Combined genome and transcriptome analysis of single disseminated cancer cells from bone marrow of prostate cancer patients reveals unexpected transcriptomes

Miodrag Gužvić¹,* , Bernhard Braun¹,²,* , Roman Ganzer³ , Maximilian Burger³ , Michael Nerlich⁴ , Sebastian Winkler⁵ , Melanie Werner-Klein⁶ , Zbigniew T. Czyż⁷ , Bernhard Polzer⁷ , Christoph A. Klein¹,⁷

¹Chair of Experimental Medicine and Therapy Research, University of Regensburg, Regensburg, Germany
²Present address: Department of Oncology, Hospital Barmherzige Brüder, Regensburg, Germany
³Department of Urology, Caritas-Hospital St. Josef, University of Regensburg, Regensburg, Germany
⁴Department of Trauma Surgery, University Hospital Regensburg, Regensburg, Germany
⁵Department of Orthopaedic Surgery, University of Regensburg, Bad Abbach, Germany
⁶Department of Immunology, University of Regensburg, Regensburg, Germany
⁷Project Group Personalized Tumour Therapy, Fraunhofer Institute of Experimental Medicine and Toxicology, Regensburg, Germany

*Miodrag Gužvić and Bernhard Braun contributed equally to this work

Running title: Transcriptional profiling of DCCs from bone marrow
Keywords: bone marrow, disseminated cancer cells, epithelial cell adhesion molecule, prostate cancer, transcriptome analysis

Grant support: This work was supported by the awards from Prostate Cancer Foundation and Dr. Josef Steiner Cancer Research Foundation Award, and grants from EU (FP6 STREP Research Consortium Migrating Cancer Stem Cells - MSCSs), Deutsche Krebshilfe (Research Consortium Tumour Stem Cells), and Bavarian Genome Research Network (BayGene) to C.A.K.

Corresponding Author: Christoph A. Klein, Chair of Experimental Medicine and Therapy Research, University of Regensburg, Franz-Josef-Strauss-Allee 11, 93055 Regensburg, Tel: +499419446720, E-mail: christoph.klein@ukr.de

Disclosure of Potential Conflicts of Interest: The authors disclose no potential conflicts of interest

Manuscript: Abstract, 217 words; Text, 4904 words; 6 figures and 1 table
Abstract

Bone is the most frequent site of metastasis in prostate cancer (PCa) and patients with bone metastases are deemed incurable. Targeting prostate cancer cells that disseminated to the bone marrow (BM) prior to surgery and before metastatic outgrowth may therefore prevent lethal metastasis. This prompted us to directly analyse the transcriptome of disseminated cancer cells (DCC) isolated from non-metastatic (UICC stage M0) prostate cancer patients. We screened 105 BM samples of M0-stage prostate cancer patients and 18 BM samples of patients without malignancy for the presence of EpCAM+ single cells. In total we isolated 270 cells from both groups by micromanipulation and globally amplified their mRNA. We used targeted transcriptional profiling to unambiguously identify DCCs for subsequent in-depth analysis. Transcriptomes of all cells were examined for the expression of *EPCAM, KRT8, KRT18, KRT19, KRT14, KRT6a, KRT5, KLK3 (PSA), MAGEA2, MAGEA4, PTPRC (CD45), CD33, CD34, CD19, GYPC, SCL4A1 (band 3), and HBA2*. Using these transcripts we found it impossible to reliably identify true DCCs. We then applied combined genome and transcriptome analysis of single cells and found that EpCAM+ cells from controls expressed transcripts thought to be epithelial-specific, while true DCCs may express haematopoietic transcripts. These results point to an unexpected transcriptome plasticity of epithelial cancer cells in bone marrow and question common transcriptional criteria to identify DCCs.
Introduction

Despite substantial investment into cancer research, metastasis - the cause of more than 90% of cancer-related deaths (1) - remains poorly understood. Delineating early steps of metastasis before detection of manifest lesions clinical imaging techniques may provide chances to prevent lethal outcome. After a short period of circulation in the blood (when they are termed circulating tumour cells - CTCs), disseminating cancer cells lodge in distant organs and are termed disseminated cancer cells (DCCs). Obviously, cancer cells remaining in the body of M0-stage patients after surgical removal of primary tumour comprise founders of later arising lethal metastases and therefore DCCs are the target population of systemic adjuvant therapies. In support of this notion are numerous reports on the prognostic impact of DCCs, which are commonly detected by the histogenetic markers, EpCAM or cytokeratins (CK), in lymph nodes and bone marrow (BM), respectively. In contrast to CTCs, which are mostly detected and studied in patients with metastases (2), DCCs are clearly relevant in early disease stages (3-5).

We previously showed that the presence of CK-positive (CK⁺) DCCs in the BM of M0-stage prostate cancer (PCa) patients is associated with shorter metastasis free survival (5). Genetic studies of isolated CK⁺ DCCs demonstrated marked heterogeneity of M0-stage DCC genomes within and between patients (5, 6). However, little is known about phenotypic characteristics of DCCs. Survival data suggest the existence of DCCs with different metastasis-initiating potential, since not all patients with DCCs develop metastasis. Furthermore, the fact that some patients without CK⁺ cells in BM develop metastasis, points to the existence of subpopulations of DCCs undetected by current markers. This is particularly true for PCa where the increase of PSA after radical prostatectomy (biochemical relapse) is not associated with increasing numbers of CK⁺ cells in BM (5). Therefore, targeting metastasis founder cells requires detailed molecular
characterization of DCCs. One hindrance to such characterization is the low frequency of M0-stage DCCs, with a median number of 1-2 DCC per $10^6$ BM cells of positive patients.

Previously, we established an approach to comprehensively characterize the transcriptome of a single cell using microarrays (7). Since cytokeratins are intracellular proteins, comprehensive analysis of a single DCC transcriptome requires use of a surface marker such as EpCAM for their detection. Furthermore, detection of EpCAM-positive (EpCAM$^+$) cells in BM has also been associated with PCa progression (8). However, unlike CK$^+$ cells (9), EpCAM$^+$ cells are often found in the BM of healthy individuals (10, 11), thereby confounding molecular studies of DCCs. Hence, there is an urgent need for an assay to reliably select true DCCs for transcriptome analysis using microarrays. Here, our aim was to detect and isolate EpCAM$^+$ cells from the BM of M0-stage PCa patients and healthy controls for detailed characterization of their gene expression. The objective of this study was to examine the expression of a panel of potentially informative genes that might discriminate individual EpCAM$^+$ cells found in BM of PCa from EpCAM$^+$ BM cells of cancer-free patients. In particular, we assessed the expression of selected epithelial (EPCAM, KRT8, KRT18, KRT19, KRT5, KRT6a, KRT14), prostate-specific (KLK3 [PSA]), tumour-specific (MAGEA2 and MAGEA4), common hematopoietic (PTPRC [CD45], CD34, CD33, CD19) and erythroid (GPYC, SCL4A1, HBA2) transcripts aiming to separate prostatic epithelial and hematopoietic lineages.
Material and methods

The workflow of this study is given in the Supplementary Figure S1.

Patients and BM samples

The local ethics committee of the University of Regensburg approved all aspects of the study (ethics vote number 07-079). Between February 2009 and November 2012, BM aspirates from 105 non-metastatic (M0-stage) PCa patients were collected shortly before radical prostatectomy (RP). Additionally, we obtained BM samples of two metastatic (M1-stage) PCa patients. As a control group, 18 BM samples from cancer-free males undergoing trauma or orthopaedic surgery were obtained. After obtaining written informed consent, the sampling was performed by aspiration of left and right iliac crest (12). The samples were transported to the laboratory within 3 hours and subjected to further processing. Baseline patient characteristics are provided in Table 1.

Processing of BM samples

Detailed description of BM processing is given in Supplementary Methods. Briefly, upon arrival, the BM sample was twice washed with Hank's salt solution to remove fat and thrombocytes. Next, cell suspension was centrifuged in density gradient. After centrifugation, the interphase containing mononuclear cells (MNCs) was carefully collected and washed with phosphate buffered saline (PBS). The number of MNCs and erythrocytes was determined on hemocytometer. To enrich the DCC-containing fraction the sample was depleted of the majority of hematopoietic cells using negative immunomagnetic selection. This was done by incubating the cell suspension with APC-conjugated antibodies against CD11b, CD33, and CD45. After incubation and washing, the cell suspension was incubated with anti-APC beads and anti-CD235a beads (glycophorin A). After incubation and washing, the cell suspension was run through the 40 μm cell sieve,
and then run on LS MACS column. The eluate, containing the unlabelled cell fraction, was collected on ice and the cell number determined using a haemocytometer.

**Staining of BM samples and cell lines**

Detailed description of staining procedure is given in Supplementary Methods. On average, two million BM cells were stained with anti-EpCAM-PE (HEA125, Miltenyi Biotec) antibody. Cell lines (WPE-Stem, DU145, PC3, WPE-Int, LNCaP and VCaP) were grown according to instructions from ATCC. At 80% confluence, the medium was discarded, cell monolayer washed with PBS, and cells fixed using 4% formaldehyde. Further immunostaining steps were performed as described with the addition of nuclear counterstaining using DAPI (Supplementary Figure S2).

**Screening of BM and isolation of single cells and cell pools**

A detailed description of the screening and cell isolation procedures is given in Supplementary Methods. Each BM sample was manually screened for the presence of EpCAM$^+$ cells on an inverted fluorescent microscope (Olympus or Zeiss), equipped with micromanipulator (Patchman NP2, Eppendorf) and pump (CellTram, Eppendorf). Single cells with preserved integrity and positively stained for EpCAM were extracted using a glass capillary attached to the micromanipulator. By visual inspection we ensured that only one cell was in the capillary. The cell was transferred to an empty field with PBS and manually isolated using a micropipette. After isolation of single EpCAM$^+$ cells from each sample we isolated a pool of ~2000-3000 cells. As a reagent control of cell isolation, 1 μl of PBS in which individual EpCAM$^+$ cells were isolated, were taken for subsequent WTA.

Photographs of the cells prior to isolation were made in 10 cases, using the software cellSens (Olympus) or AxioVision (Zeiss). Due to the rapid bleaching of PE fluorophore we abstained from taking photographs of all isolated cells.
Flowcytometric analysis of BM cells was performed with a LSR II machine equipped with FACS DIVA 5.03 software (BD Bioscience) and data was analyzed with FloJo 8.8.6 (Treestar).

**Whole transcriptome amplification (WTA) of single cells**

Isolation of mRNA from a single cell, reverse transcription, and global amplification of the first strand cDNA was carried out as previously described (7, 10). Details are provided in Supplementary methods. The quality of amplification was examined by performing end-point PCR for ACTB, EEF1A1 and GAPDH transcripts (see below and Supplementary Table S1).

**Whole transcriptome amplification (WTA) controls**

From all samples WTA controls were generated by aspirating 1 μl buffer without a cell from the slide after DCC isolation and equally subjected to WTA. After WTA we controlled for the presence of ACTB, EEF1A1 and GAPDH transcripts. Only when WTA controls were negative, cells were included into transcriptome analysis. In 2 out of 125 (2%) BM samples, the WTA control was contaminated. The WTA samples of these two patients were excluded from genome and transcriptome analysis.

**Whole genome amplification (WGA) and comparative genomic hybridization (CGH)**

Single cell genomic DNA collected during WTA procedure was precipitated and subjected to WGA, using the Ampli1™ WGA Kit (Silicon Biosystems), as described previously (10, 13). Details are provided in Supplementary methods. CGH on metaphases (cCGH, n=24 samples) and Agilent microarrays (aCGH, n=18 samples) was performed as described previously (10, 13, 14). Array files are deposited in GEO database under accession number GSE59631.
Expression analysis of selected transcripts by end-point PCR in patient samples

All WTA products were tested by end-point PCR using conditions outlined in Supplementary Methods. List of primers is given in Supplementary Table 1. To establish the correct identity of each amplified fragment, we have performed restriction digestion, using the conditions outlined in Supplementary Methods. Only after a positive result after digestion, the sample was declared positive for a given transcript.

Statistical analyses

Statistical analyses were performed using SPSS v22 (IBM). Hierarchical clustering was performed using the software tool Cluster (15).
Results

Detection of EpCAM$^+$ cells in BM samples and relation to clinical variables

We investigated the presence of putative DCCs in BM samples of 105 M0-stage PCa patients undergoing radical prostatectomy by staining for EpCAM. Additionally, we stained BM samples from two M1-stage PCa patients and 18 male control patients without cancer (Supplementary Figure S1). To enrich the epithelial target population, we first depleted BM samples of CD45$^+$, CD33$^+$, CD11b$^+$, and glycophorin A$^+$ cells. On average, 86% of MNC were removed by immunomagnetic sorting (Supplementary Figure S3). Then, we used direct immunofluorescence to detect EpCAM$^+$ cells. Screening on average 2 x 10$^6$ cells per BM sample (median 2 x 10$^6$, range 0.02 – 6.8 x 10$^6$ cells; Supplementary Figure S3) revealed that 65 (62%) of M0-stage patients, as well as 10 (56%) of control patients had EpCAM$^+$ cells in the BM (p=0.79, Fisher’s exact test, two tailed; Supplementary Table S2). Both analyzed M1-stage patients also harboured EpCAM$^+$ cells in the BM.

Clinical characteristics of the M0-stage cohort before radical prostatectomy are shown in Table 1. There was no significant association between the presence of EpCAM$^+$ cells in the BM of M0-stage PCa patients and any clinical risk factor such as Gleason’s score or pre-operative PSA (Table 1). Likewise, grading (G), resection margin (R), lymphatic invasion (L), or vein invasion (V), and perineural infiltration (PnI) were not associated with detection of EpCAM$^+$ cells (data not shown).

EpCAM staining of BM samples from patients and controls

EpCAM$^+$ cells from PCa patients (Figure 1I-X) usually displayed stronger staining intensity compared to cells from controls (Figure 1A-H). However, solely based on staining intensity, size, or morphology, one would not be able to reliably identify true DCCs. Due to rapid fading of phycoerythrin, we abstained from taking pictures of each cell or to perform
real-time measurements of fluorescence intensity. Still, comparison of available photographs of EpCAM+ cells from patients and controls, with EpCAM-stained prostate epithelial cell lines (Supplementary Figure S4) suggests that EpCAM+ cells from BM of controls and M0-stage patients display a staining pattern similar to cell lines with low levels of EpCAM expression such as PC3 or DU145 (Supplementary Figure S4E-P). To corroborate these observations, we used a BM sample of a cancer-free patient, depleted it of CD45+, CD33+, CD11b+, and glycophorin A+ cells, and divided the sample in half. We spiked one half with 2% of EpCAM-positive DU145 PCa cell line cells. Then, spiked and non-spiked samples were again divided and stained using a EpCAM-PE antibody or isotype-PE antibody. No fluorescent signal was observed in samples stained with isotype-PE antibody (Supplementary Figure S5A-H), whereas weakly EpCAM-positive cells could be seen in both samples stained with EpCAM-PE antibody (Supplementary Figure S5I-L). As expected, these cells displayed slightly weaker EpCAM-positivity compared to some of DU145 cells (Supplementary Figure S5M-P). Still, only the top 17% of the spiked-in DU145 population (i.e. 0.34% of all EpCAM+ cells) could be identified according to their higher EpCAM-positivity by flow-cytometric analysis (Supplementary Figure S5Q). This shows that the intensity of the EpCAM-staining can rarely be used to discriminate EpCAM+ true DCC from EpCAM+ BM cells.

Isolation of EpCAM+ single cells and whole genome and transcriptome amplification

On average, we isolated 4 individual EpCAM+ cells from each positive BM sample (Supplementary Figure S3). From 65 EpCAM+ BM samples of M0-stage PCa patients we isolated 220 EpCAM+ single cells (examples shown in Figure 1I-P) and 5 clusters of 2-5 cells (examples shown in Figure 1Q-T), and 12 EpCAM+ single cells and 4 clusters of 2-5 cells from two M1-stage PCa patients (Figure 1U-X). Fifty-one individual EpCAM+ cells were collected from 10 control patients (Figure 1A-H; Supplementary Table S2). BM of
controls never contained EpCAM$^+$ cell clusters. After isolation of EpCAM$^+$ cells, we isolated a pool of ~2000 - 3000 cells from each BM sample.

Next, we extracted and separated total genomic DNA (gDNA) and mRNA from each single cell, cell cluster or cell pool. Total mRNA was reversely transcribed and subjected to whole transcriptome amplification (WTA), while gDNA was subjected to whole genome amplification (WGA). For analysis of single cell transcriptomes only high quality samples were selected based on our quality control (QC) criteria (i.e. positive for the three QC PCR transcripts ACTB, EEF1A1 and GAPDH, see Materials and methods).

The three groups (M0, M1, controls) did not differ with regard to cDNA quality (p=0.75, Pearson Chi-square; Supplementary Table S2 and Supplementary Figure S7A). After WTA and QC, we selected 119 EpCAM$^+$ single cells and 5 clusters of EpCAM$^+$ cells (55% of samples) from M0-stage patients. From controls we selected 28 (55%) of EpCAM$^+$ single cells, and from M1-stage patients 7 EpCAM$^+$ single cells and 3 clusters of EpCAM$^+$ cells (62% of samples; Supplementary Table S2 and Supplementary Figure S7A). In addition, high-quality WTA products of BM cell pools from 10 randomly selected M0-stage patients, one M1-stage patient, and 10 controls were chosen for further analysis.

**Expression profiling for identification of a DCC signature in M0 stage patients**

We started the search for a DCC specific signature by expression analysis of selected genes in BM cell pools of PCa patients and controls. We could not detect significantly different frequencies for epithelial-, prostate-, tumour- and erythroid transcript expression between the analysed cell pools (Supplementary Table S3, Supplementary Figure S6 and Supplementary Figure S8D-F). Surprisingly, among common hematopoietic transcripts, only CD19 was expressed in a significantly higher number of samples from M0-stage patients as compared to pools from healthy controls (p<0.05, Fisher's exact test).
Since the analysis of BM cell pools revealed no transcript specific for cancer patients or controls, we focussed on expression profiling of selected transcripts in EpCAM+ single cells. We first analyzed the expression of EPCAM and cytokeratins transcripts. The number of EpCAM+ single cells expressing the EPCAM transcript was significantly higher among cells from M0-stage patients (88%), than among cells from controls (50%; p<0.001, Pearson Chi-square; Figure 2, Supplementary Table S3, and Supplementary Figure S7B).

Forty-two percent of cells from M0-stage patients expressed at least one KRT transcript (Supplementary Figure S7B), which was significantly different to cells from controls, where only 7% of cells expressed KRT transcripts (p<0.01, Chi-square). The difference between cancer patients and controls was even more pronounced, when we analyzed KRT expression in EPCAM-transcript-positive cells. We found that cells from M0-stage patients more frequently co-expressed EPCAM and KRT transcripts, compared to cells from controls (p<0.002, Fisher's exact test; Supplementary Table S4). Among KRT-expressing cells from M0-stage patients, the most frequently detected KRT transcripts were KRT8 and KRT18 (Figure 2, Supplementary Table S3, and Supplementary Figure S7C). These transcripts were detected significantly more often in cells from M0-stage patients than in controls (KRT8 – 30% vs. 4%, p<0.01; KRT18 – 27% vs. 7%, p<0.05; Pearson Chi-square; Supplementary Table S3). KRT19, KRT14 and KRT6a transcripts were found in less than 5% of cells from M0-stage patients, and were absent in cells from controls. No KRT5 transcripts were detected in cells of either group (Supplementary Table S3, and Supplementary Figure S7C). In total, among cells from M0-stage patients, 93% of cells expressed at least one of the tested epithelial transcripts, compared to 57% cells from controls (p<0.001, Fisher's exact test). This distribution closely resembled the frequency of EPCAM expression observed in these two groups of cells (Supplementary Figure S7B).
Next, we analyzed prostate- and tumour specific transcripts. KLK3 (PSA) transcript was found only in cells isolated from one M0-stage patient, whose BM cell pool also expressed KLK3. Overall, this transcript was detected in 9 (7%) M0-stage cells and was absent in cells from controls (p=0.143, Fisher's exact test; Figure 2 and Supplementary Table S3). MAGEA2 and MAGEA4 transcripts could not be found in any of the EpCAM+ single cells.

Given the rare detection of histogenetic or cancer-associated epithelial transcripts we sought to determine whether hematopoietic transcripts are useful to differentiate between EpCAM+ cells from cancer versus control patients. For this, we analyzed the expression of selected hematopoietic transcripts. To our surprise, among cells from M0-stage cancer patients, the PTPRC-transcript (CD45) was most frequently detected (19% of cells; Figure 2 and Supplementary Table S3). Overall, 23% of EpCAM+ single cells from M0-stage PCa patients expressed at least one of the analyzed hematopoietic transcripts (PTPRC, CD34, CD33 or CD19). Similarly, 29% of EpCAM+ single cells from control patients expressed at least one of mentioned hematopoietic transcripts, of which PTPRC was also most frequent (21% of cells; Figure 2 and Supplementary Table S3). In single cells, the frequencies of hematopoietic transcript expression did not differ significantly (p=0.5, Pearson Chi-square). We then analyzed the expression of selected erythroid lineage-specific transcripts, GYPC (CD236, glycophorin C), SCL4A1 (CD233, band 3 protein) and HBA2 (haemoglobin α2). Frequencies of erythroid transcript-expressing cells between M0-stage patients and controls were almost identical (p=0.5, Pearson Chi-square). Here, the most frequently detected transcript was HBA2 and it was detected in 93% of cells from M0-stage patients and in 86% of cells from controls (Figure 2 and Supplementary Table S3). Overall, 96% of cells from M0-stage patients expressed at least one erythroid transcript, while the frequency of such cells among control cells was 93%.
This unexpected observation prompted us to control whether extracellular erythroid transcripts may contaminate the single cell transcriptomes by mRNA carry-over or release during sample processing. Therefore, we checked the WTA controls (see Methods section for details) of 40 patients for the presence of HBA2 transcript. Since all 40 samples were negative (Supplementary Figure S9) we note a striking difference to the cell-containing samples of M0-stage prostate cancer patients (p<0.0001, Pearson Chi-square).

Overall, analysis of epithelial, hematopoietic and erythroid transcripts in EpCAM+ single cells from cancer patients and controls revealed unexpected transcriptomes (Figure 3 and Supplementary Figure S8A-B).

Expression profiling of cells from M1-stage patients
For the majority of analyzed genes, EpCAM+ single cells or cell clusters from M1-stage cancer patients displayed similar frequencies of transcript-positive cells similar to from M0-stage patients (Figure 2, Figure 3, Supplementary Table S3 and Supplementary Figure S7B). A detailed overview of co-expression of individual transcripts is given in Supplementary Figure S8C. Since we obtained only two BM samples and selected 7 EpCAM+ single cells and 3 clusters for further analysis, a statistical comparison with the other groups was not performed.

Patterns of gene expression profiles and phenotypic heterogeneity
None of the analyzed transcripts could reliably discriminate between EpCAM+ cells from patients and controls. We therefore asked whether a combination of analyzed transcripts could help us to identify true DCCs. When we performed hierarchical cluster analysis, all EpCAM+ cells (from M0- and M1-stage patients and controls) were evenly scattered among dendrogram branches (Figure 4A). The analysis of cell pools revealed two major
clusters; however, pools from PCa patients and controls were evenly distributed between the clusters (Figure 4B).

From 35 (54%) EpCAM+ M0-stage BM samples, we isolated two or more cells. Among these 35 patients, cells from 14 patients (40%) clustered together. However, 12/14 patients had at least one cell that did not cluster with others, while only in the case of 2 patients, all cells clustered together. Therefore, the great majority of samples with two or more isolated cells contained cells with heterogeneous expression of analyzed transcripts. In controls, we found 5/10 samples with more than one cell. Here 2/5 (40%) clustered together, with at least one cell that did not cluster with others ($p=1$ for comparison with M0-stage patients; Fisher’s exact test). Cells from one M1-stage patients clustered together with several cells scattered within the dendrogram (Figure 4A).

**Analysis of genomic aberrations**

The presence of hematopoietic transcripts in EpCAM+ single cells from PCa patients casted doubts if these are indeed cancer cells, although many of them co-expressed epithelial transcripts. This prompted us to analyze selected cells for the presence of genomic aberrations using CGH. While the absence of genomic aberrations at the given resolution does not prove that the analyzed cell is not a cancer cell (16), the presence of genomic aberrations confirms the malignant origin of the cell.

To demonstrate the reliability of our combined genome and transcriptome analysis, we first compared the profile of a single VCaP cell with the profile of VCaP unamplified genomic DNA and found that the majority of genomic aberrations could be retrieved at single cell genome level (Supplementary Figure S10A). In contrast, two peripheral blood leukocytes (PBLs) from a healthy volunteer displayed normal karyotypes (Supplementary Figure S10B).
Next, we analyzed the genome of 4 EpCAM+ cells from controls, and 34 EpCAM+ cells isolated from the BM of M0- and M1-stage patients using CGH. No aberrations could be detected in the genomes of 4 EpCAM+ cells from controls, (Supplementary Figure S10C) whereas 13/34 cells (38%) from patients had genomic aberrations. Strikingly, we found cells expressing hematopoietic (e.g. CD45) or erythroid (e.g. \textit{HBA2}) transcripts, and displayed genomic aberrations, both in M0- and M1-stage patients (Figure 5 and Supplementary Figure S10E). These data demonstrate that DCCs can express hematopoietic and erythroid transcripts.

Finally, we compared the 17-gene signature of EpCAM+ cells from PCa patients (M0- and M1-stage) vs. EpCAM+ cells selected for genomic aberrations. Overall, the expression patterns did not change, although we noted an increased frequency of cells expressing epithelial transcripts (\textit{KRT8, KRT18; KRT19; KRT6a, EPCAM}) among cells harbouring genomic aberrations. Furthermore, erythroid marker expression was clearly detectable in these cells similar to the whole cohort (for example, \textit{HBA2} was detected in 77% vs. 92% and CD45 in 15% vs. 18%, respectively). A potentially relevant change was observed for PSA (\textit{KLK3}). Here, cells harbouring genomic aberrations more often expressed the transcript than the whole group of EpCAM+ cells from PCa patients (p=0.02, Fisher's exact test; Figure 6).
Discussion

In this study, we performed targeted expression profiling of EpCAM⁺ single cells isolated from the BM of non-metastatic (M0-stage) PCa patients and demonstrate, contrary to our expectation, that a 17-gene mRNA signature based on selected histogenetic markers is insufficient to reliably identify DCCs. However, by combining genome and transcriptome analysis, we could show that PCa-DCCs mimic the transcriptomes of BM cells and harbour hematopoietic and particularly erythroid transcripts. These findings suggest a high phenotypic plasticity of cancer cells clearly surpassing the classical concept of epithelial-mesenchymal transition (17).

We started by analysing cell pools of CD45-, CD33-, CD11b- and glycophorin A-depleted BM cells from PCa patients and controls, reasoning that one or several markers should be able to classify cancer derived samples correctly. Since this was not the case, we focussed on individually isolated single EpCAM-positive cells expecting that this population contained bona fide DCCs with a defined transcriptional profile.

EpCAM protein expression appeared to be stronger in cells from PCa patients than controls. However, the percentage of positive samples was comparable between groups and reliable identification of DCCs based on EpCAM staining alone was not possible (10, 11, 18). Of note, EpCAM expression by putative DCCs in BM was rather low compared to the typical appearance of most PCa cell lines (19, 20). Consequently, we investigated whether our marker panel, comprising epithelial (EPCAM, KRT8, KRT18, KRT19, KRT14, KRT6a, KRT5), the prostate-specific KLK3 (PSA), tumour-associated (MAGEA2, MAGEA4) and hematopoietic transcripts (PTPRC (CD45), CD33, CD34, CD19, GYPC, SCL4A1 (band 3), HBA2), would discriminate EpCAM⁺ cells from PCa patients and controls.

For all analyzed transcripts we established single cell sensitivity. However, none of the transcripts enabled a clear DCC identification, either because very few DCCs were
positive or because control cells expressed a given transcript at comparable frequency. Similarly, we identified no expression pattern characteristic for DCCs or EpCAM-positive cells from controls.

These findings, while unexpected, corroborate previous circumstantial observations. First, EpCAM transcript detection correlated with the fluorescence-staining intensity. More than 88% of the subjectively brighter stained EpCAM+ single cells from PCa patients harbour the EpCAM transcript, while we detected EpCAM transcripts in only 50% of weakly stained EpCAM+ control cells, suggesting a high technical reliability of our study. EpCAM, a highly glycosylated protein, is known to have a long half-life (21) which may explain low transcription in control cells. Second, significantly more EpCAM+ cells from PCa patients co-expressed epithelial transcripts than EpCAM+ cells from controls (93% vs. 57%). This co-expression included particularly KRT8 and KRT18, whose protein products are recognized by anti-cytokeratin antibodies, such as A45-B/B3, commonly used to detect DCCs (12), while other KRT transcripts were essentially absent from EpCAM+ cells of PCa patients. Thus EpCAM+ cells from PCa patients may comprise cells of malignant epithelial origin, i.e. DCCs, but these DCCs do not express cytokeratins typical for basal, stem-like cells, such as KRT5, KRT6a and KRT14 (22-25). Third, the transcript for the prostate-specific differentiation marker PSA (KLK3) or cancer-associated (MAGEA) (26) transcripts were hardly or not detected in EpCAM+ cells of PCa patients. The absence of MAGEA transcripts is in line with low detection rates in analyses of bulk BM (27), and the rare finding of KLK3 expression with similar studies on CK+ cells (28, 29). However, others have found higher frequency of EpCAM+ DCCs expressing PSA (30, 31), suggesting that EpCAM+ DCCs may either be undifferentiated or that PSA expression is linked to the formation of cell clusters and micrometastases.

Even analysis of hematopoietic markers did not help to differentiate DCCs and the native EpCAM+ BM cell population. For single cells, we could not find significant
differences in the frequency of *PTPRC* (CD45), *CD19*, *CD33*, *CD34* transcript-containing cells between M0-stage and control patients. Since published data suggested that cells of the early erythroid lineage in the BM could express EpCAM (32, 33), we tested the presence of erythroid-specific transcripts, namely *HBA2* (34), *SCL4A1* (35), and *GYPC* (36). Expression was abundant in both groups of cells, with EpCAM*+* cells from controls and PCa patients expressing at least one of the three transcripts in 93% and 96%, respectively. In both groups HBA2 was expressed most frequently, which was also detected in 80% of EpCAM*+* cells from M1-stage patients.

Despite the fact that EpCAM*+* single cells from M0- and M1-stage patients are clearly different from EpCAM*+* cells from controls, the findings substantially reduce our ability to correctly identify an individual cell as DCC from this 17-gene transcriptome. We therefore first excluded the possibility that the *HBA2* detection in our samples resulted from carry-over of extracellular mRNA. We then sought for independent validation that EpCAM*+* cells of PCa patients are truly cancer cells and performed combined genome and transcriptome analysis of selected single cells. Here, we noted that cells of proven malignant origin more frequently express cytokeratin markers as compared to the total group of EpCAM*+* cells, suggesting that EpCAM*+* normal cells may have diluted true DCCs. In fact, expression of *KLK3* (PSA) was significantly enriched among aberrant EpCAM*+* cells and apparently is the only marker that correctly identifies PCa-DCCs. However, this finding must be validated since *KLK3* expressing cells were isolated from only 2 patients. Still, many genomically–proven PCa-DCCs did not express *KLK3* but a similar pattern as the remaining EpCAM*+* cells from prostate cancer patients. Thus, genomically aberrant EpCAM*+* cells contain transcripts of common hematopoietic markers, such as CD45, as well as erythroid markers, demonstrating an unexpected transcriptomic plasticity of both M0- and M1-stage PCa DCCs.
Our findings have important implications for the study of metastasis. First, an educated guess for CTC or DCC-specific markers seems highly unreliable. So far, many studies assume histogenetic specificity (37) that we could not confirm for transcriptomes of bone marrow derived DCCs. Likewise attempts to reconstruct cancer cell transcriptomes from unpurified CTCs by transcript subtraction need to be viewed with caution (38). Our findings are consistent with previous observations, such as a transcriptome study of 15 single prostate CTCs from one M1-stage patient revealing low-level expression of EPCAM, PTPRC (CD45) and CD34 (39), expression of haemoglobins in breast cancer epithelial cells (40, 41) and ovarian cancer epithelial cells trans-differentiating into erythroid cells expressing haemoglobins (42). The observed expression of hematopoietic and erythroid transcripts by BM-DCCs may reflect adaptation to the specific environment. Of note, it was shown that metastatic prostate cells compete with hematopoietic stem cells (HSCs) for the HSC niche (43). If DCCs indeed reside in the HSC niche, it is possible that crosstalk between DCCs and BM cells, mediated by cell-cell interaction or secreted factors, leads to expression of hematopoietic transcripts in DCCs. Furthermore, the transcriptomes of DCCs could be altered by the uptake of exosomes released from resident bone marrow cells (44) or our observations could be the consequence of hypoxic/oxidative stress imposed on DCCs within the BM microenvironment (45, 46). While additional studies are needed to identify the underlying causes, the observed altered transcriptomes are likely to result in altered phenotypes and cell function, which may be highly relevant for our attempts to tailor cancer specific therapies.

In summary, until a comprehensive analysis of genes expressed by EpCAM+ cells of PCa patients vs. EpCAM+ cells of controls is performed, we will struggle to identify new markers specific for DCCs or confounding cells of non-epithelial lineage. Detection of genomic aberrations currently seems to be the most reliable way to confirm cancer cell identity.
Authors' Contributions

Conception and design: M. Gužvić, B. Braun, B. Polzer, C.A. Klein

Development of methodology: M. Gužvić, B. Braun

Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M. Gužvić, B. Braun, R. Ganzer, M. Burger, M. Nerlich, S. Winkler, M. Werner-Klein, Z. T. Czyż

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M. Gužvić, B. Braun, M. Werner-Klein, Z. T. Czyż, B. Polzer

Writing, review, and/or revision of the manuscript: M. Gužvić, B. Braun, M. Werner-Klein, Z. T. Czyż, B. Polzer, C.A. Klein

Study supervision: C.A. Klein

Acknowledgements

We are grateful for excellent technical help to Isabell Blochberger, Sandra Grunewald, Manfred Meyer, Irene Nebeja, Katharina Schneider, and Thomas Schamberger, and for help with collecting the data to Elisabeth Doblinger, Gundula Haunschild, Nina Patwary, and Giancarlo Feliciello. We are grateful to Tumorzentrum Regensburg e.V. for help in collecting patient data. We are grateful to Verena Huppert, Ana Grujović, and Norbert Braun for critical reading of the manuscript.

Grant support

This work was supported by the awards from Prostate Cancer Foundation and Dr. Josef Steiner Cancer Research Foundation Award, and grants from EU (FP6 STREP Research Consortium Migrating Cancer Stem Cells - MSCSs), Deutsche Krebshilfe (Research Consortium Tumour Stem Cells), and Bavarian Genome Research Network (BayGene) to C.A.K.
References


Exosomes from bone marrow mesenchymal stem cells contain a microRNA that promotes

45. Emara M, Turner AR, Allalunis-Turner J. Hypoxia differentially upregulates the
expression of embryonic, fetal and adult hemoglobin in human glioblastoma cells. Int J

expression by oxidative stress in hepatocytes and its implication in nonalcoholic

Table 1. Patient baseline characteristics and detection of EpCAM\(^+\) cells.

<table>
<thead>
<tr>
<th>Clinical characteristics</th>
<th>Overall</th>
<th>EpCAM(^+)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Number of M0-stage PCa patients</td>
<td>105(100)</td>
<td>65 (62)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 60 years</td>
<td>41 (39)</td>
<td>23 (56)</td>
</tr>
<tr>
<td>&gt; 60 years</td>
<td>64 (61)</td>
<td>42 (66)</td>
</tr>
<tr>
<td>pre-OP PSA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 10 ng/μl</td>
<td>70 (67)</td>
<td>44 (63)</td>
</tr>
<tr>
<td>&gt; 10 ng/μl</td>
<td>35 (33)</td>
<td>21 (60)</td>
</tr>
<tr>
<td>Gleason Score</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-6</td>
<td>40 (38)</td>
<td>24 (60)</td>
</tr>
<tr>
<td>7</td>
<td>51 (49)</td>
<td>32 (63)</td>
</tr>
<tr>
<td>8-10</td>
<td>14 (13)</td>
<td>9 (64)</td>
</tr>
<tr>
<td>Pathologic tumor stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>T2</td>
<td>53 (51)</td>
<td>31 (59)</td>
</tr>
<tr>
<td>T3+4</td>
<td>52 (49)</td>
<td>34 (65)</td>
</tr>
<tr>
<td>Nodal status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N0</td>
<td>58 (55)</td>
<td>33 (57)</td>
</tr>
<tr>
<td>N+</td>
<td>14 (13)</td>
<td>10 (71)</td>
</tr>
<tr>
<td>n.a.</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>Risk classification(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>low &amp; intermediate risk</td>
<td>51 (49)</td>
<td>29 (57)</td>
</tr>
<tr>
<td>high &amp; very high risk</td>
<td>54 (51)</td>
<td>36 (67)</td>
</tr>
</tbody>
</table>

\(^a\)Chi-square test, n.a. not available, \(^b\)According to European Association of Urology (47)
Figure legends

Figure 1. Intensity of EpCAM staining of cells isolated from PCa patients and controls. Photographs of the EpCAM\(^+\) cells, isolated from two healthy donors (A-D and E-H) and three prostate cancer patients in clinical stage M0 (I–L and M–T) and M1 (U–X). Photographs A–D, E–H, I–L and M–P show individual EpCAM\(^+\) cells, and photographs Q–T and U–X show clusters of EpCAM\(^+\) cells. Images A, E, I, M, Q, and U depict bright-field photographs. Images B, C, F, G, J, K, N, O, R, S, V and W depict photographs of cells taken under fluorescent microscopy (Cy3 channel). Numbers in the lower right corner of the images B, F, J, N, R, and V ["EpCAM (original)"] represent the exposure time used for taking of the photograph. Note that photographs U-X were taken on different microscope (Zeiss), compared to photographs A–T (Olympus), hence different exposure times. Images C, G, K, O, S, and W ["EpCAM (bright.+15%/contr.+25%)"] were manipulated in a way that brightness and contrast of the original image were increased, as indicated, using image processing software, for demonstrating the different staining intensity of EpCAM\(^+\) cells of controls and cancer patients. Images D, H, L, P, T, and X are the result of merging of corresponding original bright-field and fluorescent photographs. Scale bar: 50 \(\mu\)m.

Figure 2. Gene expression of EpCAM\(^+\) cells isolated from PCa and control patients. The identities of transcripts are on the X-axis. Results for KRT5, MAGEA2 and MAGEA4 are not shown, since the transcripts were not detected in any sample. Exact numbers and percentages are given in Supplementary Table S3. M0- (grey bars) and M1-stage (black bars) prostate cancer patients and control patients (white bars). Asterisks indicate the level of significance: * – p<0.05, ** – p<0.01, *** – p<0.001; n indicates the number of analyzed single cells.
**Figure 3.** Transcript-profiling by PCR of WTA cDNA of selected EpCAM⁺ single cells or small clusters of EpCAM⁺ cells from PCa patients and controls. The photograph of each agarose gel was cropped to contain the band of interest and at least two nearest DNA size marker bands. The sizes of fragments are given in Supplementary Table S1 and Supplementary Figure S11. Identifiers of the samples are given above the wells, and the identity of analyzed transcript is given next to each cropped gel. Only those samples that contain the band of correct size were considered positive. The brightness and contrast of each individual cropped image were manipulated by using "Auto contrast" tool of Adobe Photoshop.

**Figure 4.** Hierarchical cluster analysis of gene expression.

(A) single cells and (B) pools from M0- and M1-stage patients, and healthy controls. Green colour in heatmaps denotes that the marker is detected by PCR in the sample, while red colour indicates that the sample was negative. The sample type (M0-, M1-stage or N [healthy]) is colour coded. Groups of cells that clustered together and were isolated from the same patient are also colour coded. Related Supplementary Figure S8 displays all results of expression profiling grouped by sample type (single cell or cell pool) and source (cancer patients or controls).

**Figure 5.** Combined genome and transcriptome analysis of single EpCAM⁺ cells.

(A,B), single cell CGH profiles of two M0-stage and (C) one M1-stage and (D) one control patient. Some of the detected aberrations are indicated by arrowheads. Centromeric regions known to generate false positive signals are excluded from analysis. Results of the transcript profiling of each cell is given below the CGH profile (green and 1 indicate that the transcript was detected in the sample, while red and 0 indicate that the transcript was not detected). The remaining CGH profiles and corresponding expression profiles are
given in Supplementary Figure 10. cCGH – CGH on metaphase spreads, aCGH – CGH on DNA microarrays.

**Figure 6.** Gene expression of DCCs from PCa patients.

Cells with proven malignant origin display similar profiles as the complete group of EpCAM$^+$ cells from PCa patients. Transcripts are identified by gene symbols. Results for *KRT5*, *MAGEA2* and *MAGEA4* are not shown, since the transcripts were not detected in any sample. EpCAM$^+$ cells from M0 and M1-stage- (grey bars) PCa patients, EpCAM$^+$ cells with genomic aberrations from M0 and M1-stage- (black bars) PCa patients, and EpCAM$^+$ cells from control patients (white bars). $n$ indicates the number of analyzed single cells.
Figure 1

<table>
<thead>
<tr>
<th>Bright field</th>
<th>EpCAM (original)</th>
<th>EpCAM (bright+15%/contr. +25%)</th>
<th>Merge</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>C</td>
<td>D</td>
</tr>
<tr>
<td>E</td>
<td>F</td>
<td>G</td>
<td>H</td>
</tr>
<tr>
<td>I</td>
<td>J</td>
<td>K</td>
<td>L</td>
</tr>
<tr>
<td>M</td>
<td>N</td>
<td>O</td>
<td>P</td>
</tr>
<tr>
<td>Q</td>
<td>R</td>
<td>S</td>
<td>T</td>
</tr>
<tr>
<td>U</td>
<td>V</td>
<td>W</td>
<td>X</td>
</tr>
</tbody>
</table>

Healthy donor

Prostate cancer

M0

M1
Figure 2

Percentage of given transcripts (%) for different cell types and conditions:

- **EPCAM**
- **KRT8**
- **KRT18**
- **KRT19**
- **KRT14**
- **KRT6a**
- **KLK3**
- **PTPRC**
- **CD34**
- **CD33**
- **CD19**
- **GYP**
- **SLC4A1**
- **HBA2**

**Cell Types:**
- Epithelial
- Prostate
- Haematopoietic
- Erythroid

**Conditions:**
- M0 PCa (n=124)
- M1 PCa (n=10)
- Control BM (n=28)

Significance levels:
- ***p < 0.001
- **p < 0.01
- *p < 0.05
Figure 3

DNA size marker

M0

252-852 T2
261-838 T1
281-903 T3
310-958 T4
351-1050 T1
363-1086 T2
341-1013 T2
342-804 T2
452-2244 T6

N

positive control

DNA size marker

M1

Ep2
Ep3 F
Ep2
Ep3
Ep2
Ep3
Ep3
Ep3

EPCAM
KRT8
KRT18
KRT19
KRT14
KRT6A
KRT5
KLK3
PTPRC
CD34
CD33
CD19
SLC4A1
GYPCE
HBA2
EEF1A1

Downloaded from cancerres.aacrjournals.org on April 14, 2017. © 2014 American Association for Cancer Research.
Combined genome and transcriptome analysis of single disseminated cancer cells from bone marrow of prostate cancer patients reveals unexpected transcriptomes

Miodrag Guzvic, Bernhard Braun, Roman Ganzer, et al.

Cancer Res  Published OnlineFirst October 15, 2014.

Updated version  Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-14-0934

Supplementary Material  Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2014/10/16/0008-5472.CAN-14-0934.DC1

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