Combined Genome and Transcriptome Analysis of Single Disseminated Cancer Cells from Bone Marrow of Prostate Cancer Patients Reveals Unexpected Transcriptomes

Miodrag Gužvić, Bernhard Braun, Roman Ganzèr, Maximilian Burger, Michael Nerlich, Sebastian Winkler, Melanie Werner-Klein, Zbigniew T. Czyż, Bernhard Polzer, and Christoph A. Klein

Abstract

Bone is the most frequent site of metastasis in prostate cancer and patients with bone metastases are deemed incurable. Targeting prostate cancer cells that disseminated to the bone marrow before surgery and before metastatic outgrowth may therefore prevent lethal metastasis. This prompted us to directly analyze the transcriptome of disseminated cancer cells (DCC) isolated from patients with nonmetastatic (UICC stage M0) prostate cancer. We screened 105 bone marrow samples of patients with M0-stage prostate cancer and 18 bone marrow 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, KRT15a, KRT7, KLK3 (PSA), MAGEA2, MAGEA4, PTPRC (CD45), CD33, CD34, CD19, GYPC, SCLA1 (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, whereas true DCCs may express hematopoietic transcripts. These results point to an unexpected transcriptome plasticity of epithelial cancer cells in bone marrow and question common transcriptional criteria to identify DCCs. Cancer Res; 74(24); 7383–94. ©2014 AACR.
marrow (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 to 2 DCCs per 10⁶ bone marrow cells of positive patients.

Previously, we established an approach to comprehensively characterize the transcriptome of a single cell using microarrays (7, 10). As 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 bone marrow has also been associated with prostate cancer progression (8). However, unlike CK⁺ cells (9), EpCAM⁺ cells are often found in the bone marrow 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 bone marrow of M0-stage prostate cancer 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 bone marrow of prostate cancer from EpCAM⁺ bone marrow cells of cancer-free patients. In particular, we assessed the expression of selected epithelial (EpCAM, KRT18, KRT16, KRT19, KRT5, KRT6a, KRT14), prostate-specific (KLK3 (PSA)), tumor-specific (MAGEA2 and MAGEA4), common hematopoietic (PTPRC (CD45), CD34, CD33, CD19), and erythroid (GPYC, SCLA11, HB2A) transcripts aiming to separate prostatic epithelial and hematopoietic lineages.

Materials and Methods

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

Patients and bone marrow samples

The local ethics committee of the University of Regensburg (Regensburg, Germany) approved all aspects of the study (ethics vote number 07-079). Between February 2009 and November 2012, bone marrow aspirates from 105 nonmetastatic (M0 stage) prostate cancer patients were collected shortly before radical prostatectomy. In addition, we obtained bone marrow samples of two metastatic (M1 stage) prostate cancer patients. As a control group, 18 bone marrow samples from cancer-free males undergoing trauma or orthopedic surgery were obtained. After obtaining written informed consent, the sampling was performed by aspiration of the 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 bone marrow samples

Detailed description of bone marrow processing is given in Supplementary Materials and Methods. Briefly, upon arrival, the bone marrow sample was twice washed with Hank’s balanced salt solution to remove fat and thrombocytes. Next, cell suspension was centrifuged in density gradient. After centrifugation, the interphase containing mononuclear cells (MNC) were carefully collected and washed with 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 unlabeled cell fraction, was collected on ice and the cell number determined using a hemocytometer.

Staining of bone marrow samples and cell lines

Detailed description of staining procedure is given in Supplementary Materials and Methods. On average, two million bone marrow 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, the 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 Fig. S2).

Screening of bone marrow and isolation of single cells and cell pools

A detailed description of the screening and cell isolation procedures is given in Supplementary Materials and Methods. Each bone marrow sample was manually screened for the presence of EpCAM⁺ cells on an inverted fluorescence 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 approximately 2,000 to 3,000 cells. As a reagent control of cell isolation, 1 μL of PBS, in which individual EpCAM⁺ cells were isolated, was taken for subsequent whole transcriptome amplification (WTA).

Photographs of the cells before isolation were made in 10 cases, using the software cellSens (Olympus) or AxioVision (Zeiss). Because of the rapid bleaching of PE fluorophore, we abstained from taking photographs of all isolated cells.

Flow-cytometric analysis of bone marrow cells was performed with a LSR II machine equipped with FACS DIVA 5.03 software (BD Biosciences) and data was analyzed with FlowJo 8.8.6 (Treestar).

Whole transcriptome amplification of single cells

Isolation of mRNA from a single cell, reverse transcription, and global amplification of the first strand cDNA was carried out using a WTA kit (Agencourt). The quality of the resulting cDNA was verified by gel electrophoresis and spectrophotometry (NanoDrop). The cDNA was then subjected to a seminested PCR amplification using a SmartAmp kit (Agencourt). Replicate cDNA was subjected to three rounds of PCR amplification. Each round of amplification was followed by a second strand synthesis reaction.
out as previously described (7, 10). Details are provided in Supplementary Materials and Methods. The quality of amplification was examined by performing endpoint PCR for \(ACTB\), \(EEF1A1\), and \(GAPDH\) transcripts (see below and Supplementary Table S1).

Whole transcriptome amplification controls
From all samples, WTA controls were generated by aspirating 1 \(\mu\)L 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 isolated during the same experiment were included into transcriptome analysis. In 2 of 125 (2%) bone marrow samples, the WTA control was contaminated. The WTA samples of these two patients were excluded from genome and transcriptome analysis.

Whole genome amplification and comparative genomic hybridization
Single-cell genomic DNA collected during WTA procedure was precipitated and subjected to whole genome amplification (WGA), using the Ampli1 WGA Kit (Silicon Biosystems), as described previously (10, 13). Details are provided in Supplementary Materials and 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 endpoint PCR in patient samples
All WTA products were tested by endpoint PCR using conditions outlined in Supplementary Materials and Methods. The list of primers is given in Supplementary Table S1. To establish the correct identity of each amplified fragment, we performed restriction digestion, using the conditions outlined in Supplementary Materials and 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 bone marrow samples and relation to clinical variables
We investigated the presence of putative DCCs in bone marrow samples of 105 M0-stage prostate cancer patients undergoing radical prostatectomy by staining for EpCAM.

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

<table>
<thead>
<tr>
<th>Clinical characteristics</th>
<th>Overall (n) (%)</th>
<th>Yes (n) (%)</th>
<th>No (n) (%)</th>
<th>(P^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of M0-stage prostate cancer patients</td>
<td>105 (100)</td>
<td>65 (62)</td>
<td>40 (38)</td>
<td>0.61</td>
</tr>
<tr>
<td>Age, y (\leq 60)</td>
<td>41 (39)</td>
<td>23 (56)</td>
<td>18 (44)</td>
<td></td>
</tr>
<tr>
<td>(&gt; 60)</td>
<td>64 (61)</td>
<td>42 (66)</td>
<td>22 (34)</td>
<td>0.32</td>
</tr>
<tr>
<td>Pre-OP PSA (\leq 10 \text{ ng/(\mu)L})</td>
<td>70 (67)</td>
<td>44 (63)</td>
<td>26 (37)</td>
<td></td>
</tr>
<tr>
<td>(&gt; 10 \text{ ng/(\mu)L})</td>
<td>35 (33)</td>
<td>21 (60)</td>
<td>14 (40)</td>
<td>0.77</td>
</tr>
<tr>
<td>Gleason score (2–6)</td>
<td>40 (38)</td>
<td>24 (60)</td>
<td>16 (40)</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>51 (49)</td>
<td>32 (63)</td>
<td>19 (37)</td>
<td></td>
</tr>
<tr>
<td>(8–10)</td>
<td>14 (13)</td>
<td>9 (64)</td>
<td>5 (36)</td>
<td>0.95</td>
</tr>
<tr>
<td>Pathologic tumor stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(T2)</td>
<td>53 (51)</td>
<td>31 (59)</td>
<td>22 (41)</td>
<td></td>
</tr>
<tr>
<td>(T3 +)</td>
<td>52 (49)</td>
<td>34 (65)</td>
<td>18 (35)</td>
<td>0.47</td>
</tr>
<tr>
<td>Nodal status (N0)</td>
<td>58 (55)</td>
<td>33 (57)</td>
<td>25 (43)</td>
<td></td>
</tr>
<tr>
<td>(N+)</td>
<td>14 (13)</td>
<td>10 (71)</td>
<td>4 (29)</td>
<td>0.32</td>
</tr>
<tr>
<td>n.a.</td>
<td>33 (32)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risk classification(^b)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low and intermediate risk</td>
<td>51 (49)</td>
<td>29 (57)</td>
<td>22 (43)</td>
<td></td>
</tr>
<tr>
<td>High and very high risk</td>
<td>54 (51)</td>
<td>36 (67)</td>
<td>18 (33)</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Abbreviation: n.a., not available.
\(^a\)\(\chi^2\) test.
\(^b\)According to the European Association of Urology (47).
addition, we stained bone marrow samples from two M1-stage patients with prostate cancer and 18 male control patients without cancer (Supplementary Fig. S1). To enrich the epithelial target population, we first depleted bone marrow samples of CD45+, CD33+, CD11b+, and glycophorin A+ cells. On average, 86% of MNCs were removed by immunomagnetic sorting (Supplementary Fig. S3). Then, we used direct immunofluorescence to detect EpCAM+ cells. Screening on average 2 × 10^6 cells per bone marrow sample (median 2 × 10^6, range 0.02–6.8 × 10^6 cells; Supplementary Fig. S3) revealed that 65 (62%) of M0-stage patients, as well as 10 (56%) of control patients had EpCAM+ cells in the bone marrow (P = 0.79, Fisher exact test, two-tailed; Supplementary Table S2). Both analyzed M1-stage patients also harbored EpCAM+ cells in the bone marrow.

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 bone marrow of M0-stage prostate cancer patients and any clinical risk factor such as Gleason score or preoperative 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 bone marrow samples from patients and controls**

EpCAM+ cells from patients with prostate cancer (Fig. 1I–X) usually displayed stronger staining intensity compared with cells from controls (Fig. 1A–H). However, solely based on staining intensity, size, or morphology, one would not be able to reliably identify true DCCs. Because of 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 Fig. S4) suggests that EpCAM+ cells from bone marrow 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 Fig. S4E–S4P). To corroborate these observations, we used a bone marrow sample of a cancer-free patient, depleted it of CD45+, CD33+, CD11b+, and glycophorin A+ cells, and

---

**Figure 1.** Intensity of EpCAM staining of cells isolated from patients with prostate cancer and controls. Photographs of the EpCAM+ cells isolated from two healthy donors (A–D and E–H) and three patients with prostate cancer 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 bottom right corner of the images B, F, J, N, R, and W depict the exposure time used for taking of the photograph. Note that photographs U–X were taken on a different microscope (Zeiss), 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 to demonstrate the different staining intensity of EpCAM+ cells of controls and patients with cancer. Images D, H, L, P, T, and X show clusters of EpCAM+ cells from patients with prostate cancer (Fig. 1I–X) usually displayed stronger staining intensity compared with cells from controls (Fig. 1A–H). However, solely based on staining intensity, size, or morphology, one would not be able to reliably identify true DCCs. Because of 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 Fig. S4) suggests that EpCAM+ cells from bone marrow 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 Fig. S4E–S4P). To corroborate these observations, we used a bone marrow sample of a cancer-free patient, depleted it of CD45+, CD33+, CD11b+, and glycophorin A+ cells, and...
divided the sample in halves. We spiked one half with 2% of EpCAM+ DU145 prostate cancer cell line cells. Then, spiked and nonsposted 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 Fig. S5A), whereas weakly EpCAM+ cells could be seen in both samples stained with EpCAM-PE antibody (Supplementary Fig. S5I–S5J). As expected, these cells displayed slightly weaker EpCAM positivity compared with some of DU145 cells (Supplementary Fig. S5M–S5P). 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 Fig. S5Q). This shows that the intensity of the EpCAM staining can rarely be used to discriminate EpCAM+ true DCCs from EpCAM+ bone marrow cells.

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

On average, we isolated four individual EpCAM+ cells from each positive bone marrow sample (Supplementary Fig. S3). From 65 EpCAM+ bone marrow samples of M0-stage prostate cancer patients, we isolated 220 EpCAM+ single cells (examples shown in Fig. II–P) and five clusters of two to five cells (examples shown in Fig. IQ–T), and twelve EpCAM+ single cells and four clusters of two from two M1-stage prostate cancer patients (Fig. 1U–X). Fifty-one individual EpCAM+ cells were collected from ten control patients (Fig. IA–H; Supplementary Table S2). Bone marrow of controls never contained EpCAM+ cell clusters. After isolation of EpCAM+ cells, we pooled a pool of approximately 2,000 to 3,000 cells from each bone marrow 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 transcriptome amplification of a DCC signature in M0-stage patients.

We started the search for a DCC-specific signature by expression analysis of selected genes in bone marrow cell pools of prostate cancer patients and controls. We could not detect significantly different frequencies for epithelial-, prostate-, tumor-associated, and erythroid transcript expression between the analyzed cell pools (Supplementary Table S3; Supplementary Figs. S6 and S8D–S8F). Surprisingly, among common hematopoietic transcripts, only CD19 was expressed in a significantly higher number of samples from M0-stage patients as compared with pools from healthy controls (P < 0.05, Fisher exact test).

As the analysis of bone marrow cell pools revealed no transcript specific for cancer patients or controls, we focused on expression profiling of selected transcripts in EpCAM+ single cells. We first analyzed the expression of EPCAM and cytokeratin 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 χ²; Fig. 2 and Supplementary Fig. S7B; Supplementary Table S3). Forty-two percent of cells from M0-stage patients expressed at least one KRT transcript (Supplementary Fig. S7B), 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 χ²; Supplementary Table S2 and Supplementary Fig. S7A). After WTA and QC, we selected 119 EpCAM+ single cells and five 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 seven EpCAM+ single cells and three clusters of EpCAM+ cells (62% of samples; Supplementary Table S2 and Supplementary Fig. S7A). In addition, high-quality WTA products of bone marrow cell pools from ten randomly selected M0-stage patients, one M1-stage patient, and ten controls were chosen for further analysis.

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

We next analyzed the expression of selected genes in bone marrow cell pools of prostate cancer patients and controls. We could not detect significantly different frequencies for epithelial-, prostate-, tumor-associated, and erythroid transcript expression between the analyzed cell pools (Supplementary Table S3; Supplementary Figs. S6 and S8D–S8F). Surprisingly, among common hematopoietic transcripts, only CD19 was expressed in a significantly higher number of samples from M0-stage patients as compared with pools from healthy controls (P < 0.05, Fisher exact test).

As the analysis of bone marrow cell pools revealed no transcript specific for cancer patients or controls, we focused on expression profiling of selected transcripts in EpCAM+ single cells. We first analyzed the expression of EPCAM and cytokeratin 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 χ²; Fig. 2 and Supplementary Fig. S7B; Supplementary Table S3).

Forty-two percent of cells from M0-stage patients expressed at least one KRT transcript (Supplementary Fig. S7B),
which was significantly different to cells from controls, where only 7% of cells expressed KRT transcripts ($P < 0.01; \chi^2$). The difference between patients with cancer 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 coexpressed EPCAM and KRT transcripts, compared with cells from controls ($P < 0.002$, Fisher exact test; Supplementary Table S4). Among KRT-expressing cells from M0-stage patients, the most frequently detected KRT transcripts were KRT8 and KRT18 (Fig. 2 and Supplementary Fig. S7C; Supplementary Table S3). 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 $\chi^2$; 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 Fig. S7C). In total, among cells from M0-stage patients, 93% of cells expressed at least one of the tested epithelial transcripts, compared with 57% cells from controls ($P < 0.001$, Fisher exact test). This distribution closely resembled the frequency of EPCAM expression observed in these two groups of cells (Supplementary Fig. S7B).

Next, we analyzed prostate- and tumor-specific transcripts. KLK3 (PSA) transcript was found only in cells isolated from one M0-stage patient, whose bone marrow 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 exact test; Fig. 2; 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; Fig. 2; Supplementary Table S3). Overall, 23% of EpCAM+ single cells from M0-stage prostate cancer 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; Fig. 2; Supplementary Table S3). In single cells, the frequencies of hematopoietic transcript expression did not differ significantly ($P = 0.5$; Pearson $\chi^2$). We then analyzed the expression of selected erythroid lineage–specific transcripts, GYPC (CD236, glycophorin C), SCI1A1 (CD233, band 3 protein), and HBA2 (hemoglobin $\alpha_2$). Frequencies of erythroid transcript–expressing cells between M0-stage patients and controls were almost identical ($P = 0.5$, Pearson $\chi^2$). 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 (Fig. 2; 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 carryover or release during sample processing. Therefore, we checked the WTA controls (see Materials and Methods section for details) of 40 patients for the presence of HBA2 transcript. As all 40 samples were negative (Supplementary Fig. S9) we note a striking difference to the cell-containing samples of M0-stage prostate cancer patients ($P < 0.0001$, Pearson $\chi^2$).

Overall, analysis of epithelial, hematopoietic, and erythroid transcripts in EpCAM+ single cells from cancer patients and controls revealed unexpected transcriptomes (Fig. 3 and Supplementary Fig. S8A and S8B).

**Figure 3.** Transcript profiling by PCR of WTA cDNA of selected EpCAM+ single cells or small clusters of EpCAM+ cells from patients with prostate cancer 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 Fig. 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 the “Auto contrast” tool of Adobe Photoshop.
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 those from M0-stage patients (Fig. 2, Fig. 3, and Supplementary Fig. S7B; Supplementary Table S3). A detailed overview of coexpression of individual transcripts is given in Supplementary Fig. S8C. As we obtained only two bone marrow samples and selected seven EpCAM⁺ single cells and three 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 (Fig. 4A). The analysis of cell pools revealed two major clusters; however, pools from patients with prostate cancer and controls were evenly distributed between the clusters (Fig. 4B).

From 35 (54%) EpCAM⁺ M0-stage bone marrow samples, we isolated two or more cells. Among these 35 patients, cells from 14 patients (40%) clustered together. However, 12 of 14 patients had at least one cell that did not cluster with others, whereas only in the case of two 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 of 10 samples with more than one cell. Here, 2 of 5 (40%) clustered together, with at least one cell that did not cluster with others (P = 1 for comparison with M0-stage patients; Fisher exact test). Cells from one M1-stage patients clustered together with several cells scattered within the dendrogram (Fig. 4A).

Analysis of genomic aberrations

The presence of hematopoietic transcripts in EpCAM⁺ single cells from patients with prostate cancer casted doubts whether these are indeed cancer cells, although many of them coexpressed 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 Fig. S10A). In contrast, two peripheral blood leukocytes from a healthy volunteer displayed normal karyotypes (Supplementary Fig. S10B).

Next, we analyzed the genome of four EpCAM⁺ cells from controls, and 34 EpCAM⁺ cells isolated from the bone marrow of M0- and M1-stage patients using CGH. No aberrations could be detected in the genomes of four EpCAM⁺ cells from controls, (Supplementary Fig. S10C), whereas 13 of 34 cells (38%) from patients had genomic aberrations. Strikingly, we found cells expressing hematopoietic (e.g., CD45) or erythroid (e.g., HBA2) transcripts, and harboring genomic aberrations, both in M0- and M1-stage patients (Fig. 5 and Supplementary Fig. S10E). These data demonstrate that DCCs can express hematopoietic and erythroid transcripts.

Finally, we compared the 17-gene signature of EpCAM⁺ cells from patients with prostate cancer (M0 and M1 stage) versus EpCAM⁺ cells selected for genomic aberrations. Overall, the expression patterns did not change, although we noted an increased frequency of cells expressing epithelial transcripts (KRT8, KRT18, KRT19, KRT16a, EPCAM) among cells harboring genomic aberrations. Furthermore, erythroid marker expression was clearly detectable in these cells similar to the whole cohort (e.g., HBA2 was detected in 77% vs. 92% and CD45 in 15% vs. 18%, respectively). A potentially relevant change was observed for PSA (KLK3). Here, cells harboring genomic aberrations more often expressed the transcript than the whole group of EpCAM⁺ cells from patients with prostate cancer (P = 0.02; Fisher exact test; Fig. 6).

Discussion

In this study, we performed targeted expression profiling of EpCAM⁺ single cells isolated from the bone marrow of patients with nonmetastatic (M0 stage) prostate cancer 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 prostate cancer-DCCs mimic the transcriptomes of bone marrow cells and harbor 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 analyzing cell pools of CD45⁻, CD33⁻, CD11b⁻, and glycophorin A-depleted bone marrow cells from patients with prostate cancer and controls, reasoning that one or several markers should be able to classify cancer-derived samples correctly. As 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 patients with prostate cancer 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 bone marrow was rather low compared with the typical appearance of most prostate cancer 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), tumor-associated (MAGEA2, MAGEA4) and hematopoietic transcripts (PTPRC (CD45), CD33, CD34, CD19, GYPC,
SCL4A1 (band 3), HBA2], would discriminate EpCAM^{+} cells from patients with prostate cancer 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 patients with prostate cancer harbored the EpCAM expression pattern of the sample.
transcript, whereas 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 patients with prostate cancer coexpressed epithelial transcripts than EpCAM+ cells from controls (93% vs. 57%). This coexpression included particularly KRT8 and KRT18, whose protein products are recognized by anti-CK antibodies, such as A45-B/B3, commonly used to detect DCCs (12), whereas other KRT transcripts were essentially absent from EpCAM+ cells of patients with prostate cancer. Thus, EpCAM+ cells from patients with prostate cancer 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; ref. 26) transcripts were hardly or not detected in EpCAM+ cells of patients with prostate cancer. 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+ bone marrow 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. As published data suggested that cells of the early erythroid lineage in the bone marrow 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 patients with prostate cancer 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 carryover of extracellular mRNA. We then sought for independent validation that EpCAM+ cells of patients with prostate cancer 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 with 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 prostate cancer-DCCs. However, this finding must be validated as KLK3-expressing cells were isolated from only two patients. Still, many genomically proven prostate cancer-DCCs did not express KLK3 but a similar pattern as the remaining EpCAM+ cells from patients with prostate cancer. 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 prostate cancer 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 (38) 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 hemoglobins in breast cancer epithelial cells (40, 41) and ovarian cancer epithelial cells transdifferentiating into erythroid cells expressing hemoglobins (42). The observed expression of hematopoietic and erythroid transcripts by bone marrow-DCCs may reflect adaptation to the specific environment. Of note, it was shown that metastatic prostate cells compete with hematopoietic stem cells (HSC) for the HSC niche (43). If DCCs indeed reside in the HSC niche, it is possible that crosstalk between DCCs and bone marrow 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 bone marrow 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 patients with prostate cancer versus EpCAM+ cells of controls is performed, we will struggle to identify new markers specific for DCCs or confounding cells of nonepithelial lineage. Detection of genomic aberrations currently seems to be the most reliable way to confirm cancer cell identity.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: M. Guzvić, B. Braun, C.A. Klein
Development of methodology: M. Guzvić, B. Braun, R. Ganzer, M. Nerlich
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): M. Guzvić, B. Braun, R. Ganzer, M. Burger, M. Nerlich, S. Winkler, M. Werner-Klein, Z.T. Czyż, B. Polzer
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M. Guzvić, B. Braun, Z.T. Czyż, B. Polzer
Writing, review, and/or revision of the manuscript: M. Guzvić, B. Braun, S. Winkler, M. Werner-Klein, Z.T. Czyż, C.A. Klein
Study supervision: C.A. Klein

Acknowledgments
The authors thank Isabell Blochberger, Sandra Grunewald, Manfred Meyer, Irene Nebeja, Katharina Schneider, and Thomas Schambner for excellent technical help, and to Elisabeth Doblinger, Gundula Haunschild, Nina Patwary, and Giancarlo Feliciello for help with collecting the data. The authors also thank Tumorzentrum Regensburg e.V. for help in collecting patient data and Verena Huppert, Ana Grajović, and Norbert Braun for critical reading of the manuscript.
Transcriptional Profiling of DCCs from Bone Marrow

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), Deutsche Krebshilfe (Research Consortium Tumour Stem Cells), and Bavarian Genome Research Network (Bargene) to C.A. Klein.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received April 1, 2014; revised October 7, 2014; accepted October 7, 2014; published OnlineFirst October 15, 2014.

References


