Genome Medicine

RESEARCH

Open Access

Multifocal cohort analysis unveils cell types associated with regional lymph node seeding in prostate cancer

Louise de Schaetzen van Brienen[1,2,3†], Taewoo Jung[1,2,3†], Maarten Larmuseau[1,2,3†], Kim Van der Eecken[3,4,5], Marie Van Hecke[1,2,3], Jef Haerinck[3,6], Jordy De Coninck[3,6], Jan Vanwelkenhuyzen[3,4,6], Nicolaas Lumen[3,8], Bram De Laere[3,4,10], Nele De Witte[3,5], Giles Miclotte[1,2,3], Sofie Verbeke[3,5], Geert Berx[3,6], Jo Van Dorpe[3,5], Piet Ost[3,4,9†] and Kathleen Marchal[1,2,3*†]

Abstract

Background Understanding the molecular features that underlie metastatic prostate cancer (PCa) is essential to develop prognostic markers and improve treatment decisions. However, such studies are hampered by substantial intratumor and interpatient heterogeneity.

Methods To cope with this heterogeneity, we propose a unique study design that leverages the statistical power of multifocal bulk transcriptome profiling with the resolution of a single-cell analysis to identify processes and cell types associated with regional metastatic lymph node seeding in PCa.

Results Elaborate analysis of these data allowed identifying a metric to distinguish, based on the multifocal expression data between lesions with high and low potential for regional metastatic lymph node seeding. Subsequently comparing the expression profiles of these lesions with respectively high and low metastatic potential identified an aggressiveness signature. Overlaying this signature with single cell data identified proliferative luminal cells, an adipose derived cancer-associated fibroblast (CAF) state and a specific subtype of arterial endothelial cells. Assessing the prognostic value of these cell states in an independent dataset (TCGA-PRAD) confirmed their association with regional metastatic lymph node seeding and progression free survival and unveiled a complementary role for the proliferative luminal cells and the adipose derived CAF state in driving regional metastatic lymph node seeding.

Conclusions Based on our analysis, we hypothesize that lesions with high potential for regional metastatic lymph node seeding are mostly characterized by the presence of highly proliferative luminal cells and a transitioning towards an aggressive adipose derived CAF state. Markers associated with these cell states largely explain the

†Louise de Schaetzen van Brienen, Taewoo Jung and Maarten Larmuseau contributed equally to this work.

†Piet Ost and Kathleen Marchal contributed equally to this work.

*Correspondence:
Kathleen Marchal
Kathleen.Marchal@ugent.be

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

prognostic signal of currently used commercial signatures in PCa, further supporting the role of the identified cell states in driving regional metastatic lymph node seeding and providing an in depth understanding of the success of the currently used commercial signatures.

Keywords multifocal transcriptome profiling, prostate cancer, metastatic seeding, adipose-derived stromal stem cells, cancer-associated fibroblasts, cell state transition

Background

While most prostate cancers (PCa) are indolent and localized, a small fraction of PCa evolves towards metastatic disease, leading to poor prognosis and higher mortality rates. However, the risk on metastasis after diagnosis remains hard to predict [1]. The presence of regional lymph node metastases at the time of diagnosis (referred to as N1 disease) is an important factor in determining a patient's prognosis [2]: Up to 14% of men are diagnosed with N1 disease at initial diagnosis [3] and 44% of these patients develop distant metastases within 10 years following diagnosis [4]. Lymph node involvement has been hypothesized to represent a clinical stage between localized disease and metastatic dissemination [5]. Understanding which processes in the primary tumor are associated with regional lymph node seeding can improve predicting a patient's metastatic potential [6].

A tumor's metastatic potential is not only driven by somatic mutations, but also by the interaction of the tumor cells with the tumor micro-environment (TME) [7–10]. This seems to be particularly true in the case of regional lymph node seeding, which follows an evolutionary process that is different from distant metastatic lesions [11]. The specificities that define a metastatic seeding clone might therefore, at least in part, be triggered by the plasticity of the tumor cells and the specificities of the TME [12]. Assessing transcriptomic variation has the potential to capture such phenotypic states of a tumor that are driven by the interaction between the tumor and the TME [13, 14] and that associate with metastatic potential.

Bulk profiling of large cohorts of patients already identified expression signatures obtained from primary tumor samples that associate with poor prognosis and/or lymph node involvement [13, 15–24], indicating that, in the primary prostate tumor, transcriptomic signatures must be present that reflect metastatic potential. However, most bulk profiling-based transcriptomic studies rely on samples taken at a single lesion. Given the heterogeneity of the primary prostate tumor [25], which contains multiple lesions of which only some are responsible for metastatic seeding [26–28], such single lesion studies lack the resolution to identify the specificities of the metastatic seeding process. In addition, it is difficult to decide in advance, based on clinical and histopathological characteristics which lesion contains the more aggressive cells [29]. If a lesion selected for sequencing does not correspond to the lesion that determines a tumor's metastatic potential, the risk of catching irrelevant signals or missing a true association with metastatic seeding increases.

By adopting a multifocal approach and sampling multiple lesions within the primary tumor, the chance of selecting the lesions that contain the most aggressive cells responsible for metastatic seeding increases [30, 31]. Thus, to investigate transcriptional processes in the primary tumor associated with regional lymph node seeding, we generated a large multifocal cohort of matching primary and regional metastatic lymph node samples, comprising 211 samples in total. This cohort allowed us to define, based on RNA expression a metric that enables identifying per patient the lesions with respectively the most and the least potential for regional lymph node metastasis. A subsequent direct comparison of the expression profiles of lesions with extremely different metastatic potential resulted in a gene set characterizing this potential for regional lymph node seeding. Overlaying with single cell data associated the gene set with proliferative luminal cells, an adipose derived cancer-associated fibroblast (CAF) state and a specific subtype of arterial endothelial cells. Based on these findings we hypothesized that lesions with high potential for regional metastatic lymph node seeding are characterized by the presence of highly proliferative luminal cells, a transitioning towards an aggressive adipose derived CAF state and a relative depletion of arterial endothelial cells. This was confirmed by assessing the prognostic value of these cell states for regional metastatic lymph node seeding and progression free survival in an independent dataset (TCGA-PRAD) [32]. The proliferative luminal cell state carries a signal that is highly prognostic for both regional lymph node status and progression free survival, whereas the presence of an aggressive adipose derived CAF state mostly associates with regional lymph node status. In addition, we showed that markers associated with these identified cell states largely explain the prognostic signal of the currently used commercial signatures Decipher [18] and Prostadiag [33], further confirming the role of mostly the adipose derived CAFs and proliferative luminal cells in determining a lesion's metastatic potential.

Methods

Multifocal in-house cohorts

In this study, we analyzed two multifocal in-house cohorts of patients with advanced PCa. The first cohort contains samples from 17 locally advanced treatment-naive hormone-sensitive PCa patients (mHSPC) with regional pelvic lymph node metastases at the time of diagnosis, but without detectable distant metastases (N1M0) (referred to as the locally advanced cohort). Overall, this cohort consists of bulk RNA-seq data for 62 metastatic lymph node (MLN), 85 primary tumor (RP), and 17 adjacent normal prostate (NL) FFPE samples. The second cohort contains samples from 16 treatment-naive patients with *de novo* metastatic hormone-sensitive PCa (mHSPC) with regional pelvic lymph node metastases at the time of diagnosis and detectable distant metastases (N1M1, referred to as the *de novo* cohort). This cohort contains bulk RNA-seq data for 20 regional MLN and 26 RP samples, but no adjacent normal prostate samples.

To generate the multiregion samples of these cohorts archival formalin-fixed paraffin-embedded (FFPE) tissue blocks of primary prostate tumors and regional lymph node metastases with corresponding H&E stained slides were reviewed by two urological pathologists. Tumor areas were marked and individually classified according to the current International Society of Urological Pathology (ISUP) Prostate Cancer Guidelines [34]. The soft tumor areas of anatomic location of radical prostatectomy were mapped in the PIRADS version2.1 sector maps [35]. For each patient, candidate regions for sequencing in the radical prostatectomy specimen were chosen based on the anatomic location and dimensions of the dominant tumor lesion (i.e. the highest ISUP grade and the highest percentage of tumor involvement), plus one to seven additional tumor-containing areas. Furthermore, metastatic foci measuring approximately $2mm^2$ or more and different areas within the diagnostic (i.e., preoperative) prostate biopsy blocks were selected. Tumor sampling was performed to maximize the representation of spatially distinct and/or microscopically heterogeneous tumor foci; fibrotic or hemorrhagic areas were avoided. The tissue regions selected for sequencing were sampled using 1 mm punch biopsies or 10 µm tissue sections for biopsy and surgical samples, respectively. Pathological tumor cellularity was assessed and only samples with at least 10% tumor cells were selected for RNA isolation. Informed consent was obtained from all patients who participated in the study. The clinical characteristics of the patients and their matching samples are enlisted in Additional file 1: Table S1 and S4 for the locally advanced and the *de novo* cohort.

Samples were sent to ALMAC for RNA extraction and sequencing. All samples from the same cohort were processed and sequenced at once. For both cohorts , RNA was extracted using the Qiagen RNeasy FFPE kit. RNA fragmentation was performed, and RNA integrity was assessed with High Sensitivity RNA ScreenTape from Agilent Technologies. As input for the library prep 8.5 µl (maximum volume) was used for all samples. Libraries were subsequently prepared using Illumina RNA Exome and cDNA was measured with the ThermoFisher Qubit DNA HS Assay Kit (Thermo). Paired end sequencing was performed on the NovaSeq platform.

For read alignment, quality assessment and expression quantification, we relied on the nf-core workflow (version: 3.1) [36, 37]. To facilitate comparison with TCGA-PRAD [32], we relied on the same reference genome and gene annotations as the TCGA consortium. So STAR for alignment to the human reference genome GRCh37 (Homo_sapiens.GRCh37.dna.primary_assembly.ERCC.fa) was used. Expression was quantified using StringTie [38] and normalized TPM gene expression was used as input for all subsequent analyses. Additional file 1: Table S2 and S5 provide the detailed quality assessments.

For the locally advanced cohort, targeted Next-Generation Sequencing (tNGS) data was available for five patients (that is, ID4, ID5, ID6, ID9, and ID10), in the context of an independent study conducted at the Karolinska Institute [39]. Targeted sequences were obtained using an in-house gene panel and processed as described in [39]. QiaAmp FFPE was used for DNA extraction, and DNA concentration was measured using the Qubit BR DNA assay. Library preparation was based on KAPA Hyperprep with IDT adapters. For DNA capture, TWST probes were used. The bam files for the tNGS data were aligned with the reference genome *g1k_v37_decoy.fasta* using the BWA-mem aligner [40].

For the six patients with multiple RP samples and at least one matching MLN sample in the *de novo* cohort, we also had matching tNGS data [29]. tNGS libraries were prepared using the Roche HyperCap Target Enrichment Kit, the Roche HyperCap Bead Kit, and a custom Roche SeqCap 73 PCa gene panel as previously published [41]. The targeted enrichment was performed as described in the technical manual (SeqCap EZ HyperCap Workflow User Guide; version 2.3; Chaps. 5 to 7). The amplified captured multiplex DNA sample was resuspended in 53µL nuclease-free water. In contrast to the technical manual, multiplex DNA sample library pools with a combined mass of 2 µg were prepared. Sequencing was performed on either the Illumina HiSeq 1500/2500 High Output platform or the Illumina NextSeq 500 platform.

External cohorts

To identify differentially expressed gene transcripts between normal tissues and primary prostate tumors, we also used the TCGA-PRAD bulk RNA-seq dataset.

Raw HTSeq counts were downloaded from the Xenahub portal [42]. Ensembl gene IDs were assigned to the gene names using the Xenahub annotation file.

The sample type information was downloaded from the biospecimen folder and the lymph node status information was derived from the clinical folder, both downloaded from the GDC portal [43].

The single-cell RNA-seq dataset of Chen et al. (2021) [8] consisting of 13 tumor samples (12 RPs and 1 MLN) from 12 PCa patients was downloaded from the supplementary information file of the original publication.

Identification of the progression signature

To disentangle the progression signature from the confounding tissue adaptation signal, we identified genes that changed in expression in both the NL vs. RP and in the RP vs. MLN comparisons. Hereto we used the data from TCGA-PRAD [32] and the locally advanced cohort to identify the genes that are differentially expressed between the NL vs. RP samples and the data of the locally advanced cohort and the *de novo* cohort to identify genes differentially expressed between the RP versus MLN samples.

Differentially expressed genes were identified with DESEq2 in R (ver. 1.38.3) [44] using the patient identity as a covariate and the tissue of origin as the condition of interest (design = ~ patID+sampletype, where the covariate patID reflects the ID of the patient from which the sample originates and sampletype the condition of interest, either MLN, RP or NL depending on the comparison). While the locally advanced cohort was sequenced as a single batch, *de novo* cohort was sequenced in four technical batches (Additional file 1: Table S5).

To evaluate the impact of within-cohort batch effects, we used the Likelihood Ratio Test (LRT) implemented in the DESeq2 R package [44], comparing a full model including batch as a covariate (*design* = ~ *patID* + *sampletype* + *batch*) against a reduced model without batch (*design* = ~ *patID* + *samplettype*). Only a small fraction of genes (76 genes) showed a significantly improved fit when including batch (adjusted p-value < 0.05). Importantly, none of these genes overlapped with those identified as differentially expressed in the reduced model. So we continued the analysis without correcting the *de novo* cohort for the batches.

To avoid effects originating from intrinsic differences or batch effects between the analyzed cohorts [32], differential analysis was performed for each cohort separately (Additional file 1: Table S7-S10) and results were combined at the level of the identified differentially expressed genes (meta-analysis) using the following rule (Additional file 2: Text S3):

Select = [((genes differentially expressed in NL/RP in TCGA-PRAD) OR (genes differentially expressed NL/RP in locally advanced cohort)) AND ((genes differentially expressed RP/MLN in locally advanced cohort) OR (genes differentially expressed RP/MLN *de novo* cohort))]. This resulted in a list of 575 genes (Additional file 1: Table S11).

We subsequently performed Spearman's rank principal component analysis (PCA) [45, 46] on the expression data using the 575 genes selected above as variables and the samples as observations. Selecting those genes of which the variance best associated with the tumor progression (defined in terms of their loading on the first principal component (PC1), Additional file 2: Text S3) resulted in a final progression signature of 225 genes of which the expression associates with disease progression (Additional file 1: Table S12).

Identification of the lesions with highest aggressive potential based on the progression signature

To select, per patient, the RP lesion(s) with the highest aggressive potential based on the progression signature, we defined a signature-based metric for expression similarity using the differences in scores of the samples within each patient on PC1 of the Spearman's rank PCA [45, 46] obtained with the 225 genes of the progression signature as variables (Additional file 2: Text S4). Hereto we first defined RP lesion with the least aggressive potential as the one that is the most distant from the MLN samples on PC1. An RP lesion was considered aggressive when its distance from the least aggressive lesion was greater than 25% of the distance between the least aggressive RP lesion and the median score of the MLN samples of that patient on PC1. Using this metric, we can guarantee that the most aggressive lesion is sufficiently close to the matching MLN samples and, at the same time, sufficiently different from the least aggressive RP lesion. In case all primary lesions were equally distant from their matching MLN samples, no conclusions on the identity of the most aggressive lesions were drawn.

Identification of the seeding lesions using tNGS

When tNGS data [29] were available for multiple RP lesions and matching MLN lesions, we used this information to identify the lesions in the primary tumor that were most probably responsible for metastatic seeding (seeding lesions). To this end, the most likely seeding lesion was defined as the lesion in the primary tumor that shared the highest number of mutations with the matching MLN samples. If multiple RP lesions shared the same number of mutations with the matching MLN sample, these lesions were considered to have the same seeding potential (Additional file 2: Text S5).

Identification of the aggressiveness signature, a gene set associated with a lesion's metastatic potential

To identify genes of which the expression is associated with a lesion's metastatic potential, we compared the expression of primary with the highest and the lowest metastatic potential as identified by our signature-based metric. We performed this analysis in each multifocal cohort separately (locally advanced and *de novo* cohort). To avoid diluting the differential expression signal, patient ID4 of the *de novo* cohort was omitted from the analysis as in this patient the signature-based metric and the tNGS analysis disagreed on the identity of the lesion with the lowest metastatic potential (RP2 was considered having a low metastatic potential with the signature-based metric but still identified as a seeding lesion with the NGS-based lineage reconstruction).

To identify the genes that were differentially expressed between the least and most aggressive RP lesions we used DESEq2 [44] with the patient information as model variable, design = ~ patID+sampletype, where sampletype indicates whether the RP lesion is the most or least aggressive). Genes were considered differentially expressed if their |logFC| was greater than 1 and the adjusted *p*-value lower than 0.05. To avoid batch effects each cohort was analyzed separately, resulting in a list of differentially expressed genes for the locally advanced cohort and the *de novo* cohort respectively (Additional file 1: Table S14 and S15). Both gene lists were subsequently merged: we retained genes that were found differently expressed in one of the two cohorts only using a stringent threshold (|logFC|>2, adj-*p*-value < 0.001) and added the genes that were found differentially expressed in both cohorts using a more lenient threshold (|logFC|>1, adj-*p*-value < 0.01, see Additional file 1: Table S18 and Additional file 2: Text S7).

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed using fgsea package (ver. 1.24.0) in R [47], that performed pre-ranked GSEA. Enrichment is assessed using the 50 hallmarks pathways from MSigDB [48]. The ranks used to perform enrichment were either the stat score returned by DESEq2 [44] or the average log2 fold change, provided by FindMarkers from Seurat R package (ver. 5.0.1) [49]. Only enriched gene sets with an adjusted *p*-value below 0.05 were considered statistically significant (see Additional file 2: Text S7).

Assessing the prognostic potential of the identified cell types

Prognostic potential was assessed using the expression data from primary tumor samples of TCGA-PRAD [32]. K-means clustering [50] was applied to group samples in two clusters based on their expression profiles of a prespecified gene set with ComplexHeatmap package in R (ver. 2.14.0) [51]. Enrichment in lymph node positive samples is calculated using the Fisher's exact test [52]. The cluster with a p-value < 0.05 is considered to be enriched in lymph node positive samples. Progression-free survival curves were estimated per cluster, using the Kaplan-Meier estimator [53] with survival (ver. 3.5.7) [54] and survminer (ver. 0.5.0) [55] packages in R. Survival distributions of curves of either cluster were subsequently compared using the log-rank test [56]. Moreover, the relative hazard ratio of the cluster enriched with lymph node positive samples on progression-free survival was estimated using the Cox proportional-hazard model [57], with the lymph node enrichment status of the cluster as a covariate.

As prespecified gene sets we used respectively gene signatures that mark the proliferative luminal cell state (40 genes), the adipose derived fibroblasts (24 genes), the arterial endothelial cells (26 genes), a reduced 10 gene set that characterizes the transitioning towards the aggressive adipose derived CAF state we set apart with the aggressiveness signature (in Fig. 5B) i.e. *FBLN2*, *VCAN*, *COMP*, *SFRP4*, *CTHRC1*, *COL10A1*, *CXCL14*, *COL11A1*, and *THBS2*. These markers originally associated to matrix CAFs in Bartoschek et al. (2018) [58] were described by Zhu et al. (2021) [59] as being associated with the transitioning of adipose derived stromal stem cells (ASCs) towards a fibroblast-like cell states [8]. All of the selected genes were identified by FindMarkers [49] as distinctive for the fibroblasts-like cells set apart by the aggressiveness signature, Additional file 1: Table S21). The 10 gene set was further reduced to a 5 gene set by only retaining genes that uniquely mark the transitioning to the aggressive *COL11A1* cell state and that are not expressed in any other cell type of Chen et al. (2021) dataset [8] (Fig. 6B).

As a reference we also clustered the samples according to their lymph node status and based on gene sets that define commercially used gene signatures of Decipher [18], OncotypeDx [20] and Prostadiag [33].

Using single cell data to identify the cell types captured by the aggressiveness signature

We downloaded the expression values of the 36,424 cells from the Chen et al. (2021) dataset [60]. t-SNE and cell annotation were performed as described in Chen et al. (2021) (see Additional file 3: Fig. S4) [60]. In the remainder of the analysis, only primary samples from Chen et al. (2021) were considered. Of the 427 genes in the metastatic progression signature, 380 were profiled in the Chen et al. (2021) [60]. To identify which cell types of the Chen et al. (2021) dataset [8] were represented by the genes in the aggressiveness signature, we focused on the expression of these 380 overlapping genes and assigned

each of the 380 genes of the aggressiveness signature to a cell type using FindMarkers [49]. We only focused on overexpression to ensure that we could identify genes that uniquely mark the identified cell types (see also Additional file 2: Text S8).

To identify cancerous cells in this dataset, a copy number variation (CNV) analysis was performed individually for each patient using inferCNV [61] in R (ver. 1.14.2). This tool requires a baseline, which we derived from cells not classified as luminal, basal/intermediate, or unknown clusters. inferCNV computes CNV scores, which signify copy number deviations from the baseline [61]. The top 10% of cells with the highest CNV_{scores} were selected to obtain an average aggressive baseline. Next, cancerous cells were classified as follows: cells with $CNV_{scores} > 1$ and $correlation > 0.4$ with the average aggressive baseline were labeled cancerous; cells with $CNV_{scores} < 1$ and $correlation < 0.4$were labeled non- cancerous; remaining cases were unresolved. In particular, inferCNV encountered issues in two patients, leading to missing CNV scores marked as NA [61].

Refined annotation of the single-cell data

Cells from Chen et al. (2021) [60] were re-clustered. After normalization, variable gene detection, scaling and PCA [45], we used Harmony [62] (ver. 0.1.1) in R package with $\Theta = 2$ and corrected the embeddings to reduce interpatient heterogeneity as cell types from the same origin clustered separately for some patients. Next, we clustered the cells using FindNeighbors and FindClusters functions in Seurat R package [49] and visualized them with uniform manifold approximation and projection (UMAP) [63]. The obtained clusters were overlayed with the original annotation provided by Chen et al. [8] (Additional file 3: Fig. S4). Subsequently, the epithelial and stromal (endothelial and fibroblast) cells were reclustered separately for a more refined annotation using both canonical markers as summarized in the PanglaoDB [64] and CellMarker [65] databases, as well as using previously described markers and gene expression signatures [66–68].

All used markers are indicated on the heatmaps displaying the results (Figs. 3, 4 and 5). Luminal and basal/intermediate epithelial cells were distinguished based on the expression of luminal (*KRT8*, *KLK3*, and *KLK4*) and basal (*KRT5*, *KRT14*, and *TP63*) markers. Marker genes associated with cell-cycle progression or ribosomal function used to identify the proliferative luminal epithelial cell cluster were derived from [66–68]. Within the stromal cell compartment, myofibroblasts and pericytes were distinguished based on canonical markers (E.g., *MYH11*, *ACTA2* for myofibroblasts [65] and *RGS5*, *NDUFA4L2*, and *HIGD1B* for pericytes [64]).

Preadicopyte markers (*CFD*, *APOD*, *MGP*, and *CXCL14*) were obtained from [69, 70]), markers defining adipogenic CAFs were obtained from Qian et al. (2020) [69]; To differentiate ASCs from fibroblasts we used genes related to matrix remodeling that were more upregulated in ASCs than in fibroblasts as described in [71] (*PI16*, *COMP*, *ELN*, *EFEM1*, *POSTN*, *MFAP5*, *IGFBP3*, *INHBA*, *MGP*, *FNDC1*, *MYH10*, *SRFP4*, *F2R*, *IGFBP5*, *VCAN*, *CCDC80*, *TIMP1*, *PTGIS*, and *SERPINE2*). General CAF markers *FAP*, *S100A4*, *VIM*, *PDGFRB*, *PDPN*, and *TNC*) were obtained from [72]. To define the specific subset of matrix CAFs [58] that are in Zhu et al. (2021) [59] associated with the transitioning of ASCs to an adipose derived CAF state in normal pancreatic tissue, we used *PDGFRA*, *DCN*, *LUM*, *VCAN*, *MFAP5*, *POSTN*, *FBLN1*, *FBLN2*, *LOX*, *LOXL1*, and CXCL14. Additional genes derived from Zhu et al. (2021) [59] and Qian et al. (2020) [69] that define the transitioning to the aggressive *COL11A1* CAF state were *FN1*, *CTHRC1*, *COMP*, *SFRP4*, *COL10A1*, *COL11A1*, *INHBA*, and *THBS2*.

For the endothelial cell annotation, lymphatic ECs were identified based on expression of *CCL21* and *PROX1*. Arterial ECs were based on *FBLN5*, *FN1*, and *GJA5*. The remaining vascular ECs were further divided in high endothelial venules and venous ECs (*ACKR1*, *SELP*), and endothelial tip cells (*ESM1*, *INSR*). Cells positive for both EC (*PLVAP*, *PECAM1*) and pericyte (*RGS5*, *NDUFA4L2*, and *HIGD1B*) markers were annotated as endothelial-pericyte doublets.

Results

Multifocal Profiling of locally advanced and *de novo* prostate cancer samples

To identify transcriptomic signals that associate with regional lymph node seeding, we performed an extensive multifocal RNA-seq profiling of the primary tumor and matching metastatic lymph node samples of 33 treatment-naive PCa patients. Samples were obtained from two different clinically different cohorts of patients with aggressive prostate cancer.

A first cohort (referred to as the locally advanced cohort) consists of 164 profiled tumor lesions obtained at diagnosis from 17 patients with mHSPC, with regional pelvic lymph nodes metastases at diagnosis, but no detectable distant metastases (N1M0 at diagnosis). For this cohort, per patient multiple RP lesions obtained from radical prostatectomy, multiple regional MLN regions and one tumor adjacent region proxying NL tissue were profiled (see Fig. 1A).

For the second cohort, 46 lesions were profiled from 16 mHSPC patients diagnosed with aggressive *de novo* metastatic disease (referred to as *de novo* cohort), who show distant metastases at the time of diagnosis (N1M1). It contains for 10 patients RNA-seq data for one RP and

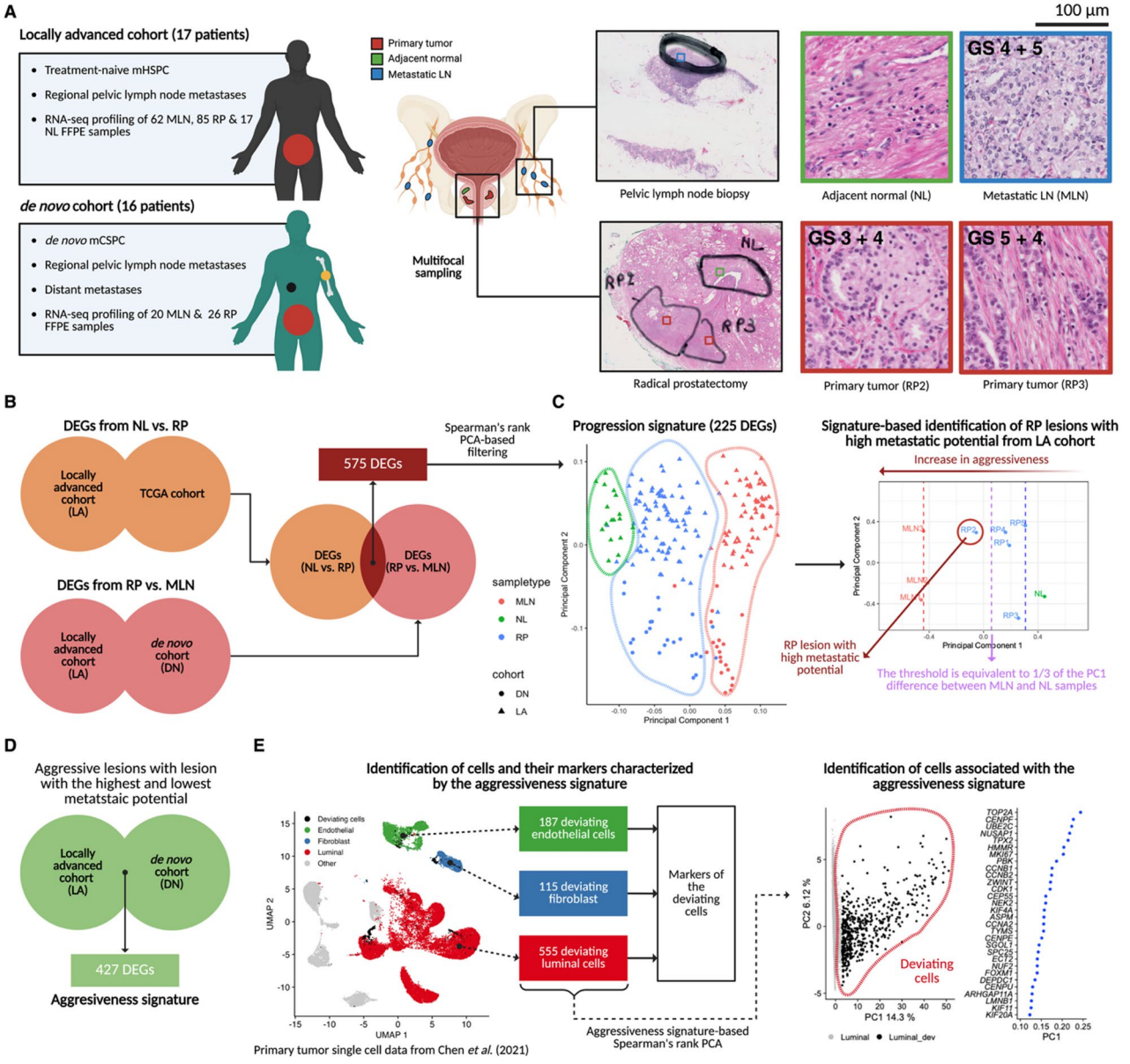

Fig. 1 Overview of the analysis. **A**. Using the expression profiling of two multifocal cohort data obtained from aggressive prostate cancer patients. **B**. A progression signature is identified containing genes that gradually change in expression during the progression from the normal to the metastatic disease stage. **C**. Primary tumor lesions that are closer to the regional metastatic lymph samples in terms of this progression signature are considered to have a higher metastatic potential. **D**. Given the identity of the lesions with the highest and lowest metastatic potential in each sample, an aggressiveness signature is identified by searching for genes of which the expression is different between the lesions with the highest and lowest metastatic potential (referred to as least versus most aggressive in the figure). **E**. To identify the specific cell types that contribute to this aggressive signature, we overlayed the signatures with single cell data of Chen et al. (2021). This allowed setting apart a specific subset of luminal, fibroblast-like and endothelial cells associating with the aggressiveness signature. Expression markers unique for these identified subsets of cells were shown to associate with lymph node staging and progression free survival, confirming the role of the identified cell types in driving the metastatic potential of a primary tumor lesion. DEG: differentially expressed gene; NL: adjacent normal samples, RP: primary sample, MLN: regional metastatic lymph node sample Created with BioRender.com released under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International license (https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en)

one regional MLN (unifocal design) and for 6 patients RNA-seq data for multiple primary regions (for an overview of the cohorts, see Additional file 1: Table S1). The Pearson correlation [73] between the gene expression values of the different primary and metastatic regions suggests a higher transcriptomic heterogeneity of primary tumors compared to metastatic tumors (Additional file 2: Text S1).

Identifying a signature-based metric to assess a lesions' metastatic potential

Exploiting this multifocal cohort to identify transcriptomic signals that associate with regional lymph node seeding first requires identifying the lesions that most likely were responsible for seeding the regional lymph node metastases. We assume that in the primary tumor these seeding lesions should exhibit expression characteristics more similar to those of the metastatic lymph node lesions than their less aggressive non-seeding counterparts. The gradual change in the progression status, ranging from a normal to a more aggressive metastatic state of the different lesions is indeed captured by the expression data, as is illustrated by the Spearman's rank PCA [45, 46] of the expression profiles of the 210 lesions sampled across both cohorts (see Additional file 2: Text S2). However, as these gradual changes in gene expression reflect, in addition to properties associated with an evolving tumor, also regular tissue adaptation, we disentangled the metastatic progression signal from these confounding tissue adaptation effects. Hereto we performed a meta-analysis of the locally advanced cohort, the *de novo* cohort and TCGA-PRAD [32] to identify those genes that were differentially expressed between the NL vs. RP samples, but additionally also between the RP vs. MLN samples and that explained mostly the variation in expression that associates with metastatic progression (see Fig. 1B, Additional file 2: Text S3 and Methods). Because they keep changing their expression during tumor progression, regardless of their tissue of origin (prostate or lymph node), the expression of the eventually selected 225 genes will be referred to as the progression signature (Additional file 1: Table S12).

As shown in Fig. 1C these 225 genes capture mostly the progression signal, regardless of the clinical characteristics of the analyzed cohorts.

Progression signature-based metric to define the metastatic potential of a lesion

Based on this progression signature, we subsequently defined a metric that allows assessing per patient the potential of the different lesions in the primary tumor for regional lymph node metastasis (further referred to as metastatic potential). Similarity in terms of the progression signature is based on the within-patient PC1 scores of the different samples obtained by performing Spearman's rank PCA [45, 46] using the genes from the progression signature as variables (Fig. 1C and Additional file 2: Text S3). We subsequently identified per patient the primary lesions with the pronouncedly higher metastatic potential as those for which the PC1 scores are closer to the scores of the matching metastatic lesions and sufficiently distinct from the scores of the other primary lesions within the same patient (see Methods and Additional file 2: Text S4). We applied this analysis to patients for whom RNA profiling is available for multiple primary samples and their corresponding lymph node metastases (i.e. 17 patients of the locally advanced cohort and 6 of the *de novo* cohort).

Overall, Fig. 2 shows that in the *de novo* cohort some lesions have a pronouncedly higher metastatic potential than the other lesions within the same patient. In the locally advanced cohort this difference also exist, but is less extreme. This shows that in patients with *de novo* metastatic disease, at least some lesions with very strong metastatic potential exist, which aligns with the clinical characteristics of those patients who already show distant metastases at diagnosis.

Based on the defined criteria, we could pinpoint in 12 of the 17 patients of the locally advanced cohort at least one primary lesion with a considerably higher metastatic potential (Fig. 2A). For five patients this was a single lesion and for seven patients multiple lesions were identified. For the remaining 5 patients all lesions had a comparable metastatic potential in terms of the progression signature.

For the *de novo* cohort, in 5 of the 6 patients with multiple primary lesions sequenced, a single primary lesion was identified with high metastatic potential (Fig. 2B). In one patient (ID3), none of the primary lesions selected for sequenced has a high metastatic potential according to our metric, suggesting that the most aggressive lesion that eventually gave rise to the regional lymph node metastases was not sequenced.

Similarity to the matching metastatic lesions captures a lesion's metastatic potential

To confirm that similarity to the matching metastatic lesions in terms of the progression signature truly reflects a lesion's metastatic potential, we used tNGS data as an alternative to identify the most likely seeding lesions. For all patients with available tNGS, we identified as a proxy for lineage reconstruction, the RP lesion within a patient that shared most mutations with its matching MLN lesions (see Methods, Additional file 1: Table S13 and Additional file 2: Text S5), as this primary lesion is the most likely ancestor of the matching lymph node metastases and, therefore, also most likely responsible for its seeding.

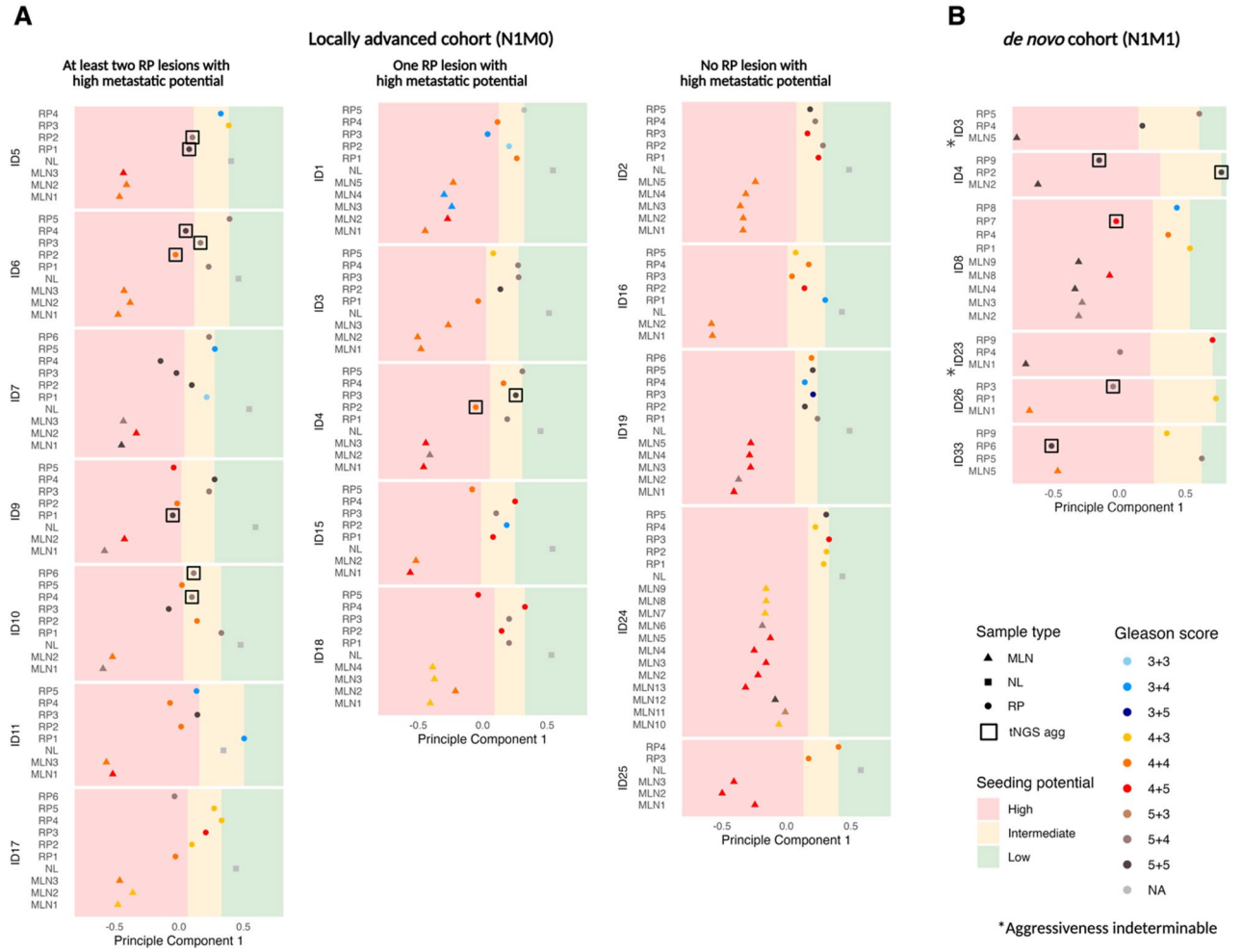

Fig. 2 Characteristics of the lesions obtained from patients in the locally advanced cohort (**A**) and the de novo cohort (**B**). Gleason score data was obtained from Additional file 1: Table S2 and S5. The lesions are colored according to their metastatic seeding potential based on the progression signature. The boundaries of the levels of metastatic seeding potential (low, intermediate and high) are set as described in Additional file 2: Text S4. The lesions identified to be seeding by tNGS data (tNGS agg) are boxed (only indicated for the subset of lesions for which targeted sequencing data is available). *Indeterminable, indicates that the tNGS was available, but that the number of identified somatic mutations was insufficient to determine the most likely seeding lesion. Lesions are characterized by different symbols and colors: Symbols reflect the origin of the sample (NL: adjacent normal tissue, RP: lesion from the primary tumor obtained after radical prostatectomy, MLN: metastatic lesion extracted from regional lymph nodes. Colors reflect the ISUP grade. The patients of the locally advanced cohort are subdivided in three columns depending on whether according to the progression signature-based metric 0, one or more lesions with high metastatic potential could be identified. The lesions with the highest metastatic potential are by definition those lesions that are located closer to the MLN lesions

In the *de novo* cohort, a clear one to one overlap exists between the seeding lesion identified with tNGS and the lesion with the highest metastatic potential, identified based on the signature-based metric (Except for ID4 where tNGS identified two lesions). This confirms that in this cohort often a single lesion has a pronouncedly higher metastatic potential than the remainder of the lesions within the same patient and it is likely also this lesion that drove metastatic seeding (Fig. 2B).

In the locally advanced cohort, all likely seeding lesions (except in patient ID10) identified based on the tNGS analysis correspond to lesions that, according to the signature-based metric show a high metastatic potential (Fig. 2A and Additional file 2: Text S6). In some cases, but not all (not in LA-ID4, LA-ID6, and LA-ID10) the lesions identified by tNGS also exhibited the highest metastatic potential according to the signature-based metric. This suggests that within the locally advanced cohort, there are lesions with relatively comparable metastatic potential. Some of those have been responsible for seeding the lymph node metastases, while the others might not have yet or the metastatic lesions seeded by those lesions have not been sequenced with tNGS (see Additional file 2: Text S5 for an in depth analysis).

Important for subsequent analyses, in all cases (except *de novo* ID4), none of the lesions that had, according to

the signature-based metric a lower metastatic potential (least aggressive) were identified as seeding lesions by the tNGS. For *de novo* ID4, both called seeding lesions (RP2 and RP9) shared an equal number of mutations with the metastatic lesion. Although this seems to indicate that both lesions might have been seeding metastasis, their large difference in metastatic potential according to the signature-based metric suggests this is unlikely.

The fact that the lesions indicated by the tNGS as responsible for metastatic seeding are also the ones that have the higher metastatic potential according to our progression signature-based metric, confirms that our metric captures a lesion's potential for seeding regional lymph node metastases.

We indicated in Fig. 2 for each of the lesions also their corresponding Gleason score. Lesions with a low Gleason score ($<3+x$) are almost always classified according to our metric as having a low metastatic potential and are never identified as seeding lesions according to tNGS. However, for higher Gleason scores ($4+3$, $4+5$, $5+4$ or $5+5$), it is surely not always the lesion with the highest Gleason score that also is the one responsible for the observed metastatic seeding (as determined by tNGS) or that has the highest metastatic potential, according to the signature-based metric. This confirms earlier observations [74]. In the locally advanced cohort, only five of the 17 patients, the lesion with the highest metastatic potential also has the highest Gleason score.

Tumor aggressiveness signature, a gene set of which the expression is associated with a lesion's metastatic potential

The ability to pinpoint with our signature-based metric per patient the lesions with respectively the highest and lowest potential for regional lymph node seeding enabled a comparative expression analysis to identify gene sets for which the expression associates with this metastatic potential without being affected by the inter-patient variability. To avoid also here batch effects and effects due to intrinsic biological differences between the analyzed multifocal cohorts, differential expression analysis between the lesions with respectively the highest and lowest metastatic potential was performed on each cohort separately (respectively the locally advanced and the *de novo* cohort) and results were combined after differential expression analysis (see Methods).

To generate a final list of genes for which expression changes are associated with a lesion's metastatic potential, we started from all genes that were differentially expressed between the lesions with the highest and lowest metastatic potential (intersection of both gene lists) in both cohorts. However, given the intrinsic differences in clinical characteristics between both cohorts, we also added the genes that were differentially expressed in one cohort only, provided they passed a strict significance threshold (Methods and Additional file 2: Text S7).

This gene set referred to as the ''tumor aggressiveness signature" consists of 268 downregulated and 159 upregulated genes (Additional file 1: Table S18) of which 51 are also part of the progression signature (overlap of 11.94%). The upregulated genes mostly belonged to pathways involved in mitotic spindle, spermatogenesis, G2M checkpoint, E2F targets hallmark pathways, whereas the downregulated were mostly related to myogenesis, KRAS signaling, apoptosis, hedgehog signaling, apical surface, adipogenesis and angiogenesis.

Cell types associated with the aggressiveness signature

To identify the cell types responsible for the observed changes in expression between lesions with high and low metastatic potential, we overlaid the metastatic progression signature with single-cell data obtained from 11 primary prostate tumors of Chen et al. (2021) [8] (Additional file 2: Text S8). From the aggressiveness signature, 380 genes were measured in Chen et al. (2021) [8] and the expression of 330 of these (86.84%) was predominantly associated with one particular cell type: 120 were relatively more expressed in luminal and basal/intermediate epithelial cells compared to the rest of the cells, 146 in stromal cells (fibroblast and endothelial cells) and a minor fraction (64 genes) in immune cells (i.e., monocytic, mast, T and B cells). This indicates that the differences between lesions with high and low metastatic potential in the primary tumor are mostly driven by changes in the expression and/or composition of epithelial and stromal cells.

Highly proliferating luminal cells are associated with a lesion's metastatic potential

To identify the subset of luminal epithelial cells that associate with the expression of genes in the aggressiveness signature, we performed PCA [45] on the Chen et al. (2021) dataset [8] using the genes from the aggressiveness signature as variables. Figure 3A, left panel, shows that based on the first PC, a subset of 555 luminal cells deviates from the profile of the overall population of luminal cells in their expression of the genes in the signature. Using as variables for the PCA [45], either the same number of randomly selected genes or all genes did not identify the same set of luminal cells (Additional file 3: Fig. S5), indicating that these identified luminal cells undergo changes specifically captured by the aggressiveness signature and are likely contributing to regional lymph node seeding.

Indeed all of the genes most relevant for subsetting these cells (i.e. all 30 genes from the aggressiveness signature with the highest loading on PC1) were specifically upregulated in the identified luminal cells as compared

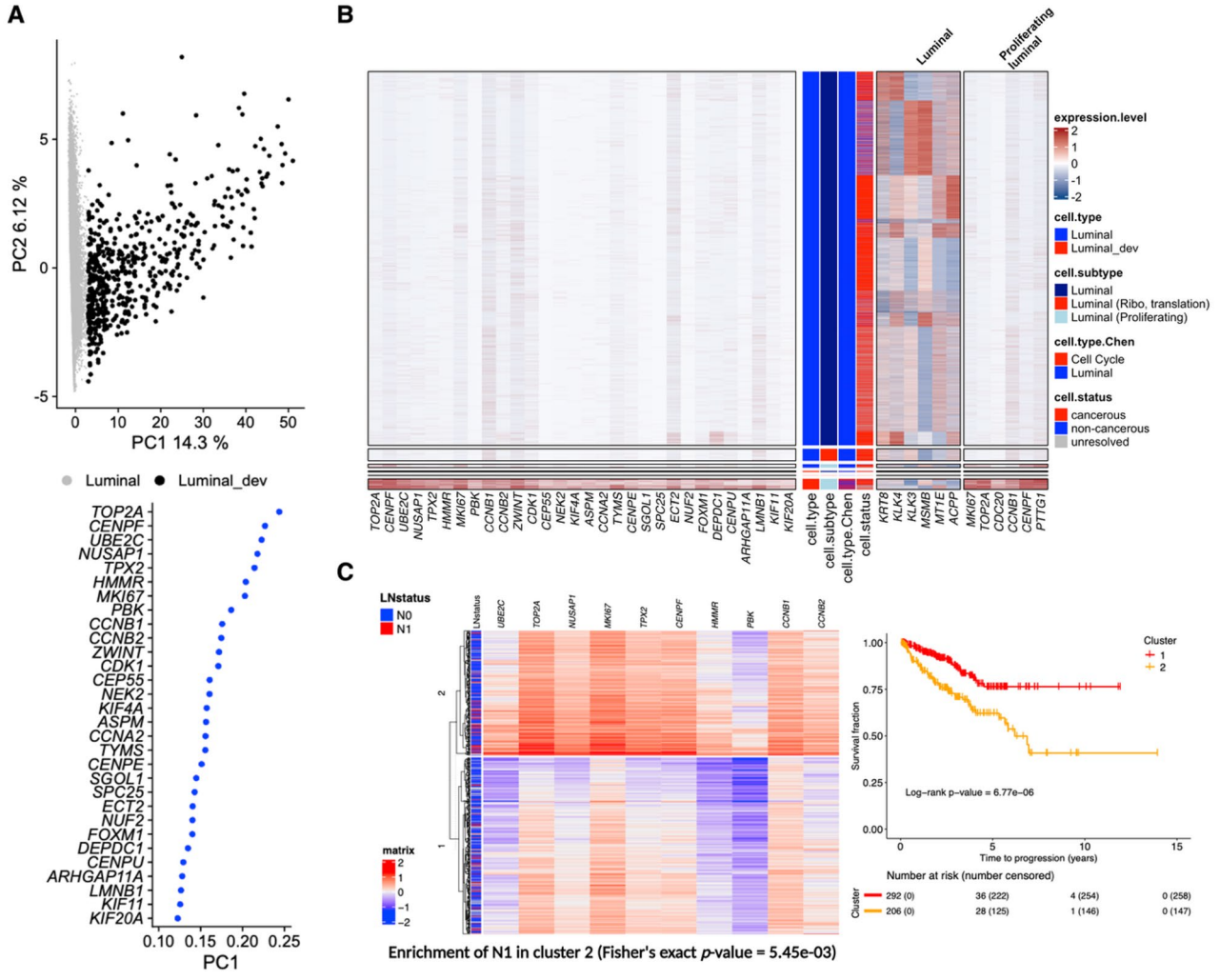

Fig. 3 Identification of luminal cells, associated with lymph node seeding. **A.** PCA of luminal cells of the primary samples from Chen et al. (2021) using the genes from the aggressiveness signature as variables (left panel). The 555 cells highlighted in black exhibit deviating expression properties based on the first PC. Loadings of the genes of the aggressive signature on the first PC (right panel). The 30 most influential genes for the first principal component are shown. **B.** Heatmap of the expression in the luminal cells of the 30 genes from the aggressiveness signature with the top 30 highest loadings on PC1 that were also specifically upregulated in the deviating (luminal_dev) (cell.type indicated in red) as compared to the remaining luminal cells in the Chen et al. dataset (cell.type indicated in blue). Middle panel with color bars in the heatmap indicates the subdivision of the luminal cells based on respectively (1) the aggressiveness signature, (2) marker genes in luminal cells (proliferating luminal cells and ribo, translation luminal cells), (3) the cell types defined by Chen et al. (2021), and (4) the cancerous status as identified by CNV analysis. Right panel in heatmap: expression of literature-based marker genes used to annotate the luminal cells. **C.** Prognostic value of the 10 gene signature, representative of identified proliferating luminal cells. Left panel clustering of TCGA-PRAD samples according to the 10 gene signature with the p-value showing the enrichment in lymph node positive samples, right panel showing the results of the survival analysis of the two clusters. The cluster enriched in lymph node positive samples is the one that exhibits the lowest survival

to the remainder of the luminal cells (FindMarkers [49], Additional file 1: Table S19) and show an expression pattern that is highly specific for these deviating luminal cells [8] (Additional file 3: Fig. S6). Of these 30 genes of the bulk-level aggressiveness signature, 19 were significantly upregulated in the lesions with high metastatic potential in both the locally advanced and *de novo* cohorts (i.e., *ARHGAP11A*, *ASPM*, *CCNA2*, *CCNB1*, *CCNB2*, *CDK1*, *CENPE*, *CENPF*, *CENPU*, *CEP55*, *DEPDC1*, *ECT2*, *HMMR*, *KIF11*, *NUF2*, *PBK*, *SGOL1*, *SPC25* and *TOP2A*) suggesting their expression is a strong signal that is present in the bulk RNA, regardless of the clinical differences between both analyzed cohorts.

Genes (FindMarkers [49], adjusted p-value < 0.001, Additional file 1: Table S19) marking these deviating luminal cells versus the remainder of the luminal cells are specifically overrepresented in proliferative pathways G2M checkpoint (with as representatives *TOP2A*, *PBK*, *UBE2C*, *TPX2*, *NUSAP1*, *HMMR*, *KIF4A*, *CENPF*, *NEK2*, *CDK1*, *CCNA2*, *MKI67*, *CCNB2*, *CENPE*, *POLQ*, *KIF11*, *LMNB1*, *ORC6* and *TTK*) and E2F targets (*TOP2A*, *SPC25*, *DLGAP5*, *HMMR*, *KIF4A*, *E2F8*, *MELK*, *CDK1*,

MKI67, *CCNB2*, *CENPE*, *BUB1B*, *LMNB1*, *RAD51AP1*, *DEPDC1* and *ORC6*) (Additional file 1: Table S20).

Based on CNV analysis (see Methods), this identified subset of ''proliferative" luminal cells is enriched in cancerous cells (p-value of 0.000535 (Fig. 3B and Additional file 2: Text S9). Interestingly, these identified proliferating luminal cells (see Fig. 3B) contain 94% of the "Cell Cycle" cells described in Chen et al. (2021) [8]. These "Cell Cycle" cells were considered "high risk cells" by Chen et al. (2021) [8] because of their association with high expression of luminal B, hypoxia, and PCS1 signatures [8]. In line with Chen et al. (2021) [8]) six of the 12 genes, specifically expressed in the PCS1 subtype [75] were present in the metastatic aggressiveness signature (i.e., *HMMR*, *MKI67*, *FOXM1*, *TPX2*, *KIF11* and *CCNB1*).

Conclusively the aggressiveness signature associates with a subset of highly proliferative luminal cells. The fact that these cells are identified by an expression signature that characterizes primary tumor lesions with high metastatic potential and involvement in the seeding of regional metastatic lymph nodes supports the hypothesis of Chen et al. (2021) [8] that proliferative luminal cells are linked to tumor aggressiveness.

Next to luminal cells, the metastatic progression signature also contains genes of which the expression is more specifically associated with basal/intermediate epithelial cells. However, using the aggressiveness signature to identify cells that distinguish themselves from the remainder of the basal /intermediate cells, identified three cells only (Additional file 3: Fig. S7). Although those cells could not be identified with a random signature nor with all genes (Additional file 3: Fig. S8), their limited number impeded an in-depth characterization.

Adipose-derived CAFs are associated with a lesion's metastatic potential

Using a similar strategy, we identified, based on the aggressiveness signature, a subpopulation of 115 fibroblast-like cells that deviate in their expression from the remaining fibroblasts in the Chen et al. (2021) dataset [8]. This subpopulation of cells set apart by the aggressiveness signature is mainly captured by the second PC (Fig. 4A) and can partially be obtained using all genes, but not with a random signature of the same size as the aggressiveness signature (Additional file 3: Fig. S9). This indicates that the identified cells are quite different from the remainder of the fibroblasts in Chen et al. (2021) [8]. To assess the specificity of the bulk signature for the identified fibroblast-like cells, we assessed how many of the genes with a high loading on the second principal component (PC2) were also specifically upregulated in the identified fibroblasts as compared to the remainder of the fibroblasts (FindMarkers [49]) in the Chen et al. dataset (2021) [8]. This was the case for 24 of the 30 (80%) (see Fig. 4B and Additional file 1: Table S21). In addition, Additional file 3: Fig. S10 show that the expression of those genes is also quite specific for the selected subset of deviating fibroblast-like cells as compared to cells other than the fibroblasts in Chen et al. (e.g., *PTGDS*, *SERPINEF1*). This shows that the aggressiveness signature captures specific expression signals marking these fibroblast-like cells.

A fine-grained annotation of the fibroblast like cells using literature derived marker genes (see Methods) shows that most of the deviating cells, set apart by our signature (Fig. 4B) differentiate from the remainder of fibroblast cells in Chen et al. (2021) [8] by markers that have previously been associated with specific subsets of matrix CAFs [58, 69], but also by the expression of pre-adipocyte markers and markers that differentiate ASCs (adipose derived stromal stem cells) from fibroblasts [71].

Interestingly, the expression of several of these markers and a subset of the 24 genes from the bulk aggressiveness signature (*PTGDS*, *MMP2*, *IGFBP6*, *GSN*, *SERPINF1*, *MRC2* and *BOC*) that characterize these identified fibroblast like cells, have been associated with a transitioning from ASCs to a fibroblast (like) cell state in normal pancreatic cells [59], which was marked by the expression of *APOD*, *PDGFRA*, *DCN*, *LUM*, and *PTGDS*. The heterogeneity in the expression of several of the ECM remodeling genes (*LUM*, *GSN* and *DCN*) marking the deviating cells in the Chen et al. (2021) identified by the aggressiveness signature reflect the different transitioning states that were also observed in this study of Zhu et al. (2021) [59].

In the same study, Zhu et al. (2021) [59] describe an alternative attractor state of those ASCs in tumor cells, indicative of a very aggressive disease state, just preceding metastasis [59]. This state is marked in Zhu et al. (2021) [59] by the expression of *COL11A1*, *INHBA* and *THBS2*. Interestingly, a subset of the cells identified by our signature in the Chen et al. (2021) dataset [8] (referred to as adipose derived CAFs in Fig. 4B) is characterized by the expression of genes that in pancreatic cancer associate with this transitioning towards this alternative COL11A1-expressing attractor state (*SERPINF1*, *MRC2*, *MMP2* and *CTGF*). This subset of cells exhibits a reduced expression of the genes that mark the transition between ASCs and a normal fibroblast like cell state (*PTGDS*, *IGFBP6*, *GSN* and *APOD*), but instead display an increased expression of an additional set of CAF-ECM markers that associate with the aggressive COL11A1-expressing state (*COL11A1*, *CTHRC1*, *COMP*, *COL10A1*, *INHBA*, *FN1*, and *THBS2*) described in pancreatic cancer [59] but also observed in a pancancer analysis of CAFs [69].

The bulk derived aggressiveness signature thus captures both the cells that undergo a transitioning from ASCs to fibroblast (like) cells that occurs in normal tissue as well

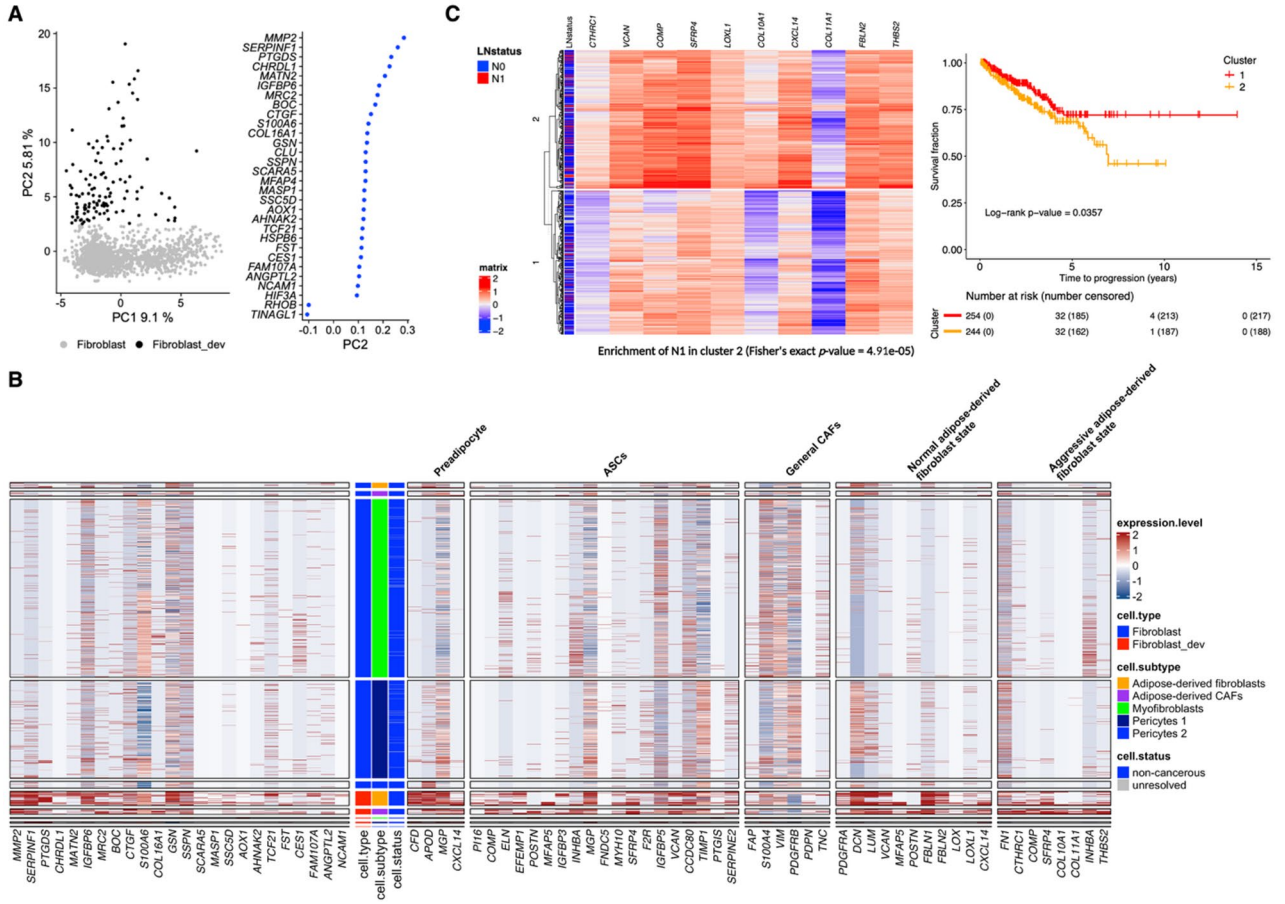

Fig. 4 Identification of adipose derived fibroblast like cells, associated with a lesion's metastatic potential. **A.** PCA of fibroblasts of the primary samples from Chen et al. (2021) based on the genes from the aggressiveness signature (left panel). The 115 cells highlighted in black exhibit deviating expression, based on the second PC. Loadings of the genes of the aggressiveness signature on the second PC (right panel). The 30 most influential genes for the second principal component are shown. **B.** Left panel: Heatmap of the expression in the fibroblasts of the 24 genes with the top 30 highest loadings on PC2 that were also specifically upregulated in the deviating (cell.type indicated in red, Fibroblast_dev) versus the remaining fibroblasts in the Chen et al. dataset (cell.type indicated in blue). Middle panel: color bars indicating the annotation obtained by (1) the aggressiveness signature (fibroblast_dev: deviating fibroblasts) and (2) a literature-based marker annotation. Right panel: Expression heatmap of the marker genes used for the annotation. **C.** Prognostic value of a 10 gene signature, representative of the transitioning towards the aggressive adipose derived CAF state (Legend as in Fig. 3)

as to a CAF state, reminiscent of a very aggressive disease state, observed in pancreatic cancer.

The 24 genes of which the expression is specifically associated with adipose derived fibroblast-like cells were found to be relatively more downregulated in the lesions with the higher than in the lesions with the lower metastatic potential in both of the analyzed cohorts (Additional file 1: Table S14 and S15). This indicates that a subset of genes we used to define the distance to the seeding lesion with our progression signatures associates with this normal fibroblast like cell state, which tends to disappear as tumor lesions in the primary tumor become relatively more aggressive. We, therefore, hypothesize that in the seeding lesion the transitioning between ASCs and the normal adipose derived fibroblast like cells shifts towards this presumably more aggressive CAF state.

This would imply that the genes marking this more aggressive disease state would be upregulated in the seeding versus the non-seeding lesions. In both locally advanced and *de novo* cohort, we observe indeed an upregulation of *COMP*, *COL10A1* and *COL11A1* in the seeding versus the non-seeding state, albeit none of them were significant (Additional file 1: Table S14 and S15). The likely small number of cells undergoing this transitioning relative to the total number of cells, in combination with the fact that the expression of some of the ECM remodeling markers that characterize this more aggressive disease state (*FN1* and *INHBA*) are not exclusive to the adipose derived CAFs (Fig. 4B), dilutes or even misses their bulk signal, making it infeasible to distinguish differences in their expression signal between lesions with high and low metastatic potential in the bulk data.

A specific subset of arterial endothelial cells is associated with a lesion's metastatic potential

We used the same PCA-based approach [45] as for the luminal and fibroblast cells to identify a subpopulation of deviating endothelial cells. Prior to the analysis, the pericyte-endothelial doublets were removed from the Chen et al. (2021) dataset [8] as their presence obviates finding an expression signal specific for endothelial cells (Additional file 3: Fig. S12). The aggressiveness signature set apart a subset of arterial endothelial cells (see Methods, Additional file 1: Table S23 and Fig. 5A), mainly captured by the second PC. The same subset of cells could partially be obtained using all genes, but not with a random signature of the same size as the aggressiveness signature (Additional file 3: Fig. S12), showing again that the aggressiveness signature is enriched in pathways that characterize those endothelial cells.

Of the 30 genes with highest loading on PC2, most relevant to subsetting these cells 27 (90%) were also identified by FindMarkers [49] as differentially upregulated in the identified arterial endothelial cells as compared to the remaining endothelial cells in the Chen et al. (2021) dataset [8] (Additional file 1: Table S23). Figure 5B and Additional file 3: Fig. S13 show how, based on these markers, the identified arterial endothelial cells differ from the remainder of the endothelial cells in their high expression of, for instance, *PTGS2*, which is involved in promoting angiogenesis in ischemic conditions [76], *MMP2*, an

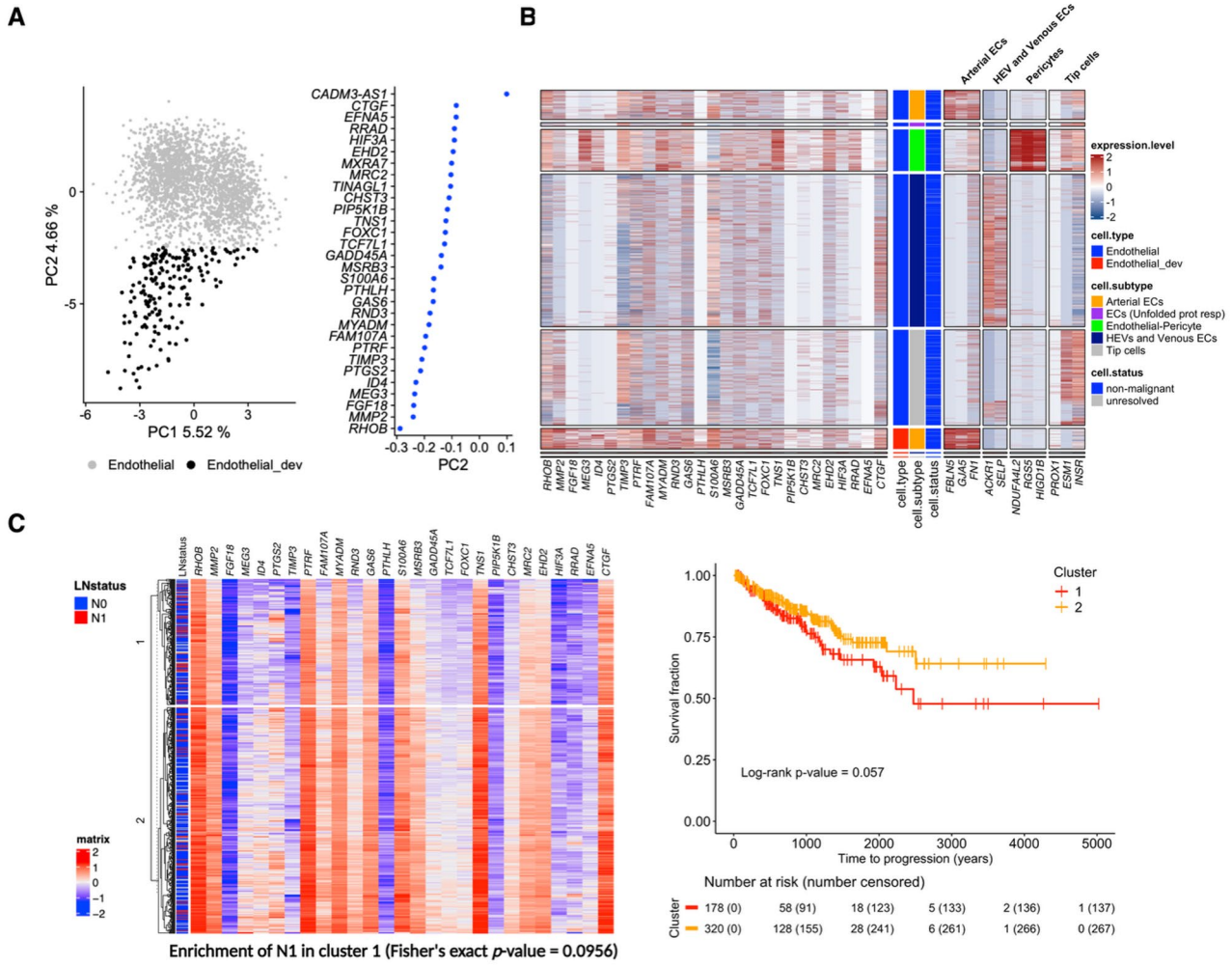

Fig. 5 Identification of arterial endothelial cells, associated with lymph node seeding. **A.** Left panel: PCA of endothelial cells without the pericyte-endothelial doublets of the primary samples from Chen et al. (2021) based on the genes from the aggressiveness signature. The 187 endothelial cells highlighted in black exhibit deviating expression, based on the second PC. Right panel: Loadings of the genes of the aggressiveness signature on the first PC. The top 30 genes, most influential for the second PC are shown. **B.** Left panel: Heatmap of the expression in the endothelial cells of the 27 genes with the top 30 highest loadings on PC2 that were also differentially upregulated between the deviating (cell.type indicated in red) and the remaining endothelial cells in Chen et al. (cell.type indicated in blue). Middle panel: color bars indicating the annotation obtained by (1) the aggressiveness signature (endothelial_dev: endothelial), (2) an independent marker-based annotation. Right panel: Heatmap of the marker genes used for the independent annotation. **C.** Prognostic value of a 27 gene signature, representative of the arterial endothelial cells associated with lymph node seeding (Legend as in Fig. 3)

ECM-degrading enzyme, which is a.o. involved in angiogenesis and of which the expression has been associated with aggressive PCa [77], ID4 of which ectopic expression results in AR-dependent tumor suppressive activity in PCa [78] and *FGF18* which is a bioactive substance, so far mostly associated with fibroblasts of which the activity has been associated with tumor progression and angiogenesis in various cancer types [76]. These genes tend to be downregulated in the lesions with high versus those with low metastatic potential (Additional file 1: Table S23).

Prognostic value of the identified cell type specific signatures associated with metastatic potential

As shown above, combining our multifocal cohort with single-cell data allowed associating a subset of luminal cells, adipose derived CAFs and arterial endothelial cells with a primary lesion's metastatic potential. To confirm the identified cell types indeed associate with metastatic progression, we assessed their prognostic potential in TCGA-PRAD [32]. Hereto we clustered the primary tumor samples in two groups, using the gene sets that mark each of the identified cell types (Methods, Additional file 2: Text S10) and assessed for each grouping the enrichment in lymph node positive (N1) samples and the association with progression free survival (Fig. 6).

The expression of the 30 genes marking the proliferative luminal cells (Fig. 6B, Additional file 3: Fig. S14) is significantly associated with both lymph node status and progression-free survival (higher than a subdivision of the samples based on the lymph node status, *p*-value of 0.011). Reducing the 30 gene signature to first the top 10 markers of these proliferative luminal cells (i.e., *TOP2A*, *CENPF*, *UBE2C*, *NUSAP1*, *TPX2*, *MKI67*, *HMMR*, *PBK*, *CCNB1* and *CCNB2*, i.e. 10 genes largely contributing to the PCA [45] in Fig. 3A and Additional file 3: Fig. S6) and subsequently to the top 5 (i.e., *TOP2A*, *CENPF*, *UBE2C*, *NUSAP1*, *TPX2*) gradually increases the association with lymph node status, at the expense of reducing the association with progression-free survival (Figs. 4C and 6B, Additional file 3: Fig. S15). This indicates that the more specific the genes set is for the subset of proliferating luminal cells we set apart with the aggressiveness signature, the more prognostic they are for the metastatic potential specifically associated with regional lymph node seeding.

The full 24 gene set that marks transitioning between ASCs and adipose derived fibroblasts is weakly associated with lymph node status, albeit not significantly, and marginally prognostic for progression-free survival (Fig. 6B, Additional file 3: Fig. S6). Assuming that this 24 gene signature reflects not only the transitioning to a more aggressive state, but also progression from ASCs to normal adipose-derived fibroblasts, we also designed a gene signature that more specifically captures the transitioning towards the presumedly more aggressive *COL11A1*-expressing state (Fig. 4C). Hereto, we selected a subset of 10 genes marking the transitioning to the *COL11A1* aggressive state (Fig. 4B), all of which were also identified by FindMarkers [49] as genes that distinguish the adipose derived fibroblast like cells in Chen et al. (2021) [8], from the remaining fibroblasts (see Additional file 1: Table S21 and Methods).

Clustering TCGA-PRAD samples [32] based on this cell state specific 10 gene signature (i.e., *CTHRC1*, *FBLN2*, *VCAN*, *COMP*, *SFRP4*, *LOXL1*, *COL10A1*, *CXCL14*, *COL11A1*, and *THBS2)* increases the prognostic performance as compared to the larger, 24 gene signature for the association with lymph node status but not for overall survival (Figs. 4C and 6B).

To exclude that the observed prognostic signal could be attributed to cell types, other than the identified CAFs we constructed a reduced signature so that the expression of the genes contained in that signature was mostly confined to the transitioning fibroblast like cells in the dataset of Chen et al. (2021) [8] (so not present in any other cell type as shown in Additional file 3: Fig. S11) (i.e., *VCAN*, *SFRP4*, *THBS2*, *COL10A1* and *COL11A1*). The high prognostic value of this reduced 5 gene signature for lymph node staging unequivocally supports the role of the identified transitioning of adipose derived fibroblasts towards aggressive adipose derived CAFs in driving a tumor's potential for regional lymph node seeding (Additional file 3: Fig. S17).

The 27 marker gene set of the arterial endothelial cells divide the primary tumor samples of TCGA-PRAD [42] in two expression clusters which show no significant association in both with progression-free survival (p-value = 0.0570) and a weak and with differential enrichment in N1 primary samples (p-value = 0.0956, Figs. 5C and 6B), confirming that the genes characterizing this subset of arterial endothelial cells carry a signal associated with metastatic progression, albeit marginal as compared to the signal carried by respectively the proliferative luminal cells and the adipose derived fibroblasts.

These results confirm that the cells transitioning towards an adipose derived aggressive CAF state identified by the aggressiveness signature carry a prognostic signal for metastatic seeding of regional lymph nodes, whereas the presence of proliferative luminal cells captures a lesion's general metastatic potential, proxied by both lymph node status and progression free survival.

To assess whether the presence of both cell states contain a complementary prognostic signal, we also tested the prognostic value of a 10 gene signature, composed of the 5 gene signatures, representing respectively the highly proliferative luminal cells and the cells transitioning towards adipose derived CAFs. This combined

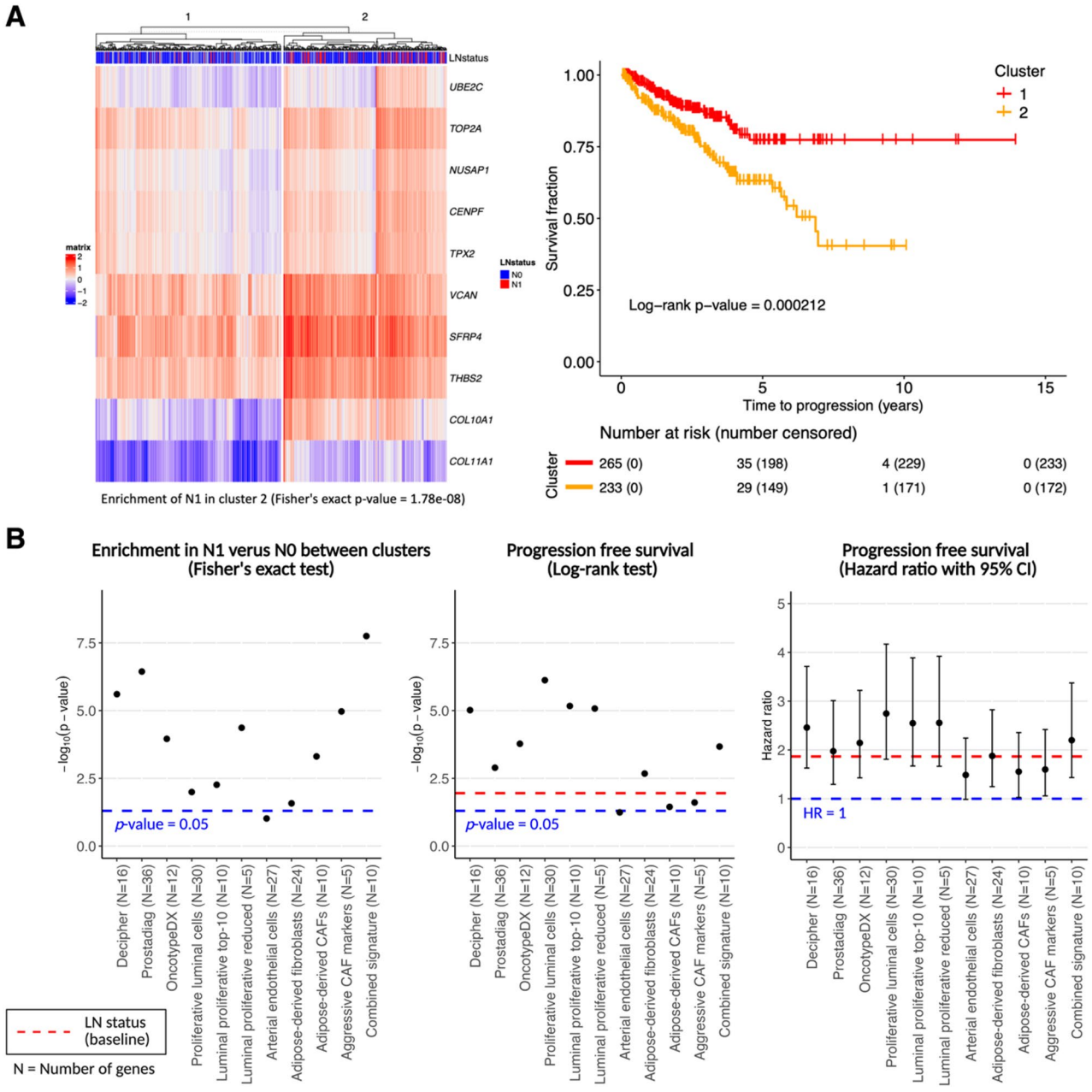

Fig. 6 Prognostic value of the tested signatures. **A.** Prognostic value of a 10 gene combined signature, representative of the proliferating luminal cells and adipose derived CAFs associated with lymph node seeding (Legend as in Fig. 3). **B.** TCGA-PRAD samples were subdivided into two groups using each of the gene signatures and LN status as a baseline. Details on the signatures can be found in Additional file 1: Table S25 and Additional file 2: Text S10. Number of genes: the number of genes in the signature. Fisher's exact test: p-value of the enrichment in lymph node positive samples (N1) of the indicated cluster. Log-rank test: p-value of the difference in progression-free survival between two groups. Hazard Ratio: relative hazard of the group enriched in lymph node positive samples compared to the other group. Note that Decipher, Prostadiag and OncotypeDX here refer to the genes contained in the commercial signatures and not to the read outs of the corresponding commercial tests

signature increases drastically the association with lymph node status as compared to the association obtained with either of the cell type specific signatures alone (Fig. 6), indicating that the expression signal derived from respectively proliferative luminal cells and adipocyte derived CAFs reinforce each other in grouping samples according to their lymph node status. For progression free survival, the signal provided by luminal cells only seems more prognostic than the combined signal. The signal from the TME which appears less informative for progression free survival here dilutes the prognostic signal of the proliferative luminal cells.

Markers associated with proliferating luminal cells and adipose derived fibroblasts determine the prognostic signal of commercially available markers

As a reference we also assessed the prognostic signal for respectively lymph node status and progression free survival contained in the genesets that characterize the commercial signatures Decipher [18], OncotypeDX [20], and Prostadiag [33] (Fig. 6B). Hereto we applied the same strategy as used for our own genesets, described above and in Methods (assessing the prognostic potential of the identified cell types).

Our 10 gene combined signature, representing both the proliferative luminal cells and the transitioning to aggressive adipose derived CAFs shows an association with lymph node status that exceeds that of the genes contained in the commercial signatures. For progression free survival our three gene sets specific for proliferative luminal cells outperform the genesets contained in commercial signatures. We, therefore, wondered whether the same cell types we associated with increased metastatic potential can explain the prognostic signal of the commercial signatures.

Hereto, we plotted the expression of the genes used by respectively Decipher [18], Prostadiag [33], and OncotypeDX [20] across the cells, marked by Chen et al. (2021) [8] as luminal, fibroblast and endothelial cells (see Additional file 3: Fig. S18-S20). Several of the genes contained in these commercial signatures indeed correspond to genes specifically expressed in the cell types, we associated with high metastatic potential (for an overlap between the genes in each of the signature and the cell type specific genes, see Additional file 1: Table S19, S21 and S23).

Decipher [18] and OncotypeDX [20] contain respectively two genes (*NUSAP1* and *UBE2C*) and one gene (*TPX2*) marking the highly proliferative luminal cell state, we identified with our aggressiveness signature (Additional file 3: Fig. S18 and S21). Prostadiag [33] also contains markers of luminal cells, but not those associated with the highly proliferative luminal cell state (Additional file 3: Fig. S18). This explains its relatively lower predictive potential for progression-free survival than e.g., Decipher [18] or than our signatures which uniquely mark these proliferative cell types. These results confirm that progression free survival is mostly driven by the properties of the proliferative luminal cells.

According to our results, genes representing proliferative luminal cells and the cells transitioning towards aggressive CAFs associate with lymph node status. Representing both cell states in a single signature largely improved this association over the use of markers of each single cell state only (Fig. 6B). Both Decipher and OncotypeDX contain genes, marking the transitioning from ASCs to the adipose derived CAFs (Additional file 3: Fig. S19 and S21): i.e. with respectively 2 genes in OncotypeDX (COLA1 and SFRP4) [20] and one gene in Decipher (THBS2) [18], while as mentioned above both signatures also contain markers of proliferative luminal cells. In Prostadiag several markers quite uniquely define the transitioning from ASCs to adipose derived CAFs (i.e.,, COL10A1, COL8A1, *COMP*, *CXCL14*, *MOXD1*, *NOX4*, *SFRP2*, *SFRP4*, *CXCL14*, *THBS2* and *VCAN* (Additional file 3: Fig. S19 and S21)) [33]. The fact that the prognostic value of these larger commercial signatures is comparable to that of our 10-gene combined signature, which was designed to specifically capture the presence of both the highly proliferative luminal cell state and the transitioning towards an aggressive adipose-derived CAF state shows that the prognostic value of these larger commercial signatures is mostly driven by genes associated with these aforementioned cell states.

Each of the commercial signatures contain also genes that mark endothelial cells but these are mostly endothelial pericytes. Only Prostadiag [33] had markers that associate with the vascular endothelial cells we set apart (*SULF1*), albeit not uniquely, as these genes are also expressed in the adipose derived CAFs (Additional file 3: Fig. S20).

Discussion

In this study, we leverage the statistical power of multifocal bulk transcriptome profiling with the cellular resolution of a single-cell based PCa study [8] to identify processes and cell types that associate with the metastatic potential of a primary tumor lesion, in particular with regional lymph node seeding.

To cope with intratumor heterogeneity, we profiled the expression of a large multifocal cohort of matching primary and metastatic regional lymph node lesions. Sampling per patient multiple lesions in the primary and metastatic tumor enables profiling and comparing the expression of lesions with different metastatic potential. To perform such comparison, we first identified a unique expression derived metric to sort lesions based on their metastatic potential. The additional availability of adjacent normal and treatment-naive regional MLN samples of the same patients allowed extracting an expression signature that captures the gradual progression from normal tissue towards regional lymph node metastases. We defined the metastatic potential of a lesion in terms of this progression signature, assuming alike [31] that the more similar a primary lesion is in its expression behavior to its matching regional MLN lesions, the higher its metastatic potential. Evaluating similarity in expression in terms of the progression signature avoids confounding effects of tissue and cellular composition. This assumption is confirmed by recent findings of the lung cancer TracerX study which showed that expression patterns in

metastases are indeed more similar to those of primary lesions seeding metastasis than to non-seeding primary lesions [79]. In cases with available data, the lesions identified as responsible for regional metastatic lymph node seeding based on tNGS were consistently classified as having a significantly elevated metastatic potential according to our expression-based metric, providing support for the metric's ability to capture true metastatic potential in the context of regional lymph node seeding.

Given that in PCa, similarly to colorectal cancer [11], regional lymph node metastases evolve earlier and more easily than distant metastases, seeding regional lymph node metastases seems selectively less constrained than seeding distant metastases [11, 80] and might, therefore, at least in part be triggered by the plasticity of the tumor cells and the specificities of the TME. To investigate the processes and cell types involved in driving this plasticity, we subsequently compared expression between lesions with the highest and lowest metastatic potential to identify a gene set associated with regional metastatic lymph node seeding, further referred to as the aggressiveness signature.

Overlaying this bulk-derived aggressiveness signature with the single-cell dataset from Chen et al. (2021) showed that it characterizes highly proliferative luminal cells, that were considered "high-risk" by Chen et al. (2021) [8]. The expression of the genes marking these highly proliferative cells is strongly predictive for progression free survival in TCGA-PRAD [42], even more so than a patient's lymph node status. This cell type also explains the prognostic value of several commercially used expression signatures (Decipher [18], Prolaris [16] and OncotypeDx [20]), and a recently described signature defining a luminal proliferative PCa subtype [81] (i.e., *TOP2A* and *MKI67*) that all contain genes, marking these proliferative luminal cells. Interestingly the association of a proliferative expression signal with the seeding state of a lesion was recently also observed in the multifocal TracerX study of lung cancer [79], indicating that the presence of highly proliferative luminal cells is a pan-cancer signature that identifies the aggressive potential of a tumor lesion and seems not only associated with regional lymph node status, but also with progression free survival.

Although the aggressiveness signature contains genes involved in TME remodeling, no luminal epithelial cells were found in the primary samples of Chen et al. (2021) [8] that showed high expression of EMT-related genes. This can be explained by the fact that (i) the number of luminal epithelial cells in the primary lesion undergoing EMT transition is too low to be detected, or that (ii) in the study of Chen et al. (2021) [8] the selected primary lesions did not correspond to the seeding lesions or lastly (iii) the metastasis to the regional lymph node is not driven by an active EMT transition. The absence of an EMT signal in the identified luminal cells is in line with previous observations that showed that most of the EMT signal in a tumor sample originates from the TME [82].

Next to proliferative luminal cells, the aggressiveness signature was also representative for a subset of cells reminiscent of ASCs transitioning towards fibroblasts, originally described in pancreatic cancer [59]. The fact that these fibroblasts like cells express marker genes that distinguish pluripotent fibroblasts from ASCs such as *MGP*, *VCAN*, *COMP* and *EFEMP1* confirms they indeed are derived from ASCs [71]. A subset of those adipose derived fibroblasts displays markers reminiscent of an alternative *COL11A1* attractor state, which has been associated with tumor aggressiveness [59]. We indeed observed that genes marking this so-called aggressive CAF state strongly associate with positive lymph node status in TCGA-PRAD [32] and drive, together with the markers of the proliferative luminal cells the prognostic value of several of the commercial prognostic signatures in PCa. Interestingly, we recently showed that a WSI-based model predictive for the presence of *TP53* mutations known to be associated with aggressive PCa [83, 84] picks up lesions in the primary prostate tumor that are characterized by the same CAF-like expression signature, containing *COL10A1*, *COMP*, *SFRP4* and *INHBA* [85], further supporting the aggressive nature of this adipose derived CAF state. Recent research underscores the critical role of adipocytes in tumorigenesis in various cancers [86, 87] and recognizes ASCs as CAF precursors, produced as a result of the interaction between tumor cells and the adipose micro-environment [87–90]. The fact that this aggressive adipose derived CAF state is associated with differences in regional lymph node seeding potential of primary PCa lesions, and has previously been observed in pancreatic, bladder and lung adeno-carcinoma [72] suggests that they might be a more universal marker of regional lymph node metastasis.

Lastly, we found the aggressiveness signature to be associated with a subset of arterial endothelial cells. Several of the genes in the aggressive signature that mark the presence of these arterial endothelial cells have previously been associated with cancer and or metastasis, but have less prognostic value in TCGA-PRAD [32], presumably because most of them are not sufficiently unique for the identified cell type. According to their functional association, they might reflect the more pronounced level of hypoxia present in the highly proliferating seeding lesions, which results in a higher need for vascularization.

Eventually, we could show that the signal of existing prognostic expression signatures (Decipher [18], Oncotype Dx [20], and Prostadiag [33]) can largely be explained by genes that mark the same proliferative luminal cells and the transitioning towards the adipose

derived CAFs that we associated with higher metastatic potential through the analysis of our multifocal cohort. We could show that for progression free survival the prognostic signal is mostly contained in the proliferative luminal cells, whereas for lymph node status the cells transitioning towards the adipose derived CAFs also contain an important signal. Given that the prognostic performance of a combined marker set, representative for both the proliferative luminal cells and cells transitioning towards the adipose derived CAF was considerably higher for regional lymph node seeding than the signatures representative of either cell type separately, suggests that an interaction between the adipose derived CAFs and the proliferative luminal cells is driving regional lymph node seeding.

Conclusions

With this study, we used the expression profiling of two large multi-focal cohorts of treatment-naive PCa patients to define a metric that allows sorting lesions according to their metastatic potential and to derive an aggressiveness signature that reflects processes associated with regional metastatic lymph node seeding. Leveraging this information with single-cell RNA-seq data identified the cell types associating with this signature, and unveiled a complementary role for proliferative luminal cells and a specific subpopulation of adipose derived CAFs in driving regional metastatic lymph node seeding in PCa. The genes that mark these specific cell states largely explain the prognostic signal of currently used commercial signatures in PCa.

Abbreviations

ASCs	Adipose derived stromal stem cells
TME	Tumor micro-environment
CAF	Cancer-associated fibroblast
CNV	Copy number variation
FFPE	Formalin-fixed paraffin-embedded
GSEA	Gene set enrichment analysis
ISUP	International Society of Urological Pathology
MLN	Metastatic lymph node
mHSPC	Treatment-naive hormone-sensitive metastatic PCa
NL	Normal prostate
PCa	Prostate cancer
PCA	Principal component analysis
PC1	First principal component
PC2	Second principal component
RP	Primary tumor
tNGS	Targeted Next-Generation Sequencing
UMAP	Uniform manifold approximation and projection

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s13073-026-01616-y.

Additional file 1. Table S1-S6: Clinical overview and tNGS-based variant calling results for the locally advanced and *de novo* metastatic prostate cancer cohorts. Table S7-S10: Differentially expressed genes identified by pairwise comparisons of RP, MLN, and NL samples. Table S11: Genes used to construct Additional file 3, Fig. S3. Table S12: Genes comprising the progression signature. Table S13: Number of shared somatic mutations between each primary tumor and its matched metastatic sample per patient. Table S14-S17: Differentially expressed genes and enriched pathways between the most and least aggressive RP lesions. Table S18: Genes comprising the aggressiveness signature. Table S19-S24: Differentially expressed genes and enriched pathways between deviating cells and the remaining luminal cells (Table S19-S20), adipose-derived fibroblasts (Table S21-S22), and endothelial cells (Table S23-S24). Table S25: Summary of the prognostic performance of the evaluated gene signatures.

Additional file 2. Text S1: Within-patient Pearson correlation coefficients of gene expression profiles. Text S2: Identification of tumor evolution using Spearman's rank PCA in the locally advanced and de novo cohorts. Text S3-S4: Generation of the progression signature and its application for assessing metastatic potential. Text S5-S6: tNGS analysis for validation of seeding lesions. Text S7: Differential gene expression analysis between lesions with the highest and lowest metastatic potential. Text S8: Single-cell analysis to assign the aggressiveness signature to specific cell types. Text S9: Identification of a subset of epithelial cells. Text S10: Gene lists used to identify and assess prognostic gene sets.

Additional file 3. Fig. S1: Within-patient Pearson correlation coefficients of gene expression profiles. Fig. S2-S3: Principal component analysis (PCA) based on Spearman's rank correlations, showing the first two principal components. Fig. S4: UMAP visualization of the single-cell dataset. Fig. S5-S13: PCA and dot plots of deviating luminal cells (Fig. S5-S8), fibroblasts (Fig. S9-S11), and endothelial cells (Fig. S12-S13) using the aggressiveness signature in the single-cell dataset. Fig. S14-S17: Gene expression heatmaps and Kaplan-Meier survival analyses of K-means clustered TCGA-PRAD samples based on gene signatures. Fig. S18-S20: Expression heatmaps of genes included in the signature in the single-cell dataset. Fig. S21: Expression of genes included in commercially available signatures.

Acknowledgements

We thank Prof. Olivier De Wever for the useful discussions.

Authors' contributions

All authors read and approved the final manuscript. Study conception: K.M., P.O., L.dSvB., J.VD., K.VdE.Study design: K.M., L.dSvB. Data acquisition: K.VdE., J.V., N.L., B.DL., S.V., J.VD., P.O. Data analysis: L.dSvB., M.L., M.VH., J.H., J.DC., T.J., J.V., K.M. Data interpretation: L.dSvB., M.L., M.VH., J.H., J.DC., T.J., K.M., G.B. Creation of new software used in the work: L.dSvB., M.L. Drafting of the manuscript: L.dSvB., K.M.Substantive revision of the work: G.M., M.VH., T.J., K.M., N.dW.

Funding

The work was supported by grants of the Fonds Wetenschappelijk Onderzoek-Vlaanderen (FWO) (3G045620, 3G046318) and UGent BOF (BOF 01J06219, BOF/IOP/2022/045BOF). LdSvB, ML and JH are supported by a personal FWO grant (3S037019, 3F013118, and 11H9523N). GB's laboratory is supported by the Geconcerteerde Onderzoeksacties Ghent University (BOF.FOA.2021.0005.02), Kom Op Tegen Kanker (STI.DIV.2020.0017.01), and Stichting tegen Kanker (STI.STK.2021.0001.01).

Data availability

All data and results generated or analyzed during this study are included in this article and/or the Additional files, which includes fully searchable Excel Tables. Created code is publicly available at GitHub [91]. Raw HTSeq counts of TCGA-PRAD bulk RNA-seq dataset were downloaded from the Xenahub portal [42]. Ensembl gene IDs were from Xenahub annotation file [92]. The TCGA-PRAD sample type information and the lymph node status information were downloaded from the GDC portal [43]. The single-cell RNA-seq dataset of Chen et al. (2021) [8] is publicly available under NCBI Gene Expression Omnibus accession number GSE141445 [60]. All de-identified sequencing data have been deposited in the European Genome–Phenome Archive database under the accession codes EGAS00001006466 (de novo tNGS) and EGAS00001006715 [93, 94].

Declarations

Ethics approval and consent to participate

The transcriptome data of *de novo* cohort is a single center dataset. It consists of samples from patients that were included in the previously described LoMP1 registry (ClinicalTrials.gov: NCT02138721 [29]) or the randomized phase II LoMP2 study (ClinicalTrials.gov: NCT03655886). This study was approved by the Committee for Ethics of Ghent University Hospital (approval no. BC-01880 and BC-07881).

The transcriptome data of the locally advanced cohort is a single center dataset, consisting of samples obtained from locally advanced metastatic prostate cancer patients. This generation and release of these data was approved by the Committee for Ethics of Ghent University Hospital (EC/2015/0260).

All participants gave written informed consent in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium
[2]Department of Information Technology, Ghent University, IMEC, iGent Toren, Technologiepark Zwijnaarde 126, Ghent 9052, Belgium
[3]Cancer Research Institute Ghent (CRIG), Ghent University, Ghent, Belgium
[4]Department of Human Structure and Repair, Ghent University, Ghent, Belgium
[5]Department of Pathology, Ghent University Hospital, Ghent, Belgium
[6]Department of Biomedical Molecular Biology, VIB Center for Inflammation research (IRC), Ghent University, Ghent, Belgium
[8]Department of Urology, Ghent University Hospital, Ghent, Belgium
[9]Department of Radiation Oncology, Iridiumnetwerk, Antwerp, Belgium
[10]Department of Medical Epidemiology and Biostatistics, Karolinska Institute, Stockholm, Sweden

Received: 11 August 2025 / Accepted: 24 February 2026
Published online: 05 March 2026

References

1. Gandaglia G, Fossati N, Zaffuto E, Bandini M, Dell'Oglio P, Bravi CA, et al. Development and Internal Validation of a Novel Model to Identify the Candidates for Extended Pelvic Lymph Node Dissection in Prostate Cancer. Eur Urol. 2017;72:632–40. https://doi.org/10.1016/j.eururo.2017.03.049.
2. Luchini C, Fleischmann A, Boormans JL, Fassan M, Nottegar A, Lucato P, et al. Extranodal extension of lymph node metastasis influences recurrence in prostate cancer: a systematic review and meta-analysis. Sci Rep. 2017;7:2374. https://doi.org/10.1038/s41598-017-02577-4.
3. Touijer KA, Sjoberg DD, Benfante N, Laudone VP, Ehdaie B, Eastham JA, et al. Limited versus Extended Pelvic Lymph Node Dissection for Prostate Cancer: A Randomized Clinical Trial. Eur Urol Oncol. 2021;4:532–39. https://doi.org/10.1016/j.euo.2021.03.006.
4. Marra G, Valerio M, Heidegger I, Tsaur I, Mathieu R, Ceci F, et al. Management of Patients with Node-positive Prostate Cancer at Radical Prostatectomy and Pelvic Lymph Node Dissection: A Systematic Review. Eur Urol Oncol. 2020;3:565–81. https://doi.org/10.1016/j.euo.2020.08.005.
5. Barry P, Vatsiou A, Spiteri I, Nichol D, Cresswell GD, Acar A, et al. The Spatio-temporal Evolution of Lymph Node Spread in Early Breast Cancer. Clin Cancer Res. 2018;24:4763–70. https://doi.org/10.1158/1078-0432.CCR-17-3374.
6. Ubellacker JM, Tasdogan A, Ramesh V, Shen B, Mitchell EC, Martin-Sandoval MS, et al. Lymph protects metastasizing melanoma cells from ferroptosis. Nature. 2020;585:113–8. https://doi.org/10.1038/s41586-020-2623-z.
7. Angelova M, Mlecnik B, Vasaturo A, Bindea G, Fredriksen T, Lafontaine L, et al. Evolution of Metastases in Space and Time under Immune Selection. Cell. 2018;175:751–65. https://doi.org/10.1016/j.cell.2018.09.018.
8. Chen S, Zhu G, Yang Y, Wang F, Xiao YT, Zhang N, et al. Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression. Nat Cell Biol. 2021;23:87–98. https://doi.org/10.1038/s41556-020-00613-6.
9. Jimenez-Sanchez A, Memon D, Pourpe S, Veeraraghavan H, Li Y, Vargas HA et al. Heterogeneous Tumor-Immune Microenvironments among Differentially Growing Metastases in an Ovarian Cancer Patient. Cell 2017; 170:927 – 38. https://doi.org/10.1016/j.cell.2017.07.025
10. Song H, Weinstein HNW, Allegakoen P, Wadsworth MH 2nd, Xie J, Yang H, et al. Single-cell analysis of human primary prostate cancer reveals the heterogeneity of tumor-associated epithelial cell states. Nat Commun. 2022;13:141. https://doi.org/10.1038/s41467-021-27322-4.
11. Reiter JG, Hung WT, Lee IH, Nagpal S, Giunta P, Degner S, et al. Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases. Nat Genet. 2020;52:692–700. https://doi.org/10.1038/s41588-020-0633-2.
12. Matsumura Y, Ito Y, Mezawa Y, Sulidan K, Daigo Y, Hiraga T, et al. Stromal fibroblasts induce metastatic tumor cell clusters via epithelial-mesenchymal plasticity. Life Sci Alliance. 2019;2:4. https://doi.org/10.26508/lsa.201900425.
13. Alarcon-Zendejas AP, Scavuzzo A, Jimenez-Rios MA, Alvarez-Gomez RM, Montiel-Manriquez R, Castro-Hernandez C, et al. The promising role of new molecular biomarkers in prostate cancer: from coding and non-coding genes to artificial intelligence approaches. Prostate Cancer Prostatic Dis. 2022;25:431–43. https://doi.org/10.1038/s41391-022-00537-2.
14. Lyubetskaya A, Rabe B, Fisher A, Lewin A, Neuhaus I, Brett C, et al. Assessment of spatial transcriptomics for oncology discovery. Cell Rep Methods. 2022;2:11. https://doi.org/10.1016/j.crmeth.2022.100340.
15. Bibikova M, Chudin E, Arsanjani A, Zhou L, Garcia EW, Modder J, et al. Expression signatures that correlated with Gleason score and relapse in prostate cancer. Genomics. 2007;89:666–72. https://doi.org/10.1016/j.ygeno.2007.02.005.
16. Cuzick J, Swanson GP, Fisher G, Brothman AR, Berney DM, Reid JE, et al. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. Lancet Oncol. 2011;12:245–55. https://doi.org/10.1016/S1470-2045(10)70295-3.
17. Dalela D, Loppenberg B, Sood A, Sammon J, Abdollah F. Contemporary Role of the Decipher(R) Test in Prostate Cancer Management: Current Practice and Future Perspectives. Rev Urol. 2016;18:1–9.
18. Erho N, Crisan A, Vergara IA, Mitra AP, Ghadessi M, Buerki C, et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. PLoS ONE. 2013;8:6. https://doi.org/10.1371/journal.pone.0066855.
19. Health Quality O. Prolaris Cell Cycle Progression Test for Localized Prostate Cancer: A Health Technology Assessment. Ont Health Technol Assess Ser. 2017;17:1–75.
20. Klein EA, Cooperberg MR, Magi-Galluzzi C, Simko JP, Falzarano SM, Maddala T, et al. A 17-gene assay to predict prostate cancer aggressiveness in the context of Gleason grade heterogeneity, tumor multifocality, and biopsy undersampling. Eur Urol. 2014;66:550–60. https://doi.org/10.1016/j.eururo.2014.05.004.
21. Knezevic D, Goddard AD, Natraj N, Cherbavaz DB, Clark-Langone KM, Snable J, et al. Analytical validation of the Oncotype DX prostate cancer assay - a clinical RT-PCR assay optimized for prostate needle biopsies. BMC Genomics. 2013;14:690. https://doi.org/10.1186/1471-2164-14-690.
22. Kuhl V, Clegg W, Meek S, Lenz L, Flake DD 2nd, Ronan T, et al. Development and validation of a cell cycle progression signature for decentralized testing of men with prostate cancer. Biomark Med. 2022;16:449–59. https://doi.org/10.2217/bmm-2021-0479.
23. Marrone M, Potosky AL, Penson D, Freedman AN. A 22 Gene-expression Assay, Decipher(R) (GenomeDx Biosciences) to Predict Five-year Risk of Metastatic Prostate Cancer in Men Treated with Radical Prostatectomy. PLoS Curr. 2015;7. https://doi.org/10.1371/currents.eogt.761b81608129ed61b0b48d42c04f92a4.
24. Nakagawa T, Kollmeyer TM, Morlan BW, Anderson SK, Bergstralh EJ, Davis BJ, et al. A tissue biomarker panel predicting systemic progression after PSA recurrence post-definitive prostate cancer therapy. PLoS ONE. 2008;3:5. https://doi.org/10.1371/journal.pone.0002318.
25. Shoag J, Barbieri CE. Clinical variability and molecular heterogeneity in prostate cancer. Asian J Androl. 2016;18:543–8. https://doi.org/10.4103/1008-682X.178852.

26. Andreoiu M, Cheng L. Multifocal prostate cancer: biologic, prognostic, and therapeutic implications. Hum Pathol. 2010;41. https://doi.org/10.1016/j.humpath.2010.02.011. :781 – 93.
27. Boutros PC. The Promise and Perils of Sequencing Individual Prostate Cancer Nuclei. Eur Urol. 2018;74:560–1. https://doi.org/10.1016/j.eururo.2018.07.028.
28. Carm KT, Hoff AM, Bakken AC, Axcrona U, Axcrona K, Lothe RA, et al. Interfocal heterogeneity challenges the clinical usefulness of molecular classification of primary prostate cancer. Sci Rep. 2019;9:13579. https://doi.org/10.1038/s41598-019-49964-7.
29. Warner EW, Van der Eecken K, Murtha AJ, Kwan EM, Herberts C, Sipola J, et al. Multiregion sampling of de novo metastatic prostate cancer reveals complex polyclonality and augments clinical genotyping. Nat Cancer. 2024;5:114–30. https://doi.org/10.1038/s43018-023-00692-y.
30. Kumar A, Coleman I, Morrissey C, Zhang X, True LD, Gulati R, et al. Substantial interindividual and limited intraindividual genomic diversity among tumors from men with metastatic prostate cancer. Nat Med. 2016;22:369–78. https://doi.org/10.1038/nm.4053.
31. Schmidt L, Moller M, Haldrup C, Strand SH, Vang S, Hedegaard J, et al. Exploring the transcriptome of hormone-naive multifocal prostate cancer and matched lymph node metastases. Br J Cancer. 2018;119:1527–37. https://doi.org/10.1038/s41416-018-0321-5.
32. Cancer Genome Atlas Research. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163:1011–25. https://doi.org/10.1016/j.cell.2015.10.025.
33. Kamoun A, Cancel-Tassin G, Fromont G, Elarouci N, Armenoult L, Ayadi M, et al. Comprehensive molecular classification of localized prostate adenocarcinoma reveals a tumour subtype predictive of non-aggressive disease. Ann Oncol. 2018;29:1814–21. https://doi.org/10.1093/annonc/mdy224.
34. van Leenders G, van der Kwast TH, Grignon DJ, Evans AJ, Kristiansen G, Kweldam CF et al. The 2019 International Society of Urological Pathology (ISUP) Consensus Conference on Grading of Prostatic Carcinoma. Am J Surg Pathol. 2020; 44:e87-e99. https://doi.org/10.1097/PAS.0000000000001497
35. Turkbey B, Rosenkrantz AB, Haider MA, Padhani AR, Villeirs G, Macura KJ, et al. Eur Urol. 2019;76:340–51. https://doi.org/10.1016/j.eururo.2019.02.033. Prostate Imaging Reporting and Data System Version 2.1: 2019 Update of Prostate Imaging Reporting and Data System Version 2.
36. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21. https://doi.org/10.1093/bioinformatics/bts635.
37. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020;38:276–8. https://doi.org/10.1038/s41587-020-0439-x.
38. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33:290–5. https://doi.org/10.1038/nbt.3122.
39. Mayrhofer M, De Laere B, Whitington T, Van Oyen P, Ghysel C, Ampe J, et al. Cell-free DNA profiling of metastatic prostate cancer reveals microsatellite instability, structural rearrangements and clonal hematopoiesis. Genome Med. 2018;10:85. https://doi.org/10.1186/s13073-018-0595-5.
40. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997(2013).
41. Annala M, Taavitsainen S, Khalaf DJ, Vandekerkhove G, Beja K, Sipola J, et al. Evolution of Castration-Resistant Prostate Cancer in ctDNA during Sequential Androgen Receptor Pathway Inhibition. Clin Cancer Res. 2021;27:4610–23. https://doi.org/10.1158/1078-0432.CCR-21-1625.
42. National Cancer Institute. TCGA-PRAD bulk RNA-seq dataset. Xenahub Portal. https://gdc-hub.s3.us-east-1.amazonaws.com/download/TCGA-PRAD.htseq_counts.tsv.gz (2023).
43. GDC Data Portal. TCGA Prostate Adenocarcinoma. https://portal.gdc.cancer.gov/projects/TCGA-PRAD. Accessed 17 Nov 2021.
44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8.
45. Hotelling H. Analysis of a complex of statistical variables into principal components. J Educ Psychol. 1933;24:417–41. https://doi.org/10.1037/h0071325.
46. Spearman C. The proof and measurement of association between two things. Int J Epidemiol. 2010;39:1137–50. https://doi.org/10.1093/ije/dyq191.
47. Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis. Preprint at https://doi.org/10.1101/060012 (2021).
48. Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 2015. https://doi.org/10.1016/j.cels.2015.12.004. 1:417 – 25.
49. Hao Y, Hao S, Andersen-Nissen E, Mauck WM 3rd, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184:3573–87. https://doi.org/10.1016/j.cell.2021.04.048.
50. Lloyd S. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982;28:129–37. https://doi.org/10.1109/TIT.1982.1056489.
51. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32:2847–9. https://doi.org/10.1093/bioinformatics/btw313.
52. Fisher RA. On the interpretation of χ 2 from contingency tables, and the calculation of P. J Roy Stat Soc. 1922;85:87–94. https://doi.org/10.2307/2340521.
53. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. J Am Stat Assoc. 1958;53:457–81. https://doi.org/10.2307/2281868.
54. survival Therneau TM. Survival Analysis (R package version 3.5.7). Preprint at https://doi.org/10.32614/CRAN.package.survival(2023).
55. Kassambara A, Kosinski M, Biecek P, survminer. Drawing Survival Curves using 'ggplot2' (R package version 0.5.0). 2024. https://doi.org/10.32614/CRAN.package.survminer
56. Mantel N. Evaluation of survival data and two new rank order statistics arising in its consideration. Cancer Chemother Rep. 1966;50:163–70.
57. Cox DR. Regression models and life-tables. J Roy Stat Soc: Ser B (Methodol). 1972;34:187–220.
58. Bartoschek M, Oskolkov N, Bocci M, Lovrot J, Larsson C, Sommarin M, et al. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. Nat Commun. 2018;9:5150. https://doi.org/10.1038/s41467-018-07582-3.
59. Zhu K, Cai L, Cui C, de Los Toyos JR, Anastassiou D. Single-cell analysis reveals the pan-cancer invasiveness-associated transition of adipose-derived stromal cells into COL11A1-expressing cancer-associated fibroblasts. PLoS Comput Biol. 2021;17:e1009228. https://doi.org/10.1371/journal.pcbi.1009228.
60. Ren S. Single cell analysis reveals onset of multiple progression associated transcriptomic remodellings in prostate cancer. NCBI Gene Expression Omnibus. (2020). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141445
61. Tickle T, Tirosh I, Georgescu C, Brown M, Haas B. Apr. inferCNV of the Trinity CTAT Project. https://github.com/broadinstitute/infercnv. Accessed 22 2024.
62. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16:1289–96. https://doi.org/10.1038/s41592-019-0619-0.
63. McInnes L, Healy J, Melville JUMAP. Uniform Manifold Approximation and Projection for Dimension Reduction. Preprint at https://doi.org/10.48550/arXiv.1802.03426(2018).
64. Franzen O, Gan LM, Bjorkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database (Oxford). 2019; 2019. https://doi.org/10.1093/database/baz046
65. Zhang X, Lan Y, Xu J, Quan F, Zhao E, Deng C, et al. CellMarker: a manually curated resource of cell markers in human and mouse. Nucleic Acids Res. 2019;47:D721–8. https://doi.org/10.1093/nar/gky900.
66. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 2000;25:25–9. https://doi.org/10.1038/75556.
67. Gene Ontology C, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, et al. The Gene Ontology knowledgebase in 2023. Genetics; 2023. p. 224. https://doi.org/10.1093/genetics/iyad031.
68. Xie Z, Bailey A, Kuleshov MV, Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene Set Knowledge Discovery with Enrichr. Curr Protoc. 2021;1:3. https://doi.org/10.1002/cpz1.90.
69. Qian J, Olbrecht S, Boeckx B, Vos H, Laoui D, Etlioglu E, et al. A pan-cancer blueprint of the heterogeneous tumor microenvironment revealed by single-cell profiling. Cell Res. 2020;30:745–62. https://doi.org/10.1038/s41422-020-0355-0.
70. Vijay J, Gauthier MF, Biswell RL, Louiselle DA, Johnston JJ, Cheung WA, et al. Single-cell analysis of human adipose tissue identifies depot and disease specific cell types. Nat Metab. 2020;2:97–109. https://doi.org/10.1038/s42255-019-0152-6.
71. Koczkowska M, Kostecka A, Zawrzykraj M, Myszczynski K, Skoniecka A, Deptula M, et al. Identifying differentiation markers between dermal fibroblasts and adipose-derived mesenchymal stromal cells (AD-MSCs) in human visceral and subcutaneous tissues using single-cell transcriptomics. Stem Cell Res Ther. 2025;16:64. https://doi.org/10.1186/s13287-025-04185-w.
72. Zhang J, Lu S, Lu T, Han D, Zhang K, Gan L, et al. Single-cell analysis reveals the COL11A1(+) fibroblasts are cancer-specific fibroblasts that promote

tumor progression. Front Pharmacol. 2023;14:1121586. https://doi.org/10.3389/fphar.2023.1121586.
73. Rodgers JL, Nicewander WA. Thirteen Ways to Look at the Correlation Coefficient. Am Stat. 1988;42:59–66. https://doi.org/10.1080/00031305.1988.10475524.
74. Kneppers J, Krijgsman O, Melis M, de Jong J, Peeper DS, Bekers E, et al. Frequent clonal relations between metastases and non-index prostate cancer lesions. JCI Insight. 2019;4. https://doi.org/10.1172/jci.insight.124756.
75. You S, Knudsen BS, Erho N, Alshalalfa M, Takhar M, Al-Deen Ashab H, et al. Integrated Classification of Prostate Cancer Reveals a Novel Luminal Subtype with Poor Outcome. Cancer Res. 2016;76:4948–58. https://doi.org/10.1158/0008-5472.CAN-16-0902.
76. Zhou Z, Lu C, Meng S, Dun L, Yin N, An H, et al. Silencing of PTGS2 exerts promoting effects on angiogenesis endothelial progenitor cells in mice with ischemic stroke via repression of the NF-kappaB signaling pathway. J Cell Physiol. 2019;234:23448–60. https://doi.org/10.1002/jcp.28914.
77. Xie T, Dong B, Yan Y, Hu G, Xu Y. Association between MMP-2 expression and prostate cancer: A meta-analysis. Biomed Rep. 2016;4:241–5. https://doi.org/10.3892/br.2015.553.
78. Komaragiri SK, Bostanthirige DH, Morton DJ, Patel D, Joshi J, Upadhyay S, et al. ID4 promotes AR expression and blocks tumorigenicity of PC3 prostate cancer cells. Biochem Biophys Res Commun. 2016;478:60–6. https://doi.org/10.1016/j.bbrc.2016.07.092.
79. Martinez-Ruiz C, Black JRM, Puttick C, Hill MS, Demeulemeester J, Larose Cadieux E, et al. Genomic-transcriptomic evolution in lung cancer and metastasis. Nature. 2023;616:543–52. https://doi.org/10.1038/s41586-023-05706-4.
80. Naxerova K, Jain RK. Using tumour phylogenetics to identify the roots of metastasis in humans. Nat Rev Clin Oncol. 2015;12:258–72. https://doi.org/10.1038/nrclinonc.2014.238.
81. Weiner AB, Liu Y, Hakansson A, Zhao X, Proudfoot JA, Ho J, et al. A novel prostate cancer subtyping classifier based on luminal and basal phenotypes. Cancer. 2023;129:2169–78. https://doi.org/10.1002/cncr.34790.
82. Szabo PM, Vajdi A, Kumar N, Tolstorukov MY, Chen BJ, Edwards R, et al. Cancer-associated fibroblasts are the main contributors to epithelial-to-mesenchymal signatures in the tumor microenvironment. Sci Rep. 2023;13:3051. https://doi.org/10.1038/s41598-023-28480-9.
83. de Schaetzen L, Miclotte G, Larmuseau M, Van den Eynden J, Marchal K. Network-Based Analysis to Identify Drivers of Metastatic Prostate Cancer Using GoNetic. Cancers (Basel). 2021;13:21. https://doi.org/10.3390/cancers13215291.
84. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015;161:1215–28. https://doi.org/10.1016/j.cell.2015.05.001.
85. Pizurica M, Larmuseau M, Van der Eecken K, de Schaetzen L, Carrillo-Perez F, Isphording S, et al. Whole Slide Imaging-Based Prediction of TP53 Mutations Identifies an Aggressive Disease Phenotype in Prostate Cancer. Cancer Res. 2023;83:2970–84. https://doi.org/10.1158/0008-5472.CAN-22-3113.
86. Bochet L, Lehuede C, Dauvillier S, Wang YY, Dirat B, Laurent V, et al. Adipocyte-derived fibroblasts promote tumor progression and contribute to the desmoplastic reaction in breast cancer. Cancer Res. 2013;73:5657–68. https://doi.org/10.1158/0008-5472.CAN-13-0530.
87. Downer MA, Griffin MF, Morgan AG, Parker JB, Li DJ, Berry CE, et al. Understanding the Role of Adipocytes and Fibroblasts in Cancer. Ann Plast Surg. 2023;91:779–83. https://doi.org/10.1097/SAP.0000000000003658.
88. Mao X, Xu J, Wang W, Liang C, Hua J, Liu J, et al. Crosstalk between cancer-associated fibroblasts and immune cells in the tumor microenvironment: new findings and future perspectives. Mol Cancer. 2021;20:131. https://doi.org/10.1186/s12943-021-01428-1.
89. El Alaa RSA, Al-Mannai W, Darwish N, Al-Mansoori L. Adipose-Derived Stromal Cells and Cancer-Associated Fibroblasts: Interactions and Implications in Tumor Progression. Int J Mol Sci. 2024;25:21. https://doi.org/10.3390/ijms252111558.
90. Kang J, Lee JH, Cha H, An J, Kwon J, Lee S, et al. Systematic dissection of tumor-normal single-cell ecosystems across a thousand tumors of 30 cancer types. Nat Commun. 2024;15:4067. https://doi.org/10.1038/s41467-024-48310-4.
91. de Schaetzen L, Jung T, Larmuseau M. Multifocal Prostate Cancer Cohort Analysis Using RNA-Seq Data. GitHub. 2026. https://github.com/MarchalLab/multi-focal-analysis
92. Luo Y, xena-GDC-ETL. GitHub. 2019. https://github.com/MarchalLab/multi-focal-analysis
93. Marchal K. Sequencing data for the manuscript Multifocal cohort analysis unveils cell types associated with regional lymph node seeding in prostate cancer. European Genome-Pheno (2026). https://ega-archive.org/studies/EGAS00001006715
94. Wyatt A. Sequencing data for the manuscript Multi-focal sampling of de novo metastatic prostate cancer reveals complex polyclonality and enables accurate clinical genotyping. European Genome-Phenome Archive. (2022). https://ega-archive.org/studies/EGAS00001006466

Publisher's Note