Science Advances
AAAS
Manuscript Template

Epithelial-type Systemic Breast Carcinoma Cells with a Restricted Mesenchymal Transition are a Major Source of Metastasis

Short title: EMT in metastatic breast cancer systemic cells

Authors

Xiao Liu[1†], Junjian Li[2†], Bruno Loureiro Cadilha[3], Anamarija Markota[3], Cornelia Voigt[3], Zhe Huang[1], Peter P. Lin[4], Daisy D. Wang[4], Juncheng Dai[5], Gisela Kranz[1], Anna Krandick[1], Darko Libl[1], Horst Zitzelsberger[6,7], Isabella Zagorski[7], Herbert Braselmann[7], Min Pan[1], Sibo Zhu[8], Yuanchi Huang[1], Sebastian Niedermeyer[1], Christoph A. Reichel[1], Bernd Uhl[1], Daria Briukhovetska[3], Javier Suárez Gosálvez[3], Sebastian Kobold[3‡], Olivier Gires[1,6‡*], and Hongxia Wang[2‡*]

Affiliations

[1] Department of Otorhinolaryngology, Grosshadern Medical Center, Ludwig-Maximilians-University of Munich, Marchioninistr. 15, 81377 Munich, Germany

[2] Department of Oncology, Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai 200080, P.R. China

[3] Center of Integrated Protein Science Munich and Division of Clinical Pharmacology, Department of Medicine IV, Klinikum der Ludwig-Maximilians-Universität München, Member of the German Center for Lung Research, Lindwurmstrasse 2a, 80337 Munich, Germany

[4] Cytelligen, San Diego, California 92121, USA

[5]Department of Epidemiology and Biostatistics, School of Public Health, Nanjing Medical University, Nanjing 211166, China

[6] Clinical Cooperation Group Personalized Radiotherapy of Head and Neck Tumors, Helmholtz Zentrum München, Neuherberg, Germany

[7] Research Unit Radiation Cytogenetics, Helmholtz Zentrum München, Neuherberg, Germany

[8] State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200438, China

* Olivier Gires Olivier.gires@med.uni-muenchen.de

or Hongxia Wang whx365@126.com

†: These authors contributed equally to this work

‡: These authors are equally contributing senior authors

Keywords: Metastatic breast cancer; Epithelial-Mesenchymal-Transition; Circulating tumor cells; Disseminated tumor cells; EpCAM

Abstract

Carcinoma cells undergo epithelial-mesenchymal transition (EMT), however contributions of EMT heterogeneity to disease progression remain a matter of debate. Here, we addressed the EMT status of *ex vivo* cultured circulating and disseminated tumor cells (CTC/DTC) in a syngeneic mouse model of metastatic breast cancer (MBC). Epithelial-type CTC with a restricted mesenchymal transition (E/m-type) possessed the strongest lung metastases formation ability, whereas mesenchymal-type CTC showed limited metastatic ability. EpCAM expression served as a surrogate marker to evaluate the EMT heterogeneity of clinical samples from MBC, including metastases, CTC, and DTC. The proportion of epithelial-type CTC and especially DTC correlated with distant metastases and poorer outcome of MBC patients. This study fosters our understanding of EMT in metastasis and underpins heterogeneous EMT phenotypes as important parameters for tumor prognosis and treatment. We further suggest that EpCAM-dependent CTC isolation systems will underestimate CTC numbers, but will quantify clinically relevant metastatic cells.

Introduction

Breast cancer mortality has decreased by 40% from 1989 to 2015, owing to the impact of early detection through screening methods and to improved therapeutic modalities (*1*). Stage I-III tumors involving breast and locoregional lymph nodes are characterized by comparably good overall survival rates at five and ten years (100% and 72%, respectively). In contrast, stage IV metastatic breast cancer (MBC), which involves colonization of distant sites, remains a major life-threatening disease with survival rates below 25% at five years. On average 5 - 10% of patients are diagnosed with stage IV disease at initial diagnosis, but 20 - 30% of stage I-III patients will eventually progress and develop distant metastases in the course of their disease. Hence, understanding basic processes of distant metastasis formation and identifying cells of origin is of paramount importance to improve the treatment of patients and ultimately their outcome (*2, 3*).

Metastasis formation initiates with the delamination of single or clusters of cancer cells from primary tumors, followed by an intravasation into the blood stream. These circulating tumor cells (CTC) may eventually extravasate from blood vessels, and disseminate to distant sites such as lungs, liver, or bone marrow, where they are referred to as disseminated tumor cells (DTC). In this novel environment, DTC can remain as single cells or generate micrometastases (*4*), which can give rise to outcome-determining metastases (*5-7*).

In the clinical setting, CTC counts evaluated through the usage of the Food and Drug Administration-approved retrieval technology CellSearch, which were as low as one cell per 7.5 mL of peripheral blood, correlated with poor outcome in a large cohort of 3,173 non-metastatic stage I-III patients (*8*). Furthermore, CTC numbers correlated with disease progression and metastases formation (*9-12*). A formal experimental proof of the metastatic potential of MBC-derived CTC was provided in a xenotransplantation model (*13*), which also demonstrated poor efficiency of metastases generation by CTC. Intrafemural transfer of CTC into the bone marrow of immunocompromised mice induced bone, lung, and liver metastases only in three out of 110 cases of progressive MBC (2.7% efficiency), with a requirement for ≥ 1,000 CTC per injection (*13*). Hence, systemic tumor cells represent a source for metastases-inducing cells (MICs), but possess low metastatic efficiency in current experimental models.

Phenotypic changes of subpopulations or even single tumor cells along an epithelial-mesenchymal transition (EMT) are postulated to decisively regulate their tumorigenic and metastatic functionality (*3, 5, 14-20*). EMT is a cellular differentiation program that is instrumental during embryonic development, which allows epithelialized cells to differentiate into mesenchymal cells and to re-locate within the developing embryo (*21*). Carcinoma cells can recapitulate EMT to variable degree, which equips them with increased migratory and invasive capacities, and thereby promotes initial steps of the metastatic cascade (*5*). A requirement for EMT, and its reversal MET, to support metastatic growth in every carcinoma type has been challenged in animal models of pancreatic and

84 breast carcinomas (*22, 23*) and is under vivid debate (*14, 16, 24*). The former two publications
85 disclosed a function of EMT in chemoresistance, but no requirement for the EMT-transcription
86 factors (EMT-TFs) Snail and Twist for the formation of metastases (*23*). In fact, lineage-tracing of
87 breast-to-lung metastases demonstrated an epithelial origin of the metastatic cells in animal models
88 (*22*). Hence, despite a substantial body of evidence in favor of EMT as a relevant switch in systemic
89 cancer and treatment resistance (*14-16*), the actual contribution of EMT phenotype(s) of CTC or
90 DTC to metastases formation remains incompletely described.
91 Typically, CTC are enriched through selection of epithelial cells from the blood *via* the cell surface
92 marker EpCAM (epithelial cell adhesion molecule). However, EpCAM expression can be lost
93 during EMT (*25*), which hampers the study of subpopulations of CTC that have potentially
94 undergone EMT. The development of CTC enrichment protocols that are independent of EpCAM
95 as a marker for retrieval allowed for the analysis of EMT features. Based on gene expression
96 profiling, epithelial, biphenotypic epithelial-mesenchymal, and mesenchymal CTC were isolated
97 from blood samples of patients suffering from various carcinomas, including breast cancer (*26*). A
98 mesenchymal status of CTC was associated with poor treatment response and disease progression
99 in MBC (*22, 26, 27*), demonstrating the relevance of a mesenchymal transition for therapy.
00 However, it remains a matter of debate which phenotype of systemic cells is required for the actual
01 induction of metastases (*14, 16, 24*).
02 In the present study, we functionally related EMT phenotypes of CTC and DTC with the ability to
03 form lung metastases in a mouse model of MBC. Systemic tumor populations with a hybrid
04 phenotype, defined as primarily epithelial with a moderate transition to mesenchymal traits (E/m-
05 type), represented the most aggressive cells in this model. Functional findings were confirmed in a
06 clinical cohort of stage III-IV breast cancer patients, in which higher proportions of $EpCAM^{+}$ cells
07 amongst CTC and DTC correlated with distant metastases. The proportion of CTC and DTC with
08 an epithelial phenotype, as measured by their expression of EpCAM, correlated with the occurrence

of lung metastases. Furthermore, EpCAM^{+} DTC predicted poor 6-months survival, and correlated with decreased overall survival. As a potential consequence, clinical modalities should consider different CTC and DTC subpopulations based on their EMT phenotype as targets for multimodal therapy to reduce treatment resistance and metastatic outgrowth.

Results

EMT phenotypes of systemic cancer cells in the syngeneic 4T1 MBC mouse model

4T1 cells are 6-thioguanine-resistent murine MBC cells derived from a lung metastasis of the 410.4 cell line, itself a fourth transplant generation of a metastatic nodule of the syngeneic 410 tumor cell line in Balb/c mice (*28, 29*). 4T1 cells generate primary tumors and spontaneously metastasize to multiple distant sites following syngeneic transplantation in immuno-competent Balb/c mice, and closely reproduce stage IV of human breast cancer progression. We used this model to isolate and characterize cellular intermediates of the metastatic cascade *ex vivo*, and to analyze the impact of EMT on their functionality *in vitro* and *in vivo*. 4T1 cells were subcutaneously transplanted in the flank of BALB/c mice and mice were sacrificed to collect primary tumors, blood, bones, and organs for the recovery of 4T1 cells through selection with 6-thioguanine (6-TG) (**Fig 1A**). Epithelial and pan-carcinoma marker EpCAM, which serves as the major marker to isolate systemic cancer cells in clinical settings (*12*), was used to characterize the epithelial status of 4T1 cells before transplantation. A majority (> 85%) of cells expressed EpCAM at high levels, with only a minority of cells being low or negative for EpCAM (**Fig S1A**).

Following syngeneic transplantation (n = 5 mice), a 4T1 cell line derived from the blood (CTC1) and a 4T1 cell line derived from the bone marrow (DTC1) were stably expanded *ex vivo* each from a separate mouse. CTC1 and DTC1 are adherent cell lines that were confirmed as authentic 4T1-derived cells through karyotyping and detailed analysis of chromosomal aberrations (**Fig S1B**), and lacked the expression of the white blood cell marker CD45 (**Fig S1C**). Additionally, 4T1, CTC1,

and DTC1 were resistant to 6-TG treatment, whereas murine NIH3T3 fibroblasts, as controls, only grew in the absence of 6-TG (**Fig S1D**).

The morphology of 4T1, CTC1, and DTC1 differed considerably. Parental 4T1 cells displayed a typical epithelial phenotype with tight cell-cell contacts (E-type), whereas CTC1 cells displayed a mesenchymal, spindle-shaped phenotype with loss of cell-cell adhesion (M-type) (**Fig 1B**). DTC1 cells were characterized by a hybrid phenotype with a majority of cells that retained an epithelial phenotype and cell-cell contact, though with reduced strength as compared to 4T1 cells, and a minor subpopulation of cells with enhanced mesenchymal appearance (E/m-type) (**Fig 1B**). Culture of 4T1 cells in selection medium over a time period of 28 days confirmed that 6-TG had no impact on the epithelial phenotype of cells and did not induce EMT (**Fig S1E**).

Immunohistochemistry (IHC) staining showed that 4T1, CTC1, and DTC1 expressed epithelial marker cytokeratin and mesenchymal marker vimentin. Furthermore, 4T1 and DTC1, but not CTC1 cells expressed high levels of epithelial markers EpCAM and E-cadherin (**Fig 1C**). Total loss of EpCAM expression in CTC1 cells was confirmed through flow cytometry analysis. DTC1 cells were characterized by an additional population of cells with approximately 10-fold reduced expression of EpCAM and an overall 50% reduction of EpCAM expression as compared to parental 4T1 cells (**Fig 1D**). mRNA levels of epithelial markers EpCAM, E-cadherin, and Rab25 as well as of mesenchymal markers N-cadherin, vimentin, Slug, Zeb1, and Zeb2 were assessed in 4T1, CTC1 and DTC1 cells. A significant decrease of epithelial markers (EpCAM, E-cadherin, Rab25, Grhl2) and a marked increase of mesenchymal makers (N-cadherin, Vimentin, Slug, Zeb1, Zeb2) was observed in CTC1 cells (**Fig 1E** and **Fig S2A**). No significant differences were observed for the expression of Ddr1, ErbB2 and ErbB3, while Krt19 was up-regulated, and Snail and Twist were down-regulated in CTC1 cells (**Fig S2A**). Measurement of mRNA levels in DTC1 reflected an overall partial loss of epithelial features, with decreased EpCAM, E-cadherin, and Rab25 levels and an increase in vimentin expression (**Fig 1E**).

Mesenchymal transition in CTC1 correlates with increased migration, but impaired proliferation and tumor formation capacity

In vitro functional and *in vivo* tumorigenic assays were performed to address the connection between EMT phenotypes and the functional behavior of 4T1-derived tumor cells. Metabolism and cell numbers were assessed in cell culture after five days. 4T1 cells displayed the highest cell metabolism compared to DTC1 (intermediate) and CTC1 cells (lowest) (**Fig 1F**), which was in line with higher cell numbers in 4T1 and lowest cell counts in CTC1, while DTC1 cells displayed intermediate counts (**Fig S2B**).

2D colony formation tests every single cell in the population for its ability to undergo unlimited division. 3D soft agar colony formation tests for anchorage-independent cell growth and repression of anoikis under non-adhesive conditions. 4T1, CTC1, and DTC1 had similar capacity in 2D anchorage-dependent cell growth (**Fig S2C**), although with an increased average colony size for CTC1 cells, which was connected to loosened cell-cell contacts within CTC1 colonies, as compared to 4T1 and DTC1 cells (**Fig S2C**). Furthermore, CTC1 had strongly and DTC1 slightly enhanced anchorage-independent cell growth capacity in 3D soft agar colony formation comparing to parental 4T1 cells (**Fig 1G** and **Fig S2D**). Generally, 4T1 cells formed smaller and highly compacted 3D colonies with sharply defined edges, whereas CTC1 cells formed bigger colonies of less defined shape and loose edges. DTC1 cells formed intermediately sized colonies with varying edge features (**Fig S2D**).

Next, adhesion of 4T1, CTC1, and DTC1 cells to murine endothelial cells, matrigel, and gelatin was assessed *in vitro*. DTC1 cells displayed significantly higher adhesion to endothelial cells than 4T1 and CTC1 cells (**Fig 1H**). Additionally, CTC1 cells were characterized by reduced adhesion to matrigel and gelatin, compared to 4T1 and DTC1 cells (**Fig S2E**). Cell migration was addressed in wound-healing experiments, demonstrating 2.3- and 1.9-fold enhanced migration of CTC1 and DTC1, respectively, as compared to parental 4T1 cells (**Fig 1I** and **Fig S2F**). Similarly, CTC1 cells

possessed the highest invasive capacity in a matrigel-coated Boyden chamber assay, while 4T1 cells displayed the lowest and DTC1 an intermediate invasive potential (**Fig 1J** and **Fig S2G**). Hence, EMT observed in CTC1 cells was accompanied by reduced proliferation and adhesion, enhanced migration, anchorage-independent growth, and invasion capacity. DTC1 cells displayed overall improved capacities, with retained proliferation, enhanced adhesion, migration, invasion, and slightly higher anchorage independent growth.

Next, the tumorigenic ability of all three cell lines was assessed *in vivo* through subcutaneous transplantation of identical cell numbers of 4T1, CTC1, or DTC1 cells into the flank of Balb/c mice. Tumor weights were quantified for all three cell lines in parallel after three weeks, and blood and bones were collected for *ex vivo* cultures. The average tumor weight and size were the highest in DTC1-transplanted mice (n = 8), with a 100% frequency of tumor formation (**Fig 2A** and **Fig S3A**). Similarly, all 4T1-transplanted mice established tumors (n = 13), however with a significantly reduced tumor weight as compared to DTC1 cells (**Fig 2A**). The size of 4T1 tumors was also reduced compared to DTC1, but differences did not reach statistical significance (**Fig S3A**). Transplantation of CTC1 cells led to tumor formation in 7 out of 17 injected mice (41.2% frequency) with reduced average tumor weight and size compared to 4T1 and DTC1 cells (**Fig 2A** and **Fig S3A**).

IHC staining of EpCAM and vimentin was performed in primary tumors of 4T1-, CTC1-, and DTC1-transplanted mice. Primary tumors generated after transplantation of CTC1 cells remained $EpCAM^-$, whereas primary tumors from 4T1, CTC1, DTC1 had comparable vimentin expression levels (**Fig S3B**). This suggests that CTC1 cells formed primary tumors without re-expression of epithelial marker EpCAM.

The tumorigenic potential of CTC1 cells was inferior to 4T1 and DTC1 at identical numbers of injected cells (*i.e.* $1.25*10^5$, **Fig 2A**). In order to further analyze the tumorigenic potential of CTC1 cells, we conducted subcutaneous injections with cells numbers in large excess of 4T1 and DTC1

transplantations (*i.e.* $5*10^5$, $1*10^6$, and $2*10^6$). An 8-fold and a 16-fold excess of CTC1 cells was required to reach tumor weights and sizes induced by injection of 4T1 and DTC1 cells, respectively (**Fig 2B** and **Fig S3C**). The metastatic potential of 4T1, CTC1, and DTC1 cells was evaluated after subcutaneous injection upon colony formation from excised lungs under 6-TG selection. 4T1 and DTC1 cells generated lung metastases at equal frequency (4/5 mice; 80%), while CTC1 cells generated lung metastases in 3/10 mice after injection of $5*10^5$ and 10^6 cells, and in 8/10 mice after injection of $2*10^6$ cells (**Fig 2C**). Average numbers of metastatic colonies after selection in 6-TG revealed the highest in DTC1-injected mice, whereas injection of $5*10^5$ and 10^6 CTC1 cells resulted in low average colony numbers, and in intermediate colony numbers after injection of $2*10^6$ cells (**Fig S3D**). The metastatic index per cell was calculated as numbers of lung metastatic colonies divided by the number of injected cells. The average metastatic index per cell of DTC1 was highest with a value of $7.04*10^{-5}$, and was significantly higher than metastatic indexes for CTC1 cells, independently of the amounts of injected cells ($1.4*10^{-6}$, $2.2*10^{-6}$, and $2.7*10^{-6}$, respectively). The metastatic index of 4T1 cells ($2.56*10^{-5}$) was 9.5- to 18-fold higher than CTC1 cells (**Fig 2C**). Hence, single CTC1 cells have substantially reduced tumorigenic and metastatic potential in comparison with 4T1 and DTC1 cells.

Frequencies of *ex vivo* cultures from primary tumors, organs (lung, liver, kidney, spleen), blood, and bone marrow, are shown per injected mice in **Figs S3E** and **F**. One blood culture (CTC1) and one bone marrow culture (DTC1) were established from 4T1-injected mice. Re-transplantation of CTC1 cells failed to establish any CTC or DTC sublines. In contrast, subcutaneous re-transplantation of DTC1 cells allowed to establish n = 26 CTC sublines from blood in 4 out of 8 mice, and to establish n = 10 DTC sublines from bone marrows of 2 out of 8 mice (**Fig S3E and F**).

In order to address potential differences between 4T1 and DTC1 cells that might explain the increased tumorigenicity and metastatic capacity of DTC1 cells, chromosomal aberrations were

analyzed in detail after karyotyping of cell lines. Differential aberrations between 4T1 and DTC1 were determined with Fischer´s exact test, resulting in two categories of aberrations: 1) aberrations occurring in both cell lines, but with significantly different frequencies, and 2) aberrations occurring exclusively in either cell line. A list of genes encoded in the genomic areas affected by aberrations was extracted using Biomart/Ensembl (http://www.ensembl.org; n = 1546 protein coding genes) and filtered for genes associated with cancer processes using a list of 419 genes deposited on the MTB (http://tumor.informatics.jax.org/mtbwi/index.do) and based on a census of human cancer genes (*30*). This resulted in a list of n = 34 genes (**Fig S3G**) that was used to perform a GO-term analysis using the functional annotation tool of the DAVID bioinformatics database (https://david.ncifcrf.gov/). All 34 cancer genes extracted from chromosomal breakpoints in 4T1 and DTC1 cells were compared with the 419 cancer-associated genes with the query name "*Goterm_BP_DIRECT*", resulting in smaller groups with improved descriptive value (**Fig S3H, I**). The breakpoint-related genes indicated in the Venn diagram were assigned to the three super-ordinated GO-terms "*Cell cycle*", "*Signal transduction*", and "*Regulation of cellular response to stress*" (**Fig S3H**). The GO-term that included the highest number of genes (n = 6) was "*Positive regulation of ERK1 and ERK2 cascade*" (**Fig S3I**).

DTC1-derived CTC lines display EMT heterogeneity

A total of 26 CTC lines were isolated from the blood of DTC1-transplanted mice, which were all CD45-negative (**Fig S1C**). Based on low numbers of tumor cells retrieved from blood, subsequent selection with 6-TG, and observation of cell growth in 96-well format, we concluded that CTC cell lines represented either mono- or oligoclones. DTC1-derived CTC sublines had substantial differences in morphology, potentially representing systemic tumor cells in different EMT stages. In order to quantify the grade of EMT in these CTC lines, we applied a scoring system implementing the percentage of mesenchymal, spindle-shaped cells (0-100%) and the level of cell-cell contact (1-4; see Materials and Methods), to obtain an EMT score ranging from 0 (epithelial,

4T1) to 400 (mesenchymal, CTC1) (**Fig S4A**). **Fig 2D** shows representative pictures of *ex vivo* cultured CTC sublines derived from 4 out of 8 DTC1-transplanted mice, demonstrating the phenotypic transition from epithelial to mesenchymal. EMT scores were evenly distributed and did not show any bias across CTC sublines (**Fig S4B**). Phenotypic heterogeneity from E-, E/m-, M/e- to an M-type was observed in CTC sublines originating from different mice, but also within one mouse (**Fig 2E** and **Fig S4B**). This demonstrates the presence of CTC with differing EMT phenotypes in the blood of individual mice.

We selected E/m-type (CTC6-6, CTC6-11, CTC8-12) and M/e-type (CTC8-6, CTC8-5, CTC8-1) DTC1-derived CTC sublines along with parental 4T1, CTC1, and DTC1 cells to decipher whether EMT traits associated with differing tumorigenic abilities. The cellular origin was analyzed by karyotyping and confirmed that all cell lines were 4T1 derivatives (**Fig S1B**). In immunohistochemistry staining, we observed a low expression of EpCAM and E-cadherin in M/e CTC, with retention of EpCAM expression in a small proportion of cells. In contrast, the majority of E/m-type CTC lines expressed high levels of EpCAM and E-cadherin protein (**Fig 2F**). All selected cell lines expressed substantial amounts of vimentin (**Fig S4C**). mRNA transcript levels confirmed a higher expression of EpCAM, E-cadherin, and Rab25 in E/m-type CTC lines, although generally reduced as compared to parental 4T1 cells, especially for the case of Rab25 (**Fig S4D**). Ddr1, Grhl2, and Krt19 expression was similar in both CTC phenotypic subtypes (**Fig S4D**). mRNA transcript levels of the EMT-related genes vimentin, Slug, and Zeb2 were significantly higher in the M/e subtype of CTC, while N-cad, Zeb1, ErbB2, ErbB3, Snail, and Twist did not show significantly differences (**Fig S4D**). Cell metabolism and proliferation rates of all DTC1-derived CTC cell sublines were generally below DTC1 and 4T1, but 2-fold higher on average than CTC1. Additionally, significantly higher cell metabolism and proliferation rates were observed in the E/m group in comparison with the M/e group (**Fig 2G** and **Fig S5A**).

The adhesion property of E/m-type CTC sublines to endothelial cells, matrigel, and gelatin was higher than that of M/e-type CTC sublines (**Fig 2H** and **Fig S5B**). Additionally, a significant but very minor increase in invasion capacity was observed for E/m-type CTC (**Fig 2I**).

Mesenchymal-type CTC sublines are more resistant to chemotherapy than epithelial-type CTC

Recent evidence suggests an association between EMT and chemoresistance, including breast and pancreatic cancer (*5*). In order to compare the resistance or vulnerability towards standard chemotherapeutics, E (4T1), E/m- (CTC6-6, CTC6-11, CTC8-12, DTC1), M/e- (CTC8-6, CTC8-5, CTC8-1), and M-type (CTC1) sublines were treated with cisplatin and doxorubicin at increasing concentrations for 48 h. Metabolic activity was detected by MTT assay. Concentration curves showed that mesenchymal-type sublines (M, M/e) had increased chemoresistance as compared to epithelial-type (E, E/m) sublines, both for cisplatin (IC50 mean values: M: 18.81 µM, M/e: 18.12 µM, E: 10.37 µM, E/m: 11.35 µM) and, more pronouncedly, for doxorubicin (IC50 mean values: M: 4.51 µM, M/e: 3.05 µM, E: 0.66 µM, E/m: 0.93 µM) (**Fig S5C**). We conclude that these mesenchymal-type CTC sublines have enhanced resistance towards clinically relevant chemotherapeutic drugs.

E/m-type CTC sublines possess highest metastasis formation ability *in vivo*

The metastatic index of 4T1, DTC1, and, especially, CTC1 differed considerably following subcutaneous injection of cells. In order to address the capacity of cells with different EMT phenotypes to generate lung metastases after intravasation into the blood stream, we performed intravenous (*i. v.*) injections. E-type cells (E; 4T1 as a control), DTC1-derived CTC sublines with an epithelial/mesenchymal (E/m; CTC6-6, CTC6-11, CTC8-12, DTC1 as control), a mesenchymal/epithelial phenotype (M/e; CTC8-6, CTC8-5, CTC8-1), and M-type cells (M; CTC1) were injected at equal cell numbers ($5*10^4$) in the tail vein of Balb/c mice (**Fig 3A**). After 19 days, formation of lung metastasis was measured by counting superficial metastases and by *ex vivo*

metastasis colony formation assay, and was eventually implemented in a metastatic index per injected cell. The results demonstrated that cells with a predominantly epithelial phenotype (E, E/m) had enhanced metastasis-inducing ability compared to mesenchymal-type cells (M, M/e). CTC lines with a hybrid E/m phenotype exhibited the highest capability to trigger metastasis, which was also higher than parental, E-type 4T1 cells (**Fig 3B** and **Fig S5D and E**).

Differences in metastatic indexes may result from a longer latency time of M/e-type CTC to develop lung metastases, and not from an inherently reduced metastatic capacity. Therefore, M/e-type clones CTC8-6, CTC8-5, and CTC8-1, and mesenchymal clone CTC1 were injected intravenously into BALB/c mice, along with E/m-type clone CTC6-6 with the highest metastatic index as a positive control (each n = 5 per cell line). Animals were observed daily for signs for an endpoint and were sacrificed if required (see Materials and Methods). After 22 days, all E/m-type CTC6-6-injected mice (100%) and one CTC8-1-injected mouse (20%) had to be sacrificed based on significant weight loss, weakness, and dyspnea (**Fig 3C**). Upon autopsy, severe lung metastases (average ≥10 metastases per lung) were observed in all animals and were confirmed by metastatic lung colony formation assay (**Fig S5F and G**). After 25 days, the remaining n = 4 mice injected with CTC8-1 (80 %) and all 5 mice injected with CTC8-5 cells (100 %) displayed similar signs for endpoint. Upon autopsy, 2 out 9 mice revealed severe lung metastases (22.2 %), 4 out of 9 mice had metastases in the proximity of larger bones (44.4 %), and 3 out of 9 mice had multiple tumor sites (33.3 %) (**Fig 3C**). At day 28, all mice injected with CTC8-6 cells displayed significant weight loss and clinical weakness. Mice injected with CTC1 cells did not show any signs of deterioration. Based on predefined endpoints, all remaining animals were sacrificed and analyzed at day 28. None of the CTC8-6-injected mice had severe lung metastases (0%), 5 out of 5 mice had metastases in the proximity of big bones (100%), and 3 out of 5 mice had multiple tumor sites (60%) (**Fig3 C and D**). None of the CTC1-injected mice displayed lung metastases, whereas 2 out of 5 mice were bearing small tumors in the tail area, *i.e.* the injection site (**Fig 3C**).

Numbers of superficial lung metastases, metastatic colonies, and the according metastatic indexes per injected cell are depicted in **Fig 3E and Figs S5F and G**. All parameters confirmed significantly higher metastatic index of E/m-type CTC6-6 cells, despite prolonged seeding times for the remaining M/e- and M-type CTC sublines. The more aggressive phenotype of E/m-type CTC6-6 cells was further underscored by premature and substantial weight loss of the injected animals (**Fig S5H**).

EMT in CTC sublines is not a reflection of 4T1 cell heterogeneity

4T1 cells and DTC1 cells display a highly variable phenotype in conjunction with numerous genetic alterations, which could account for heterogeneous EMT phenotypes observed in CTC sublines retrieved from the blood of DTC1-injected mice. In order to test this hypothesis, single cell clones (SCC) of 4T1 (n = 30), CTC1 (n = 23), and DTC1 (n = 30) were generated and compared with CTC sublines (n = 26) derived from the blood of DTC1-injected mice. Maximal EMT scores of 4T1-SCC ranged from 0-80, demonstrating a high degree of retention of their epithelial phenotype. EMT scores of CTC1-SCC ranged from 380-400, proving the steady mesenchymal phenotype of the CTC1 subline. DTC1-SCC and DTC1-derived CTC sublines had maximal EMT scores of 0-150 and 0-400, respectively, across all three independent measurements for each cell line. Resulting average EMT scores for DTC1-SCC and DTC1-derived CTC sublines were ranging from 0-90 and 0-360, respectively (**Fig S6A and B**). Hence, the range of EMT scores in circulating tumor cells is broader and not fully depicted by the heterogeneity of single cell clones. These findings were further underscored by substantially different variances (*i.e.* squared standard deviations) of 736.86 and 14428.12, and of ranges (90 and 345) for DTC1-SCC and DTC1-derived CTC sublines (**Fig S6B**). In confirmation, mean values of EpCAM expression were more broadly distributed in DTC1-derived CTC sublines than in DTC1-SCC, with increased variance and range (**Fig S6C**).

Taken together, these data demonstrate that EMT phenotypes in CTC sublines isolated from the blood of inoculated mice cannot be the sole consequence of the heterogeneity of parental cells.

EpCAM expression is a valid surrogate marker for EMT

We analyzed systematically the correlation of EpCAM expression and EMT phenotypes in the 4T1 MBC model, with the aim to use EpCAM as a surrogate marker in clinical samples of MBC. Cell surface expression of EpCAM was quantified by flow cytometry in 4T1-derived cell lines, including cell lines re-cultured from primary tumors and organ metastases, and in all CTC and DTC sublines re-isolated from the blood and bone marrow, respectively, of 4T1-, CTC1-, and DTC1-injected mice. In 4T1-derived sublines, EpCAM showed a down-regulation in a subset of primary tumor-derived sublines as compared to parental 4T1 cells, while metastasis-derived sublines displayed overall high EpCAM levels (**Fig S7A)**. EpCAM remained absent in sublines re-cultured from CTC1-derived primary tumors and organ metastases (**Fig S7B**; MFI-R=1 represents lack of expression). Accordingly, sublines from CTC1-derived primary tumors and organ metastases maintained a mesenchymal phenotype in *ex vivo* culture conditions (data not shown). In sublines of the DTC1-transplanted animals, average EpCAM expression was higher in DTC re-cultured from the bone marrow and in sublines re-cultured from organ metastases comparing with CTC sublines from the blood and primary tumors (**Fig 4A**). CTC sublines isolated from the blood of DTC1-injected mice displayed substantial heterogeneity of EpCAM expression, which was also observed across cell lines originating from the same mouse (**Fig 4A**). Strong expression of EpCAM correlated negatively with higher EMT scores (high EMT scores represent an EMT phenotype) in Spearman's rank correlation testing (**Fig 4B**, $r = -0.728$, $p < 0.001$). Thus, EpCAM expression is heterogeneous across primary and systemic 4T1-derived sublines, and high levels of EpCAM expression in CTC associated with the retention of an epithelial phenotype. Furthermore, a high degree of EMT heterogeneity was monitored in blood-derived circulating tumor cells at the level of individual animals.

Based on the described high level of EMT heterogeneity and on the correlation of EpCAM expression with the epithelial phenotype of CTC in the 4T1 MBC model, we further investigated

whether comparable observations can be made in MBC patients. Pairs of primary tumors and corresponding lymph node metastases (n = 12), liver metastases (n = 10), lung metastases (n = 8), and bone metastases (n = 8) were collected from breast cancer patients after surgery, and IHC staining of EpCAM was performed. IHC scoring results demonstrated that the expression of EpCAM was higher in metastases compared with primary tumors (**Fig 4C** and **Fig S7C**), validating results of the 4T1 MBC model.

Proportions of EpCAM$^+$ DTC predict metastases and survival of MBC patients

To extend our findings, the epithelial status of CTC and DTC in MBC patients was prospectively examined through the level of EpCAM expression by using an EpCAM-independent strategy integrating subtraction enrichment (SE) and immunostaining-FISH (iFISH) technologies (*31*). Multi-marker subtraction enrichment served to deplete WBCs from the blood and bone marrows of stage III-IV MBC patients (n = 34; see **Fig S8A**). Remaining WBCs were detected with CD45-specific staining in enriched cells in order to exclude them from further analysis. Karyotypic characterization of the ploidy status of tumor cells was performed by *in situ* hybridization with chromosome enumeration probes hybridizing to human chromosome 8 (CEP8) (**Fig S8B**). CEP8 was chosen based on the frequent alteration in chromosome 8 in cancer, including breast cancer. In addition, cell sizes for potential CTC and DTC were compared to WBC, and revealed equal or smaller for 62.2% and 95.9% of cells analyzed, respectively (**Fig S8B**). While all CTC (100%) and the vast majority of DTC (91.9%) in the fraction of smaller cells were aneuploid, the proportion of diploid CTC and DTC increased in cells with similar (8.4% and 24.5%) and larger size than WBC (34.1 and 15.4%) (**Fig S8B**). CTC in the blood and contemporaneous DTC from the bone marrow were enriched from all n = 34 first diagnosed metastatic breast cancer patients who would receive standard of care treatment.

To study genetic changes between EpCAM$^+$ and -negative CTC, single cell DNA-sequencing was performed to detect genome-wide copy number variation (CNV) in EpCAM$^+$ (EpCAM$^+$; n = 10; n

= 7 aneuploid, n = 3 diploid) and EpCAM$^-$ (EpCAM$^-$; n = 20 aneuploid) CTC, isolated from 3 out of 34 MBC patients. CNV profiles were standardized to white blood cells (n = 4). A comparison of single cell DNA sequencing from EpCAM$^+$ and EpCAM$^-$ CTC revealed a total of 657 CNVs between the two cell types (amplifications and deletions), which comprised 1255 coding genes (**Fig S9A and B**). Unsupervised clustering of the top 100 CNVs and the genes encoded within the affected genomic region discriminated EpCAM$^+$ from EpCAM$^-$ CTC (**Fig S9C)**. To examine potential functional implications of the affected genes, a Gene Ontology (GO) term enrichment analysis was performed (**Fig S9**). Within the enriched "biological process" GO terms, amplification of genes in tight junction (CLDN3, STRN, PTPN13), mitotic cell cycle (CCNB1, SHB, EIF4EBP1, DUSP3, ABL1), mammary gland epithelial cell differentiation (ERBB4), and mammary gland duct morphogenesis (GLI2, CSF1R) indicated an increase ability of cell adhesion, proliferation, and epithelial differentiation of EpCAM$^+$ CTC (**Fig 4D**). This is in line with reported functions of EpCAM in cell adhesion, proliferation, and endodermal/epithelial differentiation (*32*). All enriched GO terms with P value < 0.05 are summarized in **Fig S10**.

In the following, we assessed whether EpCAM expression levels associated with the metastatic status and disease outcome of MBC patients. Representative SE-iFISH results for single CTC and DTC, as well as clustered cells of each group, are depicted in **Fig 5A**. Total numbers of 845 CTC and 71,910 DTC were isolated from n = 34 patients; the median cell number detected per patient was 9 CTC and 413 DTC, and the correlation of CTC and DTC numbers per patient is shown in the upper panel of **Fig 5B**. Likewise, numbers of cell clusters were significantly higher in bone marrow as compared to blood (**Fig 5B**, middle panel). The proportions of EpCAM$^+$ CTC and DTC, defined as the number of EpCAM$^+$ cells divided by the total detected cell number of each patient, were higher in DTC comparing with CTC **(Fig 5B**, lower panel). Twenty out of 34 patients (58.8%) had no detectable EpCAM$^+$ CTC in the blood, while only 12 out of 34 patients (35.3%) had no EpCAM$^+$ DTC in the bone marrow (**Fig 5C**). In the context of all detected CTC and DTC, the

proportion of EpCAM$^+$ CTC and DTC was 22.4% and 65.9%, respectively (**Fig 5B**). Using EpCAM as a marker for epithelial differentiation, we conclude that MBC patients are characterized by higher proportions of mesenchymal CTC and epithelial DTC.

A correlation of the proportion of EpCAM$^+$ CTC and DTC with clinical parameters disclosed that higher rates in CTC and DTC were positively correlated with detectable organ metastasis (M_1), as compared to patients without (M_0) (**Fig 5D**). Furthermore, higher proportions of EpCAM$^+$ DTC were significantly associated with the occurrence of lung metastases (**Fig 5D**). All patients were followed for a median of 11 months, and receiver operating characteristic (ROC) curve was applied to determine the sensitivity and specificity of the proportion of EpCAM$^+$ DTC for 6 months survival. The proportion of EpCAM$^+$ DTC predicted the risk of 6 months mortality of metastatic breast cancer patients with good accuracy (**Fig 5E**; AUC = 0.785, 95% CI 0.588 – 0.983; p-value 0.018) and a cut-off value of 19.78% EpCAM-positivity (sensitivity 75.0%, specificity 82.6%) was calculated. Based on results from the ROC analysis, we applied a ≥ 20% cut-off for the proportion of EpCAM$^+$ DTC and analyzed the overall survival of patients. Patients characterized by a proportion of EpCAM$^+$ DTC ≥ 20% showed a severely decreased overall survival (**Fig 5F**). Comparable analyses were performed following the exclusion of EpCAM$^+$ diploid CTC and DTC. The proportion of EpCAM$^+$ CTC and DTC was decreased to 6.3% and 56.9%, respectively (**Fig S8C**). Importantly, proportions of EpCAM$^+$ aneuploid CTC and DTC correlated with the presence of distant and lung metastases (**Fig 5H**). Furthermore, proportions of EpCAM$^+$ DTC predicted the risk of 6 months mortality of metastatic breast cancer patients with good accuracy (**Fig 5I**; AUC = 0.793, 95% CI 0.599 – 0.988; p-value 0.015) and a cut-off value of 16.87% EpCAM-positivity. Patients characterized by a proportion of EpCAM$^+$ DTC ≥ 15% showed severely decreased overall survival (**Fig 5J**).

Hence, in strong confirmation of results derived from the 4T1 MBC model, EpCAM$^+$ CTC and DTC were associated with the generation of distant metastases and lung metastases and an EpCAM-

positivity rate above 15-20% in systemic tumor cells in the bone marrow predicted considerably decreased overall survival of MBC patients.

Discussion

In the present study, we combined the syngeneic murine 4T1 MBC model with clinical samples of systemic tumor cells from MBC patients to recapitulate different stages of tumor progression and study their association with EMT. Despite a robust generation of distant metastases in lungs, actual numbers of CTC appeared generally scarce after transplantation of 4T1 cells, as reflected by the low frequency of cell lines retrieved from the blood of transplanted animals. This is in accordance with CTC numbers of ~1 CTC per ten million white blood cells (WBCs) in 7.5 mL blood sample of patients with advanced solid cancers (*8, 9*). Despite such low numbers, systemic tumor cells, *i.e.* CTC and DTC, are considered the primary source of metastases-inducing cells (MICs) (*13, 33*), which represent a major clinical challenge, but possibly also a valuable therapeutic taryopportunity (*25*).

Morphological, molecular, and phenotypic analyses of the 4T1 model disclosed a substantial degree of inter- and intra-individual EMT heterogeneity in CTC, confirming the co-existence of CTC with epithelial and mesenchymal traits in the blood of individual animals. Earlier reports addressing EMT phenotypes in human CTC demonstrated a correlation of mesenchymal CTC with therapy resistance (*22, 23, 26*), a phenotype that was also observed in the present study of murine CTC. More recently, the notion of EMT as a central process in metastases formation in MBC was challenged using cell tracing experiments in mice, suggesting that cells responsible for metastases formation had not undergone EMT and supporting a role for mesenchymal tumor cells in chemoresistance (*22, 23*). Analysis of EMT phenotypes in prostate and bladder cancer revealed an even more intricate dependency of epithelial and mesenchymal tumor-initiating cells (TICs). An epithelial gene signature was characteristic of tumor cells with strong metastatic TIC capacity, whereas a mesenchymal signature was associated with reduced metastatic TIC activity. However,

the presence of mesenchymal TICs accelerated and enhanced the metastatic ability of epithelial TICs *in vitro* and *in vivo* (*33*).

In the murine 4T1 MBC model, we did not observe any bias in EMT phenotype in a total of n = 26 CTC cell lines that were re-isolated from blood, which indicates that our model system did not select CTC with particular EMT status. However, it must be noted that the vast majority of CTC lines originated from animals transplanted with bone marrow-derived DTC1 cells, which were characterized by an E/m phenotype and generally improved tumor and metastases formation capacity. Detailed analysis of the karyotypes of 4T1 and DTC1 cells allowed us to extract genes potentially affected by chromosomal breakpoints with significant differences in frequency between both cell lines. GO-term analysis of the extracted genes revealed that the term including the most genes (n = 6) was "*Positive regulation of ERK1 and ERK2 cascade*", which is of special interest, given the role of ERK activation status as a central integrator of EGFR signals to induce either proliferation or induction of EMT (*34*).

Based on the genetic instability of 4T1 cells and their origin from a lung metastasis derived from the 410.4 primary breast carcinoma cell line (*28, 29*), it was important to verify that the observed changes of EMT phenotype of *ex vivo* isolated CTC were not solely a reflection of highly variable phenotypes of subclones of 4T1 and DTC1. Sublines of 4T1 and DTC1 generated *in vitro* as single cell clones were characterized by a more restricted epithelial phenotype than CTC sublines retrieved from the blood of transplanted animals. Furthermore, the phenotype of CTC and DTC lines remained stable over the entire period of experimental assessment, which comprised ≥ 20 passages. Hence, these results corroborate changes in EMT phenotypes in systemic tumor cells present in the blood of transplanted mice.

The observed phenotypic diversity of CTC lines allowed us to subdivide CTC more specifically into E/m and M/e phenotype, which reflects more properly a frequently discussed partial EMT observed in tumors. Emerging evidence suggests that EMT is rarely an "all-or-nothing" condition.

Instead, cancer cells often adopt hybrid EMT phenotypes (*5, 14, 17, 18, 26, 35-37*). Hybrid phenotypes comparable to those characterized in the present study have been described, amongst others, in breast and ovarian cancer (*26, 36*). EMT heterogeneity was assessed using similar markers to ours: EpCAM, E-cadherin, keratins, fibronectin, cadherin-2 and serpine1/PAI1 (*26*), or morphological examination in combination with E-cadherin, pan-cytokeratin, and vimentin (*36*). In fact, Huang and colleagues defined identical EMT groups, *i.e.* epithelial, intermediate epithelial, intermediate mesenchymal, and mesenchymal phenotypes, with relevance to tumor progression and patients´ outcome (*36, 38*). Validation of the EMT spectrum through mRNA measurement further disclosed a 33 genes signature, where E-cadherin, cytokeratin 19, and vimentin expression confirmed the assignment of cell lines to the various EMT sub-groups (*36*).

Here, we demonstrate that epithelial-type CTC with a restricted mesenchymal transition compared to parental 4T1 cells (E/m cells) bear the strongest capacity to form lung metastases when directly inoculated in the blood stream (see scheme in **Fig 6**). Despite their ability to form metastases mostly in the vicinity of large bones after extended time periods, M/e-type CTC displayed a poor aptitude to form lung metastases. Tumor-associated deaths are primarily caused by metastases in life-supporting organs such as the lungs and liver. Therefore, enhanced adhesion ability of E/m-type CTC to endothelial cells could promote the retention at the endothelium and subsequent extravasation, while their enhanced proliferation rate could potentially facilitate renewed outgrowth in lungs to generate life-threatening metastases (**Fig 6**). It cannot be excluded that E/m-type CTC have initially undergone EMT to intravasate in blood vessels and have subsequently reverted their phenotype through mesenchymal-epithelial transition (MET) during their residency in mice. To clarify this point, options of genetic tracking of EMT during metastases formation are available in animal models (*22, 23*). However, the plethora of molecules involved in the process cannot be assessed at once, and thus definitive claims about a lack of requirement for EMT during tumor progression can hardly be made (*14, 16*). Furthermore, although subcutaneous transplantation of

4T1 cells will rather underestimate metastatic outgrowth, the impact of the two microenvironments encountered following subcutaneous *versus* orthotopic transplantation might differentially impact on EMT regulation. A multitude of parameters including soluble factors, cell-associated ligands, exosomes, and miRNAs can differ between the two primary tumor localizations, and might affect the epithelial phenotype of transplanted cells. It can nonetheless be concluded from our data that a primarily EpCAM$^+$ epithelial phenotype endorses systemic tumor cells from MBC with improved MIC capacity to generate lung metastases, which is in line with studies on MBC (*13, 19, 20*), prostate and bladder carcinomas (*33, 39*), and pancreatic carcinomas (*23*).

Importantly, findings from the 4T1 MBC model were validated in a clinical cohort of stage III-IV MBC patients (n = 34). Using the EpCAM-independent, multi-parameter enrichment SE-iFISH technology (*31*), we could demonstrate a frequent loss of epithelial phenotype in CTC and retention of epithelial phenotype in bone marrow-derived DTC, using EpCAM as a robust surrogate marker for EMT. Despite the comparably small number of patients enrolled in the study, we confirmed a significant correlation of the proportion of EpCAM$^+$ CTC and DTC with the occurrence of distant metastases and, more specifically, of lung metastases. Furthermore, the proportion of EpCAM$^+$ DTC accurately predicted 6-month survival and overall survival with a cut-off of 15-20% of EpCAM$^+$ tumor cells, which was extrapolated from the ROC analyses of the 6-months survival of the patients enrolled in the study.

The presence of single aneuploid circulating CD31$^+$ endothelial cells (CEC) was reported in the blood of cancer patients (*31*), which might impact on the interpretation of our data. Although not all systemic tumor cells enriched in the present study could be tested for CD31 expression, because the SE-iFISH CEC quantification technology was not finalized when the first patients of the study were enrolled, CD31-staining of all CTC selected for DNA-sequencing was performed and revealed negative. Additionally, proportions of EpCAM$^+$ cells were used as stratification parameter, rather than absolute numbers of systemic tumor cells, to discriminate the patients´ outcome. Thus,

depending on the numbers of CEC, the proportions of EpCAM$^+$ CTC/DTC will either remain stable or slightly decrease. We argue that this strategy does not over-estimate the ability of EpCAM$^+$ systemic cells to predict distant metastases and clinical outcome. Hence, the results from our prospective MBC patients´ cohort fortify the notion that EpCAM$^+$ systemic tumor cells represent the major source of MICs, and strongly validate results from the 4T1 animal model.

GO analysis of genes encoded by genomic regions affected by CNVs were obtained from single cell DNA-sequencing from three patients. GO terms analysis disclosed that genetic amplifications of genes involved in tight junction (including gene products that reportedly are interaction partners of EpCAM such as claudins), cell cycle regulation, and mammary epithelial cell differentiation were enriched in human EpCAM$^+$ CTC comparing to EpCAM$^-$ CTC. These findings further suggest the importance of epithelial traits, adhesion and proliferation capacities of systemic tumor cells for the process of metastases generation. It must however be noted, that despite indications for an enrichment of epithelial traits in EpCAM$^+$ CTC, no genetic alterations that would affect genes associated with the induction of a mesenchymal status (*e.g.* EMT transcription factors) were found to be enriched in EpCAM$^-$ CTC. Hence, single cell DNA sequencing confirmed EpCAM as a valid marker of the epithelial status of systemic tumor cells, but did not provide final evidence for the mesenchymal status of CTC.

Clusters of systemic tumor cells, although rare in the circulation, bear 23-50-fold increased metastatic potential compared to single CTC (*40*). In accordance, DTC isolated in stage III-IV patients in the present cohort were characterized by higher cluster formation in line with an enhanced epithelial phenotype.

Finally, a central application of CTC is their usage as liquid biopsy to harvest comprehensive instantaneous information of systemic cancer (*6, 8, 25*). To this end, EpCAM is so far the sole marker with clearance for clinical approaches within the CellSearch system. Owing to issues of EpCAM loss during EMT, concerns were raised that CTC isolation systems might underestimate

numbers and possibly oversee clinically relevant cells. Based on our results, we suggest that EpCAM-dependent enrichment systems will indeed underestimate CTC numbers, but will quantify clinically relevant cells. Accordingly, numbers of $EpCAM^+$ CTC predict clinical outcome of metastatic and non-metastatic BC patients (*8, 41*). As a synthesis, it would be beneficial to quantify total amounts of CTC (and DTC) and to subdivide systemic tumor cells according to their EMT phenotype, in order to improve the prediction of the metastatic risk and to support treatment decision-making.

In summary, our data facilitate the understanding of the role of EMT in cancer metastasis by using a mouse model of MBC to accurately recapitulate the clinical situation of MBC. We demonstrate in the MBC mouse model and verify in a clinical cohort of stage III-IV MBC patients that a subpopulation of systemic tumor cells with a hybrid E/m phenotype greatly contributes to the formation of outcome-limiting metastases.

Materials and Methods

Experimental Design

The objectives of the present study were to assess the association of differing EMT phenotypes in systemic cancer cells of metastatic breast cancer (MBC) with their ability to form lung metastases, *in vitro*, *ex vivo*, and *in vivo* in the 4T1 MBC syngeneic mouse model. Our major focus was on CTC and DTC, their EMT phenotypes, proliferation, adhesion, migration, invasion, tumorigenesis, and metastatic potential. Additionally, the objective of the study was to analyze EMT phenotypes of systemic CTC and DTC from primary MBC patients, and to correlate EMT phenotypes with the metastatic status and clinical outcome of the patients.

Ethical Statements

Mouse experiments were conducted with the approval of the Regierung von Oberbayern, Munich, Germany (Az 55.2.1.54-2532-90/12 and 177/15). The clinical study was approved by the Ethics Committee of Shanghai General Hospital (ethics #2018KY153) and was performed according to

the Declaration of Helsinki Principles. Written consent to notify blood and bone marrow samples to be applied for future research was obtained for each patient under Institutional Review Board (IRB) approved protocol.

Cell culture

Murine 4T1 cells were cultivated in Dulbecco's Modified Eagle Medium (DMEM, Biochrom GmbH, Berlin, Germany) supplemented with 10% fetal bovine serum (FCS; Biochrom AG, Heidelberg, Germany) and 1 % penicillin/streptomycin (Biochrom GmbH, Berlin, Germany). 4T1 sublines derived from primary tumors, CTC, DTC and metastases were cultured in selection medium containing 60 µM 6-thioguanine (Sigma, Saint Louis, USA) in DMEM with 10 % FCS and 1 % penicillin/streptomycin (Gibco, Planegg, Germany). DTC1-derived CTC sublines, which grew in a less adhesive manner, were propagated for both adherent and semi-adherent cells. All cell lines were grown in a 5 % CO_2 atmosphere at 37 ℃.

Mouse experiments

For tumorigenicity assay, $1.25*10^5$ cells in 100 µL PBS were subcutaneously transplanted into the flank of BALB/c mice (age-matched between 6–8 weeks). After an average of 27 days, mice were anesthetized by 0.4 % isofluorane inhalation before being sacrificed. Blood collected from the orbital sinus, and femurs and tibiae were harvested for isolation of CTC and DTC, respectively. Primary tumors and organs including lung, spleen, kidney, and liver were harvested for cryopreservation (IHC staining) and for the establishment of *ex vivo* sublines. Passage numbers below five passages were used for all derived cells for reinjection in vivo and passage numbers below ten passages were used for functional studies ex vivo.

Alternatively, $1.25*10^5$ 4T1 and DTC1, $0.5*10^6$, 10^6, and $2*10^6$ CTC1 cells in 100 µL PBS were subcutaneously transplanted into the flank of BALB/c mice (age-matched between 6–8 weeks). After 15 days, mice were sacrificed and lungs were harvested for metastatic colony formation assay.

Metastatic index/cell was calculated as numbers of lung metastatic colonies divided by the numbers of injected cells.

For the *ex vivo* culture of CTC lines, blood was taken from the retro-orbital sinus under anesthesia. Retrieved blood volumes per mouse were weight-dependent and varied within the range of 0.8-1 mL per mouse. Thereafter, full volumes of blood were first depleted of red blood cells by an ammonium chloride-based lysing reagent (BD Pharm Lyse™, BD Biosciences, Heidelberg, Germany). After washing and centrifuging, cell pellets (containing white blood cells and potential CTC) were suspended in 10 mL selection medium. Serial dilution with 2-fold dilution steps was used to plate the isolated cells in 96-well plates with a starting volume of 100 µL. Cell colonies growing to high confluence in 96-well plates were transferred independently to 12-well plates and later to 6-well plates. Eventually, selected cells were scaled up to larger volumes for further maintenance.

For the *ex vivo* culture of DTC lines, the hip and knee joints were removed from femurs and tibiae, and bone marrows were flushed with sterile PBS. After washing and depleting of red blood cells, cell pellets were resuspended in 10 mL selection medium and cell lines were generated as described for CTC.

For the *ex vivo* culture of primary tumor and metastasis cell lines, tumors or organs (lung, spleen, kidney, liver) were minced and homogenized through a 100 µm filter. After washing and centrifuging, cell pellets were resuspended in 10 mL selection medium and seeded in one culture dish. Selected cells were transferred to flasks for further maintenance upon reaching confluence.

For *i.v.* metastasis formation assay, $5*10^4$ cells in 100 µL PBS were intravenously injected into the tail vein of BALB/c mice. Mice were sacrificed at day 19, metastases in lungs were counted and lungs were collected for metastasis colony formation assay. Alternatively, $5*10^4$ cells in 100 µL PBS were intravenously injected into the tail vein of BALB/c mice. Mice were observed daily and body weights were measured every 1 - 2 days. Signs for endpoints of the experiment of each group:

1) significant weight loss (≥ 5% in more than 2 mice), 2) weakness (*i.e.* tiredness, unresponsiveness in more than 2 mice); 3) less than 2 left experimental groups.

Metastasis colony formation assay was performed to quantify 4T1-derived cells in the lungs of transplanted mice. Entire lungs were minced and incubated in RPMI medium supplemented with collagenase (5 mg/mL, Sigma, Steinheim, Germany) and DNase (1 mg/mL, Sigma, Steinheim, Germany) for 30 min. Thereafter, lung fragments were homogenized through a 100 µm filter and subsequently through a 40 µm filter, which were each rinsed with 5 mL of PBS. After centrifugation, cell pellets were incubated with erythrocyte lysis reagent (BD Pharm Lyse™, BD Biosciences, Heidelberg, Germany) for 2 min. Afterwards, cells were centrifuged and resuspended in 10 mL selection medium before being diluted 1:10 and 1:100 in selection medium. 3 mL of each concentration were pipetted in 6-well plates in triplicates. Colonies were stained after 10 days for subcutaneous and after 4 days for intravenous transplantations with 1% crystal violet/ 70 % methanol solution. Clusters of ≥ 20 cells were defined as colony. Colony numbers of 1:100 dilutions are shown in Fig S5.

Metaphase preparation and SKY (spectral karyotyping) analysis

Cells were cultivated to 80% confluency on sterile glass slides in Quadriperm chambers with 5 mL cell culture medium. For the preparation of chromosome spreads colcemid (0.1 µg/mL; Roche, Basel, Switzerland) was added to the cell culture medium for additional 3 hours at 37 °C. Afterwards, medium was removed and cells were washed with PBS. Cells were incubated with 5 mL 0.075 M KCl for 25 min at 37 °C. Subsequently, 5 mL fixation solution (methanol / acetic acid 3:1) was added for 20 min. The solution was removed and 5 mL fixation solution was added for 20 min. After another fixation step for 20 min, slides were removed from the Quadriperm chamber and air dried. Metaphase preparations were kept at room temperature for at least one week. Hybridization was performed as previously described (*42*). Briefly, the slides were dehydrated and hybridized with a denatured SKY-probe mixture (SkyPaint DNA Kit, Applied Spectral Imaging,

Carlsbad, California, USA). After hybridization, slides were washed (0.5 × SSC for 5 min at 75 °C, 4 × SSC/0.1 % Tween20 for 2 min at room temperature, H_2O for 2 min at room temperature) and probes were detected using antidigoxigenin (1 : 250; Roche), avidin-Cy-5, and avidin-Cy-5.5 antibodies (both 1 : 100; Biomol, Hamburg, Germany) according to the manufacturer's protocol. Metaphase spreads were counterstained using 0.1 % 4',6-diamidino-2-phenylindole (DAPI). Spectral imaging analysis was carried out using a fluorescence microscope (ZEISS Axioplan 2) equipped with SpectreCube device and SkyView software (Applied Spectral Imaging). A minimum of 15 metaphases were analyzed to determine the karyotype of each primary culture. Chromosome aberrations were detectable by color junctions within affected chromosomes. Image analysis was performed using the SkyView imaging software (Applied Spectral Imaging, Mannheim, Germany). Subsequent analysis of GO-term enrichment in 4T1 and DTC1 cells is described in the Results section.

Flow Cytometry

Cells were washed three times in FACS buffer (PBS, 3% FCS) before incubation with EpCAM-specific antibody (BD bioscience, Heidelberg, Germany; rat anti-mouse EpCAM G8.8; 1:50 in FACS buffer, 15 min) or CD45-specific antibody (BD bioscience, rat anti-mouse CD45 30-F11; 1:50 in FACS buffer, 15 min). After centrifugation, cells were incubated with a fluorescein isothiocyanate (FITC)-conjugated secondary antibody (BD bioscience, Heidelberg, Germany; rabbit anti-rat IgG (H+L); 1:50 in FACS buffer, 15 min). Cells were centrifuged and resuspended in FACS buffer containing propidium iodide (1 mg/mL). Cell surface expression of EpCAM was analyzed in a FACS Calibur cytometer (BD Biosciences, Heidelberg, Germany). Control staining was performed by using the secondary antibody (BD bioscience, Rabbit anti Rat IgG (H+L).

Immunocytochemistry, immunohistochemistry and EMT scoring

EpEX- (BD bioscience, Heidelberg, Germany; rat anti-mouse EpCAM G8.8) and vimentin-specific antibodies (Abcam, Cambridge, UK; Rabbit monoclonal to vimentin EPR3776) were used for

immunohistochemistry. Immunostaining was performed using the avidin-biotin-peroxidase method (Vectastain, Vector laboratories, Burlingame, CA, USA) according to the manufacturers' protocol. Immunohistochemistry intensity scores (IHC score) were calculated as the product of intensity (0 to 3+) and percentage of expressing tumor cells (score 0 = 0~5 %, 1 = 5~25 %, 2 = 25~50 %, 3 = 50~ 75%, 4 = 75~100 %). IHC scores represent averages of values independently assessed by minimum two experimenters, who were blinded to sample IDs.

EMT scores were calculated as the percentage of mesenchymal-, spindle-like cells (0-100%) and the level of cell-cell contact (disseminated cells represent 0~25% - 1, 25~50% - 2, 50~75% - 3, 75~100% - 4). Three experimenters scored independently each cell line at an average confluence of 50 – 80 % for three independent passages.

Proliferation assay, MTT assay, 2D and 3D colony formation assay

For proliferation assays, 5000 cells were plated in six-well plates in duplicates. Cell numbers were manually assessed from day 1 to day 5 upon counting in a trypan blue exclusion assay. Cell metabolism was assessed by MTT assay. 1000 cells were plated in 96-well plates in triplicates. At day 1, 3, and 5, MTT solution (Sigma, Saint Louis, USA) in medium was added to a final concentration of 0.5 mg/mL for 4 h at 37 ˚C. MTT solvent (0.1 N HCl in isopropanol) was added to solubilize formazan crystals before measuring the optical density (OD) at 570 nm wavelength and 690 nm as a reference in a microplate reader (Molecular Devices, Sunnyvale, CA, USA).

Chemoresistance was assessed as follows: Cells were seeded in a 96-well plate at 5000 cells/well and grown for 24 h. Then, cells were treated with 100 µL medium containing 120, 60, 30, 15, 7.5, 3.75, 1.875 µM doxorubicin (Sigma, Steinheim, Germany) or cisplatin (Santa Cruz Biotechnolog, Heidelberg, Germany) in triplicates for 48 h. The IC50 (drug concentration inducing 50 % death of treated cells) value was calculated by using GraphPad Prism 5 (GraphPad Software, Inc., San Diego, CA, USA). Percent of cell survivability was calculated according to the following equation:

$$\text{Cell viability} = \frac{\text{mean(OD of treated cells} - \text{OD of blank)}}{\text{mean(OD of control} - \text{OD of blank)}}.$$

For 2D colony formation assays, 50, 100, 200 cells were plated in culture dishes. After 11 days, colonies were stained with 1 % crystal violet/ 70% methanol solution, where clusters of ≥ 30 cells were defined as a colony. Plating efficiency was calculated by dividing the number of colonies by the number of plated cells. For 3D colony formation assay, 5 % and 3.5% low melting point agarose (LMP; BD Biosciences, Heidelberg, Germany) in PBS was used to generate 0.5 % agarose/medium solutions by mixing 5 % agarose with culture medium 1:10. Eight mL of the solution were quickly plated into culture dishes as base layer and solidification was completed at room temperature for 30 min. Then, $1*10^4$ cells were suspended in 9 ml growth medium and mixed with 1 ml 3.5 % LMP agarose. The resulting mixture representing 1000 cells/ml, was added onto each plate for solidification. After 11 days, colony numbers were counted under a microscope and clusters of ≥ 30 cells were defined as colony. The size of colonies was measured by image J software. All the experiments were repeated at least 3 times.

Adhesion, invasion, and scratch assay

For adhesion assays, 96-well flat bottom culture plates were coated with 50 µL of gelatin (0.2 %, Sigma), matrigel (0.9 mg/mL, Corning, Bedford, US) or kept uncoated. Cells were labeled with 10 µM calcein AM (Life Technologies, Inc., Schwerte, Germany) for 30 min at 37 °C and washed three times in PBS. $2.5*10^4$ labelled cells were seeded in 50 µL medium per well and kept for 2 h adhesion time. Thereafter, plates were washed twice with PBS (input control was not washed). Before measurement of calcein AM fluorescence in a Victor Wallac instrument, cells were lysed in 2 % Triton X-100 in distilled water. For endothelial cell adhesion assay, 10^5 bEnd.3 cells were plated in 96-well culture plates in 100 µL of medium and cultured 24 – 48 h to generate a monolayer. Endothelial cells were stimulated with 10 ng/mL TNF-α (Thermo Fisher Scientific, Bleiswijk, Netherlands) for 5 h before the assay. After removal of TNF-α-containing medium, $2.5*10^4$ labeled tumor cells were seeded in 50 µL medium per well and kept for 2 h adhesion time. Further measurements were performed as mentioned above.

Transwell invasion assay was performed using transwell chambers (8 µm, Falcon, Durham, US). 10^5 cells were seeded in the upper chamber of a 24-well plate, coated with growth factor reduced matrigel (Corning, Bedford, US), in 200 µL serum-free medium. The lower chamber was filled with 800 µL medium containing 10 % FCS. The chamber was incubated at 37 °C, for 16 h. At the end of incubation, cells in the upper surface of the membrane were removed with a cotton swab. Migrated cells on the lower surface of the membrane were stained with 1 % crystal violet/ 70 % methanol solution. Then, membranes were transferred to empty 96-wells, adding 200 µL acetic acid to dissolve crystal violet stained cells. After incubating 10 minutes on an orbital shaker, OD 590 nm was measured in a microplate reader (Molecular Devices, Sunnyvale, CA, USA).

For scratch assay, cells were seeded in six-well plates and cultured to a density of 80 %. Standard culture medium was replaced by medium without FCS and 8 h later a scratch was set with a sterile pipette tip. Cells were washed thrice with PBS and three random sections of two scratches per experimental group were marked. Pictures were taken at the indicated time points under an Axiovert 25 microscope (Zeiss, Jena, Germany) with a Samsung WB750 camera (Samsung, Schwalbach, Germany). To calculate the migration velocity, the gap area was calculated using ImageJ software.

Generation of single cell derived clones

Cells were counted by trypan blue exclusion assay and 150 cells were resuspended in 30 mL medium. Cell suspensions were seeded in 96-well (100 µL/well, corresponding to 0.5 cell/well). Colonies growing to high confluence in 96-well plates were transferred independently to 6-well plates and later to larger volumes for further maintenance.

Quantitative real-time PCR

Total mRNA was prepared using the RNeasy Mini Kit and reverse-transcribed with the QuantiTect Reverse Transcription-Kit (both Qiagen, Hilden, Germany). cDNA was amplified using SYBR-Green PCR mastermix (Qiagen) and gene-specific primers in a LightCycler 480 device (Roche, Mannheim, Germany). Normalization across samples was performed using geometric average of

constitutive gene expression of glucuronidase-beta (Gusb). Gene expression levels were calculated according to the equation $2^{-\Delta\Delta CT}$, where ΔCT was defined as $CT_{gene\ of\ interest}$ - $CT_{endogenous\ control}$, ΔΔCT was defined as $\Delta CT_{gene\ of\ interest}$ - $\Delta CT_{reference}$. ΔCT values were used for statistical comparison. The following primers were used:

Ddr1, FW: 5'-TCC ATA GAC CAG AGG GAT C-3', BW: 5'-CAG GGC ATA GCG GCA CTT GG-3;

E-cadherin, FW: 5'-CAG GTC TCC TCA TGG CTT TGC-3', BW: 5'-CTT CCG AAA AGA AGG CTG TCC-3';

Epcam, FW: 5'-CAG TGT ACT TCC TAT GGT ACA CAG AAT ACT-3', BW: 5'-CTA GGC ATT AAG CTC TCT GTG GAT CTC ACC-3';

Erbb-2, FW: 5'-TCC CCA GGG AGT ATG TGA GG-3', BW: 5'- GAG GCG GGA CAC ATA TGG AG-3';

Erbb-3, FW: 5'-GCC CAA TCC TAA CCA GTG CT-3', BW: 5'-AGC CTG TAA TCT CCC GGA CT-3';

Grlh2, FW: 5'-CAC CTC TCA AGA CTG TTA CAA GAC T-3', BW: 5'-CGA GAT GAG TGG ACT TGC TAT CTC-3';

Gusp, FW: 5'-CAA CCT CTG GTG GCC TTA CC-3', BW: 5'-GGG TGT AGT AGT CAG TCA CA -3';

Krt19, FW: 5'-CTA CCT TGC TCG GAT TGA GGA G-3', BW: 5'- AGT CTC GCT GGT AGC TCA GAT G-3';

N-cadherin, FW: 5'-AGG GTG GAC GTC ATT GTA GC-3', BW: 5'-CTG TTG GGG TCT GTC-3';

Rab25, FW: 5'-TGA GCC AAG ATG GGG AAT CG-3', BW: 5'-GGA GAA CTC AAC CCC GAT GG-3';

Slug, FW: 5'-TCC CAT TAG TGA CGA AGA-3', BW: 5'-CCC AGG CTC ACA TAT TCC-3';

Snail, FW: 5'-GCG GAA GAT CTT CAA CTG CAA ATA TTG TAA C-3', BW: 5'-GCA GTG GGA GCA GGA GAA TGG CTT CTC AC-3';

Twist, FW: 5'-CGG GTC ATG GCT AAC GTG-3', BW: 5'-CAG CTT GCC ATC TTG GAG TC-3';

Vimentin, FW: 5'-CGG AAA GTG GAA TCC TTG CA-3', BW: 5'-CAC ATC GAT CTG GAC ATG CTG T-3';

Zeb1, FW: 5'-CCA TAC GAA TGC CCG AAC T-3', BW: 5'-ACA ACG GCT TGC ACC ACA-3';

Zeb2, FW: 5'-CCG TTG GAC CTG TCA TTA CC-3', BW: 5'-GAC GAT GAA GAA ACA CTG TTG TG-3'.

Patients

Thirty-four metastatic breast cancer patients with tumor stage III- IV were recruited for analysis of CTC and DTC at Shanghai General Hospital from September 2015 to April 2017. Metastasis sites were determined by computed tomography (CT), magnetic resonance imaging (MRI), and bone scintigraphy. TNM staging was assessed and given according to the Union for International Cancer Control (UICC) 2009 guidelines. Patients were followed up until January 2018.

Subtraction enrichment of CTC and DTC

Enrichment of CTC and DTC was performed according to the manufacturers´ instructions (Cytelligen, San Diego, CA, USA) (*31*). For each patient, 6 mL peripheral blood sample and 3 mL bone marrow were collected prior to therapy. The first 2 mL of sample were discarded to avoid epithelial cell contamination. Briefly, samples were washed with PBS and subsequently loaded on the non-hematopoietic cell separation matrix, followed by centrifugation at 450 g for 5 min to remove sedimented RBCs. Supernatants containing WBCs and tumor cells were incubated with anti-CD45 monoclonal antibody-coated magnetic beads (Promega, Madison, WI). Beads-bound WBCs were depleted using a magnetic separator. WBC-free supernatants containing tumor cells were centrifuged and resuspended in PBS, followed by spreading onto a Cytelligen-formatted slide and dried overnight at 30 – 32°C for immediate use or stored at −20 °C for long-term storage.

Identification of CTC and DTC by iFISH

Experiment was performed according to the product manufacture's instruction (Cytelligen, San Diego, CA, USA) (*31*). Briefly, samples on the coated slides were subjected to Vysis Centromere Probe (CEP8) Spectrum Orange (Abbott Laboratories, Abbott Park, IL, USA) hybridization for 3 hours using a S500 StatSpin ThermoBrite Slide Hybridization/Denaturation System (Abbott Molecular, Des Plaines, IL, USA). Samples were subsequently incubated with Alexa Fluor 594-conjugated monoclonal anti-CD45 and Alexa Fluor 488-conjugated monoclonal anti-EpCAM antibodies (Cytelligen, San Diego, CA, USA). Cell nuclei were stained with 4'-6-diamidino-2-phenylindole (DAPI) (Life Technologies, Carlsbad, CA, USA). Images of the identified tumor cells were automatically acquired using a Metafer-iFISH automated CRC image scanning and analysis system (*31*). CTC/DTC were defined as $DAPI^+$, $CD45^-$, heteroploid chromosome 8 cells with or without visible EpCAM; or diploid CEP8 signal with visible EpCAM. Cell clusters were defined as ≥ 2 contacted cells. To avoid bias, blood sample collection and CTC/DTC detection were co-performed by cross-blinded physicians and research scientists.

Collection of single CTC and amplification

Single CTC enriched and identified by SE-iFISH staining was isolated by means of a non-laser microscopic single cell manipulator (NMSCM, Cytelligen) followed by transferring into a clean PCR tube containing 7 μL single cell lysis buffer on ice from the REPLI-g Single Cell Kit (QIAGEN, Hilden, Germany). Each sample was incubated at 65 °C for 10 min. Whole-genome amplification (WGA) was performed using a REPLI-g Single Cell Kit according to the standard protocol in a total volume of 50 μL at 30°C for 8 h, and was terminated at 65 °C for 3 min. Amplified DNA products were stored at -20 °C.

Quality control of whole-genome amplification products and library preparation

The concentration of amplified DNA was measured using the Qubit Quantization Platform (Invitrogen, USA) and amplified DNA was purified using Agencourt Ampure XP beads (Beckman Coulter, USA). A total of 41 single-cell samples showed whole-genome DNA successfully amplified and were selected for library preparation.

Whole-genome sequencing, CNV and Gene Ontology (GO) analysis

One ng of amplified DNA was disintegrated by Tn5 at 55 ℃ for 5 min. Products of Tn5 disintegration were used to construct DNA library using the Nextera XT DNA Sample Preparation Index Kit (Illumina, USA) and were subjected to pair-end sequencing on Illumina Nextseq 500 platform with sequence depth of 0.1X. 565.5 Million raw reads were obtained from 41 cells; low quality reads and adapter sequences were trimmed by Trimmomatic V0.35 before mapping. Qualified reads were processed with the in-house bioinformatics pipeline, which was followed by best practice treatment suggested by the Genome Analysis Toolkit V 3.7 (GATK: https://software.broadinstitute.org/gatk/). In brief, FASTQ sequences were mapped to the human genome GRCh37 using BWA-MEM (https://nchcdl.sourceforge.net/project/bio-bwa/bwa-0.7.15.tar.bz2). In the end, 30 cells (including 10 $EpCAM^+$ and 20 $EpCAM^-$ cells) with mapping rates > 50%, mapping quality > 20 and duplicates > 30 were selected for copy number variation

analysis. The CNVkit software (https://cnvkit.readthedocs.io/en/stable/index.html) was used to analyze sequencing coverage and copy number in the aligned sequencing reads from targeted amplicon sequencing of $EpCAM^+$ and $EpCAM^-$ cell samples. Copy number segments were annotated to genes and regions bearing a log2 ratio of at least ± 0.4 were identified as suggestive of shallow deletions or gains. Segments with $\log2 < -1.2$ were classified as deep deletions, and those with $\log2 > 2$ were classified as amplifications. To functionally analyze genes located in our CNVs, we used the DAVID tool (ver. 6.7 Beta; http://david.abcc.ncifcrf.gov/) to perform GO classification.

Statistical analysis

Statistical calculations were performed using Prism5 (GraphPad Software, La Jolla, CA, USA), Microsoft Excel (Microsoft, Redmond, WA, USA), and SPSS statistical software, ver. 13.0.1 (SPSSInc., Chicago, IL). The Kolmogorov–Smirnov test was used to test normal distribution of data sets. If data sets were subjected to a normal distribution, unpaired or paired t-test was used for comparing between two groups, one-way ANOVA and *post hoc* Bonferroni t-test for multiple comparisons. Mann–Whitney U test was used to compare the difference between two independent groups with a non-parametric data distribution. The Wilcoxon signed-rank test was used to compare two related samples with a non-parametric data distribution. Kruskal-Wallis test with *posthoc* Dunn´s test was applied for multiple comparisons among groups with a non-parametric data distribution. The correlation between EpCAM expression and EMT score was analyzed using Spearman's rank test. Receiver operating characteristic (ROC) curves were constructed to calculate the area under the curve (AUC) for the capability of the EpCAM positivity rate of cells to predict patients' outcome. The Kaplan–Meier method was used to analyze patients' overall survival. Bars and error bars in the histograms represent mean values ± standard deviation (s.d.) of at least three independent experiments. P-values < 0.05 were considered significant.

References and Notes

1. C. E. DeSantis, J. Ma, A. Goding Sauer, L. A. Newman, A. Jemal, Breast cancer statistics, 2017, racial disparity in mortality by state. *CA Cancer J Clin* **67**, 439-448 (2017).
2. B. Lim, G. N. Hortobagyi, Current challenges of metastatic breast cancer. *Cancer Metastasis Rev* **35**, 495-514 (2016).
3. Y. Kang, K. Pantel, Tumor cell dissemination: emerging biological insights from animal models and cancer patients. *Cancer Cell* **23**, 573-581 (2013).
4. K. Pantel, C. Alix-Panabieres, S. Riethdorf, Cancer micrometastases. *Nat Rev Clin Oncol* **6**, 339-351 (2009).
5. A. W. Lambert, D. R. Pattabiraman, R. A. Weinberg, Emerging Biological Principles of Metastasis. *Cell* **168**, 670-691 (2017).
6. S. A. Joosse, T. M. Gorges, K. Pantel, Biology, detection, and clinical implications of circulating tumor cells. *EMBO molecular medicine* **7**, 1-11 (2015).
7. C. L. Chaffer, R. A. Weinberg, A perspective on cancer cell metastasis. *Science* **331**, 1559-1564 (2011).
8. W. Janni *et al.*, Pooled Analysis of the Prognostic Relevance of Circulating Tumor Cells in Primary Breast Cancer. *Clin Cancer Res*, (2016).
9. M. Cristofanilli, Circulating tumor cells, disease progression, and survival in metastatic breast cancer. *Semin Oncol* **33**, S9-14 (2006).
10. W. Janni *et al.*, Persistence of disseminated tumor cells in the bone marrow of breast cancer patients predicts increased risk for relapse--a European pooled analysis. *Clin Cancer Res* **17**, 2967-2976 (2011).
11. V. Muller *et al.*, Prognostic impact of circulating tumor cells assessed with the CellSearch System and AdnaTest Breast in metastatic breast cancer patients: the DETECT study. *Breast Cancer Res* **14**, R118 (2012).
12. K. Pantel, C. Alix-Panabieres, Circulating tumour cells in cancer patients: challenges and perspectives. *Trends Mol Med* **16**, 398-406 (2010).
13. I. Baccelli *et al.*, Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat Biotechnol* **31**, 539-544 (2013).
14. T. Brabletz, R. Kalluri, M. A. Nieto, R. A. Weinberg, EMT in cancer. *Nat Rev Cancer*, (2018).
15. J. P. Thiery, C. T. Lim, Tumor dissemination: an EMT affair. *Cancer Cell* **23**, 272-273 (2013).
16. X. Ye *et al.*, Upholding a role for EMT in breast cancer metastasis. *Nature* **547**, E1-E3 (2017).
17. I. Pastushenko *et al.*, Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463-468 (2018).
18. E. W. Thompson, S. H. Nagaraj, Transition states that allow cancer to spread. *Nature* **556**, 442-444 (2018).
19. Y. Chao, Q. Wu, M. Acquafondata, R. Dhir, A. Wells, Partial mesenchymal to epithelial reverting transition in breast and prostate cancer metastases. *Cancer Microenviron* **5**, 19-28 (2012).
20. M. Korpal *et al.*, Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nat Med* **17**, 1101-1108 (2011).
21. E. D. Hay, An overview of epithelio-mesenchymal transformation. *Acta Anat (Basel)* **154**, 8-20 (1995).
22. K. R. Fischer *et al.*, Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472-476 (2015).

23. X. Zheng *et al.*, Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525-530 (2015).
24. N. M. Aiello *et al.*, Upholding a role for EMT in pancreatic cancer metastasis. *Nature* **547**, E7-E8 (2017).
25. H. Wang, N. H. Stoecklein, P. P. Lin, O. Gires, Circulating and disseminated tumor cells: diagnostic tools and therapeutic targets in motion. *Oncotarget* **8**, 1884-1912 (2017).
26. M. Yu *et al.*, Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**, 580-584 (2013).
27. N. Bednarz-Knoll, C. Alix-Panabieres, K. Pantel, Plasticity of disseminating cancer cells in patients with epithelial malignancies. *Cancer Metastasis Rev* **31**, 673-687 (2012).
28. C. J. Aslakson, F. R. Miller, Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res* **52**, 1399-1405 (1992).
29. B. E. Miller, F. R. Miller, G. H. Heppner, Interactions between tumor subpopulations affecting their sensitivity to the antineoplastic agents cyclophosphamide and methotrexate. *Cancer Res* **41**, 4378-4381 (1981).
30. T. Santarius, J. Shipley, D. Brewer, M. R. Stratton, C. S. Cooper, A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* **10**, 59-64 (2010).
31. P. P. Lin, O. Gires, D. D. Wang, L. Li, H. Wang, Comprehensive in situ co-detection of aneuploid circulating endothelial and tumor cells. *Sci Rep* **7**, 9789 (2017).
32. O. Gires, N. H. Stoecklein, Dynamic EpCAM expression on circulating and disseminating tumor cells: causes and consequences. *Cell Mol Life Sci* **71**, 4393-4402 (2014).
33. T. Celia-Terrassa, Y. Kang, Distinctive properties of metastasis-initiating cells. *Genes Dev* **30**, 892-908 (2016).
34. M. Pan *et al.*, EpCAM ectodomain EpEX is a ligand of EGFR that counteracts EGF-mediated epithelial-mesenchymal transition through modulation of phospho-ERK1/2 in head and neck cancers. *PLoS Biol* **16**, e2006624 (2018).
35. N. V. Jordan, G. L. Johnson, A. N. Abell, Tracking the intermediate stages of epithelial-mesenchymal transition in epithelial stem cells and cancer. *Cell Cycle* **10**, 2865-2873 (2011).
36. R. Y. Huang *et al.*, An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). *Cell Death Dis* **4**, e915 (2013).
37. S. S. Sikandar *et al.*, Role of epithelial to mesenchymal transition associated genes in mammary gland regeneration and breast tumorigenesis. *Nature communications* **8**, 1669 (2017).
38. T. Z. Tan *et al.*, Epithelial-mesenchymal transition spectrum quantification and its efficacy in deciphering survival and drug responses of cancer patients. *EMBO molecular medicine* **6**, 1279-1293 (2014).
39. S. de Wit *et al.*, EpCAM(high) and EpCAM(low) circulating tumor cells in metastatic prostate and breast cancer patients. *Oncotarget* **9**, 35705-35716 (2018).
40. N. Aceto *et al.*, Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110-1122 (2014).
41. B. Rack *et al.*, Circulating tumor cells predict survival in early average-to-high risk breast cancer patients. *J Natl Cancer Inst* **106**, (2014).
42. H. Zitzelsberger *et al.*, Cytogenetic changes in radiation-induced tumors of the thyroid. *Cancer Res* **59**, 135-140 (1999).

Acknowledgments

Funding: Parts of the work have been supported by grants GI 540/3-1 and GI 540/3-2 from the Deutsche Forschungsgemeinschaft to OG, by the international doctoral program "i-Target: Immunotargeting of cancer" funded by the Elite Network of Bavaria (to SK), the Marie-Sklodowska-Curie "Training Network for the Immunotherapy of Cancer (IMMUTRAIN)" funded by the H2020 program of the European Union (to SK) and the European Research Council Starting Grant (grant number 756017 to SK), by grant 81772802 from the China National Natural Science Funds and 16ZR1427400 from the Shanghai Science and Technology Innovation Action Plan to HW.

Author contributions:

XL: performed and analyzed all *in vitro* experiments with the help of ZH, performed *in vivo* animal experiments together with BC, AM, CV, MP, YH, SN, DB, and JSG

JL: conducted the clinical study under supervision of HW, analyzed SE-iFISH results

XL, BC, AM, CV, ZH, MP, YH, SN, DB, JSG: performed all *in vivo* animal experiments

PPL, DW: performed SE-iFISH

CAR and BU: helped with adhesion assays

JD, SZ: performed the sequence data analysis of DNA-sequencing

GK, DL: performed immunohistochemistry and cytochemistry staining

HZ, IZ, HB: performed karyotyping, gene extraction and GO-term analysis of 4T1 cells and sublines

SK: supervised the animal experiments, helped writing the manuscript

SK, OG, HW: supervised the study

OG: generated figures and wrote the manuscript with the help of XL, SK and HW, and support of all other authors.

HW: supervised the clinical and DNA-sequencing studies, helped writing the manuscript.

Competing interests: All others declare no conflicts of interest, except Peter P. Lin. PPL has a financial, management/advisory relationship as President of Cytelligen, San Diego, US.

Data and materials availability:

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

Figures Legends

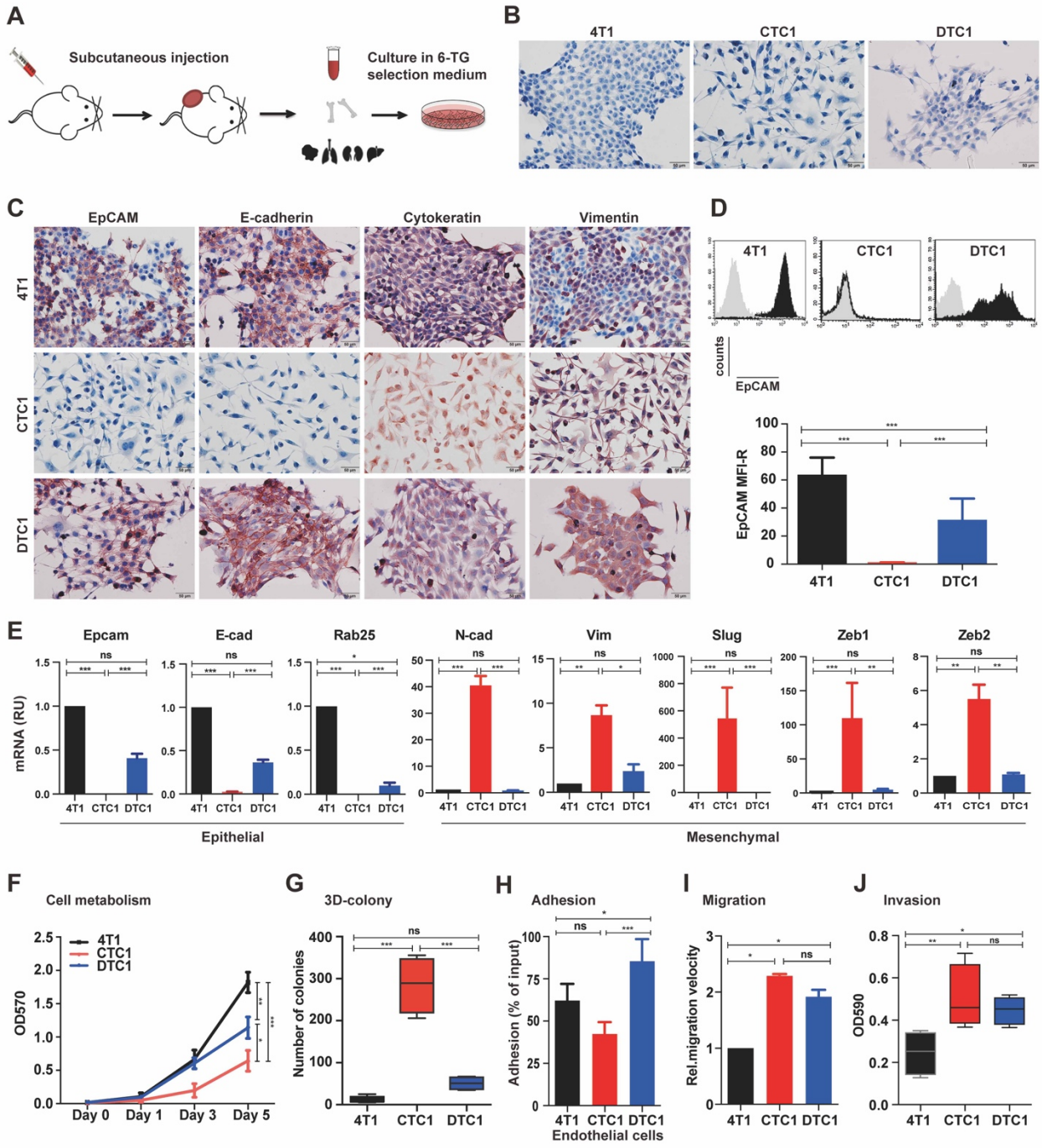


Figure 1: EMT phenotype and *in vitro* functional characteristics of systemic cancer cells in the 4T1 metastatic breast cancer mouse model

(**A**) Schematic representation of the experimental set-up. 4T1 cells were subcutaneously transplanted in the flank of BALB/c mice. After 2 – 4 weeks, mice were sacrificed and primary tumor, blood, bone, and organs were harvested for further culture. 4T1 tumor sublines were recovered in 6-thioguanin (6TG)-containing selection medium.

(**B**) Morphology of 4T1, CTC1, and DTC1 lines after syngeneic transplantation and recovery from the blood (CTC1) and bone marrow (DTC1). Shown are representative pictures of each cell line.

(**C**). Immunohistochemistry staining of EpCAM, E-cadherin, cytokeratin, and vimentin in 4T1, CTC1, and DTC1. Shown are representative pictures from n = 3 independent experiments.

(**D**) EpCAM expression in 4T1, CTC1, and DTC1 was determined by flow cytometry. Upper panel shows representative histograms with EpCAM staining in black and controls in grey. Quantification of EpCAM expression on 4T1, CTC1, and DTC1 is presented as the mean fluorescence intensity ratio (MFI-R with SD) from n ≥ 5 independent experiments performed in unicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; *** < 0.001.

(**E**) mRNA transcript levels of epithelial markers Epcam, E-cadherin and Rab25, and of EMT markers N-cadherin, vimentin, Slug, and Zeb1/2 in 4T1, CTC1, and DTC1 were assessed upon qRT–PCR with specific primers and GUSP as a house-keeping gene. Shown are mean with SD from n = 3 independent experiments performed in triplicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, * < 0.05, ** < 0.01, *** < 0.001.

(**F**) Cell metabolism of 4T1, CTC1, DTC1 was assessed by MTT assay (initial cell number 1000 cells). Shown are mean with SD from n ≥ 3 independent experiments performed in triplicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; * < 0.05, *** < 0.001.

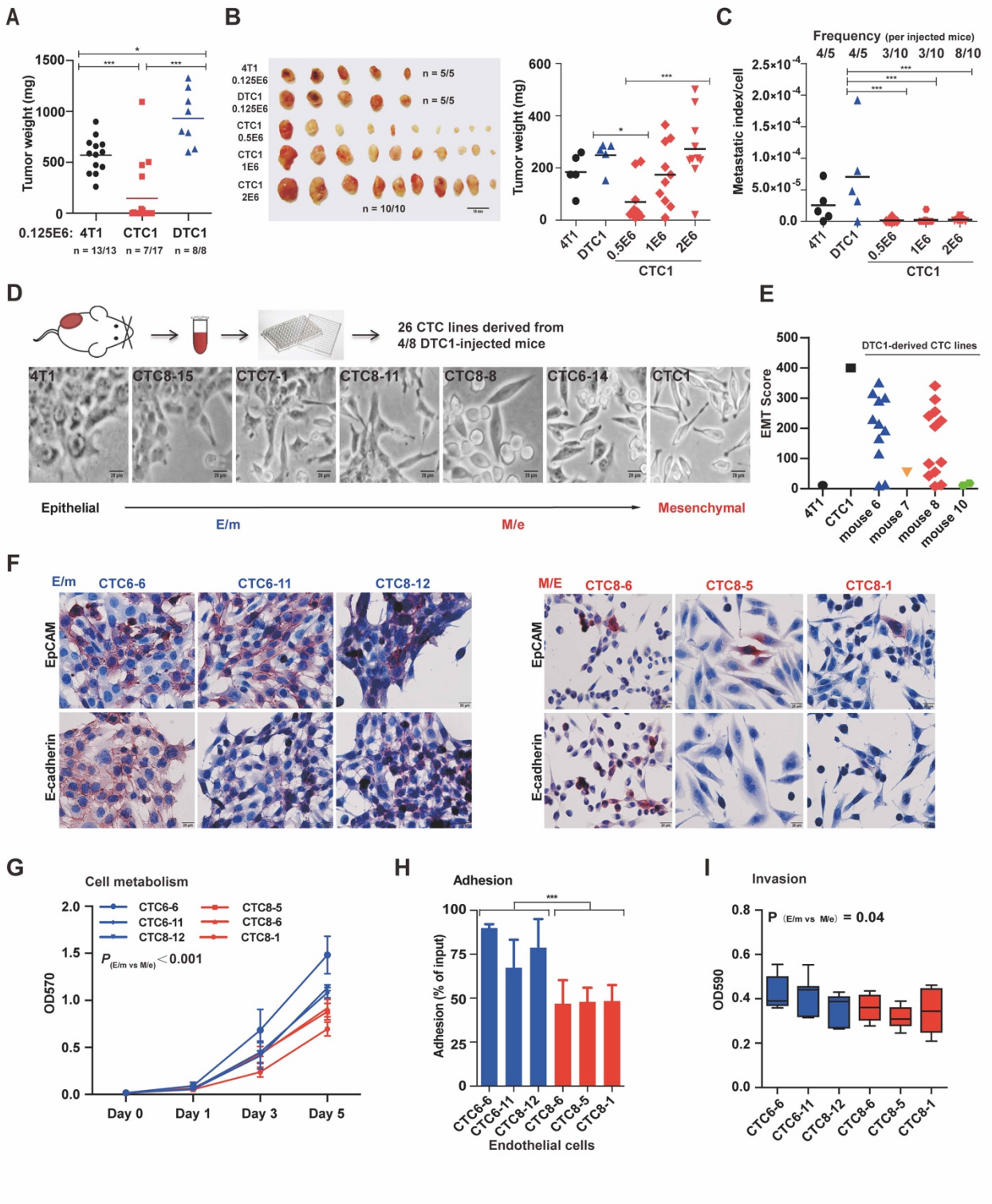
(**G**) 3D colony formation assay was performed with 4T1, CTC1, and DTC1 cells. Numbers of colonies are shown as box-plot whiskers graph with mean from n = 4 independent experiments

performed in unicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, *** < 0.001.

(**H**) Adhesion of 4T1, CTC1, and DTC1 cells to bEnd.3 endothelial cells was assessed. Shown are mean adhesion rate with SD from n ≥ 3 independent experiments performed in triplicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, * < 0.05, *** < 0.001.

(**I**) Migration capacity of 4T1, CTC1, and DTC1 was assessed in a scratch assay. Migration velocity is given as mean µm/h with SD from n = 3 independent experiments performed in unicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, * <0.05.

(**J**) Invasion capacity of 4T1, CTC1, DTC1 cells was assessed by transwell invasion assay. Shown are mean OD590 quantifications of invaded cells as box-plot whiskers graph with SD from n = 3 independent experiments performed in duplicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, * < 0.05, ** < 0.01.


A
Tumor weight (mg)
0.125E6: 4T1 CTC1 DTC1
n = 13/13 n = 7/17 n = 8/8
B
4T1 0.125E6
DTC1 0.125E6
CTC1 0.5E6
CTC1 1E6
CTC1 2E6
n = 5/5
n = 5/5
n = 10/10
Tumor weight (mg)
4T1 DTC1 0.5E6 1E6 2E6
CTC1
C
Frequency (per injected mice)
4/5 4/5 3/10 3/10 8/10
Metastatic index/cell
4T1 DTC1 0.5E6 1E6 2E6
CTC1
D
26 CTC lines derived from
4/8 DTC1-injected mice
4T1 CTC8-15 CTC7-1 CTC8-11 CTC8-8 CTC6-14 CTC1
Epithelial
Mesenchymal
E/m
M/e
E
DTC1-derived CTC lines
EMT Score
4T1 CTC1 mouse 6 mouse 7 mouse 8 mouse 10
F
E/m CTC6-6 CTC6-11 CTC8-12
M/E CTC8-6 CTC8-5 CTC8-1
EpCAM
E-cadherin
G
Cell metabolism
CTC6-6 CTC6-11 CTC8-12 CTC8-5 CTC8-6 CTC8-1
P (E/m vs M/e) < 0.001
OD570
Day 0 Day 1 Day 3 Day 5
H
Adhesion
Adhesion (% of input)
CTC6-6 CTC6-11 CTC8-12 CTC8-6 CTC8-5 CTC8-1
Endothelial cells
I
Invasion
P (E/m vs M/e) = 0.04
OD590
CTC6-6 CTC6-11 CTC8-12 CTC8-6 CTC8-5 CTC8-1

)54

Figure 2: *In vivo* tumorigenicity of 4T1, CTC1, and DTC1 cells, and EMT traits of DTC1-derived CTC lines

(**A**) 4T1, CTC1, and DTC1 ($1.25*10^5$ cells) were transplanted subcutaneously into BALB/c mice. Dot plot shows individual and mean tumor weights for each group at the end of the experiment including numbers of transplanted mice. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; * < 0.05, *** < 0.001.

(**B**) 4T1 ($1.25*10^5$ cells; n = 5), CTC1 ($5*10^5$ cells, 10^6, $2*10^6$ cells, each n = 10) and DTC1 ($1.25*10^5$ cells; n = 5) were transplanted subcutaneously into BALB/c mice. Shown are pictures of primary tumors. Dot plot shows individual and mean tumor weights for each group at the end of the experiment. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; * < 0.05, *** < 0.001.

(**C**) Metastatic index/cell was calculated as numbers of lung metastatic colony divided by initially injected cell numbers. Dot plot shows metastatic index/cell and frequencies of lung metastasis per mouse. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; *** < 0.001.

(**D**) Schematic representation of the establishment of CTC sublines from DTC1-transplanted mice. Shown are representative pictures of CTC1 and 4T1, and DTC1-derived CTC displaying various degrees of EMT.

(**E**) Dot plot shows mean EMT score grouped by mouse from n = 3 independent scoring results. 4T1 (epithelial, score 0) and CTC1 (mesenchymal, score 400) are included as controls.

(**F**) Immunohistochemistry staining of EpCAM and E-cadherin in E/m-type (CTC6-6, CTC6-11, CTC8-12) and M/e-type (CTC8-6, CTC8-5, CTC8-1) CTC-derived from DTC1 transplantations. Shown are representative pictures from n = 3 independent staining.

(**G**) Cell metabolism of E/m-type (CTC6-6, CTC6-11, CTC8-12) and M/e-type (CTC8-6, CTC8-5, CTC8-1) CTC was assessed by MTT assay (initial cell number 1000 cells). Shown are mean with

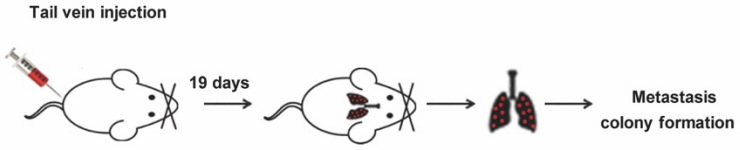
SD from n ≥ 3 independent experiments performed in triplicates. T-test E/m versus M/e cells is indicated.

(**H**) Adhesion assay to bEnd.3 endothelial cells was performed with E/m-type (CTC6-6, CTC6-11, CTC8-12) and M/e-type (CTC8-6, CTC8-5, CTC8-1) CTC. Shown are mean adhesion rates with SD from n ≥ 3 independent experiments performed in triplicates. T-test E/m versus M/e cells is indicated; *** < 0.001.

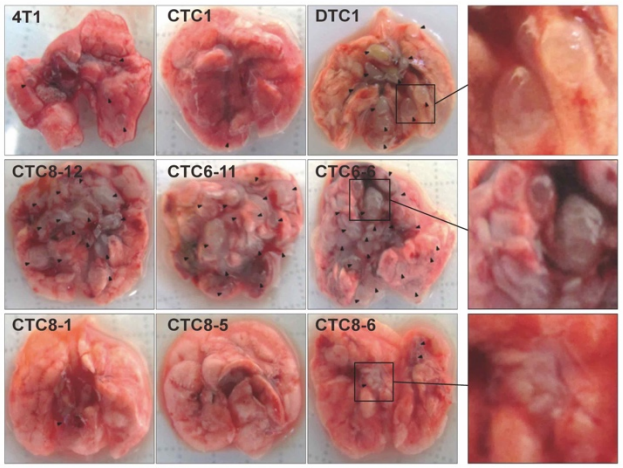
(**I**) Invasion capacity of E/m-type (CTC6-6, CTC6-11, CTC8-12) and M/e-type (CTC8-6, CTC8-5, CTC8-1) CTC was detected by transwell invasion assay. Shown are mean OD 590 quantifications of invaded cells as box-plot whiskers graph with SD from n ≥ 4 independent experiments performed in duplicates. T-test E/m versus M/e cells is indicated.

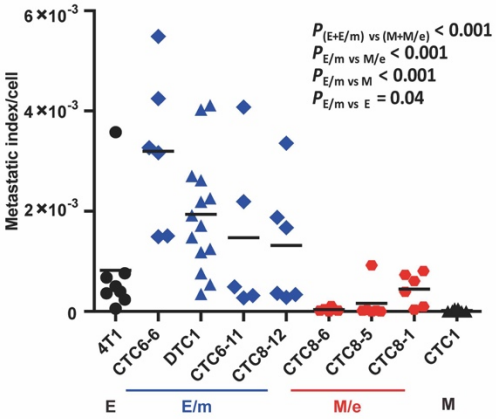
A

Phenotype	Cell line
E	4T1
E/m	CTC6-6 CTC6-11 CTC8-12 DTC1
M/e	CTC8-6 CTC8-5 CTC8-1
M	CTC1



B





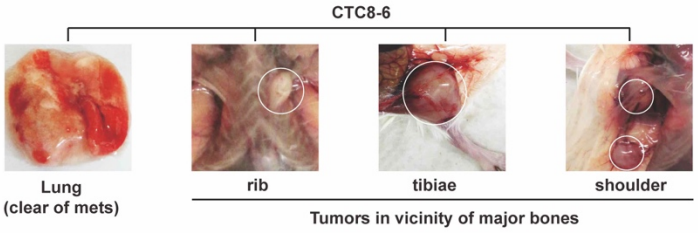
C

Phenotype	Cell line
E/m	CTC6-6
M/e	CTC8-6 CTC8-5 CTC8-1
M	CTC1

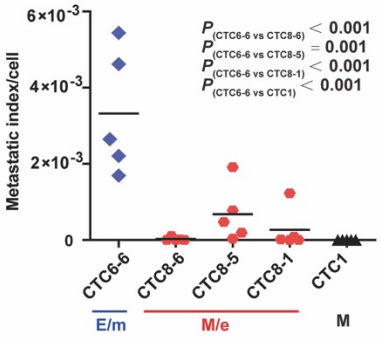
Tail vein injection — Day22 — Day25 — Day28

Group	CTC6-6 One CTC8-1 mouse	CTC8-1 CTC8-5	CTC8-6	CTC1
Sign for endpoint	Significant weight loss, weakness, dyspnea	Significant weight loss, weakness	Significant weight loss, weakness	no signs
Autopsy	Severe lung metastasis (6/6 mice)	2/9 severe lung metastasis, 4/9 metastasis near big bones, 3/9 with multiple tumor sites	0/5 severe lung metastasis, 5/5 metastasis near big bones, 3/5 with multiple tumor sites	2/5 small tumor in tail

D



E



189

Figure 3: Metastasis formation of 4T1, CTC1, DTC1, and CTC sublines of DTC1-transplanted animals

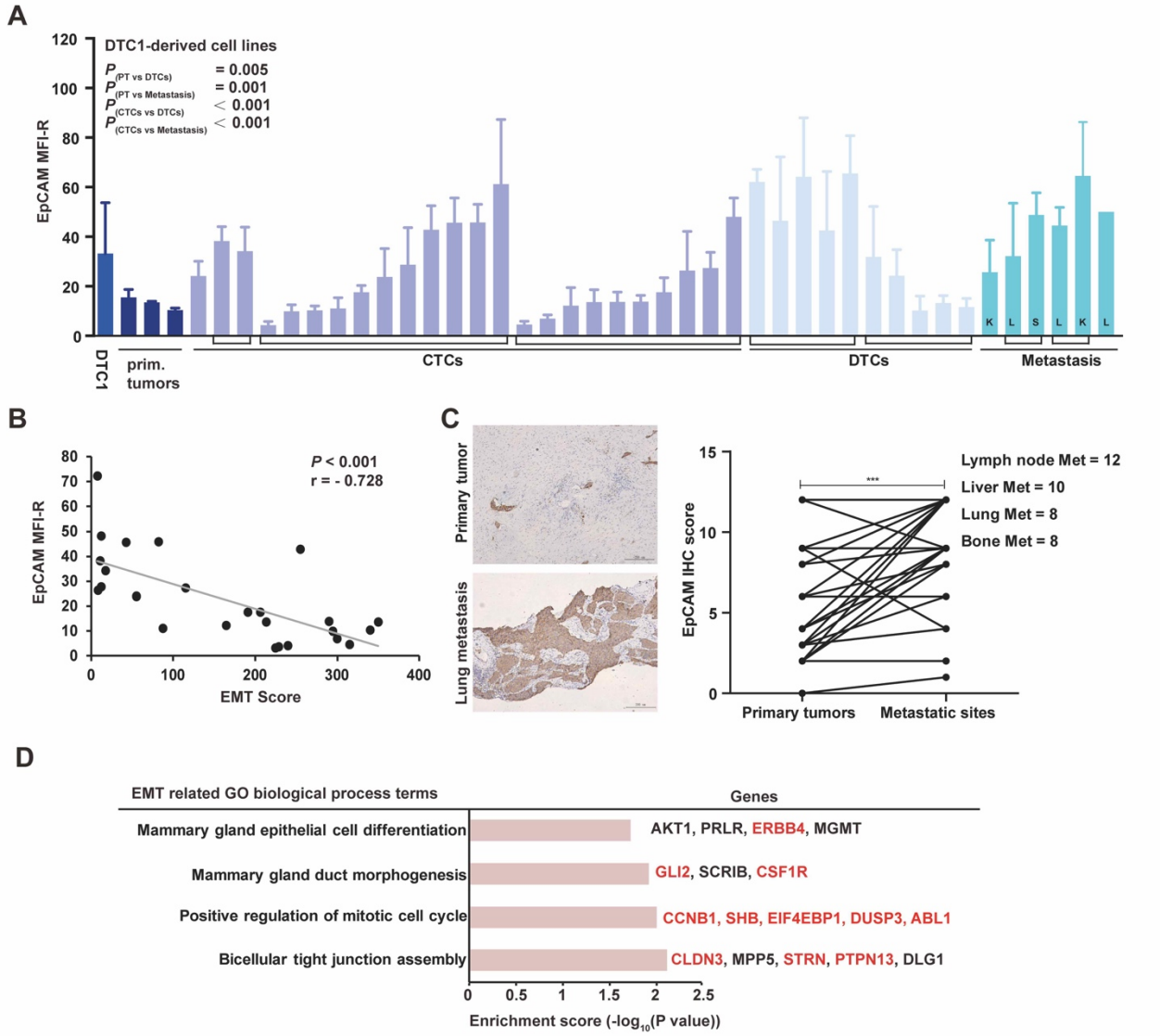
(**A**) Scheme of syngeneic *i.v.* injections: Epithelial E-type (4T1, n = 8 mice), E/m-type (CTC6-6, n = 6 mice; CTC6-11, n = 5 mice; CTC8-12, n = 6 mice; DTC1, n = 14 mice), M/e-type (CTC8-6, n = 6 mice; CTC8-5, n = 6 mice; CTC8-1, n = 6 mice), and mesenchymal M-type (CTC1, n = 8 mice) cells ($5*10^4$) were transplanted into BALB/c mice through tail vein. After 19 days, numbers of superficial lung metastasis were counted and lungs were harvested for further metastasis colony formation assay.

(**B**) Left panel: Representative pictures of lungs following *i. v.* injection of cells bearing different EMT phenotypes. Black dots indicate metastatic lung nodules. Right panel: Dot plot shows metastatic index/cell including mean (line) and p-values for each cell line after *i. v.* injection. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction*, p-value is indicated.

(**C**) Scheme of syngeneic *i.v.* injections: E/m-type (CTC6-6), M/e-type (CTC8-6, CTC8-5, CTC8-1), and mesenchymal M-type (CTC1) cells were transplanted into BALB/c mice through tail vein injection. Each experimental group was ended at the day of the indicated signs for endpoint. Results from autopsy at the given time points are described.

(**D**) Pictures of autopsy results from CTC8-6 injected mice displaying the lack of lung metastasis, and the presence of tumors in the vicinity of rib, tibiae, and shoulder blade.

(**E**) Dot plot shows metastatic index/cell of E/m-type (CTC6-6), M/e-type (CTC8-6, CTC8-5, CTC8-1), and mesenchymal M-type (CTC1) cells including mean values (line). One-way ANOVA with *posthoc* multiple testing and Bonferroni correction*, p-value is indicated.


A
DTC1-derived cell lines
P(PT vs DTCs) = 0.005
P(PT vs Metastasis) = 0.001
P(CTCs vs DTCs) < 0.001
P(CTCs vs Metastasis) < 0.001
EpCAM MFI-R
DTC1
prim. tumors
CTCs
DTCs
Metastasis
K L S L K L
B
P < 0.001
r = - 0.728
EpCAM MFI-R
EMT Score
C
Primary tumor
Lung metastasis
EpCAM IHC score
***
Lymph node Met = 12
Liver Met = 10
Lung Met = 8
Bone Met = 8
Primary tumors
Metastatic sites
D
EMT related GO biological process terms
Genes
Mammary gland epithelial cell differentiation
AKT1, PRLR, ERBB4, MGMT
Mammary gland duct morphogenesis
GLI2, SCRIB, CSF1R
Positive regulation of mitotic cell cycle
CCNB1, SHB, EIF4EBP1, DUSP3, ABL1
Bicellular tight junction assembly
CLDN3, MPP5, STRN, PTPN13, DLG1
Enrichment score (-log10(P value))

11

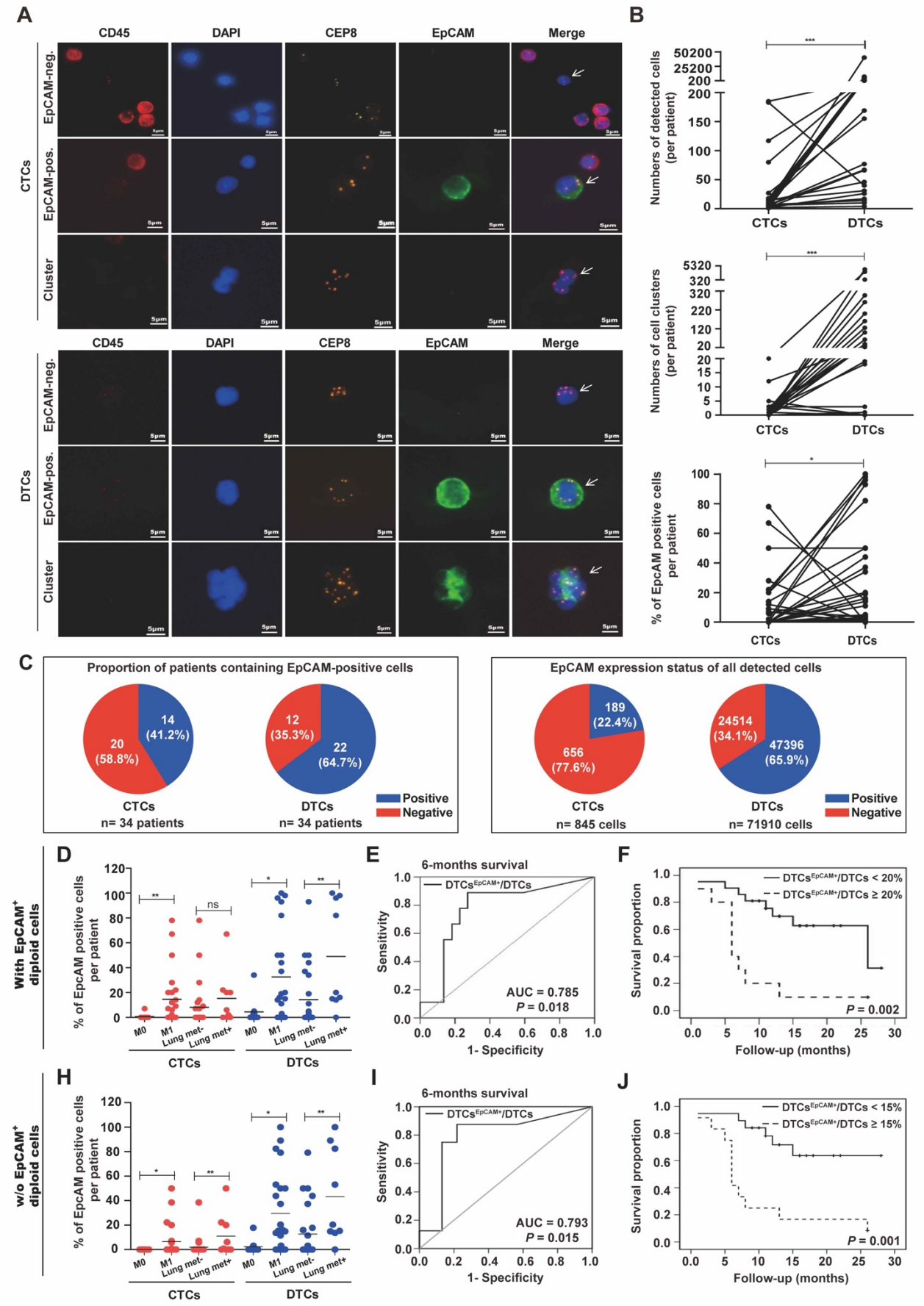
Figure 4: EpCAM expression profiles correlate with an epithelial phenotype of experimental cells and human CTC

(**A**) EpCAM expression was measured by flow cytometry in cell lines from primary tumors, CTC, DTC, and metastases isolated from DTC1-injected BALB/c mice. Brackets demark cell lines originating from one individual mouse. L: lung, S: spleen, K: kidney. Data is presented as mean fluorescence intensity ratios (EpCAM/iso) with SD from $n \geq 3$ independent experiments performed in unicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction*, p-value is indicated.

(**B**) Cluster plot analysis of Spearman's rank correlation between EpCAM expression and EMT score in DTC1-derived CTC sublines. Correlation coefficient (r) and p-value are included.

(**C**) EpCAM expression level was assessed in n = 38 human breast tumors and associated lymph node metastases (n = 12), liver metastases (n = 10), lung metastases (n = 8), and bone metastases (n = 8). Shown are representative immunohistochemistry (IHC) staining of EpCAM in primary tumors and corresponding lung metastasis, and quantifications of IHC scores for all samples as paired samples (see Materials and Methods). Paired T-test, *** < 0.001.

(**D**) Enrichment analysis of GO biological process terms of genes extracted from CNVs of $EpCAM^+$ (n = 10) versus $EpCAM^-$ CTC (n = 20) from patients suffering from MBC (n = 3). GO terms related to epithelial differentiation are listed. Enrichment score was -log10(p-value) of more than 1.3 was considered significant. Gene names in red font: amplified in $EpCAM^+$ CTC; gene names in black font: mutated in $EpCAM^+$ or $EpCAM^-$ CTC.


A
CD45
DAPI
CEP8
EpCAM
Merge
CTCs
DTCs
EpCAM-neg.
EpCAM-pos.
Cluster
B
Numbers of detected cells (per patient)
Numbers of cell clusters (per patient)
% of EpcAM positive cells per patient
CTCs
DTCs
C
Proportion of patients containing EpCAM-positive cells
14 (41.2%)
20 (58.8%)
12 (35.3%)
22 (64.7%)
CTCs n= 34 patients
DTCs n= 34 patients
Positive
Negative
EpCAM expression status of all detected cells
189 (22.4%)
656 (77.6%)
24514 (34.1%)
47396 (65.9%)
CTCs n= 845 cells
DTCs n= 71910 cells
D
With EpCAM+ diploid cells
% of EpcAM positive cells per patient
M0
M1
Lung met-
Lung met+
E
6-months survival
DTCsEpCAM+/DTCs
Sensitivity
1- Specificity
AUC = 0.785
P = 0.018
F
DTCsEpCAM+/DTCs < 20%
DTCsEpCAM+/DTCs ≥ 20%
Survival proportion
Follow-up (months)
P = 0.002
H
w/o EpCAM+ diploid cells
I
AUC = 0.793
P = 0.015
J
DTCsEpCAM+/DTCs < 15%
DTCsEpCAM+/DTCs ≥ 15%
P = 0.001

32

Figure 5: Proportions of EpCAM-positive systemic tumor cells correlate with the clinical outcome of MBC patients

(**A**) CTC and DTC were detected by SE-iFISH in pairs of blood and bone marrow samples from n = 34 breast cancer patients**.** Shown are representative examples of aneuploid CTC, DTC, and cell clusters identified by iFISH (white arrows). DAPI, immunofluorescence (CD45 - red, DAPI - blue, EpCAM - green), and FISH (CEP8 - orange) are depicted as indicated. Shown are representative staining for each antigen.

(**B**) Shown are numbers of detected CTC and DTC (upper panel), numbers of cell clusters of CTC and DTC (middle panel), and percentages of $EpCAM^+$ CTC and DTC in each patient (lower panel). Paired CTC and DTC values are connected by solid lines. Wilcoxon signed-rank test, * < 0.05, *** < 0.001.

(**C**) Shown are proportions of $EpCAM^+$ CTC and DTC in individual patients (n = 34) (left pie charts) and in overall numbers of CTC (n = 845) and DTC cells (n = 71,910) (right pie charts). Blue: $EpCAM^+$, red: $EpCAM^-$.

(**D and H**) Dot-plots depict percentages of $EpCAM^+$ CTC and DTC per patient (n=34) stratified for distant metastases status M_0 and M_1, and for the absence or presence of lung metastases. (D) Includes $EpCAM^+$ diploid cells, whereas (H) excludes $EpCAM^+$ diploid cells, as indicated. Mean percentage of EpCAM expression is indicated by a horizontal line. Mann-Whitney U test, ns: not significant, * < 0.05, ** < 0.01.

(**E and I**) Specificity and sensitivity of the EpCAM-positivity rate of DTC to predict 6 months survival rate is depicted as receiver operating characteristic (ROC) curve. Area under the curve (AUC) and p-value are indicated. (Sensitivity = true positives; 1-specificity = false positives). (E) Includes $EpCAM^+$ diploid cells, whereas (I) excludes $EpCAM^+$ diploid cells.

(**F and I**) Overall survival of patients with stage III-IV MBC (n = 34) was stratified according to the presence of $EpCAM^+$ DTC in bone marrow with a cut-off of 20 % (F) and 15% (I) (deduced

from ROC analysis), and is depicted as Kaplan-Meier survival curves with p-value. (F) Includes $EpCAM^+$ diploid cells, whereas (J) excludes $EpCAM^+$ diploid cells, as indicated.

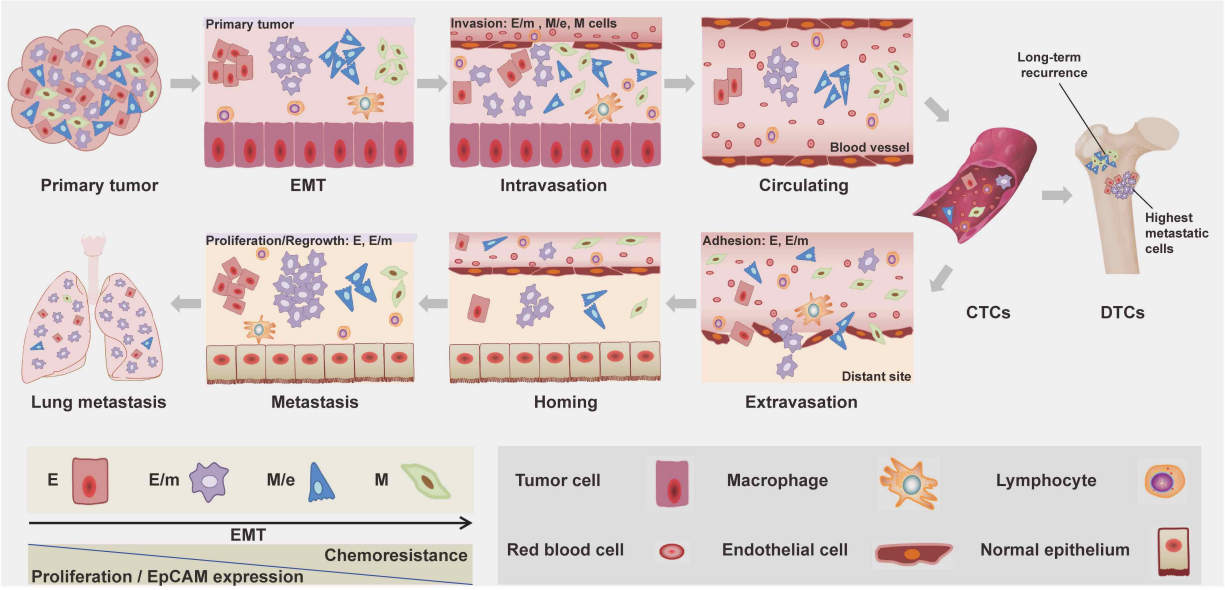


Figure 6: Schematic representation of EMT during the metastatic cascade

Tumor cells can undergo gradual or full EMT (E, E/m, M/e, M) that is associated with decreased proliferation, loss of EpCAM expression, and increased migration. In the metastatic cascade, M > M/e > E/m cells have increased ability to intravasate into the lymphovascular system. Once tumor cells intravasate into blood vessels, they are termed CTC. E, E/m CTC have enhanced capacity to adhere, hence to extravasate into distant site. After homing into distant organs, including the bone marrow, systemic tumor cells are termed DTC. E, E/m systemic tumor cells are associated with improved capacity in proliferation and metastatic outgrowth, M/e, M cells are related to long-term tumor recurrence.

Supplementary materials for:

Epithelial-type Systemic Breast Carcinoma Cells with a Restricted Mesenchymal Transition are a Major Source of Metastasis

Authors: *Xiao Liu[1†], Junjian Li[2†], Bruno Cadilha[3], Anamarija Markota[3], Cornelia Voigt[3], Zhe Huang[1], Peter P. Lin[4], Daisy D. Wang[4], Juncheng Dai[5], Gisela Kranz[1], Anna Krandick[1], Darko Libl[1], Horst Zitzelsberger[6,7], Isabella Zagorski[7], Herbert Braselmann[7], Min Pan[1], Sibo Zhu[8], Yuanchi Huang[1], Sebastian Niedermeyer[1], Christoph A. Reichel[1], Bernd Uhl[1], Daria Briukhovetska[3], Javier Suárez Gosálvez[3], Sebastian Kobold[3‡], Olivier Gires[1,6‡*], and Hongxia Wang[2‡*]*

[†] These authors have equally contributed

[‡] These senior authors have equally contributed

To whom correspondence should be addressed: Olivier Gires Olivier.gires@med.uni-muenchen.de (OG) or Hongxia Wang whx365@126.com (HW)

Keywords: Metastatic breast cancer; Epithelial-Mesenchymal-Transition; Circulating tumor cells; Disseminated tumor cells; EpCAM

List of Supplementary materials:

Figures S1: Immunohistochemistry of EpCAM in 4T1 cells; Karyograms of 4T1 cells and sublines; CD45 flow cytometry analysis of 4T1 cells and sublines; Cell proliferation in 6-thioguanine selection medium; Morphology of 4T1 cells under 6-thioguanine selection.

Figure S2: Quantitative RT-PCR analysis of EMT markers in 4T1, CTC1, and DTC1 cells; Growth curves of 4T1, CTC1, and DTC1 cells; 2D and 3D colony formation assays with 4T1, CTC1, and DTC1 cells; Adhesion of 4T1, CTC1, and DTC1 cells to matrigel and gelatin; Representative pictures of scratch assays; Representative pictures of invasion assays.

Figure S3: Tumor growth curves for 4T1, CTC1, and DTC1 cells *in vivo*; EpCAM and vimentin expression in 4T1-, CTC1-, and DTC1-derived tumor; Average tumor size for 4T1, CTC1, and DTC1 cells; Metastatic colony numbers for 4T1, CTC1, and DTC1 cells; Scheme of *s.c.* transplantation and resulting *ex vivo* cultured sublines; Analysis of genes and GO-terms from chromosomal breakpoints differentially represented in 4T1 versus DTC1 cells.

Figure S4: EMT scores of 4T1, CTC1, and DTC1-derived CTC lines; Vimentin expression in 4T1, CTC1, and DTC1-derived CTC lines; Quantitative RT-PCR analysis of EMT markers in 4T1 and DTC1-derived CTC sublines.

Figure S5: Cell proliferation of E/m- and M/e-type DTC1-derived CTC sublines; Sensitivity to cisplatin and doxorubicin of E/m- and M/e-type DTC1-derived CTC sublines; Numbers of lung metastases following *i. v.* injection of 4T1, CTC1, and E/m- and M/e-type DTC1-derived CTC sublines; Numbers of lung metastases colonies following *i. v.* injection of 4T1, CTC1, and E/m- and M/e-type DTC1-derived CTC sublines; Mouse weight following *i. v.* injection of CTC1, and E/m- and M/e-type DTC1-derived CTC sublines

Figure S6: Morphology, EMT scores, and EpCAM mean fluorescence intensity ratios with variances and ranges of 4T1-, CTC1-, and DTC1-derived single cell clones, and DTC1-derived CTC sublines.

Figure S7: EpCAM expression in 4T1- and CTC1-derived primary tumors and metastases; EpCAM expression with correlated IHC scores in primary tumors, lymph node and distant metastases of clinical samples of MBC patients.

Figure S8: MBC patients´ characteristics; Ploidy and cell size in CTC and DTC from MBC patients (n = 34).

Figure S9: Copy number variations (CNV) in $EpCAM^+$ (n = 10) and $EpCAM^-$ CTC (n = 20) from MBC patients (n = 3). Shown are CNV on all chromosomes with numbers of CNV regions, genes, and length of CNV regions. Unsupervised clustering of top 100 CNV in in $EpCAM^+$ and $EpCAM^-$ CTC.

Figure S10: Go-term enrichment analysis of biological processes of CNVs from $EpCAM^+$ and $EpCAM^-$ CTC.

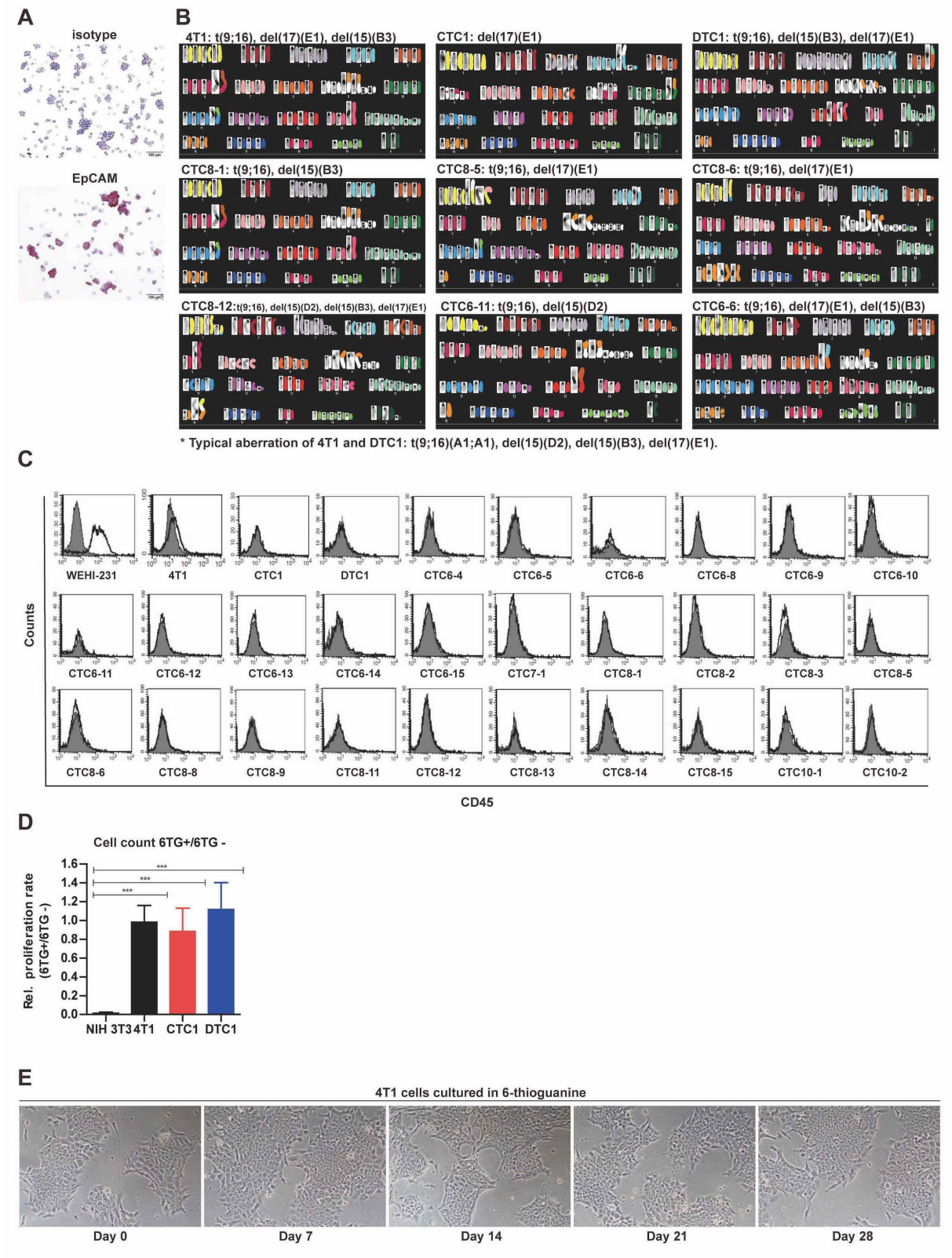


Figure S1:

(**A**) Immunocytochemistry staining of EpCAM on cytospins of 4T1 cells. Shown are representative staining from n = 3 independent experiments.

(**B**) Karyotype analysis of 4T1, CTC1, DTC1, and DTC1-derived CTC sublines CTC8-1, CTC8-5, CTC8-6, CTC8-12, CTC6-11 and CTC6-6. Shown are representative karyotypes including color-coded

chromosomes, chromosome numbers, and marker mutations. Typical aberrations of 4T1 and DTC1 (t(9;16)(A1;A1), del(15)(D2), del(15)(B3), del(17)(E1)) were also found in CTC sublines. del: deletion; der: derived; dmin: double minute chromosomes; mar: marker chromosome; rob: Robertsonian translocation; t: translocation.

(**C**) CD45 expression on 4T1-derived sublines. Expression of leukocyte marker CD45 on the cell surface of 4T1, CTC1, DTC1 and DTC1-derived CTC sublines was measured by flow cytometry with CD45-specific antibodies(black) and isotype controls (grey). Murine B cell lymphoma cell line WEHI-231 was used as a positive control. Shown are representative histograms.

(D) 6-TG resistance of 4T1, CTC1 and DTC1 cell lines. 4T1, CTC1, and DTC1 cell lines and murine NIH3T3 fibroblast cells were plated at equal cell numbers (initial cell number 5000 cells). Relative proliferation rates were calculated at day 5 as cell numbers in 6-thioguanine (6-TG) containing medium divided by cell numbers in medium without 6-TG. NIH 3T3 cells were used as a negative control based on their sensitivity towards 6-TG. Shown are mean ratios with SD of $n \geq 3$ independent experiments performed in duplicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; *** < 0.001.

(E) 4T1 cells were cultured for the indicated 28 days in the presence of 6-TG and cell morphology was assessed at the indicated time points. Shown are representative points of treated cultures.

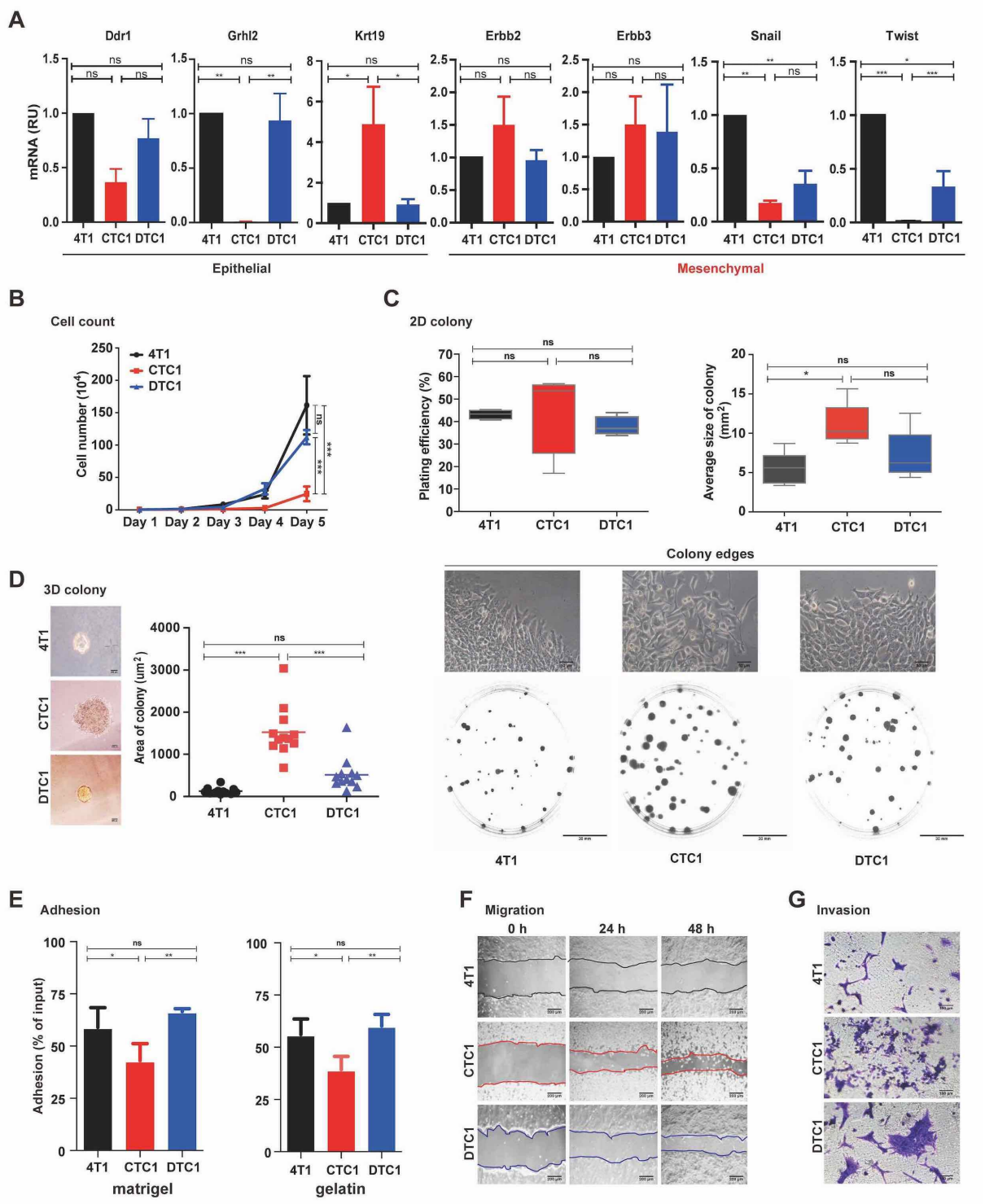


Figure S2:

(**A**) mRNA transcript levels of epithelial markers Ddr1, Grhl2 and Krt19, and of EMT markers Erbb2, Erbb3, Snail and Twist in 4T1, CTC1 and DTC1 were assessed upon qRT–PCR with specific primers and GUSP as a house-keeping gene. Shown are mean with SD from n = 3 independent experiments performed in triplicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, * < 0.05, ** < 0.01, *** < 0.001.

(**B**) Proliferation rate of 4T1, CTC1, DTC1 was assessed by cell counting (initial cell number 5000 cells). Shown are mean with SD from n ≥ 4 independent experiments. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, *** < 0.001.

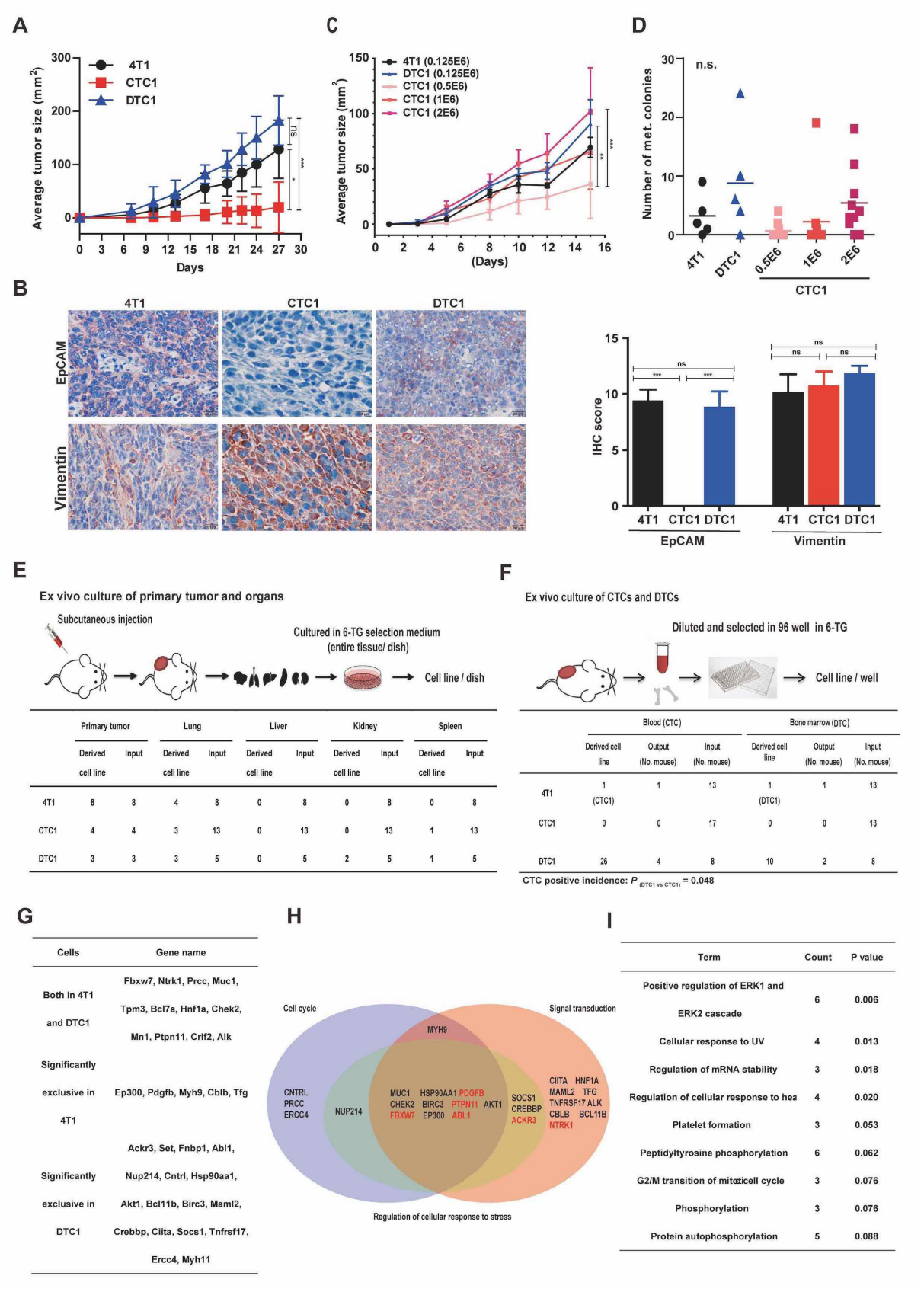
(**C**) 2D colony formation assay was performed with 4T1, CTC1, and DTC1 cells. Left panel: Plating efficiency is shown as box-plot whiskers graph with mean and SD from n = 4 independent experiments performed in unicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant. Lower panels: Representative images of crystal violet-stained colonies from each cell line and 2D colony edges are shown from 4T1, CTC1, and DTC1 cells at an initial seeding density of 200 cells. Right panel: Colony sizes were calculated using Image J software and are represented as box-plot whiskers graph with mean and SD. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, * < 0.05.

(**D**) 3D colony formation assay was performed with 4T1, CTC1, and DTC1 cells. Representative images of colonies from 4T1, CTC1, and DTC1 cells are shown. Quantification of 3D colony size is shown (right panel) as dot plots with mean from n = 12 randomly selected colonies for each cell lines. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, *** < 0.001.

(**E**) Adhesion assay to matrigel and gelatin was performed with 4T1, CTC1, and DTC1 cells. Shown are mean adhesion rate with SD from n ≥ 3 independent experiments performed in triplicates. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, * <0.05, ** < 0.01.

(**F**) Migration capacity of 4T1, CTC1 and DTC1 was assessed in a scratch assay. Representative images of cellular migration were taken at times 0 h, 24 h and 48 h.

(**G**) The invasion capacity of 4T1, CTC1, DTC1 cells was detected by transwell invasion assay. Representative images of invasive cells are shown (n = 3 independent experiments).



	Primary tumor		Lung		Liver		Kidney		Spleen	
	Derived cell line	Input	Derived cell line	Input	Derived cell line	Input	Derived cell line	Input	Derived cell line	Input
4T1	8	8	4	8	0	8	0	8	0	8
CTC1	4	4	3	13	0	13	0	13	1	13
DTC1	3	3	3	5	0	5	2	5	1	5

	Blood (CTC)			Bone marrow (DTC)		
	Derived cell line	Output (No. mouse)	Input (No. mouse)	Derived cell line	Output (No. mouse)	Input (No. mouse)
4T1	1 (CTC1)	1	13	1 (DTC1)	1	13
CTC1	0	0	17	0	0	13
DTC1	26	4	8	10	2	8

CTC positive incidence: $P_{(DTC1\ vs\ CTC1)} = 0.048$

Cells	Gene name
Both in 4T1 and DTC1	Fbxw7, Ntrk1, Prcc, Muc1, Tpm3, Bcl7a, Hnf1a, Chek2, Mn1, Ptpn11, Crlf2, Alk
Significantly exclusive in 4T1	Ep300, Pdgfb, Myh9, Cblb, Tfg
Significantly exclusive in DTC1	Ackr3, Set, Fnbp1, Abl1, Nup214, Cntrl, Hsp90aa1, Akt1, Bcl11b, Birc3, Maml2, Crebbp, Ciita, Socs1, Tnfrsf17, Ercc4, Myh11

Term	Count	P value
Positive regulation of ERK1 and ERK2 cascade	6	0.006
Cellular response to UV	4	0.013
Regulation of mRNA stability	3	0.018
Regulation of cellular response to hea	4	0.020
Platelet formation	3	0.053
Peptidyltyrosine phosphorylation	6	0.062
G2/M transition of mitotic cell cycle	3	0.076
Phosphorylation	3	0.076
Protein autophosphorylation	5	0.088

Figure S3:

(**A**) 4T1, CTC1, and DTC1 ($1.25*10^5$ cells) were transplanted subcutaneously into BALB/c mice. Line charts show tumor growth curves for each group as mean with SD. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, * <0.05, *** < 0.001.

(**B**) Shown are representative immunohistochemistry (IHC) staining of EpCAM and vimentin in primary tumors from 4T1, CTC1 and DTC1 cells injected group (left panel) and quantified IHC scores (see Materials and Methods) across all tumors. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, *** < 0.001.

(**C**) 4T1 and DTC1 cell number in $1.25*10^5$, whereas CTC cell numbers in $0.5*10^6$, 10^6, and $2*10^6$ were transplanted subcutaneously into BALB/c mice. Line charts show tumor growth curves for each group as mean with SD. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction; ns: not significant, ** < 0.01, *** < 0.001.

(**D**) Metastasis colony formation assay was performed as described in Materials and Methods. Dot plots show numbers of colonies including mean (line). No significant difference was detected between groups.

(**E**) *Ex vivo* establishment of primary tumor and metastatic cell lines. Schematic representation of *ex vivo* set-up of cell lines from primary tumors and metastatic sites (lung, spleen, liver and kidney). Table shows frequencies of successfully established cell lines from input.

(**F**) *Ex vivo* establishment of CTCs and DTCs. Schematic representation of *ex vivo* set-up of CTC and DTC lines from blood and bone marrow. Table shows frequencies of successfully established cell lines from input.

(**G**) Cancer genes extracted from genomic regions affected by chromosomal aberrations defined as significantly different between 4T1 and DTC1 cells. Shown are genes from aberrations present in 4T1 and DTC1 with different frequencies and genes from exclusive aberrations.

(**H**) Venn diagram representing 27 potentially affected breakpoint genes that are assigned to three superordinated GO-terms (cell cycle, signal transduction, regulation of response to stress). Six genes marked in red are part of the most important downstream GO-term "*Positive regulation of ERK1 and ERK2 cascade*". The remaining genes are part of other GO-terms.

(**I**) GO-term enrichment analysis of 34 genes listed in (G). GO-terms were referenced to the selected cancer genes. The top 10 GO-terms are depicted.

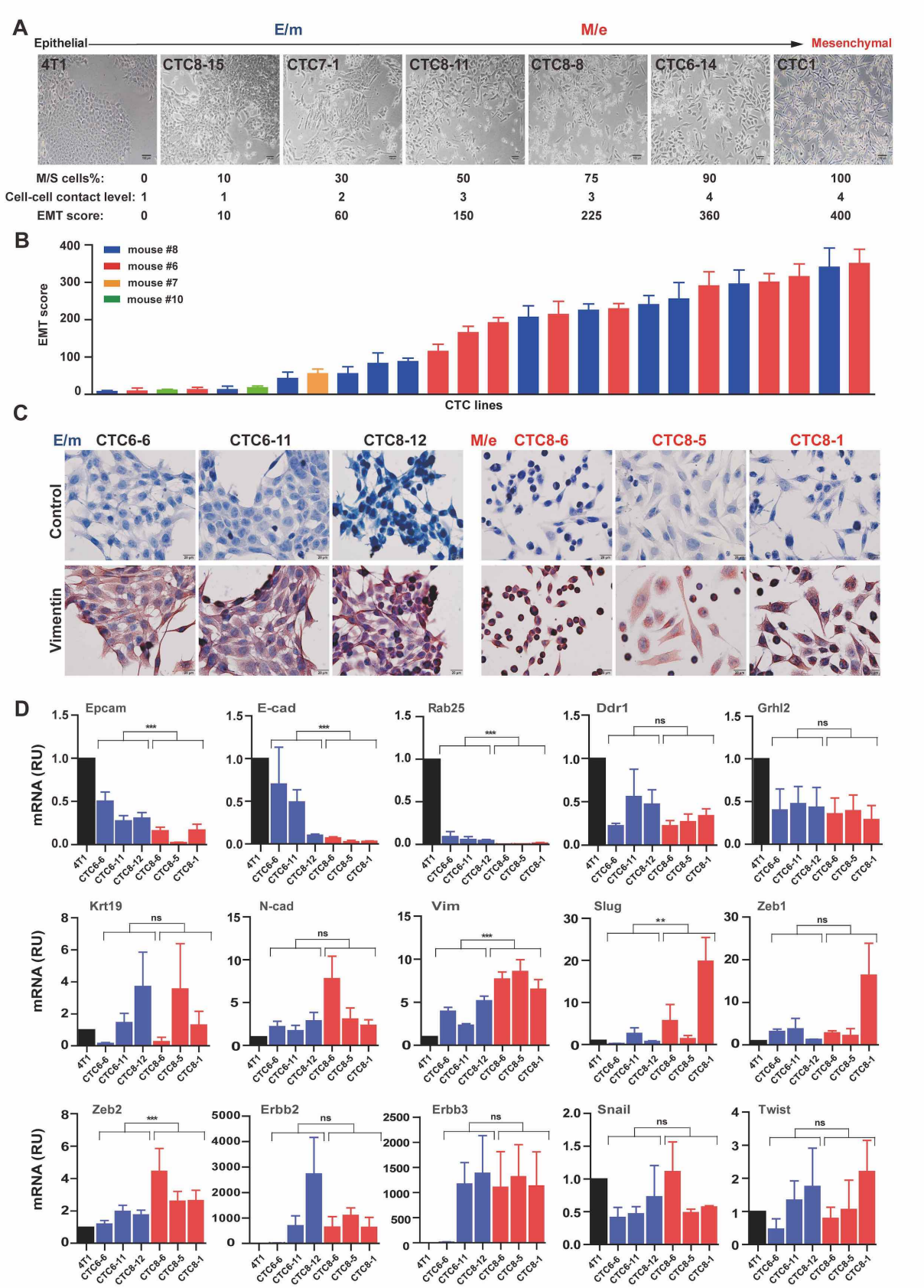


Figure S4:

(**A**) Epithelial and mesenchymal phenotypes of 4T1, CTC1, and CTC lines generated from mice re-transplanted with DTC1 cells were assessed. Shown are representative pictures of CTC1, 4T1 and CTCs

displaying various degrees of EMT (upper panels). EMT score is presented as product of percentage of mesenchymal/spindle shape cells and cell-cell contact level (see Materials and Methods).

(**B**) EMT score for 4T1, CTC1, and DTC1-derived CTC lines is presented as mean with SD from n = 4 experiments. Mice of origin are color-coded in the bar graph.

(**C**) Immunohistochemistry staining of vimentin and control in E/m-type (CTC6-6, CTC6-11, CTC8-12) and M/e-type (CTC8-6, CTC8-5, CTC8-1) CTCs derived from DTC1 transplantations. Shown are representative pictures from n = 3 staining.

(**D**) mRNA transcript levels of epithelial markers Epcam, E-cad, Rab25, Ddr1, Grhl2, Krt19, and mesenchymal maker N-cad, Vim, Slug, Zeb1/2, Erbb2/3, Snail, Twist in E/m-type (CTC6-6, CTC6-11, CTC8-12) and M/e-type (CTC8-6, CTC8-5, CTC8-1) CTCs derived from DTC1 transplantations, with 4T1 cells as a reference set to 1. Shown are mean with SD from n = 3 independent experiments performed in triplicates. T-test E/m versus M/e cells is indicated; ns: not significant, * <0.05, ** < 0.01, *** < 0.001.

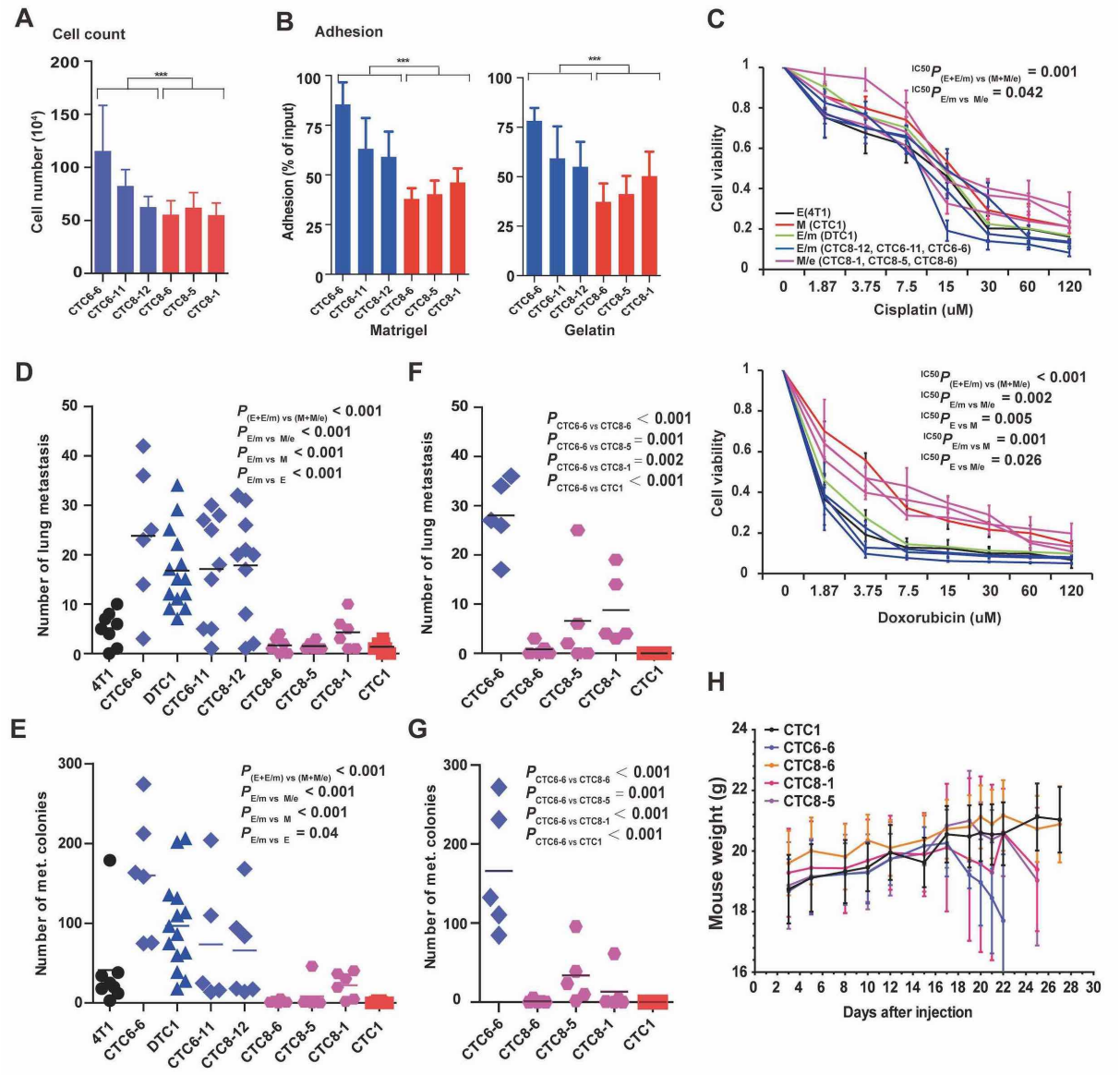


Figure S5:

(**A**) Cell proliferation of E/m-type (CTC6-6, CTC6-11, CTC8-12) and M/e-type (CTC8-6, CTC8-5, CTC8-1) CTCs derived from DTC1 transplantations (initial seeding number 5000 cells, cell numbers were counted on day 5). Shown are mean with SD from n=4 independent experiments. T-test E/m versus M/e cells is indicated; *** < 0.001.

(**B**) Adhesion assay to matrigel and gelatin was performed with E/m-type (CTC6-6, CTC6-11, CTC8-12) and M/e-type (CTC8-6, CTC8-5, CTC8-1) CTCs. Shown are mean adhesion rate with SD from n ≥ 3 independent experiments performed in triplicates. T-test E/m versus M/e cells is indicated; *** < 0.001.

(**C**) Chemoresistance towards cisplatin and doxorubicin was tested across a concentration range of 1.875 - 120 µM by MTT assay in the indicated cell lines. Shown are viability curves with SD from n = 3 independent experiments performed in triplicates. Statistical analysis was done by comparing IC_{50} values. One-way ANOVA with posthoc multiple testing and Bonferroni correction*, p-values is indicated.

(**D-E**) **D**: Epithelial E-type (4T1), E/m-type (CTC6-6, CTC6-11, CTC8-12, DTC1), M/e-type (CTC8-6, CTC8-5, CTC8-1) and mesenchymal M-type (CTC1) cells were transplanted into mice through tail vein injection. After 19 days, numbers of superficial lung metastasis were counted and lungs were harvested for further metastasis colony formation assay. Dot plots show numbers of metastasis counted in lungs including mean (line) and p-values. **E**: Metastasis colony formation assay was performed (Materials and Methods) and is displayed as dot plots showing numbers of colonies including mean (line). One-way ANOVA with *posthoc* multiple testing and Bonferroni correction*, p-values is indicated.

(**F-G**) E/m-type (CTC6-6), M/e-type (CTC8-6, CTC8-5, CTC8-1) and mesenchymal M-type (CTC1) cells were transplanted into mice through tail vein injection. Each experimental group was ended at indicated signs of endpoint. **F:** Dot plots show numbers of metastasis counted in lungs including mean (line) and p-values. **G**: Metastasis colony formation assay was performed and is displayed as dot plots showing numbers of colonies including mean (line). One-way ANOVA with *posthoc* multiple testing and Bonferroni correction*, p-values is indicated.

(**H**) Line charts shows mouse weight curves for each group as mean with SD.

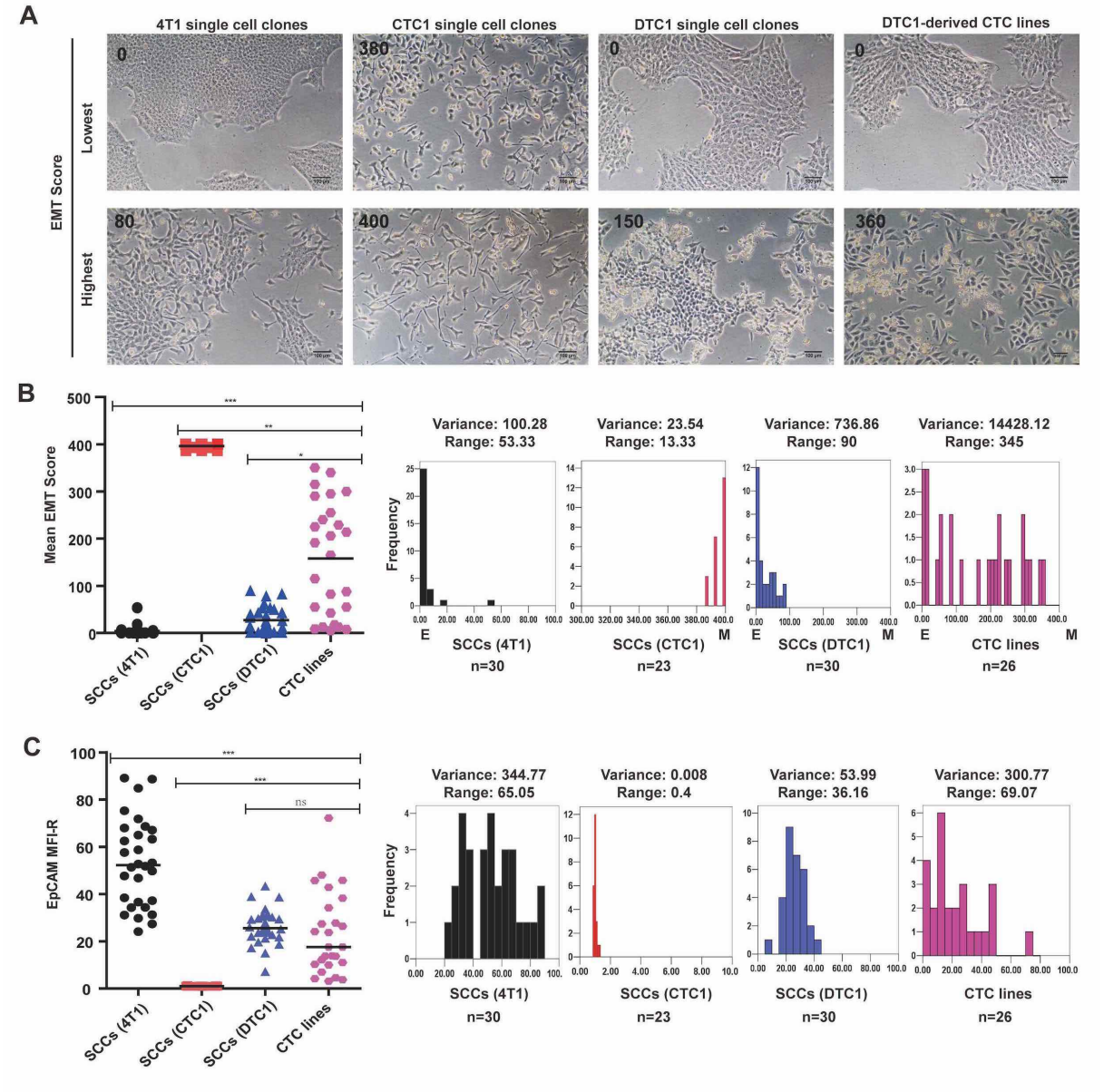


Figure S6:

(**A**) Single cell clones were generated from 4T1 (n = 30), CTC1 (n = 23), and DTC1 cells (n = 30). Shown are representative pictures of 4T1, CTC1, DTC1 single cell clones and DTC1-derived CTC lines (n = 26) with highest and lowest EMT scores.

(**B**) Dot plots shows mean values of EMT scores in 4T1, CTC1, DTC1 single cell clones and DTC1-derived CTC lines from n = 3 independent scoring. The degree of EMT score dispersion in each group is shown as frequency diagrams with variance (squared value of standard deviation) and range (difference between lowest and highest values). Kruskal-wallis test with *posthoc* multiple testing and Dunn´s correction, * < 0.05, ** < 0.01, *** < 0.001.

(**C**) Dot plots shows means of EpCAM expression in 4T1, CTC1, DTC1 single cell clones and DTC1-derived CTC lines from n = 2 initial measurements. The degree of EpCAM expression dispersion in each group is shown as frequency diagrams with variance and range. One-way ANOVA with *posthoc* multiple testing and Bonferroni correction, *** < 0.001.

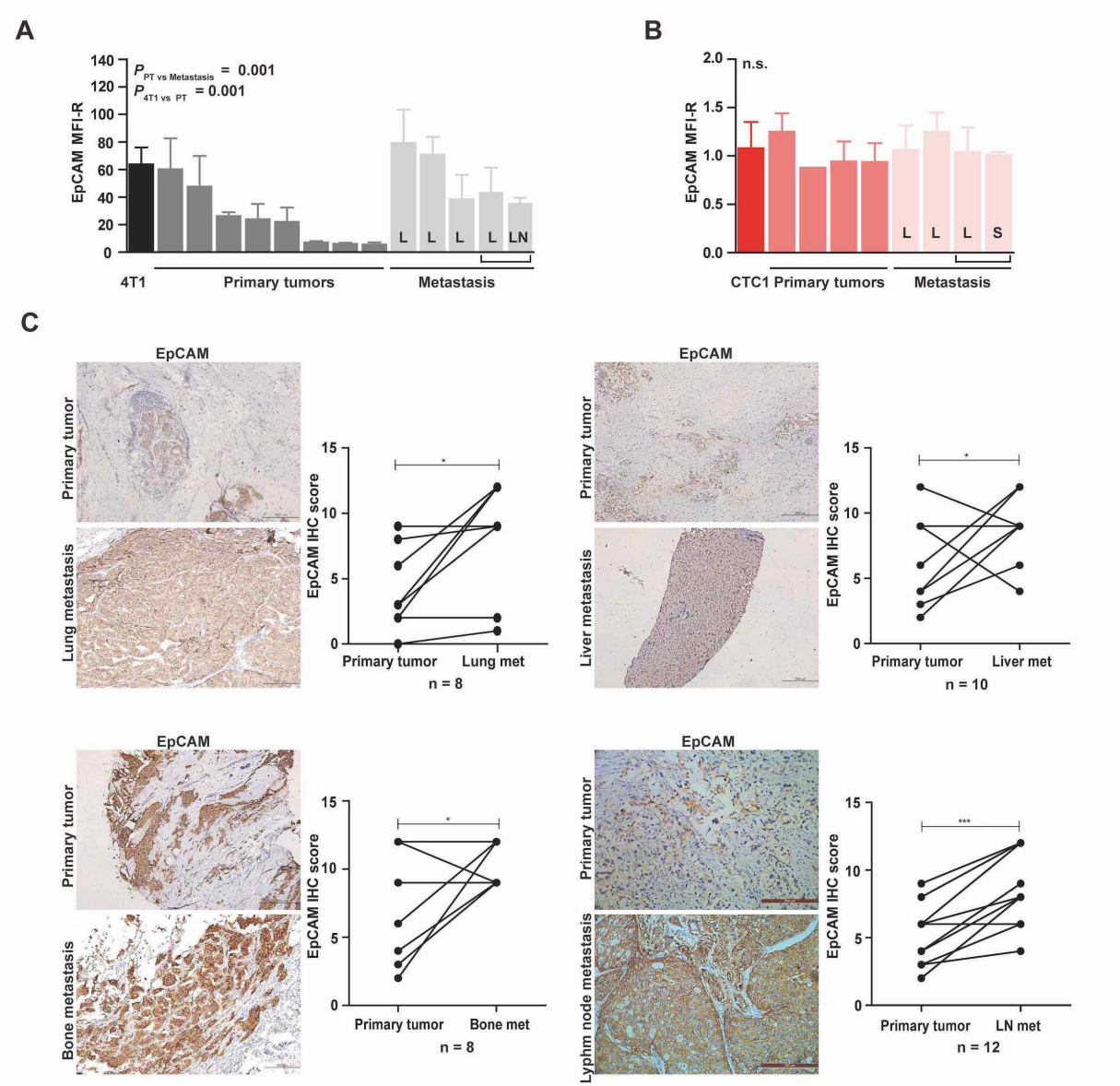


Figure S7:

(**A-B**) EpCAM expression in 4T1 derivative cell lines. EpCAM expression was measured by flow cytometry in permanent cell lines originating from primary tumor and metastases generated following re-transplanted of 4T1 (left panel) and CTC1 cells (right panel) into BALB/c mice. Brackets demark cell lines originating from one individual mouse. L: lung, S: spleen, LN: lymph node. Data is presented as mean fluorescence intensity ratios (EpCAM/control) with SD. Shown are mean MFI-R with SD from n ≥ 3 independent experiments performed in unicates. One-way ANOVA with posthoc multiple testing and Bonferroni correction*; p-values is indicated ns: not significant.

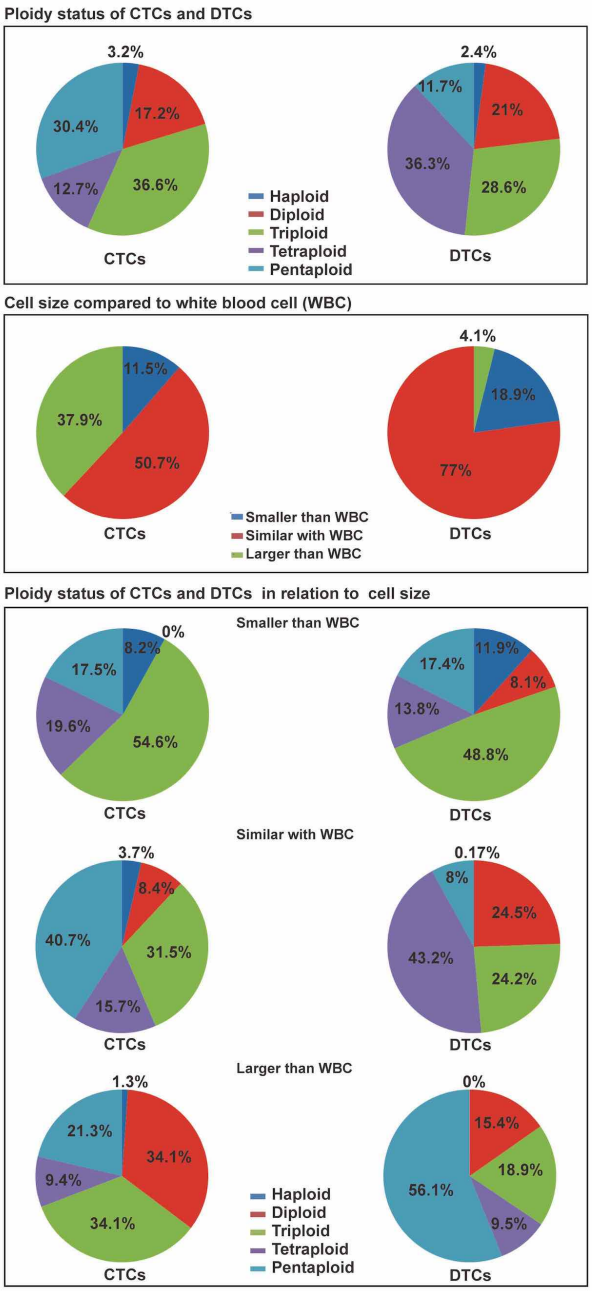
(**C**) EpCAM expression level was assessed in n = 38 human breast tumors. Shown are representative immunohistochemistry (IHC) staining of EpCAM in primary tumors and corresponding metastatic sites (lung, liver, bone marrow, lymph node) and the quantification of EpCAM IHC scores of paired tumor and metastases samples (see Materials and Methods). Paired T test; * < 0.05, *** < 0.001.

A

Characteristic	No. (%)
Total no. of patients	34 (100)
Gender	
Female	34 (100)
Age (years)	
Median	57
Range	28-76
Tumor stage	
Stage IIIa	1 (2.94)
Stage IIIc	10 (29.41)
Stage IV	23 (67.65)
Intrinsic Subtype	
Luminal A	6 (17.65)
Luminal B	15 (44.12)
HER2 +	9 (26.47)
HER2 -	6 (17.65)
HER2 positive (non-luminal)	7 (20.59)
Triple negative (basal-like)	4 (11.76)
Normal-like	2 (5.88)
Follow-up (Months)	
Median	11
Range	1-28
Distant detectable metastases	
Lung	5 (14.71)
Bone	5 (14.71)
Liver	4 (11.76)
Others	2 (5.88)
Multiple	7 (20.59)
Numbers of detected CTC per patient	
Median	9
Range	1-185
Numbers of detected DTC per patient	
Median	413
Range	4-40240

HER2 = human epidermal growth factor receptor 2.

B



C

w/o EpCAM+ diploid cells

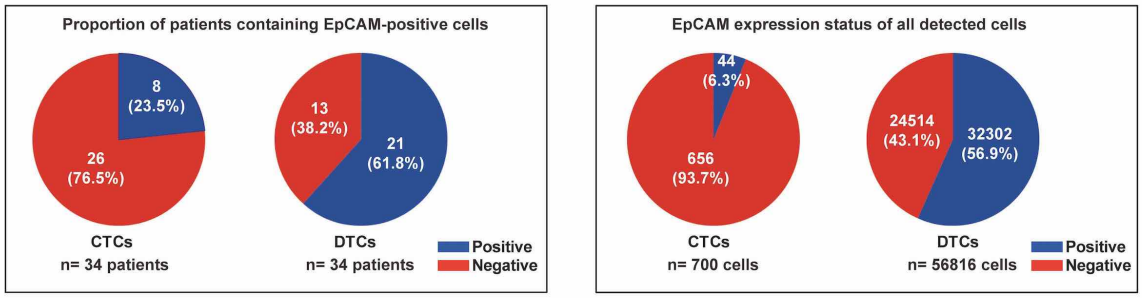
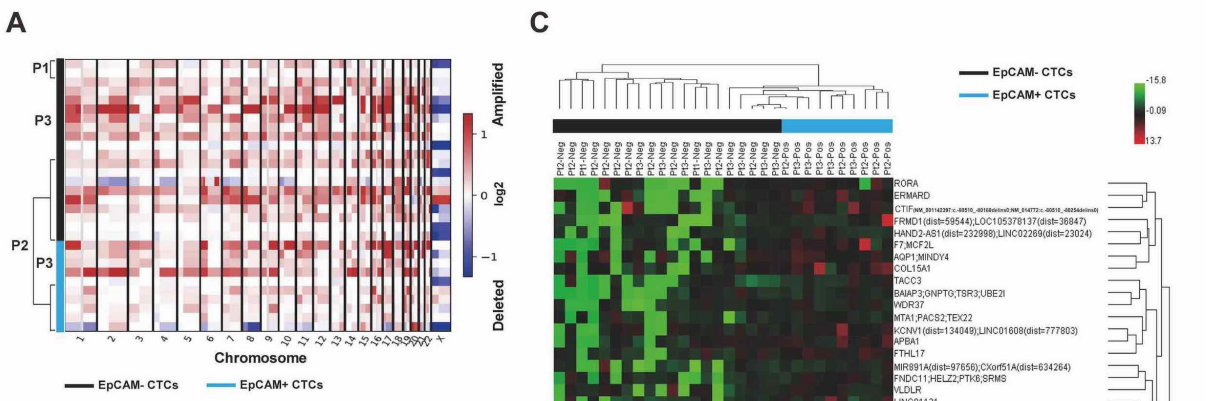


Figure S8:

(**A**) Patients' characteristics. Shown are clinical parameters for n=34 MBC patients including gender, age, tumor stage, subtype, follow-up time in months, distant metastases, CTC and DTC numbers.

(**B**) Characterization of chromosome 8 ploidy and cell size compared to white blood cells (WBC) in CTCs and DTCs isolated from breast cancer patients. Shown are frequencies of differential ploidy statuses and cell size compared to WBC in percent for CTC and DTC from n=34 MBC patients. Additionally, the ploidy status is depicted in relation to CTC and DTC size compared to WBC (lower panels).
(**C**) Shown are proportions of $EpCAM^+$ CTC and DTC in individual patients (n = 34) (left pie charts) and in overall numbers of CTC (n = 700) and DTC cells (n = 56,816) (right pie charts). Blue: $EpCAM^+$, red: $EpCAM^-$. Data exclude $EpCAM^+$ diploid CTC and DTC, and represent exclusively aneuploid cells.



B

Chromosome	NO. of CNVRs	No. of genes	Total length of CNVRs (bp)
1	65	128	3249107
2	64	94	3199142
3	44	82	2200491
4	29	46	1450565
5	45	78	2250786
6	32	52	1597679
7	28	52	1400012
8	24	54	1202875
9	33	55	1656598
10	27	42	1350461
11	29	65	1450727
12	36	61	1800333
13	18	22	900461
14	13	22	649922
15	27	61	1351108
16	25	59	1249948
17	31	108	1549902
18	10	14	499777
19	20	77	999275
20	16	25	799888
21	9	18	449844
22	10	33	500390
X	16	15	800046
Y	6	8	299910
Total	657	1255 (Duplication removed)	32859247

Figure S9:

(A) Visualization of copy number variation (CNV) profiles in n=10 EpCAM^{+} (light blue) and n=20 EpCAM^{-} single CTCs (black). Within the chromosome plots, red indicates DNA copy number amplifications and blue indicates deletions in log2 scale.

(**B**) Chromosome distribution of significantly different CNV regions (CNVRs) between EpCAM^{+} and EpCAM^{-} CTCs. Listed are chromosomes, numbers of CNVRs, numbers of genes affected and total length of CNV regions in base pairs (bp).

(**C**) Shown is an unsupervised clustering of the top 100 CNV, including genes encoded in the genomic regions affected by the CNV between $EpCAM^+$ (light blue) and $EpCAM^-$ CTCs (black). CTCs originated from n = 3 patients, which are denoted as Pt1-3.

Term	Gene count	%	P Value	Enrichment Score	Genes
Peptidyl-tyrosine dephosphorylation	12	1.447527	0.000346	3.460445	MTM1, PTPN6, DUSP3, PTPRD, CDC14A, PTPN3, UBASH3B, PTPRN2, PTPN13, PTPRT, PTPRQ
Endoplasmic reticulum organization	6	0.723764	0.001541	2.812319	TOR1AIP2, DNM1L, BNIP1, VAPB, ATL1, EIF2AK3
Negative regulation of canonical Wnt signaling pathway	14	1.688782	0.002451	2.610573	CSNK1A1, NKD2, BICC1, PARK2, WWTR1, TMEM64, GLI1, SFRP5, RGS20, GPC3, SOST, PSMB6, SCYL2, PSMD1
Microtubule-based movement	9	1.085645	0.004561	2.340949	KIF22, CLTA, KIF1A, DNAH14, DNAH17, KIF9, SH3GL2, DYNC1I2, DNAH6
Nervous system development	19	2.291918	0.005729	2.241948	ERBB4, MAFB, NLGN1, DPYSL5, DPYSL2, ARID1B, SLC7A5, SMN1, ARHGAP26, P2RX5, MEF2D, MYT1L, NAV2, MTR, CNTN4, NAIP, DLG2, ZNF423, DLG1
Mammary gland epithelial cell differentiation	4	0.482509	0.00751	2.124341	AKT1, PRLR, ERBB4, MGMT
Protein dephosphorylation	11	1.3269	0.007694	2.113842	MTM1, PTPN6, PTPRD, PTPN3, CPPED1, PTPRN2, PPP3R1, PTPN13, PTPRT, LHPP, DLG1
Phospholipid biosynthetic process	6	0.723764	0.008908	2.050225	PGS1, LPCAT1, PEMT, PCYT1A, AGPAT3, LPCAT3
Neural retina development	4	0.482509	0.009332	2.030015	RAB11FIP4, SLC17A8, ACTL6A, TGFB2
Microspike assembly	3	0.361882	0.009747	2.011145	MTSS1, ACTN2, ABL1
Mammary gland duct morphogenesis	3	0.361882	0.009747	2.011145	GLI2, SCRIB, CSF1R
Neuronal action potential	5	0.603136	0.011882	1.925105	P2RX4, CATSPER4, SCN1A, CHRNA1, GPR88
Positive regulation of mitotic cell cycle	5	0.603136	0.011882	1.925105	CCNB1, SHB, EIF4EBP1, DUSP3, ABL1
Insulin secretion involved in cellular response to glucose stimulus	3	0.361882	0.014309	1.844389	RAB11FIP5, PTPRN2, RAB11B
Positive regulation of establishment of protein localization to plasma membrane	5	0.603136	0.015121	1.820428	AKT1, EZR, NKD2, DPP10, DLG1
Mammary gland alveolus development	4	0.482509	0.016228	1.789727	PRLR, ERBB4, PHB2, ESR1
Transport	20	2.412545	0.018401	1.735169	CREBRF, SLC20A2, GRIK4, SLC12A5, UNC50, CACNG2, SLC7A5, SEC14L1, ABCG2, P2RX5, RAB11FIP4, P2RX4, SEC22A, SLC13A2, PITPNC1, CLVS1, SLC25A39, SLC51A, CHRNA1, SLC27A4
Bicellular tight junction assembly	5	0.603136	0.018866	1.724313	CLDN3, MPP5, STRN, PTPN13, DLG1
Peptidyl-serine phosphorylation	10	1.206273	0.020018	1.698588	CSNK1A1, AKT1, STK32A, CSNK1G1, TTBK2, MORC3, LMTK2, PRKCH, EIF2AK3, DMPK
Regulation of mitophagy	5	0.603136	0.020936	1.679113	DNM1L, ATG7, BNIP3L, ACTL6A, SREBF2
Cell differentiation	24	2.895054	0.027298	1.563874	SRPK2, PTPN6, RMDN3, APOLD1, PRKCH, CDHR5, EDAR, NHS, SLC7A5, FOXN3, PURB, PTPRQ, SFRP5, AKT1, SHB, MYT1L, RGS20, RNF151, YIPF3, CAND1, JAK2, ETV6, ZNF423, ANGPTL4
Toxin transport	5	0.603136	0.027954	1.55356	DNAJC17, TCP1, TRIP4, RAB28, ATP6V0A1
Apoptotic process	28	3.377563	0.028827	1.540204	ZFAND6, DCC, SEPT4, GULP1, SAV1, HINT2, RASSF7, CTNNBL1, PRUNE2, SHB, DOCK1, CDCA7, CSE1L, CASP9, API5, PTPN6, RMDN3, EDAR, NOA1, NLRP1, SFRP5, MEF2D, BNIP1, BNIP2, JAK2, NAIP, GADD45A, PUF60
Positive regulation of chemokine secretion	3	0.361882	0.032209	1.492029	IL4R, C5, CSF1R
Receptor localization to synapse	3	0.361882	0.032209	1.492029	NLGN1, DLG2, DLG1
Response to endoplasmic reticulum stress	7	0.844391	0.034011	1.468384	TTC23L, CREBRF, ALOX15, PDIA5, PARK2, ABL1, EIF2AK3
Regulation of membrane potential	7	0.844391	0.034011	1.468384	KCNMA1, ASIC2, ACTN2, PXK, FAM19A4, CHRNA1, DLG1
Rransmembrane receptor protein tyrosine kinase signaling pathway	8	0.965018	0.035665	1.447763	NTRK3, CNKSR1, MTSS1, ERBB4, PTPRT, SHC3, CSF1R, BLNK
Regulation of cell shape	10	1.206273	0.037814	1.422344	CSNK1A1, EZR, CSNK1G1, TTBK2, DIAPH1, C15ORF62, BAMBI, ARAP1, CSF1R, DLG1
Cellular response to DNA damage stimulus	13	1.568154	0.037815	1.422335	VAV3, KIAA0101, MCM10, CHCHD6, RAD50, RAD51, SETX, AKT1, RAD1, CUL4A, CASP9, IRF7, ABL1
Brain development	12	1.447527	0.044552	1.351133	SEPT4, CD9, SCT, SLC17A8, VCY1B, MDGA1, CNTNAP2, CNTN4, DPYSL2, MBD3, CTNS, SLC7A11
Signal peptide processing	4	0.482509	0.045428	1.342675	SEC11A, PROZ, SPPL2A, SPCS2
Regulation of release of sequestered calcium ion into cytosol	3	0.361882	0.047162	1.326405	PTPN6, UBASH3B, DIAPH1
Signal transduction in response to DNA damage	3	0.361882	0.047162	1.326405	CASP9, ABL1, GADD45A
Positive regulation of nitric oxide biosynthetic process	5	0.603136	0.049196	1.308071	AKT1, P2RX4, ESR1, JAK2, TLR5

Figure S10:

Enrichment analysis of GO biological process terms was carried out with in 657 CNVs comprising 1255 genes obtained from $EpCAM^{+}$ vs $EpCAM^{-}$ CTCs. Shown are all terms with p-value < 0.05. Genes marked in red font were amplified, genes marked in green font were deleted.