Androgen deprivation therapy-resistant club cells are linked to myeloid cell-driven immunosuppression in the prostate tumor microenvironment

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], Leena Latonen[18], G. Steven Bova[1,2], Ian G. Mills[19,20], Steven Joniau[7,8], Johannes V. Swinnen[5,6], Tuomas Mirtti[21,22,23], Gerhardt Attard[9,24], Frank Claessens[25], Tapio Visakorpi[1,2,26], Kirsi J. Rautajoki[1,2], May-Britt Tessem[3,4], Alfonso Urbanucci[1,2,27], 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 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 Institute of Biomedicine, University of Eastern Finland, Kuopio, Finland
19 Nuffield Department of Surgical Sciences, University of Oxford, Oxford, UK
20 Patrick G Johnston Centre for Cancer Research, Queen's University of Belfast, Belfast, UK
21 Department of Pathology, University of Helsinki & Helsinki University Hospital, Helsinki, Finland
22 Research Program in Systems Oncology, Faculty of Medicine, University of Helsinki, Helsinki, Finland
23 iCAN–Digital Precision Cancer Medicine Flagship, Helsinki, Finland
24 University College London Hospitals, London, UK
25 Molecular Endocrinology Laboratory, Cellular and Molecular Medicine, KU Leuven, Leuven, Belgium
26 Fimlab Laboratories, Ltd, Tampere University Hospital, Tampere, Finland
27 Department of Tumor Biology, Institute for Cancer Research, Oslo University Hospital, Oslo, Norway

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. We identify prostate club cells as a key prostate epithelial cell subtype that acts as an interface between the prostate and the immune system. Club cell-enriched tissue areas have depleted androgen signaling, increased epithelial senescence and inflammatory signaling, and a resemblance to a luminal progenitor subtype found in castrated mice. Club cells are associated with increased polymorphonuclear myeloid-derived suppressor cell (PMN-MDSC) activity that has recently emerged as a viable therapeutic target. We propose that the club cells are central to understanding and overcoming prostate cancer therapy resistance.

Introduction

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 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 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 tumors[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 likeness 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], whereas 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 and its treatment-induced changes in prostate TME. We generated spatial transcriptomics (ST) data from benign prostatic hyperplasia, treatment-naive 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 prostatectomy specimens in total. Our study provides an exhaustive map of the prostate TME, ranging from benign tissue to treatment-naive 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 data were divided into two groups, constituting a discovery and a validation cohort (**Figure 1a)**. One section per patient was included in the discovery cohort, while 32 sections in the validation cohort were produced from 8 patients. We used the discovery cohort, consisting of treatment-naive 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 to cross-compare results and test hypotheses arising from the discovery cohort.

Sample-independent deconvolution of spatial transcriptomics data using 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 structure into distinct, biologically meaningful regions[31]. We hypothesized that by inferring the cell type counts simultaneously across multiple ST sections, we could create a sample-agnostic feature space that could be used to find biologically similar regions across the data.

To achieve truly sample-independent cell type deconvolution, we set out to build a comprehensive single-cell atlas of healthy and diseased prostate. 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 cell-type reference that captured a similar phenotype distribution as our discovery cohort. Using unsupervised clustering and non-negative matrix factorization (NMF) to define shared intra-tumor heterogeneity 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].

The most abundant cell states were fibroblasts, myocytes, and luminal epithelial cells, together accounting for 55.1% of the total inferred number of cells (**Supplementary Table S3**). The inferred abundance of tumor cells per spot varied across sample categories ($p < 0.05$; one-way ANOVA). A high inferred number of tumor cells was found in TRNA (average weight 5.7) and CRPC (5.7) samples, while lower in NEADT (3.4) and BPH (0.9) samples. 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. To address this, we categorized individual spots into eight categories based on their majority cell state contributors. Briefly, we performed iterative NMF with an increasing number of components, stopping at the highest number of components where each factor had a unique cell state as the highest-contributing feature (**Methods**). This resulted in commonly co-localized cell populations, hereon called single-cell mapping-derived regions (SCM-regions, **Figure 1b**).

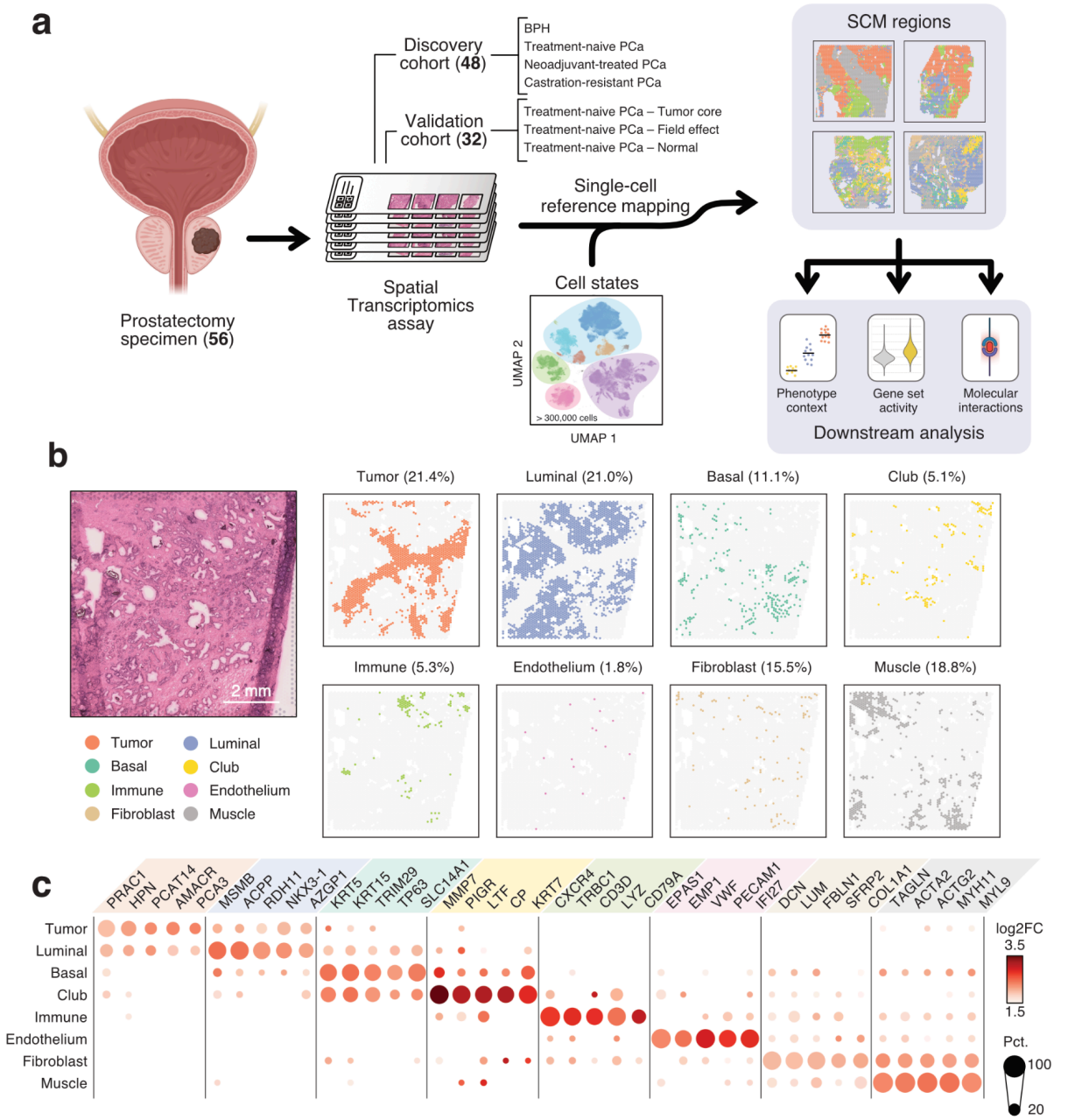

Figure 1. *Integration of scRNA-seq and ST data reveals the organizational patterns of the prostate TME.* ***a)*** *Sample collection and analysis pipeline overview. Brackets indicate the number of samples in each category.* ***b)*** *An untreated primary tumor sample with eight SCM regions shown separately. 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 adjusted p-value < 0.05 & log 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. To find differentially expressed genes between these regions, while controlling for inter-sample variance and the relatively poor coverage of ST, we adopted the following strategy. First, differentially expressed genes were calculated between the regions within each sample to determine genes that were highly expressed in a region (DEGs, two-sided Wilcoxon rank-sum test, adjusted p-value < 0.05, log-fold change ≥ 1, **Supplementary Table S4**). We then performed enrichment analysis to determine genes that were overrepresented as DEGs in each of the regions (one-sided Fisher's exact test, adjusted p-value < 0.05). We termed these genes as region-specific marker genes (**Supplementary Table S5)**.

We interrogated the region-specific marker genes to identify expression programs previously described in the prostate (**Figure 1c**). Prostate cancer-associated genes *AMACR*[33], *PCA3*[34], and *PCAT14*[35] were markers for the Tumor region. Canonical luminal cell markers *MSMB, ACPP,* and *CD38* were markers for the Luminal region, similar to KRT5, KRT15, and TP63 in the Basal region. The SCM region annotated as the Immune region contained upregulated expression of 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). Three stromal regions, namely Endothelium (VWF, EPAS1, EMP1), Fibroblast (DCN, FBLN1, LUM), and Muscle (ACTA2, TAGLN, MYL9) were likewise named according to the expression of canonical gene markers. One SCM region contained overexpression of 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. While the Endothelium region identified in the discovery cohort was not detected, a novel region expressing *FOS, JUN*, and *EGR1* was identified and named the *Stressed luminal* region. Both Endothelium and the Stressed regions were the smallest in their cohorts representing 1.8% and 2.2% of all spots, respectively.

Samples in the validation cohort were categorized as low (cancer % < 5%, 11 samples), mid (5% ≤ cancer % < 50 %, 11 samples), and high (50% ≤ cancer %, 10 samples) cancer content according to the pathologists' annotation (**Supplementary Figure S1**). This categorization was recapitulated by the SCM regions, as *Low cancer %* samples had a lower fraction of Tumor region spots than *Mid cancer %* ($p = 1.2e^{-4}$, two-sided Wilcoxon rank-sum test) and *High cancer %* samples ($p = 1.1e^{-4}$). Similarly, *High cancer %* samples had lower Basal region fraction than

Mid cancer % ($p = 3.5e^{-3}$) or *High cancer %* ($p = 1.4e^{-4}$) samples. The fraction of Luminal ($p = 1.6e^{-2}$, Kruskal-Wallis test) and Muscle ($p = 2.9e^{-4}$) regions also varied across sampling locations while the fraction of Club, Immune, Stressed, and Fibroblast did not. Although the pipeline used in producing the SCM regions was naive to spatial location, the neighborhood of each spot was enriched for spots of the same region (**Supplementary Figure S1**). The overall percentage of Club region spots (5.1% in discovery, 7.4% in validation) was in line with the previously reported 5% prevalence for club cells in the prostate[27].

SCM regions agreed with the pathologists' assessment of sample histology (**Figure 2a**). In the validation cohort, spots annotated as ISUP grade groups were enriched for the Tumor region ($p = 2.6e^{-270}$, two-sided Fisher's exact test). 29% of the spots deemed normal by the pathologists' were Luminal region spots, and an additional 38% were either 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 epithelial phenotypes

ADT and modern androgen signaling inhibitors (ARSIs) are a staple of prostate cancer treatment. Our discovery cohort contained 22 NEADT samples treated with goserelin, bicalutamide, degarelix, and a combination of degarelix & apalutamide (**Methods**). To understand how the SCM regions were affected by ADT, we analyzed the gene expression of the epithelial SCM regions in pre- and post-treatment samples. We observed a 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 lowly expressed in the Basal and Club regions in pre- and post-treatment samples. Conversely, canonical basal cell markers KRT5, KRT15, and TP63 remained highly expressed in the Basal and Club regions both pre- and post-treatment. These genes were expressed at a higher level in the Luminal, Basal, and Club region spots of NEADT samples than in TRNA samples ($p < 0.05$, two-sided Wilcoxon rank-sum test).

LTF was the only club cell marker gene whose expression in the Club region decreased in response to treatment ($p = 2.0e^{-46}$, two-sided Wilcoxon rank-sum test). Human orthologs of mouse luminal progenitor marker genes[36] were overexpressed in the Club region, and their expression in the Club region was upregulated as a response to ADT. WFDC2, CLU, KRT19, KLF5, ATP1B1, and KRT4 were all expressed at a higher level in the NEADT Club region than

in the TRNA ($p < 0.05$, two-sided Wilcoxon rank-sum test). Of these genes all but CLU were also markers for the Club region (**Supplementary Table S5**).

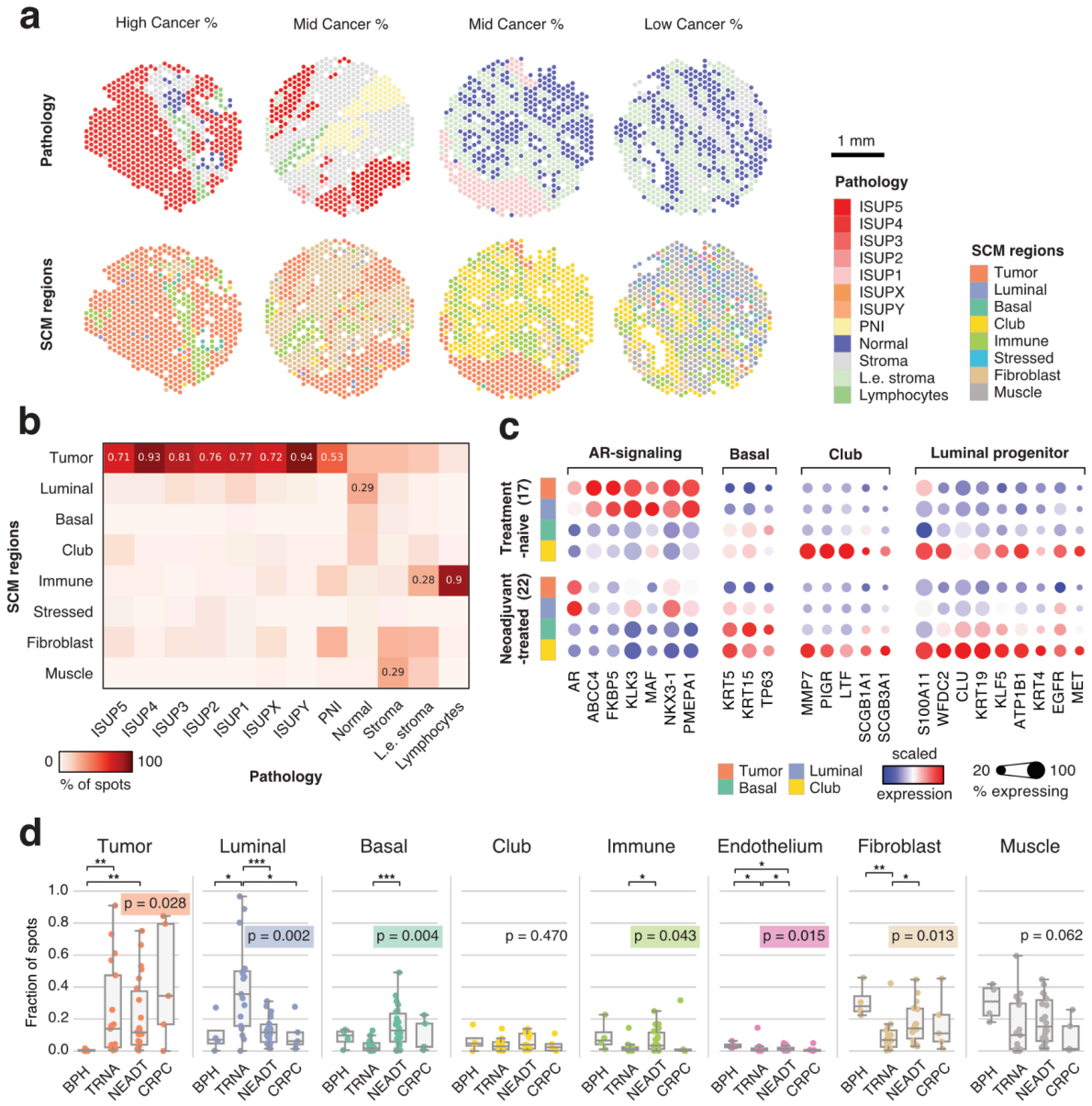

Figure 2. *SCM regions match histopathological features and reflect treatment-dependent shifts in gene expression.* ***a)*** *Histopathology classes and SCM regions of four representative samples in the validation cohort* ***b)*** *Percentage of SCM regions in each histopathology category in the validation cohort. Values are shown for each region with the highest percentage.* ***c)*** *Scaled gene expression in SCM regions divided 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; *: $p < 0.05$; **: $p < 0.01$; ***: $p < 0.001$. PNI: perineural invasion; L.e. stroma: lymphocyte-enriched stroma; BPH: Benign prostatic hyperplasia; TRNA: Treatment-naive; NEADT: Neoadjuvant-treated; CRPC: Castration-resistant prostate cancer*

The effect of ADT was also reflected in the fraction of spots belonging to each region, as the percentage of Tumor, Luminal, Basal, Immune, Endothelium, and Fibroblast spots varied across the treatment categories ($p < 0.05$, Kruskall-Wallis test, **Figure 2d**). TRNA ($p = 7.2e^{-3}$) and NEADT samples had a higher percentage of tumor spots than BPH samples ($p = 4.5e^{-3}$, two-sided Wilcoxon rank-sum test). A lower percentage of Luminal region was present in the NEADT samples than in the TRNA samples ($p = 9.2e^{-4}$), while an opposite effect was observed for the Basal region ($p = 4.4e^{-4}$). This difference in region abundance, combined with the increased expression of basal cell markers, reflects 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 further elucidate the different roles performed by the SCM regions in the prostate TME, we calculated spot-level gene set activity scores on a host of biologically relevant gene sets (**Supplementary Table S6**). These included prostate-specific[16,37,38], pan-cancer associated[39,40], hallmark signaling[41,42], and immune activity-related gene sets[12,43–46]. As the fraction of Club region spots persisted throughout treatment and club cell function is not well understood, we initially compared spot-wise activity scores between the Club region and other epithelial regions. Expectedly, the originally described *club cell* signature scored higher in the Club region than in other epithelial regions (**Figure 3a**). The Club region also scored higher for *epithelial senescence* (EpiSen) as defined in a pan-cancer study of transcriptomics heterogeneity[40], *IL6/JAK/STAT3 signaling*, and high neutrophil-to-lymphocyte ratio associated (*high NLR-associated*) gene sets ($p < 0.05$, two-sided independent samples t-test & additional quantile thresholds). Gene sets 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 in the Club (**Supplementary Figure S3**). Multiple inflammation-related hallmark gene signatures scored higher in the Club region, including *inflammatory response*, *interferon-gamma*, *interferon-alpha*, and *KRAS signaling UP*.

The increased activity of high NLR-associated genes in the Club region was deemed particularly interesting. This set of eight genes has been described in late-stage CRPC patient samples, where it correlates with the neutrophil-to-lymphocyte ratio 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**). There was no overlap between the two signatures (**Figure 3c**).

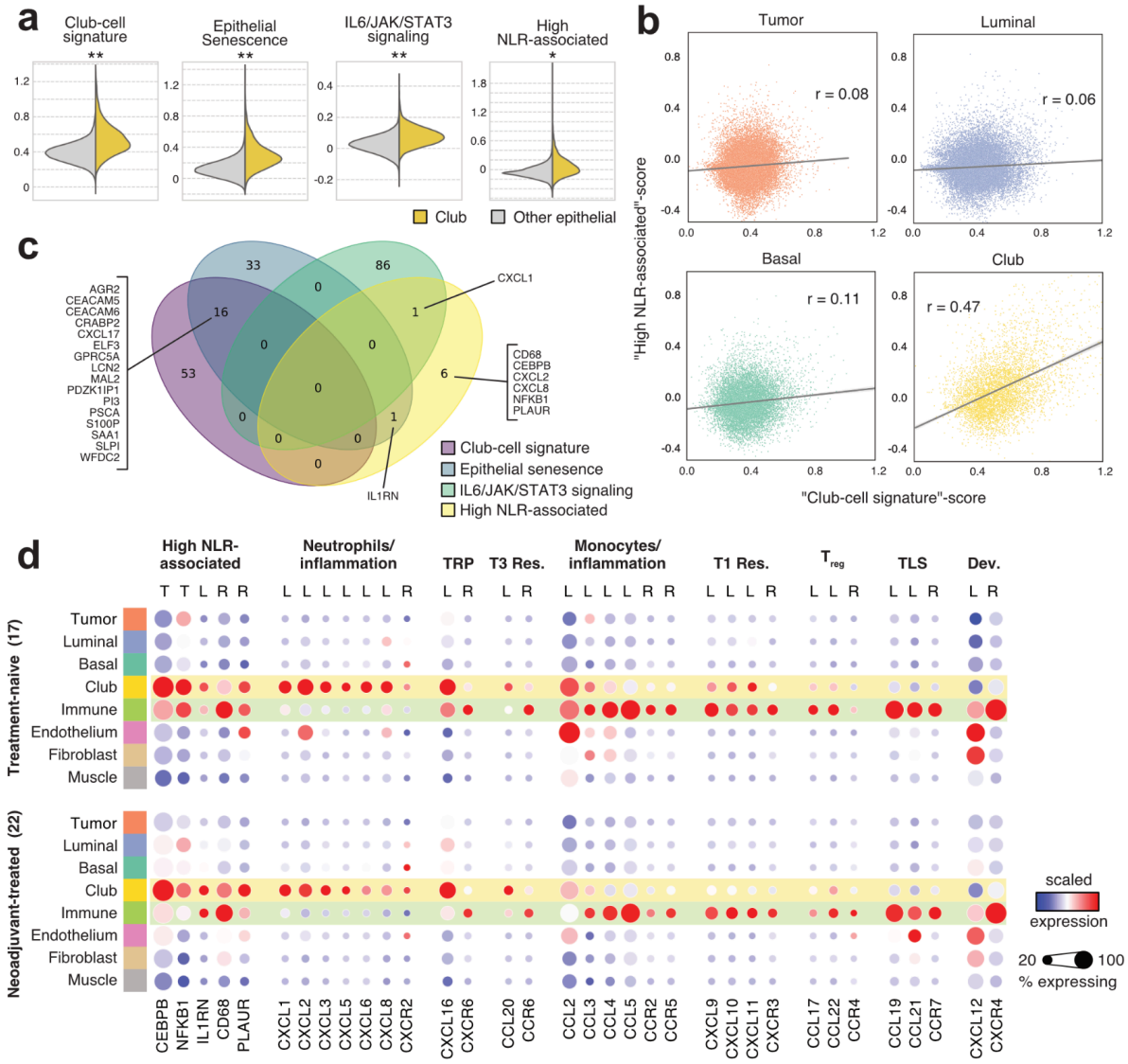

Figure 3. *The Club region has increased expression of genes linked to a senescence-associated secretory phenotype* ***a)*** *Spotwise gene set activity scores in the Club region compared to other epithelial regions. Asterisks indicate two-sided independent samples t-test p-value ($p < 0.05$) and effect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

We consolidated the scoring results by performing enrichment analysis between the gene sets and region-specific gene markers (**Supplementary Table S7**). Genes of the *high NLR-associated* signature were overrepresented in the list of Club region marker genes (adjusted $p = 1.5e^{-3}$, one-sided Fisher's exact test). Interestingly, *TNF-α-signaling via NF-κB* ($p = 2.7e^{-26}$) was among the most significantly enriched gene sets for the Club region, while the *NFKB1* gene coding for the *nuclear factor NF-κB* is included in the high NLR-associated signature. The p53 pathway ($p = 2.3e^{-7}$) hallmark gene set was uniquely enriched in the Club region, supportive of the supposed senescent phenotype of this region[50]. Overall, the enrichment analysis recapitulated observations made based on the spot-wise scoring results, while extending the comparison across all regions and better accounting for data sparsity (**Supplementary Figure S4**).

Increased senescent and inflammatory gene expression activity indicates the presence of senescence-associated secretory phenotype (SASP) cells in the Club region[51]. SASP is particularly tied with the secretion of chemokines, small molecules responsible for inducing chemotaxis of immune cells that have a significant role in TME organization[52]. To investigate this connection, 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* (adjusted $p = 1.5e^{-13}$, one-sided Fisher's exact test), *CXCL2* ($p = 7.4e^{-13}$), and *CXCL8* ($p = 1.7e^{-5}$) were region-specific markers for the Club region, as were *CXCL16* ($p = 1.52e^{-13}$) and *CCL20* ($p = 2.0e^{-4}$). CCL20 was expressed at a higher level in the NEADT Club region compared to the TRNA Club region ($p = 1.9e^{-2}$, two-sided Wilcoxon rank-sum test). The canonical receptors of these chemokines[52] were Immune region markers: *CXCR6* ($p = 2.7e^{-2}$) for CXCL16 and *CCR6* ($p = 4.1e^{-4}$) for CCL20. While the neutrophil chemotactic receptor *CXCR2* was detected in only 0.4% of all the spots in the discovery cohort, a potential reflection of the poor coverage of ST, the extent of neutrophil-attracting chemokines markers within the Club region was indicative of neutrophil recruitment.

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

The observed upregulation of chemokines in the Club and Immune regions prompted us to closely examine the immunological activity taking place in these regions. Gene set enrichment analysis of Gene Ontology: Biological Process (GO:BP) terms highlighted enrichment of myeloid cell chemotaxis-associated gene sets for the Club region (**Figure 4a**). The Immune region was enriched for activation-related processes of both lymphoid and myeloid immune cells. Custom gene set enrichment analysis showed that the Immune region was enriched for two prostate-specific gene sets for myeloid-derived suppressor cell (MDSC) activity derived from a mouse model[43] ($p = 4.3e^{-7}$) and human tumor samples[12] ($p = 2.0e^{-6}$) (**Supplementary Table S7**). Furthermore, a CRISPR-Cas9 validated signature of polymorphonuclear MDSC (PMN-MDSC) immunosuppression[46] was uniquely enriched in the Immune region ($p = 1.9e^{-2}$). These results suggest that the Club region is actively recruiting immunosuppressive myeloid cells to the TME.

To test this hypothesis, we calculated the sample-specific correlation between the Club region fraction and PMN-MDSC activity score in all but the Club region spots in both sample cohorts. Club region fraction was associated with higher PMN-MDSC activity in both the discovery cohort (coef = 0.68, adjusted $p = 2.14e^{-6}$, two-sided Spearman correlation test) and the validation cohort (coef = 0.61, $p = 1.6e^{-3}$, **Figure 4b**). Tumor region spots had a higher PMN-MDSC activity score when proximal to Club region spots (**Figure 4c**).

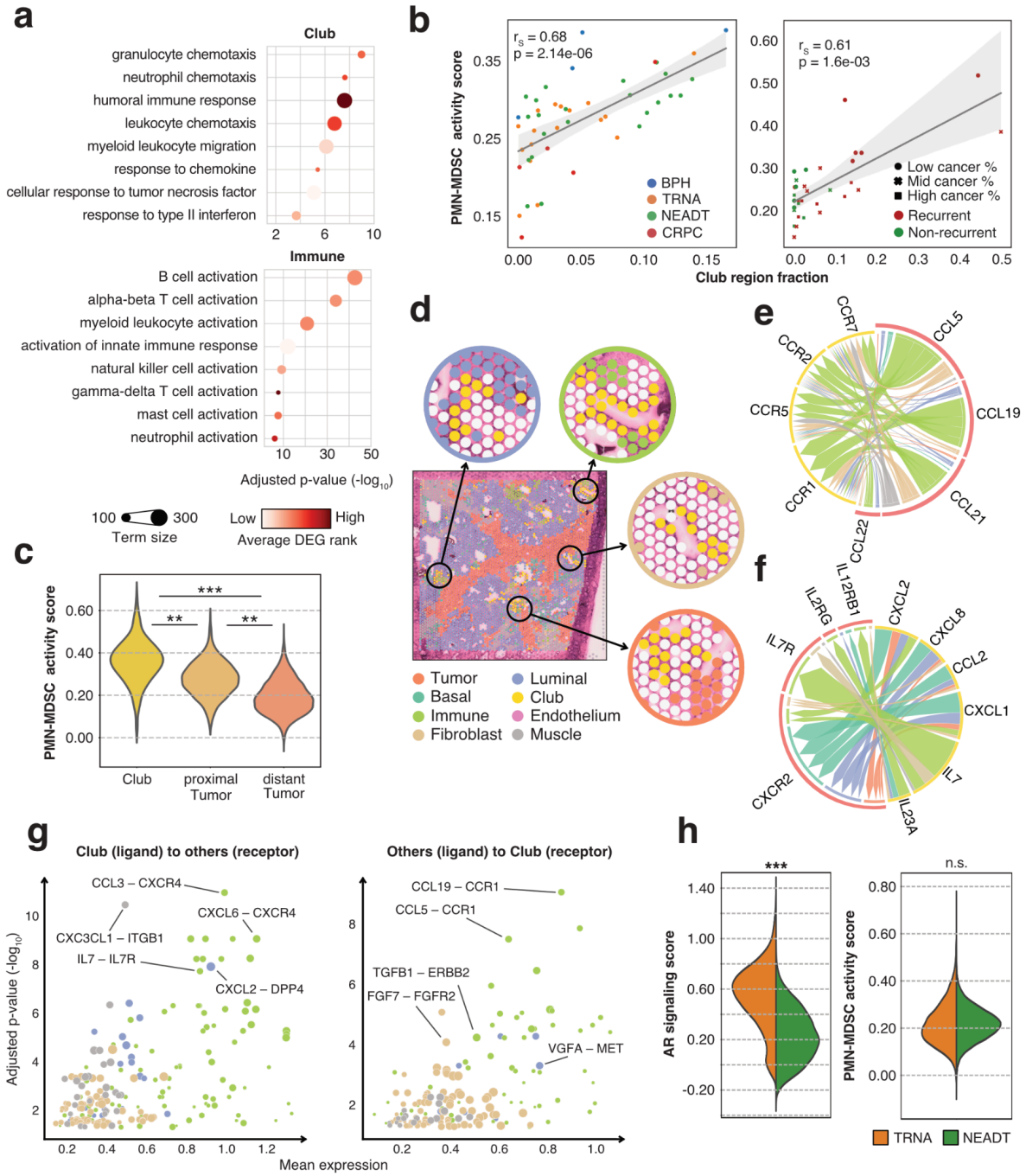

Figure 4. *The Club region is associated with increased myeloid-derived suppressor cell activity.* ***a)*** *GO:BP term enrichments for genes with upregulated expression in the Club (top) and Immune (bottom) regions.* ***b)*** *Correlation between Club region fraction and PMN-MDSC activity scores in 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 divided according to their proximity to the Club region. Asterisks indicate two-sided independent samples t-test p-value ($p < 0.05$) 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,f)*** *Chord*

diagrams for region interface ligand-receptor signaling activity. Edge weights represent the number of samples where a signaling connection was active. The Club region expresses the receptor (d) or the ligand (e), edge is colored according to the interacting region. ***g)*** *Ligand-receptor interaction activity in Club region interfaces for outgoing and incoming signaling. Dot size represents the number of interfaces where interaction was active. Dot color represents the interacting region.* ***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*

Similar correlations were identified for three other MDSC signatures in both cohorts (**Supplementary Figure S5**). In the validation cohort, where the patients were categorized as having recurrent (n = 5) or non-recurrent (n = 3) disease, one prostate MDSC signature ($p = 1.2e^{-4}$, two-sided Wilcoxon rank-sum test) scored higher in the recurrent samples. Recurrent samples of the validation cohort had higher mean scores for all of the tested signatures. There was no difference in PMN-MDSC activity scores between TRNA and NEADT samples of the discovery cohort ($p \geq 0.05$, two-sided independent samples t-test, **Figure 4h**).

To elucidate the specific molecular interactions occurring between the Club and other regions, we performed ligand-receptor binding analysis in region interfaces (**Figure 4d, Methods**). Multiple chemokines actively participated in the signaling between the Club and other regions (**Figure 4e, 4f**). Most notably, the interaction between *interleukin 7 (IL7)* and *interleukin 7 receptor (IL7R)* was significantly more active in the Club-Immune interface than in other Club interfaces (adjusted $p = 1.8e^{-8}$, Fisher's exact test, **Figure 4g**). This was due to the high expression of IL7R in the Immune region, where it was among the top region-specific markers ($p = 2.3e^{-26}$). While IL7 is generally regarded for its anti-immunosuppressive function[53], elevated IL7 and IL7R expression has been associated with tumor progression and poor survival in prostate cancer[54,55].

Club-like senescence is associated with immunosuppressive PMN-MDSC activity

Finally, we examined the connection between PMN-MDSC activity and SASP in the Club region. The pan-cancer *EpiSen* signature that scored high in the Club region was also enriched in its region-specific markers ($p = 3.2e^{-22}$, **Figure 5a**). We termed the 26-gene overlap between these two signatures *club-like senescence*. Genes from the *EpiSen* signature were also overrepresented in the PMN-MDSC activity signature ($p = 2.1e^{-5}$, two-sided Fisher's exact test, **Figure 5a**). While the PMN-MDSC activity signature genes were not overrepresented in the

Club region markers (p = 4.2e^{-1}), the overlapping genes of these three gene sets were highly expressed in the neutrophil and monocyte populations of metastatic CRPC (mCRPC) scRNA-seq data[56] (**Figure 5b**).

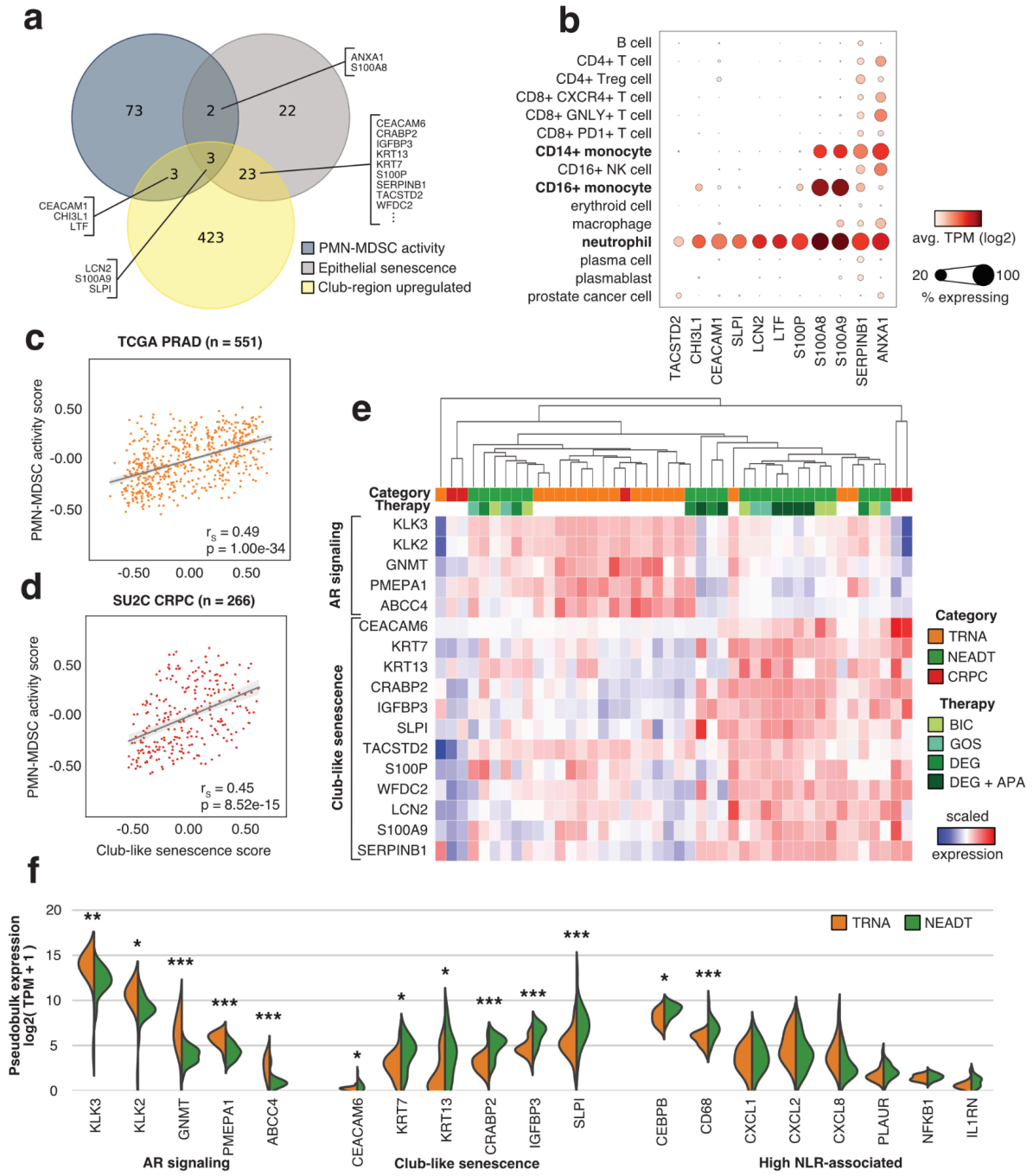

Figure 5. *Club-like senescence is associated with immunosuppressive PMN-MDSC activity.* ***a)*** *Venn diagram showing overlaps between PMN-MDSC activity, epithelial senescence, and*

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,d)*** *Score correlation for the Club-like senescence and PMN-MDSC activity signatures in TCGA PRAD (c) and SU2C (d) cohorts.* ***e)*** *Scaled gene expression of pseudo-bulk spatial transcriptomics samples. Columns have been clustered using Euclidean distance. Color annotations indicate the treatment category (first row) and administered therapy (second row). Therapy is only shown for neoadjuvant-treated samples.* ***f)*** *Violin plot of normalized gene expression in pseudo-bulk spatial transcriptomics samples. Asterisks indicate differential gene expression test significance levels between treatment-naive and neoadjuvant-treated samples (*: $p<0.05$, **: $p<0.01$, ***: $p<0.001$). PMN-MDSC: polymorphonuclear myeloid-derived suppressor cell; TRNA: treatment-naive, NEADT: neoadjuvant-treated; CRPC: castration-resistant prostate cancer; BIC: bicalutamide; GOS: goserelin; DEG: degarelix; APA: apalutamide*

Club-like senescence and the PMN-MDSC activity signature GSVA scores correlated across untreated primary and mCRPC tumors in the TCGA-PRAD and SU2C-PCF cohorts, respectively (**Figure 5c, 5d**). The club-like senescence signature score also correlated positively with MDSC signature scores and the high NLR-associated score across these cohorts (**Supplementary Figure S6**). The club-like senescence score had a negative correlation with AR-signaling in both the TCGA-PRAD (coef = -0.30, $p = 1.2e^{-12}$, two-sided Spearman correlation test) and the SU2C-PCF (coef = -0.28, $p = 5.0e^{-6}$) cohorts.

TPM-normalized pseudo-bulk expression of AR-signaling, club-like senescence, and PMN-MDSC activity-related genes revealed nested groups within the different treatment categories of the discovery cohort (**Figure 5e**). 15 out of the 22 NEADT samples were characterized by decreased AR-signaling and increased club-like senescence activity. Pseudo-bulk level 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 5f, Supplementary Table S8**). None of the genes in these two signatures were downregulated following treatment. Taken together, these results show that club-like senescence signaling counterbalances AR-signaling activity and associates with PMN-MDSC activity.

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 pipeline uses the same methodological approaches that have captured recurrent gene expression programs across different tumor types[39,40,57]. This NMF-based approach emphasizes gene set variation within individual samples before cross-sample integration, avoiding batch effect issues related to commonly used clustering workflows[58,59]. Through this methodology, we have ensured an unbiased interrogation of prostate cell states from scRNA-seq data and used them to yield insights into the organization of prostate TME.

Our analysis reveals that club cells display a pronounced SASP-expression profile. The upregulated expression of chemokines CXCL1, CXCL2, CXCL8, CXCL16, and CCL20 is indicative of SASP[60] and highlights the immune-interactive nature of these cells. In a prostate-specific context, the expression of CXCL1, CXCL2, and CXCL8 is independently predictive of overall survival in metastatic CRPC patients, and blocking their target receptor CXCR2 can reverse therapy resistance[45]. High CCL20 expression in the prostate is associated with more aggressive disease[61] 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[62,63]. Intriguingly, 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 a possible metastasis mechanisms acting through CXCL16[63].

Increased cytokine signaling activity in the Club region is in line with reports that club cells are enriched in proliferative inflammatory atrophy (PIA)[24,64,65]. PIA has been regarded as a potential precursor for prostate cancer[64], with DNA damage resulting from cytokine-induced oxidative stress reported as one possible mechanism[14,66]. We found increased expression of PIA gene markers[65] KRT5, KRT8, and MET, and PIA-associated club 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 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[67], whereas high IFNγ signaling is associated with worse outcomes in high-risk localized disease[68]. These results indicate that the club cells are interchangeably linked with inflammation in the prostate TME. Our findings corroborate the link between club cells and the luminal progenitor phenotype reported in mice[27]. Luminal progenitor cells are an inherently castration-resistant cell population in the mouse prostate[25,69] that have also been proposed as a possible cellular origin of prostate cancer[70]. 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 is supported by our data[36]. The hillock cell marker KRT13 is expressed in the Club and Basal regions, implying that while hillock cells do not constitute a distinct area of the prostate, they are present in the tissue. Our analysis shows that the Club region is unperturbed by ADT, indicating that human club cells are similarly resistant to castration. Upregulated expression of transcription factors SOX9, KLF5, and ELF3 in the Club region is also indicative of the stem-like characteristics of these cells[71]. Whether these cells actively participate in the post-castration regeneration of the human prostate is beyond the scope of this study.

Finally, we postulate a connection between club cells and clinically significant immune cell dysfunction. Immunosuppressive PMN-MDSCs are a driver of CRPC that can reactivate androgen signaling through IL-23 secretion[43]. The secretion of their main chemoattractant IL-8, is increased post-castration. More generally, tumor sites contain expanded PMN-MDSC populations compared to healthy tissue[72,73]. The success of CXCR2 inhibitors in the treatment of late-stage mCRPC patients proves that castration resistance can be tackled by inhibiting PMN-MDSC function[45].

Taken together, our work represents the largest spatially resolved gene expression landscape of prostate tumors published to date. Our approach of leveraging single-cell gene expression data enabled us to perform a joint analysis of our whole dataset, a task deemed challenging due to the convoluted nature of ST data. Our exploratory analysis replicates the abundance and expression of club cells identified in single-cell RNA-seq data and ties their presence to the immunosuppressive activity of PMN-MDSCs. The observed combination of club cells' inherent castration resistance and PMN-MDSC chemotaxis-inducing capability implicates them as a possible contributor to castration resistance, warranting further research into their function both in tumor initiation and resistance to therapy.

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, two uropathologists (T.M. and C.P.) 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).

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 was converted into spot-wise histology classes is given elsewhere[74].

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*[75]. 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*[76,77].

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 S7**). These cell abundances can be considered as an alternative to gene counts, where instead of thousands of genes the feature space is 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. NMF was run for multiple iterations with *n_components* parameter values 5 to 12. 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 S8**). Components (regions) inherited their names from cell states that were among the highest contributors. Each ST data point was then categorized according to its highest contributing component, resulting in eight regions across the whole ST dataset (**Supplementary Figure S9)**.

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 less 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 adjusted p-value < 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 adjusted p-value < 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[78,79]. 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 adjusted $p < 0.05$ were considered enriched. g:Profiler was used to perform enrichment analysis of region-specific gene markers on the Gene Ontology: Biological Process database[80,81].

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 less 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 less 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.

squidpy's implementation of the *CellphoneDB* method was used to calculate ligand-receptor interactions in the region interfaces. More specifically, *squidpy.gr.ligrec* was run with parameter settings *complex_policy='all'*, *threshold=0.01*, and *n_perms=1000*. Ligand-receptor pairs with adjusted p-value < 0.05 were considered to be active in each interface at a sample level. Finally, *consensus_score_intercell_target* > 10 and *consensus_score_intercell_source* > 10 were used

as filters to remove redundant ligand-receptor pairs. The final result was the number of samples where a ligand-receptor pair was active for each region interface.

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*.

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 expression data of the SU2C-PCF mCRPC cohort was downloaded from a public repository (https://www.cbioportal.org/study/summary?id=prad_su2c_2019)[82]. *gseapy.gsva* (v1.1.0) was used to calculate GSVA scores for each sample[83]. 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)[84] 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

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

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.

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. Mackall, C. L., Fry, T. J. & Gress, R. E. Harnessing the biology of IL-7 for therapeutic application. *Nat. Rev. Immunol.* **11**, 330–342 (2011).
54. Schroten, C. *et al.* The additional value of TGFβ1 and IL-7 to predict the course of prostate cancer progression. *Cancer Immunol. Immunother.* **61**, 905–910 (2012).
55. Qu, H. *et al.* IL-7/IL-7 receptor axis stimulates prostate cancer cell invasion and migration via AKT/NF-κB pathway. *Int. Immunopharmacol.* **40**, 203–210 (2016).
56. He, M. X. *et al.* Transcriptional mediators of treatment resistance in lethal prostate cancer. *Nat. Med.* **27**, 426–433 (2021).
57. Kinker, G. S. *et al.* Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat. Genet.* **52**, 1208–1218 (2020).
58. Stein-O'Brien, G. L. *et al.* Enter the Matrix: Factorization Uncovers Knowledge from Omics. *Trends Genet.* **34**, 790–805 (2018).
59. Kotliar, D. *et al.* Identifying gene expression programs of cell-type identity and cellular activity with single-cell RNA-Seq. *Elife* **8**, e43803 (2019).
60. 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).
61. Kfoury, Y. *et al.* Human prostate cancer bone metastases have an actionable immunosuppressive microenvironment. *Cancer Cell* **39**, 1464–1478.e8 (2021).

62. Lu, Y. *et al.* CXCL16 functions as a novel chemotactic factor for prostate cancer cells in vitro. *Mol. Cancer Res.* **6**, 546–554 (2008).
63. Jung, Y. *et al.* Recruitment of mesenchymal stem cells into prostate tumours promotes metastasis. *Nat. Commun.* **4**, 1–11 (2013).
64. 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).
65. 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).
66. Mani, R. S. *et al.* Inflammation-Induced Oxidative Stress Mediates Gene Fusion Formation in Prostate Cancer. *Cell Rep.* **17**, 2620–2631 (2016).
67. 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).
68. 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.
69. Karthaus, W. R. *et al.* Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* **159**, 163–175 (2014).
70. 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).
71. Bian, X. *et al.* Integration Analysis of Single-Cell Multi-Omics Reveals Prostate Cancer Heterogeneity. *Adv. Sci.* e2305724 (2024) doi:10.1002/advs.202305724.
72. Gabrilovich, D. I. & Nagaraj, S. Myeloid-derived suppressor cells as regulators of the immune system. *Nat. Rev. Immunol.* **9**, 162–174 (2009).

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. 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.

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

76. 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).

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

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

79. 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).

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

81. 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).

82. Abida, W. *et al.* Genomic correlates of clinical outcome in advanced prostate cancer. *Proceedings of the National Academy of Sciences* **116**, 11428–11436 (2019).

83. Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, (2023).

84. Muzellec, B., Teleńczuk, M., Cabeli, V. & Andreux, M. PyDESeq2: a python package for bulk RNA-seq differential expression analysis. *Bioinformatics* **39**, (2023).