the manuscript. All authors reviewed and approved the final version of the manuscript.

## Competing interests

C.T.A.P.'s employer may gain commercially from licensing data to Artera AI. G.A. received personal fees, grants, and travel support from Janssen and Astellas Pharma; personal fees or travel support from Pfizer, Novartis/AAA, Bayer Healthcare Pharmaceuticals, AstraZeneca, and Sanofi-Aventis; in addition, G.A.'s former employer, The Institute of Cancer Research, receives royalty income from abiraterone and G.A. receives a share of this income through the Institute's Rewards to Discoverers Scheme. G.A. has received research funding (institutional) from Janssen, Astellas Pharma, and Novartis. All other authors declare no potential conflicts of interest.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-54364-1>.

**Correspondence** and requests for materials should be addressed to Alfonso Urbanucci or Matti Nykter.

**Peer review information** *Nature Communications* thanks Arianna Calcinotto, Yunshun Chen, and Li Xin for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

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

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

© The Author(s) 2024

**Antti Kiviaho** <sup>1,2</sup>, **Sini K. Eerola** <sup>1,2</sup>, **Heini M. L. Kallio** <sup>1,2</sup>, **Maria K. Andersen** <sup>3,4</sup>, **Miina Hoikka** <sup>1,2</sup>, **Aliisa M. Tiihonen** <sup>1,2</sup>, **Iida Salonen** <sup>1,2</sup>, **Xander Spotbeen** <sup>5,6</sup>, **Alexander Giesen** <sup>7,8</sup>, **Charles T. A. Parker** <sup>9</sup>, **Sinja Taavitsainen** <sup>1,2</sup>, **Olli Hantula** <sup>1,2</sup>, **Mikael Marttinen** <sup>1,2,10,11</sup>, **Ismail Hermelo** <sup>1,2</sup>, **Mazlina Ismail** <sup>9</sup>, **Elise Midtbust** <sup>3,4</sup>, **Maximilian Wess** <sup>3,4</sup>, **Wout Devlies** <sup>6,15</sup>, **Abhibhav Sharma** <sup>13</sup>, **Sebastian Krossa** <sup>3,14</sup>, **Tomi Häkkinen** <sup>1,2</sup>, **Ebrahim Afyounian** <sup>1,2</sup>, **Katy Vandereyken** <sup>6,15</sup>, **Sam Kint** <sup>6,15</sup>, **Juha Kesseli** <sup>1,2</sup>, **Teemu Tolonen** <sup>2,16</sup>, **Teuvo L. J. Tammela** <sup>1,17</sup>, **Trond Viset** <sup>18</sup>, **Øystein Størkersen** <sup>18</sup>, **Guro F. Giskeødegård** <sup>4,13</sup>, **Morten B. Rye** <sup>4,19</sup>, **Teemu Murtola** <sup>1,2</sup>, **Andrew Erickson** <sup>20,21,22</sup>, **Leena Latonen** <sup>23</sup>, **G. Steven Bova** <sup>1,2</sup>, **Ian G. Mills** <sup>22,24</sup>, **Steven Joniau** <sup>7,8</sup>, **Johannes V. Swinnen** <sup>5,6</sup>, **Thierry Voet** <sup>5,15</sup>, **Tuomas Mirtti** <sup>20,21,25</sup>, **Gerhardt Attard** <sup>9,26</sup>, **Frank Claessens** <sup>12</sup>, **Tapio Visakorpi** <sup>1,2,27</sup>, **Kirsi J. Rautajoki** <sup>1,2</sup>, **May-Britt Tessem** <sup>3,4</sup>, **Alfonso Urbanucci** <sup>1,2,28,29</sup> ✉ & **Matti Nykter** <sup>1,2,29</sup> ✉

<sup>1</sup>Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. <sup>2</sup>Prostate Cancer Research Center, Tampere University and TAYS Cancer Center, Tampere, Finland. <sup>3</sup>Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway. <sup>4</sup>Clinic of Surgery, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway. <sup>5</sup>Laboratory of Lipid Metabolism and Cancer, KU Leuven and Leuven Cancer Institute (LKI), Leuven, Belgium. <sup>6</sup>KU Leuven Institute for Single Cell Omics (LISCO), KU Leuven, Leuven, Belgium. <sup>7</sup>Department of Urology, University Hospitals Leuven, Leuven, Belgium. <sup>8</sup>Department of Development and Regeneration, KU Leuven, Leuven, Belgium. <sup>9</sup>University College London Cancer Institute, London, UK. <sup>10</sup>European Molecular Biology Laboratory, Structural and Computational Biology Unit, Heidelberg, Germany. <sup>11</sup>European Molecular Biology Laboratory, Genome Biology Unit, Heidelberg, Germany. <sup>12</sup>Molecular Endocrinology Laboratory, Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium. <sup>13</sup>Department of Public Health and Nursing, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. <sup>14</sup>Central staff, St. Olavs Hospital HF, 7006 Trondheim, Norway. <sup>15</sup>Laboratory of Reproductive Genomics, Department of Human Genetics, KU Leuven, Leuven, Belgium. <sup>16</sup>Department of Pathology, Fimlab Laboratories, Ltd, Tampere University Hospital, Tampere, Finland. <sup>17</sup>Department of Urology, Tampere University Hospital, Tampere, Finland. <sup>18</sup>Department of Pathology, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway. <sup>19</sup>Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway. <sup>20</sup>Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland. <sup>21</sup>ICAN-Digital Precision Cancer Medicine Flagship, Helsinki, Finland. <sup>22</sup>Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK. <sup>23</sup>Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland. <sup>24</sup>Patrick G Johnston Centre for Cancer Research, Queen's University of Belfast, Belfast, UK. <sup>25</sup>Department of Pathology, University of Helsinki & Helsinki University Hospital, Helsinki, Finland. <sup>26</sup>University College London Hospitals, London, UK. <sup>27</sup>Fimlab Laboratories, Ltd, Tampere University Hospital, Tampere, Finland. <sup>28</sup>Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway. <sup>29</sup>These authors jointly supervised the work: Alfonso Urbanucci, Matti Nykter.

✉ e-mail: [alfonso.urbanucci@tuni.fi](mailto:alfonso.urbanucci@tuni.fi); [matti.nykter@tuni.fi](mailto:matti.nykter@tuni.fi)