# Dissection of prostate tumour, stroma and immune transcription reveals a key contribution by the microenvironment to disease progression

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

[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]Institute of Microbiology of the Czech Academy of Sciences, Prague, Czech Republic.

[5]Molecular Immunology Division, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia

[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]Australian Prostate Cancer Research Centre Epworth, Richmond, Victoria, Australia

[9]Flow cytometry facility, The Walter and Eliza Hall Institute of Medical Research, Parkville, Victoria, Australia

[10]Cancer Program, Biomedicine Discovery Institute, Monash University, Victoria, Australia; Department of Anatomy and Developmental Biology, Monash University, Victoria, Australia.

[11]The Olivia Newton-John Cancer Research Institute, Heidelberg, Melbourne, Australia

[12]Department of Urology, Frankston Hospital, Frankston, Victoria, Australia

[13]Peter MacCallum Cancer Centre, Melbourne, VIC 3000, Australia.

[14]Department of Medical Biology, University of Melbourne, Melbourne, Victoria, Australia.

[15]Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria, Australia.

[16]School of Mathematics and Statistics, University of Melbourne, Melbourne, VIC 3010, Australia.

[*]These authors contributed equally.

To whom correspondence should be addressed:
Christopher Hovens (chovens@unimelb.edu.au), Anthony T Papenfuss (papenfuss@wehi.edu.au), Niall Corcoran (con@unimelb.edu.au).

## 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 and the development a Bayesian inference model for continuous differential transcript abundance analyses permitted definition of the transcriptional landscape of the prostate cancer microenvironment across the disease progression spectrum. An important role of monocytes and macrophages in prostate cancer progression and disease recurrence was reinforced by both our transcriptional landscape findings and by differential tissue composition analyses.

*Conclusions*

This study contributes to understanding of monocyte-derived recruitment in primary prostate cancer, and supports a clear direction for further investigation into mechanisms of the immune system that contribute to disease progression.

## 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. Our inability to discern progressive disease at an early stage leads to substantial overtreatment of localised disease, as well as significant costs clinically for the patient and economically for 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, which have been consistently associated with disease recurrence [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 characterization, it remains unclear what mechanisms drive prostate cancer progression in most patients [7]. However, it is hypothesised 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 (FACS) 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] (EMT) 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 combined cell-type enrichment and ultra-low-input RNA sequencing using an optimised protocol, which allowed probing of four key cell types across 13 fresh prostate tissue

and spanning a wide spectrum of clinical disease based on the Cancer of the Prostate Risk Assessment Post-Surgical (CAPRA-S) score. The CAPRA-S score is used clinically to estimate risk of prostate cancer recurrence after prostatectomy. It ranges from 1 to 12 and is calculated for each individual patient based on pre-surgical PSA level, the pathologic Gleason score, the presence or absence of positive surgical margins, of extracapsular extension, of seminal vesicles extension, of regional lymph nodes extension. Motivated by the pseudo-continuous properties of 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 was able to model changes in transcription without *a priori* patient risk stratification, and robustly map transcriptional change events to cancer risk states. Using the TCGA prostate cancer cohort [7], we show that the microenvironment transcriptional signatures identified are clinically relevant. Several hallmarks of prostate cancer were identified among the signatures of cell-surface and secreted protein coding genes. An emergent signature for monocyte-derived cell recruitment was identified and tested on the TCGA dataset with two orthogonal analyses.

## Results

*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 cell populations was guided by their predominant role in prostate and other cancers progression [15–19]. Technical and practical experimental limitations prevented the consideration of 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 sequenced (Fig. S1), generating a median of 22 million reads per library (Fig. S2). Conservative cell-type purity estimation using Cibersort and a Bayesian estimator (see Material and Methods), demonstrated a high overall enrichment in the four cell type categories: 87% for epithelial samples, 76% for fibroblasts, 69% for myeloid (53% for neutrophils, 11% for monocytes, 5% for

macrophages), and 85% for T cells (Fig. S3). On average across the four cell types, 40% of genes had 0 sequenced reads in more than half of samples and were removed from further analysis.

*A new model for analysis of differential transcription*

Dimensionality reduction (multidimensional scaling; MDS in Materials and Methods) of the filtered transcript abundance revealed. A clear gradient in risk score was seen for epithelial and fibroblasts (Bonferroni adjusted p-value of 0.019 and 0.0097 respectively), while a weaker pattern was apparent for myeloid and T cells (Bonferroni adjusted p-value of 0.030 and 0.37 respectively); possibly due to the greater heterogeneity of the two immune cell populations compared to epithelial and fibroblasts.

We performed differential gene transcript abundance analysis for each cell type independently, seeking associations between transcript abundance across subjects with CAPRA-S risk score treated as pseudo-continuous variable. In order to perform differential transcription 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 robust generalised sigmoid regression. 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 the *a priori* stratification of patients in low-/high-risk groups based on an arbitrary threshold. In principle, the use of a generalised sigmoid function permits modeling linear, exponential and sigmoid-like trends of transcriptional alterations (Fig. 1B and 1C); however, to provide robust modeling for RNA sequencing data we re-parameterized the generalised sigmoid function to better suit the numerical properties of transcript abundance (Fig. 1B; Materials and Methods). In addition to robustness, the sigmoid function allows the mapping of each differential transcriptional event with a clinical risk state, effectively providing a new developmental dimension to the analyses. This is possible because the inflection point represents the CAPRA-S risk score at which the transcriptional alteration is most pronounced. The location of 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 inference of transcriptional change, an average of 10% of genes were removed based on the posterior predictive check across all samples [20], as not satisfying the assumptions of our model (Table 1; Materials and Methods). A total of 1,626 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 common practices in Bayesian statistics [21]; Table 1; Supplementary file 1). The distributions of differential transcription events along the CAPRA-S risk score range were 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).

*Gene annotation and pathway analyses*

We next examined the overlap with known cancer-related genes, 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 (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 (Table 1). To generate a high-level summary the biological information contained in our data, we performed a cell-type agnostic, ensemble gene set enrichment analysis (EGSEA; Supplementary file 2). For epithelial cells, the top-ranked hallmarks were identified as cholesterol homeostasis, apoptosis, androgen response and TNF-α signaling. The most highly ranked gene ontology categories were negative regulation of glycoprotein biosynthesis and response to parathyroid hormone and the main KEGG pathway was fatty acid degradation. For fibroblasts, hypoxia, TNF-α signaling, and TGF-β signaling were the top ranked hallmarks; glycosphingolipid biosynthesis, ascorbate and aldarate metabolism and oxidative phosphorylation were the main KEGG pathways. For myeloid cells, pancreatic beta cells, myogenesis [22] and spermatogenesis were indicated as hallmarks; neuron cell adhesion [23], male genitalia development and regulation of myoblast proliferation were the top-ranked gene ontology categories; and nicotine addiction, ascorbate and aldarate metabolism and steroid hormone biosynthesis [24] were the main KEGG pathways. For T cells, the top-ranked hallmarks included TNF-α signaling, apoptosis and Wnt/β-catenin

signaling. The top gene ontology categories were regulation of macrophage differentiation, hepatocyte proliferation and negative regulation of B cell apoptosis.

*Stratification of patients by cell-type specific signatures*

We next considered if any cell-type specific transcriptional signatures were clinically relevant in tumour progression by testing for association with disease-free survival on an expanded independent patient cohort with both disease-free survival and CAPRA-S score information. After filtering for cell-type specific gene transcripts (see Material and Methods) and dimensionality reduction based removal of highly transcriptionally correlated genes (Pearson correlation > 0.6; see Material and Methods) of the 1,626 differentially transcribed genes, we tested a total of 188 genes, 18 of which were significantly associated with disease-free survival independently of the CAPRA-S risk score (p-value < 0.05; multivariate cox regression; Fig. 2A). For most of these genes, the transcript abundance was positively associated with the disease-free survival. These genes encode for both intracellular and cellular surface/secreted proteins. For example, within the epithelial tumour cell population, Sphingolipid Transporter 2 (SPNS2), Cystatin E/M (CST6), and Kidney Associated Antigen (KAAG1) were the most associated with survival. Within the fibroblast cell population, an exonuclease and RNase (EXO1), an uncharacterised protein (C1orf220) and a transcription factor (IRF8) that promotes transcription of a pro-fibrotic gene (neurofibromin) [25] were the top associated genes. Within the myeloid cell population, the most associated genes were nuclear hormone receptor (NR2C2), a Phosphoinositide-3-Kinase Regulatory Subunit (PIK3R6) linked to aldosterone synthesis and secretion [26] and collagen COL1A2. Within the T cell population, the top associated genes were a Ribonucleotide Reductase Regulatory Subunit (RRM2), Complement Factor Properdin (CFP) and a TATA-Box Binding Protein Associated Factor (TAF13).

As expected, patient stratification based on the number of deleterious dysregulation events (i.e., abundant deleterious genes or lack of beneficial genes; see Materials and Methods for further details) revealed an inverse relationship with disease-free survival (Fig. 2B, left). Dichotomising patients in the top and bottom third of deleterious events (Fig. 2B, right), all four cell types were able to group patients based on their survival outcome (p-value $4.3\times10^{-4}$; $1.0\times10^{-4}$; $2.3\times10^{-2}$ and $3.2\times10^{-4}$ for epithelial, fibroblast, myeloid and T cell respectively). The integration of

microenvironmental-based classifications (including fibroblast, myeloid and T cell, using the majority rule) significantly improved upon CAPRA-S risk score alone (dotted grey line; Fig. 2C-bottom), identifying a high survival patient cohort (blue trend; Fig. 2C-bottom). Importantly, this signature outperformed the combination of epithelial classification and CAPRA-S risk score (Fig. 2C-top), characterised by an overall good stratification ability but a poor stratum size balance.

While our goal was to gather evidence for clinically relevant information in the microenvironmental transcriptional signature, rather than building an improved classifier, we performed a comparison with a signature selected through a naive differential-transcription based approach (using edgeR [27]) on the whole tissue TCGA primary prostate dataset. Although both derived signatures performed similarly, the integrated non-epithelial signature identified a patient group with substantially tighter confidence intervals (Fig. S5).

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

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 have direct influence on 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 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).

To enable 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 [28] with more than 1500 scientific articles (Supplementary file 3). This allowed us to identify six recurring hallmarks of cancer (Fig. 3): (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 genes. This balance appears to be dynamic along the disease progression course. The epithelial cell

migration hallmark includes three main functional clusters: tissue remodeling, tissue fibrosis and 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 signaling appears to be sustained along the whole disease progression, where a gene alteration signature of platelet recruitment linked with endothelial cell migration is 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, which is a key molecular hallmark in prostate cancer [29]. Within this set, the most frequently recurring metabolite that is linked with differentially transcribed genes is cholesterol. While for most hallmarks all four cell types contributed similarly to the signatures, a clear bias was present for cancer cell growth, osteogenesis and hormone modulation signatures, which were enriched with genes that were 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 signatures from the four cell types tracks with the risk score covariate (Fig. 4). While the magnitude of the pro-inflammatory transcriptional signature remains roughly constant through the risk range, with 18 genes for CAPRA-S risk score $\leq 2$ and 14 for CAPRA-S risk score $>2$; the anti-inflammatory signature significantly expands (p-value 0.015; t-test) in 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 were involved in the recruitment of monocytes and macrophages [30–37] (highlighted in yellow in Fig. 4A). These include CAMKK2 [38], ORM1 [39] and DCN [40,41] in epithelial; IL2RB [42–44], ICAM4 [45,46], DCN [40,41] and MDK [47,48] in myeloid cells; and CSF1 [30] and PDGFD [49] in T cells. In addition, we identified a known fibroblast-macrophage chemotactic interaction

including the regulation of the cytokines CXCL10 [33], CXCL14 [32] and the receptor SLAMF1 [34,35] for fibroblasts—with COL1A2 [50] and CYR61 [51] (for CAPRA-S 6-8) altered in myeloid cells known to function as co-stimulatory loop. A smaller cluster of genes were linked with T cell recruitment and inflammation, including CFP [52], IL24 [53], PROK2 [54], SELL [55]. Interestingly, epithelial cells upregulate a cluster of receptor genes normally involved in antigen recognition and presentation in immune cells [56], including an MHC class II cell surface receptor (i.e., HLA-DRB5) and three Fc receptors (i.e., FCER1G, FCGR1A and FCGR2A).

The anti-inflammatory signatures target a more heterogeneous set of cell types then the pro-inflammatory signature. Monocyte-derived cells were mainly targeted by genes that were differentially transcribed in epithelial and myeloid cells. These included the receptor genes SPNS2 [57,58], IL10RA [59] and ICAM5 by epithelial cells; and the receptor genes CPM [60] and PEX13 [61] and the secreted protein genes FN1 [62] and ANGPT2 [63] by myeloid cells. Another cluster of genes targets predominantly T cells, including AREG [64,65], CD200 [66], LRCH1 [67], CD47 [68]. 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 (DTC) analysis (i.e. a test for difference in cell-type abundance between conditions) based on an independent methodology and independent data. We used a hierarchical Bayesian inference model [69] on an independent cohort of 134 patients from the primary prostate cancer TCGA dataset [7] that included both disease-free survival and CAPRA-S score information. This algorithm used a collection of 250 curated publicly available transcriptional profiles (including BLUEPRINT [70], ENCODE [71], GSE89442 [72] and GSE107011 [73]), encompassing transcriptional signatures for 18 hierarchical cell-types (of which nine major cell types were tested for differences) and used those reference signatures to understand the contribution of each cell type to the observed mixed transcriptional signal. This analysis provides tissue composition estimates as well as their association with risk score [69]. Overall, we estimated a median of 88% for epithelial cellular fraction across samples

(consistent with public literature [74], Fig. S6), 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. 5A).

In order to test whether the enrichment in monocyte-derived cells is clinically relevant, we generated Kaplan–Meier curves using the estimated cell-type abundances. The stratification of patients based on the extent of monocyte-derived cells infiltration revealed significant separation in survival (Fig. 5B). Within the primary prostate cancer TCGA cohort selected, the disease-free survival of 50.5% of the patients seemed to be affected by high monocyte-derived infiltration within the primary tumour. For comparative purposes, we tested patient stratification for the other cell types included in the model. Only granulocytes and B cells (with the poor outcome cohort including only few patients) showed a significant negative association, while no other significant associations were detected for other cell types including epithelial, endothelial, fibroblasts, and immune cell types such as including T cells and natural killers (Fig. 5C).

## 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. It is clear that non-malignant cells within the tumour microenvironment also contribute to cancer progression, and that they are often modulated by cancer cells toward pro-tumorigenic behaviours. In this study, we found concurrent transcriptional changes in epithelial, fibroblast, myeloid and T cells along the CAPRA-S risk score range, adding to understanding of the molecular landscape of primary prostate tumour microenvironment.

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 CAPRA-S risk score, we developed a novel statistical inference method (TABI), which permitted modeling of transcript

abundance natively on continuous factors of interest with a limited number of parameters, avoiding loss of information due to the dichotomization of patients into risk score groups. As suggested by multidimensional-scaling plots and supported by our inference, transcriptional change events were 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 TABI is based on, can be extrapolated for a wide range of applications.

Our study supports the evidence that molecular profiles of non-malignant cell populations within the tumour microenvironment hold clinically relevant information. Although a limitation of this analysis is the multi-cell-type mixed signal within TCGA RNA sequencing samples, it is clear that a gene selection based on our cohort provides improved stratification even compared to a direct prioritisation based on differential transcript abundance analysis on all genes, proving the clinical relevance of the data generated in this study. For example, our integrated non-epithelial signature improved patient stratification based on clinical variables alone and performed better than a signature derived solely from the tumour epithelial cells. 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 transcribed genes linked to monocyte-derived cells recruitment and modulation. The association of monocyte-derived cell recruitment with increased risk score was reflected in an orthogonal differential tissue composition analysis on an expanded independent cohort. 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 to clinical variables. This approach is particularly relevant considering the substantial noise associated with such inferences. To test the clinical significance of quantifying monocyte-derived cell numbers within the tumour mass, we performed survival analyses using the inferred cell-type proportions from the large, independent TCGA prostate cancer dataset. 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 [75].

In prostate cancer, specific and overall survival analyses have identified an elevated monocyte count as an independent prognostic factor for poor outcome [76–79]. 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 [80].

## Conclusions

In prostate cancer, there has been limited benefit observed through the unselected use of novel immune checkpoint inhibitors based on T cell receptor blockade (e.g., PD-1, PD-L1 and CTLA-4) [81]. Such failure may in part be driven by our limited understanding of the dynamic interplay between immune components of the microenvironment and tumour cells. 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. This study provides a clear direction for further investigation into mechanisms of the immune system (particularly, monocyte-derived cells) that contribute to disease progression; for example, changing the hormonal and growth-factor homeostasis through a sustained inflammatory state. Both the methodological advances and the novel findings presented in this study provide a research framework for improved immune interventions.

## Declarations

*Ethics approval and Consent to participate*

The collection and use of tissue for this study had Melbourne Health institutional review board approval and patients provided written informed consent (Melbourne Health Local Project Number: 2016.087).

*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 has been deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAD00001004948.

*Acknowledgements*

We wish to thank Dr. Stephen Wilcox for technical assistance and are grateful to the staff of the Flow cytometry facility and Clinical Translational Centre (Walter and Eliza Hall Institute of Medical Research). We thank Professor Gail Risbridger (Monash University) for the great support with the initial phase of the study. We thank Dr. Damiano Spina (RMIT) for the support with machine learning. We thank Dr. Bob Carpenter and Prof. Andrew Gelman (Columbia University) for incredible mentorship in Bayesian inference, and all the Stan community at discourse.mc-stan.org for technical support. Most importantly, we thank all the patients who participated in the study.

*Author contributions*

SM conceived and designed the study, performed part of the cellular biology procedures, implemented the statistical methods and performed data analysis and visualisation, under the supervision of AJC, NH, NMC, ATP and CMH. MK, JP and PD contributed with sample harvesting. PM, FSFG, DB, NL, CN, MLP and SIM contributed to the cellular biology procedures; and PM and BP contributed to the molecular procedures. MM contributed to statistical model implementation. All authors contributed in manuscript writing.

*Competing interests*

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

*Funding*

SM was supported by the David Mayor PhD Scholarship from the Prostate Cancer Research Foundation, and by the Galli Next Generation Cancer Discoveries Initiative. KC was supported by a Postgraduate Medical Research Scholarship from the Prostate Cancer Research Fund, and the Australian Government Research Training Program Scholarship provided by the Australian Commonwealth Government and the University of Melbourne. NDH was a recipient of a Melanoma Research Grant from the Harry J Lloyd Charitable Trust. FSFG was supported by a grant #1158085 awarded through the Priority-driven Collaborative Cancer Research Scheme and

funded by Cure Cancer Australia with the assistance of Cancer Australia. MM was supported by ELIXIR CZ research infrastructure project (MEYS Grant No: LM2015047) including access to computing and storage facilities.. NMC was supported by a David Bickart Clinician Research Fellowship from the Faculty of Medicine, Dentistry and Health Sciences at the University of Melbourne, as well as Movember — Distinguished Gentleman's Ride Clinician Scientist Award through Prostate Cancer Foundation of Australia's Research Programme. ATP was supported by the Lorenzo and Pamela Galli Charitable Trust and by an Australian National Health and Medical Research Council (NHMRC) Program Grant (1054618) and NHMRC Senior Research Fellowship (1116955). The research benefitted by support from the Victorian State Government Operational Infrastructure Support and Australian Government NHMRC Independent Research Institute Infrastructure Support.

## Material and methods

*Tissue sampling and processing*

Following the prostatectomy of 13 patients, ranging 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), a four millimeter tissue core was collected from the prostate tumour site, conditional to histopathological verification [82,83]. If not otherwise specified, all procedures were carried out at 4 °C. Tissue blocks were washed in Phosphate-buffered saline (PBS) solution for 2 minutes and minced for 2 minutes 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 (4g), through three steps of 5, 10 and 10 minutes of duration, with the final 3 minutes 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 300gfor five minutes, with the supernatant collected and the cell pellet resuspended into 1 ml 2% PBS-serum prior to labeling (Fig. S1).

*Antibody labeling, 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), Thy-1 (CD90) PerCP-Cy5.5 (BD Biosciences, USA). All antibodies were used at concentrations according to manufacturer's recommendations and incubated for 30 mins at 4°C. Following labeling the cells were diluted to 5ml and centrifuged at 300g for 5 minutes. The supernatant was removed and the cell pellet was resuspended in dPBS-serum. Viability die (7AAD) was added to the suspension to a final concentration of 5µg/ml. Epithelial, fibroblasts, myeloid and T cells were sorted using a FACS Aria III cell sorter (Becton Dickinson San Jose, Ca). The cell sorting strategy utilized a robust 3 stage design: (i) a series of gates based or forward and side scatter to exclude debris, cell clumps and doublets. (ii) a gate to exclude all dead cells and

(iii) combination of the fluorescent antibodies to allow purification of the above cell types. The four cells 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 1.5ml 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 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 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 buffers, RWT and RPE, before eluting the total RNA with 14µl of RNase-free water. RNA estimation was carried out using Tapestation (Agilent).

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), according to manufacturer's protocol. The first-strand cDNA synthesis utilised 3' SMART-Seq CDS Primer II A and 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 minute, 12 cycles of 98 °C 10 seconds, 65 °C 30 seconds and 68 °C 3 minutes; 72 °C for 10 minutes 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 end repair process, the addition of 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 [84]. 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 [85] with default settings. The quality control on the alignment was performed with RNA-SeQC [86]. For each sample, the gene transcription abundance was quantified in terms of nucleotide reads per gene (read-count) using FeatureCounts [87] 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 [88] with default settings.

*Statistical inference*

Changes of transcriptional levels along CAPRA-S risk score [89] were estimated independently for each cell 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 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 [90] (Eq.1), but re-parameterised to improve numerical stability (Eq. 2). In particular, the standard parameterisation suffers from non-determinability issue in case the slope is close to zero; furthermore, in case of an exponential-like trend the upper plateau is not supported by data and tends to infinity.

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

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

The new parameter $y_0$ represents the intercept on the y axis, η represents the point of inflection on the x-axis, β 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 [91] was used to encode the joint probability function of the model (Eq. 3), partitioning the transcriptomic dataset in 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) of the raw transcript abundance. In order to account for diverse sequencing depths across samples a sample-wise normalisation parameter was added to the negative binomial expected value. To increase the robustness of the inference of changes in transcription and better help to anchor data from different samples for normalisation, the slope parameter for the main covariate ($\beta_1$) was subject to a regularised horseshoe [92] prior. The role of this prior is to impose a sparsity assumption on the gene-wise transcriptional changes; that is, most genes are not differentially transcribed. The overall distribution of the gene intercepts

follows a gamma probability function. The statistical model is defined by the following joint probability density.

$$(3)\ P(\gamma)\,P(\delta)\,P(\sigma)\,P(\eta)\,P(\xi)\,P\left(\dot{\beta}|\xi\right)$$

$$\left(\prod_{r=2}^{R} P(\beta_r|\sigma)\right)\left(\prod_{g=1}^{G} P\left(y_{\circ g}|\gamma',\gamma''\right)\right)$$

$$\left(\prod_{g=1}^{G}\prod_{s=1}^{S} P\left(Y_{g,s}|\hat{Y},\delta,\omega\right)\right)$$

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

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

$$(6)\ \beta_{g,1} \sim RegHorseshoe\left(...\right)$$

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

$$\sigma_k \sim HalfN\left(0,1\right)$$

$$(8)\ \ y_{0g} \sim Gamma\left(\gamma_1+1,\gamma_2\right)$$

$$\gamma_i \sim Exponential(1)$$

$$\omega \sim Gamma(1.02,2)$$

$$(9)\ \ \eta_g \sim N\left(0,1\right)$$

$$\delta_t \sim N(0,1); \sum \delta_t \sim N\left(0, 0.001 * T\right)$$

where Y represents raw transcript abundance, $\hat{\underline{Y}}$ represents the expected values of transcript abundance and X represents the design matrix (with no intercept term and with 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 re-parameterised sigmoid function (Eq. 2), while γ represent 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 $\beta$) is defined by the hyperparameter list $\xi$ [92] (i.e., nu_local = 1; nu_global = 1; par_ratio = 0.8; slab_df = 4; slab_scale = 0.5), while $\sigma$ represents the standard deviations of the other factors (in our case only the batch). The algorithm multidimensional scaling (MDS) [93] 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 [20]). Among the well fitted genes, those for which the 0.95 credible interval 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 common 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 common practices in Bayesian statistics [21].

In order to interpret the inflection points over the CAPRA-S risk score (i.e., the point of 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 (in order 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 ($\dot{X}$) 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 re-parameterised sigmoid function (Eq. 10).

$$(10)\ \dot{X} = \frac{\beta_1\eta - log\left(e^{\frac{y_0}{2}}\sqrt{e^{y_0\eta}+1} - 1\right)}{y_0}$$

Genes were functionally annotated with gene ontology categories [28] using BiomaRt [94]. Furthermore, genes were functionally annotated with the protein atlas database [95] 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 by: (i) a first deconvolution step that infers tissue composition from whole tissue gene transcript abundances based on reference transcriptional profiles of single cell types; and (ii) integrated beta regression on the inferred proportions. The use of a Bayesian inference framework (GitHub: stemangiola/ARMET@v0.7.1) allowed to transfer the uncertainty from (i) to (ii). The probabilistic framework Stan **[91]** was used to encode the joint probability function of the model. The 0.95 credible interval of the posterior distributions was used for result interpretation. This algorithm bases the inference on 18 hierarchical cell-types defined by a collection of 250 curated publicly available transcriptional profiles (including BLUEPRINT **[70]**, ENCODE **[71]**, GSE89442 **[72]** and GSE107011 **[73]**).

## Survival analyses

The study of the association between gene transcription abundances and disease-free survival was performed using survfit from the survival R package [96] in a multivariate fashion, including CAPRA-S risk score as covariate. These analyses were performed on the TGCA primary prostate cancer dataset [7] using DFS_MONTHS as factor of interest and is_recurred as censor variable. Transcript abundance was normalised across TCGA samples using the TMM normalisation [97]. The gene features within the TCGA cohort were selected among the differentially transcribed genes performed on the original cell-type specific cohort (n=52). In order to focus on the contribution of each of the four specific cell types in analysis, the differentially transcribed genes were filtered for those which transcription was higher than the other three cell types. For example, genes for fibroblasts were considered for further analysis if the normalised abundance was higher than for epithelial cells, higher than T cells, and higher than Myeloid cells. For example, genes for fibroblasts were considered for further analysis if the normalised abundance was higher than for epithelial cells, higher than T cells, and higher than Myeloid cells. Dimensionality reduction based on removal of highly transcriptionally correlated genes was performed on the filtered genes in

order to avoid excessive redundancy for the multivariate Cox regression. This procedure was carried on independently for each of the four cell-type specific signatures. Briefly, gene pairs with highly correlated (Pearson correlation > 0.6) transcript abundance within the TCGA dataset were identified, and genes with the highest transcriptional change within the original cohort (n=52; observing the distance of the credible interval of the slope from zero) were selected at each comparison., based on the inferred gene means (using TABI). Similarly, only samples with high cell-type abundance for each of the four cell types (i.e., above median, inferred with a Bayesian inference model) were selected for the analyses.

Kaplan–Meier curves were drawn using the R survminer [98] and survival [99] packages [98]. Kaplan–Meier curves for the cell type abundance were performed on the inferred cell-type proportions by a Bayesian inference method (see Material and Methods section, differential tissue composition analyses sub-section). For patient stratification (Fig. 2B and 2C), deleterious transcriptomic changes were identified. Genes were classified as deleterious if their transcription abundance was directly correlated with CAPRA-S risk score in our cohort and inversely with disease-free survival in the TCGA database, and therefore possibly contributing to disease progression. On the contrary, genes were classified as beneficial if their transcription abundance is inversely correlated with CAPRA-S risk score in our cohort and directly with disease-free survival in the TCGA database and therefore possibly playing a anti-cancer role. Deleterious regulations were defined integrating those two concepts, as being either a deleterious gene that is highly transcribed in one patient (defined by the binary split as previously described), or a beneficial gene that lowly transcribed in one patient (defined by the binary split as previously described).

## 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. Ippolito L, Morandi A, Taddei ML, Parri M, Comito G, Iscaro A, et al. Cancer-associated fibroblasts promote prostate cancer malignancy via metabolic rewiring and mitochondrial transfer. Oncogene [Internet]. 2019; Available from: http://dx.doi.org/10.1038/s41388-019-0805-7
16. 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. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2013;4:1795.
17. 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.
18. Sharma A, Suleyman N, Jones O, Vasdev N. Immunotherapy in Urological Tumors. Rev Urol. 2019;21:15–20.
19. 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.
20. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Bayesian Data Analysis, Third Edition. CRC Press; 2013.
21. Gelman A, Hill J, Yajima M. Why We (Usually) Don't Have to Worry About Multiple Comparisons. J Res Educ Eff. Routledge; 2012;5:189–211.
22. Sloboda DD, Brown LA, Brooks SV. Myeloid Cell Responses to Contraction-induced Injury Differ in Muscles of Young and Old Mice. J Gerontol A Biol Sci Med Sci. 2018;73:1581–90.
23. Bitzer-Quintero OK, González-Burgos I. Immune system in the brain: a modulatory role on

dendritic spine morphophysiology? Neural Plast. 2012;2012:348642.
24. Reboldi A, Dang E. Cholesterol metabolism in innate and adaptive response. F1000Res [Internet]. 2018;7. Available from: http://dx.doi.org/10.12688/f1000research.15500.1
25. Zhu C, Saberwal G, Lu Y, Platanias LC, Eklund EA. The interferon consensus sequence-binding protein activates transcription of the gene encoding neurofibromin 1. J Biol Chem. 2004;279:50874–85.
26. Bateman A, Pearson WR, Stein LD, Stormo GD, Yates JR III, editors. The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Current Protocols in Bioinformatics. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2002. p. 1.30.1–1.30.33.
27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
28. 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.
29. 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.
30. 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.
31. 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.
32. Lu J, Chatterjee M, Schmid H, Beck S, Gawaz M. CXCL14 as an emerging immune and inflammatory modulator. J Inflamm . 2016;13:1.
33. 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.
34. 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.
35. 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.
36. Rovis TL, Brlic PK, Kaynan N, Lisnic VJ. Inflammatory monocytes and NK cells play a crucial role in DNAM-1–dependent control of cytomegalovirus infection. Journal of [Internet]. jem.rupress.org; 2016; Available from: http://jem.rupress.org/content/213/9/1835.short
37. O'Connell PA, Surette AP, Liwski RS, Svenningsson P, Waisman DM. S100A10 regulates plasminogen-dependent macrophage invasion. Blood. 2010;116:1136–46.
38. Racioppi L. CaMKK2: a novel target for shaping the androgen-regulated tumor ecosystem. Trends Mol Med. 2013;19:83–8.
39. 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.
40. 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.
41. 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.
42. 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.
43. ImmuNet [Internet]. [cited 2018 Aug 26]. Available from: http://immunet.princeton.edu/genes/detail/homo-sapien/IL2RB/
44. 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.
45. 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.
46. 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.
47. 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.

48. 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.
49. 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.
50. 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.
51. Bai T, Chen C-C, Lau LF. Matricellular protein CCN1 activates a proinflammatory genetic program in murine macrophages. J Immunol. 2010;184:3223–32.
52. Blatt AZ, Pathan S, Ferreira VP. Properdin: a tightly regulated critical inflammatory modulator. Immunol Rev. 2016;274:172–90.
53. Persaud L, De Jesus D, Brannigan O, Richiez-Paredes M, Huaman J, Alvarado G, et al. Mechanism of Action and Applications of Interleukin 24 in Immunotherapy. Int J Mol Sci [Internet]. 2016;17. Available from: http://dx.doi.org/10.3390/ijms17060869
54. 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.
55. 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.
56. Kobayashi KS, van den Elsen PJ. NLRC5: a key regulator of MHC class I-dependent immune responses. Nat Rev Immunol. 2012;12:813–20.
57. 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.
58. 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.
59. 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.

60. Denis CJ, Lambeir A-M. The potential of carboxypeptidase M as a therapeutic target in cancer. Expert Opin Ther Targets. 2013;17:265–79.
61. 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.
62. Pinto AR, Godwin JW, Rosenthal NA. Macrophages in cardiac homeostasis, injury responses and progenitor cell mobilisation. Stem Cell Res. 2014;13:705–14.
63. 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.
64. 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.
65. Arosa FA. On the Origin and Function of Human NK-like CD8+ T Cells: Charting New Territories. Frontiers Media SA; 2018.
66. 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.
67. 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.
68. 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.
69. Mangiola S, Stuchbery R, McCoy PJ, Chow K, Kurganovs N, Kerger M, et al. Androgen deprivation therapy promotes an obesity-like microenvironment in periprostatic fat. Endocrine Connections [Internet]. 2019; Available from: https://ec.bioscientifica.com/view/journals/ec/aop/ec-19-0029.xml
70. Stunnenberg HG, International Human Epigenome Consortium, Hirst M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. Cell. 2016;167:1897.
71. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489:57–74.

72. 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.
73. 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.
74. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. Nat Commun. 2015;6:8971.
75. 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.
76. 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.
77. 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.
78. Zhang E, Dai F, Mao Y, He W, Liu F, Ma W, et al. Differences of the immune cell landscape between normal and tumor tissue in human prostate. Clin Transl Oncol [Internet]. 2019; Available from: http://dx.doi.org/10.1007/s12094-019-02128-5
79. 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.
80. 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.
81. Rescigno P, de Bono JS. Immunotherapy for lethal prostate cancer. Nat Rev Urol. 2019;16:69–70.
82. 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.
83. 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.
84. Andrews S, Others. FastQC: a quality control tool for high throughput sequence data. 2010;
85. 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.
86. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. Bioinformatics. 2012;28:2184–5.
87. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.
88. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 2014;15:R46.
89. Cooperberg MR, Hilton JF, Carroll PR. The CAPRA-S score: a straightforward tool for improved prediction of outcomes after radical prostatectomy. Cancer [Internet]. Wiley Online Library; 2011; Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/cncr.26169
90. Richards FJ. A Flexible Growth Function for Empirical Use. J Exp Bot. Oxford University Press; 1959;10:290–301.
91. 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.
92. Piironen J, Vehtari A. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electron J Stat. The Institute of Mathematical Statistics and the Bernoulli Society; 2017;11:5018–51.
93. 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.
94. 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.
95. 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.
96. Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model. Springer Science & Business Media; 2013.
97. Robinson MD, Oshlack A. A scaling normalization method for differential expression

analysis of RNA-seq data. Genome Biol. 2010;11:R25.
98. Kassambara A, Kosinski M, Biecek P, Others. survminer: Drawing Survival Curves using'ggplot2'. R package version 0 3. 2017;1.
99. Grambsch PM, Therneau TM. Modeling survival data: extending the Cox model. Stat Biol Health. 2000;
100. Forbes SA, Beare D, Gunasekaran P, Leung K, Bindal N, Boutselakis H, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer [Internet]. Nucleic Acids Research. 2015. p. D805–11. Available from: http://dx.doi.org/10.1093/nar/gku1075

## Tables

| Cell type | Total genes | Genes filtered (zeros) | Genes filtered (PPC) | Differentially transcribed | | | Differentially transcribed in interface (curated annotation) | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Total (up/down) | Of which cancer genes | Of which PC genes | Total (up/down) | Of which cancer genes, consistent | Of wl genes |
| Epithelial | 21,618 | 5,408 | 189 | 171 (139/32) | 45 (26%) | 29 (64%) | 80 (67/13) | 35 (44%) | 23 (6 |
| Fibroblast | 21,510 | 7141 | 651 | 267 (156/111) | 27 (10%) | 9 (33%) | 97 (58/39) | 17 (18%) | 7 (41 |
| Myeloid | 22,507 | 13836 | 2695 | 900 (827/73) | 56 (6%) | 11 (20%) | 261 (238/23) | 32 (12%) | 10 (3 |
| T cell | 21,716 | 8807 | 540 | 288 (195/93) | 42 (15%) | 18 (42%) | 83 (55/28) | 26 (31%) | 15 (5 |

**Table 1:** Summary statistics of the differential transcription analysis, including 52 samples from 13 patients and 4 enriched cell types. 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 database [100] 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 dataset [100] or labelled as such in our manually curated cell-type specific database (Supplementary file 2).

## Figures

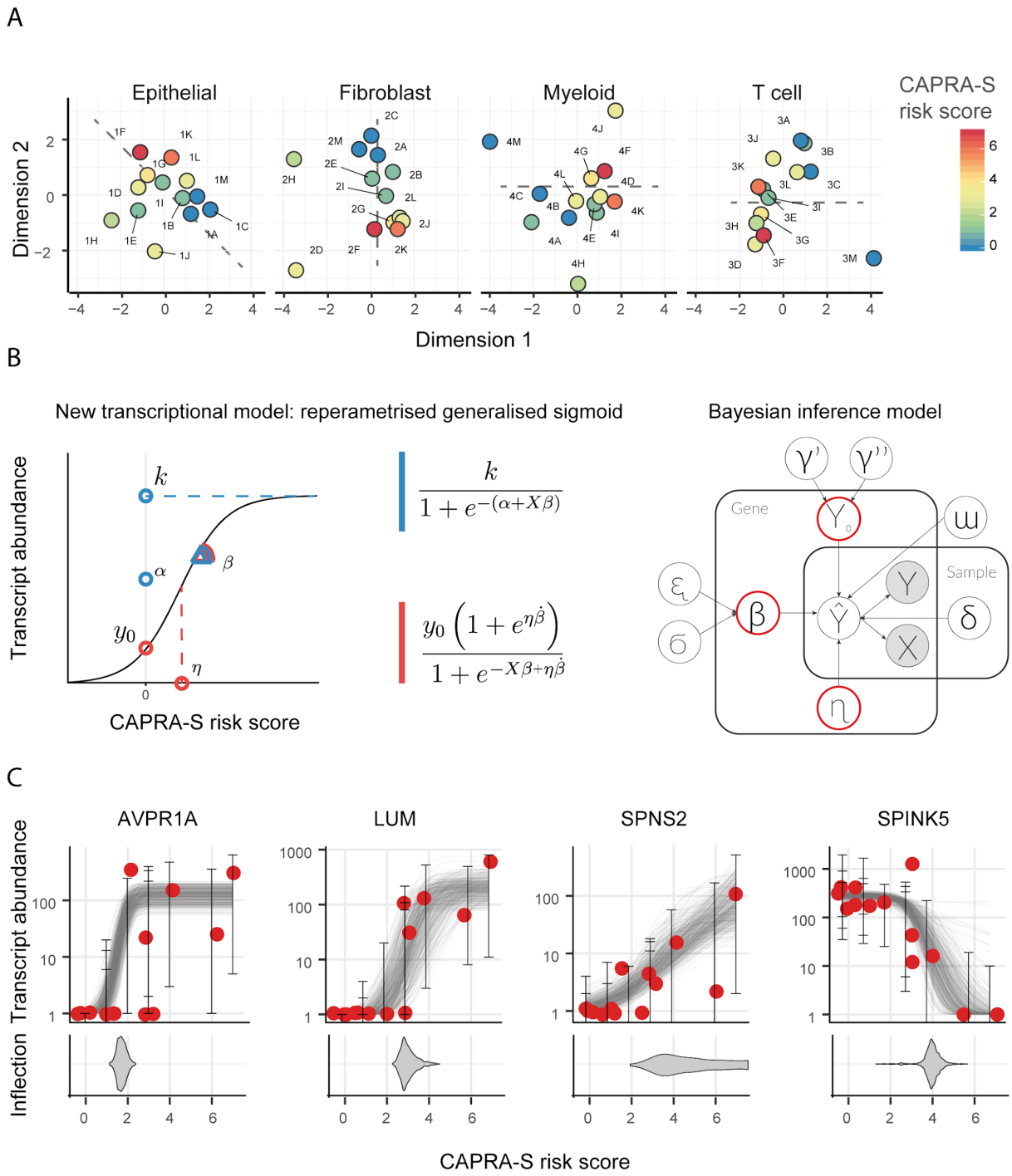


**Figure 1:** Data exploration and implementation of a model reflecting the continuous (i.e., non-discrete) relation between the CAPRA-S risk score and gene transcript abundance. **A** — Multidimensional scaling (MDS) plots of transcript abundance grouped by cell type. The colour

coding represents the CAPRA-S risk score. The risk score shows a correlation (tested with linear regression, lm function from R) across the first and/or the second dimension (Bonferroni adjusted p-value of 0.0187, 0.00971, 0.0306 and 0.367, respectively), particularly in epithelial and fibroblast cells. Alphanumeric-codes refer to patient identifier (Supplementary material). The dashed lines indicate the main direction of correlation between the first and/or the second dimension with CAPRA-S risk score. **B —** Representation of the re-parameterisation of the generalised sigmoid function and the resulting probabilistic model (Material and Methods). Left-panel: In blue are the three reference parameters for the standard parameterisation, in red are those for the alternative robust parameterisation. Right-panel: graphic representation of the probabilistic model TABI. **C —** Example for the continuous associations 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 section displays the inferred distribution of possible values (as posterior distribution) of the inflection point for each gene sigmoid trend.

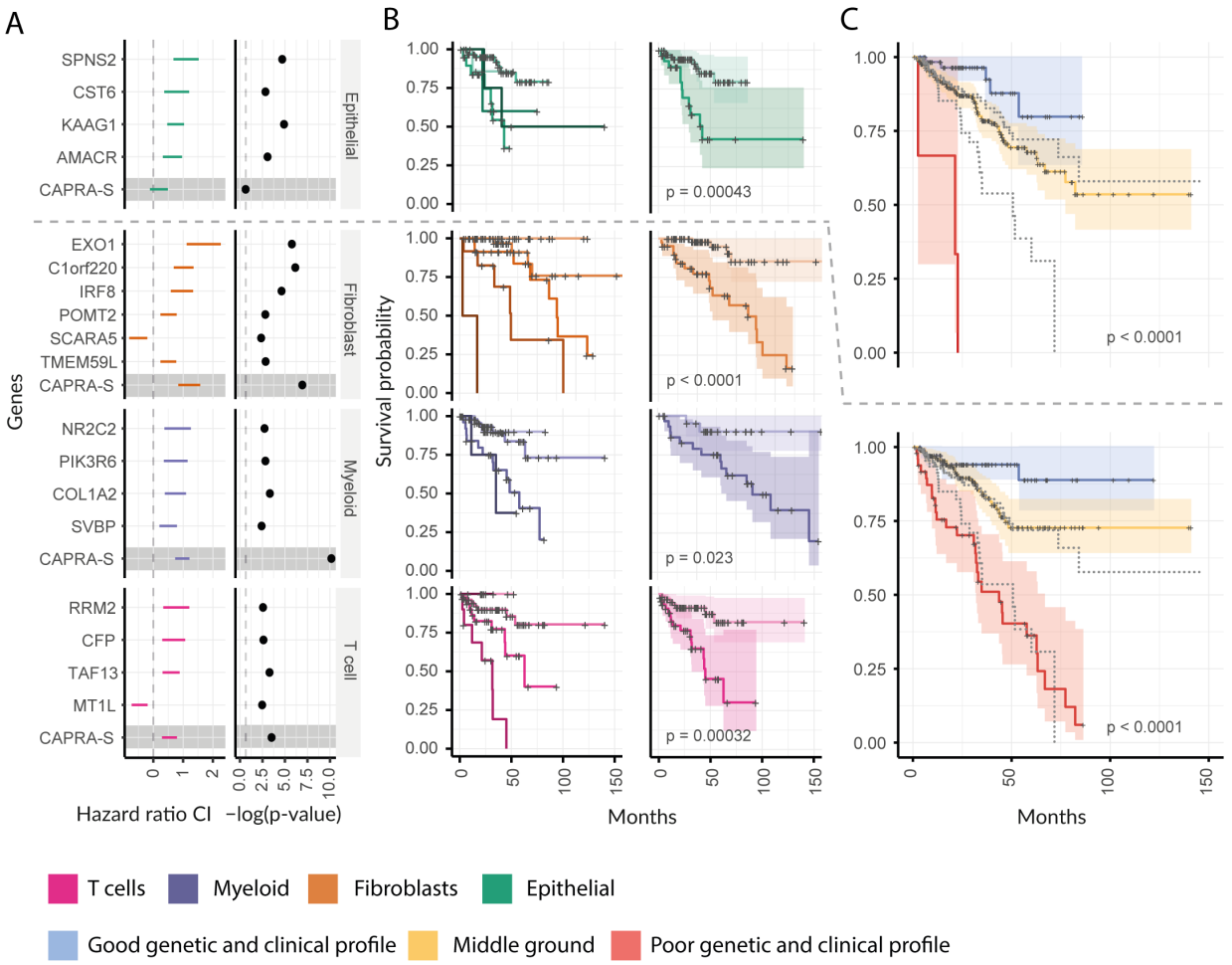


**Figure 2.** Ability of the cell-type specific signatures to improve patient stratification, compared to clinical variables alone or epithelial signature. Kaplan–Meier curves of patient outcomes in primary prostate cancer TCGA database (n = 134) after stratification based on epithelial and integrated non-epithelial signature (i.e., including fibroblasts, myeloid and T cells); compared with clinical variables alone (i.e., CAPRA-S risk score). **A** — Hazard ratio confidence interval and -log(p-value) of genes associated with disease-free survival independently (i.e., p-value < 0.05 with a multivariate cox regression test) from CAPRA-S risk score (displayed for comparison with the grey shade), grouped by cell types. The dashed line corresponds to the origin of the axis (left side) and the p-value threshold 0.05 (right side) **B left** — Kaplan-Meier curves of the associated genes, stratifying patients based on the number of deleterious transcriptomic changes (clear colour represent no deleterious changes, dark colours represent large number of deleterious changes; see Material and Methods for further details). **B right** — Kaplan-Meier curves of patients stratified by

the top and bottom third of number of deleterious transcriptional changes. **C —** Kaplan-Meier curves of the epithelial (top) or microenvironmental-based (bottom) stratifications, integrated with the clinical stratification (using CAPRA-S risk score). The blue line includes patient with co-occurring good transcriptomic and clinical classifications, while the red line includes patient with co-occurring bad transcriptomic and clinical classifications. The dashed lines represent the classification using CAPRA-S risk score alone (cut-off = 2). The confidence intervals of the Kaplan-Meier curves were calculated using survfit [96]

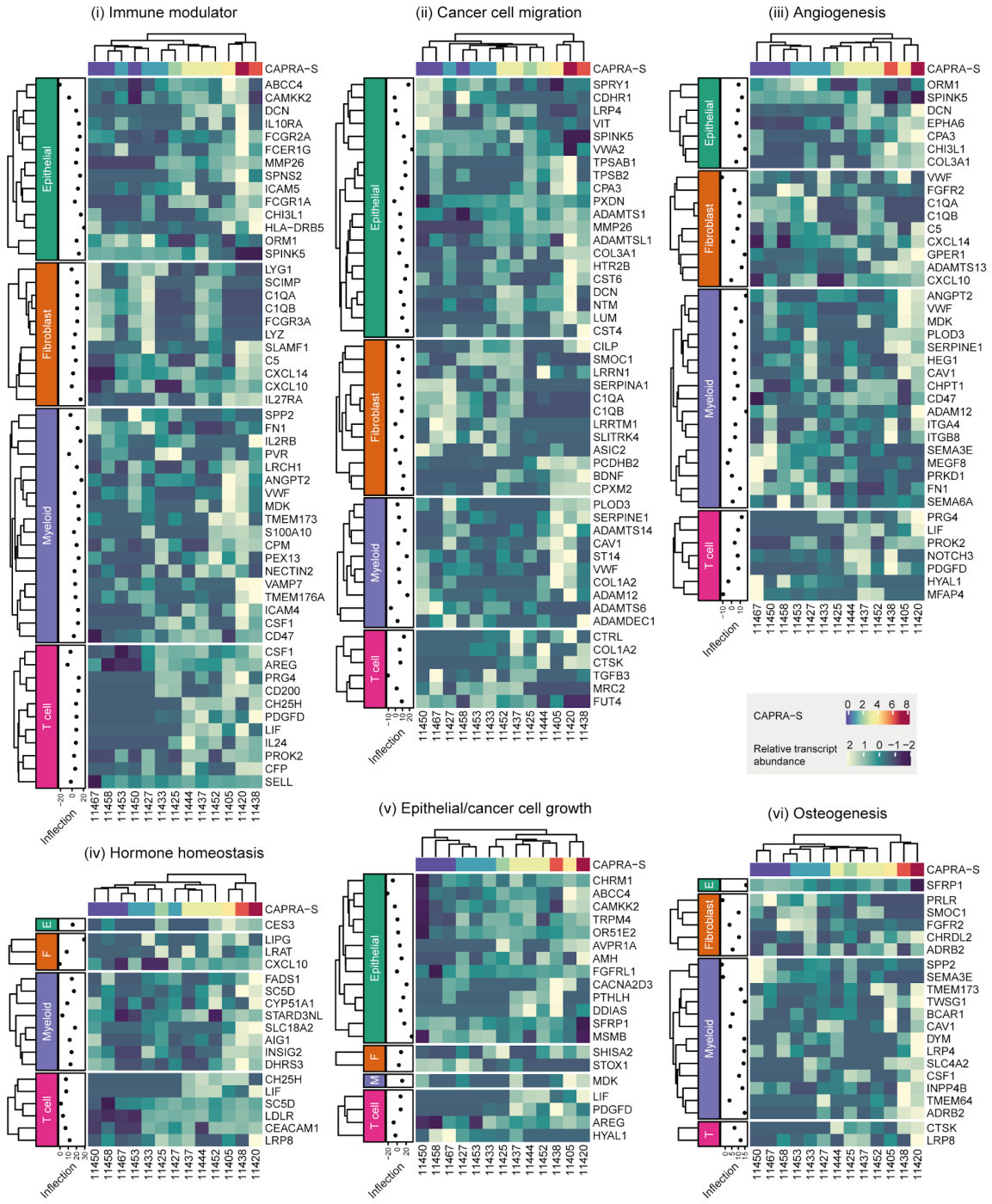


**Figure 3**. Recurrent functional categories identified in the differentially transcribed cellular interface-protein-coding (secreted and transmembrane) genes. The estimated inflection point for each gene shows the CAPRA-S risk score at which the transcriptional change was inferred to be fastest; values < 0 or > 7 indicate an early or late change, respectively.

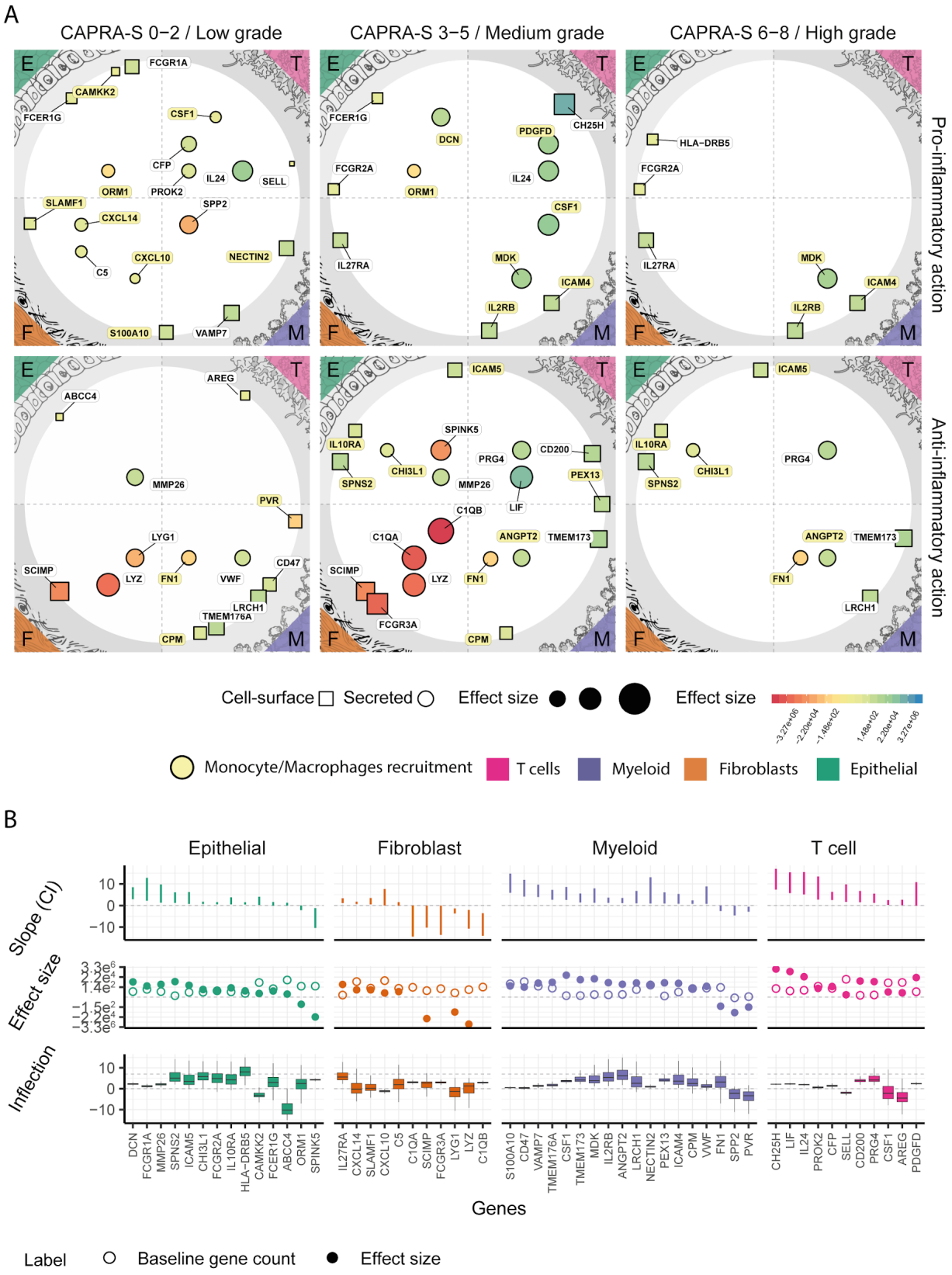


**Figure 4:** Multi cell-type immune-modulation evolves with risk progression and is mainly targeted

at monocyte-derived cells. Landscape of the immune-modulation related genes that encode for cellular interface-proteins (i.e., cell-surface or secreted) inferred by TABI 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 with 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).

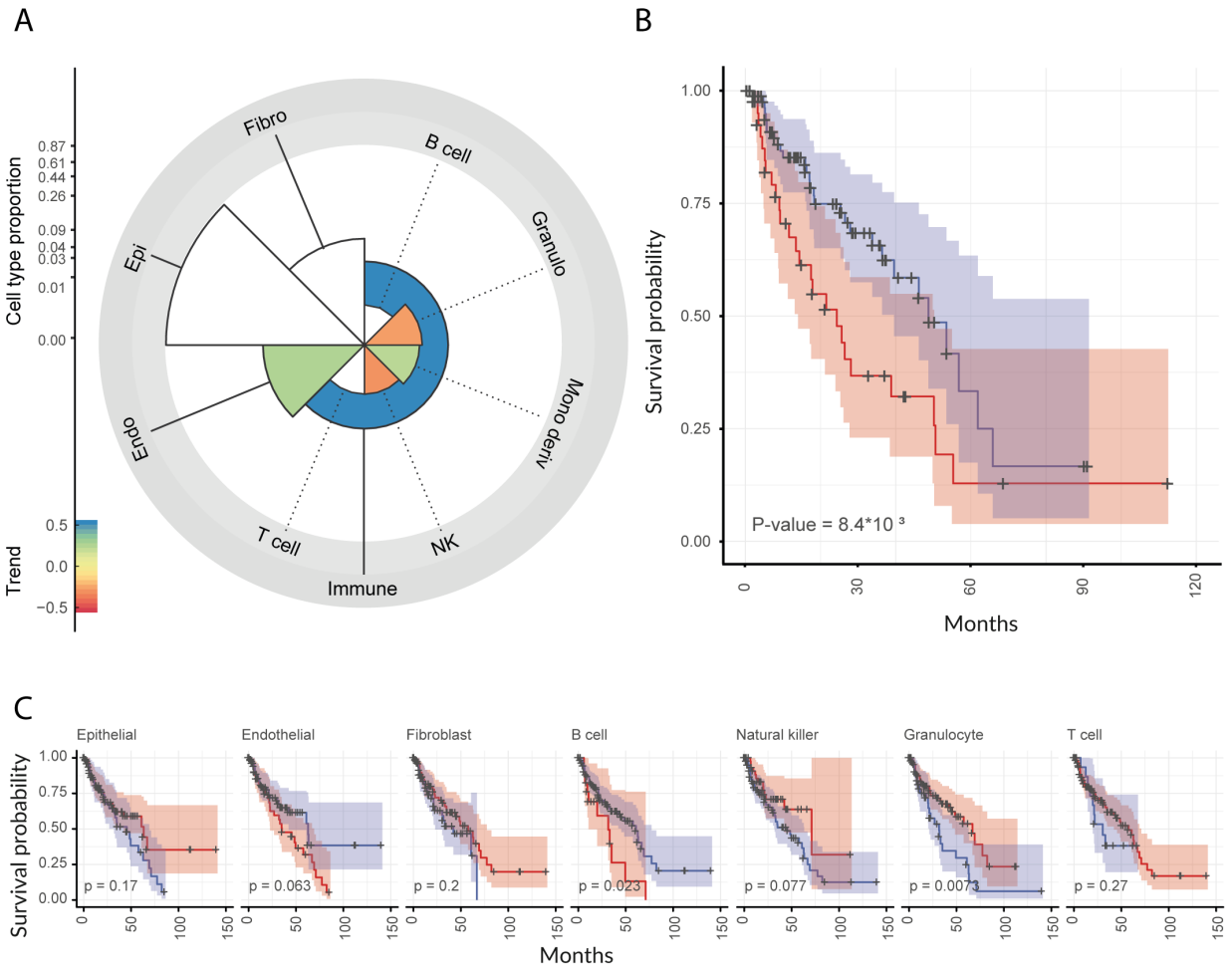


**Figure 5.** The abundance of monocyte-derived cells is positively associated with CAPRA-S risk score and negatively associated with disease-free survival. Association analysis of the abundance of monocyte-derived cells (inferred with a Bayesian model) with disease-free survival, performed on the independent primary prostate cancer TCGA dataset (n = 134). **A** — Polar plot for the differential tissue composition analysis of primary prostate cancer TCGA samples (n=134) 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 Material and Methods section, Survival analyses subsection). **C —** Kaplan–Meier plot for the other cell types included in the analysis.