**BMC Cancer**

**Open Access**

# Transcriptome sequencing and multi-plex imaging of prostate cancer microenvironment reveals a dominant role for monocytic cells in progression

Stefano Mangiola[1,2,3,4], Patrick McCoy[2,3], Martin Modrak[5], Fernando Souza-Fonseca-Guimaraes[6], Daniel Blashki[7], Ryan Stuchbery[3], Simon P. Keam[8,9], Michael Kerger[3], Ken Chow[2,3], Chayanica Nasa[10], Melanie Le Page[10], Natalie Lister[11,12], Simon Monard[10], Justin Peters[13], Phil Dundee[13], Scott G. Williams[8,9], Anthony J. Costello[2,3], Paul J. Neeson[8,9], Bhupinder Pal[14], Nicholas D. Huntington[11,12], Niall M. Corcoran[2,3,15†], Anthony T. Papenfuss[1,4,8,9,16*†] and Christopher M. Hovens[2,3†]

## Abstract

**Background:** Prostate cancer is caused by genomic aberrations in normal epithelial cells, however clinical translation of findings from analyses of cancer cells alone has been very limited. A deeper understanding of the tumour microenvironment is needed to identify the key drivers of disease progression and reveal novel therapeutic opportunities.

**Results:** In this study, the experimental enrichment of selected cell-types, the development of a Bayesian inference model for continuous differential transcript abundance, and multiplex immunohistochemistry permitted us to define the transcriptional landscape of the prostate cancer microenvironment along the disease progression axis. An important role of monocytes and macrophages in prostate cancer progression and disease recurrence was uncovered, supported by both transcriptional landscape findings and by differential tissue composition analyses. These findings were corroborated and validated by spatial analyses at the single-cell level using multiplex immunohistochemistry.

**Conclusions:** This study advances our knowledge concerning the role of monocyte-derived recruitment in primary prostate cancer, and supports their key role in disease progression, patient survival and prostate microenvironment immune modulation.

**Keywords:** Prostate cancer, Transcriptomics, FACS, Immunohistochemistry, Deconvolution, Bayes, Differential gene expression, CAPRA-S, Microenvironment, Epithelial, Myeloid, Macrophages, Cholesterol, PDL1

* Correspondence: papenfuss@wehi.edu.au
†Niall M. Corcoran, Anthony T. Papenfuss and Christopher M. Hovens contributed equally to this work.
[1]Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia
[4]Department of Medical Biology, University of Melbourne, Melbourne, Victoria, Australia
Full list of author information is available at the end of the article

## Background

Prostate cancer is the second most commonly diagnosed cancer in men globally [1]. Although most cancers follow an indolent clinical course, an unpredictable 10–15% of tumours progress to metastases and death. The inability to discern progressive disease at an early stage results in substantial overtreatment of localised disease, leading to a significant clinical cost to the patient and economic cost to the healthcare system. Selecting patients for treatment is usually reliant on a small number of well-established clinical and pathological factors, such as tumour grade, prostate serum antigen (PSA) level and clinical stage [2], the development of metastases [3] and prostate cancer-specific death [4]. Although comprehensive molecular analyses have linked clinical outcomes with rates of genomic alterations [5, 6], such as somatic changes in copy number, nucleotide sequence and methylation, it is yet to be demonstrated that such measures are able to consistently outperform standard clinico-pathological risk scoring across a broad range of grades and stages. Despite many years of tumour evolution characterization, it still remains unclear what mechanisms drive prostate cancer progression in most patients [7]. It is believed that reciprocal interactions between malignant epithelium and surrounding non-cancerous cells within the tumour microenvironment are responsible for driving disease progression [8, 9].

Selected targets in the prostate tumour microenvironment have been extensively studied through in vitro and in vivo experiments, such as migration assays [10] and xenograft mouse models [11] respectively. More recently, several studies that integrated fluorescence-activated cell sorting or laser microdissection with RNA sequencing increased the gene and sample throughput while maintaining a degree of resolution of the tissue heterogeneity [8, 12, 13]. Additionally, the use of spatial transcriptomics has identified gradients of benign-cell gene transcription around tumour foci [14]. However, these studies mainly focused on the process of epithelial to mesenchymal transition [12, 13] or were limited to the overall stromal contribution to disease progression [8]. An integrative investigation of immune, stromal and cancer cell transcriptional changes associated with clinical risk is still lacking.

In this study, we applied an optimised protocol for combined cell-type enrichment and ultra-low-input RNA sequencing, which allowed the probing of four key cell types across 13 fresh prostate tissue spanning a broad clinical disease spectrum. Motivated by the pseudo-continuous properties of the CAPRA-S risk score, we developed a novel statistical inference model for differential transcription analyses on continuous covariates, TABI (Transcriptional Analysis through Bayesian Inference). Our inference model estimated changes in transcription without a priori patient risk stratification and robustly mapped transcriptional change events to cancer risk states. Among the list of significant genes for the four cell types coding for cell-surface and secreted proteins, we identified several hallmarks of prostate cancer. These hallmarks included a dominant signal for monocyte-derived cell recruitment. We tested this with tissue deconvolution on the extensive Cancer Genome Atlas (TCGA) cohort and multiplex immunohistochemistry on an independent patient cohort. The latter single-cell resolution spatial analysis revealed the relationship between macrophages and epithelial and T cells with progression. For prostate cancer, prioritising targets for immunotherapies is far from settled in the literature, with monocytic cells being an under-represented player [15, 16]. In this scenario, the parallel lines of evidence we provide from unbiased analyses contribute to shaping future research directions.

## Methods

### Tissue sampling and processing

Following the prostatectomy of 13 patients, a four-millimetre tissue core was collected from the prostate tumour site, conditional to histopathological verification [17, 18]. The patient cohort ranged from 52 to 78 years of age and from CAPRA-S risk score of 0 (attributed to benign tissue samples, harvested from a site far from a low grade, low volume cancer) to 7 (Supplementary file 4), If not otherwise specified, all procedures were carried out at 4 °C. Tissue blocks were washed in phosphate-buffered saline (PBS) solution for 2 min and minced for 2 min with a scalpel. Homogenised tissue was added to a solution (total volume of 7 ml) composed by of 1 mg/ml collagenase IV (Worthington Biochemical Corp, USA), 0.02 mg/ml DNase 1 (New England Biolabs, USA), 0.2 mg/ml dispase (Merck, USA). The homogenised tissue was serially digested in the shaker incubator at 37 °C at 180 rpm (4 g), through three steps of 5, 10 and 10 min of duration. The final 3 min were dedicated to sedimentation at 0 rpm. After each digestion step, the supernatant was aspirated and filtered through a 70 μm strainer into a pre-chilled tube, diluting the solution with 15 ml of Dulbecco's PBS containing 2% Bovine serum (dPBS-serum) to quench the enzymatic reaction. The resulting cumulative solution was then centrifuged at 300g for five minutes, with the supernatant collected and the cell pellet resuspended into 1 ml 2% PBS-serum before labelling (Fig. S1).

### Antibody labelling, flow cytometry and cell storage

The cell preparation was labelled with the following antibodies: CD3-BV711 (Becton Dickinson San Jose Ca), EpCAM-PE (BD Biosciences, USA), CD31-APC (BD Biosciences, USA), CD90-PerCP-Cy5.5 (Becton

Dickinson San Jose Ca), CD45 APC-CY7, and CD16 Pacific Blue (BD Biosciences, USA). All antibodies were used at concentrations according to manufacturers recommendations and incubated for 30 mins at 4 °C. Following labelling, the cells were diluted to 5 ml and centrifuged at 300 g for 5 min. The supernatant was removed, and the cell pellet was resuspended in dPBS-serum. The viability die (7AAD) was added to the suspension to a final concentration of 5 μg/ml. Epithelial, fibroblasts, myeloid and T cells were sorted using a fluorescence-activated cell sorting Aria III cell sorter (Becton Dickinson San Jose, Ca). The cell sorting strategy utilised a robust 3 stage design: (i) a series of gates based on forward and side scatter to exclude debris, cell clumps and doublets. (ii) a gate to exclude all dead cells and (iii) a combination of the fluorescent antibodies to allow purification of the above cell types. The four cell types were identified as follows: T Cells: FSC and SSC lo, PI negative, EpCAM and CD31 negative, CD3 and CD45 positive. Epithelial cells: FSC and SSC high, PI negative, CD31 and CD90 negative and EpCAM positive. Myeloid cells: FSC and SSC hi and medium, PI negative, CD31 and EpCAM negative and CD16 positive. Fibroblasts: FSC and SSC hi, PI negative, EpCAM and CD31 negative, CD90 positive. The four purified populations were sorted directly into 1.5 ml conical tubes and stored on dry ice immediately after collection before permanent storage at − 80 °C.

### RNA extraction, library preparation and RNA sequencing

RNA extraction was performed in two batches (comprising 6 and 7 patients, for a total of 24 and 28 samples, respectively) on consecutive days. In order to eliminate time-dependent methodological biases, the two patient batches included a balanced distribution of Gleason score (means 2.00 and 2.71, standard deviations 2.50, 1.86; Supplementary file 4) and days elapsed from tissue processing (means 197 and 222, standard deviations 46.3 and 71.9; Supplementary file 4). The RNA extraction was performed using the miRNeasy Micro Kit (Qiagen; Cat #217084), according to the manufacturer's protocol. Briefly, cell pellets were lysed with QIAzol lysis reagent, treated with chloroform, and centrifugation carried out to separate the aqueous phase. Total RNA was precipitated from the aqueous phase using absolute ethanol, filtered through the MinElute spin column and treated with DNase I to remove genomic DNA. The RNA bound columns were washed with the buffers RWT and RPE before eluting the total RNA with 14 μl of RNase-free water. RNA estimation was carried out using Tapestation (Agilent).

According to the manufacturer's protocol, transcriptome sequencing on low input total RNA samples (up to 10 ng) was carried out using SMART-Seq v4 Ultra Low

Input RNA Kit (Clontech). The first-strand cDNA synthesis utilised 3′ SMART-Seq CDS Primer II-A. The SMART-Seq v4 Oligonucleotide together with the cDNA amplification was carried out on Thermocycler using PCR Primer II-A and PCR conditions: 95 °C for 1 min, 12 cycles of 98 °C 10 s, 65 °C 30 s and 68 °C 3 min; 72 °C for 10 min and 4 °C until completion. The PCR-amplified cDNA was purified using AMPure XP beads and processed with the Nextera XT DNA Library Preparation Kits (Illumina, Cat. # FC-131-1024 and FC- 131-1096) as per the protocol provided by the manufacturer.

Sequencing library preparation (10–100 ng) was carried out using Truseq RNA Sample Preparation Kit v2. The poly-A containing mRNA was purified using oligo-dT bound magnetic beads followed by fragmentation. The first-strand cDNA synthesis utilised random primers, and second-strand cDNA synthesis was carried out using DNA Polymerase I. The cDNA fragments then underwent an end-repair process, adding a single 'A' base and ligation of the RNA adapters. The adaptor-ligated cDNA samples were bead-purified and enriched with PCR (15 cycles) to generate the final RNAseq library.

The SMART-Seq v4 RNA and Truseq RNA libraries were sequenced on an Illumina Nextseq 500 to generate 15–20 million 75 base pairs paired-end reads for each sample. The batch effect due to sequencing runs was minimised by pooling all 52 libraries and carrying out three sequential runs on a Nextseq500 sequencer.

### Sequencing data quality control, mapping and read counting

The quality of the sequenced reads for each sample was checked using the Fastqc [19]. Reads were trimmed for custom Nextera Illumina adapters; low-quality fragments and short reads were filtered out from the pools using BBDuk (jgi.doe.gov) according to default settings. All remaining reads were aligned to the reference genome hg38 using the STAR aligner [20] with default settings. The quality control on the alignment was performed with RNA-SeQC [21]. For each sample, the gene transcription abundance was quantified in terms of nucleotide reads per gene (read-count) using FeatureCounts [22] with the following settings: isPairedEnd = T, requireBothEndsMapped = T, checkFragLength = F, useMetaFeatures = T. All sequenced reads that did not align to the reference human genome were assigned to bacterial and viral reference genomes using Kraken [23] with default settings.

### Statistical inference of differential gene transcript-abundance

Changes of transcriptional levels along CAPRA-S risk score [24] were estimated independently for each cell

Mangiola *et al. BMC Cancer* (2021) 21:846

Page 4 of 18

type (epithelial, fibroblast, myeloid and T cell). The CAPRA-S risk score is a combination of (i) concentration of blood prostate serum antigen (PSA); (ii) presence of surgical margin (SM); (iii) Gleason score; (iv) presence of seminal vesicle invasion (SVI); (v) the extent of extracapsular extension (ECE); and (vi) lymph node involvement. The RNA extraction batch was used as a further covariate. Due to the absence of publicly available models for non-linear monotonic regression along a continuous covariate, a new Bayesian inference model was implemented. This model is based on the simplified Richard's curve [25] (Eq.1) but re-parameterised to improve numerical stability (Eq. 2). In particular, the standard parametrisation suffers from non-determinability issues if the slope is close to zero; furthermore, in the case of an exponential-like trend, the upper plateau is not supported by data and tends to infinity.

$$GL\left(X, \alpha, \beta, \kappa\right) = \frac{k}{1 + e^{-(\alpha + X\beta)}} \qquad (1)$$

$$GLA\left(X, y_0, \beta, \eta\right) = \frac{y_0\left(1 + e^{\eta\beta_1}\right)}{1 + e^{\eta\beta_1 - X\beta}} \qquad (2)$$

The new parameter $y_0$ represents the intercept on the y axis, $\eta$ represents the point of inflection on the x-axis, $\beta$ represents the matrix of coefficients (i.e. slope coefficients, without the intercept term), $\beta_1$ represents the coefficient of interest (i.e. main slope), and k the upper plateau of the generalised sigmoid function.

Bayesian inference was used to infer the values of all parameters of the model, with TABI (GitHub: stemangiola/TABI@v0.1.3). The probabilistic framework Stan [26] was used to encode the joint probability function of the model (Eq. 3). We partitioned the transcriptomic dataset into blocks of 5000 genes to decrease the analysis run-time. This Bayes model is based on a negative binomial distribution (parameterised as mean and overdispersion). In order to account for various sequencing depths across samples, a sample-wise normalisation parameter was added to the negative binomial expected value. The slope parameter for the main covariate ($\beta_1$) was subject to a regularised horseshoe prior [27] to increase the robustness of the inference of transcription changes and help anchor data from different samples for normalisation. The role of this prior is to impose a sparsity assumption on the gene-wise transcriptional changes; that is, most genes are not Tdifferentially transcribed. The overall distribution of the gene intercepts follows a gamma probability function. The following joint probability density defines the statistical model.

$$P(\gamma)P(\delta)P(\sigma)P(\eta)P(\xi)P(\beta|\xi)$$
$$\left(\prod_{r=2}^{R} P(\beta_r|\sigma)\right)\left(\prod_{g=1}^{G} P\left(y_{og}|\gamma', \gamma''\right)\right) \qquad (3)$$
$$\left(\prod_{g=1}^{G}\prod_{s=1}^{S} P\left(Y_{g,s}|\hat{Y}, \delta, \omega\right)\right)$$

$$Y_{t,g} \sim NB\left(exp(\delta_t)\hat{Y}_{t,g}, \omega\right) \qquad (4)$$

$$\hat{Y}_{t,g} = GLA\left(X_t, y_{0_g}, \beta_g, \eta_g\right) \qquad (5)$$

$$\beta_{g,1} \sim RegHorseshoe(\ldots) \qquad (6)$$

$$\beta_{g,k} \sim N(0, \sigma_k); k > 1$$
$$\sigma_k \sim HalfN(0, 1) \qquad (7)$$

$$y_{0_g} \sim Gamma(\gamma_1 + 1, \gamma_2)$$
$$\gamma_i \sim Exponential(1) \qquad (8)$$
$$\omega \sim Gamma(1.02, 2)$$

$$\eta_g \sim N(0, 1)$$
$$\delta_t \sim N(0, 1); \sum \delta_t \sim N(0, 0.001 * T) \qquad (9)$$

Y represents raw transcript abundance, $\underline{\hat{Y}}$ represents the expected values of transcript abundance, and X represents the design matrix (with no intercept term and scaled covariates). The regression function also includes β, which represents the gene-wise matrix of factors (i.e. slopes excluding the intercept term), $\underline{y}$ and η, which represent the gene-wise y-intercept and the inflection point of the generalised reparameterised sigmoid function (Eq. 2). γ represents the hyperparameters of $\underline{y}$. Other parameters of the negative binomial function are δ, which represents the normalisation factors, and ω, which represents overdispersion. The regularising prior (for imposing the sparsity assumption) over the covariate of interest $\beta_1$ (first column of β) is defined by the hyperparameter list ξ [27] (i.e. nu_local = 1; nu_global = 1; par_ratio = 0.8; slab_df = 4; slab_scale = 0.5), while σ represents the standard deviations of the other factors (in our case only the batch). The algorithm multidimensional scaling [28] was used to map the data in two-dimensional space.

## Gene annotation
Each gene (g) was considered well fitted by the model if it had read counts outside the 95th percentile of the generated quantities for three or fewer samples (according to posterior predictive checks standards [29]). Among the well-fitted genes, those for which the 0.95 credible intervals of the posterior distribution of the factor of interest $\beta_{1g}$ did not include the value 0 were labelled as differentially transcribed. The credible interval is a numerical range within which an unobserved

parameter value falls within a certain probability. As distinct from standard practices for frequentist models operating on confidence intervals and *p*-values, for this study, the credible interval probability threshold was not altered for multiple hypothesis testing, consistently with standard practices in Bayesian statistics [30].

In order to interpret the inflection points over the CAPRA-S risk score (i.e. the point of the maximum slope; at what stage of the disease a transcriptional change happens) covariate in a biologically meaningful way, the inflection point was adjusted to the log-scale. Considering that the lower plateau of our generalised sigmoid function was set to 0 (to limit the number of parameters needed to model it), the inflection point of the logarithm-transformed function is not defined. Therefore, we calculated the inflection point (*X*) of the log sigmoid forcing a plateau at 1 (i.e. log (0) = 1; Eq. 10; Fig. S7). This new inflection point can now be calculated as the value of the x-axis at half distance between zero and the upper plateau of the generalised reparameterised sigmoid function (Eq. 10).

$$\dot{X} = \frac{\beta_1 \eta - \log\left(e^{\frac{y0}{2}}\sqrt{e^{y0\eta} + 1 - 1}\right)}{y0} \tag{10}$$

Genes were functionally annotated with gene ontology categories [31] using BiomaRt [32]. Furthermore, genes were functionally annotated with the protein atlas database [33] for identifying those that interface with the extracellular environment, encoding for cell-surface and secreted proteins. For a more in-depth analysis of possible interactions between cell types, we compiled a cell-type-specific annotation database for cell-surface and secreted protein-coding genes (Supplementary file 3).

### Differential tissue composition analyses

The differential tissue composition analysis is composed of two integrated modules. First, a module infers tissue composition from whole-tissue gene transcript abundances based on reference transcriptional profiles of pure cell types (deconvolution). Second, a module for beta regression on the inferred proportions along the factor of interest (and additional covariates). Bayesian inference allows the transfer of the uncertainty between the two modules (GitHub: stemangiola/ARMET@v0.7.1). The probabilistic framework Stan [26] was used to encode the joint probability function of the model [34]. The 0.95 credible interval of the posterior distributions was used as a significance threshold.

The supervised deconvolution was based on deconvolution signatures created using a curated collection of 250 publicly available transcriptional profiles (included in BLUEPRINT [35], ENCODE [36], GSE89442 [37] and GSE107011 [38]) encompassing of 8 broad categories of

cell types and 18 cell phenotypes. Genes whose transcription varied across datasets (detected using Limma [28]) were used to identify highly correlated datasets. The Pearson correlation was calculated for all-versus-all samples. The samples with a Pearson correlation greater than 0.99 were discarded as redundant. Each cell-type category was classified as belonging to a node of the cell-differentiation tree, which includes epithelial, fibroblasts, endothelial and immune cells in the first level, and B-, T-, natural killer, monocyte-derived, and granulocyte cells. For each cell type in the differentiation tree, the gene-transcript abundance was modelled using a negative binomial distribution (parameterised by mean and overdispersion). Differences in sequencing depth across biological replicates were modelled with a biological replicate-wise exposure rate term $\epsilon$ that multiplies the transcripts expected abundance (mean). For each cell-type pair of the same level, 40 genes (20 for each direction) were selected that (i) were abundant (had a mean value higher than the median of all genes), and (ii) segregated the two cell types (having the largest gap between the upper quantile of one cell-type and the lower quantile of the other; 95% credible interval). The gene selection for each level was represented by the union of marker genes for all cell-type pairs. The inference was carried out along the two levels of the hierarchy structure, and the inference for each node (e.g. T-cells) was relative to its parent (e.g. immune cells).

### Analysis of tumour microenvironment using multiplex immunohistochemistry

Slides (3 μm) from formalin-fixed and paraffin-embedded (FFPE) tissue were taken from a total of 63 core biopsies of localised prostate cancer across 17 patients. A pathological evaluation was done to define the tumour and surrounding benign tissue areas for each biopsy. The methodology for performing multiplex immunohistochemistry, cell type classification and localisation has been detailed by Keam et al. [39]. Briefly, slides were deparaffinised and rehydrated with xylene and ethanol. The fluorochrome-coupled antibodies against human CD68 (macrophages and dendritic cells), high molecular weight cytokeratin (HMWCK; epithelial basal cells), CD3 (T cells), CD20 (B cells), CD11c (dendritic cells), and PDL1 were used. The dye DAPI was used for nuclei staining. Vectra 3.0 Automated Quantitative Pathology Imaging System (Perkin Elmer, MA) was used for imaging, as Keam et al. [39] detailed. The software HALO was used for cell segmentation and phenotyping. Stromal cells were defined with the negative selection of all antibodies (DAPI positive) and with filtering by large size (cell area > 70) and highly elongated shape (ratio of largest dimension and smallest dimension > 2; 0.9 percentile; 0.9 percentile; Fig. S6).

Mangiola *et al. BMC Cancer* (2021) 21:846

Page 6 of 18

Cell type proximity was quantified as the number of cells within a radius of 20 cells sizes from a selected cell, averaged per tissue area (5 cell size units) for smoothing and avoiding information duplication due to tight cell clusters. Cell relative size was calculated at 15 units as the observed median length units in the coordinate system. The statistics were summarised at the biopsy level. When the distance between two cell types was measured, only the biopsies including both cell types were selected. The robust regression analyses were performed using the R heavy package [40] on log-transformed proximity measure. The co-proximity analysis between epithelial basal cells and PDL1+ macrophages and T cells was performed at the single cell level (averaged by tissue area of 5 cell size units). We calculated the proximity on a radius of 50 relative cell sizes for ensuring good coverage of both T cells and PDL1+ macrophages and decrease sparsity. Only the epithelial basal cells in immune rich areas (with > 5 neighbour T cells) were considered.

## Result

### Data generation and quality

To investigate the role of the tumour microenvironment in patient outcome, we enriched for four cell populations (epithelial: EpCAM$^+$; fibroblasts CD90$^+$/CD31$^-$; T cells: CD45$^+$/CD3$^+$; and myeloid: CD45$^+$/CD16$^+$) from fresh prostatectomies of 13 prostate cancers, ranging from benign tissue (labelled as CAPRA-S score 0) to high-risk tumours (CAPRA-S risk score 7). The choice of those cell populations was guided by their predominant role in the progression of prostate and other cancers [41–45]. Technical and practical experimental limitations prevented considering other key cell types such as luminal and basal epithelial compartments, endothelial, smooth muscle and other lymphocytes such as B and natural killer cells. RNA extracted from the four cell populations was then sequenced, generating a median of 22 million reads per library (Fig. S1 and S2). Overall for the four cell type categories, a conservative cell-type purity inference (using Cibersort and a Bayesian estimator; Material and Methods) estimated high enrichment: 99% for epithelial samples; 99% for fibroblasts; 97% for myeloid (83% for neutrophils; 14% for monocyte-derived cells); and 95% for T cells (Fig. S3). On average, across the four cell types, 40% of genes had 0 sequenced reads in more than half of the samples and were removed from further analysis.

### Differential transcription and model fitting

Dimensionality reduction (multidimensional scaling; in Materials and Methods) of the filtered transcript abundance revealed an association of CAPRA-S risk score with either the first and the second principal components (Fig. 1a; with an indicative direction represented

by the dashed grey line; tested with linear regression, lm function from R). A clear gradient in risk score was seen for epithelial and fibroblasts (Bonferroni adjusted $p$-value of $1.0 \times 10^{-2}$ and $9.7 \times 10^{-3,}$ respectively). A weaker pattern was apparent for myeloid and T cells (Bonferroni adjusted p-value of $3.0 \times 10^{-2}$ and 0.37 respectively), possibly due to the greater heterogeneity of the two immune cell populations than epithelial and fibroblasts.

We performed a differential transcript abundance analysis (at the gene level) for each cell type independently, seeking associations between transcript abundance across subjects and CAPRA-S risk score, treated as pseudo-continuous variables. In order to perform differential analyses that would robustly model the pseudo-continuous properties of the CAPRA-S risk score (Fig. 1a), we developed a Bayesian inference model (TABI) that implements a robust generalised sigmoid regression (i.e. sigmoid function extending from zero to any positive value). TABI was used to model the gene transcript abundance as a continuous function of CAPRA-S risk score (from 0 representing benign to 7 representing high risk); this avoids the loss of information caused by patients' a priori stratification in low–/high-risk groups based on an arbitrary threshold. In principle, using a generalised sigmoid function permits modelling linear, exponential and sigmoid-like trends of transcriptional alterations (Fig. 1b and c). However, to provide robust modelling for RNA sequencing data, we reparameterised the generalised sigmoid function to suit better the numerical properties of transcript abundance (Fig. 1b; Materials and Methods). In addition to robustness, the sigmoid function allows mapping each differential transcriptional event with a clinical risk state, effectively providing a new developmental dimension to the analyses. This mapping is possible because the inflection point represents the CAPRA-S risk score at which the transcriptional alteration is most pronounced. The location of the most rapid change can be highly localised in the case of a dramatic change in transcription at a specific risk score or can be diffused in the case of a gradual change of transcription along the risk score range (Fig. 1c).

Following the statistical inference of transcriptional alterations, an average of 10% of genes was removed based on the posterior predictive check across all samples [29], as not satisfying the assumptions of our model (Table 1; Materials and Methods). A total of 1626 genes were identified as differentially transcribed across the four cell type categories (i.e. 95% credible interval excluding zero; with no need for multiple test adaptation, consistent with standard practices in Bayesian statistics [30]; Table 1; Supplementary file 1). The distributions of differential

**Fig. 1** (See legend on next page.)

Mangiola *et al. BMC Cancer*     (2021) 21:846

Page 8 of 18
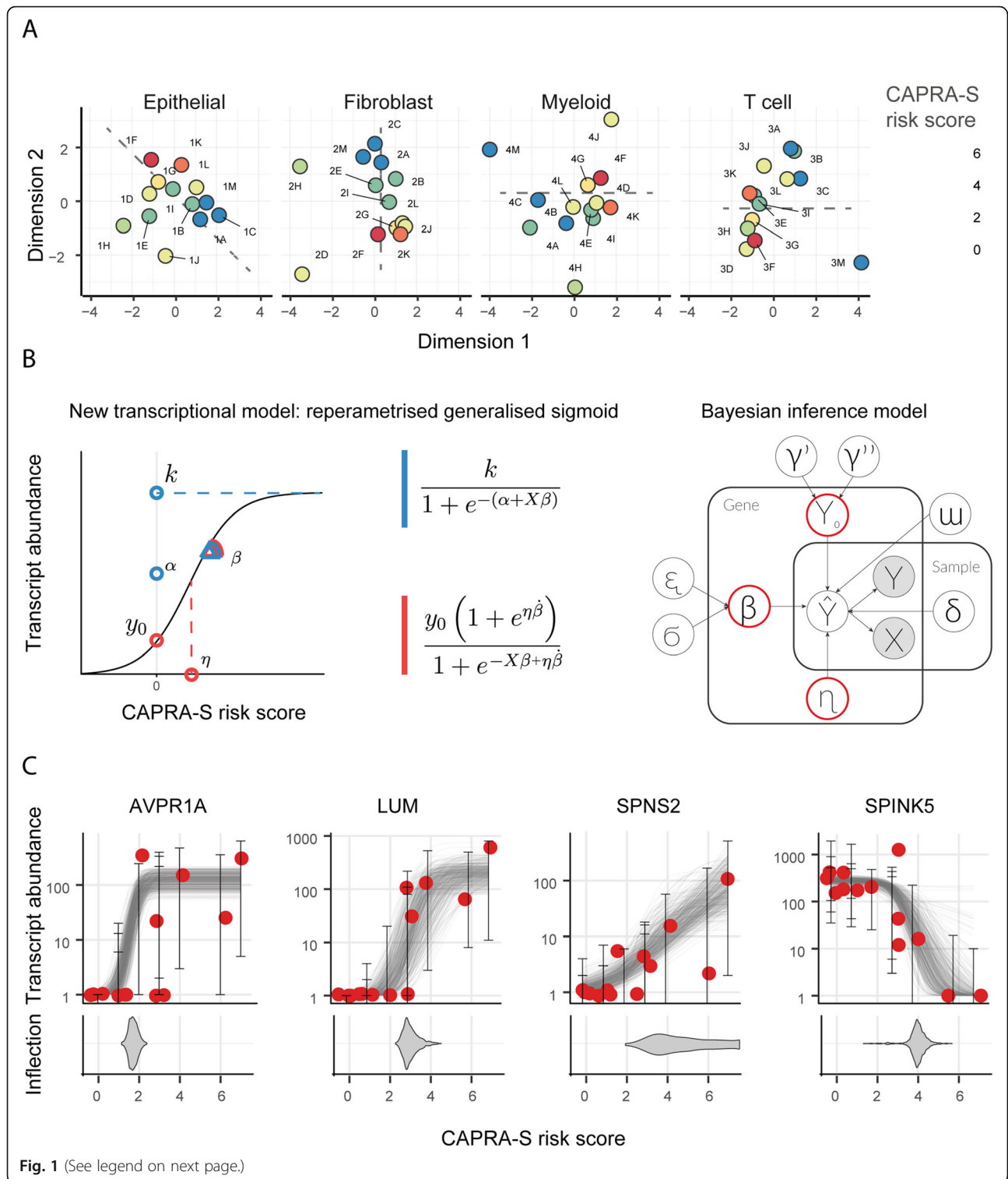
(See figure on previous page.)

**Fig. 1** The continuous relationship between the CAPRA-S risk score and gene transcript abundance. **A** Multidimensional scaling plots of transcript abundance grouped by cell type. The colour coding represents the CAPRA-S risk score. The risk-score is correlated with the first and second dimension, particularly in epithelial and fibroblast cells (linear regression performed using lm in R; Bonferroni adjusted *p*-value of 0.0187, 0.00971, 0.0306 and 0.367, respectively). Alphanumeric codes refer to patient identifiers (Supplementary Table S1). The dashed lines indicate the correlation between the first and the second dimension with the CAPRA-S risk score. **B** Re-parameterisation of the generalised sigmoid function and probabilistic model (Material and Methods). Left-panel: The three reference parameters for the standard parameterisation (blue). Alternative robust parameterisation (red). Right-panel: a graphic representation of the probabilistic model TABI. **C** Examples of continuous relationships between transcript abundance of four representative genes and CAPRA-S risk score (for epithelial cell population), from more discrete-like to more linear-like. The bottom panel displays the inferred distribution of possible values (as posterior distribution) of the inflection point for each gene sigmoid trend

transcription events along the CAPRA-S risk score range are concentrated on low-risk scores (Fig. S4) for the four cell types, indicating that most transcriptional changes occur early in cancer developmental stages (including benign prostate tissue).

### Differentially transcribed cell-surface and secreted protein-coding genes are linked with recurring cancer hallmarks

In order to provide an initial biological evaluation of the resultant differentially transcribed genes, we sought the overlap with cancer-related gene datasets and calculated the enrichment of gene sets against functional and clinical gene annotation databases. On average, across the four cell types, 14% of all the differentially transcribed genes have been previously identified as cancer-related; of these, 24% have been previously described as prostate cancer-related genes [46] (Table 1). For differentially transcribed cell-surface and secreted protein-coding genes, an average of 33 and 51% have been previously described as cancer and prostate cancer-related genes, respectively [46] (Table 1). In order to investigate possible cell-cell interactions within the primary prostate tumour microenvironment, we focused on genes encoding for cell-surface and secreted proteins, which may
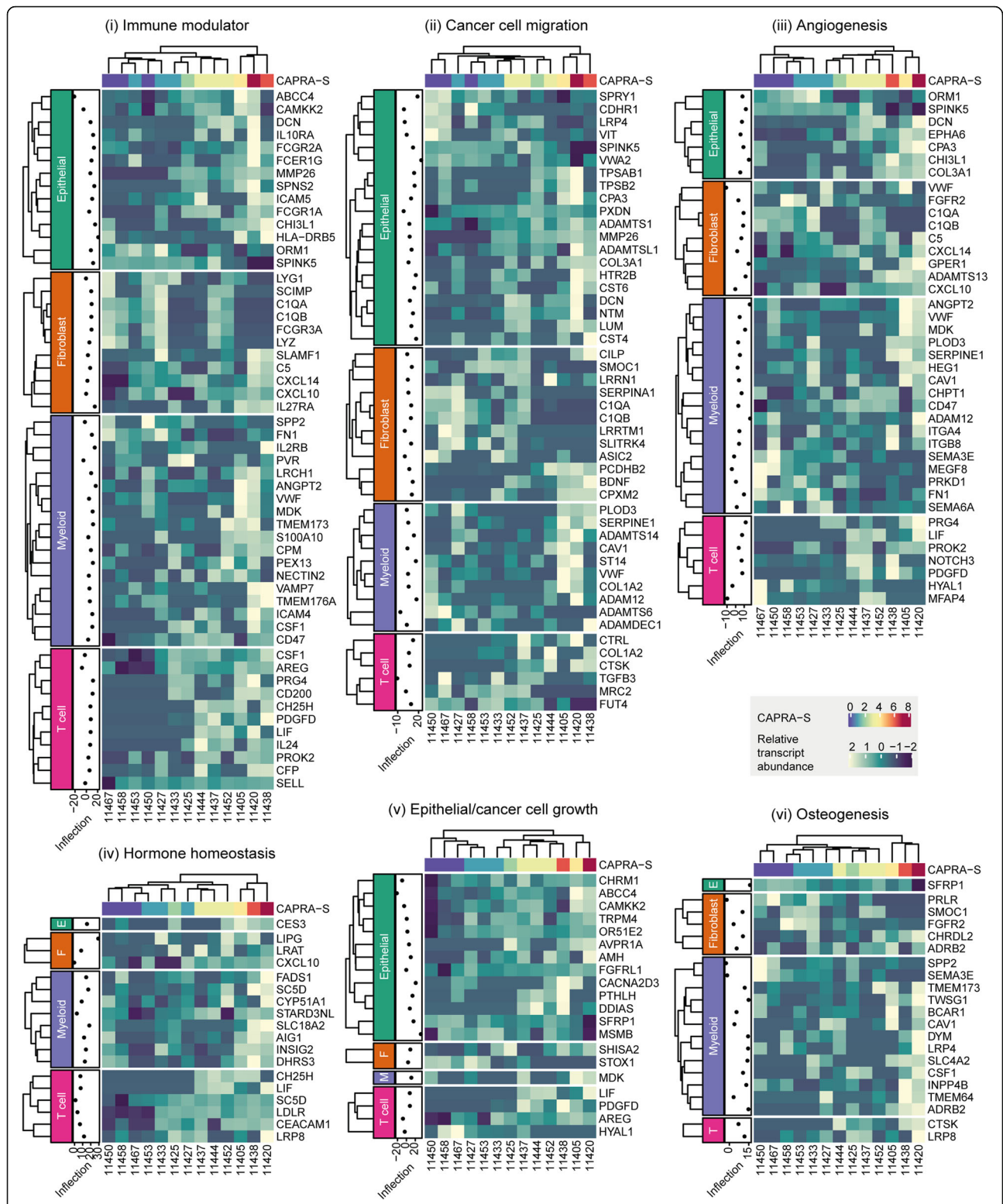
directly influence other cell types. On average, across the four cell types, 35% of differentially transcribed genes encode for cellular-interface proteins; of those, 148 genes have been previously described as cancer-related genes. For all cell types, most cancer genes have a direction of change for all cell types consistent with the direction reported in the literature (35 vs 13 for epithelial; 17 vs 8 for fibroblasts; 32 vs 6 for myeloid cells; and 26 vs 11 for T cells; Supplementary file 3).

In order to allow an in-depth interpretation of the concurrent transcriptional differences for cell-surface and secreted protein-coding genes across cell-types, we produced a cell-type and disease-specific annotation database integrating curated cell-specific Gene Ontology information [31] with more than 1500 scientific articles (Supplementary file 3). This database allowed us to identify six recurring hallmarks of cancer (Fig. 2): (i) immune modulation; (ii) cancer cell migration; (iii) angiogenesis; (iv) hormonal homeostasis; (v) epithelial/cancer cell growth; and (vi) osteogenesis. Among the immune modulation related genes, a balance exists between pro and anti-inflammatory. This balance appears to be dynamic along the disease progression course. The epithelial cell migration hallmark includes three main functional clusters: tissue remodelling, tissue fibrosis and

**Table 1** Summary statistics of the differential transcription analysis, including 52 samples from 13 patients and 4 enriched cell types

| Cell type | Total genes | Genes filtered (zeros) | Genes filtered (PPC) | Differentially transcribed | | | Differentially transcribed in the interface (curated annotation) | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Total (up/down) | Of which cancer genes | Of which PC genes | Total (up/down) | Of which cancer genes, consistent | Of which PC genes, consistent |
| Epithelial | 21,618 | 5408 | 189 | 171 (139/32) | 45 (26%) | 29 (64%) | 80 (67/13) | 35 (44%) | 23 (67%) |
| Fibroblast | 21,510 | 7141 | 651 | 267 (156/111) | 27 (10%) | 9 (33%) | 97 (58/39) | 17 (18%) | 7 (41%) |
| Myeloid | 22,507 | 13,836 | 2695 | 900 (827/73) | 56 (6%) | 11 (20%) | 261 (238/23) | 32 (12%) | 10 (31%) |
| T cell | 21,716 | 8807 | 540 | 288 (195/93) | 42 (15%) | 18 (42%) | 83 (55/28) | 26 (31%) | 15 (58%) |

*PPC* posterior predictive check, *PC* prostate cancer. "Of which" refers to the gene selection relative to the category adjacent on the left. "Interface" refers to cell-surface and secreted protein-coding genes. "Curated" refers to the curated database for cellular-interface genes produced in our study (Supplementary file 2). "Consistent" refers to a consistent direction of transcriptional change according to the curated database. Genes were labelled as "cancer genes" if present in the tier1 COSMIC databasehttps://paperpile.com/c/BQQ95X/zLPNs [46] or labelled as such in our manually curated cell-type-specific database (Supplementary file 2). Genes were labelled as "prostate cancer genes" if present in the tier1 COSMIC prostate cancer database datasethttps://paperpile.com/c/BQQ95X/zLPNs [46] or labelled as such in our manually curated cell-type-specific database (Supplementary file 2)

**Fig. 2** Recurrent functional categories identified in differentially transcribed secreted and transmembrane genes. The estimated inflection point for each gene shows the CAPRA-S risk score at which the transcriptional change was fastest; values < 0 or > 7 indicate an early or late change, respectively

Mangiola *et al. BMC Cancer* (2021) 21:846

Page 10 of 18

direct epithelial-to-mesenchymal transition. The differential transcription events of those three classes do not appear to be concentrated on any particular stage of disease progression.

Similarly, angiogenesis signalling appears to be sustained along the whole disease progression. A cluster of genes linked to platelet recruitment and endothelial cell migration is differentially expressed in synergy by both myeloid and T cells. Several transcriptional alterations from both epithelial and immune cells were linked with hormonal and lipid homeostasis, a key molecular hallmark in prostate cancer [47]. Within this set, the most recurring metabolite that is linked with differentially transcribed genes is cholesterol. While all four cell types contributed similarly for most hallmarks, a clear bias is present for cancer cell growth, osteogenesis and hormone modulation, which genes are differentially transcribed in epithelial and immune cells types, respectively. As the most compelling signal, immune modulation was selected for further investigation.

### Immune modulation is associated with cancer grade and targets predominantly monocyte-derived cells

In order to elucidate the role of the four cell types in the immune response to primary prostate cancer and their potential interactions, we focused on genes that encode for cell-surface and secretory proteins involved in immune modulation. In doing so, we again used the fitted inflection point of the sigmoid model to distinguish between early (i.e. low CAPRA-S risk score) and late (i.e. high) transcriptional changes. The balance between pro and anti-inflammatory signalling from the four cell types tracks the risk score covariate (Fig. 3). The number of differentially abundant gene-transcripts encoding for pro-inflammatory proteins remains roughly constant through the risk range, with 18 genes for CAPRA-S risk score ≤ 2 and 14 for CAPRA-S risk score > 2. On the contrary, the number of altered anti-inflammatory-related genes significantly expands (*p*-value 0.015; t-test) for more advanced stages of the disease, with 12 genes against 20 for the two risk score categories, respectively.

Overall, a large proportion (14 genes of 27) of the inflammatory-related transcriptional alterations across all four cell types is involved in the recruitment of monocytes and macrophages [48–55] (highlighted in yellow in Fig. 3a). These include CAMKK2 [56], ORM1 [57] and DCN [58, 59] in epithelial; IL2RB [60, 61], ICAM4 [62, 63], DCN [58, 59] and MDK [64, 65] in myeloid cells; and CSF1 [48] and PDGFD [66] in T cells. In addition, we identified a known fibroblast-macrophage chemotactic interaction including the regulation of the cytokines CXCL10 [51], CXCL14 [50] and the receptor SLAMF1 [52, 53] for fibroblasts; with COL1A2 [67] and CYR61 [68] (for

CAPRA-S 6–8) altered in myeloid cells known to function as a co-stimulatory loop. A smaller cluster of genes was linked with T cell recruitment and inflammation, including CFP [69], IL24 [70], PROK2 [71], SELL [72]. Interestingly, epithelial cells upregulate a cluster of receptor genes normally involved in antigen recognition and presentation in immune cells [73], including an MHC class II cell surface receptor (i.e. HLA-DRB5) and three Fc receptors (i.e. FCER1G, FCGR1A and FCGR2A).

Overall, the significant genes associated with anti-inflammation across the four cell types targets a more heterogeneous set of cell types than the pro-inflammatory ones. Monocyte-derived cells are mainly targeted by genes that are differentially transcribed in epithelial and myeloid cells. These include the receptor genes SPNS2 [74, 75], IL10RA [76] and ICAM5 by epithelial cells; and the receptor genes CPM [77] and PEX13 [78] and the secreted protein genes FN1 [79] and ANGPT2 [80] by myeloid cells. Another cluster of genes targets predominantly T cells, including AREG [81, 82], CD200 [83], LRCH1 [84], CD47 [85]. Fibroblasts mainly downregulate pro-inflammatory cell-surface and secreted protein genes, such as FCGR3A and C1QA/B.

### Increased monocyte-derived cell infiltration in tumours is associated with lowered disease-free survival

In order to test the relevance of recruitment of monocyte-derived cells suggested by our integrated transcriptional analysis, we performed a differential tissue composition analysis (i.e. a test for difference in cell-type abundance between conditions) based on an independent methodology and independent patient cohort. We used a higher-order Bayesian inference model [34] that integrates deconvolution and downstream regression modules in a joint model (that showed superiority compared with the serial use of deconvolution and regression; Fig. S8) on an independent cohort of 134 patients from the primary prostate cancer.

The Cancer Genome Atlas (TCGA) dataset [7] included both disease-free survival and CAPRA-S score information. The deconvolution module of this algorithm bases its supervised inference of cell-type proportions on a collection of 250 curated publicly available transcriptional profiles (including BLUEPRINT [35], ENCODE [36], GSE89442 [37] and GSE107011 [38]), encompassing 8 broad cell categories and 18 phenotypes. The deconvolution model uses those reference deconvolution signatures to estimate the contribution of each cell type to the observed mixed transcriptional signal (i.e. TCGA tissue RNA sequencing data). This analysis provides tissue composition estimates as well as their association with risk score [34]. Overall, we estimated a median of 88% for epithelial cellular fraction across samples
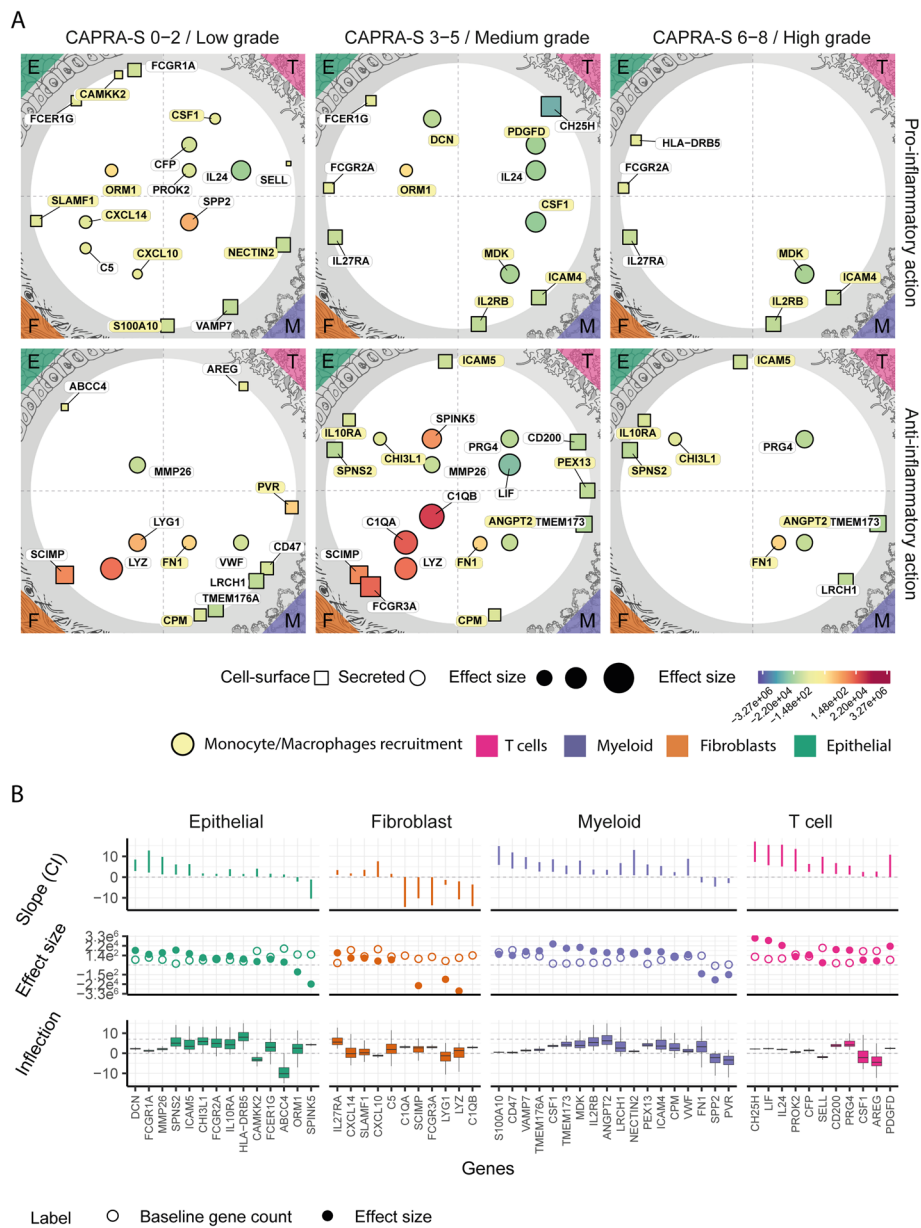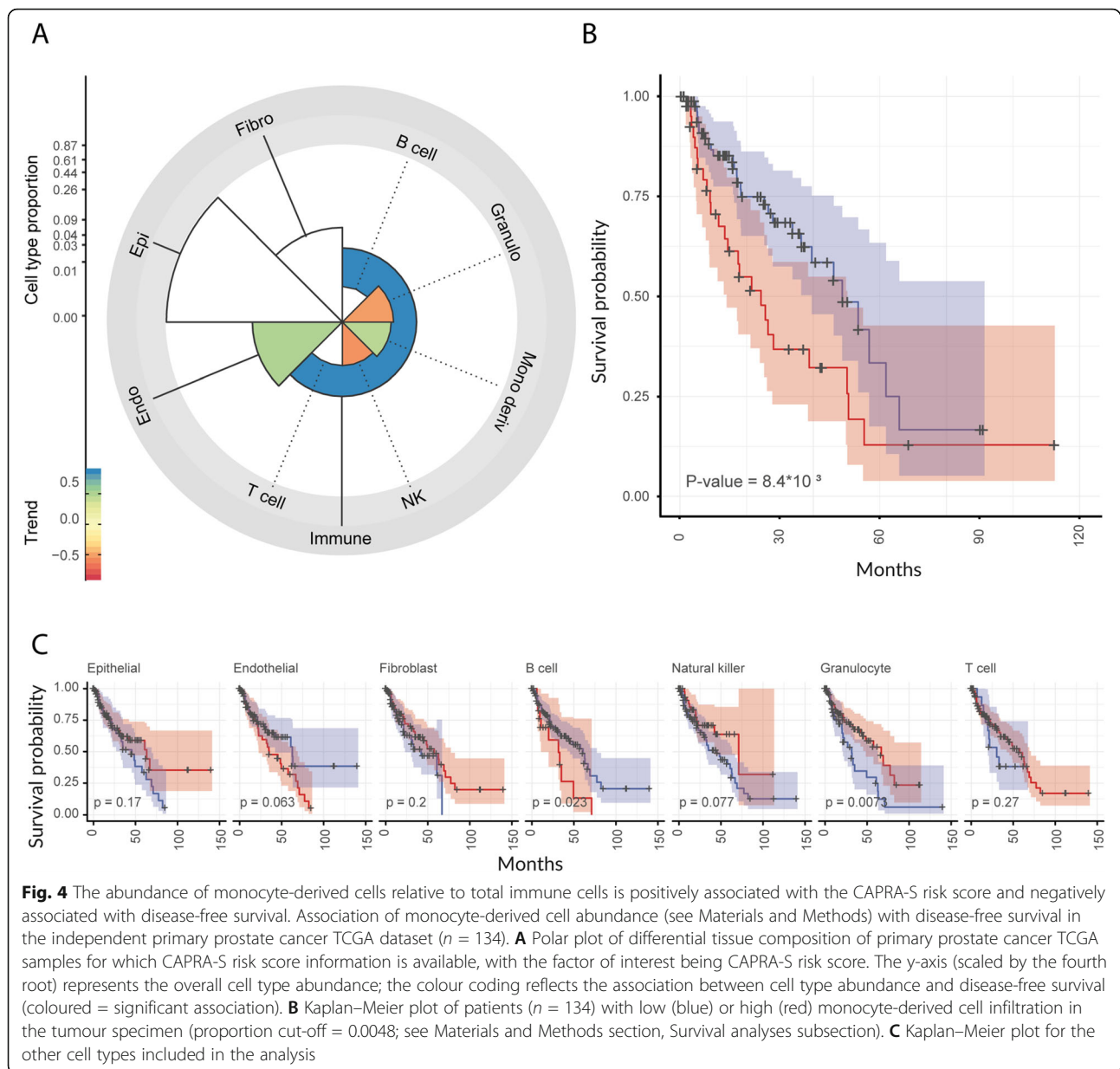
**Fig. 3** Multi cell-type immune-modulation changes with risk progression and is mainly targeted at monocyte-derived cells. The landscape of the immune-modulation related genes encoding cellular interface-proteins (i.e. cell-surface or secreted) inferred to be differentially transcribed across CAPRA-S risk scores, grouped by cell type. **A** Map of the secretory (represented as circles) and cell-surface (represented as squares) protein-coding genes that are differentially transcribed across the four cell types. The data point size is proportional to the baseline transcript abundance. The colour coding represents the effect size. Genes with a similar inflection point (i.e. at what stage of the disease a transcriptional change happens) are clustered vertically (CAPRA-S risk score < =2, > 2 and < =5 and > 5). Genes are split horizontally according to their pro- or anti-inflammatory role. Genes encoding for proteins that target monocyte-derived cells are highlighted in yellow. **B** Statistics of the differentially transcribed genes displayed in panel (**A**). Top: credible interval of the association between transcript abundance and CAPRA-S risk score. Middle: inferred effect size (full dots) and baseline transcription (empty dots). Bottom: credible interval of the CAPRA-S value for the transcriptomic change (i.e. inflection point; e.g., the gene HLA − DRB5 is upregulated in late stages of the disease)

(consistent with public literature [86], Fig. S5), 4.8% for endothelial, 4.8% for fibroblasts, and 1.6% for immune cells. The differential tissue composition analysis showed a significant positive association with the CAPRA-S risk score of the monocyte-derived and a negative association

of the natural killer and granulocyte cells (95% credible interval excluding 0; Fig. 4a).

In order to test whether the enrichment in monocyte-derived cells is clinically relevant, we generated Kaplan–Meier curves using the estimated cell-

**Fig. 4** The abundance of monocyte-derived cells relative to total immune cells is positively associated with the CAPRA-S risk score and negatively associated with disease-free survival. Association of monocyte-derived cell abundance (see Materials and Methods) with disease-free survival in the independent primary prostate cancer TCGA dataset (*n* = 134). **A** Polar plot of differential tissue composition of primary prostate cancer TCGA samples for which CAPRA-S risk score information is available, with the factor of interest being CAPRA-S risk score. The y-axis (scaled by the fourth root) represents the overall cell type abundance; the colour coding reflects the association between cell type abundance and disease-free survival (coloured = significant association). **B** Kaplan–Meier plot of patients (*n* = 134) with low (blue) or high (red) monocyte-derived cell infiltration in the tumour specimen (proportion cut-off = 0.0048; see Materials and Methods section, Survival analyses subsection). **C** Kaplan–Meier plot for the other cell types included in the analysis

type abundances. The stratification of patients based on the extent of monocyte-derived cells infiltration revealed a significant separation of the survival probabilities (Fig. 4b). For comparative purposes, we tested patient stratification for the other cell types included in the model. As a result, only granulocytes and B cells (with the poor outcome cohort including only a few patients) showed a significant negative association. In contrast, no other significant associations were detected for other cell types, including epithelial, endothelial, fibroblasts, and immune cell types, including T cells and natural killers (Fig. 4c). The negative association for the epithelial component might be due to the absence of the mesenchymal component in

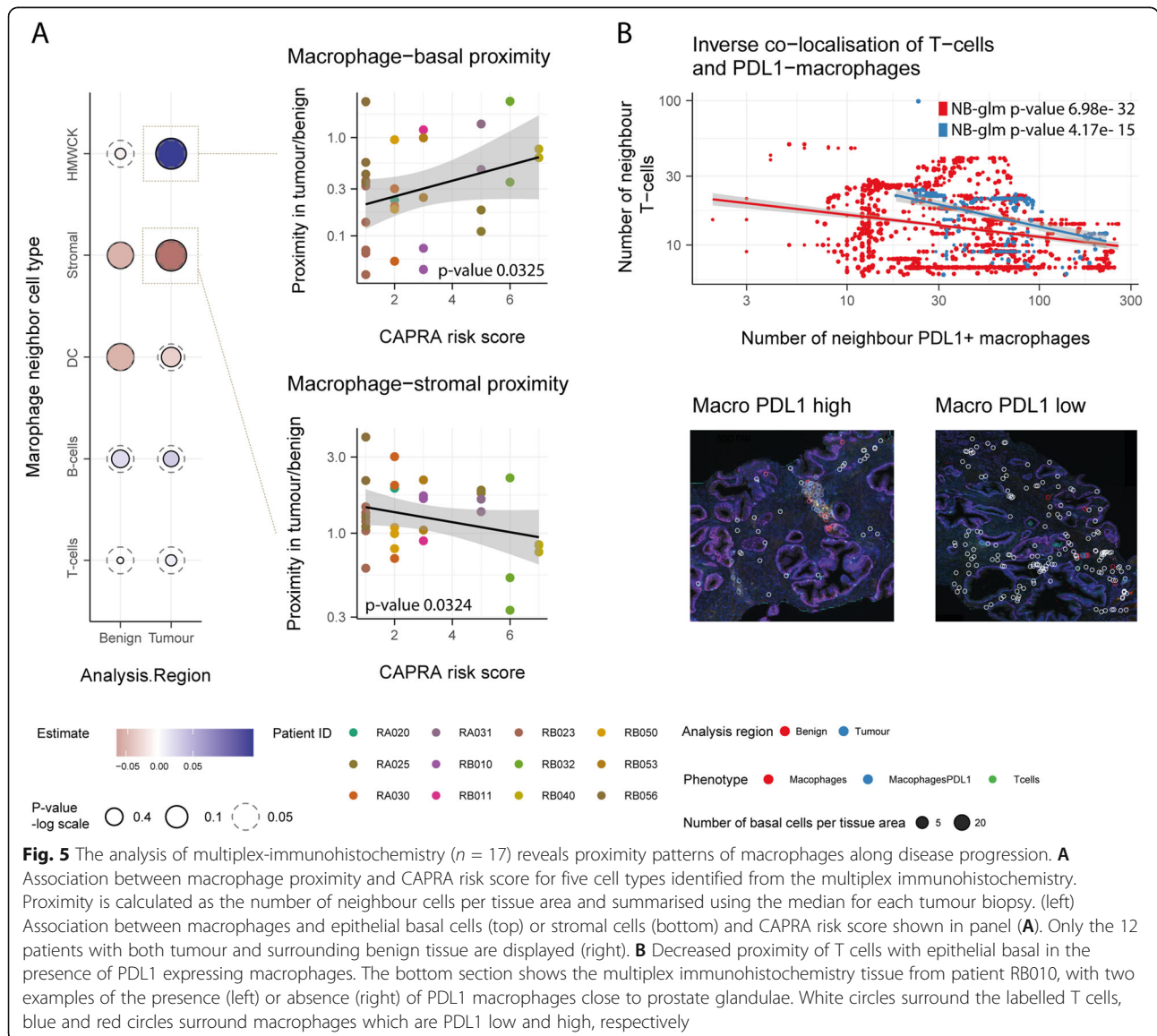the epithelial deconvolution-signature derived from public data.

## Macrophage proximity to epithelial glandular clusters increases with tumour progression

In order to validate our findings and gather more in-depth knowledge about the role of macrophages in disease progression, we performed a spatial analysis at the single-cell level of 63 prostate biopsies from the independent RadBank cohort of 17 patients with localised prostate cancer, spanning a wide range of CAPRA risk scores (Table S1). Using 7 immuno-fluorescent dyes (including DAPI) with 6 others linked to antibodies (CD3, CD20, CD68, CD11c and HMWCK), cell size and shape,

we were able to identify 6 major cell types: T cells, B cells, macrophages, dendritic myeloid, epithelial basal and stromal cells. Epithelial basal cells define prostate glands, where cancer cells originate and co-localise (confirmed by pathological evaluation). PDL1 expressing cells were also categorised using a PDL1-linked dye. Overall, we sampled an average of $41 \times 10^3$ cells per biopsy (standard deviation $2.7 \times 10^4$). The most abundant of the categorised cell types was epithelial basal (8.06% on average), followed by T cells (5.03% on average).

We estimated the association between macrophage proximity to five other cell types and CAPRA risk score (Fig. 5a). Across the five cell types, the average number of neighbour cells to macrophages ranges from 1.4 to 23.0. The strongest positive association was between macrophage and epithelial basal cells in tumour areas

(*p*-value 0.0325; Fig. 5a-top-right), while the strongest negative association was for stromal cells in tumour areas (*p*-value 0.0324; Fig. 5a-bottom). Overall, the average proximity between macrophages and other cells, aggregated by biopsy, did not strongly associate with the CAPRA risk score. In order to gather evidence that the increased proximity of macrophages to the prostatic gland structures in advanced cancer stages had some direct effect on the local immune microenvironment, we tested the hypothesis that macrophages in proximity to gland structures would displace T cells. We observed an inverse association between the number of neighbour macrophage expressing PDL1 to epithelial basal cells and the number of T-cell neighbours (Fig. 5b). Epithelial basal cells that are close to clusters of PDL1 expressing macrophages tend to be further away from T cells.



**Fig. 5** The analysis of multiplex-immunohistochemistry (n = 17) reveals proximity patterns of macrophages along disease progression. **A** Association between macrophage proximity and CAPRA risk score for five cell types identified from the multiplex immunohistochemistry. Proximity is calculated as the number of neighbour cells per tissue area and summarised using the median for each tumour biopsy. (left) Association between macrophages and epithelial basal cells (top) or stromal cells (bottom) and CAPRA risk score shown in panel (**A**). Only the 12 patients with both tumour and surrounding benign tissue are displayed (right). **B** Decreased proximity of T cells with epithelial basal in the presence of PDL1 expressing macrophages. The bottom section shows the multiplex immunohistochemistry tissue from patient RB010, with two examples of the presence (left) or absence (right) of PDL1 macrophages close to prostate glandulae. White circles surround the labelled T cells, blue and red circles surround macrophages which are PDL1 low and high, respectively

## Discussion

To date, in-depth analyses of genomic features of prostate cancer alone, including single nucleotide variants and small and large structural rearrangements, have not been sufficient to provide transformative prognostic tools or unveil the full complexity of this disease. Non-malignant cells within the tumour microenvironment make an integral contribution to the mechanisms that cause cancer progression. They are often modulated by cancer cells toward pro-tumorigenic behaviours. In this study, we significantly improved the knowledge of the molecular landscape of the primary prostate tumour microenvironment, revealing concurrent transcriptional changes in epithelial, fibroblast, myeloid and T cells along the CAPRA-S risk score range.

We optimised a combined fluorescence-activated cell sorting and ultra-low-input RNA sequencing protocol, allowing us to obtain high-quality sequencing data from inputs down to 1000 cells. Such a strategy is of general utility as it enables studies of rare cell types from both fresh tissue cores and biopsies. In order to optimally detect changes in transcription along the CAPRA-S risk score, we developed a novel statistical inference method, TABI. This model permitted modelling transcript abundance natively on continuous factors of interest with a minimal number of parameters ($n$ = 4), avoiding loss of information due to the dichotomisation of the risk score into low –/high-risk patient groups. As suggested by multidimensional-scaling plots and supported by our inference, transcriptional change events are indeed continuously distributed along the whole risk score range. This method is of broad utility in all cases where a continuous (or pseudo-continuous) factor of interest is present (e.g., risk score, time and chemical concentration) and a monotonic change in transcript abundance is of interest. Furthermore, the novel parametrisation of the generalised sigmoid function that TABI is based on can be extrapolated for a wide range of applications. On the contrary, publicly available statistical models for continuous regression of transcript abundance, such as for temporal RNA dynamics [87], cannot be used in the context of risk score. This is because they require a large biological replication, specific experimental design and are not constrained to monotonic trends, affecting practical interpretability.

The compilation of a curated cell-type-specific database of gene functions for cell-surface and secreted protein-coding transcripts enabled the detection of several recurrent hallmarks of prostate cancer, characterised by the involvement of multiple cell types. The most striking aspect to emerge was the large number of differentially abundant gene-transcripts linked to monocyte-derived cell recruitment and modulation. The association of monocyte-derived cell recruitment with increased risk score was reflected in an orthogonal differential tissue composition analysis on the extensive Cancer Genome Atlas (TCGA) independent patient cohort against CAPRA-S risk score through a differential tissue composition analysis. This analysis was enabled by a robust Bayesian inference model, able to transfer the uncertainty of the estimation of tissue composition for each sample to the linear model linking cell-type proportion (across samples) to clinical variables. This aspect is particularly relevant considering the substantial noise associated with the inference of tissue composition (i.e. deconvolution). To test the clinical significance of quantifying monocyte-derived cell numbers within the tumour mass besides the CAPRA-S risk score, we produced Kaplan-Meier estimates using the inferred cell-type proportions along progression-free survival on the TCGA cohort. For both analyses, we identified the strongest association with clinical variables being monocyte-derived cells. The infiltration of cell types such as monocyte-derived cell populations has previously been shown to be linked to the extent of proliferative inflammatory atrophy lesions, chronic prostatic inflammation and cancer grade https://paperpile.com/c/BQQ95X/GLf0R [88]. In prostate cancer, specific and overall survival analyses have identified an elevated monocyte count as an independent prognostic factor for poor outcome https://paperpile.com/c/BQQ95X/vyBlX+pxdwu+xaA8T+2mcuI [89–92]. Furthermore, the infiltration of tumour-associated macrophages in prostate needle biopsy specimens has been shown to have potential as a predictive factor for PSA failure or disease progression after hormonal therapy https://paperpile.com/c/BQQ95X/eaIrr [93].

In order to validate further our hypothesis and enrich our knowledge about the relation of macrophages with epithelial and a range of immune cells along disease progression, we used multiplex immune-histochemistry to determine the immune context at the single-cell level in an independent cohort. This data supports the hypothesis of a weakened relation of macrophages with stromal compartments and a strengthened association with epithelial glandular clusters along the disease progression spectrum; glandular clusters being generally colocalised with cancer cells in tumour tissues. This aspect becomes highly relevant as the glandular areas in both tumour and adjacent benign compartments rich in PDL1 macrophages are poorer in T cells. PDL1 expressing macrophages have been associated with their M2 wound-healing phenotype https://paperpile.com/c/BQQ95X/VvaW8 [94]. The relationship between PD-L1 expression in intratumoral macrophages and prognosis in cancer patients is still controversial. The two competing

hypotheses are (i) PDL1 intratumoral macrophages lead to dysfunctional T cellshttps://paperpile.com/c/BQQ95X/2mglb [95] or (ii) not having significant effectshttps://paperpile.com/c/BQQ95X/BJESQ [96].

Nonetheless, the expression of PDL1 in macrophages has been shown to induce anti-inflammatory cytokines such as IL-10https://paperpile.com/c/BQQ95X/VvaW8 [94]. Although PDL1 in macrophages may primarily function as protection against induced cell death, our study supports the hypothesis that it may have the effect of inducing an anti-inflammatory, immune cold local microenvironment, with adverse effects on disease progression. Although the role of several immune cell types has been widely investigated in prostate cancer, the driving forces of immune modulation, clinically relevant for immunotherapy resilience, are still under investigation. From the myeloid compartment, the main focus has been myeloid-derived suppressor cellshttps://paperpile.com/c/BQQ95X/i6jgh [15], that have been investigated mainly through blood and in-vitro analyses [97, 98].

## Conclusion
There has been limited benefit observed in prostate cancer through the unselected use of novel immune checkpoint inhibitors based on T cell receptor blockade (e.g., PD-1, PD-L1 and CTLA-4) https://paperpile.com/c/BQQ95X/i6jgh [15]. Such failure may, in part, be driven by our limited understanding of the dynamic interplay between immune components of the microenvironment and tumour cells. This study provides a clear direction for further investigation into mechanisms of the immune system, monocyte-derived cells in particular, that contribute to disease progression; for example, through changing the hormonal and growth-factor homeostasis through a sustained inflammatory state. Furthermore, this study provides a novel and robust method for detecting monotonic changes in transcript abundance over a continuous factor of interest such as risk and time that has broad applicability to other research areas. The methodological advances and the novel findings presented in this study provide a research framework for improved immune interventions.

### Abbreviations
TCGA: The Cancer Genome Atlas; PBS: Phosphate-buffered saline

## Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12885-021-08529-6.

**Additional file 1.**
**Additional file 2.**
**Additional file 3.**
**Additional file 4.**

**Additional file 5.**

### Availability of data and materials
The code used to conduct the analyses is available at github.com/stemangiola under the following repositories: prostate-TME-N52–2019; TABI@v0.1.3; ARMET@v0.7.1. Sequence data was deposited at the European Genome-phenome Archive (EGA), hosted by the EBI and the CRG, under accession number EGAD00001004948.

## Declarations

### Ethics approval and consent to participate
The collection and use of tissue for this study had Melbourne Health institutional review board approval. Patients provided written informed consent (Melbourne Health Local Project Number: 2016.087; and PMCC; HREC approvals 10/68, 13/167, 18/204).

### Consent for publication
Not applicable.

Mangiola *et al. BMC Cancer* (2021) 21:846

Page 16 of 18

## Author details
[1]Bioinformatics Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. [2]Department of Surgery, The University of Melbourne, Parkville, Victoria, Australia. [3]Department of Urology, Royal Melbourne Hospital, Parkville, Victoria, Australia. [4]Department of Medical Biology, University of Melbourne, Melbourne, Victoria, Australia. [5]Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic. [6]University of Queensland Diamantina Institute, Translational Research Institute, University of Queensland, Brisbane, QLD, Australia. [7]The Peter Doherty Institute for Infection and Immunity, Parkville, Victoria, Australia. [8]Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria, Australia. [9]Peter MacCallum Cancer Centre, Melbourne, VIC 3000, Australia. [10]Flow Cytometry Facility, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia. [11]Cancer Program, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia. [12]Department of Anatomy and Developmental Biology, Monash University, Clayton, Victoria, Australia. [13]Epworth Center of Cancer Research, Clayton, Victoria, Australia. [14]The Olivia Newton-John Cancer Research Institute, Heidelberg, Melbourne, Australia. [15]Department of Urology, Frankston Hospital, Frankston, Victoria, Australia. [16]School of Mathematics and Statistics, University of Melbourne, Melbourne, VIC 3010, Australia.

## References
1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018;68:394–424.
2. D'Amico AV, Whittington R, Malkowicz SB, Fondurulia J, Chen MH, Kaplan I, et al. Pretreatment nomogram for prostate-specific antigen recurrence after radical prostatectomy or external-beam radiation therapy for clinically localized prostate cancer. J Clin Oncol. 1999;17:168–72.
3. Bhindi B, Karnes RJ, Rangel LJ, Mason RJ, Gettman MT, Frank I, et al. Independent validation of the American Joint Committee on Cancer 8th Edition Prostate Cancer Staging Classification. J Urol. 2017;198:1286–94.
4. Lu-Yao GL, Albertsen PC, Moore DF, Lin Y, DiPaola RS, Yao S-L. Fifteen-year Outcomes Following Conservative Management Among Men Aged 65 Years or Older with Localized Prostate Cancer. Eur Urol. 2015;68:805–11.
5. Chua MLK, Lo W, Pintilie M, Murgic J, Lalonde E, Bhandari V, et al. A Prostate Cancer "Nimbosus": Genomic Instability and SChLAP1 Dysregulation Underpin Aggression of Intraductal and Cribriform Subpathologies. Eur Urol. 2017;72:665–74.
6. Fraser M, Sabelnykova VY, Yamaguchi TN, Heisler LE, Livingstone J, Huang V, et al. Genomic hallmarks of localized, non-indolent prostate cancer. Nature. 2017;541:359–64.
7. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. Cell. 2015;163:1011–25.
8. Tyekucheva S, Bowden M, Bango C, Giunchi F, Huang Y, Zhou C, et al. Stromal and epithelial transcriptional map of initiation progression and metastatic potential of human prostate cancer. Nat Commun. 2017;8:420.
9. Strasner A, Karin M. Immune Infiltration and Prostate Cancer. Front Oncol. 2015;5:128.
10. Salvatore V, Teti G, Focaroli S, Mazzotti MC, Mazzotti A, Falconi M. The tumor microenvironment promotes cancer progression and cell migration. Oncotarget. 2017;8:9608–16.
11. Katt ME, Placone AL, Wong AD, Xu ZS, Searson PC. In vitro tumor models: advantages, disadvantages, variables, and selecting the right platform. Front Bioeng Biotechnol. 2016;4:12.
12. Smith BA, Sokolov A, Uzunangelov V, Baertsch R, Newton Y, Graim K, et al. A basal stem cell signature identifies aggressive prostate cancer phenotypes. Proc Natl Acad Sci U S A. 2015;112:E6544–52.
13. Zhang D, Park D, Zhong Y, Lu Y, Rycaj K, Gong S, et al. Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. Nat Commun. 2016;7:10798.
14. Berglund E, Maaskola J, Schultz N, Friedrich S, Marklund M, Bergenstråhle J, et al. Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. Nat Commun. 2018;9:2419.
15. Rescigno P, de Bono JS. Immunotherapy for lethal prostate cancer. Nat Rev Urol. 2019;16:69–70.
16. Venturini NJ, Drake CG. Immunotherapy for Prostate Cancer. Cold Spring Harb Perspect Med. 2019;9(5):a030627. https://doi.org/10.1101/cshperspect.a030627.
17. Kerger M, Hong MKH, Pedersen J, Nottle T, Ryan A, Mills J, et al. Microscopic assessment of fresh prostate tumour specimens yields significantly increased rates of correctly annotated samples for downstream analysis. Pathology. 2012;44:204–8.
18. Mangiola S, Stuchbery R, Macintyre G, Clarkson MJ, Peters JS, Costello AJ, et al. Periprostatic fat tissue transcriptome reveals a signature diagnostic for high-risk prostate cancer. Endocr Relat Cancer. 2018;25:569–81.
19. Andrews S. FastQC: a quality control tool for high throughput sequence data; 2010.
20. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.
21. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28:2184–5.
22. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
23. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15:R46.
24. Cooperberg MR, Hilton JF, Carroll PR. The CAPRA-S score: A straightforward tool for improved prediction of outcomes after radical prostatectomy. Cancer. 2011;117(22):5039–46.
25. Richards FJ. A flexible growth function for empirical use. J Exp Bot. 1959;10: 290–301 Oxford University Press.
26. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, et al. Stan: A probabilistic programming language. J Stat Softw. 2016;20:1–37.
27. Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electron J Stat. 2017;11:5018–51 The Institute of Mathematical Statistics and the Bernoulli Society.
28. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43:e47.
29. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis. 3rd ed. US: Taylor & Francis Ltd; 2013.
30. Gelman A, Hill J, Yajima M. Why we (usually) don't have to worry about multiple comparisons. J Res Educ Eff. 2012;5:189–211 Routledge.
31. Botstein D, Cherry JM, Ashburner M, Ball CA, Blake JA, Butler H, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25: 25–9.
32. Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. Bioinformatics. 2005;21:3439–40.
33. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. Science. 2015; 347:1260419.
34. Mangiola S, Stuchbery R, McCoy P, et al. Androgen deprivation therapy promotes an obesity-like microenvironment in periprostatic fat. Endocr Connect. 2019;8(5):547–58.
35. Stunnenberg HG, International Human Epigenome Consortium, Hirst M. The international human epigenome consortium: a blueprint for scientific collaboration and discovery. Cell. 2016;167:1897.
36. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.
37. Basit F, Mathan T, Sancho D, de Vries IJM. Human dendritic cell subsets undergo distinct metabolic reprogramming for immune response. Front Immunol. 2018;9:2489.
38. Xu W, Monaco G, Wong EH, Tan WLW, Kared H, Simoni Y, et al. Mapping of γ/δ T cells reveals Vδ2+ T cells resistance to senescence. EBioMedicine. 2019;39:44–58.
39. Keam SP, Halse H, Nguyen T, et al. High dose-rate brachytherapy of localized prostate cancer converts tumors from cold to hot. J Immunother Cancer. 2020;8(1):e000792.
40. Osorio F. Heavy: Robust estimation using heavy-tailed distributions. R package version 0 38 ed; 2016.

41. Ippolito L, Morandi A, Taddei ML, et al. Cancer-associated fibroblasts promote prostate cancer malignancy via metabolic rewiring and mitochondrial transfer. Oncogene. 2019;38(27):5339–55.

42. Jung Y, Kim JK, Shiozawa Y, Wang J, Mishra A, Joseph J, et al. Recruitment of mesenchymal stem cells into prostate tumours promotes metastasis. Nat Commun. 2013;4:1795 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.

43. Kaminski A, Hahne JC, Haddouti E-M, Florin A, Wellmann A, Wernert N. Tumour-stroma interactions between metastatic prostate cancer cells and fibroblasts. Int J Mol Med. 2006;18:941–50.

44. Sharma A, Suleyman N, Jones O, Vasdev N. Immunotherapy in Urological Tumors. Rev Urol. 2019;21:15–20.

45. Seif F, Sharifi L, Khoshmirsafa M, Mojibi Y, Mohsenzadegan M. A Review of Preclinical Experiments Toward Targeting M2 Macrophages in Prostate Cancer. Curr Drug Targets. 2019;20:789–98.

46. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acids Res. 2015; 43(Database issue):D805–D811.

47. Long J, Zhang C-J, Zhu N, Du K, Yin Y-F, Tan X, et al. Lipid metabolism and carcinogenesis, cancer development. Am J Cancer Res. 2018;8:778–91.

48. Zhu Y, Knolhoff BL, Meyer MA, Nywening TM, West BL, Luo J, et al. CSF1/ CSF1R blockade reprograms tumor-infiltrating macrophages and improves response to T-cell checkpoint immunotherapy in pancreatic cancer models. Cancer Res. 2014;74:5057–69.

49. Kumar V, Donthireddy L, Marvel D, Condamine T, Wang F, Lavilla-Alonso S, et al. Cancer-Associated Fibroblasts Neutralize the Anti-tumor Effect of CSF1 Receptor Blockade by Inducing PMN-MDSC Infiltration of Tumors. Cancer Cell. 2017;32:654–68.e5.

50. Lu J, Chatterjee M, Schmid H, Beck S, Gawaz M. CXCL14 as an emerging immune and inflammatory modulator. J Inflamm. 2016;13:1.

51. Liu M, Guo S, Hibbert JM, Jain V, Singh N, Wilson NO, et al. CXCL10/IP-10 in infectious diseases pathogenesis and potential therapeutic implications. Cytokine Growth Factor Rev. 2011;22:121–30.

52. Latour S, Gish G, Helgason CD, Humphries RK, Pawson T, Veillette A. Regulation of SLAM-mediated signal transduction by SAP, the X-linked lymphoproliferative gene product. Nat Immunol. 2001;2:681–90.

53. Wang G, van Driel BJ, Liao G, O'Keeffe MS, Halibozek PJ, Flipse J, et al. Migration of myeloid cells during inflammation is differentially regulated by the cell surface receptors Slamf1 and Slamf8. PLoS One. 2015;10:e0121968.

54. Lenac Rovis T, Kucan Brlic P, Kaynan N, et al. Inflammatory monocytes and NK cells play a crucial role in DNAM-1-dependent control of cytomegalovirus infection. J Exp Med. 2016;213(9):1835–50.

55. O'Connell PA, Surette AP, Liwski RS, Svenningsson P, Waisman DM. S100A10 regulates plasminogen-dependent macrophage invasion. Blood. 2010;116: 1136–46.

56. Racioppi L. CaMKK2: a novel target for shaping the androgen-regulated tumor ecosystem. Trends Mol Med. 2013;19:83–8.

57. Gemelli C, Martello A, Montanari M, Zanocco Marani T, Salsi V, Zappavigna V, et al. The Orosomucoid 1 protein is involved in the vitamin D - mediated macrophage de-activation process. Exp Cell Res. 2013;319:3201–13.

58. Zhang W, Ge Y, Cheng Q, Zhang Q, Fang L, Zheng J. Decorin is a pivotal effector in the extracellular matrix and tumour microenvironment. Oncotarget. 2018;9:5480–91.

59. Järvinen TAH, Ruoslahti E. Target-seeking antifibrotic compound enhances wound healing and suppresses scar formation in mice. Proc Natl Acad Sci U S A. 2010;107:21671–6.

60. Jounaidi Y, Cotten JF, Miller KW, Forman SA. Tethering IL2 to Its Receptor IL2Rβ Enhances Antitumor Activity and Expansion of Natural Killer NK92 Cells. Cancer Res. 2017;77:5938–51.

61. Espinoza-Delgado I, Bosco MC, Musso T, Gusella GL, Longo DL, Varesio L. Interleukin-2 and human monocyte activation. J Leukoc Biol. 1995;57:13–9.

62. Kim M, Lee S-J, Shin S, Park K-S, Park SY, Lee CH. Novel natural killer cell-mediated cancer immunotherapeutic activity of anisomycin against hepatocellular carcinoma cells. Sci Rep. 2018;8:10668.

63. Ihanus E, Uotila LM, Toivanen A, Varis M, Gahmberg CG. Red-cell ICAM-4 is a ligand for the monocyte/macrophage integrin CD11c/CD18: characterization of the binding sites on ICAM-4. Blood. 2007;109:802–10.

64. Narita H, Chen S, Komori K, Kadomatsu K. Midkine is expressed by infiltrating macrophages in in-stent restenosis in hypercholesterolemic rabbits. J Vasc Surg. 2008;47:1322–9.

65. Fan N, Sun H, Wang Y, Zhang L, Xia Z, Peng L, et al. Midkine, a potential link between obesity and insulin resistance. PLoS One. 2014;9:e88299.

66. Yang P, Manaenko A, Xu F, Miao L, Wang G, Hu X, et al. Role of PDGF-D and PDGFR-β in neuroinflammation in experimental ICH mice model. Exp Neurol. 2016;283:157–64.

67. Saito T, Hara M, Kumamaru H, Kobayakawa K, Yokota K, Kijima K, et al. Macrophage Infiltration Is a Causative Factor for Ligamentum Flavum Hypertrophy through the Activation of Collagen Production in Fibroblasts. Am J Pathol. 2017;187:2831–40.

68. Bai T, Chen C-C, Lau LF. Matricellular protein CCN1 activates a proinflammatory genetic program in murine macrophages. J Immunol. 2010;184:3223–32.

69. Blatt AZ, Pathan S, Ferreira VP. Properdin: a tightly regulated critical inflammatory modulator. Immunol Rev. 2016;274:172–90.

70. Persaud L, De Jesus D, Brannigan O, et al. Mechanism of Action and Applications of Interleukin 24 in Immunotherapy. Int J Mol Sci. 2016;17(6): 869.

71. Lattanzi R, Maftei D, Marconi V, Florenzano F, Franchi S, Borsani E, et al. Prokineticin 2 upregulation in the peripheral nervous system has a major role in triggering and maintaining neuropathic pain in the chronic constriction injury model. Biomed Res Int. 2015;2015:301292.

72. Mohammed RN, Watson HA, Vigar M, Ohme J, Thomson A, Humphreys IR, et al. L-selectin is essential for delivery of activated CD8(+) T cells to virus-infected organs for protective immunity. Cell Rep. 2016;14:760–71.

73. Kobayashi KS, van den Elsen PJ. NLRC5: a key regulator of MHC class I-dependent immune responses. Nat Rev Immunol. 2012;12:813–20.

74. Tran HB, Jersmann H, Truong TT, Hamon R, Roscioli E, Ween M, et al. Disrupted epithelial/macrophage crosstalk via Spinster homologue 2-mediated S1P signaling may drive defective macrophage phagocytic function in COPD. PLoS One. 2017;12:e0179577.

75. Rodriguez YI, Campos LE, Castro MG, Aladhami A, Oskeritzian CA, Alvarez SE. Sphingosine-1 Phosphate: A new modulator of immune plasticity in the tumor microenvironment. Front Oncol. 2016;6:218.

76. Shouval DS, Biswas A, Goettel JA, McCann K, Conaway E, Redhu NS, et al. Interleukin-10 receptor signaling in innate immune cells regulates mucosal immune tolerance and anti-inflammatory macrophage function. Immunity. 2014;40:706–19.

77. Denis CJ, Lambeir A-M. The potential of carboxypeptidase M as a therapeutic target in cancer. Expert Opin Ther Targets. 2013;17:265–79.

78. Vijayan V, Srinu T, Karnati S, Garikapati V, Linke M, Kamalyan L, et al. A new Immunomodulatory role for Peroxisomes in Macrophages activated by the TLR4 Ligand Lipopolysaccharide. J Immunol. 2017;198:2414–25.

79. Pinto AR, Godwin JW, Rosenthal NA. Macrophages in cardiac homeostasis, injury responses and progenitor cell mobilisation. Stem Cell Res. 2014;13: 705–14.

80. Wu X, Giobbie-Hurder A, Liao X, Connelly C, Connolly EM, Li J, et al. Angiopoietin-2 as a Biomarker and Target for Immune Checkpoint Therapy. Cancer Immunol Res. 2017;5:17–28.

81. Wang S, Zhang Y, Wang Y, Ye P, Li J, Li H, et al. Amphiregulin confers regulatory T Cell suppressive function and tumor invasion via the EGFR/ GSK-3β/Foxp3 Axis. J Biol Chem. 2016;291:21085–95.

82. Arosa FA. On the origin and function of human NK-like CD8+ T Cells: Charting new territories: Frontiers Media SA; 2018.

83. Aref S, Azmy E, El-Gilany AH. Upregulation of CD200 is associated with regulatory T cell expansion and disease progression in multiple myeloma. Hematol Oncol. 2017;35:51–7.

84. Xu X, Han L, Zhao G, Xue S, Gao Y, Xiao J, et al. LRCH1 interferes with DOCK8-Cdc42-induced T cell migration and ameliorates experimental autoimmune encephalomyelitis. J Exp Med. 2017;214:209–26.

85. Kaur S, Singh SP, Elkahloun AG, Wu W, Abu-Asab MS, Roberts DD. CD47-dependent immunomodulatory and angiogenic activities of extracellular vesicles produced by T cells. Matrix Biol. 2014;37:49–59.

86. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun. 2015;6:8971.

87. Oh S, Song S. Bayesian Modeling Approaches for Temporal Dynamics in RNA-seq Data. In: New Insights into Bayesian Inference: BoD--Books on Demand; 2018. p. 7.

88. Lanciotti M, Masieri L, Raspollini MR, Minervini A, Mari A, Comito G, et al. The role of M1 and M2 macrophages in prostate cancer in relation to extracapsular tumor extension and biochemical recurrence after radical prostatectomy. Biomed Res Int. 2014;2014:486798.

Mangiola *et al. BMC Cancer*     (2021) 21:846

Page 18 of 18

89. Wang Y-Q, Zhu Y-J, Pan J-H, Xu F, Shao X-G, Sha J-J, et al. Peripheral monocyte count: an independent diagnostic and prognostic biomarker for prostate cancer - a large Chinese cohort study. Asian J Androl. 2017;19:579–85.

90. Lundholm M, Hägglöf C, Wikberg ML, Stattin P, Egevad L, Bergh A, et al. Secreted Factors from Colorectal and Prostate Cancer Cells Skew the Immune Response in Opposite Directions. Sci Rep. 2015;5:15651.

91. Zhang E, Dai F, Mao Y, et al. Differences of the immune cell landscape between normal and tumor tissue in human prostate. Clin Transl Oncol. 2020;22(3):344–50.

92. Gannon PO, Poisson AO, Delvoye N, Lapointe R, Mes-Masson A-M, Saad F. Characterization of the intra-prostatic immune cell infiltration in androgen-deprived prostate cancer patients. J Immunol Methods. 2009;348:9–17.

93. Nonomura N, Takayama H, Nakayama M, Nakai Y, Kawashima A, Mukai M, et al. Infiltration of tumour-associated macrophages in prostate biopsy specimens is predictive of disease progression after hormonal therapy for prostate cancer. BJU Int. 2011;107:1918–22.

94. Zhang Y, Du W, Chen Z, Xiang C. Upregulation of PD-L1 by SPP1 mediates macrophage polarization and facilitates immune escape in lung adenocarcinoma. Exp Cell Res. 2017;359:449–57.

95. Smith P, Walsh CM, Mangan NE, Fallon RE, Sayers JR, McKenzie ANJ, et al. *Schistosoma mansoni* worms induce anergy of T cells via selective up-regulation of programmed death ligand 1 on macrophages. J Immunol. 2004;173:1240–8.

96. Singhal S, Stadanlick J, Annunziata MJ, et al. Human tumor-associated monocytes/macrophages and their regulation of T cell responses in early-stage lung cancer. Sci Transl Med. 2019;11(479):eaat1500.

97. Idorn M, Køllgaard T, Kongsted P, Sengeløv L, Thor SP. Correlation between frequencies of blood monocytic myeloid-derived suppressor cells, regulatory T cells and negative prognostic markers in patients with castration-resistant metastatic prostate cancer. Cancer Immunol Immunother. 2014;63:1177–87.

98. Chi N, Tan Z, Ma K, Bao L, Yun Z. Increased circulating myeloid-derived suppressor cells correlate with cancer stages, interleukin-8 and -6 in prostate cancer. Int J Clin Exp Med. 2014;7:3181–92.

## Publisher's Note