Immunosuppression in the prostate tumor microenvironment is tied to androgen deprivation therapy-resistant club-like cells

Authors

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

Author affiliations

1 Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland
2 Prostate Cancer Research Center, Tampere University and TAYS Cancer Center, Tampere, Finland
3 Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway
4 Clinic of Surgery, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
5 Laboratory of Lipid Metabolism and Cancer, KU Leuven and Leuven Cancer Institute (LKI), Leuven, Belgium
6 KU Leuven Institute for Single Cell Omics (LISCO), KU Leuven, Leuven, Belgium
7 Department of Urology, University Hospitals Leuven, Leuven, Belgium
8 Department of Development and Regeneration, KU Leuven, Leuven, Belgium
9 University College London Cancer Institute, London, UK
10 European Molecular Biology Laboratory, Structural and Computational Biology Unit
11 European Molecular Biology Laboratory, Genome Biology Unit
12 Department of Public Health and Nursing, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
13 Central staff, St. Olavs Hospital HF, 7006, Trondheim, Norway
14 Department of Pathology, Fimlab Laboratories, Ltd, Tampere University Hospital, Tampere, Finland
15 Department of Urology, Tampere University Hospital, Tampere, Finland
16 Department of Pathology, St. Olav's Hospital, Trondheim University Hospital, Trondheim, Norway

17 Department of Clinical and Molecular Medicine, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
18 Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland
19 ICAN-Digital Precision Cancer Medicine Flagship, Helsinki, Finland
20 Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK
21 Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland
22 Patrick G Johnston Centre for Cancer Research, Queen's University of Belfast, Belfast, UK
23 Laboratory of Reproductive Genomics, Department of Human Genetics, KU Leuven, Leuven, Belgium
24 Department of Pathology, University of Helsinki & Helsinki University Hospital, Helsinki, Finland
25 University College London Hospitals, London, UK
26 Molecular Endocrinology Laboratory, Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium
27 Fimlab Laboratories, Ltd, Tampere University Hospital, Tampere, Finland
28 Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway

* These authors jointly supervised the work

Abstract

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 110 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 partake in inducing myeloid inflammation previously associated with androgen deprivation therapy resistance, providing a rationale for their therapeutic targeting.

Main

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[1]. Genomics and transcriptomics data from clinical samples have revealed the various aberrations underlying primary[2,3] and treatment-resistant tumors[4,5]. 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[6–9]. 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[10], malignant cells in histologically benign areas[11], and immunosuppressive regulatory T-cell activity in coordination with myeloid cells[12], 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[13–15].

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[16,17]. 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[18,19]. Hillock cells are a source of discrepancy as they have been claimed to be found[9,12,16,20] or to be undiscernible[17], 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[7,21].

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*[16,22]. Initial observations noted that club cells reside in the urethral epithelium and proximal prostate ducts[20,23]. In contrast, more recent work has highlighted their presence in the peripheral zone and inflamed regions of the prostate[24]. Intriguingly, human club cells have a transcriptomic profile similar to the castration-resistant luminal progenitor phenotype in mice[25–27], raising questions regarding their function in response to androgen deprivation therapy (ADT) in humans.

We set out to investigate the cellular composition of the prostate TME and its response to androgen deprivation. We generated 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 110 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 1,835 tissue-covered spots, 3,275 median genes per spot, and 9,549 median UMIs per spot on average (**Supplementary Table 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, **Figure 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 µm 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[28,29]. A commonly adopted analysis strategy uses scRNA-seq-derived transcriptomics profiles to infer the most probable cell type compositions in each ST spot[30]. These inferred cell type counts can further be used to divide tissue structures into distinct, biologically meaningful regions[31].

We re-analyzed previously published scRNA-seq datasets from normal prostate tissue[12,32], low- and high-grade primary tumors[7,9,12,17], and both locally recurrent and metastatic castration-resistant prostate cancer tumors[6,8] 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 Table S2**). We used this cell state reference to model the most probable cell state composition in spatial locations of our discovery cohort[31].

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 Table 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, **Figure 1b**).

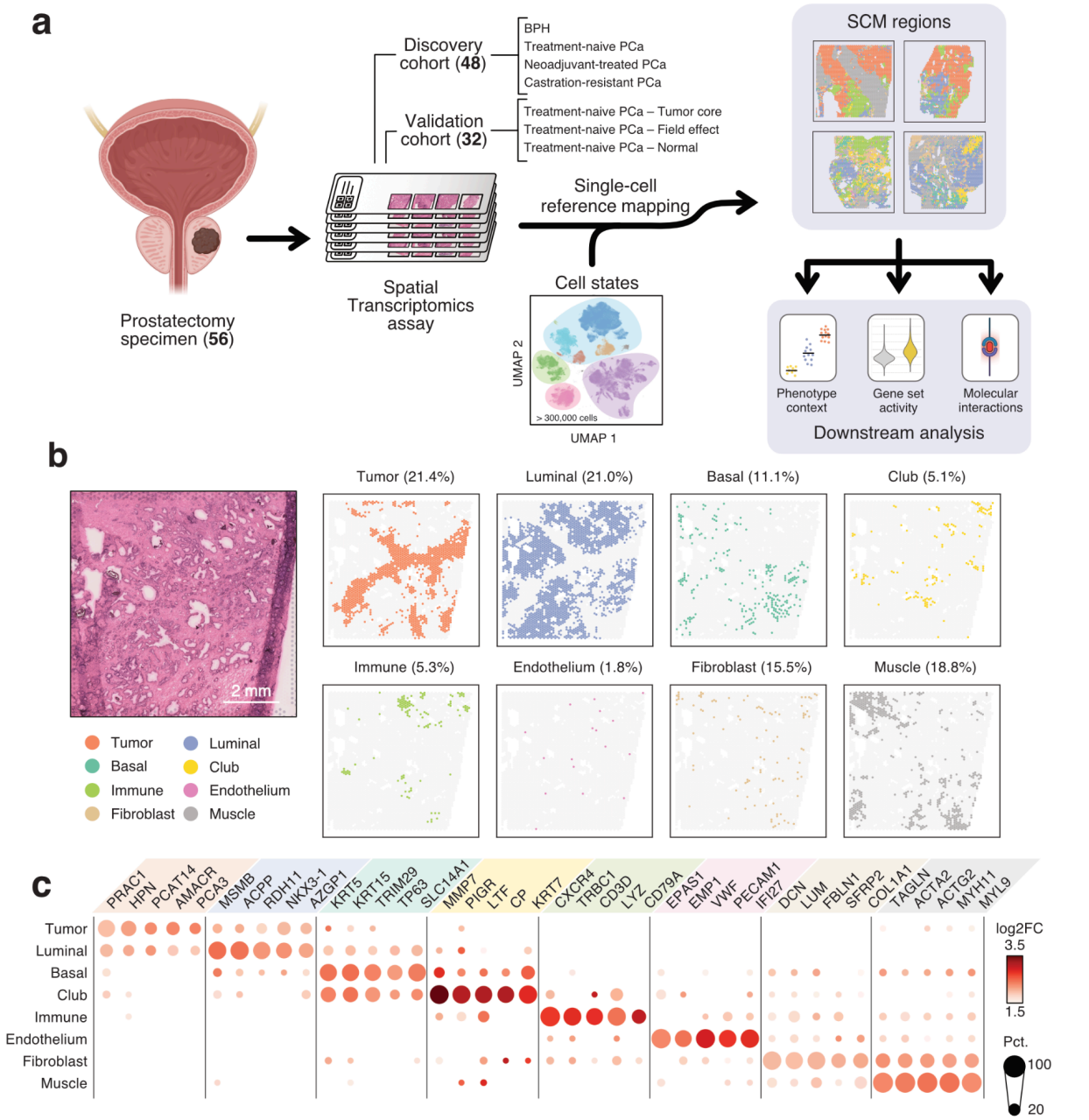

Figure 1: *Integration of scRNA-seq and ST data reveals the organizational patterns of the prostate TME.* ***a)*** *Sample collection and analysis pipeline overview. Brackets indicate the number of samples in each category.* ***b)*** *An untreated primary tumor sample with eight SCM regions shown separately. Percentages represent the share of spots across the discovery cohort* ***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* $p_{adj} < 0.05$ *&* $\log_2$ *fold change ≥ 1). Dot color represents the average log-fold change. SCM regions: single-cell mapping-based regions; log2FC: log2 fold-change.*

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 (**Methods, Supplementary Table S4**). We then determined genes that were enriched among the overexpressed genes in each region ($p_{adj} < 0.05$, one-sided Fisher's exact test), terming these *region-specific markers* (**Supplementary Table S5)**.

We annotated the SCM regions according to their region-specific markers (**Figure 1c**). The *Tumor* region overexpressed known PCa-specific markers such as *AMACR*[33], *PCA3*[34], and *PCAT14*[35]. 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*, *C1QA*, *C1QB*). Stromal regions *Endothelium* (*VWF*, *EPAS1*, *EMP1*), *Fibroblast* (*DCN*, *FBLN1*, *LUM*), and *Muscle* (*ACTA2*, *TAGLN*, *MYL9*) were similarly annotated. One SCM region overexpressed recently proposed markers for club cells[17,24] *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 *EGR1* 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 Figure S1**).

SCM regions agreed with the pathologists' assessment of sample histology (**Figure 2a**). To quantitively 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 Figure S1**). *Low cancer %* samples had a lower fraction of Tumor region spots than *Mid cancer %* ($p = 1.2\times10^{-4}$, two-sided Wilcoxon rank-sum test) and *High cancer %* samples ($p = 1.1\times10^{-4}$). *High cancer %* samples had lower Basal region fraction than *Mid cancer %* ($p = 3.5\times10^{-3}$) or *Low cancer %* ($p = 1.4\times10^{-4}$) samples. The fraction of Luminal ($p = 1.6\times10^{-2}$, Kruskal-Wallis test) and Muscle ($p = 2.9\times10^{-4}$) 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 ($p < 2.2\times10^{-16}$, two-sided Fisher's exact test). 67% of the spots deemed normal by the pathologists' were Luminal, Basal, or Club region spots (**Figure 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 Figure 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 ($p < 0.05$, two-sided Wilcoxon rank-sum test, **Figure 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.

The expression of canonical club cell marker genes was high in the Club region both pre- and post-treatment. Mouse luminal progenitor marker genes[36] *WFDC2*, *KRT19*, *KLF5*, *ATP1B1*, and *KRT4* were upregulated in the Club region as a response to ADT ($p < 0.05$, two-sided Wilcoxon rank-sum test). These genes were also markers for the Club region (**Supplementary Table S5**).

The percentage of Tumor, Luminal, Basal, Immune, Endothelium, and Fibroblast spots varied across the treatment categories (**Figure 2d**). TRNA ($p = 7.2\times10^{-3}$, two-sided Wilcoxon rank-sum test) and NEADT ($p = 4.5\times10^{-3}$) samples had a higher percentage of tumor spots than BPH samples. NEADT samples contained a lower percentage of Luminal region than TRNA samples ($p = 9.2\times10^{-4}$), while a higher proportion of Basal region ($p = 4.4\times10^{-4}$). These results indicate an increase in the fraction of basal phenotype epithelial cells in response to ADT.

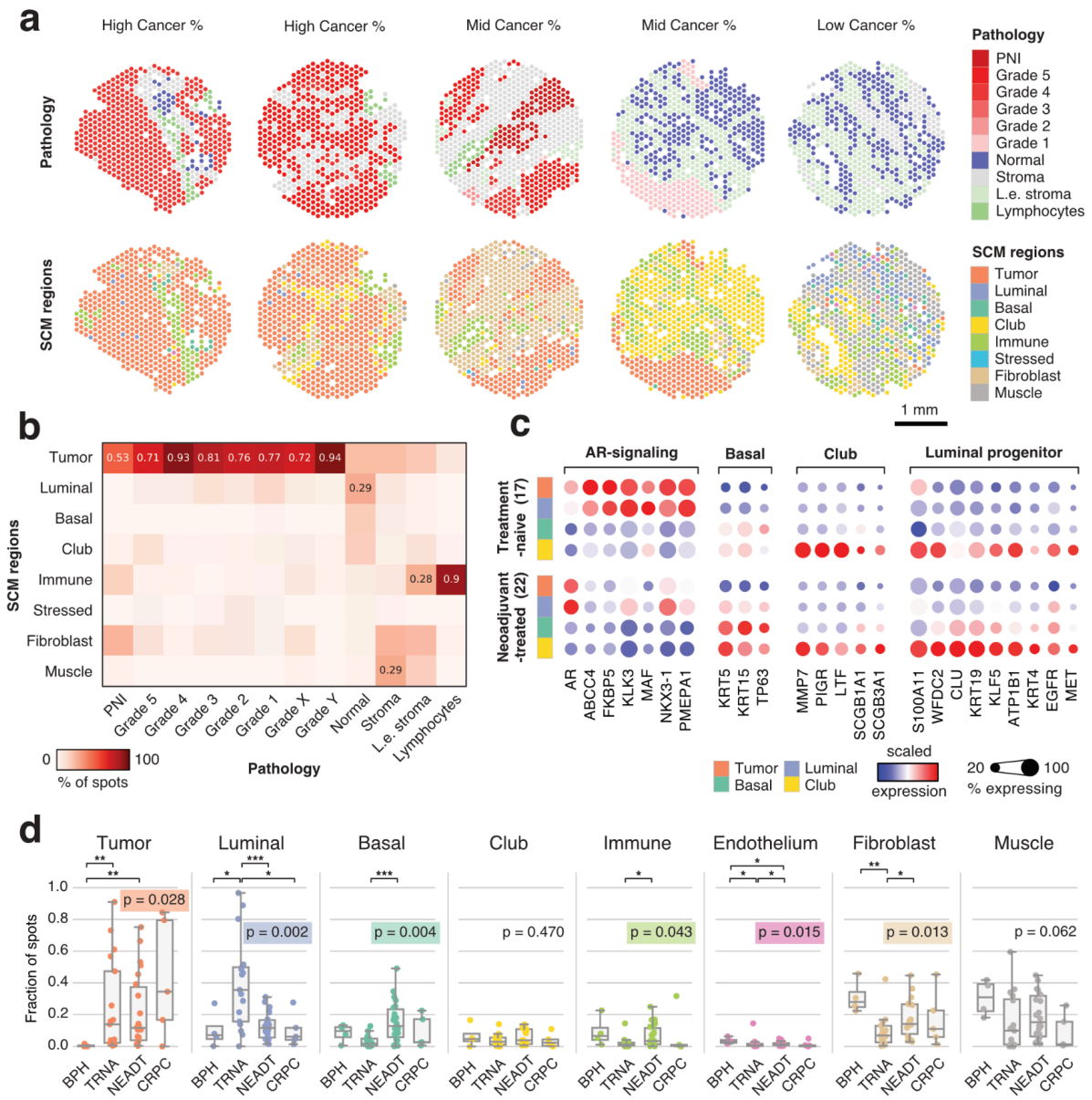

Figure 2. *SCM regions match histopathological features and reflect treatment-dependent shifts in gene expression.* ***a)*** *Histopathology classes and SCM regions of five representative samples from the validation cohort.* ***b)*** *Percentage of SCM regions in each histopathology category in the validation 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. Kruskal-Wallis test p-value is displayed for each category, while asterisks indicate Wilcoxon rank-sum test significance level; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. Grade X and Grade Y are cancer spots with indecisive grading (see Methods). PNI: perineural invasion; L.e. stroma: lymphocyte-enriched stroma; BPH: Benign prostatic hyperplasia; TRNA: Treatment-naïve; NEADT: Neoadjuvant-treated; CRPC: Castration-resistant prostate cancer*

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[16,37,38], pan-cancer associated[39,40], hallmark signaling[41,42], and immune activity-related gene sets[12,43–46] and compared the scores across epithelial cell regions (**Supplementary Table S6**). *Club cell*[16], *epithelial senescence* (*EpiSen*)[40], *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, **Figure 3a**). Genes related to androgen response scored lower in the Club region, while the mouse *luminal progenitor*[36], *lung club cell*[47], *stem cell-like CRPC*[38], and *PROSGenesis*[37] signatures scored higher (**Supplementary Figure S3**). 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 Figure S3**).

The expression of *high NLR-associated* genes in CRPC primary tumors correlates with the NLR measured from peripheral blood[45]. High blood NLR is associated with shorter survival and treatment resistance[45,48,49]. 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 (**Figure 3b**), despite no overlap between the two signatures (**Figure 3c**). We consolidated gene set activity scores with an enrichment analysis between region-specific gene markers and these gene sets (**Supplementary Figure S4**, **Supplementary Table S7**). Genes of the *high NLR-associated* signature were overrepresented among Club region markers ($p_{adj} = 1.5\times10^{-3}$, one-sided Fisher's exact test). One of the genes in *high NLR-associated* signature, *NFKB1*, codes for nuclear factor κB (*NF-κB*). *TNF-α-signaling via NF-κB* was among the most significantly enriched gene sets for the Club region ($p = 2.7\times10^{-26}$). The p53 pathway ($p = 2.3\times10^{-7}$) hallmark gene set was uniquely enriched in the Club region, in line with its putative senescent nature[50].

We hypothesized that the club cells display senescence-associated secretory phenotype (SASP)[51]. 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[52]. We examined the expression of key chemokines across all SCM regions in a pre- and post-treatment setting (**Figure 3d**). Neutrophil chemotaxis-inducing chemokines *CXCL1* (p_{adj} =

1.5×10^{-13}, one-sided Fisher's exact test), *CXCL2* ($p_{adj} = 7.4\times10^{-13}$), and *CXCL8* ($p_{adj} = 1.7\times10^{-5}$) were region-specific markers for the Club region, along with *CXCL16* ($p_{adj} = 1.5\times10^{-13}$) and *CCL20* ($p_{adj} = 2.0\times10^{-4}$). *CCL20* was expressed at a higher level in the NEADT Club region compared to the TRNA Club region ($p_{adj} = 1.9\times10^{-2}$, two-sided Wilcoxon rank-sum test). Canonical receptors of these chemokines[52] were region-specific markers in the Immune region: *CXCR6* ($p_{adj} = 2.7\times10^{-2}$) for *CXCL16* and *CCR6* ($p_{adj} = 4.1\times10^{-4}$) for *CCL20*. While the sensitivity of ST was not sufficient to find *CXCR2* expression (no transcripts in 99.6% of all spots), the abundant expression of inflammatory chemokines indicated active neutrophil recruitment.

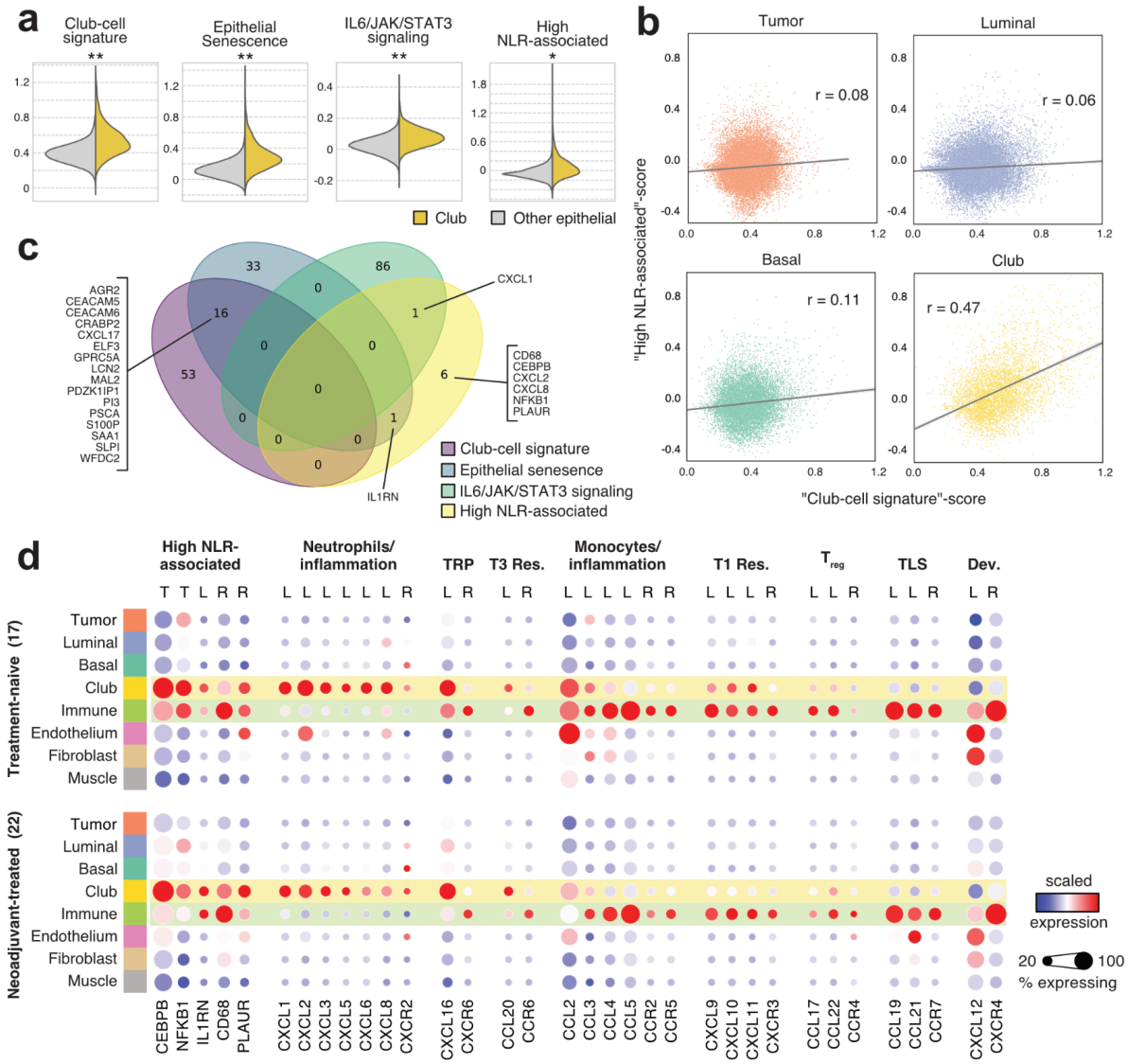

Figure 3. *The Club region has increased expression of genes linked to a senescence-associated secretory phenotype* ***a)*** *Spotwise gene set activity scores in the Club region compared to other epithelial regions. Asterisks indicate two-sided independent samples t-test p-value (p < 0.05) and effect 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 overlaps of gene sets shown in a.* ***d)*** *Scaled gene expression in SCM regions divided by ADT exposure. 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*

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 (**Figure 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[43] ($p = 4.3\times10^{-7}$) and human[12] ($p = 2.0\times10^{-6}$) were also enriched in the Immune region (**Supplementary Table S7**). A CRISPR-Cas9 validated signature of polymorphonuclear MDSC (PMN-MDSC) immunosuppression[46] was enriched in the Immune region ($p = 1.9\times10^{-2}$), indicating an accumulation of immunosuppressive myeloid cells in the prostate TME.

The fraction of the Club region in a sample correlated positively with its PMN-MDSC activity score calculated from non-Club region spots in both cohorts (**Figure 4b**). Tumor region spots had a higher PMN-MDSC activity score when proximal to Club region spots (**Figure 4c**). Similar results were acquired for three other MDSC signatures (**Supplementary Figure S5**). One of the MDSC activity signatures scored higher in the recurrent samples ($n = 20$) than in non-recurrent samples of the validation cohort ($n = 12$) ($p = 1.2\times10^{-4}$, two-sided Wilcoxon rank-sum test). Pseudobulk-level PMN-MDSC activity scores didn't differ between the TRNA and NEADT groups ($p \geq 0.05$, two-sided independent samples t-test, **Figure 4h**).

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 (**Figure 4d, Supplementary Table S8**). 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.

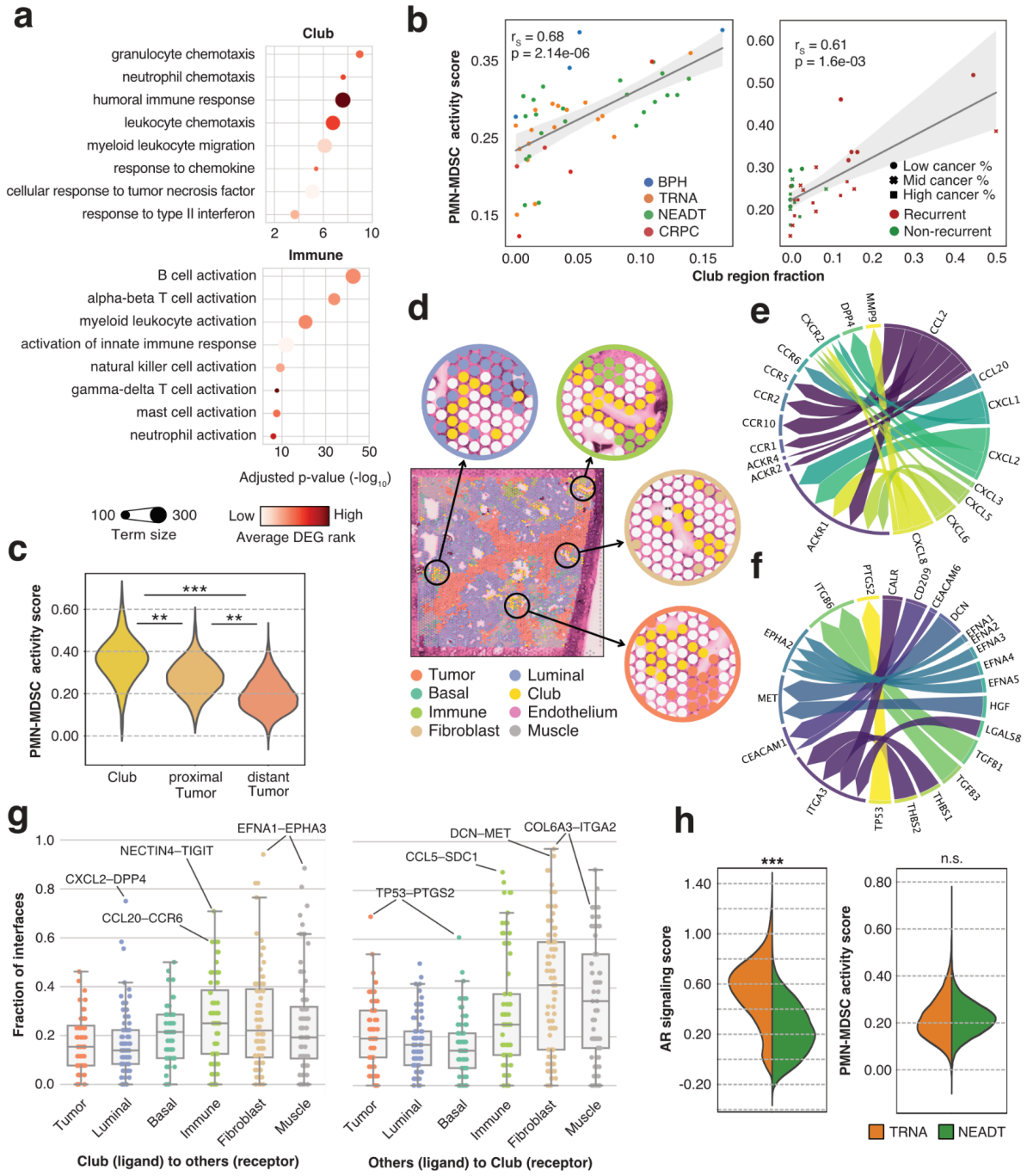

Figure 4. *The Club region is associated with increased myeloid-derived suppressor cell activity.* ***a)*** *GO:BP term enrichments for genes with upregulated expression in the Club (top) and Immune (bottom) regions.* ***b)*** *Correlation between Club region fraction and PMN-MDSC activity scores in pseudo-bulk for the discovery (left) and the validation (right) cohorts. Spearman correlation coefficient and p-value are shown.* ***c)*** *PMN-MDSC activity score in 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.* ***d)*** *Illustration of region interfaces for the Club region against Luminal, Immune, Fibroblast, and Tumor regions in a single sample.*

e) *Ligand-receptor signaling activity for ligands expressed in the Club region. Edge weights represent the number of interfaces where a ligand-receptor interaction was active, colors represent unique ligands.* ***f)*** *Same as e), but for receptors expressed in the Club region.* ***g)*** *Ligand-receptor interaction prevalence on Club region interfaces.* ***h)*** *Violin plots of gene set activity scores across ST spots irrespective of region. Violins are split according to treatment category. TRNA: treatment-naive; NEADT: neoadjuvant-treated; CRPC: Castration-resistant prostate cancer*

This analysis identified 142 unique ligand-receptor interactions occurring at the Club region interface, 59 (41.5%) of which were enriched at specific interfaces ($p_{adj} < 0.05$, one-sided Fisher's exact test). *Atypical chemokine receptor 1 (ACKR1)* was the most common target for neutrophil chemotaxis-inducing chemokines *CXCL1, 2, 5, 6,* and *8* (**Figure 4e**). The *CXCL2* to *ACKR1* interaction was enriched at the Club-Fibroblast interface ($p_{adj} = 2.3\times10^{-2}$). *Dipeptidyl peptidase-4* (*DPP4*) was a unique target of *CXCL2* and enriched at the Club-Luminal interface ($p_{adj} = 5.7\times10^{-8}$, **Figure 4g**). *CCL20* and *CCR6* formed an interaction that was enriched in the Club-Immune interface ($p_{adj} = 2.2\times10^{-3}$).

The Club region overexpressed receptors with known oncogenic properties *(MET*, *EPHA2*)[53,54], inflammatory activities (*CEACAM1*, *PTGS2*)[55,56], and adhesive function (*ITGA3*, *ITGA6*) (**Figure 4f, Figure 4g**). The Fibroblast-Club interface was enriched for *MET* interaction with *decorin* (*DCN,* $p_{adj} = 1.0\times10^{-6}$) and *hepatocyte growth factor* (*HGF,* $p_{adj} = 5.6\times10^{-5}$). *EPHA2* was the target of multiple *ephrin*-family ligands, of which *EFNA1* and *EFNA5* signaling interactions were enriched at the Luminal ($p_{adj} = 1.0\times10^{-6}$) and Muscle ($p_{adj} = 2.4\times10^{-2}$) interfaces, respectively. *TP53* and *PTGS2* interaction was enriched at the Tumor-Club interface ($p_{adj} = 3.0\times10^{-2}$). An interaction between Immune region-expressed *CCL5* and Club region-expressed *SDC1* was the single most enriched interaction between these two regions ($p_{adj} = 5.9\times10^{-5}$). These results elucidate the region-specific signaling relied upon by the club 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 ($p = 3.2\times10^{-22}$, one-sided Fisher's exact test), as well as the *EpiSen* signature and PMN-MDSC activity signature ($p = 2.1\times10^{-5}$, **Figure 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[57], suggesting that similar senescence-driven processes could be taking place in the metastatic tumor environment **(Figure 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 Table S9)**. We determined expression-based clusters for each sample separately to find spots with similar expression and underlying cell populations (**Figure 5c, Supplementary Figure S6**). We then performed enrichment analysis and calculated gene set activity scores to determine whether any of these clusters corresponded to the club cell population identified in primary tumors.

Cluster 1 in *Met A* (liver metastasis) and 5 in *Met B* (subdural metastasis) displayed elevated scores for the Club region markers and the *club-like senescence* signature (**Figure 5d**). The corresponding sites in the tissue overexpressed genes that were enriched for Club region markers (*Met A1* $p_{adj} = 3.2\times10^{-7}$ and *Met B5* $p_{adj} = 1.3\times10^{-13}$, one-sided Fisher's exact test) and *PMN-MDSC activity* markers (*Met A1* $p_{adj} = 3.2\times10^{-2}$, *Met B5* $p_{adj} = 2.3\times10^{-2}$, **Figure 5e**, **Supplementary Table S9**). Genes overexpressed in the *Met B1* and *Met B2* clusters were similarly enriched for the Club region markers ($p_{adj} = 3.4\times10^{-7}$ and $p_{adj} = 1.5\times10^{-4}$, 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 overexpressed in *Met B1* ($\log_2$ fold-change = 1.06, $p_{adj} = 4.5\times10^{-24}$) and *Met B5* ($p_{adj} = 8.8\times10^{-21}$). *PIGR* was similarly overexpressed in *Met B5* ($p_{adj} = 1.2\times10^{-2}$). Other Club region markers overexpressed in *Met B5* included *CLDN1* ($p_{adj} = 7.8\times10^{-13}$), *SCNN1A* ($p_{adj} = 2.9\times10^{-6}$), and *TACSTD2* ($p_{adj} = 8.5\times10^{-61}$), all of which are exclusively expressed in epithelial cells. These results suggest the presence of club-like cells in metastatic prostate cancer tumors.

C*lub-like senescence* and *PMN-MDSC activity* signature GSVA scores correlated across untreated primary and mCRPC tumors in the TCGA-PRAD and SU2C-PCF cohorts (**Figure 5g, 5h**). The score of *club-like senescence* signature also correlated with two other MDSC signatures, and the *high NLR-associated* signature (**Supplementary Figure S7**). *Club-like senescence* score correlated negatively with AR-signaling in both TCGA-PRAD ($r_s = -0.30$, $p = 1.2\times10^{-12}$, two-sided Spearman correlation test) and SU2C-PCF ($r_s = -0.28$, $p = 5.0\times10^{-6}$) cohorts.

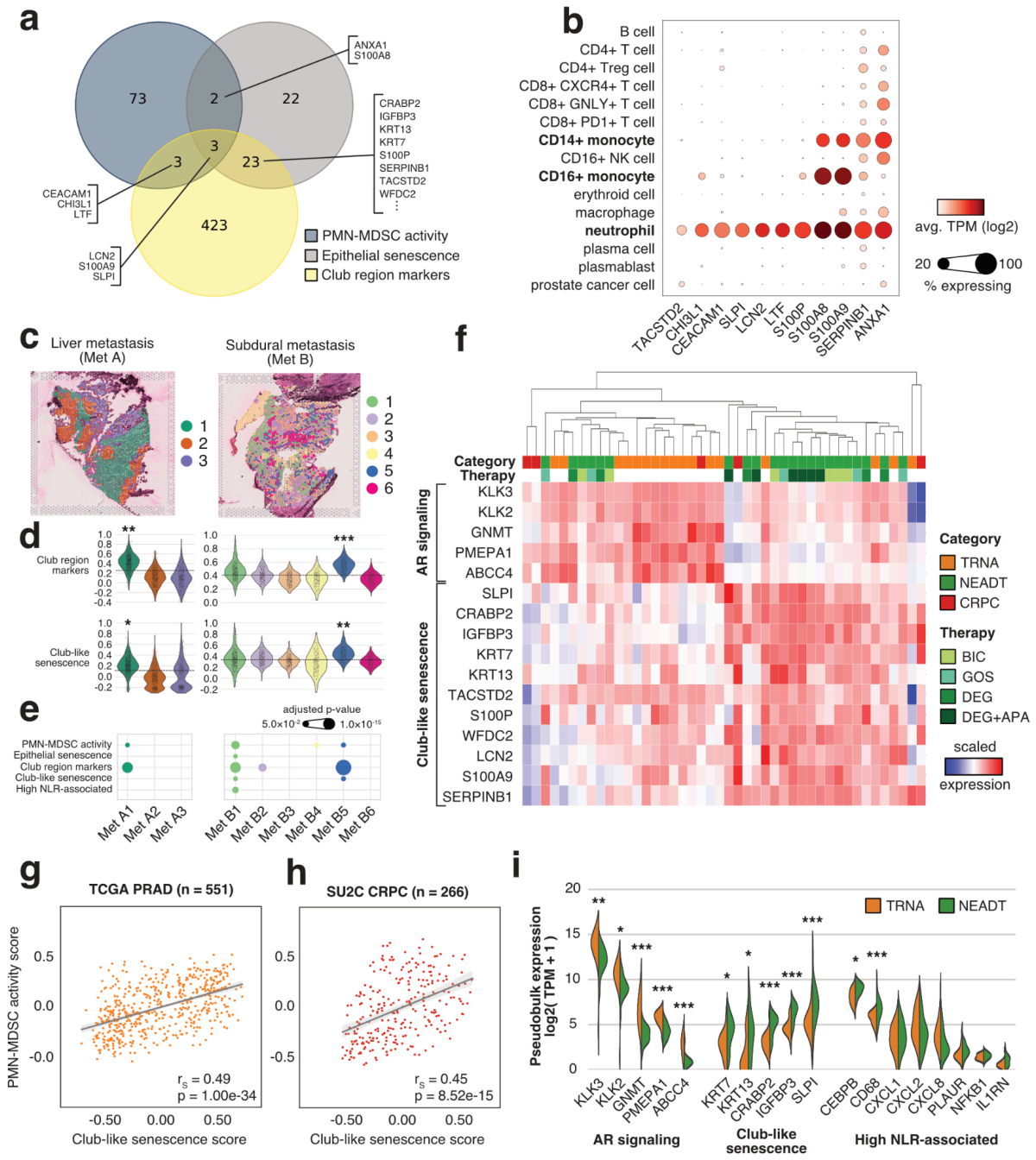

Figure 5. *Club-like senescence is associated with immunosuppressive PMN-MDSC activity in primary and metastatic tumors.* ***a)*** *Venn diagram showing overlaps between PMN-MDSC activity, epithelial senescence, and club-region upregulated gene sets.* ***b)*** *Dot plot of normalized gene expression in metastatic castration-resistant prostate cancer samples. Data and clusters as reported in He et al. 2021.* ***c)*** *Expression-based clustering overlaid on two prostate cancer metastasis ST samples.* ***d)*** *Cluster-specific gene set scores. 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 30th percentile > overall median; **: cluster 20th percentile > overall median; *: cluster 10th percentile > overall median)* ***e)***

Log-transformed overrepresentation p_{adj} *among each clusters overexpressed genes.* ***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 and SU2C cohorts.* ***i)*** *Violin plot of normalized gene expression in pseudo-bulk spatial transcriptomics samples. Asterisks indicate differential gene expression test significance levels between treatment-naïve and neoadjuvant-treated samples (*: p<0.05, **: p<0.01, ***: p<0.001). PMN-MDSC: polymorphonuclear myeloid-derived suppressor cell; TRNA: treatment-naïve, NEADT: neoadjuvant-treated; CRPC: castration-resistant prostate cancer; BIC: bicalutamide; GOS: goserelin; DEG: degarelix; APA: apalutamide*

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 (**Figure 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 (**Figure 5i, Supplementary Table S10**). None of the genes in these signatures were downregulated post-treatment.

Discussion

We used spatial transcriptomics to explore the microenvironment of benign, untreated, neoadjuvant-treated, and castration-resistant prostate tumors. We employed a novel bioinformatics approach 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[39,40,58] 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[59,60], resulting in an unbiased cell-state reference.

Here we report that club-like cells display a pronounced SASP expression profile. The upregulated expression of chemokines *CXCL1*, *CXCL2*, *CXCL8*, *CXCL16*, and *CCL20* is indicative of SASP[61] and highlights the myeloid-interactive nature of these cells. In a prostate-specific context, the expression of *CXCL1*, *CXCL2*, and *CXCL8* is independently predictive of poor overall survival in metastatic CRPC patients[45]. High *CCL20* expression in the prostate is associated with more aggressive disease[62] and its blockade slows tumor progression in mice[12]. *In vitro* studies have shown that *CXCL16* expression induces tumor cell and mesenchymal stem cell migration, promoting cancer metastasis to the bone[63,64]. 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*[64].

Increased cytokine signaling activity in the Club region is in line with reports that club-like cells are enriched in proliferative inflammatory atrophy (PIA)[24,65,66]. PIA has been regarded as a potential precursor for prostate cancer[65], with DNA damage resulting from cytokine-induced oxidative stress reported as one possible mechanism[14,67]. We found increased expression of PIA gene markers[66] *KRT5*, *KRT8*, and *MET*, and PIA-associated club-like cell gene markers[17,24] *CP*, *MMP7*, *LTF*, and *PIGR* in the Club region. Inflammation-related molecular signaling routes *IL6/JAK/STAT3*, *IFNγ response*, and *TNFα signaling via NFκB* 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[9]. Inflammation-associated low-CD38 expressing luminal cells have previously been associated with biochemical recurrence[68], whereas high IFNγ signaling is

associated with worse outcomes in high-risk localized disease[69]. 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[27]. Luminal progenitor cells are an inherently castration-resistant cell population in the mouse prostate[25,70] that have also been proposed as a possible cellular origin of prostate cancer[71]. These cells are mainly present in the proximal region of the prostatic duct, with scattered cells found in the distal lobes[20,23]. The previously described transcriptomic similarity of the mouse progenitor and human club cells is supported by our data[36]. Our analysis also shows that the Club region is unperturbed by ADT, indicating that human club 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[72].

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 (PMN-MDSCs)[14,46,73]. We show that club-like cells are associated with PMN-MDSC activity in primary tumors and that club-like cells can be found in metastatic tumors. Immunosuppressive PMN-MDSCs are a driver of CRPC that can reactivate androgen signaling through IL-23 secretion[43]. In general, tumor sites contain expanded PMN-MDSC populations compared to healthy tissue[73,74], and the inhibition of their chemotactic receptor *CXCR2* can reverse castration resistance[45].

Taken together, our analysis highlights club-like cells as an underappreciated TME constituent that is associated with myeloid inflammation and immunosuppressive PMN-MDSC activity. The combination of their inherent castration resistance and PMN-MDSC chemotaxis-inducing capability implicates them as a possible contributor to treatment resistance and warrants further research into their function in tumor initiation and progression.

Methods

Sample collection

A total of 48 prostatectomy samples were collected for the discovery cohort. 37 freshly frozen prostate tissue samples from 37 patients of different disease stages, including benign prostatic hyperplasia (BPH, n = 4), treatment-naïve (n = 11) and neoadjuvant-treated primary PCa (n = 17), and locally recurrent castration-resistant prostate cancer (CRPC, n = 5) tissue samples from radical prostatectomies were obtained from Tampere University Hospital (TAUH, Tampere, Finland) and utilized for the analysis. 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, 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 a part of the ARNEO trial (ClinicalTrials.gov ID NCT03080116). 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 Table S1**).

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 non-cancerous morphology close to the cancerous region (Field effect), and one sample with non-cancerous 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 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 µm thickness to the Visium spatial gene expression slide, fixed in ice-cold 100% methanol for 30 min at -20 °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 µm; 1–10 cells) as recommended by the manufacturer.

For the validation cohort samples, 10 µm-thick tissue sections were cut at -20°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°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.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[75].

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*[76]. 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*[77,78].

Single-cell reference mapping

The scRNA-seq cell-state reference created from publicly available data (**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 Figure S8**). 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 smallest 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 Figure S9**). Each ST data point was then categorized according to its highest contributing component, resulting in eight regions across the whole ST dataset (**Supplementary Figure S10)**.

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 ≥ 1 and $p_{adj} < 0.05$ 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 $p_{adj} < 0.05$ 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 153 gene sets on the normalized data[79,80]. 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 $p_{adj} < 0.05$ were considered enriched. g:Profiler was used to perform enrichment analysis of region-specific gene markers on the Gene Ontology: Biological Process database[81,82].

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.

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 μm) 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 (Figure 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*[83] 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*[84] 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)[85]. Subjects A14 (Met A), A3 (Met C), and A16 (Met D) included in this study consented 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[86].

ST libraries were generated identically to the discovery cohort primary tumor samples generated in Tampere. An identical data preprocessing 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 *fold-change* ≥ 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://xenabrowser.net/datapages/?dataset=survival%2FPRAD_survival.txt&host=https%3A%2F%2Ftcga.xenahubs.net&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443) 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)[87]. *gseapy.gsva* (v1.1.0) was used to calculate GSVA scores for each sample[88]. Possible gene overlap was removed when calculating scores for the correlation analysis of two signatures.

Spatial transcriptomics pseudo-bulk GSVA 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 GSVA scores 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)[89] with TRNA used as the reference level.

Statistics

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.

Acknowledgments

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.

Funding

Antti Kiviaho

Tampere University Doctoral School, Finnish Cultural Foundation; Cancer Foundation Finland

Sinja Taavitsainen

Tampere University Doctoral School; Relander Foundation

Charles T.A. Parker

Jean Shanks Foundation; Pathological Society Clinical PhD Fellowship

Tuomas Mirtti

Finnish Cancer Institute; Helsinki University Hospital State Funding (VTR); Sigrid Juselius Foundation; Cancer Foundation Finland

May-Britt Tessem, Maria K. Andersen, Sebastian Krossa, Morten B. Rye, Maximillian Wess, Elise Midtbust, Guro F. Giskeødegård and Abhibhav Sharma

The generation of the validation cohort was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 758306), the Liaison Committee between the Central Norway Regional Health Authority (RHA) and NTNU, the Norwegian Cancer Society and Terje Eugen Johnsen funds.

Kirsi Rautajoki

Cancer Foundation Finland, Sigrid Jusélius Foundation, Emil Aaltonen Foundation, Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital,

Väre Research foundation, Aamu Pediatric Cancer Foundation

Alfonso Urbanucci

Academy of Finland project no. 349314; Cancer Foundation Finland; Norwegian Cancer Society project no. 198016-2018 & project no. 333723-2023, Tampere Institute for Advanced Study.

Matti Nykter

Academy of Finland Center of Excellence programme (project no. 312043), Cancer Foundation Finland, Sigrid Juselius Foundation, Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital.

Competing interests

Charles T.A. Parker

Artera, Inc., Los Altos CA USA

References

1. Attard, G. *et al.* Prostate cancer. *Lancet* **387**, 70–82 (2016).
2. Taylor, B. S. *et al.* Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
3. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011–1025 (2015).
4. Gundem, G. *et al.* The evolutionary history of lethal metastatic prostate cancer. *Nature* **520**, 353–357 (2015).
5. Robinson, D. *et al.* Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
6. Dong, B. *et al.* Single-cell analysis supports a luminal-neuroendocrine transdifferentiation in human prostate cancer. *Commun Biol* **3**, 778 (2020).
7. 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).
8. 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).
9. 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).
10. Berglund, E. *et al.* Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun.* **9**, 2419 (2018).
11. Erickson, A. *et al.* Spatially resolved clonal copy number alterations in benign and malignant tissue. *Nature* vol. 608 360–367 Preprint at https://doi.org/10.1038/s41586-022-05023-2 (2022).
12. 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).

13. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19**, 1423–1437 (2013).
14. 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).
15. Yuan, S., Almagro, J. & Fuchs, E. Beyond genetics: driving cancer with the tumour microenvironment behind the wheel. *Nat. Rev. Cancer* (2024) doi:10.1038/s41568-023-00660-9.
16. Henry, G. H. *et al.* A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell Rep.* **25**, 3530–3542.e5 (2018).
17. 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).
18. Wang, X. *et al.* A luminal epithelial stem cell that is a cell of origin for prostate cancer. *Nature* **461**, 495–500 (2009).
19. 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).
20. Joseph, D. B. *et al.* Urethral luminal epithelia are castration-insensitive cells of the proximal prostate. *Prostate* **80**, 872–884 (2020).
21. Germanos, A. A. *et al.* Defining cellular population dynamics at single-cell resolution during prostate cancer progression. *Elife* **11**, e79076 (2022).
22. Manyak, M. J., Kikukawa, T. & Mukherjee, A. B. Expression of a uteroglobin-like protein in human prostate. *J. Urol.* **140**, 176–182 (1988).
23. Crowley, L. *et al.* A single-cell atlas of the mouse and human prostate reveals heterogeneity and conservation of epithelial progenitors. *Elife* **9**, (2020).

24. Huang, F. W. *et al.* Club-like cells in proliferative inflammatory atrophy of the prostate. *J. Pathol.* **261**, 85–95 (2023).
25. Karthaus, W. R. *et al.* Regenerative potential of prostate luminal cells revealed by single-cell analysis. *Science* **368**, 497–505 (2020).
26. Chan, J. M. *et al.* Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling. *Science* **377**, 1180–1191 (2022).
27. Baures, M. *et al.* Prostate luminal progenitor cells: from mouse to human, from health to disease. *Nat. Rev. Urol.* **19**, 201–218 (2022).
28. Lu, S., Fürth, D. & Gillis, J. Integrative analysis methods for spatial transcriptomics. *Nature methods* vol. 18 1282–1283 (2021).
29. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
30. 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).
31. Kleshchevnikov, V. *et al.* Cell2location maps fine-grained cell types in spatial transcriptomics. *Nat. Biotechnol.* **40**, 661–671 (2022).
32. Chen, Y. *et al.* Single-cell transcriptomics reveals cell type diversity of human prostate. *J. Genet. Genomics* **49**, 1002–1015 (2022).
33. Rubin, M. A. *et al.* α-Methylacyl Coenzyme A Racemase as a Tissue Biomarker for Prostate Cancer. *JAMA* **287**, 1662–1670 (2002).
34. Hessels, D. *et al.* DD3PCA3-based Molecular Urine Analysis for the Diagnosis of Prostate Cancer. *Eur. Urol.* **44**, 8–16 (2003).
35. 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**, (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. U. S. A.* **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**, (2020).
45. Guo, C. *et al.* Targeting myeloid chemotaxis to reverse prostate cancer therapy resistance. *Nature* 1–3 (2023) doi:10.1038/s41586-023-06696-z.
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* 1–8 (2024) doi:10.1038/s41586-024-07113-9.
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. Kastenhuber, E. R. & Lowe, S. W. Putting p53 in Context. *Cell* **170**, 1062–1078 (2017).
51. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12**, 31–46 (2022).
52. Mempel, T. R., Lill, J. K. & Altenburger, L. M. How chemokines organize the tumour microenvironment. *Nat. Rev. Cancer* **24**, 28–50 (2024).
53. Zhang, Y. *et al.* Function of the c-Met receptor tyrosine kinase in carcinogenesis and associated therapeutic opportunities. *Mol. Cancer* **17**, 45 (2018).
54. Xiao, T. *et al.* Targeting EphA2 in cancer. *J. Hematol. Oncol.* **13**, 114 (2020).
55. Liu, B., Qu, L. & Yan, S. Cyclooxygenase-2 promotes tumor growth and suppresses tumor immunity. *Cancer Cell Int.* **15**, 106 (2015).
56. 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).
57. He, M. X. *et al.* Transcriptional mediators of treatment resistance in lethal prostate cancer. *Nat. Med.* **27**, 426–433 (2021).
58. Kinker, G. S. *et al.* Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet.* **52**, 1208–1218 (2020).
59. Stein-O'Brien, G. L. *et al.* Enter the Matrix: Factorization Uncovers Knowledge from Omics. *Trends Genet.* **34**, 790–805 (2018).
60. Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife* **8**, e43803 (2019).
61. 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).

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* (2023) doi:10.21203/rs.3.rs-2488586/v1.

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. Bian, X. *et al.* Integration Analysis of Single-Cell Multi-Omics Reveals Prostate Cancer Heterogeneity. *Adv. Sci.* e2305724 (2024) doi:10.1002/advs.202305724.

73. 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).

74. Gabrilovich, D. I. & Nagaraj, S. Myeloid-derived suppressor cells as regulators of the immune system. *Nat. Rev. Immunol.* **9**, 162–174 (2009).

75. Andersen, M. K. *et al.* Spatial transcriptomics reveals strong association between SFRP4 and extracellular matrix remodeling in prostate cancer. *Research Square* (2023) doi:10.21203/rs.3.rs-2922325/v1.

76. Deroulers, C. *et al.* Analyzing huge pathology images with open source software. *Diagn. Pathol.* **8**, 92 (2013).

77. Lun, A. T. L., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75 (2016).

78. Luecken, M. D. *et al.* Benchmarking atlas-level data integration in single-cell genomics. *Nat. Methods* **19**, 41–50 (2021).

79. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

80. 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).

81. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).

82. 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).

83. 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).

84. Türei, D. *et al.* Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol. Syst. Biol.* **17**, e9923 (2021).

85. 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).
86. 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).
87. Abida, W. *et al.* Genomic correlates of clinical outcome in advanced prostate cancer. *Proceedings of the National Academy of Sciences* **116**, 11428–11436 (2019).
88. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, (2023).
89. Muzellec, B., Teleńczuk, M., Cabeli, V. & Andreux, M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics* **39**, (2023).