Dissection of prostate tumour, stroma and immune transcriptional components reveals a key contribution of the microenvironment for disease 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], Michael Kerger[3], Ken Chow[2,3], Chayanica Nasa[9], Melanie Le Page[9], Natalie Lister[10], Simon Monard[9], Justin Peters[11] and Phil Dundee11, Anthony J. Costello[2,3], Paul J. Neeson[8], Scott G. Williams[8], Bhupinder Pal[12], Nicholas D. Huntington[6], Niall M. Corcoran[2,3,9,13,*], Anthony T. Papenfuss[1,13,14,4,8,*], Christopher M. Hovens[2,3,9,*]

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. 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. Epworth Center of Cancer Research, Victoria Australia
12. The Olivia Newton-John Cancer Research Institute, Heidelberg, Melbourne, Australia
13. Department of Urology, Frankston Hospital, Frankston, Victoria, Australia
14. Peter MacCallum Cancer Centre, Melbourne, VIC 3000, Australia.
15. School of Mathematics and Statistics, University of Melbourne, Melbourne, VIC 3010, Australia.

* These authors contributed equally

Correspondence:

Christopher Hovens (chovens@unimelb.edu.au)

Anthony T Papenfuss (papenfuss@wehi.edu.au)

Niall Corcoran (con@unimelb.edu.au)

Abstract

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. In this study, the experimental enrichment of selected cell-types and the development of a Bayesian inference model for continuous differential transcript abundance 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. 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.

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 leads to 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, 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 has 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 (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 applied an optimised protocol for combined cell-type enrichment and ultra-low-input RNA sequencing, which allowed probing of four key cell types across 13 fresh prostate tissue spanning a wide clinical disease spectrum. 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. 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 both with tissue deconvolution on the extensive Cancer Genome Atlas (TCGA) cohort and with multiplex immunohistochemistry on an independent patient cohort. The spatial analysis at the single-cell level allowed us to better identify relations between macrophages and epithelial and T cells along cancer progression.

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 those cell populations was guided by their predominant role in the progression of prostate and other cancers[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 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 samples and were removed from further analysis.

Differential transcription and model fitting

Dimensionality reduction (multidimensional scaling; MDS in Materials and Methods) of the filtered transcript abundance revealed an association of CAPRA-S risk score with either the first and/or 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 1.0×10^{-2} and 9.7×10^{-3} respectively), while a weaker pattern was apparent for myeloid and T cells (Bonferroni adjusted p-value of 3.0×10^{-2} and 0.37 respectively); possibly due to the greater heterogeneity of the two immune cell populations compared to epithelial and fibroblasts.

We performed a differential gene transcript abundance analysis 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 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 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 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 modelling linear, exponential and sigmoid-like trends of transcriptional alterations (Fig. 1B and 1C);

however, to provide robust modelling for RNA sequencing data we reparameterized 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 the statistical inference of transcriptional alterations, 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 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 (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). 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).

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[22] with more than 1500 scientific articles (Supplementary file 3). This 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 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, 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[23]. Within this set, the most 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 is present for cancer cell growth, osteogenesis and hormone modulation signatures, which are enriched with genes that 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 signatures from the four cell types tracks with the risk score covariate (Fig. 3). While the magnitude of the pro-inflammatory transcriptional signature remains roughly constant through the risk range, with 18 genes for CAPRA-S risk score ≤ 2 and 14 for CAPRA-S risk score >2; the anti-inflammatory signature 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[24–31] (highlighted in yellow in Fig. 3A). These include CAMKK2[32], ORM1[33] and DCN[34,35] in epithelial; IL2RB[36–38], ICAM4[39,40], DCN[34,35] and MDK[41,42] in myeloid cells; and CSF1[24] and PDGFD[43] in T cells. In addition, we identified a known fibroblast-macrophage chemotactic interaction including the regulation of the cytokines CXCL10[27], CXCL14[26] and the receptor SLAMF1[28,29] for fibroblasts; with COL1A2[44] and CYR61[45] (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[46], IL24[47], PROK2[48], SELL[49]. Interestingly, epithelial cells upregulate a cluster of receptor genes normally involved in antigen recognition and presentation in immune cells[50], 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 than the pro-inflammatory signature. Monocyte-derived cells are mainly targeted by genes that are differentially transcribed in epithelial and myeloid cells. These include the receptor genes SPNS2[51,52], IL10RA[53] and ICAM5 by epithelial cells; and the receptor genes CPM[54] and PEX13[55] and the secreted protein genes FN1[56] and ANGPT2[57] by myeloid cells. Another cluster of genes

targets predominantly T cells, including AREG[58,59], CD200[60], LRCH1[61], CD47[62]. 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 patient cohort. We used a hierarchical Bayesian inference model[63] 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 uses a collection of 250 curated publicly available transcriptional profiles (including BLUEPRINT[64], ENCODE[65], GSE89442[66] and GSE107011[67]), encompassing transcriptional signatures for 18 hierarchical cell-types (of which nine major cell types were tested for differences) and uses 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[63]. Overall, we estimated a median of 88% for epithelial cellular fraction across samples (consistent with public literature[68], 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-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 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. 4C).

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 PC - 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. PDL1 expressing cells were also categorised, using a PDL1-linked dye. Overall, we were able to sample an average of $41x10^3$ cells per biopsy (standard deviation 2.7×10^4). Of the categorised cell types, the most abundant 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 0.031; Fig. 5A-top-right), while the strongest negative association was for stromal cells in tumour areas (p-value 0.093; Fig. 5A-bottom). Overall, the average proximity between macrophages and other cells, aggregated by biopsy did not show any strong association with 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). That is, epithelial basal cells that are close to clusters of PDL1 expressing macrophages tend to be further away from T-cells.

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 make an integral contribution to the mechanisms that cause cancer progression, and that they are often modulated by cancer cells toward pro-tumorigenic behaviours. In this study, we significantly improved the knowledge of the molecular landscape of 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 CAPRA-S risk score, we developed a novel statistical inference method (TABI), which permitted modelling of transcript abundance natively on continuous factors of interest with a limited number of parameters, avoiding loss of information due to the dichotomization 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.

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 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 to clinical variables. Such an aspect is particularly relevant considering the substantial noise associated with such inference. To test the clinical significance of quantifying monocyte-derived cell numbers within the tumour mass beside the CAPRA-S risk score, we performed survival analyses from the same TCGA cohort, from the inferred cell-type proportions. 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[69]. In prostate cancer, specific and overall survival analyses have identified an elevated monocyte count as an independent prognostic factor for poor outcome[70–73]. 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[74].

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 (IHC) 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, while distanciating from stromal compartments; the latter being colocalized with cancer cells in tumour tissues This aspect becomes highly relevant as the glandular areas in both tumour and adjacent benign compartments that are rich in PDL1 macrophages become consequently poorer in T-cell numbers. PDL1 expressing macrophages have been associated with their M2 wound-healing phenotype[75]. The relationship between PD-L1 expression in intratumoral macrophages and prognosis in cancer patients is still controversial, with the two competing hypotheses being: (i) PDL1 intratumoral macrophages

lead to dysfunctional T-cells[76] or (ii) not having major effects[77]. Nonetheless, expression of PDL1 in macrophages has been shown to induce anti-inflammatory cytokines such as IL-10[75]. 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 negative effects on disease progression.

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)[78]. 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 (particularly, monocyte-derived cells) that contribute to disease progression; for example, 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. 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; and PMCC; HREC approvals 10/68, 13/167, 18/204).

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 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 an 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. We thank the anonymous reviewers whose comments have greatly improved this manuscript.

Authors' contribution

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. SPK led the multiplex histochemistry experiments supervised by PJN and SGW. 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[79,80]. 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 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), Thy-1 (CD90) PerCP-Cy5.5 (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 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 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 1.5ml 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 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 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 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 an 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[81]. 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[82] with default settings. The quality control on the alignment was performed with RNA-SeQC[83]. For each sample, the gene transcription abundance was quantified in terms of nucleotide reads per gene (read-count) using FeatureCounts[84] 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[85] with default settings.

Statistical inference

Changes of transcriptional levels along CAPRA-S risk score[86] 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[87] (Eq.1), but re-parameterised to improve numerical stability (Eq. 2). In particular, the standard parametrisation suffers from non-determinability issues 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), β_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[88] 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 (parameterized 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 (β_1) was subject to a regularised horseshoe[89] 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, $\hat{Y}$ 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), $\bar{y}$ and η which represent the gene-wise y-intercept and the inflection point of the generalised reparameterised sigmoid function (Eq. 2), while γ represent the hyperparameters of $\bar{y}$. 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 β_1 (first column of β) is defined by the hyperparameter list ξ[89] (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 (MDS)[90] 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 β_{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 reparameterised 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[22] using BiomaRt[91]. Furthermore, genes were functionally annotated with the protein atlas database[92] 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[88] 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[64], ENCODE[65], GSE89442[66] and GSE107011[67]).

Analysis of tumour microenvironment using multiplex immunohistochemistry

Slides (3µm) 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 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. [93]. Briefly, slides were deparaffinised and rehydrated with xylene and ethanol. The fluorochrome-coupled antibodies against human CD68 (macrophages and dendritic

cells; DC), high molecular weight cytokeratin (HMWCK; epithelial basal cells), CD3 (T cells), CD20 (B cells), CD11c (dendritic cells), and PDL1. The dye DAPI was used for nuclei staining. Vectra 3.0 Automated Quantitative Pathology Imaging System, 200 slide (Perkin Elmer, MA) was used for imaging as detailed by Keam et al.[93]. The software HALO was used for cell segmentation and phenotyping. Stromal cells were defined with 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).

Cell type proximity was quantified as the amount 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 data 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[94] on log transformed proximity measure. The co-proximity analysis between epithelial basal cells and PDL1+ macrophages and T cells, was performed at single cell level (averaged by tissue area of 5 cell size units) calculating the proximity on a radius of 50 relative cell sizes for ensuring a 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.

References

1. Bray, F. *et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* **68**, 394–424 (2018).
2. D'Amico, A. V. *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.* **17**, 168–172 (1999).
3. Bhindi, B. *et al.* Independent Validation of the American Joint Committee on Cancer 8th Edition Prostate Cancer Staging Classification. *J. Urol.* **198**, 1286–1294 (2017).
4. Lu-Yao, G. L. *et al.* Fifteen-year Outcomes Following Conservative Management Among Men Aged 65 Years or Older with

Localized Prostate Cancer. *Eur. Urol.* **68**, 805–811 (2015).

5. Chua, M. L. K. *et al.* A Prostate Cancer 'Nimbosus': Genomic Instability and SChLAP1 Dysregulation Underpin Aggression of Intraductal and Cribriform Subpathologies. *Eur. Urol.* **72**, 665–674 (2017).
6. Fraser, M. *et al.* Genomic hallmarks of localized, non-indolent prostate cancer. *Nature* **541**, 359–364 (2017).
7. Cancer Genome Atlas Research Network. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* **163**, 1011–1025 (2015).
8. Tyekucheva, S. *et al.* Stromal and epithelial transcriptional map of initiation progression and metastatic potential of human prostate cancer. *Nat. Commun.* **8**, 420 (2017).
9. Strasner, A. & Karin, M. Immune Infiltration and Prostate Cancer. *Front. Oncol.* **5**, 128 (2015).
10. Salvatore, V. *et al.* The tumor microenvironment promotes cancer progression and cell migration. *Oncotarget* **8**, 9608–9616 (2017).
11. Katt, M. E., Placone, A. L., Wong, A. D., Xu, Z. S. & Searson, P. C. In Vitro Tumor Models: Advantages, Disadvantages, Variables, and Selecting the Right Platform. *Front Bioeng Biotechnol* **4**, 12 (2016).
12. Smith, B. A. *et al.* A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E6544–52 (2015).
13. Zhang, D. *et al.* Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. *Nat. Commun.* **7**, 10798 (2016).
14. Berglund, E. *et al.* Spatial maps of prostate cancer transcriptomes reveal an unexplored landscape of heterogeneity. *Nat. Commun.* **9**, 2419 (2018).
15. Ippolito, L. *et al.* Cancer-associated fibroblasts promote prostate cancer malignancy via metabolic rewiring and mitochondrial transfer. *Oncogene* (2019) doi:10.1038/s41388-019-0805-7.
16. Jung, Y. *et al.* Recruitment of mesenchymal stem cells into prostate tumours promotes metastasis. *Nat. Commun.* **4**, 1795 (2013).
17. Kaminski, A. *et al.* Tumour-stroma interactions between metastatic prostate cancer cells and fibroblasts. *Int. J. Mol. Med.* **18**, 941–950 (2006).
18. Sharma, A., Suleyman, N., Jones, O. & Vasdev, N. Immunotherapy in Urological Tumors. *Rev. Urol.* **21**, 15–20 (2019).
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* **20**, 789–798 (2019).
20. Gelman, A. *et al. 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.* **5**, 189–211 (2012).

22. Botstein, D. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

23. Long, J. *et al.* Lipid metabolism and carcinogenesis, cancer development. *Am. J. Cancer Res.* **8**, 778–791 (2018).

24. Zhu, Y. *et al.* CSF1/CSF1R blockade reprograms tumor-infiltrating macrophages and improves response to T-cell checkpoint immunotherapy in pancreatic cancer models. *Cancer Res.* **74**, 5057–5069 (2014).

25. Kumar, V. *et al.* Cancer-Associated Fibroblasts Neutralize the Anti-tumor Effect of CSF1 Receptor Blockade by Inducing PMN-MDSC Infiltration of Tumors. *Cancer Cell* **32**, 654–668.e5 (2017).

26. Lu, J., Chatterjee, M., Schmid, H., Beck, S. & Gawaz, M. CXCL14 as an emerging immune and inflammatory modulator. *J. Inflamm.* **13**, 1 (2016).

27. Liu, M. *et al.* CXCL10/IP-10 in infectious diseases pathogenesis and potential therapeutic implications. *Cytokine Growth Factor Rev.* **22**, 121–130 (2011).

28. Latour, S. *et al.* Regulation of SLAM-mediated signal transduction by SAP, the X-linked lymphoproliferative gene product. *Nat. Immunol.* **2**, 681–690 (2001).

29. Wang, G. *et al.* Migration of myeloid cells during inflammation is differentially regulated by the cell surface receptors Slamf1 and Slamf8. *PLoS One* **10**, e0121968 (2015).

30. Rovis, T. L., Brlic, P. K., Kaynan, N. & Lisnic, V. J. Inflammatory monocytes and NK cells play a crucial role in DNAM-1–dependent control of cytomegalovirus infection. *Journal of* (2016).

31. O'Connell, P. A., Surette, A. P., Liwski, R. S., Svenningsson, P. & Waisman, D. M. S100A10 regulates plasminogen-dependent macrophage invasion. *Blood* **116**, 1136–1146 (2010).

32. Racioppi, L. CaMKK2: a novel target for shaping the androgen-regulated tumor ecosystem. *Trends Mol. Med.* **19**, 83–88 (2013).

33. Gemelli, C. *et al.* The Orosomucoid 1 protein is involved in the vitamin D - mediated macrophage de-activation process. *Exp. Cell Res.* **319**, 3201–3213 (2013).

34. Zhang, W. *et al.* Decorin is a pivotal effector in the extracellular matrix and tumour microenvironment. *Oncotarget* **9**, 5480–5491 (2018).

35. Järvinen, T. A. H. & Ruoslahti, E. Target-seeking antifibrotic compound enhances wound healing and suppresses scar formation in mice. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 21671–21676 (2010).

36. Jounaidi, Y., Cotten, J. F., Miller, K. W. & Forman, S. A. Tethering IL2 to Its Receptor IL2Rβ Enhances Antitumor Activity

and Expansion of Natural Killer NK92 Cells. *Cancer Res.* **77**, 5938–5951 (2017).

37. ImmuNet. http://immunet.princeton.edu/genes/detail/homo-sapien/IL2RB/.
38. Espinoza-Delgado, I. *et al.* Interleukin-2 and human monocyte activation. *J. Leukoc. Biol.* **57**, 13–19 (1995).
39. Kim, M. *et al.* Novel natural killer cell-mediated cancer immunotherapeutic activity of anisomycin against hepatocellular carcinoma cells. *Sci. Rep.* **8**, 10668 (2018).
40. Ihanus, E., Uotila, L. M., Toivanen, A., Varis, M. & Gahmberg, C. G. Red-cell ICAM-4 is a ligand for the monocyte/macrophage integrin CD11c/CD18: characterization of the binding sites on ICAM-4. *Blood* **109**, 802–810 (2007).
41. Narita, H., Chen, S., Komori, K. & Kadomatsu, K. Midkine is expressed by infiltrating macrophages in in-stent restenosis in hypercholesterolemic rabbits. *J. Vasc. Surg.* **47**, 1322–1329 (2008).
42. Fan, N. *et al.* Midkine, a potential link between obesity and insulin resistance. *PLoS One* **9**, e88299 (2014).
43. Yang, P. *et al.* Role of PDGF-D and PDGFR-β in neuroinflammation in experimental ICH mice model. *Exp. Neurol.* **283**, 157–164 (2016).
44. Saito, T. *et al.* Macrophage Infiltration Is a Causative Factor for Ligamentum Flavum Hypertrophy through the Activation of Collagen Production in Fibroblasts. *Am. J. Pathol.* **187**, 2831–2840 (2017).
45. Bai, T., Chen, C.-C. & Lau, L. F. Matricellular protein CCN1 activates a proinflammatory genetic program in murine macrophages. *J. Immunol.* **184**, 3223–3232 (2010).
46. Blatt, A. Z., Pathan, S. & Ferreira, V. P. Properdin: a tightly regulated critical inflammatory modulator. *Immunol. Rev.* **274**, 172–190 (2016).
47. Persaud, L. *et al.* Mechanism of Action and Applications of Interleukin 24 in Immunotherapy. *Int. J. Mol. Sci.* **17**, (2016).
48. Lattanzi, R. *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**, 301292 (2015).
49. Mohammed, R. N. *et al.* L-selectin Is Essential for Delivery of Activated CD8(+) T Cells to Virus-Infected Organs for Protective Immunity. *Cell Rep.* **14**, 760–771 (2016).
50. Kobayashi, K. S. & van den Elsen, P. J. NLRC5: a key regulator of MHC class I-dependent immune responses. *Nat. Rev. Immunol.* **12**, 813–820 (2012).
51. Tran, H. B. *et al.* Disrupted epithelial/macrophage crosstalk via Spinster homologue 2-mediated S1P signaling may drive defective macrophage phagocytic function in COPD. *PLoS One* **12**, e0179577 (2017).
52. Rodriguez, Y. I. *et al.* Sphingosine-1 Phosphate: A New Modulator of Immune Plasticity in the Tumor Microenvironment. *Front. Oncol.* **6**, 218 (2016).

53. Shouval, D. S. *et al.* Interleukin-10 receptor signaling in innate immune cells regulates mucosal immune tolerance and anti-inflammatory macrophage function. *Immunity* **40**, 706–719 (2014).

54. Denis, C. J. & Lambeir, A.-M. The potential of carboxypeptidase M as a therapeutic target in cancer. *Expert Opin. Ther. Targets* **17**, 265–279 (2013).

55. Vijayan, V. *et al.* A New Immunomodulatory Role for Peroxisomes in Macrophages Activated by the TLR4 Ligand Lipopolysaccharide. *J. Immunol.* **198**, 2414–2425 (2017).

56. Pinto, A. R., Godwin, J. W. & Rosenthal, N. A. Macrophages in cardiac homeostasis, injury responses and progenitor cell mobilisation. *Stem Cell Res.* **13**, 705–714 (2014).

57. Wu, X. *et al.* Angiopoietin-2 as a Biomarker and Target for Immune Checkpoint Therapy. *Cancer Immunol Res* **5**, 17–28 (2017).

58. Wang, S. *et al.* Amphiregulin Confers Regulatory T Cell Suppressive Function and Tumor Invasion via the EGFR/GSK-3β/Foxp3 Axis. *J. Biol. Chem.* **291**, 21085–21095 (2016).

59. Arosa, F. A. *On the Origin and Function of Human NK-like CD8+ T Cells: Charting New Territories*. (Frontiers Media SA, 2018).

60. Aref, S., Azmy, E. & El-Gilany, A. H. Upregulation of CD200 is associated with regulatory T cell expansion and disease progression in multiple myeloma. *Hematol. Oncol.* **35**, 51–57 (2017).

61. Xu, X. *et al.* LRCH1 interferes with DOCK8-Cdc42-induced T cell migration and ameliorates experimental autoimmune encephalomyelitis. *J. Exp. Med.* **214**, 209–226 (2017).

62. Kaur, S. *et al.* CD47-dependent immunomodulatory and angiogenic activities of extracellular vesicles produced by T cells. *Matrix Biol.* **37**, 49–59 (2014).

63. Mangiola, S. *et al.* Androgen deprivation therapy promotes an obesity-like microenvironment in periprostatic fat. *Endocrine Connections* (2019) doi:10.1530/EC-19-0029.

64. Stunnenberg, H. G., International Human Epigenome Consortium & Hirst, M. The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* **167**, 1897 (2016).

65. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

66. Basit, F., Mathan, T., Sancho, D. & de Vries, I. J. M. Human Dendritic Cell Subsets Undergo Distinct Metabolic Reprogramming for Immune Response. *Front. Immunol.* **9**, 2489 (2018).

67. Xu, W. *et al.* Mapping of γ/δ T cells reveals Vδ2+ T cells resistance to senescence. *EBioMedicine* **39**, 44–58 (2019).

68. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).

69. Lanciotti, M. *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**, 486798 (2014).

70. Wang, Y.-Q. *et al.* Peripheral monocyte count: an independent diagnostic and prognostic biomarker for prostate cancer - a large Chinese cohort study. *Asian J. Androl.* **19**, 579–585 (2017).

71. Lundholm, M. *et al.* Secreted Factors from Colorectal and Prostate Cancer Cells Skew the Immune Response in Opposite Directions. *Sci. Rep.* **5**, 15651 (2015).

72. Zhang, E. *et al.* Differences of the immune cell landscape between normal and tumor tissue in human prostate. *Clin. Transl. Oncol.* (2019) doi:10.1007/s12094-019-02128-5.

73. Gannon, P. O. *et al.* Characterization of the intra-prostatic immune cell infiltration in androgen-deprived prostate cancer patients. *J. Immunol. Methods* **348**, 9–17 (2009).

74. Nonomura, N. *et al.* Infiltration of tumour-associated macrophages in prostate biopsy specimens is predictive of disease progression after hormonal therapy for prostate cancer. *BJU Int.* **107**, 1918–1922 (2011).

75. 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.* **359**, 449–457 (2017).

76. Smith, P. *et al.* Schistosoma mansoni worms induce anergy of T cells via selective up-regulation of programmed death ligand 1 on macrophages. *J. Immunol.* **173**, 1240–1248 (2004).

77. Singhal, S. *et al.* Human tumor-associated monocytes/macrophages and their regulation of T cell responses in early-stage lung cancer. *Sci. Transl. Med.* **11**, (2019).

78. Rescigno, P. & de Bono, J. S. Immunotherapy for lethal prostate cancer. *Nat. Rev. Urol.* **16**, 69–70 (2019).

79. Kerger, M. *et al.* Microscopic assessment of fresh prostate tumour specimens yields significantly increased rates of correctly annotated samples for downstream analysis. *Pathology* **44**, 204–208 (2012).

80. Mangiola, S. *et al.* Periprostatic fat tissue transcriptome reveals a signature diagnostic for high-risk prostate cancer. *Endocr. Relat. Cancer* **25**, 569–581 (2018).

81. Andrews, S. & Others. FastQC: a quality control tool for high throughput sequence data. (2010).

82. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

83. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).

84. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

85. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).

86. Cooperberg, M. R., Hilton, J. F. & Carroll, P. R. The CAPRA-S score: a straightforward tool for improved prediction of outcomes after radical prostatectomy. *Cancer* (2011).

87. Richards, F. J. A Flexible Growth Function for Empirical Use. *J. Exp. Bot.* **10**, 290–301 (1959).

88. Carpenter, B. *et al.* Stan: A probabilistic programming language. *J. Stat. Softw.* **20**, 1–37 (2016).

89. Piironen, J. & Vehtari, A. Sparsity information and regularization in the horseshoe and other shrinkage priors. *Electron. J. Stat.* **11**, 5018–5051 (2017).

90. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

91. Durinck, S. *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).

92. Uhlén, M. *et al.* Proteomics. Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

93. Keam, S. P. *et al.* High dose-rate brachytherapy of localized prostate cancer converts tumors from cold to hot. *J Immunother Cancer* **8**, (2020).

94. Osorio, F. heavy: Robust estimation using heavy-tailed distributions. *R package version 0. 38. ed* (2016).

95. Forbes, S. A. *et al.* COSMIC: exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Research* vol. 43 D805–D811 (2015).

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 which PC genes, consistent
Epithelial	21,618	5,408	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	13836	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%)

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[95] 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[95] or labelled as such in our manually curated cell-type specific database (Supplementary file 2).

Figures

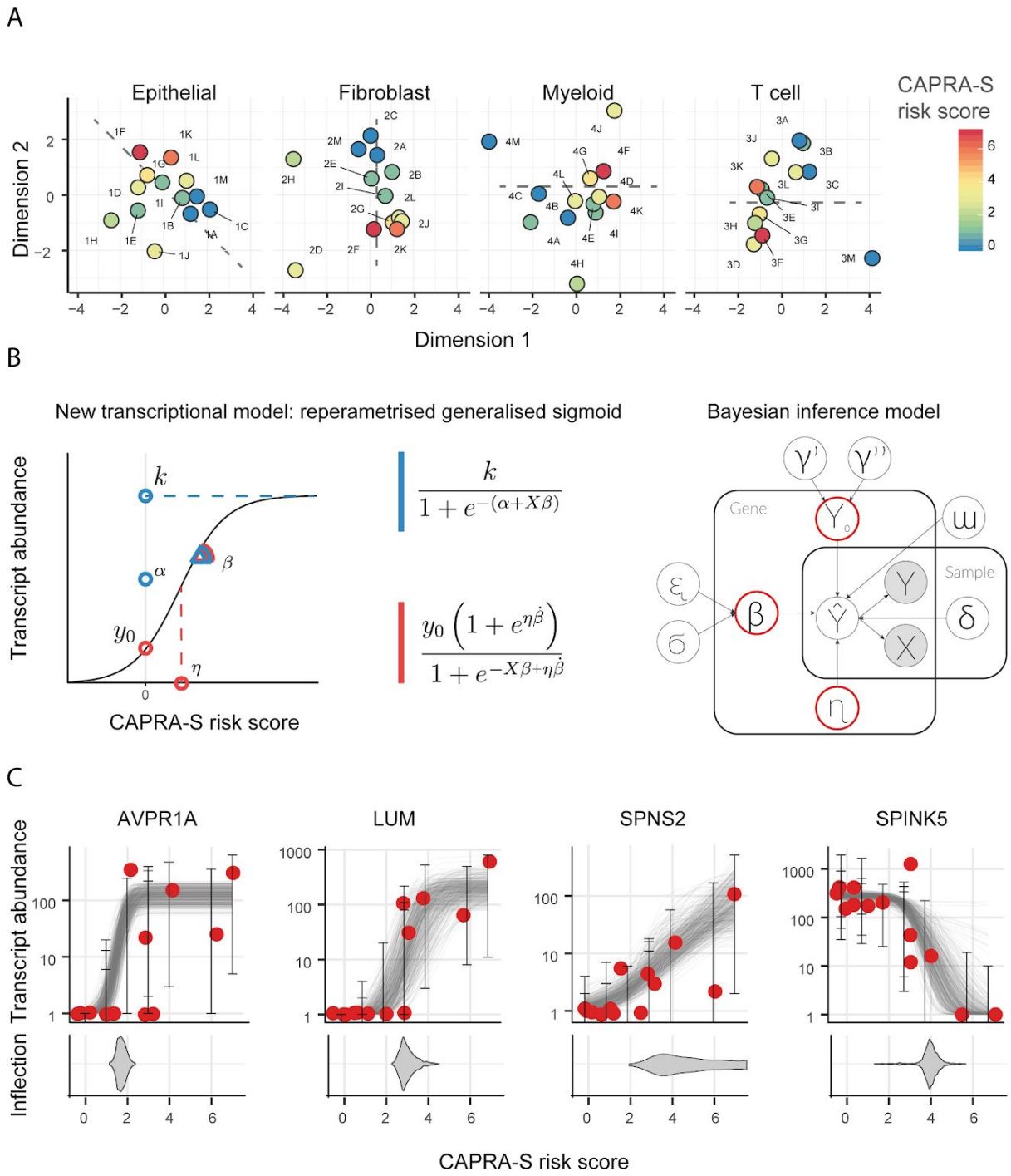


Figure 1: Data exploration and implementation of a model reflecting the continuous (i.e., non-discrete) relation between the CAPRA-S risk score and gene transcript abundance. **A** —

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 identifiers (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.

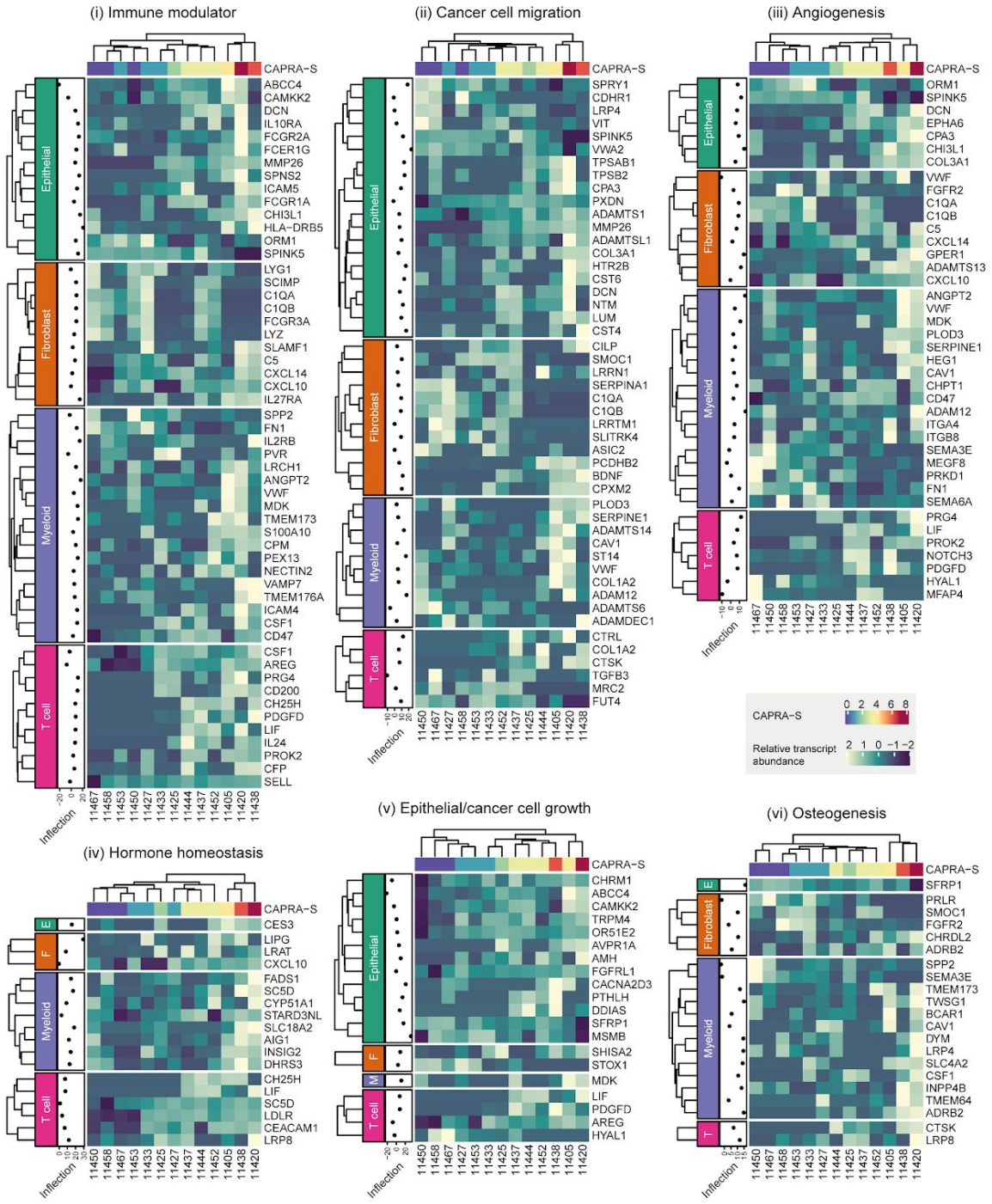
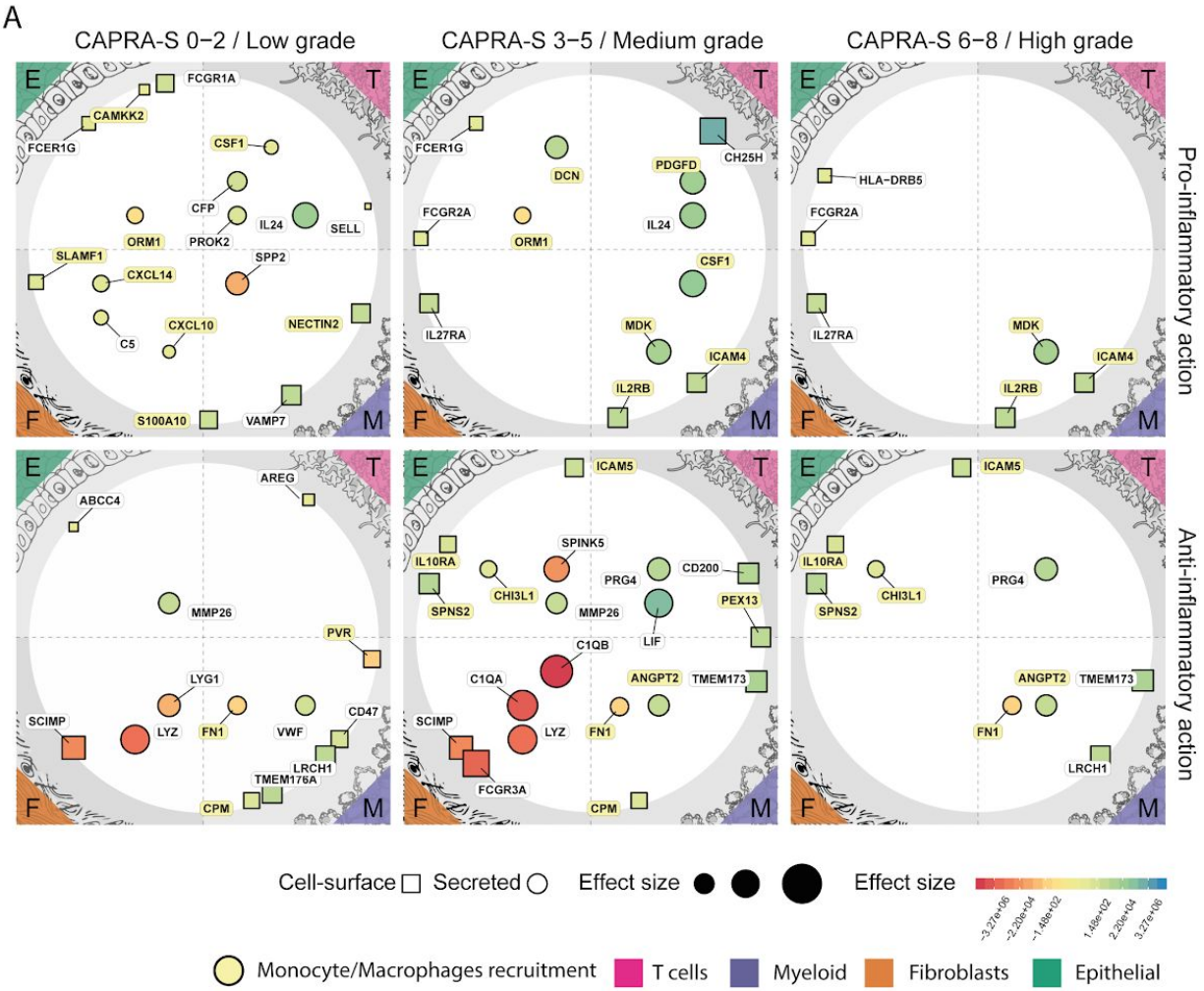


Figure 2. Recurrent functional categories identified in the differentially transcribed cellular interface-protein-coding (secreted and transmembrane) genes. The estimated inflection point for each gene shows the CAPRA-S risk score at which the transcriptional change was inferred to be fastest; values < 0 or > 7 indicate an early or late change, respectively.


A
CAPRA-S 0−2 / Low grade
CAPRA-S 3−5 / Medium grade
CAPRA-S 6−8 / High grade
Pro-inflammatory action
Anti-inflammatory action
E
T
F
M
FCGR1A
CAMKK2
FCER1G
CSF1
CFP
ORM1
PROK2
IL24
SELL
SLAMF1
CXCL14
SPP2
C5
CXCL10
NECTIN2
S100A10
VAMP7
DCN
PDGFD
CH25H
FCGR2A
IL27RA
MDK
ICAM4
IL2RB
HLA−DRB5
AREG
ABCC4
MMP26
PVR
LYG1
SCIMP
LYZ
FN1
VWF
CD47
LRCH1
TMEM176A
CPM
ICAM5
SPINK5
IL10RA
SPNS2
CHI3L1
PRG4
CD200
PEX13
LIF
C1QB
C1QA
ANGPT2
TMEM173
FCGR3A
Cell-surface
Secreted
Effect size
Effect size
−3.27e+06
−2.20e+04
−1.48e+02
1.48e+02
2.20e+04
3.27e+06
Monocyte/Macrophages recruitment
T cells
Myeloid
Fibroblasts
Epithelial

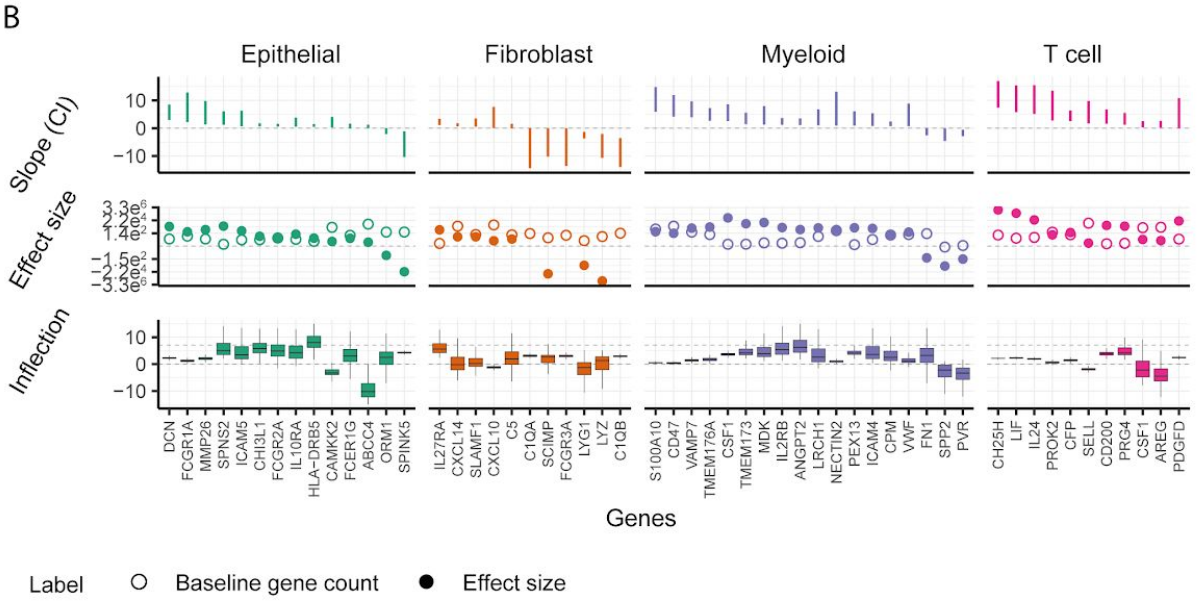

B
Epithelial
Fibroblast
Myeloid
T cell
Slope (CI)
Effect size
Inflection
10
0
−10
DCN
FCGR1A
MMP26
SPNS2
ICAM5
CHI3L1
FCGR2A
IL10RA
HLA−DRB5
CAMKK2
FCER1G
ABCC4
ORM1
SPINK5
IL27RA
CXCL14
SLAMF1
CXCL10
C5
C1QA
SCIMP
FCGR3A
LYG1
LYZ
C1QB
S100A10
CD47
VAMP7
TMEM176A
CSF1
TMEM173
MDK
IL2RB
ANGPT2
LRCH1
NECTIN2
PEX13
ICAM4
CPM
VWF
FN1
SPP2
PVR
CH25H
LIF
IL24
PROK2
CFP
SELL
CD200
PRG4
CSF1
AREG
PDGFD
Genes
Label
Baseline gene count
Effect size

Figure 3: 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).

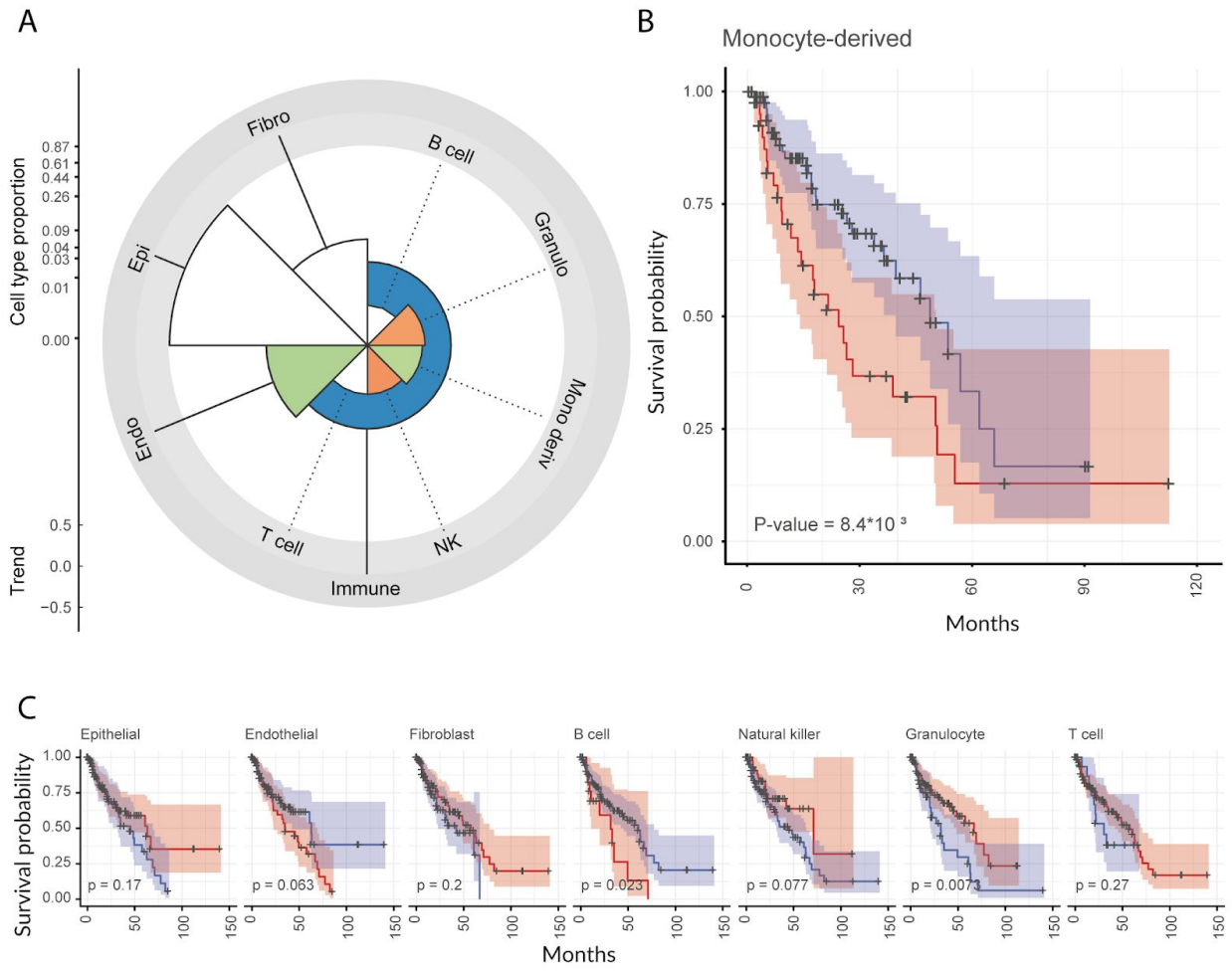


Figure 4. The abundance of monocyte-derived cells is positively associated with CAPRA-S risk score and negatively associated with disease-free survival. Association analysis of the abundance of monocyte-derived cells (inferred with a Bayesian model) with disease-free survival, performed on the independent 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.

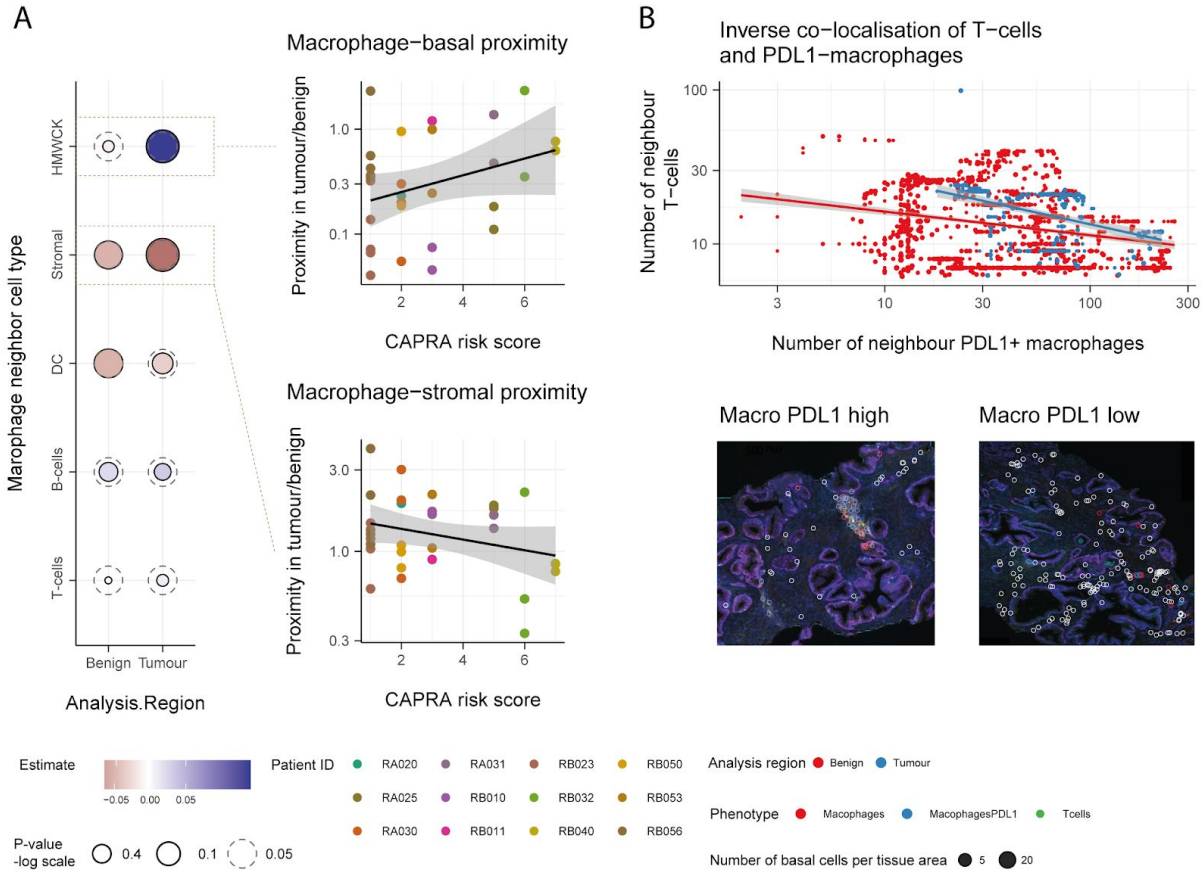


Figure 5. A left - Association between macrophage proximity and CAPRA risk score for five cell types identified from the multiplex immunohistochemistry. Proximity is calculated as number of neighbour cells per tissue area, and summarised using the median for each tumour biopsy. **A right -** Association 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. **B -** Decreased proximity of T-cells with epithelial basal in presence of PDL1 expressing macrophages. The bottom section shows the multiplex immunohistochemistry tissue from patient RB010, with two examples of 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.