



Single-cell analysis reveals transcriptomic remodellings in distinct cell types that contribute to human prostate cancer progression

Sujun Chen^{1,2,3,4,23}, Guanghui Zhu^{1,2,5,6,23}, Yue Yang^{1,2,3}, Fubo Wang^{1,2,3}, Yu-Tian Xiao¹, Na Zhang⁷, Xiaojie Bian⁸, Yasheng Zhu¹, Yongwei Yu¹, Fei Liu⁹, Keqin Dong¹, Javier Mariscal¹⁰, Yin Liu¹¹, Fraser Soares¹², Helen Loo Yau^{2,3}, Bo Zhang¹², Weidong Chen¹², Chao Wang¹², Dai Chen¹², Qinghua Guo¹², Zhengfang Yi⁷, Mingyao Liu⁷, Michael Fraser², Daniel D. De Carvalho^{12,3}, Paul C. Boutros^{13,14,15,16,17}, Dolores Di Vizio^{10,18,19,20}, Zhou Jiang²¹, Theodorus van der Kwast¹², Alejandro Berlin^{12,22}, Song Wu^{5,6}, Jianhua Wang¹²✉, Housheng Hansen He^{12,3}✉ and Shancheng Ren¹✉

Prostate cancer shows remarkable clinical heterogeneity, which manifests in spatial and clonal genomic diversity. By contrast, the transcriptomic heterogeneity of prostate tumours is poorly understood. Here we have profiled the transcriptomes of 36,424 single cells from 13 prostate tumours and identified the epithelial cells underlying disease aggressiveness. The tumour microenvironment (TME) showed activation of multiple progression-associated transcriptomic programs. Notably, we observed promiscuous *KLK3* expression and validated the ability of cancer cells in altering T-cell transcriptomes. Profiling of a primary tumour and two matched lymph nodes provided evidence that *KLK3* ectopic expression is associated with micrometastases. Close cell-cell communication exists among cells. We identified an endothelial subset harbouring active communication (activated endothelial cells, aECs) with tumour cells. Together with sequencing of an additional 11 samples, we showed that aECs are enriched in castration-resistant prostate cancer and promote cancer cell invasion. Finally, we created a user-friendly web interface for users to explore the sequenced data.

Prostate cancer remains the most prevalent male malignancy worldwide. In 2018, nearly 1.2 million new cases were projected to be diagnosed¹, accounting for 13.5% of all newly diagnosed cancers in men. Prostate cancer has a highly heterogeneous clinical course. Patients with indolent disease can live for years without progression, while aggressive disease can quickly metastasize and become incurable. Although the incidence of prostate cancer, as a whole, has decreased in recent years, a steady increase in advanced or metastatic prostate cancer has been observed concurrently², advocating for improvement in treatment strategies. Thus, there is an urgent need to further our understanding of prostate cancer heterogeneity.

Cancer subtyping provides valuable insights into cancer biology and informs treatment planning. Genomic and transcriptomic studies have revealed subtypes of prostate cancer characterized by mutations and aberrant transcription^{3–6}. Recently, the PAM50 classifier was used to subtype prostate cancer into prognostic groups with predictive power for treatment sensitivity⁶. However, previous studies were based on bulk sequencing, which probably omits rare populations that may drive disease development and progression. Moreover, bulk data represent the collective reflection of signals from both the tumour and the microenvironment (TME), the latter of which can be a strong driver of tumour aggression.

¹Department of Urology, Shanghai Changhai Hospital, Shanghai, P. R. China. ²Princess Margaret Cancer Centre, University Health Network, Toronto, Ontario, Canada. ³Department of Medical Biophysics, University of Toronto, Toronto, Ontario, Canada. ⁴Ontario Institute for Cancer Research, Toronto, Ontario, Canada. ⁵Urology Institute of Shenzhen University, The Third Affiliated Hospital of Shenzhen University, Shenzhen University, Shenzhen, P. R. China. ⁶Department of Urological Surgery, The Affiliated Luohu Hospital of Shenzhen University, Shenzhen University, Shenzhen, P. R. China. ⁷East China Normal University, Shanghai, P. R. China. ⁸Cancer Institute, Shanghai Urological Cancer Institute, Fudan University Shanghai Cancer Center, Fudan University, Shanghai, P. R. China. ⁹Department of Urology, Cancer Hospital, Chinese Academy of Medical Sciences, Shanghai, P. R. China. ¹⁰Department of Surgery, Division of Cancer Biology and Therapeutics, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ¹¹Department of Laboratory Medicine, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai, P. R. China. ¹²Novel Bioinformatics Co., Ltd, Shanghai, P. R. China. ¹³Department of Human Genetics, University of California Los Angeles, Los Angeles, CA, USA. ¹⁴Department of Urology, University of California Los Angeles, Los Angeles, CA, USA. ¹⁵Jonsson Comprehensive Cancer Centre, University of California Los Angeles, Los Angeles, USA. ¹⁶Institute for Precision Health, University of California Los Angeles, Los Angeles, CA, USA. ¹⁷Department of Pharmacology & Toxicology, University of Toronto, Toronto, Ontario, Canada. ¹⁸Samuel Oschin Comprehensive Cancer Institute, Division of Cancer Biology and Therapeutics, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ¹⁹Department of Medicine, David Geffen School of Medicine, University of California Los Angeles, Los Angeles, CA, USA. ²⁰Department of Biomedical Sciences, Division of Cancer Biology and Therapeutics, Cedars-Sinai Medical Center, Los Angeles, CA, USA. ²¹Department of Pathology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai, P. R. China. ²²Department of Radiation Oncology, University of Toronto, Toronto, Ontario, Canada. ²³These authors contributed equally: Sujun Chen, Guanghui Zhu, Yue Yang, Fubo Wang. ✉e-mail: jianhuaw2007@qq.com; Hanshenhe@uhnresearch.ca; renshancheng@gmail.com

Single-cell RNA-sequencing (scRNA-seq) technology has created unprecedented opportunities to simultaneously assess thousands of cells within a sample, enabling heterogeneity among tumour cells and the complexity of the TME to be revealed^{7–9}. Furthermore, scRNA-seq provides unique opportunities to assess the regulation, evolution and interaction of individual cells^{9–13}. Unravelling cellular communication and evolution is key to understanding the roles of different tumour and TME components¹⁴, as well as establishing their collaboration network.

Although scRNA-seq is increasingly being adopted, its application to prostate cancer has been limited to circulating tumour cells and immortalized cell lines. This limited representation of human tumours and lack of TME profiling leaves the intra-tumoral transcriptomic heterogeneity of the most common cancer in men largely unknown. In this Article, we have performed scRNA-seq on 13 prostate tumours and obtained transcriptomic profiles for 36,424 cells. We have collected and analysed an additional 14 samples consisting of normal prostate tissue, NCCN (National Comprehensive Cancer Network) low/high-risk and castration-resistant prostate cancers (CRPCs). In total 111,914 single cells from 27 tissue samples were analysed. We have provided a comprehensive single-cell gene expression atlas for the human prostate and uncovered a highly heterogeneous prostate cancer-specific transcriptome landscape harbouring multiple malignant programs.

Results

Single-cell transcriptome landscape in prostate cancer. To understand prostate cancer heterogeneity at single-cell resolution, we collected 13 (12 primary and 1 lymph node metastasis) tissue samples from 12 patients (Supplementary Table 1, data batch 1) and performed single-cell RNA sequencing (scRNA-seq). After standard data processing and quality control procedures (Methods), we obtained transcriptomic profiles for 36,424 cells (Extended Data Fig. 1a).

We identified 16 subclusters (Extended Data Fig. 1b) from the scRNA-seq profiles using a graph-based clustering method (Methods). The QuSAGE¹⁵ method was used to quantitatively assign the broad lineage of epithelial and stromal cells¹⁶ (Extended Data Fig. 1c) for each cluster. Inference of copy number aberrations (CNAs) was performed as previously described⁹ (Methods). The inferred CNAs identified canonical prostate cancer genome alterations^{4,17}, including gains of chromosome 8q and losses of chromosome 8p, 13 and 16q (Extended Data Fig. 1d). We observed an inflection point that separates putative malignant and non-malignant cells in a few samples (Extended Data Fig. 1e). However, the distinction was less precise in other samples, resulting in many cells being classified as unresolved. There are probably a number of reasons for this. First, some localized prostate tumours are well described to have a quiet genome. Indeed, in The Cancer Genome Atlas (TCGA) study, a subset of prostate cancers harboured zero CNAs⁴. Our data also show a significantly smaller proportion of the genome altered with strong CNA relative to melanoma (Extended Data Fig. 1f), making the signal weaker and more susceptible to dilution and subthreshold CNA scores. Second, prostate cancer is highly heterogeneous, and previous DNA sequencing studies have revealed a wide range of CNA profiles (0–50% of the genome altered through CNA). As a matter of fact, CNA burden is prognostic for prostate cancer¹⁸. Third, most prostate tumours harbour multiple clones with highly divergent CNA burdens, and the subclonal CNA load is far more than clonal in most localized tumours¹⁹. Despite these caveats, we identified a linear CNA cutoff from samples having clear separation, and applied it to our full cohort. This allowed us to broadly distinguish malignant from non-malignant clusters. Generally, the TME cell clusters exclusively contain cells classified as non-malignant (Extended Data Fig. 1g), while luminal clusters (Fig. 1a) contain a mixture of cells, mostly classified as malignant along with a subset of cells with unresolved classification, probably reflecting subclonal

diversity. The basal/intermediate cell clusters mainly contain cells classified as non-malignant and unresolved.

We curated 20 cell type signatures to generate a detailed cell identity annotation (Fig. 1a,b). Interestingly, besides canonical keratin genes, we observed expression of T-cell co-stimulatory genes in luminal cells (Fig. 1b and Extended Data Fig. 2a), hinting at a potential role of epithelial cells involved in antigen presentation. The TME components (non-epithelial) show a generally lower expression complexity compared to epithelial cells (Fig. 1c), consistent with a previous study¹⁰. Although the inferred CNA deviation for luminal cells is significantly higher (Mann–Whitney U test, $P < 2.2 \times 10^{-16}$, common language effect size (CLES) = 0.97) than the rest (Fig. 1d), values for individual cells span a wide range (0–0.47). Nevertheless, it is important to note that non-malignant epithelial cells may also contribute to the CNA profiles of luminal cells.

The cell type composition of individual tumours differs substantially (Fig. 1e). As expected, luminal is the dominant epithelial type for most tumours (Fig. 1e and Extended Data Fig. 2b). However, basal/intermediate is the major epithelial type for sample 154 (basal/intermediate:luminal = 7; Fig. 1e). Previous work using the PAM50 signature classified prostate cancer into luminal A, luminal B and basal subtypes and was able to inform disease outcome⁶. However, the PAM50 signature was created for breast cancer and might not be fully representative of prostate cancer. In addition, a high signal for the PAM50 basal signature is observed for non-epithelial components in our data (Extended Data Fig. 2c), suggesting contamination of the stromal signal in the signature¹⁰. Similarly, a high Gleason score (GS)-related gene signature derived from a previous microdissection analysis²⁰ was highest in stroma, with 3 of 19 genes preferentially expressed in TME cells (Extended Data Fig. 2d,e). Considering the perhaps dominant contribution of the stroma component to certain bulk sequencing-derived signatures²¹, it is important to assess different cell types independently. Therefore, we next sought to dissect the epithelial subpopulations underlying tumour aggressiveness and to create a purified signature for prostate cancer.

Intrinsic epithelial cell subpopulations underlying tumour subtypes. Epithelial cells, including both basal/intermediate and luminal types, were re-analysed, yielding 16 clusters (Fig. 2a). Most clusters consisted of cells from multiple patients, indicating an unbiased cell subpopulation representation (Extended Data Fig. 3a). To identify the cell subpopulations associated with differential clinical outcomes, we compared the expression of multiple prostate-cancer-subtype-related signatures across clusters. The PAM50 luminal A and B signatures show a high signal for most of the clusters simultaneously, except clusters 10 and 12 (Extended Data Fig. 3b). Cluster 10 has a high signal for luminal A, while cluster 12 is enriched for the luminal B signature, suggesting the two signatures are specifically determined by these two clusters of cells, respectively. Cluster 12 also has the highest scores of hypoxia signalling²² and PCS1, a metastasis associated luminal subtype derived from the prostate cancer transcriptome⁵. Considering that all four signatures have important clinical associations, clusters 10 and 12 are potentially critical cell subpopulations associated with prostate cancer progression. We therefore next derived signatures representative of these subpopulations (Methods and Extended Data Fig. 3c). Cluster 10 alone has high expression of basal/intermediate specific marker genes (*KRT5*, *KRT14*, *KRT19* and *TP63*; Fig. 2b), and was named ‘Basal/Intermediate cells’. Interestingly, marker genes of cluster 12, including the known cell-cycle-related genes *CDC20*, *CCNB1*, *CENPF* and *PTTG1*, are enriched with terms related to cell division (Extended Data Fig. 3d), indicating differences in cell-cycle regulation. We thus used the cyclone function²³ (Methods) to assign cells into their cell-cycle phases. Cluster 12 showed significant enrichment (Fisher’s exact test, $P = 6.7 \times 10^{-41}$, odds ratio (OR) = 5.7) for cells in G2/M (Extended Data Fig. 3d),

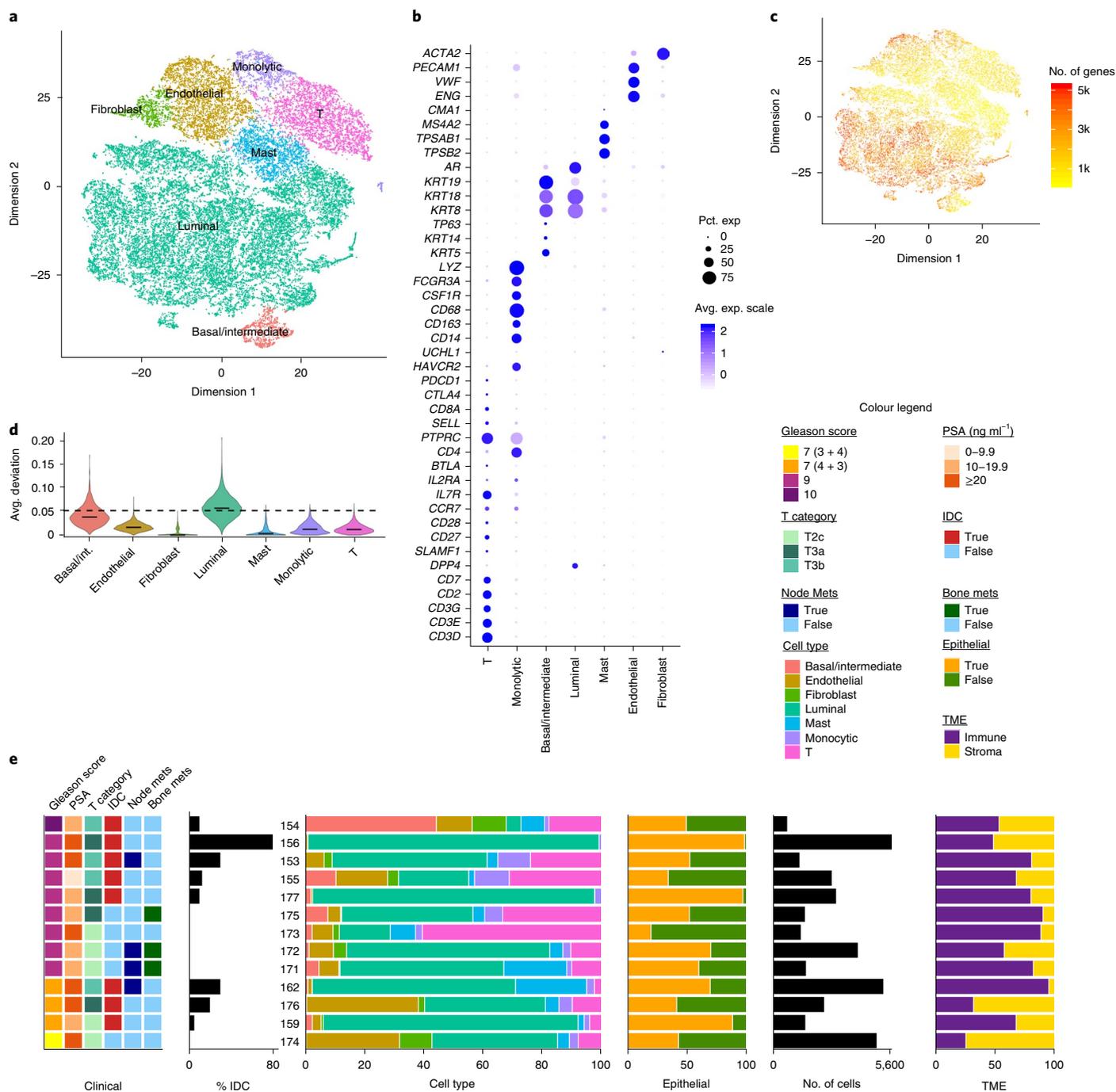


Fig. 1 | Overview of the single-cell landscape for prostate cancer. a, A *t*-distributed stochastic neighbour embedding (tSNE) view of 36,424 single cells, colour-coded by assigned cell type. **b**, Marker gene expression for each cell type, where dot size and colour represent percentage of marker gene expression (pct. exp) and the averaged scaled expression (avg. exp. scale) value, respectively. **c**, tSNE view of all cells, colour-coded by number of genes detected in each cell. **d**, CNA for each cell type, calculated using the inferred copy number. The colour scheme is the same as in **a**. Cells with score >0.2 are shown as 0.2. The *P* value (Mann–Whitney U test, two-sided) for the luminal (*n* = 23,674 cells) group compared to all the rest (*n* = 12,750 cells) is <2.2 × 10⁻¹⁶, common language effect size (CLES) = 0.97. **e**, Cell composition distribution for each patient sample. Samples are ordered according to Gleason score (GS). TME, tumour microenvironment; IDC, intraductal carcinoma/ciribriform architecture. See also Extended Data Figs. 1 and 2 and Supplementary Table 1.

representing a population driven by cell-cycle-related features, and was thus termed ‘CellCycle’. Compared with the other luminal cells, CellCycle and Basal/Intermediate cells show distinctive cell functions (Fig. 2c). The CellCycle population is enriched for functions such as oxidative phosphorylation and DNA replication, while depleted for general luminal cell functions. The Basal/Intermediate

cells are particularly associated with antigen processing and presentation (Fig. 2c).

Consistent with a previous study²⁴, CellCycle is predictive of disease outcome across multiple cohorts (Fig. 2d and Extended Data Fig. 3e). This is concordant with the observation that CellCycle cells show high expression of luminal B, hypoxia and PCS1 signatures^{5,6,22}

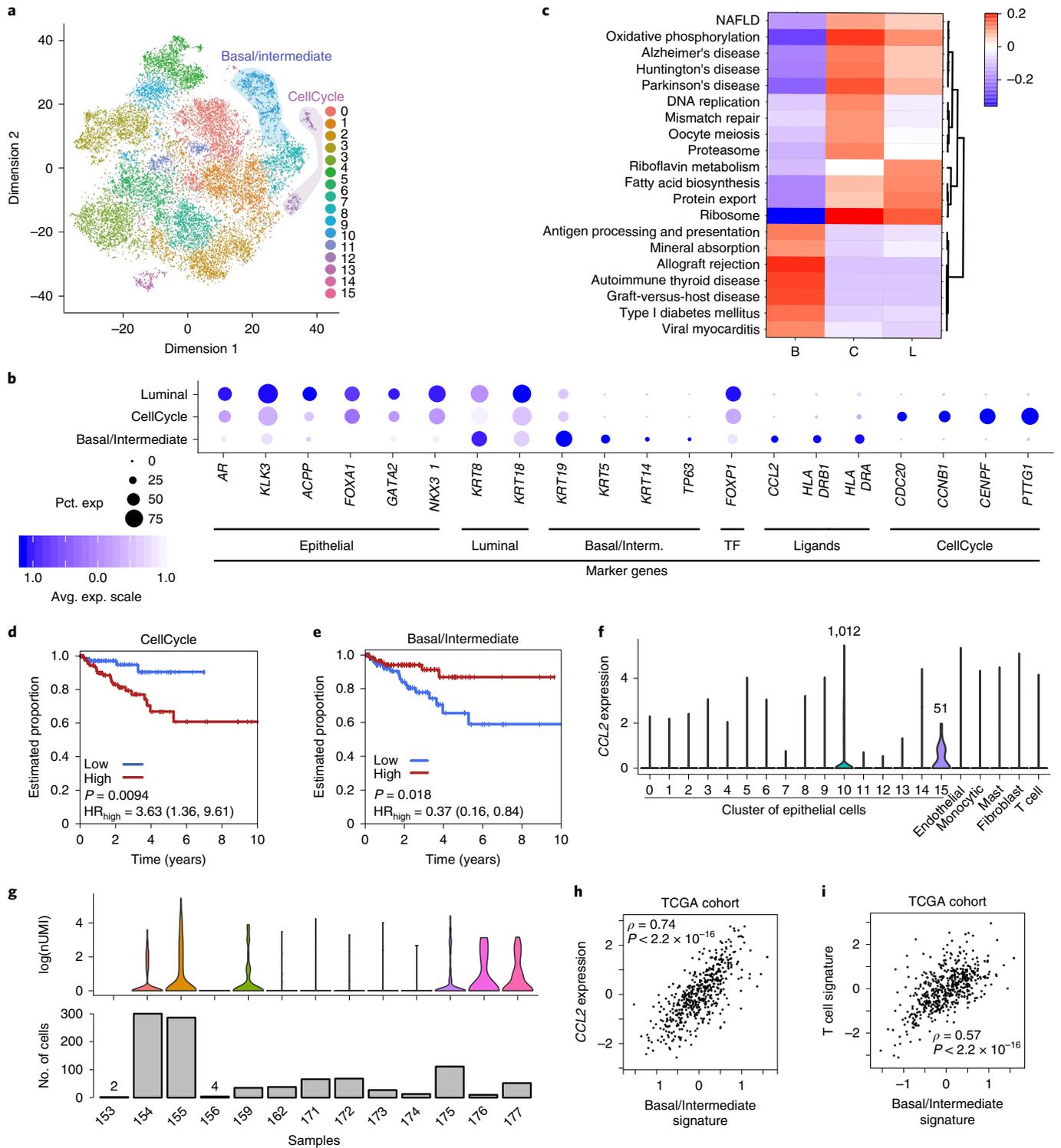


Fig. 2 | Purified signature derived from single-cell data and the association with survival. **a**, tSNE view of 23,674 epithelial cells, colour-coded by re-evaluated clusters, with colour shading highlighting clusters of CellCycle and Basal/Intermediate subtypes. **b**, Marker gene expression for the three epithelial subtypes. **c**, Differentially activated pathways among the epithelial subtypes, with 'B', 'C' and 'L' representing Basal/Intermediate, CellCycle and Luminal, respectively. **d,e**, Comparison of biochemical recurrence (BCR)-free rates for the high and low groups, stratified using the CellCycle (**d**) and Basal/Intermediate (**e**) signatures in TCGA. P values are calculated using the Cox proportional hazard model (CoxPH) and are not adjusted for multiple comparisons. Numbers in parentheses show 95% confidence intervals (CIs) for hazard ratios (HRs). **f**, Smoothed distribution of *CCL2* expression, with TME and epithelial cells grouped by their assigned cell types and clusters, respectively. Cell numbers are shown for epithelial clusters 10 and 15 above the violin plots. A one-sided Fisher's exact test is used to examine enrichment of *CCL2* expression in cluster 10. $P < 2.2 \times 10^{-16}$; OR = 21.64. **g**, Smoothed distribution of *CCL2* expression in different samples (top) and number of basal/intermediate cells as a bar plot (bottom). **h**, Correlation between *CCL2* expression and basal/intermediate signature in TCGA. **i**, Correlation between T cell and basal/intermediate signature. Two-sided P values are calculated for Spearman's rank correlation and not adjusted for multiple comparisons. See also Extended Data Figs. 3 and 4.

(Extended Data Fig. 3b), as all these three signatures are associated with unfavourable outcomes. Convergence of all three high-risk signatures in the CellCycle population suggests common subpopulations underlying different high-risk subtypes identified from bulk data. Nevertheless, it is yet to be determined whether cell-cycling status is the main driving force of this population.

In line with luminal A being indicative of longer recurrence-free survival⁶, a high basal/intermediate signature is associated with better survival (Fig. 2e and Extended Data Fig. 4a). As basal/intermediate cells show high expression of antigen processing and presentation genes (Fig. 2c), we investigated immune regulation-related genes. Both HLA class II and the chemokine gene *CCL2* are specifically expressed in basal/intermediate cells (Fig. 2b). Notably, basal/intermediate cells are the main cell population expressing *CCL2* (Fig. 2f,g), and in the two samples with the least basal/intermediate cell content, *CCL2* expression is not detected. Expression of *CCL2* is also significantly associated with content of basal/intermediate cells in bulk sequencing data (Fig. 2h). Such a high correlation persists after removing the effect of infiltrated stromal cells on *CCL2*²⁵ (Extended Data Fig. 4b). Considering *CCL2* is a well-known chemokine that attracts immune cells including macrophage and T cells, the relationship between content of basal/intermediate cells and infiltration of macrophage and T cells was investigated. As expected, strong correlations were observed across multiple cohorts (Fig. 2i and Extended Data Fig. 4c). Together, these results suggest that basal/intermediate cells attract immune cells, which might partially contribute to the good prognosis of patients with high basal/intermediate score.

Tumour-associated macrophages show osteoclast-like features.

Immune cells in the TME have complicated roles in tumour progression. We thus used our single-cell data to investigate the heterogeneity of infiltrating immune cells. Two distinct myeloid populations—monocytic and mast cells (Fig. 1a)—were identified. To further characterize the monocytic clusters (Fig. 3a), we compared their averaged abundance with bulk profiles on sorted cells²⁶ and inspected the expression of canonical marker genes. Cluster 1 showed the highest correlation with myeloid dendritic cells, while all the others were most highly associated with monocytes (Fig. 3b). All clusters showed high levels of *HLA-DRA* and *CD68*, and all the monocytes, but not dendritic cells, showed high levels of *CD14* and *FCGR3A* (Fig. 3c). Additionally, clusters 0, 2, 4, 5 and 6 are high in C1q protein family genes and *MSR1* (CD204), showing characteristics of macrophages. Together, we assigned the cells into three subtypes: monocyte cells, dendritic cells (DCs) and tumour-associated macrophages (TAMs) (Fig. 3a).

Two major types of macrophage activation, the classic (M1) and alternative (M2) types, have been described, with M2 primarily thought to have a pro-cancer role. We observed a mixture of M1 and M2 signal activation among TAMs (Fig. 3c), with a strong positive correlation between the two signals (Fig. 3d), in line with recent studies in human cancers showing that TAMs co-express M1 and M2 signals^{7,27}. Nevertheless, notable heterogeneity exists among TAM clusters, and important pathways involved in macrophage function and activation, including tumour necrosis factor (TNF), nuclear factor (NF)- κ B, nucleotide oligomerization domain- and Toll-like receptor signalling, are regulated differently (Fig. 3e).

Cluster 6 (C6) showed higher activation of osteoclast (OC)-related pathways such as mineral absorption and lysosome. This is intriguing, as bone is one of the most common metastatic sites for prostate cancer and OCs contribute to the process²⁸. Monocyte/macrophage lineage differentiation into OCs is associated with active ligand-receptor interactions, so we performed cell communication analysis using CellPhoneDB¹⁴ (Methods). We focused on interactions unique to C6 (Extended Data Fig. 5a) and find they are mostly related to OC regulation. Among the unique pairs, *CDH1*-encoded E-cadherin has

been reported to be an important modulator of osteoclastogenesis²⁹. Activins (*INHBA/INHBB*) and their receptors (*ACVR1B/ACVR2a*) are involved in modulating OC regulation, including enhancing RANKL-mediated differentiation, which is a major route inducing macrophage into OCs^{30,31}. OC inhibition has been shown to delay prostate cancer bone metastasis²⁸. Indeed, C6 TAM is detectable across all samples (Extended Data Fig. 5b), indicating early onset of TME remodelling that contributes to tumour spread.

Infiltrated CD8⁺ effector T cells express tumour marker genes.

T cells can be subtyped into CD4⁺ T conventional (Tconv), CD4⁺ T regulatory (Treg), CD8⁺ T naive and CD8⁺ T effector cells according to marker gene expression (Fig. 3f,g). Decreased lipid and amino-acid metabolism with increased glycolysis level was observed in Treg cells, consistent with previous studies of *FOXP3*-positive thymus-derived Treg cells³² (Extended Data Fig. 5c,d). Interestingly, the three clusters (2, 3 and 5) consisting of CD8⁺ T effector cells show substantial heterogeneity (Extended Data Fig. 5c). Cluster 5 is the least activated subtype, with the lowest levels of most metabolism and immune pathways, while cluster 3 shows the characteristics of highly activated T cells (Fig. 3h and Extended Data Fig. 5c). Strikingly, we see a high level of *KLK3*, the gene encoding prostate specific antigen (PSA), in clusters 3 and 5 as well as Treg cells. To understand the underlying mechanism for this unexpected *KLK3* expression, we performed gene co-expression (Methods) analysis to identify highly correlated gene modules and focused on ones that show subtype- or cluster-specific activation.

Analysis of subtype-specific modules showed significant enrichment of terms associated with their respective functions (Fig. 3h and Supplementary Table 2). Interestingly, more than half of the subtype-specific modules are unique to Tregs, indicating a distinct transcriptomic program for Tregs compared to other T-cell subtypes. The top enriched terms for Treg-specific modules include lineage regulatory processes such as lymphoid progenitor differentiation, regulation of regulatory T-cell differentiation and metabolism of various carbohydrate and amino-acid substrates³³. Terms related to pathways previously reported to be involved in Foxp3 and Treg regulation, such as Wnt-protein³⁴ and RAGE receptor³⁵ binding, are also among the top enrichments. Cell adhesion mediated by integrin is also enriched for Tregs. As integrin expression on T cells can affect immune infiltration³⁶, such enrichment thus points to a possible mechanism for Treg homing into the tumour region.

Module 61 is the only one specific to *KLK3*-high T-cell clusters, with gradient activation corresponding to *KLK3* abundance (Fig. 3h). Notably, the statistically enriched terms in this module are related to extracellular vesicle (EV) and exosome (Fig. 3h and Supplementary Table 2). Moreover, other androgen receptor (*AR*) signature genes are mostly not expressed (Extended Data Fig. 5e), indicating *KLK3* abundance might be of exogenous origin, rather than from activation of *AR* signalling in T cells. We thus hypothesized that the accumulation of tumour-specific gene expression in CD8⁺ T cells is mediated by tumour-derived EVs. We detected *KLK3* expression in all T-cell subgroups, hinting that the effect of EVs might accompany different stages of tumour-infiltrating T cells (Fig. 3h).

We next asked if the accumulation of epithelial tumour markers in T cells is restricted to prostate cancer. We analysed public scRNA-seq data for four additional cancer types (Methods): head and neck squamous cell carcinoma (HNSCC), non-small cell lung cancer (NSCLC), colon rectal cancer (CRC) and hepatocellular carcinoma (HCC). Expression of tumour marker genes in T cells was identified in all four cancer types (Fig. 4a), suggesting that this may be a general feature of the transcriptome of infiltrating T cells.

EVs originating in prostate cancer cells induce *KLK3* expression in T cells. We next sought to validate the presence and source of tumour marker genes in infiltrated T cells from patient samples.

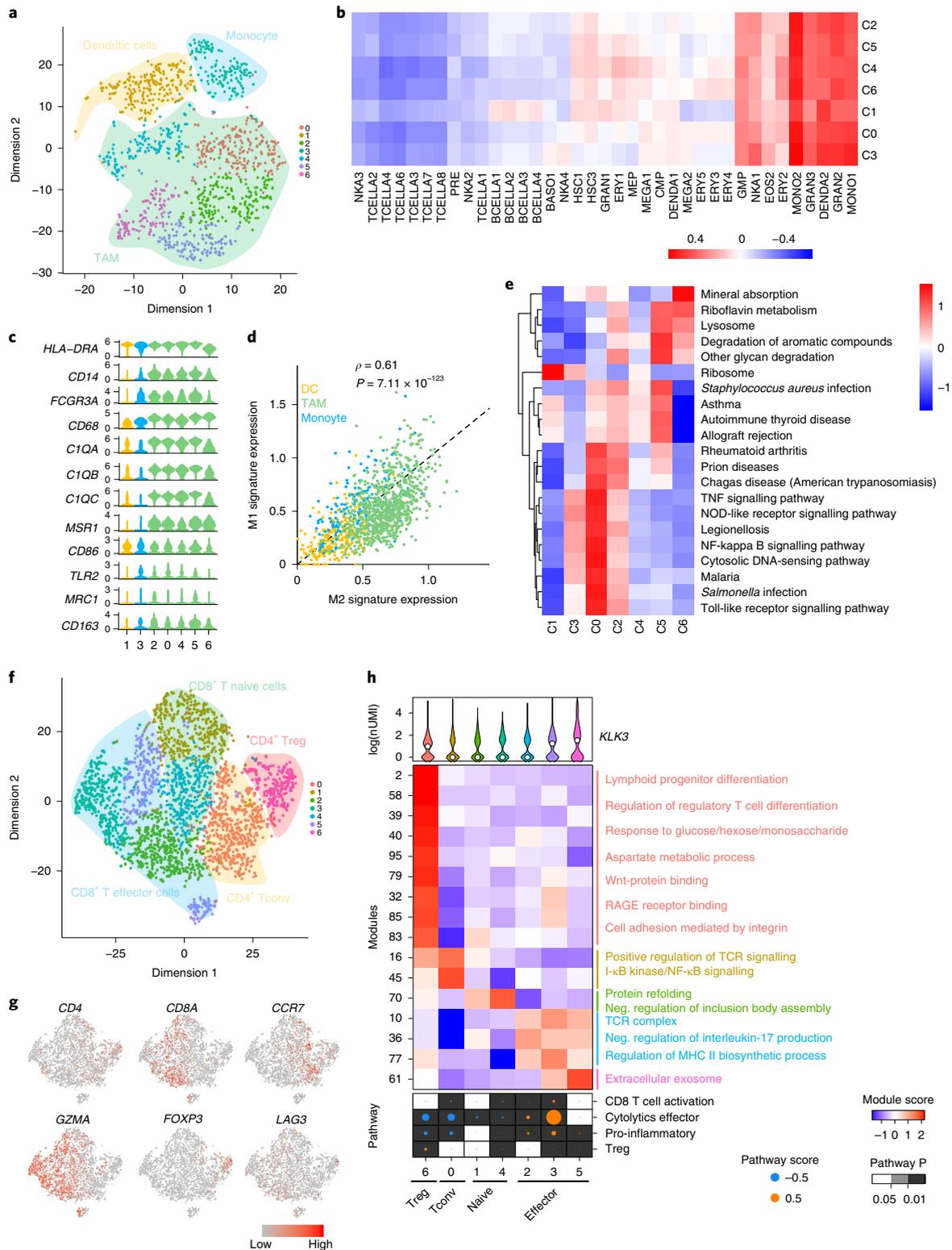


Fig. 3 | scRNA-seq reveals heterogeneity in immune components. **a**, tSNE view of monocytes, colour-coded by subtypes. **b**, Correlation of monocyte cell clusters with bulk RNA-seq on different sorted immune cells²⁶. **c**, Smoothed distribution of monocyte-related marker gene abundance. **d**, Correlation between the M1 and M2 signatures, where each dot represents a cell. Two-sided P values are calculated for Spearman's rank correlation and not adjusted for multiple comparisons. **e**, Differentially activated pathways among all the TAM clusters. Only pathways with activation score higher than 0.1 are shown. **f**, tSNE view of 3,116 T cells, colour-coded by re-evaluated clusters (**f**) and expression level of marker genes (**g**). **h**, Characterization of different T-cell populations, with clusters ordered and grouped by subtypes. Top: smoothed distribution of *KLK3* abundance. Middle: subtype-specific modules, with representative enrichment of module genes, grouped by subtype, shown at the right. Bottom: relative activation of immune-related pathways in each cluster. See also Extended Data Fig. 5 and Supplementary Table 2.

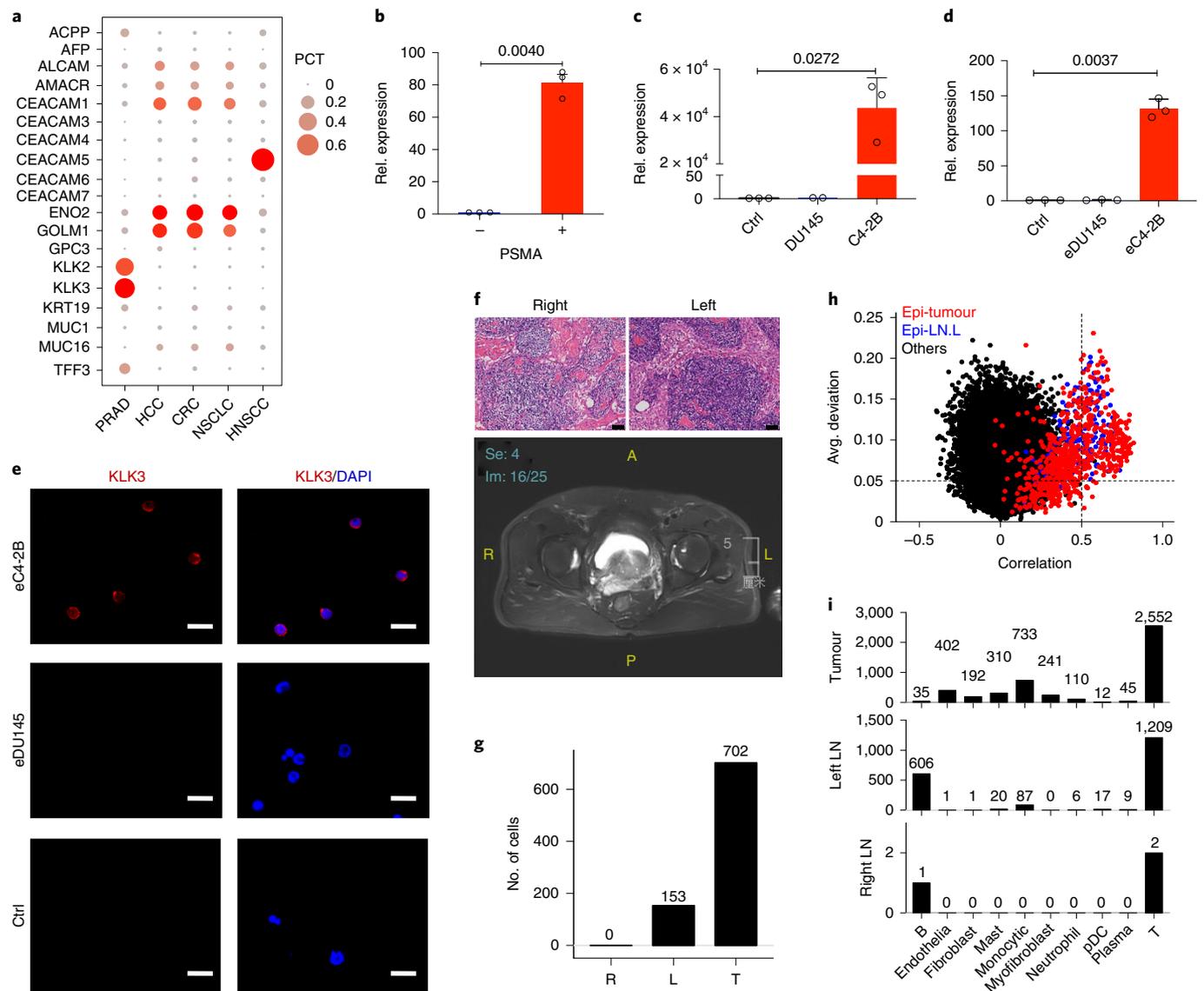


Fig. 4 | Tumour-derived EVs convey ectopic *KLK3* expression in T cells. **a**, Expression of epithelial tumour markers in T cells across different cancer types. PRAD, prostate adenocarcinoma. **b**, rtPCR analysis of *KLK3* level in PSMA positive (+) and negative (-) T cells isolated from fresh prostate tumour. $P=0.004$ (Student's *t*-test, $n=3$ independent experiments, two-sided). **c**, rtPCR analysis of *KLK3* in T cells after co-incubation with prostate cancer cell lines DU145 and C4-2B compared with control. $P=0.0272$ (Student's *t*-test, $n=3$ independent experiments, two-sided) for C4-2B compared to the Ctrl group. **d**, rtPCR analysis of *KLK3* in T cells after co-incubation with secreted EVs from prostate cancer cells DU145 (eDU145) and C4-2B (eC4-2B) compared with control. $P=0.0037$ (Student's *t*-test, $n=3$ independent experiments, two-sided) for eC4-2B compared to the Ctrl group. Data are presented as mean values \pm s.d. **e**, RNA fluorescence in situ hybridization (FISH) analysis showing localization of *KLK3* to T cells. Scale bars, 20 μ m. T cells were isolated using anti-CD8 microbeads. Purified T cells were cultured alone (Ctrl) or co-cultured with EVs derived from C4-2B (eC4-2B) and DU145 (eDU145). Data show representative results from two repeats. **f**, Haematoxylin and eosin (H&E) staining (top; scale bars, 50 μ m) and magnetic resonance imaging (MRI; bottom) results, showing no metastasis to either the left or right pelvic lymph nodes. The analysis was not repeated. IM, image; Se, sequence. **g**, Number of epithelial cells in right (R) and left (L) lymph node as well as primary tumour (T). **h**, Average CNA deviation (avg. deviation) plotted against correlation with the averaged CNA profiles of epithelial cells in primary tumour sample for each cell. Epi-tumour, epithelial cells in tumour; Epi-LN.L, epithelial cells in left lymph node. **i**, Number of *KLK3*-positive cells in the three samples. See also Extended Data Fig. 6. Statistical data relevant to **b-d** are provided in the source data. pDC, plasmacytoid dendritic cell.

The *KLK3*-positive T-cell-specific module 61 contains other prostate-cancer-specific genes, including *FOLH1*, which encodes the prostate specific membrane antigen (PSMA; Extended Data Fig. 5f). We used flow cytometry to isolate PSMA-positive T cells from tumour tissues (Extended Data Fig. 6a,b). Real-time quantitative polymerase chain reaction with reverse transcription (rtPCR) analysis showed a very high level of *KLK3* in the PSMA-positive but

not -negative T cells (Fig. 4b), validating the presence of tumour marker genes in T cells.

To test the capacity of tumour cells to alter T-cell gene expression profiles, we cultured T cells sorted from healthy donors (Methods) together with two prostate cancer cell lines (Extended Data Fig. 6c,d) and examined the *KLK3* abundance in T cells after 24h. Co-culturing with the AR-positive C4-2B, but not the AR-negative

DU145 cell line, led to a significant increase of *KLK3* expression in T cells (Fig. 4c). Moreover, administration of EV-containing pellet isolated from C4-2B, but not DU145 cell culture supernatant, also led to detection of *KLK3* in T cells, as determined by both rtPCR and RNA fluorescence in situ hybridization (RNA-FISH; Fig. 4d,e). Additionally, immunofluorescent staining revealed localization of PSA in CD8⁺ T cells co-cultured with EVs from C4-2B, but not DU145 cells (Extended Data Fig. 6e). A discontinuous ultracentrifugation iodixanol gradient of the EV-containing pellet shows that both the *KLK3* RNA and protein are predominantly present in the vesicles but not in the EV free fractions, which are known to contain soluble material (Methods and Extended Data Fig. 6f,g). Taken together, these data show that EVs secreted by prostate cancer cells can alter the transcriptome of infiltrating T cells.

Ectopic expression of *KLK3* is associated with micrometastases. Cancer-derived EVs are increasingly recognized for their important roles in the tumour–stroma interaction as well as disease development and progression³⁷. Because we observed that tumour-derived EVs can directly alter T-cell mRNA and protein abundance in localized tumours (Fig. 4d,e), we next asked whether and to what extent EVs affect nearby lymph nodes (LNs). Primary prostate tumour and two external iliac obturator LN samples were collected from an NCCN high-risk patient (Supplementary Table 1, data batch 2) treated with radical prostatectomy and pelvic lymph node dissection. Both magnetic resonance imaging (MRI) and histopathological analysis showed no metastasis in the LNs (Fig. 4f).

Cells from tumour show distinct separation from LN samples (Extended Data Fig. 6h,i). Of interest, we identified a small number of epithelial cells in the left but not right LN (Fig. 4g), despite the fact that no observable metastasis was detected by imaging or pathological examination (Fig. 4f). CNA inference analysis showed that a subset of the epithelial cells (84/153) harbour CNAs concordant with the averaged tumour CNA profile, strongly suggesting lymphatic micrometastases in the left LN (Fig. 4h). We observed widespread *KLK3* expression across all the cell types, with the largest number (1,209) being T cells, in the left but not the right lymphatic samples (Fig. 4i). In line with the expression being EV-mediated, the abundance of *KLK3* in T cells and the rest of the TME are much lower than in epithelial cells (Extended Data Fig. 6j–l). It is worth mentioning that in the right LN sample where no micrometastases were observed, we detected two T cells and one B cell with *KLK3* expression. This suggests alteration of immune cell gene expression in LNs might occur before the actual metastasis takes place, potentially thereby establishing a pre-metastatic niche.

scRNA-seq reveals heterogeneity for cancer-associated fibroblasts in the TME. We next investigated the non-immune component of the TME, including fibroblasts and endothelial cells (ECs). Increasing evidence suggests an important role for cancer-associated fibroblasts (CAFs) in prostate cancer progression, and recent scRNA-seq-based studies have expanded our understanding of their heterogeneity in other cancer types³⁸. We identified 948 fibroblast cells (Fig. 5a). The five distinct clusters were classified into three subtypes according to similarities among each cluster and key marker genes expression (Fig. 5a,b and Extended Data Fig. 7a).

Most markers show unique expression patterns across the three subtypes, with only the general mesenchymal gene *VIM* showing a high and universal signal (Fig. 5b). Although *ACTA2* and its protein product α -actin are common CAF markers in other cancer types, immunohistochemistry staining in prostate cancer showed depleted expression³⁹. However, here *ACTA2* showed abundance comparable to that of *VIM* in the CAF subsets, and the signal co-expressed with other markers. Depleted *ACTA2* expression in the previous study might have resulted from changes in the stroma composition, such as an increase in *ACTA2*-negative ECs or expansion of the *ACTA2*-low CAF population. In addition, a positive trend, although not significant (Spearman's $\rho=0.38$, $P=0.202$), was observed between the percentage of *ACTA2*-positive CAF and the epithelial to mesenchymal (EMT) transitioning score in epithelial cells (Extended Data Fig. 7b), suggesting EMT being a possible source for *ACTA2*-positive CAFs. Together, these differences highlight the complexity of the TME and the need for single-cell-level analysis in larger patient cohorts.

We performed gene enrichment analysis on the top 150 upregulated genes in each CAF subtype. Not surprisingly, the angiogenesis-associated genes are enriched in all subtypes (Fig. 5c), while myofibroblastic, cell adhesion and extracellular matrix (ECM) related ones (Fig. 5c and Supplementary Table 3) are more subtype-specific. Transcription factor (TF) analysis (Methods) showed the most enriched TFs for S2 are *CREB3L1* and *PLAGL1* (Extended Data Fig. 7c), genes that control ECM production and composition^{40,41}, suggesting ECM activation as an important aspect of CAF differentiation. *HOXB2* and *MAFB* have the highest activation in S3, while *ETS1* is slightly higher in S1. Interestingly, *MAFB* inhibits transcription of myeloid lineage genes mediated by *ETS1*⁴². This is similar to the transdifferentiation of myeloid cells into fibroblast-like cells in pathological conditions⁴³. Although more work is needed to illustrate the actual source of myofibroblastic CAFs in prostate cancer, our data provide evidence for a shared regulatory network between CAF and non-fibroblastic lineages in the TME.

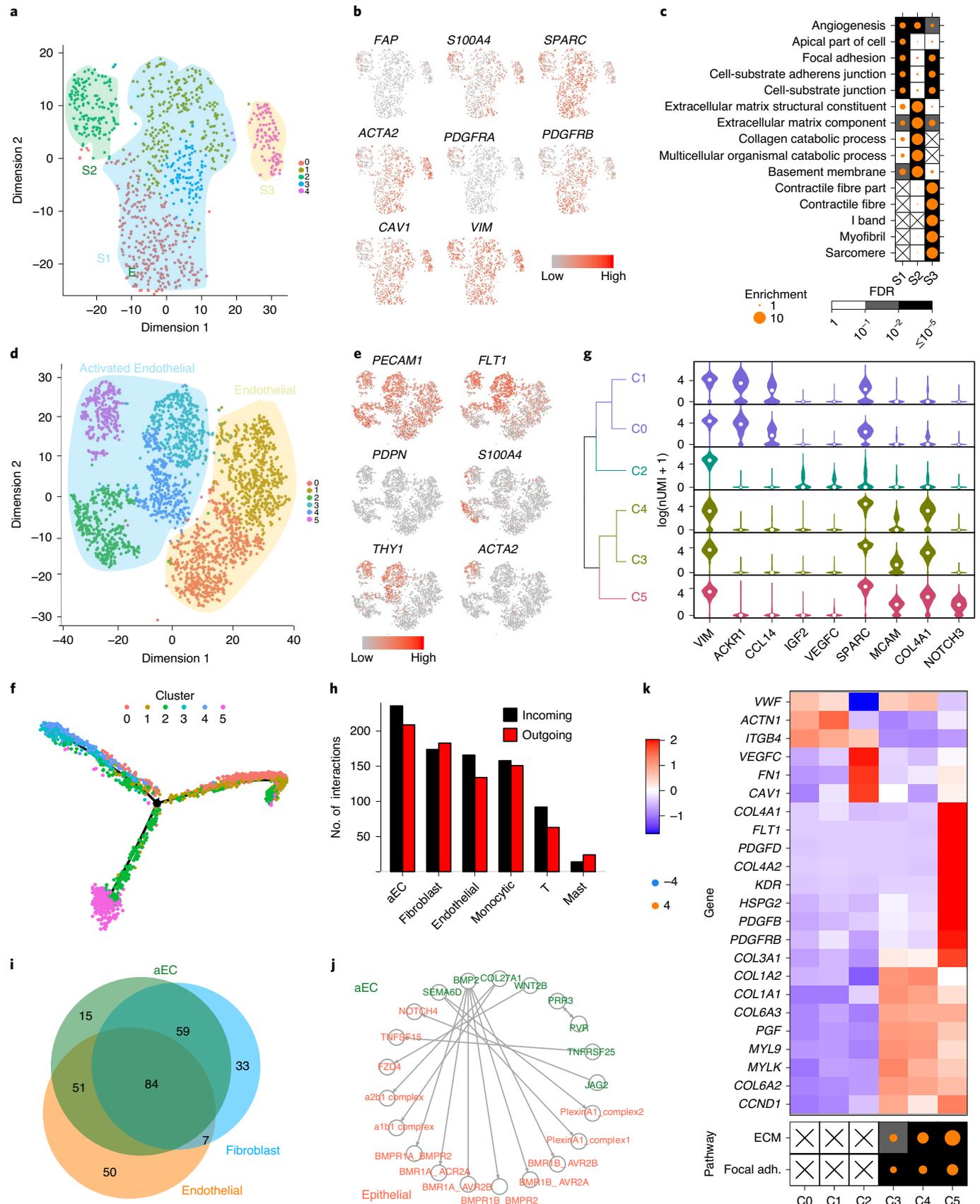
Activated ECs modify tumour extracellular matrix. We detected 3,115 ECs, which are further classified into six subclusters (Fig. 5d). We observed universal expression of the endothelial marker gene *PECAM1*. The presence of *FLT1* and absence of *PDPN* expression suggest that the ECs are derived from vascular rather than lymphatic vessels (Fig. 5e). Interestingly, we observed that genes related to activated CAFs are highly expressed in EC subsets (Fig. 5e). Specifically, cluster 2 expressed *S100A4*, a metastasis-associated gene previously reported to promote tumour angiogenesis⁴⁴. Clusters 3 and 4 are positive for *THY1*, the expression of which on microvascular ECs promotes metastasis in melanoma⁴⁵. Cluster 5 has high abundance of *THY1*, *ACTA2* and *S100A4*, showing a highly CAF-like feature. Together, we named these four CAF-marker-expressing EC clusters aECs.

Pseudotime analysis of the aECs using Monocle¹³ (Methods) suggested two diverging cell fates, starting at clusters 0 and 1, progressing towards clusters 3 and 4 at one end and cluster 5 at the other, with cluster 2 being a transitioning state spreading along the axis (Fig. 5f). Indeed, differential gene expression analysis attributed the six clusters to the four subtypes concordant with the pseudotime states (Fig. 5g). A more detailed inspection of the differen-

Fig. 5 | Identifying an activated EC cell subset in stromal cells. **a**, tSNE view of CAF cells, colour-coded by subtype. **b**, tSNE view of CAF cells, colour-coded by the indicated marker gene abundance. **c**, Top five most enriched gene ontology (GO) terms for each CAF subset. **d,e**, tSNE view of 3,115 ECs, colour-coded by re-evaluated clusters (**d**) and expression level of marker genes (**e**). **f**, All ECs ordered along pseudotime trajectories, with the cells colour-coded by cluster. **g**, Smoothed distribution of representative membrane-protein-coding genes that are differentially expressed among EC subtypes. The left dendrogram was calculated using all differential genes and is coloured according to group membership. **h**, Number of unique cell communication pairs in each cell type. **i**, Overlap among the unique outgoing cell communication pairs from fibroblasts, canonical ECs and activated ECs (aECs) to epithelial cells. **j**, aEC specific outgoing cell communication pairs. **k**, Genes in ECM-receptor interaction and focal adhesion pathways that show differential expression among canonical ECs and aECs. See also Extended Data Fig. 7 and Supplementary Table 3.

tially expressed genes revealed significant dysregulation of ligands and receptors between ECs and aECs. Thus, we next investigated cell–cell interactions with CellphoneDB¹⁴ (Methods). Notably, aECs

have higher numbers of inferred interactions with epithelial cells (Fig. 5h). Additionally, aECs possess a large portion of the interactions from both classic ECs and fibroblasts (Fig. 5i and Extended



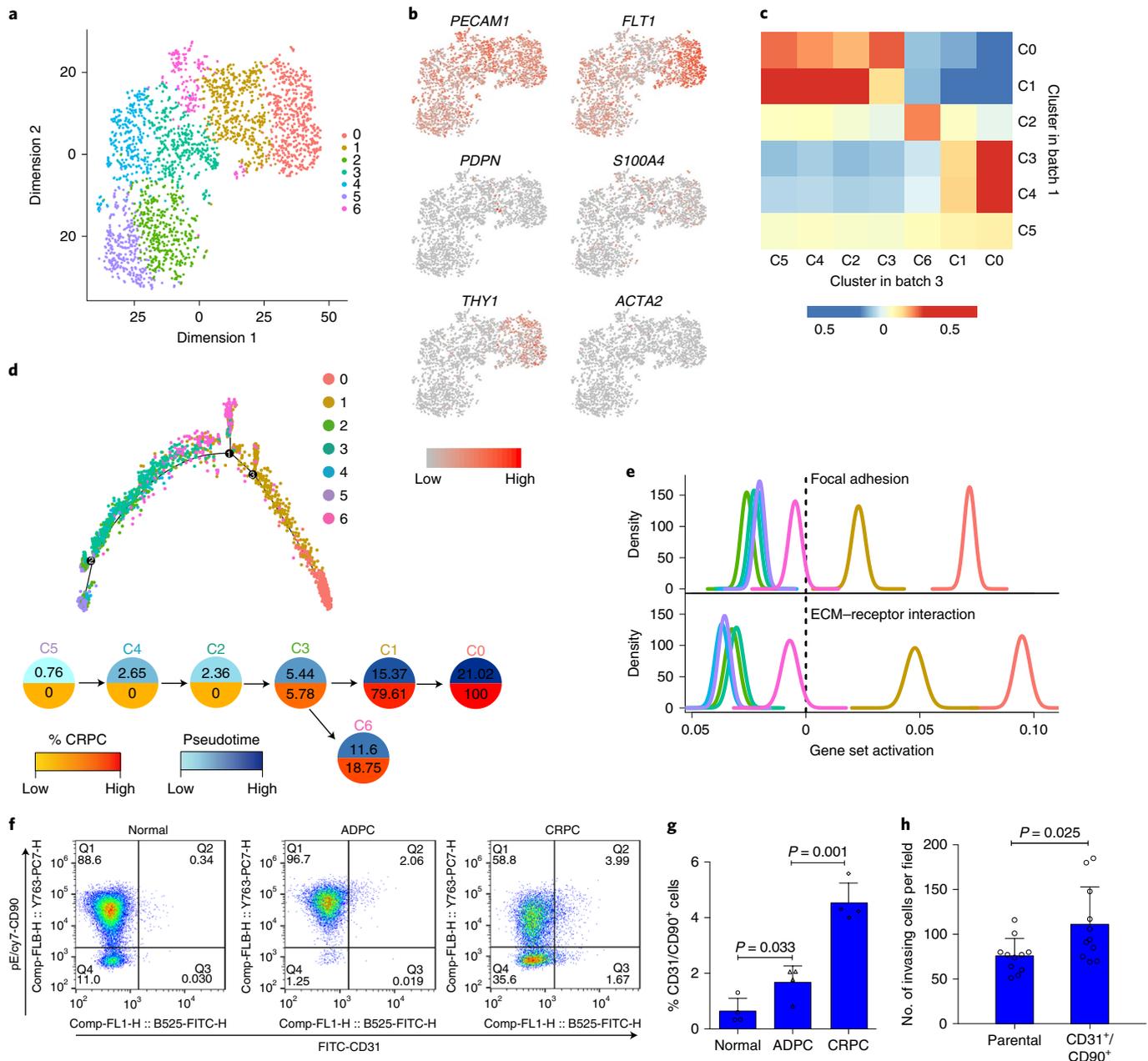


Fig. 6 | aEC cells undergo ECM remodelling and are enriched in CRPC. a, b, tSNE view of ECs from normal-CRPC samples, colour-coded by clusters (**a**) and marker gene expression (**b**). **c,** QuSAGE scores, indicating similarities among EC clusters in batch 1 and 3 data. **d,** Top: all ECs from normal-CRPC samples ordered along pseudotime trajectories, colour-coded by clusters. Bottom: schematics of the trajectories, with pie chart colour-coded by average pseudotime and percentage of cells from CRPC samples in the indicated cluster. Two-sided *P* values are calculated for Spearman's rank correlation and not adjusted for multiple comparisons. **e,** Differential activation of focal adhesion and ECM-receptor interaction pathways among all seven EC clusters. **f,** Dot plots of flow cytometry showing percentage of CD31⁺ and CD90⁺ aECs in normal, primary prostate adenocarcinoma (ADPC) and castration-resistant prostate cancer (CRPC) samples. **g,** Quantification of aEC percentage in **f**. *P* values (Student's *t*-test, *n* = 4 independent samples, two-sided) for normal versus ADPC and ADPC versus CRPC are 0.033 and 0.001, respectively. **h,** Number of invading fields after PC3 co-culture with parental or CD31⁺/CD90⁺ aEC cells. Data are taken from three replicates. *P* = 0.025 (Student's *t*-test, *n* = 11 fields examined, two-sided). Statistical data relevant to **g** and **h** are provided in the source data.

Data Fig. 7d). Among the aEC unique pairs are interactions between *BMP2* and the type I/II BMP receptors (*BMPRIA*, *BMPRII*), which induce migration and invasion in human ECs (Fig. 5j)⁴⁶. By contrast, EC-specific interactions are enriched for chemokines including *CCL3*, *CCL5*, *CCL18* and *CCL23*. Immune-activation-related genes such as *CCL2* and *BIRC3* are at low abundance in classic ECs,

consistent with the previous finding that tumour-associated ECs downregulate immune attraction pathways⁸.

Pathway activation comparison among EC clusters further supports our analysis (Extended Data Fig. 7e). The top enriched pathways in cluster 0 and 1 are mostly immune-related, although the most common pathways in aECs are ECM-receptor signalling and

focal adhesion (Fig. 5k). Together, these data suggest that aECs modify the ECM while further downregulating immune activation.

aECs are enriched in CRPC and promote cancer cell invasion.

Given that aECs possess metastasis-associated signalling and that cancer cells can induce remodelling of the ECM to facilitate disease spreading⁴⁷, we hypothesize that aECs are associated with more aggressive forms of prostate cancer. We performed scRNA-seq of six prostate tissues obtained during radical cystectomy for bladder cancer and five tumours from CRPC patients (Fig. 6a,b and Supplementary Table 1, batch 3). For patients with bladder cancer, pathology analysis did not identify prostatic malignancy for patient SC009B, but identified bladder cancer invasion for SC007B and low-risk prostate cancer (GS 6) for SC008B. These samples capture different elements of the prostate life history and serve as controls. For better representation of the spatial heterogeneity of the prostate, samples from the left and right peripheral zones of each prostate were collected and separately sequenced. We detected *THY1*-positive subsets resembling the aEC population in this normal-CRPC cohort (Fig. 6c). Strikingly, the two aEC clusters (0 and 1) consisted mostly of cells from CRPC samples (Fig. 6d). Pseudotime analysis revealed a progression from classic EC to aEC, which is highly correlated with an increase of CRPC cell percentage in the clusters ($\rho=0.95$, $P=0.001$). Activation of the ECM–receptor interaction and focal adhesion pathways was also observed in the aEC subsets (Fig. 6e).

We next sought to validate the association of aECs with CRPC in fresh human tumour samples. We collected primary prostate adenocarcinoma (ADPC) and CRPC samples, as well as normal prostate tissues from bladder cancer patients ($n=4$ for each condition). aECs were isolated by flow cytometry using CD31 (*PECAM1*, endothelial cell marker) and CD90 (*THY1*, CAF cell marker) antibodies. Consistent with our analysis, a significantly higher fraction of aECs were isolated from CRPC compared to ADPC and normal prostate samples (Fig. 6f,g; Student's test, $P=0.001$ and 0.033 , respectively). To understand the effect of aECs on cancer cells, a co-culture experiment was performed using the CD31⁺/CD90⁺ aEC population followed by examination of cell invasion. As expected, a significantly (Student's test, $P=0.025$) increased invasion was observed in the aEC co-culture condition (Fig. 6h).

Discussion

Individual prostate tumours are highly spatially and clonally heterogeneous, and the disease varies dramatically among patients¹⁹. Bulk transcriptome sequencing acquires averaged signals across cell mixtures and therefore limits evaluation of these heterogeneities. We used scRNA-seq to produce a single-cell transcriptomic atlas including primary prostate cancer, metastases, as well as non-cancerous ‘normal’ prostate tissues. These data allowed us to identify unexpected biological features in distinct cell types.

The observation that the basal/intermediate cell type is the dominant epithelial class in sample 154 appears to conflict with previous findings showing the absence of basal cells and expansion of luminal cells in prostate cancer⁴⁸. Yet the basal/intermediate cells identified here are transcriptomically distinct from non-transformed prostate basal cells with less abundant *KRT5*, *KRT14* and *TP63*, probably leading to negative detection in histopathology analysis. Such dissociation of molecular and phenotypic features of basal cells is also supported by the detection of basal-like signatures in bulk sequencing of prostate tumours^{5,6}. Moreover, the fact that p63 protein expression is rare but associated with high-grade tumours⁴⁹ aligns well with our observation.

We showed that transcriptomic signatures derived independently converge on the same cell population (CellCycle). This population only contains 364 cells, accounting for ~1% of all cells studied. Nevertheless, 12 of the 13 samples contribute to this

subset. Such universal existence of aggressiveness ‘seeding’ or an active cycling population synergizes with observations in the TME, where multiple metastasis-associated programs, including those for osteoclasts-like TAMs and aECs, are activated across patients. This led to the question as to what factors contribute to such malignant remodelling.

The TME, especially the immune component, certainly plays a role. Understanding the mechanisms underlying the immune-suppressive environment of prostate cancer is of great interest. We observed tumour-derived perturbations on immune cells. A high level of *KLK3* was found in the non-epithelial subsets. There is a nonlinear relationship whereby, as the abundance of *KLK3* in T cells increases, an increase of effector cell cytolytic activity is observed before a decrease occurs. This might result from the complex interactions between T-cell-activating and -suppressing signals. In addition to *KLK3*, we observed PSA in EVs and EV-co-cultured T cells. As the levels of *KLK3* in T cells are much lower than in the epithelial cells, it is challenging to detect PSA in situ in the background of high PSA in epithelial cells. It is also unclear whether and how *KLK3* expression affects T-cell function. Fully illustrating the impact of *KLK3* in TME cells is a complex and delicate task. Factors such as the amount, size and type of EVs generated, as well as the duration of TME cell exposure, need to be taken into consideration.

Ectopic *KLK3* expression can provide additional clinical inference. Pre-clinical studies have shown promises of quantifying *KLK3* in pelvic LN for detecting micrometastases⁵⁰. Here our work has hinted at a possible mechanism underlying this molecular analysis of LN metastasis, where a favourable niche in the TME is prepared before actual metastasis happens, allowing tumour-derived EVs and/or tumour-educated immune cells to prelude tumour cells in migrating into nearby LNs. Such a hypothesis is further supported by the fact that ectopic *KLK3* expression in immune cells was detected for the right LN we tested where no micrometastases were observed.

Here we have used benign prostate tissues from bladder cancer patients as controls. Although challenging to obtain, normal prostate tissues from healthy donors will be important to further validate these observations and to exclude potential influences of bladder tumours on the prostate microenvironment. Taken together, we have revealed a transcriptome landscape in primary tumours where multiple metastasis-associated transcriptional programs become activated. Our data call for larger single-cell sequencing of prostate cancer linked to robust clinical outcome data to advance the identification of therapeutic targets and the development of biomarkers of aggressive disease.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41556-020-00613-6>.

Received: 15 September 2020; Accepted: 25 November 2020; Published online: 8 January 2021

References

1. Ferlay, J. et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int. J. Cancer* **144**, 1941–1953 (2019).
2. Weiner, A. B., Matulewicz, R. S., Eggen, S. E. & Schaeffer, E. M. Increasing incidence of metastatic prostate cancer in the United States (2004–2013). *Prostate Cancer Prostatic Dis.* **19**, 395–397 (2016).
3. Chen, S. et al. Widespread and functional RNA circularization in localized prostate cancer. *Cell* **176**, 831–843 (2019).
4. Cancer Genome Atlas Research Network The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).

5. You, S. et al. Integrated classification of prostate cancer reveals a novel luminal subtype with poor outcome. *Cancer Res.* **76**, 4948–4958 (2016).
6. Zhao, S. G. et al. Associations of luminal and basal subtyping of prostate cancer with prognosis and response to androgen deprivation therapy. *JAMA Oncol.* **3**, 1663–1672 (2017).
7. Azizi, E. et al. Single-cell map of diverse immune phenotypes in the breast tumor microenvironment. *Cell* **174**, 1293–1308 (2018).
8. Lambrechts, D. et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat. Med.* **24**, 1277–1289 (2018).
9. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
10. Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck. *Cancer Cell* **171**, 1611–1624 (2017).
11. Kumar, M. P. et al. Analysis of single-cell RNA-seq identifies cell–cell communication associated with tumor characteristics. *Cell Rep.* **25**, 1458–1468 (2018).
12. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
13. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
14. Vento-Tormo, R. et al. Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature* **563**, 347–353 (2018).
15. Yaari, G., Bolen, C. R., Thakar, J. & Kleinstein, S. H. Quantitative set analysis for gene expression: a method to quantify gene set differential expression including gene–gene correlations. *Nucleic Acids Res.* **41**, e170 (2013).
16. Henry, G. H. et al. A cellular anatomy of the normal adult human prostate and prostatic urethra. *Cell Rep.* **25**, 3530–3542 (2018).
17. Taylor, B. S. et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
18. Fraser, M. et al. Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359–364 (2017).
19. Espiritu, S. M. G. et al. The evolutionary landscape of localized prostate cancers drives clinical aggression. *Cell* **173**, 1003–1013 (2018).
20. Pressinotti, N. C. et al. Differential expression of apoptotic genes PDIA3 and MAP3K5 distinguishes between low- and high-risk prostate cancer. *Mol. Cancer* **8**, 130 (2009).
21. Isella, C. et al. Stromal contribution to the colorectal cancer transcriptome. *Nat. Genet.* **47**, 312–319 (2015).
22. Ragnum, H. B. et al. The tumour hypoxia marker pimonidazole reflects a transcriptional programme associated with aggressive prostate cancer. *Br. J. Cancer* **112**, 382–390 (2015).
23. Scialdone, A. et al. Computational assignment of cell-cycle stage from single-cell transcriptome data. *Methods* **85**, 54–61 (2015).
24. Cuzick, J. 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.* **12**, 245–255 (2011).
25. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).
26. Novershtern, N. et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).
27. Muller, S. et al. Single-cell profiling of human gliomas reveals macrophage ontogeny as a basis for regional differences in macrophage activation in the tumor microenvironment. *Genome Biol.* **18**, 234 (2017).
28. Hayes, A. R., Brungs, D. & Pavlakis, N. Osteoclast inhibitors to prevent bone metastases in men with high-risk, non-metastatic prostate cancer: a systematic review and meta-analysis. *PLoS ONE* **13**, e0191455 (2018).
29. Fiorino, C. & Harrison, R. E. E-cadherin is important for cell differentiation during osteoclastogenesis. *Bone* **86**, 106–118 (2016).
30. Tsuchida, K. et al. Activin signaling as an emerging target for therapeutic interventions. *Cell Commun. Signal.* **7**, 15 (2009).
31. Sugatani, T. et al. Ligand trap of the activin receptor type IIA inhibits osteoclast stimulation of bone remodeling in diabetic mice with chronic kidney disease. *Kidney Int.* **91**, 86–95 (2017).
32. Alon, R. A sweet solution: glycolysis-dependent Treg cell migration. *Immunity* **47**, 805–807 (2017).
33. Xu, T. et al. Metabolic control of TH17 and induced Treg cell balance by an epigenetic mechanism. *Nature* **548**, 228–233 (2017).
34. van Loosdregt, J. et al. Canonical Wnt signaling negatively modulates regulatory T cell function. *Immunity* **39**, 298–310 (2013).
35. Wild, C. A. et al. HMGB1 conveys immunosuppressive characteristics on regulatory and conventional T cells. *Int. Immunol.* **24**, 485–494 (2012).
36. Harjunpaa, H., Lloret Asens, M., Guenther, C. & Fagerholm, S. C. Cell adhesion molecules and their roles and regulation in the immune and tumor microenvironment. *Front. Immunol.* **10**, 1078 (2019).
37. Tung, K. H., Ernstoff, M. S., Allen, C. & Shu, S. A review of exosomes and their role in the tumor microenvironment and host–tumor ‘macroenvironment’. *J. Immunol. Sci.* **3**, 4–8 (2019).
38. Bartoschek, M. et al. Spatially and functionally distinct subclasses of breast cancer-associated fibroblasts revealed by single cell RNA sequencing. *Nat. Commun.* **9**, 5150 (2018).
39. Ayala, G. et al. Reactive stroma as a predictor of biochemical-free recurrence in prostate cancer. *Clin. Cancer Res.* **9**, 4792–4801 (2003).
40. Vellanki, R. N., Zhang, L. & Volchuk, A. OASIS/CREB3L1 is induced by endoplasmic reticulum stress in human glioma cell lines and contributes to the unfolded protein response, extracellular matrix production and cell migration. *PLoS ONE* **8**, e54060 (2013).
41. Varrault, A. et al. Identification of Plagl1/Zac1 binding sites and target genes establishes its role in the regulation of extracellular matrix genes and the imprinted gene network. *Nucleic Acids Res.* **45**, 10466–10480 (2017).
42. Sieweke, M. H., Tekotte, H., Frampton, J. & Graf, T. MafB is an interaction partner and repressor of Ets-1 that inhibits erythroid differentiation. *Cell* **85**, 49–60 (1996).
43. Sinha, M. et al. Direct conversion of injury-site myeloid cells to fibroblast-like cells of granulation tissue. *Nat. Commun.* **9**, 936 (2018).
44. Ochiya, T., Takenaga, K. & Endo, H. Silencing of S100A4, a metastasis-associated protein, in endothelial cells inhibits tumor angiogenesis and growth. *Angiogenesis* **17**, 17–26 (2014).
45. Schubert, K., Gutknecht, D., Koberle, M., Anderegg, U. & Saalbach, A. Melanoma cells use Thy-1 (CD90) on endothelial cells for metastasis formation. *Am. J. Pathol.* **182**, 266–276 (2013).
46. Finkenzeller, G., Hager, S. & Stark, G. B. Effects of bone morphogenetic protein 2 on human umbilical vein endothelial cells. *Microvasc. Res.* **84**, 81–85 (2012).
47. de Sousa Mesquita, A. P., de Araujo Lopes, S., Pernambuco Filho, P. C. A., Nader, H. B. & Lopes, C. C. Acquisition of anoikis resistance promotes alterations in the Ras/ERK and PI3K/Akt signaling pathways and matrix remodeling in endothelial cells. *Apoptosis* **22**, 1116–1137 (2017).
48. Wojno, K. J. & Epstein, J. I. The utility of basal cell-specific anti-cytokeratin antibody (34βE12) in the diagnosis of prostate cancer. A review of 228 cases. *Am. J. Surg. Pathol.* **19**, 251–260 (1995).
49. Parsons, J. K., Gage, W. R., Nelson, W. G. & De Marzo, A. M. p63 protein expression is rare in prostate adenocarcinoma: implications for cancer diagnosis and carcinogenesis. *Urology* **58**, 619–624 (2001).
50. Heck, M. M. et al. Topography of lymph node metastases in prostate cancer patients undergoing radical prostatectomy and extended lymphadenectomy: results of a combined molecular and histopathologic mapping study. *Eur. Urol.* **66**, 222–229 (2014).

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

© The Author(s), under exclusive licence to Springer Nature Limited 2021

Methods

Cell lines. PC3 and DU145 cell lines were obtained from Stem Cell Bank, Chinese Academy of Sciences, and were cultured in RPMI-1640 supplemented with 9% fetal bovine serum (FBS, Gibco) and 1× antibiotic-antimycotic (Gibco). C4-2B cells were provided by L. Chung (Cedars-Sinai) and were cultured in RPMI-1640 supplemented with 9% FBS (Gibco) and 1× antibiotic-antimycotic. For EV harvesting, the culture medium was supplemented with 9% EV-depleted FBS (SBI) d instead of normal FBS.

Human specimens. The prostate tumour samples used in this study were collected from patients who had undergone radical prostatectomy at Shanghai Changhai Hospital. The non-cancerous prostate samples and an incidental prostate cancer sample used in this study were collected from three male patients with bladder cancer who had undergone radical cystectomy at Shanghai Changhai Hospital. The lymph node samples were collected by lymph node dissection performed along with radical prostatectomy. Samples of CRPC were collected using surgical procedures outlined in Supplementary Table 1. Human peripheral blood was collected from healthy donors. The study is compliant with all relevant ethical regulations regarding research involving human participants. The Institutional Review Board (IRB) of Shanghai Changhai Hospital approved the tissue acquisition procedures and experimental protocols. Informed consent was obtained from all participants. Gleason scores and tumour cellularity were evaluated by two genito-urinary pathologists on scanned H&E-stained slides. For patient SC001H, MRI and bone scans were performed during routine clinical practice. Fresh tissue samples were immediately dissected into fractions for (1) flash freezing, (2) fixation in 4% paraformaldehyde solution followed by paraffin embedding and (3) enzymatic digestion into single cells as described below.

Tissue processing. Fresh tissue samples were surgically removed and kept in MACS tissue storage solution (Miltenyi Biotec) until processing. Briefly, samples were first washed with phosphate-buffered saline (PBS), minced into small, ~1-mm cubic pieces on ice and enzymatically digested with the following protocol: DNase I (Worthington; 30 U ml⁻¹), collagenase IV (Worthington; 195 U ml⁻¹), collagenase I (Worthington; 10 U ml⁻¹) and 30% FBS for 1 h at 37°C, with agitation to achieve single cells.

After digestion, samples were sieved through a 70-µm cell strainer (Corning) and centrifuged at ~300–500g for 8 min. After the supernatant was removed, the pelleted cells were suspended in red blood cell lysis buffer and incubated for 2 min to lyse red blood cells. After washing with RPMI-1640 medium, the cell pellets were resuspended in RPMI-1640 medium and re-filtered through a 35-µm cell strainer (Corning).

For the BD Rhapsody platform, the viability of dissociated single cells was assessed using Calcein-AM (Thermo Fisher) and Draq7 (BD Biosciences). The single-cell suspension was further enriched with a dead cell removal kit (Miltenyi Biotec).

Single-cell RNA sequencing. The 13 samples in batch 1 were sequenced with 10X Genomics and the rest of the samples, including batches 2 and 3, were sequenced with BD Rhapsody. Detailed methods are outlined in the following sections.

Sequencing with 10X Genomics. Cell capture and cDNA synthesis was carried out using the Single Cell 3' Library and Gel Bead Kit V2 (10X Genomics, cat. no. 120237) and a Chromium Single Cell A Chip Kit (10X Genomics, 120236). The cell suspension (300–600 living cells per microlitre, determined by Count Star) was loaded onto the Chromium single-cell controller (10X Genomics) to generate single-cell gel beads in the emulsion, according to the manufacturer's protocol. In short, single cells were suspended in PBS containing 0.04% bovine serum albumin (BSA). Approximately 7,000 cells were added to each channel, and the target cell recovery rate was estimated to be 3,000 cells. Captured cells were lysed and the released RNA were barcoded through reverse transcription in individual gel bead in emulsions (GEMs)³¹.

Using an S1000TM Touch Thermal Cycler (Bio-Rad) to reverse transcribe, the GEMs were programmed at 53°C for 45 min, followed by 85°C for 5 min, then held at 4°C. Complementary DNA (cDNA) was generated and then amplified, and the quality was assessed using an Agilent 4200 system (performed by CapitalBio, Beijing).

For scRNA-seq library preparation, according to the manufacturer's protocol, scRNA-seq libraries were constructed using a Single Cell 3' Library Gel Bead Kit V2. Sequencing was performed on an Illumina NovaSeq 6000 sequencer with a sequencing depth of at least 100,000 reads per cell and 150-bp (PE150) paired-end reads (performed by CapitalBio, Beijing).

Sequencing with a BD Rhapsody system. Single-cell capture was achieved by random distribution of a single-cell suspension across >200,000 microwells through a limited dilution approach. Beads with oligonucleotide barcodes were added to saturation so that a bead was paired with a cell in a microwell. Cell-lysis buffer was added so that poly-adenylated RNA molecules hybridized to the beads. Beads were collected into a single tube for reverse transcription. Upon cDNA synthesis, each cDNA molecule was tagged on the 5' end (that is, the 3' end of an mRNA transcript) with a unique molecular identifier (UMI) and cell label indicating its

cell of origin. Whole transcriptome libraries were prepared using the BD Rhapsody single-cell whole-transcriptome amplification workflow. In brief, second-strand cDNA was synthesized, followed by ligation of the WTA adaptor for universal amplification. Eighteen cycles of PCR were used to amplify the adaptor-ligated cDNA products. Sequencing libraries were prepared using random priming PCR of the whole-transcriptome amplification products to enrich the 3' end of the transcripts linked with the cell label and UMI. Sequencing libraries were quantified using a High Sensitivity DNA chip (Agilent) on a Bioanalyzer 2200 system and a Qubit High Sensitivity DNA assay (Thermo Fisher Scientific). The library for each sample was sequenced by HiSeq X Ten (Illumina) on a 150-bp paired-end run.

Flow cytometry and sorting. Prostate cancer tissues were minced and digested with collagenase (Sangon Biotech) and DNase (Solarbio). The cell suspension was filtered through 100-µm cell strainers (BD Falcon) and centrifuged at 1,200 r.p.m. for 5 min. The precipitate was resuspended in flow cytometry and sorting (fluorescence-activated cell sorting; FACS) buffer (1% FBS/1640 medium) and filtered through 70-µm cell strainers (BD Falcon) and washed once with the FACS buffer. The resuspended cells were then stained for PSMA (Alexa Fluor 488; ab187570, 1:150, Abcam) and CD8 (PerCP/Cy5.5; ab157306, 1:150, Abcam). After washing three times with FACS buffer, the prepared cell suspension was filtered through 40-µm cell strainers (BD Falcon) and processed by a Sony cell sorter SH800S. PSMA⁺CD8⁺ cells and PSMA-CD8⁺ cells were collected, respectively, in 15-ml centrifuge tubes. The gating strategy to identify the different cell types is shown in Extended Data Fig. 6a.

To assay the expression of CD31 and CD90, 10⁶ cells were incubated at room temperature for 30 min with 5 µl of FITC anti-human CD31 (cat. no. 303104, Biolegend) and 5 µl of PE/cy7 anti-human CD90 (cat. no. 328124, Biolegend), while taking 5 µl of non-specific isotype-matched control immunoglobulin-G incubated cells as control. Unbound antibodies were removed by washing the cells twice in the PBS buffer. The gating strategy to identify different cell types is shown in Fig. 6f.

Cells were analysed by a flow cytometer (FC500, Beckman Coulter). FlowJo V10 was used for data analysis. All experiments were repeated three times.

Isolation of CD8⁺ T cells. CD8⁺ T cells were collected from healthy donor human peripheral blood using anti-CD8 microbeads (MACS, Miltenyi) according to the manufacturer's recommendations and were activated for 24 h with TransAct T cell reagents (Miltenyi, 130-111-160). Activated T cells were cultured in X-VIVO 15 culture medium (Lonza, 04-418-QCN) supplemented with 10% FBS (Gibco, 10091148), 1% penicillin-streptomycin (BasalMedia, J40602) and anti-human interleukin-2 (Peprotech, 200-21-100; 100 U ml⁻¹).

RNA-FISH analysis. RNA-FISH was conducted using a Ribo fluorescence in situ hybridization kit (C10910, RiboBio) in accordance with the manufacturer's directions and as previously described³². In brief, the cells were seeded and fixed with 4% paraformaldehyde and treated with 0.5% Triton in PBS followed by pre-hybridization. The fixed cells were then hybridized at 5 mM probe concentration overnight. KLK3 FISH probes (LNC1101594, RiboBio) were designed and synthesized by RiboBio. All images were visualized and obtained by a confocal microscope (Philips).

Cell line immunofluorescent staining. After co-culturing, T cells were mounted on adhesion microscope slides (CITOTEST, cat. no. 188105), then 0.3% Triton X-100 was used to permeabilize the cell membrane. Immunol staining blocking buffer (Beyotime, cat. no. P0102) was added to block non-specific binding for 1 h. Slides were incubated overnight with antibodies against CD8 (Abcam, ab237709, 1:100) and PSA (Abcam, ab140337, 1:100) at 4°C in a moisture chamber. Slides were washed in PBS three times then incubated for 1 h with fluorescence-labelled secondary antibodies (CST 4409 and 4412, 1:1,000, Invitrogen) at room temperature in a light-proof moisture chamber. Slides were washed three times in PBS, then the nuclei were stained with DAPI (Beyotime, cat. no. C1005) for 10 min. The slides were then examined and captured using a fluorescence microscope (Olympus).

Isolation and culture of primary fibroblasts. Primary prostate fibroblasts were isolated from freshly removed prostate samples from eight patients with prostate cancer (PCa) and four cases of patients undergoing radical cystectomy for invasive transitional cell carcinoma of the bladder without any histopathological evidence of carcinoma in the prostate. All tissues were minced with scalpels and then enzymatically dissociated in DMEM/F12 medium (cat. no. 11330-032, Gibco) supplemented with 2% BSA, 10 ng ml⁻¹ cholera toxin (cat. no. c8052, Sigma), 300 U ml⁻¹ collagenase (cat. no. 17018-029, Invitrogen) and 100 U ml⁻¹ hyaluronidase (cat. no. H3506, Sigma) at 37°C for 18 h. The digested suspension was centrifuged at 600 r.p.m. for 4 min to separate the epithelial and fibroblast cells. To pellet the fibroblasts, the supernatant was collected and centrifuged at 800 r.p.m. for 10 min. The cell pellet containing the fibroblasts was resuspended in DMEM/F12 medium supplemented with 10% FBS (Gemini) and 5 µg ml⁻¹ insulin (cat. no. 40112ES25, Yeasen). All tissues were obtained from the Shanghai Changhai Hospital (see 'Human specimens' section).

Co-culture experiment. A total of 2×10^6 activated T cells were co-cultured with 2×10^6 AR-positive (C4-2B) or -negative (DU145) PCa cells in six-well plates for 24 h. Activated T cells without co-culture of PCa cells served as a negative control. CD3⁺ T cells were collected by immunostaining with PE-conjugated anti-CD3 antibodies (cat. no. 300307, BioLegend, 1:20) and sorted using a FACSAria III cell sorter (BD Biosciences) following the manufacturer's protocols.

EV isolation. The process of EV isolation was conducted as previously described^{53,54}, with minor modifications. Briefly, PCa cells (DU145, C4-2B) were cultured in RPMI-1640 (Gibco) supplemented with 10% exosome-depleted FBS (SBI). After 48 h, the cell culture medium was collected and centrifuged at 400g at room temperature (RT) to remove cell debris. To further remove apoptotic bodies and microvesicles, the medium supernatant was collected and subjected to the following centrifugation steps at 4°C: (1) 2,000g, 20 min; (2) 15,000g, 40 min. The resulting supernatant was passed through a 0.22- μ m Millex-GP filter (Millipore, cat. no. SLGP033RB) and subjected to ultracentrifugation at 120,000g for 4 h to pellet EVs. The supernatant was carefully removed, and crude EV-containing pellet was resuspended in a large volume of ice-cold PBS, followed by ultracentrifugation at 120,000g to wash the EV product. EV protein concentration was evaluated using a Pierce BCA protein assay kit (Thermo Fisher) and used to estimate EV quantity. EV isolation for gradient centrifugation analysis was performed with an SW28 swinging rotor (Beckman Coulter) with minor modifications.

EV administration experiment. Approximately 2×10^6 activated T cells were seeded in six-well plates and subsequently treated with EVs derived from PCa cells DU145 or C4-2B for 24 h. RNA of T cells was isolated using RNAiso Plus Reagent (Takara, cat. no. 9109).

Analysis of EV isolates with gradient centrifugation. *Gradient centrifugation of EV.* Freshly pelleted EVs were resuspended in filtered PBS and deposited at the bottom of an ultracentrifuge tube, then 30% (4.3 ml, 1.18 g ml⁻¹), 25% (3 ml, 1.15 g ml⁻¹), 15% (2.5 ml, 1.10 g ml⁻¹) and 5% (6 ml, 1.05 g ml⁻¹) iodixanol solutions were sequentially layered in decreasing density to form a discontinuous gradient⁵⁵. Separation was performed by ultracentrifugation at 100,000g for 3 h 50 min (4°C, k-factor 254.7) and density fractions were collected separately and diluted up to 17 ml in PBS. Each density fraction was then split into two for the collection of RNA and protein, spun at 100,000g (60 min, 4°C) and resuspended in the appropriate buffer. All ultracentrifugation spins were performed in an SW28 swinging rotor (Beckman Coulter).

Immunoblotting. For immunoblotting analysis, freshly pelleted material from the flotation of small EVs was resuspended in 100 μ l of 4% SDS/Tris-HCl lysis buffer and 10 μ l of sample was loaded in each lane. Whole cell lysate (1 μ g) from parental cells was included as a control. Immunoblotting analysis was performed as described previously⁵⁶. The primary antibodies used were PSA (sc-7638 (C-19), 1:5,000 dilution) from Cell Signaling and GM130 (ab52649 (EP892Y), 1:10,000 dilution) and CD81 (ab79559 (M38), 1:10,000 dilution) from Abcam.

Extraction of total RNA and quantitative PCR. Total RNA was extracted with Trizol reagent (Invitrogen) according to the manufacturer's directions. Briefly, freshly pelleted material from each of the density fractions was resuspended in 1 ml of Trizol, mixed with 200 μ l of chloroform and centrifuged at 12,000g (4°C, 15 min) for collection of the total RNA fraction within the aqueous phase. Total RNA was then precipitated in one volume of ice-cold 2-propanol in the presence of 20 μ g ml⁻¹ glycogen (Invitrogen) for 24 h. RNA was then collected (12,000g, 10 min, 4°C), washed in 75% RNase-free ethanol and resuspended in 20 μ l of RNase-free water. Samples were stored at -20°C until use.

cDNA from 10 μ l of total RNA was obtained with the iScript cDNA synthesis kit (Bio-Rad) following the manufacturer's instructions. The resultant 20 μ l of cDNA was then subjected to TaqMan qPCR in a QuantStudio 5 real-time PCR device (Applied Biosystems) for the probes ACTB (Hs01060665_g1), GAPDH (Hs99999905_m1) and KLK3 (Hs02576345_m1). Samples were run in duplicate and negative controls were included.

Quantitative real-time PCR. Total RNA was isolated from cells using RNAiso Plus reagent (Takara) and reverse transcribed into first-strand cDNA with a PrimeScript II 1st Strand cDNA synthesis kit (Takara). The samples were then analysed in an Applied Biosystems 7500 Real-Time PCR system. β -actin was used as an endogenous control.

Cell invasion assay. To measure the effect of aECs on prostate cancer cell invasion ability, aECs (1×10^5) were seeded into the upper chamber and PC-3 cells (5×10^5) were added into the lower chamber of a six-well cell culture insert with 0.4- μ m pore size (Merck Millipore, PHT30R48), taking parental fibroblasts as the control group. The cells were co-cultured for 96 h before invasion assays.

BD Falcon cell culture inserts (8 μ m; cat. no. 353097, BD) precoated with Matrigel (cat. no. 356234, BD) were used for examining cell invasion. A suspension

of PC-3 cells co-cultured with parental cells or with CD31⁺/CD90⁺ aECs was placed into the upper chamber in 0.2 ml of F12k serum-free medium (1×10^5 cells per filter). F12k medium supplemented with 10% FBS was placed in the lower chamber as a chemoattractant. Invasion was scored following 48 h. Cells at the lower surface of the inserts were then fixed in 4% paraformaldehyde (Sangon Biotech, cat. no. A500684) for 30 min at room temperature and stained with Coomassie brilliant blue (BBI, cat. no. A602151) for 30 min. Values for cell invasion were expressed as the mean number of cells per microscopic field over five fields per one insert for triplicate experiments.

Statistics and reproducibility. Statistical analyses were performed using the R statistical environment (v3.5.1). For cell line experiments, data were shown as mean \pm s.d., *n* represents the number of technical replicates, and biological replicates were shown separately unless otherwise stated in figure legends. Experiments were repeated at least twice with similar results unless otherwise stated in the legends. All tests were two-sided unless otherwise specified. CLES was calculated using the R package canprot (v0.1.2). The type of test method used for statistical analysis is specified in the text where the results are described and details for the test are explained in the relevant figure legend and Methods section. For the Student's *t*-test, a Shapiro-Wilk test was used to test for normality. For survival analysis, the assumptions of the Cox proportional hazards model were tested using the 'cox.zph' function in the R Survival package (v2.41-3) with 0.1 as cutoff. A log-rank test was used when the cox.zph test failed. For samples where *KLK3* could not be detected after 40 cycles in rtPCR analysis, 40 was assigned as the Ct (cycle threshold) value in calculations.

Sequencing data preprocessing. We applied fastp⁵⁷ with default parameter filtering of the adaptor sequence and removing the low-quality reads.

For data generated with the BD platform, Umi-tools (v1.0.0) was applied to identify the cell barcode whitelist. Raw data were mapped using STAR (v2.5.2b) with default parameters from the UMI-tools standard pipeline to obtain the UMI counts of each sample. For data generated with the 10X platform, raw data were processed using cellranger (v2.1.0) into the UMI matrix. For all analysis, human genome (ensemble hg38, v91) was used as the reference. The maximum number of cells (6,152) in one sample was estimated by cellranger across 13 samples and was used as the estimation of cell number. The top 6,152 cells with the highest number of genes from each sample are included for further analysis.

For each cell, we calculated three quality measures: percent of mitochondrial genes, number of total genes and housekeeping genes expressed⁹. We removed cells that had more than 20% expression on mitochondrial genes, fewer than 200 or more than 5,218 total genes expressed, or fewer than 56 housekeeping genes expressed⁹.

The resultant gene expression matrix was imported into the R (v3.5.1) statistical environment for further analysis. Data normalization and annotation were performed using the R package scan (v1.10.2)⁵⁸. To calculate the pool-based normalization factor, cells for each sample were first split into sensible clusters using the quickCluster function (max.size = 3,000). The fastMNN function (*k* = 5, *d* = 50, approximate = TRUE, auto.order = TRUE) was used to apply the mutual nearest-neighbour method to correct for batch effect among samples.

Major cell type annotation and CNA inference. The normalized and batch corrected data were imported into Seurat (v2.3.4)⁵⁹ for downstream analysis and visualization. Dimensionality reduction was then performed with principal component analysis on the batch corrected data. We then performed tSNE dimensionality reduction on the calculated principal components (PCs) to obtain a two-dimensional representation for data visualization. A shared nearest-neighbour graph was also calculated using the top eight PCs to determine clusters using the function FindClusters.

To annotate the identity of each cluster, we performed QuSAGE (2.16.1)¹⁵ analysis as described before¹⁶. Briefly, script was adapted from this previous work, then (1) a list of marker genes for different cell types was collected, (2) the activation score for each cell type was calculated using the quusage function and (3) the cell type with the highest positive activation score was assigned to each cluster. We used the previously reported signature¹⁶ to assign broad lineage (epithelia and stroma) for cells. For detailed annotation, a list of marker genes was collected from a public database and previous work^{10,60,61}.

Copy number analysis was performed as previously described⁹ with the R package inferCNV (v0.8.2) (cutoff = 0.1). The obtained residual expression for each gene is a surrogate for CNAs of the gene and the extent of CNAs for each sample was defined as the mean of squares of deviation across the genome.

Differential gene expression analysis. To identify differentially expressed (marker) genes for clusters or subtypes, the functions FindAllMarkers (multiple condition comparisons) and FindMarkers (two condition comparison) from the Seurat package were used with default parameters. Significant differentially expressed genes (markers) were selected as those with adjusted *P* values less than 0.05, average fold-change larger than 1.5 and percentage of cells with expression higher than 0.1.

Cell-cycle analysis. The cyclone function from the R package *scran*⁵⁸ was used for cell-cycle analysis.

Survival and correlation analysis in bulk RNA-seq data. For survival analysis, the assumption of the Cox proportional hazards model was tested using the *cox.zph* function in the R Survival package (v2.41-3) with 0.1 as cutoff. The Cox proportional hazards model was then fit using patient groups dichotomized by median level of signature, with BCR as endpoint. Data visualization was performed using the R library BPG (v5.9.8)⁶².

Marker genes for different cell populations were identified as described in the 'Differential gene expression analysis' section. A signature score was calculated as the mean expression of all marker genes.

Bulk RNA-seq data from previous studies were collected for survival analysis. For GSE70770, data were downloaded with the R package GEOquery (v2.50.5)⁶³. Memorial Sloan-Kettering Cancer Center (MSKCC) data and BCR information were downloaded with the R package *cgdsr* (v1.2.10). Other clinical information was downloaded from GSE21034¹⁷ with GEOquery. For all five datasets, z-score normalization was performed on the non-log₂-transformed expression matrix.

For correlation analysis, z-score normalization was performed on the log₂-transformed expression matrix. For correlation analysis in the TCGA dataset, data for 499 prostate cancer samples were downloaded using the R package TCGA2STAT (v1.2)⁶⁴.

Reanalysis of major cell types. For subtype assessment within the major cell types, we re-analysed cell subsets separately. For each TME cell subset, we used the top 10 MNNs of 20 re-analysed MNNs for tSNE construction, using a resolution of 0.8 for graph-based clustering to identify the cell cluster. All 20 MNNs were used for epithelial cell reanalysis.

Pseudotime analysis. We applied the Single-Cell Trajectories analysis with Monocle2¹³ using DDR-Tree and default parameters. For Monocle analysis, positive marker genes for each cluster were used. Based on the pseudotime analysis, branch expression analysis modelling (BEAM analysis) was applied for branch fate determined gene analysis.

Cell communication analysis. To enable a systematic analysis of cell–cell communication molecules, we applied cell communication analysis based on CellPhoneDB (v1.1.0)¹⁴, a public repository of ligands, receptors and their interactions. Membrane, secreted and peripheral proteins of the cluster of different time points were annotated. Significant mean and cell communication significance ($P < 0.05$) was calculated based on the interaction and the normalized cell matrix achieved by *scran* normalization.

Transcription factor analysis. To assess TF regulation strength, we applied the single-cell regulatory network inference and clustering (pySCENIC, v0.9.5)¹² workflow, using the 20-thousand motifs database for RcisTarget and GRNboost.

Gene enrichment analysis. For gene enrichment analysis, Fisher's exact test was applied to calculate the P value for each gene set. Raw P values were adjusted for multiple hypothesis tests using the Benjamini and Hochberg method. Such enrichment analysis was applied to annotations including GO (v2.5.13), the Kyoto Encyclopedia of Genes and Genomes (KEGG, 201900613) and customized gene sets as indicated in the relevant figure legends and main text. We downloaded the GO annotations from NCBI (<http://www.ncbi.nlm.nih.gov/>), UniProt (<http://www.uniprot.org/>) and Gene Ontology (<http://www.geneontology.org/>). KEGG annotation was purchased from the database (<https://www.genome.jp/kegg/>). A customized 41-signature gene set including immune-, cytokine- and neurobiology-related terms was collected from the CellPhoneDB database, the neurotransmitters receptor gene of Genebank (<https://www.ncbi.nlm.nih.gov/genbank/>) and the immune scoring gene set from previous work⁷.

To characterize the relative activation of a given gene set, the R package QuSAGE was used as described in the 'Major cell type annotation and CNA inference' section.

Co-regulated gene analysis. To discover the gene co-regulation network, the *find_gene_modules* function of Monocle3⁶⁵ was used with default parameters.

Analysis of public scRNA-seq data. To analyse the epithelial gene expression in T cells, several data matrices were collected from NCBI GEO datasets, including NSCLC⁶⁶, CRC⁶⁷, HCC⁵¹ and HNSCC¹⁰. T cells were identified as defined in the Supplementary Information.

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

Data availability

Data have been deposited in the Gene Expression Omnibus (GEO) under accession no. GSE141445 and the Genome Sequence Archive for Human (GSA-Human) under accession HRA000312 and can be accessed at www.pradcellatlas.com.

For gene expression analysis in T cells, scRNA-seq data from the following NCBI GEO accessions were used: GSE99254 (NSCLC)⁶⁶, GSE108989 (CRC)⁶⁷, GSE98638 (HCC)⁵¹ and GSE103322 (HNSCC)¹⁰. For survival analysis, bulk RNA-seq data from the following studies were used: TCGA (333 samples, <http://firebrowse.org/?cohort=PRAD>)⁴, ref. 17 (131 samples, MSKCC, <https://doi.org/10.1016/j.ccr.2010.05.026>), ref. 68 (294 samples, GSE70770), ref. 69 (79 samples, <https://doi.org/10.1172/JCI20032/>) and Changhai 2020 (136 samples, www.cpgea.com)⁷⁰. All other data supporting the findings of this study are available from the corresponding author on reasonable request. Source data are provided with this paper.

Code availability

All R packages used are available online, as described in the Methods. Customized code for data analysis and plotting are available on GitHub (<https://github.com/chensujun/scRNA>).

References

- Zheng, C. et al. Landscape of infiltrating T cells in liver cancer revealed by single-cell sequencing. *Cell* **169**, 1342–1356 (2017).
- Guo, X. et al. A linc1405/Eomes complex promotes cardiac mesoderm specification and cardiogenesis. *Cell Stem Cell* **22**, 893–908 (2018).
- Jeppesen, D. K. et al. Reassessment of exosome composition. *Cell* **177**, 428–445 (2019).
- Minciacci, V. R. et al. Large oncosomes contain distinct protein cargo and represent a separate functional class of tumor-derived extracellular vesicles. *Oncotarget* **6**, 11327–11341 (2015).
- Vagner, T. et al. Large extracellular vesicles carry most of the tumour DNA circulating in prostate cancer patient plasma. *J. Extracell. Vesicles* **7**, 1505403 (2018).
- Zhou, B. et al. Low-background acyl-biotinyl exchange largely eliminates the coisolation of non-S-acylated proteins and enables deep S-acylproteomic analysis. *Anal. Chem.* **91**, 9858–9866 (2019).
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
- Lun, A. T., McCarthy, D. J. & Marioni, J. C. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with BioConductor. *F1000Res.* **5**, 2122 (2016).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Taylor, R. A., Toivanen, R. & Risbridger, G. P. Stem cells in prostate cancer: treating the root of the problem. *Endocr. Relat. Cancer* **17**, R273–R285 (2010).
- Zhang, X. et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res.* **47**, D721–D728 (2019).
- P'ng, C. et al. BPG: seamless, automated and interactive visualization of scientific data. *BMC Bioinf.* **20**, 42 (2019).
- Davis, S. & Meltzer, P. S. GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* **23**, 1846–1847 (2007).
- Wan, Y. W., Allen, G. I. & Liu, Z. TCGA2STAT: simple TCGA data access for integrated statistical analysis in R. *Bioinformatics* **32**, 952–954 (2016).
- Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
- Guo, X. et al. Global characterization of T cells in non-small-cell lung cancer by single-cell sequencing. *Nat. Med.* **24**, 978–985 (2018).
- Zhang, L. et al. Lineage tracking reveals dynamic relationships of T cells in colorectal cancer. *Nature* **564**, 268–272 (2018).
- Ross-Adams, H. et al. Integration of copy number and transcriptomics provides risk stratification in prostate cancer: a discovery and validation cohort study. *EBioMedicine* **2**, 1133–1144 (2015).
- Glinsky, G. V., Glinskii, A. B., Stephenson, A. J., Hoffman, R. M. & Gerald, W. L. Gene expression profiling predicts clinical outcome of prostate cancer. *J. Clin. Invest.* **113**, 913–923 (2004).
- Li, J. et al. A genomic and epigenomic atlas of prostate cancer in Asian populations. *Nature* **580**, 93–99 (2020).

Acknowledgements

This work was supported by the National Key R&D Plan of the China Precision Medicine Project (2017YFC0908002 to S.R.), the National Natural Science Foundation of China (81872105 to S.R.), the Princess Margaret Cancer Foundation (88612001223 to H.H.H.), the Canada Foundation for Innovation and Ontario Research Fund (CFI32372 to H.H.H.), a NSERC discovery grant (498706 to H.H.H.), a Canadian Cancer Society Innovation Grant (703800 to H.H.H.), Prostate Cancer Canada (TAG2018-2061, RS2016-1022 and D2016-1115 to H.H.H.), CIHR operating grants (142246, 152863, 152864 and 159567 to H.H.H.) and a Terry Fox New Frontiers Program Project Grant (1090 P3 to H.H.H.). H.H.H. was supported by an OMIR Early Researcher Award and CIHR New Investigator Award. H.H.H. holds the Joey and Toby Tanenbaum Brazilian Ball Chair in Prostate Cancer. J.W. was partially funded by the National Natural Funding of China (81272404, 81772806 and 81972744). G.Z. and S.W. were supported by the Sanming

Project of Medicine in Shenzhen (SZSM201601043). Y.L. was funded by the National Natural Science Foundation of China (31801111) and the Dream Mentor–Outstanding Young Talent Program (fkyq1910). P.C.B. was supported by the NIH/NCI under award nos. P30CA016042, 1U01CA214194-01 and 1U24CA248265-01. We thank L.W. Chung from Cedars-Sinai for providing the C4-2B cell line.

Author contributions

The studies were designed by S.C., G.Z., J.W., M.F., D.D.V., H.H.H. and S.R. Experiments were performed by Y.-T.X., N.Z., X.B., Y. Yang, F.W., C.W., Y.Z., Y. Yu, K.D., J.M., Y.L., F.S., H.L.Y., M.L. and W.C. Data analysis was carried out by S.C., G.Z., B.Z., F.L., W.C., D.C., Q.G., Z.Y., S.W., M.F., P.C.B., D.D.D.C., T.v.d.K., Z.J. and A.B. The first draft of the manuscript was written by S.C., G.Z., S.R. and H.H.H. All authors revised and approved the manuscript.

Competing interests

B.Z., W.C., C.W., D.C. and Q.G. are co-founders for Novel Bioinformatics Co., Ltd. All other authors declare no competing interests.

Additional information

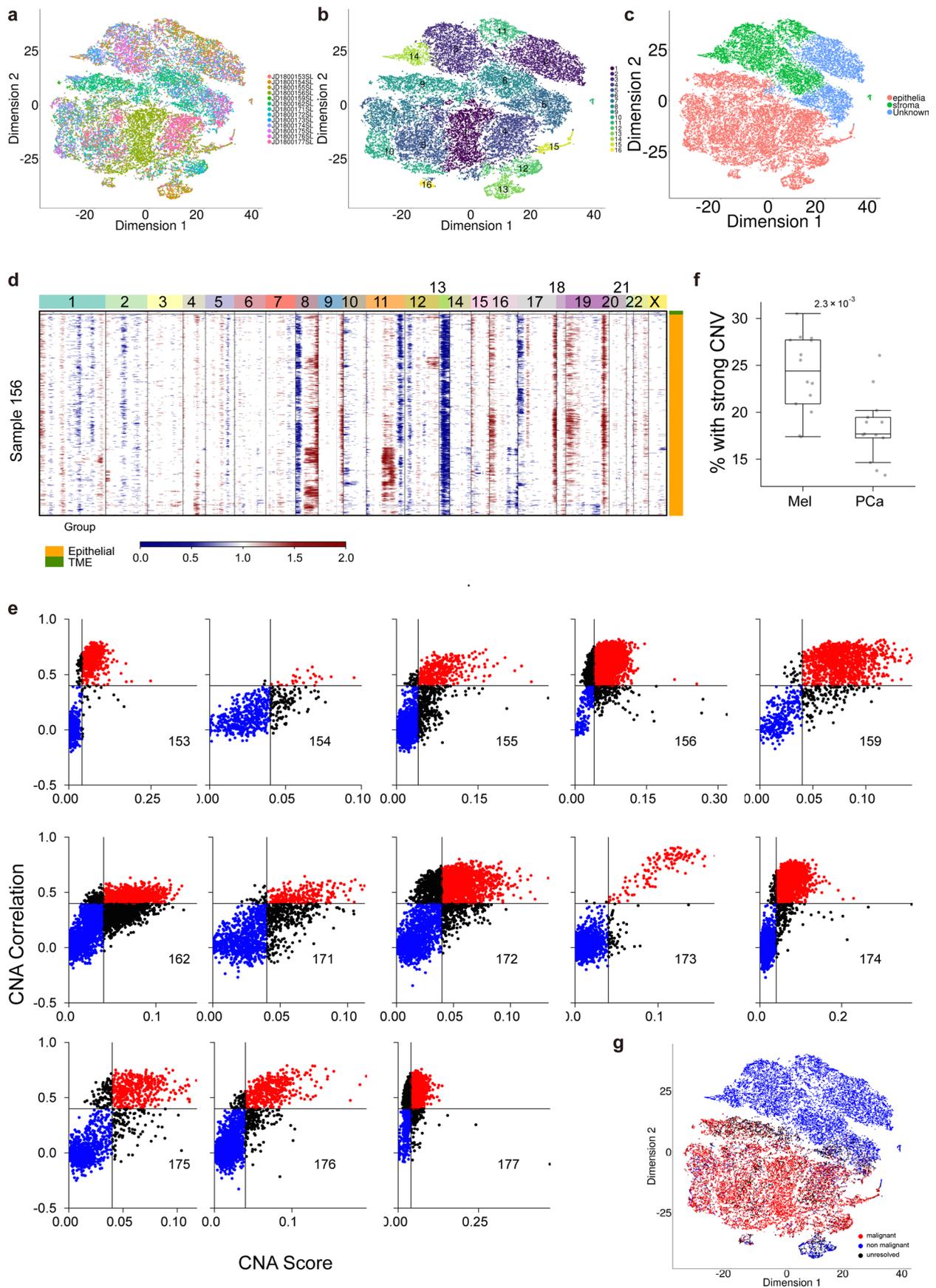
Extended data is available for this paper at <https://doi.org/10.1038/s41556-020-00613-6>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41556-020-00613-6>.

Correspondence and requests for materials should be addressed to J.W., H.H.H. or S.R.

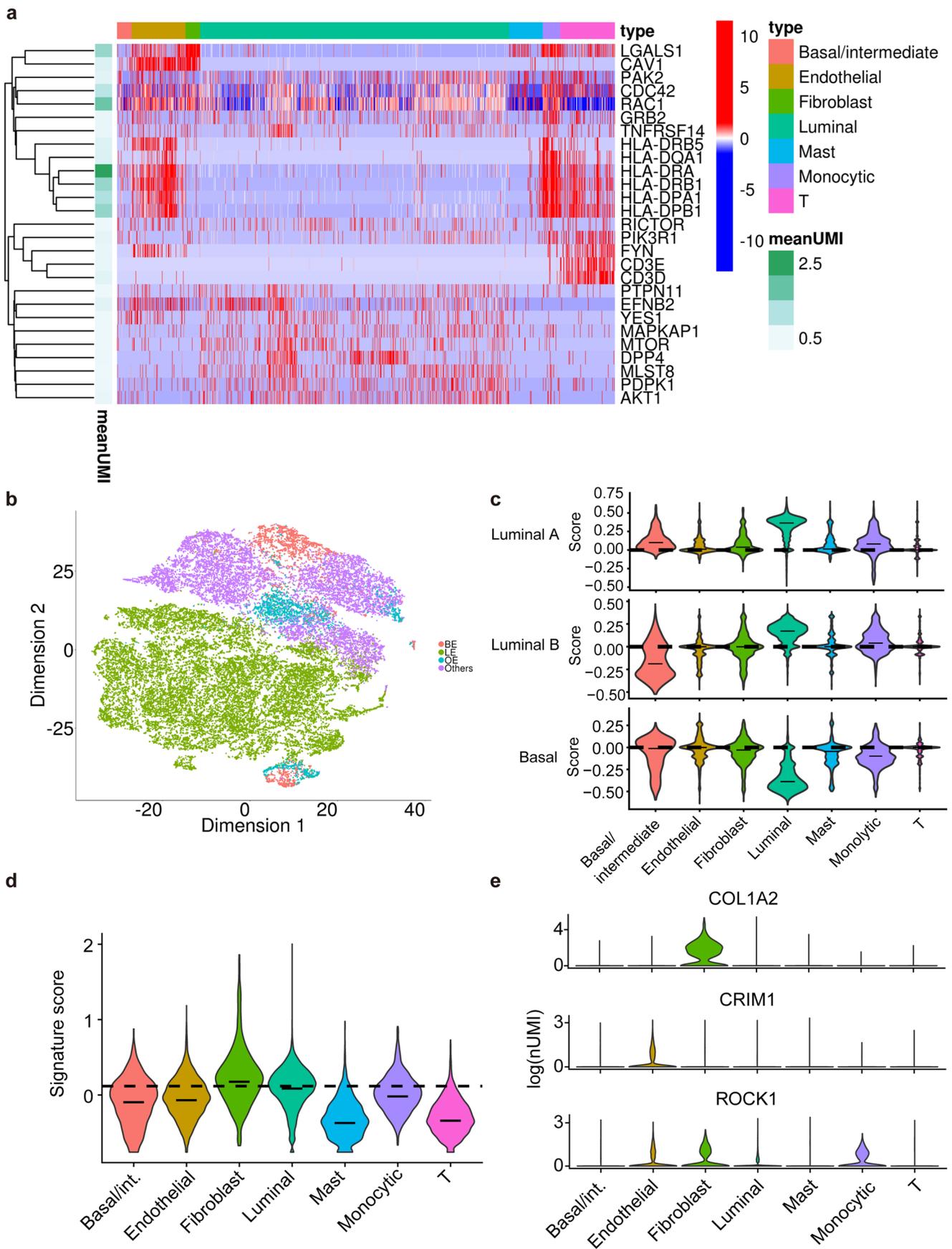
Peer review information *Nature Cell Biology* thanks Mark A. Rubin for his contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.



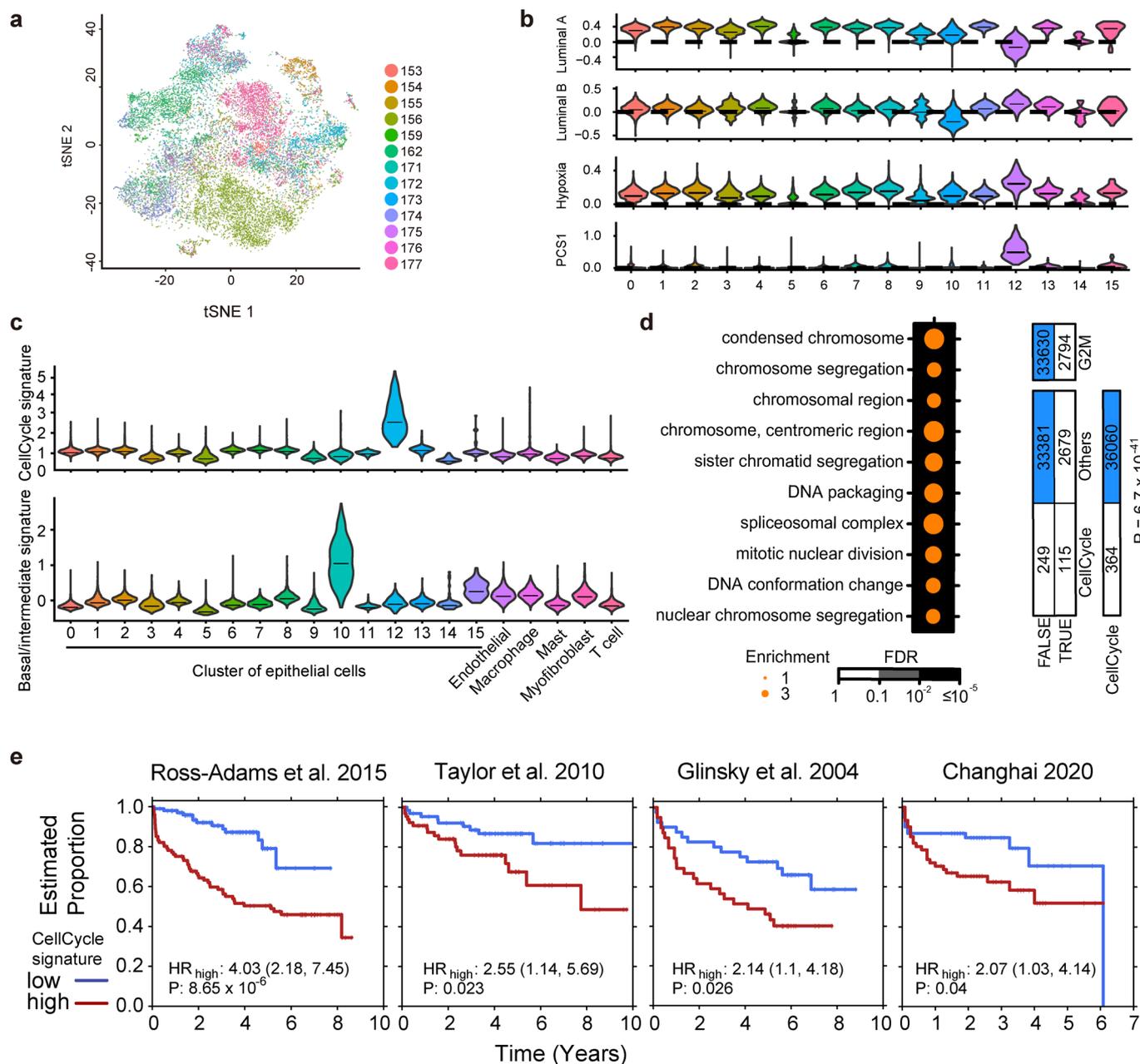
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Single cell characterization and CNV inference analysis of prostate tumours. tSNE view of 36,424 single cells, color coded by sample (**a**), cluster (**b**) and broad lineage (**c**). **d**, Inferred CNA for cells in tumour 156. **e**, the CNA score and correlation for each cell in the indicated sample. Red, CNA score > 0.04 and CNA correlation > 0.4 ; blue, CNA score < 0.04 and CNA correlation < 0.4 ; black, all remaining cells. **f**, Percentage of genes showing strong CNV (averaged CNA score > 0.1 in putative malignant cells) in our data (PCa, $n = 13$ tumor samples) and that from Tirosh et al. (Mel, $n = 14$ tumor samples). P value calculated with two-sided Mann-Whitney U test. **g**, tSNE view of 36,424 single cells, color coded by inferred cell malignancy identity. For box plots, center line represents the median and box limits represent upper and lower quartiles, and whiskers depicts 1.5 \times the interquartile range (IQR), extreme values outside of this range is shown as individual points. Statistical data for Extended Data Fig. 1f are provided in the source data.

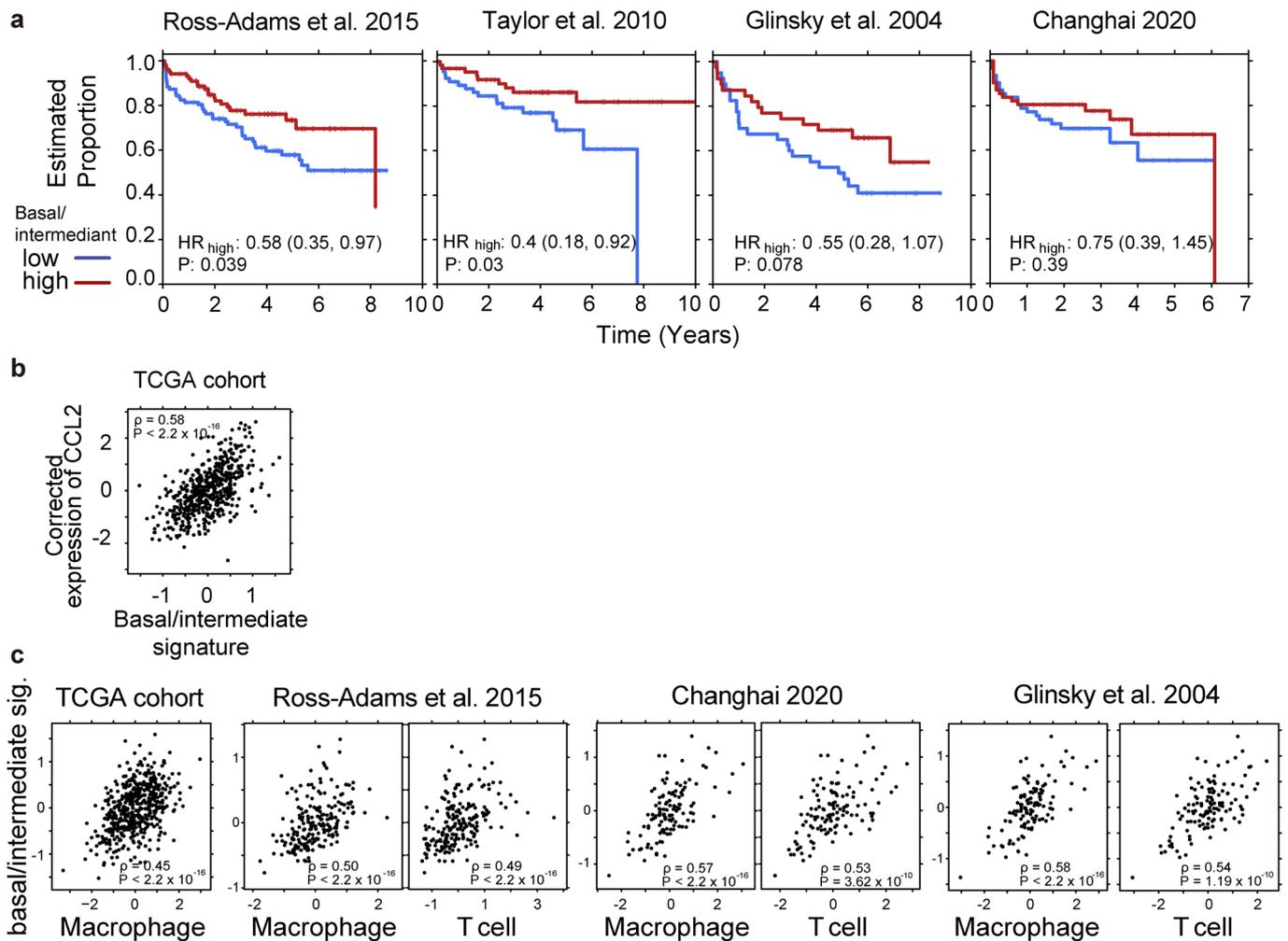


Extended Data Fig. 2 | See next page for caption.

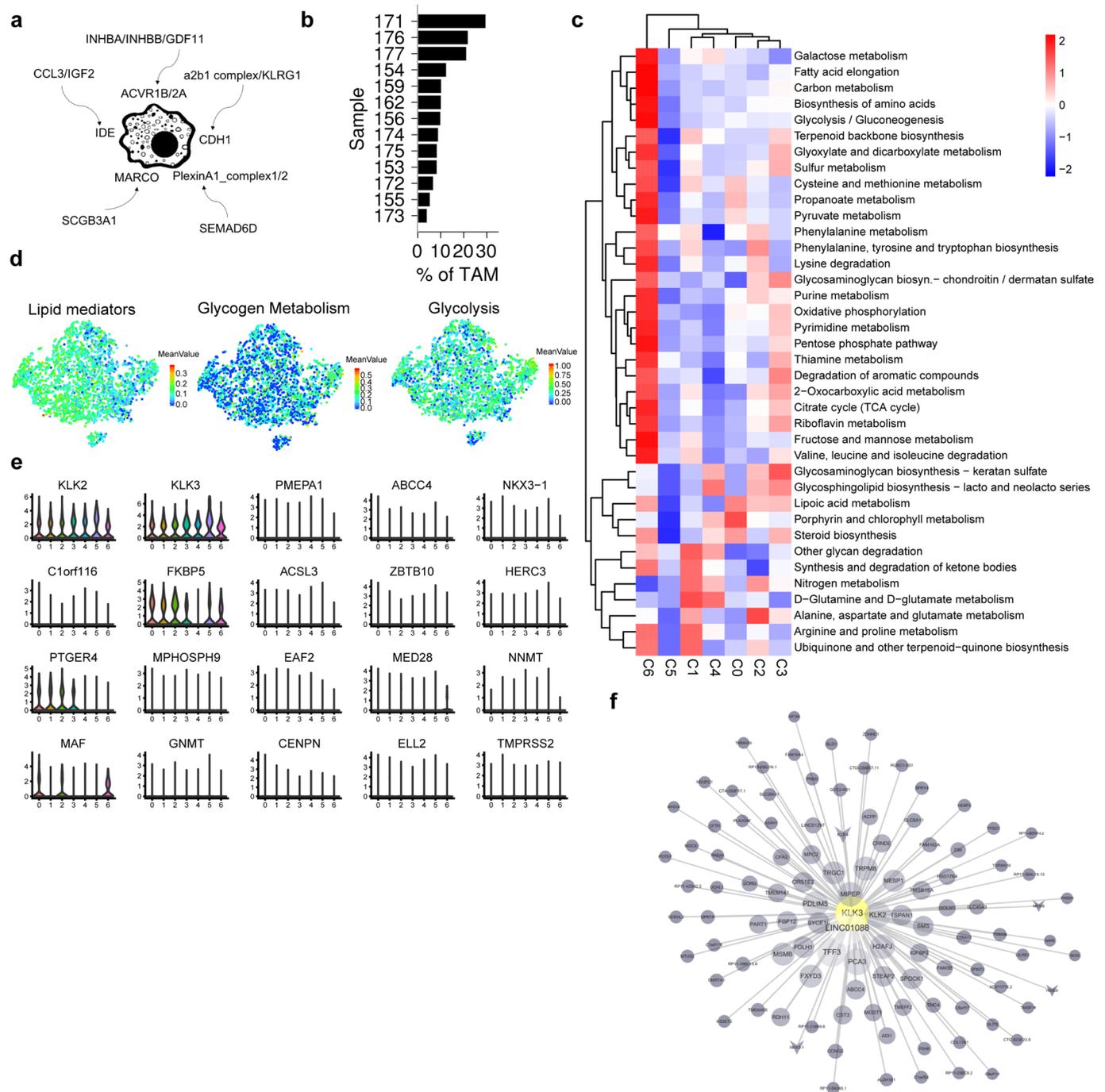
Extended Data Fig. 2 | Gene expression and signature analysis for different cell types. **a**, Heatmap shows the relative abundance of 27 genes in the 'T cell costimulation' process in each cell. The top color bar indicates cell types and the left color bar shows the mean UMI of genes. **b**, tSNE view of 36,424 single cells, color coded by epithelial subtypes. **c**, Smoothened distribution of PAM50 signature score, cells grouped by annotated cell type. **d**, Smoothened distribution of high Gleason Score (GS) related signature, cells grouped by annotated cell type. Signature score calculated as the mean of z-score for the 19 high GS related genes identified by Pressinotti et al.²⁰. **e**, Smoothened distribution of 3 high GS related individual genes; cells grouped by annotated cell type. Y-axis shows normalized UMI (nUMI) in logarithm scale.



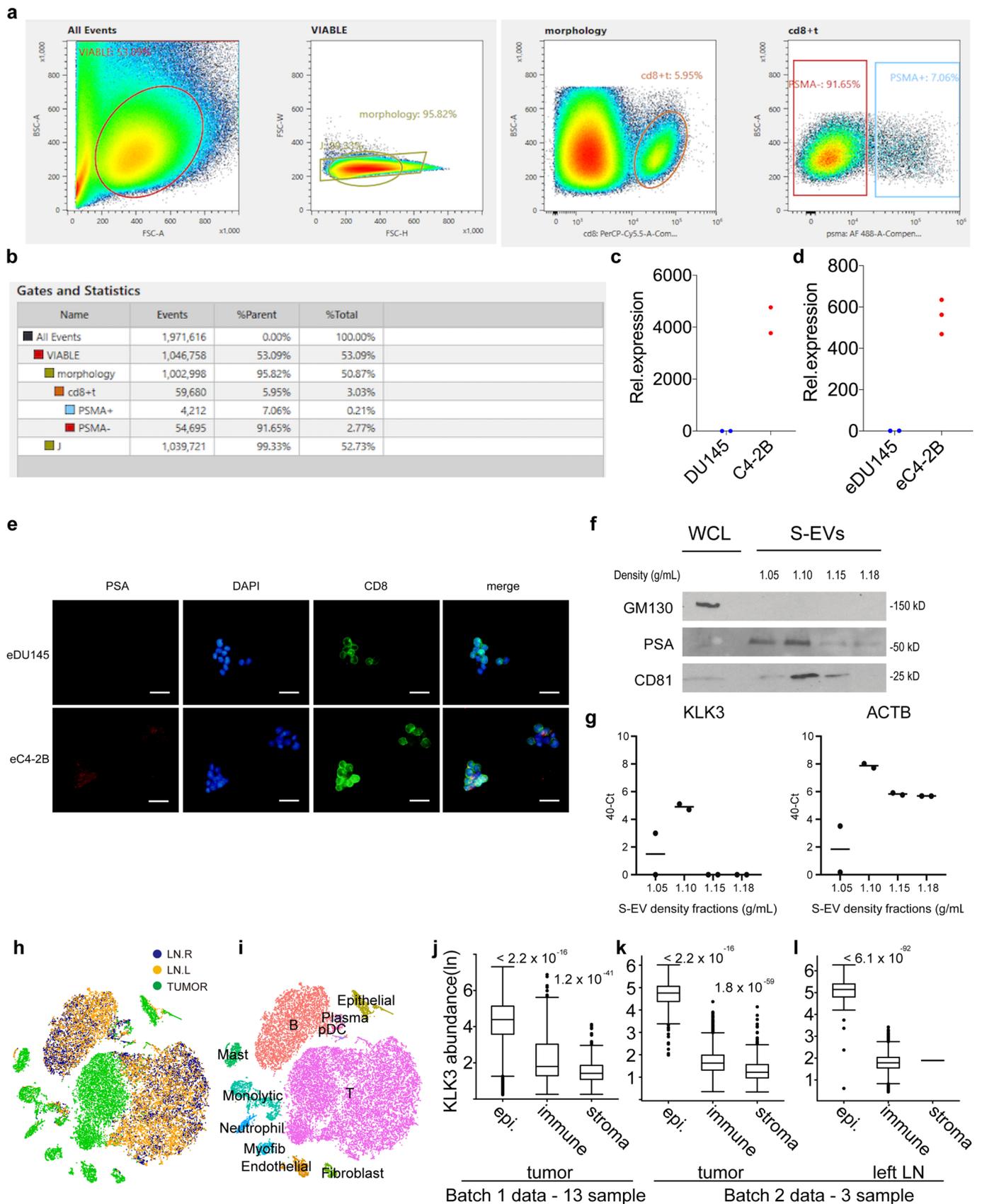
Extended Data Fig. 3 | Characterizing different epithelial cell derived programs. a, tSNE view of 23,674 epithelial cells, colour-coded by sample. **b**, Smoothened distribution for four prostate cancer related signatures in epithelial cell clusters as defined in Fig. 2a. P-values for cluster 10 compared to all other cells for Luminal A, Luminal B are $<2.2 \times 10^{-16}$; CLEs are 0.26 and 0.14, respectively; P-values for cluster 12 compared to all other cells for Luminal A, Luminal B, Hypoxia and PCS1 are $<2.2 \times 10^{-16}$; CLEs are 0.04, 0.79, 0.88 and 0.995, respectively. P values are two-sided and not adjusted for multiple comparisons. **c**, Smoothened distribution of the indicated signature score, TME and epithelial cells grouped by their assigned cell types and clusters, respectively. P-values (Mann-Whitney U test, two-sided) for cluster 12 CellCycle and cluster 10 basal/intermediate signatures compared to all the rest cells are $<2.2 \times 10^{-16}$; CLEs are 0.998 and 0.98, respectively. **d**, GO terms enriched in CellCycle subtype (left) and contingency table showing number of cells in G2/M for CellCycle subtype compared to all the other cells (right), One-sided P value calculated with Fisher's exact test. OR = 5.7. **e**, Comparison of BCR-free rate between the high and low groups stratified using CellCycle signature across multiple datasets. P-values are calculated using Cox proportional hazard model (CoxPH) and not adjusted for multiple comparisons. Numbers in brackets show 95% CI for PH.



Extended Data Fig. 4 | Characterizing the basal/intermediate signature. **a**, Comparison of BCR-free rate between the high and low groups stratified using Basal/Intermediate signature in the indicated datasets. P-values are calculated using Cox proportional hazard model (CoxPH) and not adjusted for multiple comparisons. Numbers in brackets show 95% confidence interval (CI) for hazard ratio (HR). **b**, Correlation between tumour purity corrected CCL2 expression and basal/intermediate signature in TCGA. **c**, Correlation between macrophage, T cell and basal/intermediate signature across multiple datasets. Two-sided P values calculated for Spearman's rank correlation and not adjusted for multiple comparisons.

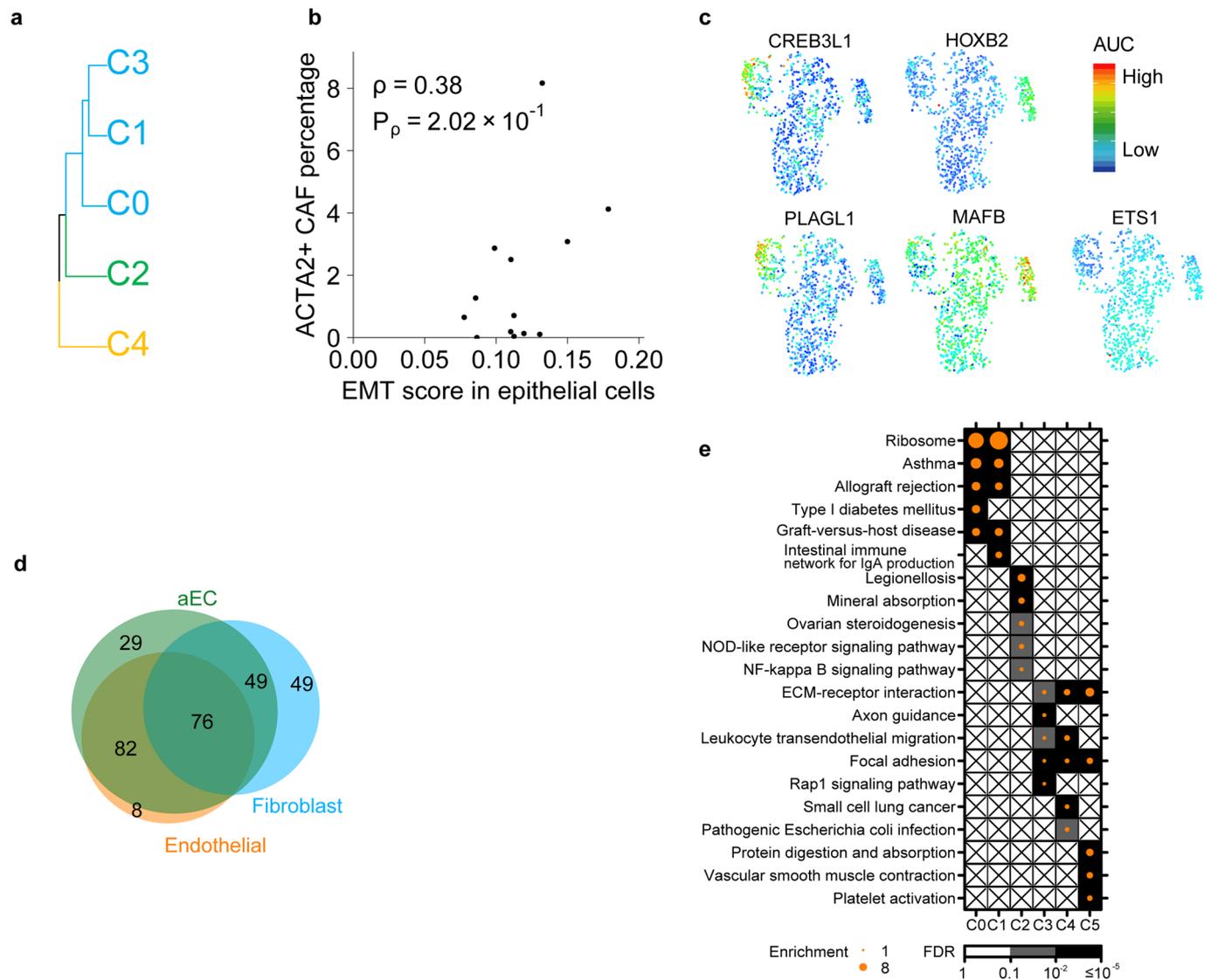


Extended Data Fig. 5 | Single cell transcriptome reveals immune cell heterogeneity. **a**, Schematics for all C6 specific incoming signals. **b**, Percentage of C6 TAM cells in each sample. **c**, Differentially activated metabolism-related pathways. **d**, tSNE view of 3,116 T cells, color coded by the average expression of lipid mediator, glycogen metabolism and glycolysis genes. **e**, Smoothed distribution of AR signature gene abundance, cells grouped by cluster. **f**, Genes in module 61. Line length and circle size corresponds to expression correlation between *KLK3* and the indicated gene. Statistical data relevant to Extended Data Fig. 5b are provided in the source data.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Analysis of *KLK3* expression in T cells. Flow cytometry sorting strategy (**a**) and statistics (**b**) of PSMA+ CD8 T cells. Relative expression of *KLK3* in the indicated prostate cancer cell lines (**c**) and EVs (**d**) derived from them. **e**, Examination of *KLK3* protein (PSA, red) in T cells (CD8+, green) after co-culture with EVs derived from the indicated prostate cancer cell lines by fluorescence microscopy. Scale bars, 10 μ m. Data show representative results of two repeats. **f**, Immunoblot for small EV-enriched (CD81) and depleted proteins (GM130) in C4-2B EVs. EVs collected by differential centrifugation followed by density gradient purification. 1 μ g of protein was loaded for whole cell lysate (WCL). Same volume was loaded for S-EV fractions. **g**, rtPCR for the indicated probes in C4-2B small EVs collected by differential centrifugation and floated in different density fractions. tSNE view showing cells from the high-risk prostate cancer patient (SC001H), color coded by the tissue source (**h**) or cell type (**i**). Distribution of *KLK3* expression in different cell types for samples from Batch 1 data (**j**) (n = 22,667, 7,495 and 629 cells for epi., immune and stroma groups, respectively), tumour tissue (**k**) (n = 700, 4,038 and 594 cells for epi., immune and stroma groups, respectively) and left LN (**l**) (n = 153, 1,954 and 2 cells for epi., immune and stroma groups, respectively) from Batch 2. Two-sided P values are calculated using Mann-Whitney U test. Y-axis of **j-l** represents natural logarithm scale. For box plots, center line represents the median and box limits represent upper and lower quartiles, and whiskers depicts 1.5 \times the IQR, extreme values outside of this range is shown as individual points. Statistical data relevant to Extended Data Fig. 6c-d, g are provided in the source data. Unprocessed Western Blots relevant to Extended Data Fig. 6f are provided in the source data.



Extended Data Fig. 7 | Single cell characterization of stromal components. **a**, Hierarchical clustering using all subtype marker genes, color coded by subtypes. **b**, Correlation between EMT score in epithelial cells and percentage of ACTA2+ CAF. Two-sided P values calculated for Spearman's rank correlation and not adjusted for multiple comparisons. **c**, tSNE view CAF cells, color coded by activation level of the indicated TFs (AUC). **d**, Overlap of the unique incoming cell communication pairs from epithelial cells to fibroblast, regular EC and aEC. **e**, Top 5 most enriched pathways for each cluster.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

For GSE70770, data was downloaded with R package GEOquery (v2.50.5). MSKCC data and BCR information was downloaded with R package cgdscr (v1.2.10). Other clinical information was downloaded from GSE2103433 with GEOquery (v2.50.5). Data for 499 prostate cancer samples were downloaded using R package TCGA2STAT (v1.2).

Data analysis

All R packages used are available online as described in the method section. Customized code for data analyzing and plotting can be found on GitHub: <https://github.com/chensujun/scrna>.
Software used: R statistical environment (v3.5.1); Umi-tools (v1.0.0); Cellranger (v2.1.0); STAR (v2.5.2b); FlowJo (v10)
R packages: Survival (v2.41-3); BPG (v5.9.8); Seurat (v2.3.4); scran (v1.10.2); QuSAGE (2.16.1); InferCNV (v0.8.2); Monocle2; Monocle3; GEOquery (v2.50.5); cgdscr (v1.2.10); TCGA2STAT (v1.2); canprot (v0.1.2)
Python package: pySCENIC (v0.9.5); CellphoneDB (v1.1.0)
Website: CellphoneDB (v1.1.0)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Data were deposited in Gene Expression Omnibus (GEO) under the accession GSE141445 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE141445>), the Genome Sequence Archive for Human (GSA-Human) under the accession HRA000312 (<https://bigd.big.ac.cn/gsa-human/browse/HRA000312>), and can be accessed at www.pradcellatlas.com.

For survival analysis, bulk RNA-seq data from the following study were used: TCGA (333 samples, <http://firebrowse.org/?cohort=PRAD>), Taylor et al. (131 samples, MSKCC, <https://doi.org/10.1016/j.ccr.2010.05.026>), Ross-Adams et al. (294 samples, GSE70770), Glinsky et al. (79 samples, <https://doi.org/10.1172/JCI20032/>), and Changhai 2020. (136 samples, www.cpgea.com).

For gene expression analysis in T cells, scRNA-seq data from the following GEO accessions were used: GSE99254 (NSCLC), GSE108989 (CRC), GSE98638 (HCC) and GSE103322 (HNSCC).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For single cell analysis, the thousands of individual profiles obtained from each tissue sample provide sufficient power. There is thus no formal sample size calculation. We do seek proper representation for prostate cancer and required at least three patients in groups classified according to major clinical features: e.g. patients with Gleason score no greater/less than 7, patients with T2/T3/T4 stage, patients with/without intraductal carcinoma and patients with/without lymphatic/bone metastasis. All patients meeting the clinical criteria with proper consent were included. For functional studies, statistical significance can be derived from the results, representing sufficient power using the given sample size.
Data exclusions	The single cell expression profiles obtained were subjected to quality control and low quality cells were excluded from analysis. Exclusion criteria were established in previous work or according to the manufacturer's instruction and were described in detail in Methods section: We removed cells that have higher than 20% expression on mitochondrion genes, lower than 200 or higher than 5218 total genes expressed, or lower than 56 housekeeping genes expressed.
Replication	For human biological specimens, entire samples were used for processing, hence, no additional replicates were available. For functional studies, we conducted at least 2 independent experiments. We performed at least 3 independent experiments where statistics is needed. Experimental observations were reproducible and successful among the independent experiments.
Randomization	The work required no randomization: Human specimen were allocated into groups according to disease status (benign, prostate adenocarcinoma, CRPC etc.). For functional studies, specimens were subjected to different treatment and were allocated to different groups accordingly.
Blinding	Blinding was not relevant to our study: Patient sample allocation was predetermined by their known disease status. Specimens in functional experiments were grouped according to the treatments they were subjected to.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input type="checkbox"/>	<input checked="" type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input type="checkbox"/>	<input checked="" type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used	Rabbit monoclonal anti-CD8 (Cat# ab237709, 1:100, Abcam); Mouse monoclonal anti-PSA (Cat# ab140337, 1:100, Abcam); Goat anti-mouse IgG (Cat# 4412; 1:1000, Invitrogen); Goat anti-rabbit IgG (Cat# 4409; 1:1000, Invitrogen); PE Conjugate anti-CD3 (Cat# 300307, 1:20, BioLegend); PSMA (ab187570, 1:150, Abcam); CD8 (ab157306, 1:150, Abcam); FITC anti-human CD31 (Cat. #303104; 1:20, Biolegend);
-----------------	---

CD90 (Cat. #328124; 1:20, Biolegend);
GM130 (ab52649 [EP892Y], 1:10,000, Abcam);
CD81 (ab79559 [M38], 1:10,000, Abcam);
PSA (sc-7638 [C-19]; 1:5,000, Cell Signaling)

Validation

Rabbit monoclonal anti-CD8 (Cat# ab237709, 1:100, Abcam), Mouse monoclonal anti-PSA (Cat# ab140337, 1:100, Abcam), Goat anti-mouse IgG (Cat# 4412; 1:1000, Invitrogen) and Goat anti-rabbit IgG (Cat# 4409; 1:1000, Invitrogen) are validated by vendor for IF;
PE Conjugate anti-CD3 (Cat# 300307, 1:20, BioLegend) is validated by vendor for flow;
PSMA (Alexa Fluor® 488) (ab187570, 1:150, Abcam) is validated by vendor for flow;
CD8 (ab157306, 1:150, Abcam) is validated by vendor for flow, IP;
FITC anti-human CD31 (Cat. #303104; 1:20, Biolegend) and CD90 (Cat. #328124; 1:20, Biolegend) is validated by vendor for flow;
GM130 (ab52649 [EP892Y], 1:10,000, Abcam) is validated by vendor for WB;
CD81 (ab79559 [M38], 1:10,000, Abcam) has been validated in publication PMID:32341353 for WB;
PSA (sc-7638 [C-19]; 1:5,000, Cell Signaling) has been validated in publication PMID:25864123 for WB

Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)

DU145 (Stem Cell Bank, CAS, TCHU222); PC3 (Stem Cell Bank, CAS, TCHU158); C4-2B was originally provided by Dr. Leland Chung (Cedars-Sinai), who established the cell line (PMID: 8168083).

Authentication

DU145 has been authenticated by STR analysis at GENEWIZ Inc. PC3 cells have been authenticated by STR analysis, as declared by Stem Cell Bank, Chinese Academy of Sciences (CAS). C4-2B cell line has been authenticated by STR analysis at Shanghai Integrated Biotech Solutions Co, Ltd.

Mycoplasma contamination

DU145 and PC3 cells have been tested negative for Mycoplasma contamination, as declared by Stem Cell Bank, Chinese Academy of Sciences (CAS). C4-2B cells have been tested negative for Mycoplasma contamination using MycAway™ -Color One-Step Mycoplasma Detection Kit (40611ES25, Yeasen).

Commonly misidentified lines
(See [ICLAC](#) register)

No commonly misidentified cell lines were used in the study.

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics

All patients involved in this study were male aged 55-81 with prostate or bladder cancer undergone radical prostatectomy or radical cystectomy.
In experiments using healthy donors, population characteristics are controlled by allocating specimens from the same patient into treatment and control groups, and are thus not relevant covariates.

Recruitment

Patients who were consented and met the clinical criteria were included. The healthy donors for collection of peripheral blood are consecutive healthy participants with consent. There are no other criteria for selecting patients to this study. There are no self-selection bias or other biases in recruitment.

Ethics oversight

The Institutional Review Board (IRB) of Shanghai Changhai Hospital approved the tissue acquisition procedures and experimental protocols

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

Prostate cancer tissues were minced and digested with collagenase (Sangon Biotech) and DNase (Solarbio). The cell suspension was filtered through 100µm cell strainers (BD Falcon) and centrifuged at 1200rpm for 5min. The precipitate was resuspended in FACS buffer (1% fetal bovine serum/1640 medium) and filtered through 70µm cell strainers (BD Falcon) and washed once with FACS buffer. The resuspended cells were then stained for PSMA (Alexa Fluor® 488) (ab187570, 1:150, Abcam) and CD8 (PerCP/

Cy5.5[®]) (ab157306, 1:150, Abcam). After washed for three times with FACS buffer, the prepared cell suspension was filtered through 40µm cell strainers (BD Falcon) and processed by Sony cell sorter SH800S. PSMA+CD8+ cells and PSMA-CD8+ cells were collected respectively in 15ml centrifuge tubes.

To assay the expression of CD31 and CD90, 10⁶ cells were incubated at room temperature for 30 min with 5 µl of FITC anti-human CD31 (Cat. #303104; Biolegend) and 5 µl of PE/cy7 anti-human CD90 (Cat. #328124; Biolegend). Taking 5 µl of non-specific isotype-matched control IgG incubated cells as control. Unbound antibodies were removed by washing the cells twice in PBS buffer.

Instrument

flow cytometer(FC500; Beckman).

Software

FlowJo V10 was used for data analysis.

Cell population abundance

Cells collected for following experiments were both CD31 and CD90 positive via flow cytometer sorting.

Gating strategy

CD31 and CD90 were used to identify aECs from non-aECs.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.