The Prognostic Potential of Human Prostate Cancer-Associated Macrophage Subtypes as Revealed by Single-cell Transcriptomics

Joseph C Siefert¹,²,#, Bianca Cioni¹,#, Mauro J Muraro³,⁴, Mohammed Alshalalfa⁵,
Judith Vivié³,⁴, Henk van der Poel⁶, Ivo G Schoots⁷, Elise Bekers⁸, Felix Y Feng⁵,
Lodewyk Wessels²,⁹,* Wilbert Zwart¹,⁹,* and Andries M Bergman¹,¹⁰,*

¹ Division of Oncogenomics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX, Amsterdam, The Netherlands, ² Division of Molecular Carcinogenesis, The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX, Amsterdam, The Netherlands, ³ Single Cell Discoveries B.V., Uppsalalaan 8, 3584CT, The Netherlands, ⁴ Hubrecht Institute-KNAW and University Medical Center Utrecht, Uppsalalaan 8, 3584CT, The Netherlands, ⁵ Department of Radiation Oncology, UCSF Helen Diller Family Comprehensive Cancer Center, San Francisco, CA, USA, ⁶ Division of Urology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX, Amsterdam, The Netherlands, ⁷ Department of Radiology and Nuclear Medicine The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX, Amsterdam, The Netherlands, ⁸ Department of Pathology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX, The Netherlands, ⁹ Oncode Institute, the Netherlands,¹⁰ Division of Medical Oncology, The Netherlands Cancer Institute, Plesmanlaan 121, 1066CX, Amsterdam, The Netherlands

Conflict of Interest Statement: The authors declare no potential conflicts of interest

# These authors contributed equally
Macrophages in the tumor microenvironment are causally linked with prostate cancer development and progression, yet little is known about their composition in neoplastic human tissue. By performing single cell transcriptomic analysis of human prostate cancer resident macrophages, three distinct populations were identified in the diseased prostate. Unexpectedly, no differences were observed between macrophages isolated from the tumorous and non-tumorous portions of the prostatectomy specimens. Markers associated with canonical M1 and M2 macrophage phenotypes were identifiable, however these were not the main factors defining unique subtypes. The genes selectively associated with each macrophage cluster were used to develop a gene signature which was highly associated with both recurrence-free and metastasis-free survival. These results highlight the relevance of tissue-specific macrophage subtypes in the tumour microenvironment for prostate cancer progression and demonstrates the utility of profiling single-cell transcriptomics in human tumor samples as a strategy to design gene classifiers for patient prognostication.
Implications

The specific macrophage subtypes present in a diseased human prostate have prognostic value, suggesting that the relative proportions of these populations are related to patient outcome. Understanding the relative contributions of these subtypes will not only inform patient prognostication, but will enable personalised immunotherapeutic strategies to increase beneficial populations or reduce detrimental populations.

Introduction

Blood-derived monocytes reach the majority of the tissues in the body, both cancer affected and normal, where they become tissue-resident macrophages [1]. However, not all tissue-resident macrophages arise from circulating monocytes, as resident macrophages are present during embryonic development and persist during adulthood [2, 3]. Consequently, as in other tissues, the prostate is composed of both embryonic-derived and blood-derived macrophages, where it remains unclear whether macrophages of distinct origins have distinct functions [4].

Macrophages are extremely plastic and phenotypically heterogeneous immune cells, whose diversity is largely influenced by the microenvironment in which they reside [5]. Several studies showed that in vitro blood monocyte-derived macrophages can acquire a large spectrum of phenotypes depending on different stimuli present in the cell culture [6, 7]. However, these models do not capture the dynamic nature of macrophages in their native microenvironment. Human tissue-specific characterization of tumor associated macrophages (TAMs) is limited to glioma, skin and hepatocellular carcinoma [8-10], and there are no studies addressing prostate cancer (PCa)-specific macrophage phenotypic diversity.
The PCa tumor microenvironment (TME) is composed of various cells, including stromal, endothelial and immune cells, with tissue-resident macrophages representing one of the most predominant immune cell populations [11, 12]. Macrophages are critical mediators of tissue homeostasis and have the capacity to suppress cancer-associated processes, including tumor cell proliferation, angiogenesis and metastasis [13]. Multiple studies have shown a correlation between high infiltration of TAMs in the PCa microenvironment and poor prognosis, which suggests a role of these cells in cancer progression [14-18]. Given the prognostic significance of macrophages in the TME, strategies aiming to target these cells have emerged as strong candidates for cancer treatment [19-22].

It is also thought that immune cell type rather than sheer numbers of immune cells present in the TME relates to efficacy [23]. Various macrophage phenotypes have been described, including the pro-inflammatory/anti-tumor M1 state, and the anti-inflammatory/pro-tumor M2 state, both characterized by expression of specific markers [24]. However, macrophage diversity is likely not a binary division, but rather a continuum of phenotypes between M1 and M2 extremes [6, 7]. However, the diversity of human macrophage populations in PCa is not yet explored and, therefore, tissue specific markers of macrophage populations in prostate cancer are not yet defined.

To address this, we applied single-cell mRNA sequencing on myeloid cells isolated from prostatectomy specimens. Here we describe novel phenotypes of PCa associated macrophages and their distinct prognostic potential. Moreover, we propose new molecular markers for identification of these phenotypes in patients with localized disease. Understanding the unique macrophage populations in individual prostate tumors and their effect on outcome will not only enhance our
knowledge of prostate cancer biology and progression, but can better inform clinicians regarding a patients’ prognosis and treatment options.

**Materials and methods**

*Patients, tumor specimens and Ethics Statement*

Prostate biopsies were collected from post robotic-assisted laparoscopic prostatectomy (RALP) surgical specimen of PCa patients who did not receive any prior therapy (Gleason score 3+4 and 3+3). Biopsies were collected from both the tumorous and the non-tumorous site of the prostate, which were estimated by a pre-surgery multi parametric magnetic resonance imaging (mpMRI) scan of the pelvis and palpation by the surgeon of the prostatectomy specimen. An average of three 18G biopsies were taken from the areas of the prostate with a very high likelihood of containing clinically relevant prostate cancer (tumorous site) and from areas of the prostate without suspected prostate cancer (non-tumorous site) on mpMRI. Accuracy of the mpMRI to identify tumorous and non-tumorous areas of the prostate was histologically verified in H&E stained whole mount formalin fixed and paraffin embedded slides. Fresh biopsies from the tumorous site and non-tumorous site of the prostatectomy specimen from 4 patients were processed separately for cell-surface markers CD14+ and/or CD11b+ myeloid cell isolation and submitted for single-cell RNA sequencing.

The occurrence of identified clusters of macrophage populations in prostatectomy specimen was estimated by immunohistochemistry staining for cluster specific marker genes, which were selected from the top ten list of most differentially expressed genes between the clusters, that are not secreted and had the highest specificity for myeloid cells. Stained cores for the selected markers of the clusters,
Cluster 0: SLC40A1 (antibody HPA065634; antibody 11 cores), Cluster 1: PLAC8 (antibody HPA040465; 21 cores), Cluster 2: FCN1 (antibody HPA001295; 32 cores) and the pan-macrophage marker CD68 (antibody CAB000051; 54 cores), were downloaded from the human protein atlas (www.proteinatlas.org) and visually estimated for percentage surface containing tumor and non-tumor and for number of marker positive cells in both compartments. Marker positive cells were identified and quantified by specific marker staining and morphology by a pathologist specialized in uro-oncology. A score (0, 1, 2, or 3) was assigned based on the number of macrophages present (0, 1-10, 11-100, >100), respectively. Density of marker positive cells in tumorous and non-tumorous tissue, was estimated by dividing the number of marker positive cells in a compartment by the fraction of the total core surface containing the compartment.

The use of patient data and biopsies from fresh prostatectomy specimens for research purposes at the Netherlands Cancer Institute have been executed pursuant to Dutch legislation and international standards. Prior to 25 May 2018, national legislation on data protection applied, as well as the International Guideline on Good Clinical Practice. From 25 May 2018 on, we also adhere to the GDPR. Within this framework, patients are informed and have always had the opportunity to object or actively consent to the (continued) use of their personal data & biospecimens in research. For the current study, informed consent was obtained from all patients. Hence, the procedures comply both with (inter-) national legislative and ethical standards.

Tissue dissociation and CD11b+ and/or CD14+ cells sorting
Single-cell suspension was prepared from fresh PCa biopsies by mechanical dissociation within two hours after surgery. Biopsies were transported from the operation room on ice and minced with a scalpel in cold PBS + 0.5% BSA. Tissue was then mechanically dissociated using a gentleMACS™ Dissociator (MACS Milteny Biotec) using C-tubes (MACS Milteny Biotec) for 2 minutes as previously described [25]. Subsequently, the samples were filtered through a 70 μm strainer (BD Falcon) and spun down for 5 min at 300 g at 4°C. Cells were re-filtered through a 40 μm strainer (BD Falcon), spun down for 5 min at 300 g at 4°C and re-suspended in cold PBS + 0.5% BSA. To compare the efficacy of GentleMACS with enzymatic digestion, biopsies were chopped with scalpel and tweezers and transferred to a 15ml Falcon tube containing 100U/ml of Collagenase I (SCR103, Sigma Aldrich), 0.05% of DNAse I (89836, Thermo Fisher) and 5U/ml of Hyaluronidase (H3506 Sigma Aldrich) and incubated at 37°C for 2 hours. Enzymatic digestion was then stopped by adding 1:1 volume of FBS-containing RPMI medium to the cell suspension. A 70 μm strainer was used to filter the cell suspension, followed by centrifugation at 1200 rpm for 10 minutes. Supernatant was removed and cells were resuspended in RPMI medium containing 10% FBS.

Cells of the dissociated biopsies were incubated with APC-CD45, PE-CD14, PE-CD11b and FITC-CD3 (all Ebioscience) for 20 minutes and washed before sorting using a Moflo Astrios (Beckman Coulter) or FACSaria IIu (BD BioSciences). As a first step, cell doublets/multiplets and dead cells were excluded using FSC height vs. area. Then, CD45+ leukocytes were selected, while small CD45+ cells (low SSC) were discarded as possible lymphocytes. Subsequently, CD14+ and/or CD11b+ single cells lacking CD3 expression (lymphocytes) were selected. Living single CD45+CD3-CD14+ and/or CD11b+ macrophages (based on DAPI and scatter
properties) were sorted into eight 384 wells plates (Biorad) where cDNA synthesis was performed as previously described [26].

**Single-cell sequencing with SORT-seq**

Single-cell mRNA sequencing was performed according to an adapted version of the SORT-seq protocol [26], using barcoded poly-A primers described by van den Brink *et al.* [27]. In short, single cells were FACS sorted into 384-well plates containing 384 different poly-A barcoded primers and Mineral oil (Sigma). After sorting, plates were snap-frozen on dry ice and stored at -80°C. Subsequently, cells were heat-lysed at 65°C followed by cDNA synthesis using the CEL-Seq2 protocol [28] using robotic liquid handling platforms Nanodrop II (GC Biotech) and Mosquito (TTP Labtech). After second strand cDNA synthesis of poly-A transcripts, the barcoded material was pooled into libraries of 384 cells and amplified using *in vitro* transcription [28]. Following amplification, the rest of the CEL-seq2 protocol was followed for preparation of the amplified cDNA library, using TruSeq small RNA primers (Illumina) as previously described [26]. The DNA library was paired-end sequenced on an Illumina Nextseq™500, high output, 1x75 bp.

**Single-cell sequencing data analysis**

After Illumina sequencing, read 1 was assigned 26 base pairs and was used for identification of the Illumina library barcode and cell barcode. Unique molecular identifiers (UMI) tags were added to each read. These are molecular tags used in order to detect and quantify unique mRNA transcripts. More specifically, mRNA libraries were generated by fragmentation and reverse transcribed to cDNA with tag-specific primers. Read 2 was assigned 60 base pairs and used to map to the
reference transcriptome of Hg19. Data was demultiplexed as described previously [29].

The Seurat v3.1.4 package was used in R v3.6.1 for processing the data [30]. To exclude dead, dying, or otherwise low-quality cells, cells with less than 1000 features (genes) and cells with greater than 1% mitochondrial reads were removed from analysis. To exclude cell doublets/multiplets, cells with greater than 6000 features (genes) were removed from analysis. The data were log-normalized with a scale factor of 10,000, and the 2000 most variable features were identified using a variance stabilizing transformation. The data was scaled according to all detected genes and principal component analysis was performed on the most variable genes. For linear dimensionality reduction, the number of principal components (20) was selected based on combined Jackstraw analysis, examination of elbow plots, difference in variation between subsequent principle components, and cumulative percent variation explained [31]. To identify clusters, a K-nearest neighbors graph was constructed from the selected principal components and clusters were identified from this using the Louvain algorithm at a resolution of 0.5 in Seurat. These were then projected with Uniform Manifold Approximation and Projection (UMAP) using uwot v0.1.5 and umap v0.2.4.1 packages in R.

A reciprocal PCA method was used for data integration [30]. In this procedure, the data from each patient was normalized, variable features were selected, the data was scaled, and principal component analysis was performed independently. The PCA space of each patient was then projected into each other patient to identify anchor points. The anchors represent matching cell states identified by pairwise correspondence between cells from different patients and are used to transform the datasets into a shared space. The integrated data was then scaled and principal
component analysis was performed as before. Clustering at various shared nearest neighbors (SNN) resolutions was evaluated and plotted with clustree_0.4.3 [32].

**Differential expression and Gene Set Enrichment Analysis**

Differential expression analysis was performed on normalized RNA values with minimum percentage (min.pct) and log fold-change (logfc) thresholds of 0.25 to identify marker genes specific to each cluster. Of note, while upregulation of genes in a cluster relative to other clusters is generally reported, this implies that in the latter clusters those genes are relatively downregulated. Significant differentially expressed genes were defined by Bonferroni adjusted p-value <0.05. Gene set enrichment analysis was performed using clusterProfiler v3.12.0 with msigdb v7.1.1 database in R [33]. Hallmark gene set enrichment was performed by calculating logfc for all genes in each cluster as compared to the other two clusters, without any thresholds for min.pct or logfc, and ranking genes based on logfc.

**Generation and evaluation of gene signatures**

Differentially expressed genes between macrophage clusters identified in the integrated dataset were used to construct prognostic signatures for biochemical relapse-free survival of PCa patients in a published MKSCC PCa dataset (GSE21032) [34]. This dataset, comprised of 131 primary PCa patients with RNA expression and biochemical recurrence (BCR) as determined by serum prostate specific antigen (PSA) levels, was used to assess biochemical relapse-free survival (RFS). RFS was defined as time from prostatectomy to BCR (rise of PSA ≥0.2 ng/ml on two occasions).
The gene signature (classifier) was generated employing Elastic net Cox regression using glmnet v4.0 in R [35]. The prognostic performance of the selected gene set was assessed in the MSKCC data by performing a nested 10-fold cross-validation (10FCV), with the full dataset split randomly into 10 folds with each fold stratified for the number of events and Gleason score. The Elastic net regularization parameters (alpha and lambda) were optimized as follows. A preselected set of values for alpha were tested (0 to 1 by 0.1). For each value of alpha in this set, we performed 10FCV to determine the optimal value of lambda. To this end, we selected the value of lambda associated with the minimum, average mean-squared error (MSE) across the folds. This procedure delivered, for each value of alpha, the optimal value of lambda and the associated MSE. We then selected the value of alpha that delivered the lowest MSE across all values of alpha in the set. Hazard ratios (HR), confidence intervals (CI), p-values, and Harrel’s C-index (concordance index) for the MSKCC evaluation were generated using the coxph function in the survival v3.1-12 package in R [36]. Survival plots were made by selecting the high (top 25%), low (bottom 25%) and intermediate (middle 50%) risk of relapse from the CV predictions, fit with event censoring and BCR-free time from MSKCC dataset using the survfit function in the survival package. Receiver operating characteristic (ROC) curves and AUC were calculated using predictions from the CV with ROCR v1.0-7 in R [37]. The final macrophage gene signature used for validation was generated by fitting the full MSKCC dataset with the optimized parameter values.

Validation of gene signatures in independent cohorts

Gene signatures were tested in three independent cohorts by first extracting the coefficients (betas) for the selected genes from the model fit, then multiplying the
scaled gene expression data in the independent datasets by these coefficients. The prospective Decipher cohort contains anonymized genome-wide expression profiles from clinical use of the Decipher test in the radical prostatectomy (RP) setting, between February 2014 to August 2017, retrieved from the Decipher GRID™ (NCT02609269). The retrospective natural history cohort from Johns Hopkins Medical Institute is comprised of men treated with RP, with a median follow up time of 108 months [38]. The Mayo clinic cohort is a retrospective cohort of men treated with RP, with a median follow-up time of 156 months [39, 40]. Model discriminatory capability was assessed based on the AUC. Cox proportional hazards was used to estimate the Hazard Ratio of metastasis-free survival for patients with high signature (top25%).

Public availability of data

Limited and specific single cell RNA sequencing data of patient macrophages can be found at GSE133094. RNA expression data for Mayo (GSE46691, GSE62116) and JHMI (GSE79957) cohorts are also available.

Results

Single-cell analysis of myeloid cells isolated from PCa biopsies

To identify the macrophage populations present in diseased human prostates, fresh biopsies were collected from ‘tumorous’ and ‘non-tumorous’ areas of post radical prostatectomy (RP) specimens. Four previously untreated PCa patients, aged 50-74 years, with an initial serum PSA between 7.6 and 9.3 ng/ml and diagnosed with a Gleason score 6-7 and a clinical stage pT2-3 adenocarcinoma of the prostate were included in this study (Supplementary Table 1). In order to obtain a single cell
suspension for FACS sorting to isolate macrophages, two isolation methods were compared: mechanical dissociation versus enzymatic digestion. The yield of CD45+ leukocytes and CD45+/CD14+ and/or CD11b+ myeloid cells was proportionally higher in mechanically digested tissue compared to enzymatically digested tissue (Supplementary Figure 1). In addition, the percentage of live cells (DAPI-negative) was higher in mechanically digested tissue compared to enzymatically digested tissue (82.8% vs 65.0%). This is in agreement with studies reporting loss of epitopes in immune cells, especially myeloid cells that are known to be sensitive to higher temperature [41-43]. Based in these observations, mechanical dissociation was selected as the optimal method to isolate macrophages from the prostate cancer samples for scRNA-seq analyses.

The procedure for obtaining native PCa associated myeloid cells is depicted in Figure 1A. Tissue resident macrophages were isolated from the biopsies by successively FACS sorting a single-cell suspension of the biopsies for CD45+ leukocytes, followed by isolation of CD3-CD14+ and/or CD11b+ myeloid cells [44-46]. In total, 1920 cells, including 911 cells isolated from the tumorous and 1009 cells isolated from the non-tumorous areas of the prostates were sequenced on 8 plates (Supplementary Table 1). Cells with less than 2000 UMIIs were not considered, while only genes that were detected with at least 4 UMIIs in at least 3 cells, were used for further analysis. In plates 4 and 7, very few cells above the 2000 UMI cut-off were found. Furthermore, high levels of technical artefact genes like KCNQ1OT1, which is a non-coding region rich in poly-A repeats and often found in cell transcripts of poor quality were also detected [47]. For these reasons, plates 4 and 7 were excluded from further analysis. The range (202 – 12,107) and mean (1,806) of genes (features) detected per cell are displayed in Supplementary Figure 2A. After
additional quality control filtering to remove dead, dying, or otherwise low-quality
cells and cell doublets/multiplets, plate 8 was found to contain very few remaining
cells and was also removed from analysis. The removal of plates 7 and 8 resulted in
the loss of all cells from patient 4. Therefore, the subsequent analysis contains cells
from three patients. The remaining 751 cells retained for subsequent analysis
showed a range and mean of 1,026 – 5,876 and 2,787 genes per cell, respectively
(Supplementary Figure 2B).

Subsequently, the accuracy of the mpMRI to annotate tumorous and non-
tumorous areas in the prostate was evaluated by histological assessments of the
H&E stained prostatectomy specimen of patients 1-3. The pre-surgery mpMRI of all
3 patients, indicated areas with a high likelihood of containing tumor and an area
less likely containing tumor in patient 3 correctly. In patient 1, an area less likely
containing tumor was suggested, however, presence of tumor could not be
confirmed (Supplementary Figure 3). Since, biopsies were taken from areas with a
high likelihood of containing tumor and from areas not suspected for containing
tumor in the prostate, we conclude that the biopsies were labelled correctly as
‘tumorous and non-tumorous’.

Clustering of myeloid cells to identify PCa macrophage subtypes

To surmount the implicit noise of individual features in scRNA data, principal
component (PC) analysis was used to reduce dimensionality, followed by graph-
based clustering to identify populations of highly-interconnected cells [30]. Initial
clustering of the data yielded 6 independent clusters, with cluster 3 being
substantially divergent from the remaining clusters (Supplementary Figure 4A). This
was also evident in the first principal component (Supplementary Figure 4B).
Examination of the genes within PC1, as well as markers for various cell types across all clusters revealed that cells in cluster 3 expressed the well-known natural killer (NK) cell markers NKG7 and GNLY (Figure 1B, Supplementary Figures 4C-D) [48]. The presence of these cells after FACS sorting is likely residual from the CD11b (ITGAM) selection (Figure 1B). Since the focus of this study is macrophages, these NK cells were removed from further analysis. The 641 cells in the remaining clusters are considered macrophages as all clusters express various macrophage markers (CD68, CD86, CD163), while lacking expression of established markers for other immune cell types (T-cell, B-cell, NK-cell), prostate epithelium (FOLH1, KLK3), and mesenchymal cells (PDGFRB, FAP) (Figure 1B) [49, 50].

After removal of NK cells, the remaining macrophages were reanalyzed as above, and 20 PCs were selected for further analysis (Supplementary Figure 5). Clustering these PCs yielded 5 populations, however the clustering was highly patient specific (Figures 2A-B). To remove these patient-specific batch effects, a reciprocal PCA method was employed to integrate the patient datasets (see “Methods” for details). This method will effectively integrate correspondence cells between datasets even in the presence of extensive biological or technical differences [30]. Clustering of the final integrated dataset revealed 3 distinct macrophage subtypes (Figure 2C).

The number of macrophage clusters identified will depend on a variety of factors specific to the dataset, including the number of cells and cell types, tissue with which the macrophages are associated, method of digestion, the number of principal components used, and the resolution of the shared nearest neighbors (SNN) graph. The number of macrophage clusters identified in this study was largely comparable to those reported in previous scRNA-seq studies across a variety
of tissues (Supplementary Table 2) [51-58]. The number of PCs used in this analysis were carefully selected to ensure the appropriate amount of biological variation was included (Supplementary Figure 5). To assess the number of clusters identified, increasing resolutions for the SNN graph were tested (Supplementary Figure 6A). This analysis indicated that while increasing the SNN resolution could force additional clusters, the new clusters were in fact subclusters of the three primary subtypes, and no new distinct clusters were identified. While the primary subtypes may indeed have subclusters of specific function, this analysis revealed that subclusters were often defined by a relatively small number of cells or genes in this dataset, and appeared unstable as they often merged back together at different resolutions. Therefore, in this study we focused only on the three primary macrophage subtypes identified.

In the integrated dataset, the cells from each patient were no longer forming isolated or dominant clusters, but were instead distributed across all three clusters, indicating that the reciprocal PCA method was effective at removing the patient-specific batch effects (Figures 2D). Unexpectedly, the macrophages from the tumor and non-tumorous biopsies showed nearly identical distributions among the clusters, and no differences were observed between the macrophage subtypes present in the tumorous and non-tumorous portions of the diseased prostate (Figures 2E-F). In addition, the distributions of CD11b (ITGAM) and CD14 expressing cells were assessed in the three clusters and found to be comparable between the tumorous and non-tumorous cells (Supplementary Figure 6B). Cumulatively these results indicate that there are three biologically distinct macrophage subtypes present in the tumorous and non-tumorous portions of diseased human prostates.
As a first step to examine the identity of these macrophage clusters, previously established markers associated with M1-like and M2-like phenotypes were investigated [24]. Plotting all detectable M1 and M2 marker genes for each cell in a heatmap revealed that many markers are not readily detectable in all cells and there is no clear M1/M2 separation between these clusters (Figure 3A). Given the varying expression levels and the sparsity of marker expression, averaging individual markers within each cluster was not useful in evaluating M1/M2 identity within the clusters, though it appeared that M2 markers were generally expressed at higher levels (Supplementary Figures 7A-B). For these reasons, all M1 and M2 markers were separately combined by averaging the RNA expression of all M1 or M2 markers within each cell. From this analysis it is evident that the mean expression level of all combined M2 markers per cell are higher than the mean M1 marker expression levels (Figure 3B, Supplementary Figure 7C). Furthermore, the mean M1 expression levels were slightly but significantly higher in cluster 2, while the mean M2 expression levels were significantly elevated in cluster 0 (Figures 3C-D). These results demonstrate that while a slightly elevated expression of the averaged M1 and M2 markers can be detected in certain clusters, these are not the main factors contributing to the variation that separates these macrophage populations.

To determine the biological differences between these three macrophage clusters, differential expression analysis was performed to detect marker genes in each cluster. This analysis identified 468 significantly differentially expressed genes, with 164 genes identified as markers in cluster 0, 199 genes in cluster 1, and 105
genes in cluster 2 (Supplementary Table 3). Examining this list of genes showed only 11/68 M1 and M2 markers to be differentially expressed, with 3/33 M1 markers upregulated in cluster 2 and 6/35 M2 markers upregulated in cluster 0 (Supplementary Table 4). These results agree with the slight enrichments observed in Figures 3C-D, however there are also two M2 markers upregulated in cluster 2, further exemplifying the need for better stratification. The most differentially expressed genes from each cluster show either expression only in their cluster, or elevated expression as compared to the other clusters (Figure 4). These genes represent ideal markers for these novel macrophage subtypes.

To evaluate the localization of macrophage subtypes in prostatectomy samples, we scored immunohistochemistry staining of a representative ideal marker from each macrophage cluster, as well as a pan-macrophage marker (CD68), in multiple independent patient samples (Supplementary Figure 8) [59]. Analysis of CD68 control, SLC40A1 for cluster 0, PLAC8 for cluster 1, and FCN1 for cluster 2 suggested that the three macrophage subtypes are present in both the tumorous and non-tumorous portions of prostate cancer patient samples at approximately equal percentages (Supplementary Figure 9A-D). When correcting for the overall percentage of tumor, the density of CD68 and PLAC8 (Cluster 1) positive cells was higher in the non-tumorous portions of the prostatectomy samples than in the tumorous portions, while there was no difference in density of SLC40A1 (Cluster 0) and FCN1 (Cluster 2) positive cells between the two portions (Supplemental Figure 9E).

To explore the functional pathways that genes associated with each cluster are involved in, gene set enrichment analysis (GSEA) was performed (Supplementary Figure 10). Cluster 0 genes showed activation of the hallmark TNFα
signalling via NFkB as well as WNT β-catenin signalling, and suppression of interferon pathways (IFN-α and IFN-γ), MTORC1 signalling, and complement pathways, among others (Supplementary Figure 10A). Conversely, cluster 2 showed activation of multiple inflammatory pathways including IFN-α, IFN-γ, TNFα, and complement, while showing suppression of WNT β-catenin signalling and cell cycle pathways (Supplementary Figure 10B). Cluster 1, however, showed suppression of multiple immune pathways (IFN-α, IFN-γ, TNFα, among others), and activation of cell cycle pathways (E2F targets, MYC targets, G2M checkpoint) as well as MTORC1 signalling (Supplementary Figure 10C). To explore the possibility of PCa TAM regulation of T-cells, known markers of T-cell regulation by TAMs were interrogated in the data [60]. Very few of these markers were readily detectable and only one, CSF1R, was found to be significantly differentially expressed (Supplementary Figure 11). Collectively, these results suggest that each macrophage population is involved in unique biological functions.

**Generation and evaluation of macrophage gene signature**

To develop a prospective gene signature, all genes in the integrated dataset found to be significantly differentially expressed between macrophage clusters were included in the model (Figure 5A). Using a published PCa dataset from Memorial Sloan Kettering Cancer Center (MSKCC) [34], a 217-gene prognostic signature (Supplementary Table 5) predicting biochemical relapse-free survival of PCa patients was selected (see “Methods” for details). The performance of the macrophage gene signature was evaluated employing 10-fold nested cross-validation on the MSKCC dataset using receiver-operator characteristic (ROC) as a performance measure (Figure 5B). As expected, in a Cox regression analysis this classifier showed a
significant association with relapse-free survival (hazard ratio (HR) = 4.1, p = 1.7e-05) (Figure 5C). Furthermore, in a multi-variate analysis including the signature with Gleason score (biopsy and pathological), pre-diagnosis biopsy PSA levels, seminal vesicle invasion (SVI), extracapsular extension (ECE), and clinical stage the signature was found to be an independent predictor of outcome (Figure 5D).

Using the Cox model linear predictor as a prognostic index (PI), the relative prognostic value of each macrophage cluster was assessed by summing the product of the model coefficients (betas) and the scaled gene expression values in each cell (Supplementary Figure 12). This analysis demonstrated that cells from cluster 2 had a low prognostic index, while cells from clusters 0 and 1 had a high prognostic index. This result indicates that higher numbers of cells from cluster 2 are associated with better outcome, while higher numbers of cells form clusters 0 and 1 are associated with worse outcome. Taken together with the pathway analysis in Supplementary Figure 10, these results indicate that the pro-inflammatory macrophage subsets are associated with better outcome, while anti-inflammatory and proliferative macrophage subtypes are associated with worse outcome.

Validation of gene signature in independent PCa cohorts

To further assess the prognostic value of the macrophage gene signature, it was validated in three independent cohorts from the Decipher GRID registry. The first cohort is a prospective Decipher GRID cohort containing RNA expression data from >5,000 radical prostatectomy (RP) patients and includes basic demographic and pathological data, but not longitudinal clinical outcomes. This cohort was used to associate the signature to Decipher risk groups and pathological Gleason score (Figure 6A-B). Since this cohort has no metastasis outcome yet, high Decipher group
was used as a surrogate of metastasis potential since it was heavily validated for that endpoint [38, 40, 61]. The second cohort is a retrospective natural history cohort (n=355) comprised of men treated with RP at Johns Hopkins Medical Institutions (JHMI) [38]. The third cohort is a retrospective cohort (n=780) of men treated with RP at the Mayo Clinic [40, 61]. All three cohorts are described in Supplementary Table 6.

The strength of association of the macrophage signature with metastasis-free survival was tested using a Cox regression analysis on the Mayo and JHMI cohorts and the signature showed significant association with metastasis-free survival (Mayo: HR = 1.89, p-value = 1.0e-06; JHMI: HR = 2.25, p-value = 3.3e-05) (Figure 6C-D). In both cohorts, the classifier was also found to be an independent predictor of metastasis in multivariate analysis (Supplementary Table 7). Taken together, these results indicate that profiling single-cell RNA expression in PCa associated macrophages and identifying subpopulations present in the diseased prostate can have significant prognostic value in predicting patients' likelihood of biochemical relapse and metastasis. These results lay the foundation for profiling macrophage populations in prostate cancer and other cancer types, and will inform future studies investigating the immune systems’ role in cancer progression.

**Discussion**

Macrophages can either promote or suppress cancer development and progression depending on their specific phenotype and function. In this study, we defined the degree of human PCa-specific macrophage diversity through single-cell sequencing with the aim to identify PCa-specific macrophage populations. Three macrophage subtypes were identified, and while some canonical M1 and M2...
markers were present, these were not adequate to define the clusters. The distinction between inflammatory M1 and anti-inflammatory / proliferative M2 macrophages was based on in vitro cell line models and describe the two extremes of the spectrum of macrophage differentiation [21]. Our findings suggest that the M1/M2 distinction of macrophage differentiations does not accurately recapitulate native prostate cancer associated macrophages, which should be considered in future studies.

Our results are largely in line with a recent single-cell study describing cell types in human prostate cancer [58]. Here the authors showed evidence that TAMs express a mixture of M1 and M2 markers, and that two of the main sources of variation between TAM subtypes are TNFα and NFkB pathways. This study identified five macrophage clusters, suggesting that surveying more patients and sequencing higher numbers of cells may further our understanding of the role TAMs play in prostate cancer. Of note for comparison of our study with previous studies, are the different methods of obtaining single cell solutions. While, most previous studies performed enzymatic digestion, we performed mechanical dissociation. This difference in techniques, might introduce a bias for particular cell populations. However, in agreement with previous studies [41-43], we demonstrated that mechanical digestion leads to a higher number of live myeloid cells as compared to enzymatic digestion. This becomes particularly relevant in poorly infiltrated tumors, such as prostate cancer, where the immune cell content, specifically macrophages, is already reduced.

Multiple reports demonstrated a correlation between TAMs and poor prognosis of PCa patients [62-64], however, the vast majority of these studies only focused on a small selection of TAM-associated markers, including IL-10, CD163
and MRC1 (CD206). Other than previous studies, we constructed signatures and clustering of macrophages based on the transcriptional profile of macrophages as they occur in all their complexity in the human PCa-microenvironment. Using the genes differentially expressed between the clusters, we were able to develop a gene signature with significant prognostic value in multiple independent PCa cohorts. These results advance the field not only by defining TAMs subtypes, but also by demonstrating that the genes differentially expressed between these subtypes can predict a PCa patient’s prognostic outlook in terms of biochemical recurrence and metastasis.

Macrophages were isolated from biopsies from the ‘tumorous’ and ‘non-tumorous’ sites of the prostate. Sites of the biopsies were identified using pre-surgery mpMRI images. Although the presence of tumor was not histologically confirmed in the biopsies themselves, histological evaluation of the whole prostatectomy specimen confirmed that the mpMRI images correctly identified tumorous areas in all three patients. This confirms correct labelling of biopsies as tumorous or non-tumorous. Remarkably, no differences were observed between macrophage subtypes found in the tumorous and the non-tumorous sites, suggesting that tumorigenic factors may also affect distant non-tumorigenic sites. Furthermore, this suggests that these macrophage populations could in theory be detected from a biopsy regardless to tumor cell percentage. This is important because the prognostic value of the gene signature outweighs the biopsy Gleason score and pre-diagnosis biopsy PSA levels, and is approaching the significance of pathological Gleason, suggesting a possible path to identifying high-risk patients without necessitating radical prostatectomy. Additionally, the macrophage signature and pathological Gleason score were both independent predictors in our multivariate analysis,
suggesting that the signature can provide additional prognostic information. However, these finding will require further experimental validation before such measures could be employed.

The gene set enrichment analysis performed in this study suggests that each macrophage subtypes is involved in unique biological processes. Cluster 1 does not appear to be participating in inflammatory pathways and may represent a proliferative feeder cell type, or otherwise less differentiated macrophage subtype. Cluster 0 appears to be largely anti-inflammatory, while cluster 2 appears primarily pro-inflammatory. These results agree with the notion that macrophages can broadly adopt either a pro-inflammatory or anti-inflammatory phenotype, and this could either potentiate or mediate cancer progression [6, 7, 24]. Moreover, since there was no difference in the occurrence of the three clusters in biopsies from the tumorous and non-tumorous site of the prostate, the differences in functions of the macrophages in the three clusters may be related to the origin of the macrophages, which could either be embryonic-derived or blood-derived [4].

Furthermore, using a prognostic index we demonstrate that pro-inflammatory cluster 2 cells are associated with better prognosis, while the anti-inflammatory cluster 0 and anti-inflammatory/proliferative cluster 1 cells are associated with worse outcome. This suggests that targeting cluster 2 cells to increase their numbers and targeting clusters 0 and 1 to decrease their numbers is a potential therapeutic strategy for PCa patients. To this purpose, our findings provide novel drug targetable genes specific for each cluster. It will be important for future studies to explore the role these macrophage populations play in prostate cancer, and to investigate targeting these subtypes and their associated pathways with immunotherapy.
Cancer immunotherapies, specifically those inhibiting T cell immune checkpoints, have generated significant impact in recent years, with established efficacy in advanced melanoma [65, 66], non-small cell lung cancer [67, 68] and bladder cancer patients [69, 70]. However, in other cancers, including PCa, immunotherapy efficacy is limited [71, 72]. The uncertain therapeutic efficacy of immunotherapy in PCa is partly due to a poor infiltration of immune cells in the TME [16, 73-76]. Moreover, TAMs display an ability to modulate tumor immunity by suppression of T cell recruitment and function, though the precise mechanisms have yet to be elucidated [60]. Several direct and indirect suppressive actions of macrophages on T cell functions have been suggested, including involvement of immune checkpoints ligands (e.g.: PDL1, B7-H4), cytokines (e.g.: IL-10, CXCL10, CCL22) and cell surface receptors (e.g.: CD206, CSF1R) [60]. However, in this study, only the colony-stimulating factor 1 receptor (CSF1R), which is a key regulator of immunosuppressive macrophage expansion, was found to be enriched in cluster 0. Whether the macrophage subtypes discovered in this study play a role in T cell regulation will be an important question for future studies.

Limitations of this study include the small number of patients included in the study and the absence of assessment of protein expression of the key selected genes. To this end, future studies should include immunohistochemistry analysis to further support our findings.

In conclusion, in this study we demonstrate the relevance of using single-cell transcriptomics from PCa-associated macrophages as a prognostication strategy for individual patients. We propose that targeting unique tumor-associated macrophage subtypes, as opposed to all macrophages, can provide a therapeutic avenue to combat prostate cancer and potentially other cancer types.
Acknowledgements

We thank our collaborators from Single Cell Discoveries B.V. supported by the Hubrecht Institute and the Oncode Institute.

Funding details

This work was supported by the FP7 MCA-ITN Grant agreement 317445 – TIMCC and KWF Dutch Cancer Society Grant number 2009-4356 awarded to Andre Bergman.

References


**Figure 1. Single cell sequencing of myeloid cells from prostate cancer patients.** (A) Experimental procedure: single cell suspensions of multiple biopsies from the tumor and tumor-adjacent portions of radical prostatectomy specimens were sequentially FACS sorted to isolate CD45+/CD14+ and/or CD45+/CD11b+ myeloid cells, after negative selection for CD45+CD3+ lymphocytes. Subsequently, single cells were processed for SORT-seq analysis. (B) Dot plot for average expression levels of markers for macrophages, other immune cells, and prostate cell types present in the clusters. Cells were positively identified as macrophages and NK cells.

**Figure 2. Identification of prostate cancer-associated macrophage subtypes.** (A) UMAP projections of macrophage clustering after NK cell removal. Colors indicate five distinct clusters. (B) UMAP projections of cells from each patient reveals strong patient-specific batch effects. Each color represents cells from a different patient. (C) UMAP projections after reciprocal-PCA integration removing batch effects between patients. Colors indicate three distinct clusters. (D) UMAP projection of integrated data shows all patient samples are comparable and do not cluster separately. Each color represents cells from a different patient. (E) Distributions of cells within each cluster, with percentage of cells from each cluster show on the y-
axis. (F) UMAP projection of macrophages isolated from tumor (pink) and tumor
adjacent (blue-green) biopsies are comparable and do not cluster separately.

Figure 3. Canonical M1 and M2 macrophage marker expression in prostate
cancer associated macrophages. (A) Heatmap of all detectable canonical markers
used to differentiate M1 and M2 macrophages by gene expression in relation to the
three macrophage clusters. Cell surface markers are highlighted in bold. Color bar
indicates RNA expression level scaled by z-score. (B) Mean expression of all M1
(or M2) markers from (A) averaged together within each cell. p-
values from Welch’s t-test. (C) Average normalized expression of all M1 markers in
(A) per cell within each cluster. p-values from One-way ANOVA with post-hoc Tukey
HSD. (D) Average normalized expression of all M2 markers in (A) per cell within
each cluster; p-values from One-way ANOVA with post-hoc Tukey HSD.

Figure 4. Identification of marker genes for each macrophage cluster. (A)
Heatmap of the top 10 most differentially expressed genes in each macrophage
cluster with expression level scaled to z-score. (B) Violin plots for the top 5 most
differentially expressed genes in each cluster showing normalized RNA expression
levels. Colors indicate macrophage clusters and each black dot represents a cell.

Figure 5. Gene signature from differentially expressed macrophage genes has
significant prognostic value. (A) Heatmap of all significant differentially expressed
genes present in the integrated dataset by macrophage cluster (Bonferroni adjusted
p-value <0.05). Color bar indicates integrated expression level scaled by z-score. (B)
Receiver operating characteristic (ROC) curve for 10-fold cross-validation of
macrophage signature in MSKCC dataset, colorized by threshold value. (C)

Macrophage gene signature outcome predictions from 10-fold cross-validation in MSKCC dataset. Colors indicate patient subpopulations of predicted high (red; 25%), medium (blue; 50%) and low (green; 25%) risk of recurrence. p-value is from log-rank test. (D) Multivariate Cox regression analysis for macrophage signature in MSKCC cohort showing hazard ratio with 95% confidence interval and p-value from log-rank test.

**Figure 6. Validation of the gene signature in independent prostate cancer cohorts.** (A) Macrophage gene signature predictions correlate in a prospective cohort of >5000 patients with high Decipher GRID scores (blue) vs. low and intermediate scores (grey). (B) Macrophage gene signature predictions correlate with pathological Gleason score in Decipher GRID. p-value from linear model between Gleason and Macrophage signature. (C-D) Gene signature outcome predictions in two independent PCa cohorts from JHMI (C) and Mayo Clinic (D), stratifying patients on low (green, bottom 25%), medium (blue; middle 50%) or high (red; top 25%) risk of metastatic disease. Y-axis shows metastasis-free survival, number of patients at risk are indicated, p-values from log-rank test.
Figure 3

A

B

C

D

Figure 3
Figure 4

A

B
Figure 5

A: Differentially expressed genes

B: 10-fold nested cross-validation ROC

C: Strata

D: Hazard ratio

<table>
<thead>
<tr>
<th></th>
<th>Hazard ratio</th>
<th>p-value</th>
<th>CI</th>
<th>Number at risk</th>
<th>Time</th>
<th>Survival</th>
</tr>
</thead>
<tbody>
<tr>
<td>Macrophage_signature</td>
<td>1.38 (1.03 - 1.9)</td>
<td>0.029 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biopsy_Gleason_score</td>
<td>1.07 (0.59 - 1.9)</td>
<td>0.818</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pathological_Gleason_score</td>
<td>2.46 (1.27 - 4.8)</td>
<td>0.007 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PSA_at_diagnosis</td>
<td>0.75 (0.23 - 2.5)</td>
<td>0.645</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seminal_vesicle_invasion</td>
<td>1.81 (0.58 - 5.7)</td>
<td>0.309</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extracapsular_extension</td>
<td>0.85 (0.30 - 2.4)</td>
<td>0.752</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surgical_margins</td>
<td>1.61 (0.65 - 4.0)</td>
<td>0.308</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Events: 27; Global p-value (Log-Rank): 7.0175e-06**

AIC: 218.07; Concordance Index: 0.8
Figure 6

A

Prospective Decipher GRID cohort (n=5,239)

Cumulative Percentage %

<table>
<thead>
<tr>
<th>Low 10%</th>
<th>Macrophage signature (10%)</th>
<th>Top 10%</th>
</tr>
</thead>
<tbody>
<tr>
<td>90%</td>
<td>86%</td>
<td>69%</td>
</tr>
<tr>
<td>91%</td>
<td>80%</td>
<td>73%</td>
</tr>
<tr>
<td>90%</td>
<td>27%</td>
<td>58%</td>
</tr>
<tr>
<td>14%</td>
<td>22%</td>
<td>31%</td>
</tr>
<tr>
<td>10%</td>
<td>50%</td>
<td>67%</td>
</tr>
</tbody>
</table>

B

Prospective Decipher GRID cohort (n=5,239)

Pathological Gleason

Macrophage signature

p < 0.0001

C

JHMI cohort (n=355)

Metastasis-free Survival

Time (months)

<table>
<thead>
<tr>
<th>Number at risk</th>
<th>Low 25%</th>
<th>Middle 50%</th>
<th>Top 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low 25%</td>
<td>89</td>
<td>78</td>
<td>40</td>
</tr>
<tr>
<td>Middle 50%</td>
<td>177</td>
<td>148</td>
<td>76</td>
</tr>
<tr>
<td>Top 25%</td>
<td>89</td>
<td>59</td>
<td>25</td>
</tr>
</tbody>
</table>

D

Mayo cohort (n=780)

Metastasis-free Survival

Time (months)

<table>
<thead>
<tr>
<th>Number at risk</th>
<th>Low 25%</th>
<th>Middle 50%</th>
<th>Top 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low 25%</td>
<td>195</td>
<td>171</td>
<td>124</td>
</tr>
<tr>
<td>Middle 50%</td>
<td>390</td>
<td>320</td>
<td>231</td>
</tr>
<tr>
<td>Top 25%</td>
<td>195</td>
<td>139</td>
<td>79</td>
</tr>
</tbody>
</table>

p < 0.0001
Molecular Cancer Research

The Prognostic Potential of Human Prostate Cancer-Associated Macrophage Subtypes as Revealed by Single-cell Transcriptomics

Joseph C Siefert, Bianca Cioni, Mauro J Muraro, et al.

Mol Cancer Res  Published OnlineFirst June 15, 2021.

Updated version  Access the most recent version of this article at:
doi:10.1158/1541-7786.MCR-20-0740

Supplementary Material  Access the most recent supplemental material at:
http://mcr.aacrjournals.org/content/suppl/2021/06/11/1541-7786.MCR-20-0740.DC1

Author Manuscript  Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, use this link http://mcr.aacrjournals.org/content/early/2021/06/14/1541-7786.MCR-20-0740. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.