Cell

Circulating Tumor Cell Clusters Are Oligoclonal Precursors of Breast Cancer Metastasis

Nicola Aceto,[1,2] Aditya Bardia,[1,2] David T. Miyamoto,[1,5] Maria C. Donaldson,[1,2] Ben S. Wittner,[1,2] Joel A. Spencer,[3,4] Min Yu,[1,2] Adam Pely,[3,4] Amanda Engstrom,[1,2] Huili Zhu,[1,2] Brian W. Brannigan,[1,2] Ravi Kapur,[6] Shannon L. Stott,[1,2,6] Toshi Shioda,[1,2] Sridhar Ramaswamy,[1,2] David T. Ting,[1,2] Charles P. Lin,[3,4] Mehmet Toner,[6,7] Daniel A. Haber,[1,2,8,*] and Shyamala Maheswaran[1,7,*]

[1]Massachusetts General Hospital Cancer Center, Harvard Medical School, Boston, MA 02129, USA
[2]Department of Medicine, Harvard Medical School, Boston, MA 02129, USA
[3]Advanced Microscopy Program, Wellman Center for Photomedicine and Center for Systems Biology, Massachusetts General Hospital, Boston, MA 02114, USA
[4]Harvard Stem Cell Institute, 1350 Massachusetts Avenue, Cambridge, MA 02138, USA
[5]Department of Radiation Oncology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02129, USA
[6]Center for Bioengineering in Medicine, Harvard Medical School, Boston, MA 02129, USA
[7]Department of Surgery, Harvard Medical School, Boston, MA 02129, USA
[8]Howard Hughes Medical Institute, Chevy Chase, MD 20815, USA
*Correspondence: dhaber@mgh.harvard.edu (D.A.H.), maheswaran@helix.mgh.harvard.edu (S.M.)
http://dx.doi.org/10.1016/j.cell.2014.07.013

SUMMARY

Circulating tumor cell clusters (CTC clusters) are present in the blood of patients with cancer but their contribution to metastasis is not well defined. Using mouse models with tagged mammary tumors, we demonstrate that CTC clusters arise from oligoclonal tumor cell groupings and not from intravascular aggregation events. Although rare in the circulation compared with single CTCs, CTC clusters have 23- to 50-fold increased metastatic potential. In patients with breast cancer, single-cell resolution RNA sequencing of CTC clusters and single CTCs, matched within individual blood samples, identifies the cell junction component plakoglobin as highly differentially expressed. In mouse models, knockdown of plakoglobin abrogates CTC cluster formation and suppresses lung metastases. In breast cancer patients, both abundance of CTC clusters and high tumor plakoglobin levels denote adverse outcomes. Thus, CTC clusters are derived from multicellular groupings of primary tumor cells held together through plakoglobin-dependent intercellular adhesion, and though rare, they greatly contribute to the metastatic spread of cancer.

INTRODUCTION

The metastatic spread of breast cancer, typically to bone, lung, liver, and brain, accounts for the vast majority of cancer-related deaths (Nguyen et al., 2009). Our understanding of epithelial cancer metastasis is derived primarily from mouse models and it is thought to involve a series of sequential steps: epithelial-to-mesenchymal transition (EMT) of individual cells within the primary tumor leading to their intravasation into the bloodstream, survival of such circulating tumor cells (CTCs) within the bloodstream, and finally their extravasation at distant sites, where mesenchymal-to-epithelial transition (MET) culminates in their proliferation as epithelial metastatic deposits (Hanahan and Weinberg, 2011). While EMT has indeed been demonstrated in human breast cancer cells in the circulation (Yu et al., 2013), the requirement for EMT to initiate metastasis has been debated (Ledford, 2011; Tarin et al., 2005). Alternative models proposed include tumor-derived microemboli that may break off from primary tumors, lodging into distal capillaries where they initiate metastatic growth (Fidler, 1973; Liotta et al., 1976; Molnar et al., 2001). Using diverse technological platforms, we and others have indeed detected clusters of CTCs, ranging from 2–50 cancer cells, within the circulation of patients with metastatic epithelial cancers (Cho et al., 2012; Fidler, 1973; Molnar et al., 2001; Stott et al., 2010; Yu et al., 2013).

Studies of cancer metastasis have emphasized the concept of "seed versus soil" as a key determinant of metastatic propensity (Fidler, 2003). This model matches the importance of mutated genetic drivers within tumor cells conferring proliferative and invasive properties, with that of the microenvironment of the distant organ or "niche," which may facilitate metastatic growth. However, the physical characteristics of single CTCs and CTC clusters may also contribute to metastatic propensity, especially as they impact the ability of epithelial tumor cells to survive the loss of cell adherence and shear forces in the blood stream, i.e., different survival signals among the cancer cell "seeds" may be important. For instance, in a mouse endogenous pancreatic cancer model, noncanonical Wnt signaling is elevated within CTCs, where it appears to suppress anoikis (Yu et al., 2012),

while in a subcutaneous tumor xenograft model, the admixture of tumor and stromal cells within microemboli may contribute stromal-derived survival signals (Duda et al., 2010). In a recent study of human breast cancer, mesenchymal markers indicative of EMT were expressed within the cancer cells comprising CTC clusters (Yu et al., 2013). Taken together, both human and mouse modeling studies point to the complexity of bloodborne metastasis and the need to capture and characterize CTCs to better understand this process.

CTCs have been detected in the majority of epithelial cancers, where they represent cancer cells captured as they transit through the bloodstream (Alix-Panabières and Pantel, 2013; Yu et al., 2011). As such, they hold the key to understanding critical pathways that mediate the bloodborne dissemination of cancer, which may not be readily evident through analyses of bulk primary or metastatic tumor populations. Factors leading to the generation of CTCs from a primary tumor are unknown, including the fraction derived from cancer cells that have actively intravasated into the bloodstream, versus those that are passively shed as a result of compromised tumor vasculature. Although exceedingly rare compared with normal blood cells, the number of CTCs in the bloodstream far exceeds the number of metastatic lesions in patients, indicating that the vast majority CTCs die in the bloodstream, with only a minor fraction representing viable metastatic precursors. Epithelial cells that have lost adhesion-dependent survival signals rapidly undergo anoikis, a fate likely to meet most CTCs in the bloodstream. It is in this context that either mesenchymal transformation, stromal-derived factors, or persistent interepithelial cell junctions may provide survival signals that attenuate this apoptotic outcome (Duda et al., 2010; Mani et al., 2008; Robson et al., 2006; Yu et al., 2012). Dissecting the contributions of these various mechanisms to human cancer requires the ability to isolate individual CTCs from the bloodstream and subject these to detailed molecular analyses.

Multiple technologies have been developed for CTC capture, taking advantage of tumor-specific epitopes absent in normal blood cells, variations in their physical properties such as size, density, and electromechanical characteristics, or by applying high throughput imaging to unpurified blood cell preparations (for review, see Yu et al., 2011). The fact that CTCs are extremely rare, even in patients with advanced metastatic cancers (estimated at one CTC/billion normal blood cells), and that they may be poised on the verge of apoptosis, has made their analysis contingent upon technological constraints. We have introduced a series of microfluidic devices that have the advantage of low-shear, yet high throughput, interrogation of unprocessed whole blood, providing highly enriched and unfixed CTCs that are suitable for detailed molecular analysis (Nagrath et al., 2007; Ozkumur et al., 2013; Stott et al., 2010). Among these, the herringbone (^{HB}CTC-Chip) makes use of grooves within the ceiling of the microfluidic chamber to generate turbulent microfluidic flow, directing cells against antibody-coated walls of the device, where CTCs are captured (Stott et al., 2010). This device, whose highly efficient design enabled our initial detection of large CTC clusters, requires on-chip cell lysis for nucleic acid extraction and hence provides an enriched but heterogeneous CTC population for analysis (Yu et al., 2012, 2013). In contrast, our recently described ^{neg}CTC-iChip achieves highly efficient depletion of erythrocytes and leukocytes from blood specimens, yielding untagged CTCs and small CTC clusters in solution, where they can be micromanipulated for single-cell RNA sequencing (Ozkumur et al., 2013). Here, we use both of these devices, along with in vivo flow cytometry and next generation RNA sequencing, to interrogate CTCs from both patients with metastatic breast cancer and mouse tumor models. We find, using mouse models, that CTC clusters are derived from oligoclonal clumps of primary tumor cells and constitute a rare but very highly metastasis-competent subset of CTCs, compared with single circulating breast cancer cells. RNA sequencing of human breast CTC clusters identifies plakoglobin as a key mediator of tumor cell clustering, which is expressed in a heterogeneous pattern within the primary tumor. Knockdown of plakoglobin expression in the mouse model suppresses CTC cluster formation and reduces metastatic spread.

RESULTS

Endogenous CTC Clusters Have Increased Metastatic Potential Compared to Single CTCs

To define the origin and functional properties of CTC clusters, compared with single CTCs, we made use of mouse models, where tumor cell composition, transit of CTCs through the bloodstream, and metastatic deposits can be monitored and quantified. We first established a model to test the generation of endogenous CTCs and metastases from a primary orthotopic tumor xenograft. These experiments were designed both to test the metastatic propensity of CTC clusters versus single CTCs, as well as to determine whether CTC clusters originate from an oligoclonal grouping of primary tumor cells or from the clonal progeny of an individual tumor cell. MDA-MB-231-LM2 (LM2) cells, a lung-metastatic variant of MDA-MB-231 human breast cancer cells (Minn et al., 2005), were engineered to express either green fluorescent protein (LM2-GFP) or mCherry (LM2-mCherry), and a 1:1 mixture of these differentially tagged cells was injected into the mammary fat pad of immunodeficient (NSG) mice. As expected, overt primary breast tumors were observed after 5 weeks and these retained an equal distribution of LM2-GFP and LM2-mCherry tagged cells, as confirmed by IHC staining (Figures S1A and S1B available online). We sampled the blood of tumor-bearing animals for presence of single or clustered CTCs using a terminal bleed and simultaneously harvested the lungs for analysis of metastatic deposits. In addition to enumeration of CTCs, we reasoned that clonally-derived CTC clusters would uniformly express either GFP or mCherry, whereas aggregations of cells from the primary tumor would be heterogeneous for the two markers (Figure 1A). We observed a mean of 2,486 CTC events per mouse (n = 5 mice), of which a mean of 65 (2.6%) were CTC clusters and 2,421 (97.4%) were single CTCs (Figures 1B, 1C, and S2A). Virtually all (91%) CTC clusters were dual positives for GFP and mCherry. A mean of 5.6 (9%) CTC clusters per mouse (with fewer than three cells per cluster) were comprised of cells expressing only one of the two markers, consistent with expected probabilities given a 1:1 mixture of GFP/mCherry expressing cells in the primary tumor (Figures 1C and S2B). Thus, CTC clusters do not result from the proliferation of a single tumor cell in the vasculature, instead

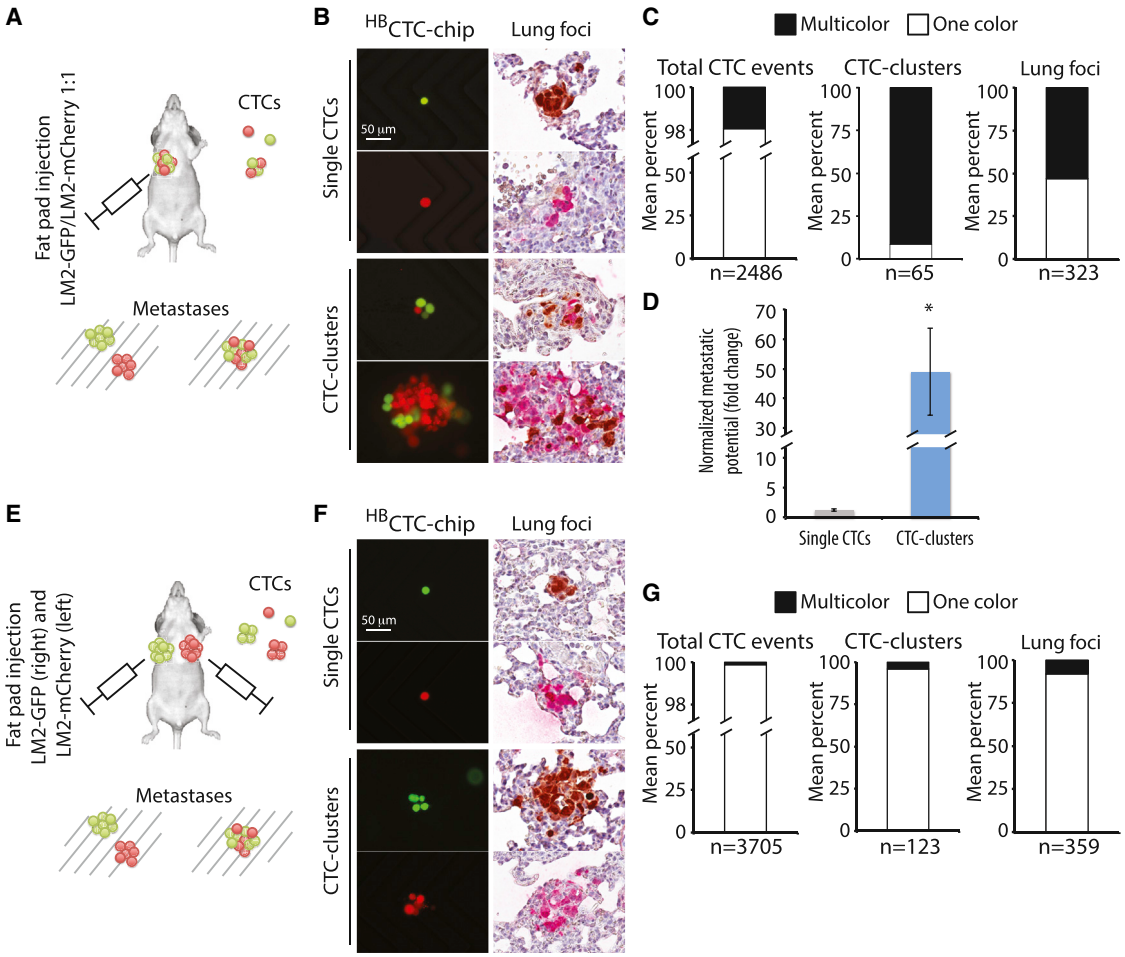

Figure 1. CTC Clusters Demonstrate Increased Metastatic Potential Compared to Single CTCs

(A) Schematic of the experiment. MDA-MB-231-LM2 (LM2) cells expressing GFP (LM2-GFP) or mCherry (LM2-mCherry) cells were mixed at 1:1 ratio and injected in the right mammary gland of immunodeficient mice to generate one-color single CTCs and multicolor CTC clusters. Accordingly, one-color metastatic foci are derived from a single CTC, while multicolor foci arise predominantly from a CTC cluster.

(B) Representative images of single CTCs (GFP- or mCherry-positive) and CTC clusters (GFP- and mCherry-positive) captured on the ^{HB}CTC-Chip (left). Lung metastatic foci derived from a single CTC (GFP- or mCherry-positive) or a CTC cluster (GFP- and mCherry-positive) are shown (right). GFP (brown), mCherry (red). Blood samples and lung specimens were isolated 5 weeks after primary tumor development. n = 5.

(C) Bar graphs showing the mean percentage of one-color versus multicolor CTC events captured by the ^{HB}CTC-Chip (left), the mean percentage of one-color versus multicolor CTC clusters (middle), as well as the mean percentage of one-color versus multicolor lung foci (right). n = 5.

(D) Bar graph showing the normalized metastatic potential of single CTCs and CTC clusters. Error bars represent SEM. n = 5, *p = 0.031 by Student's t test.

(E) Schematic of the experiment. LM2-GFP cells were injected in the right mammary gland while LM2-mCherry cells were injected in the left mammary gland of immunodeficient mice to generate tumors that give rise to one-color single CTCs and CTC clusters, as well as rare multicolor CTC clusters (resulting from aggregation events). Accordingly, one-color metastatic foci are derived from a single CTC or a CTC cluster, while multicolor foci derive from CTC aggregates.

(F) Representative images of single CTCs (GFP- or mCherry-positive) and CTC clusters (GFP- or mCherry-positive) captured on the ^{HB}CTC-Chip (left). Lung metastatic foci derived from a single CTC (GFP- or mCherry-positive) or a CTC cluster (GFP- or mCherry-positive) are shown (right). GFP (brown), mCherry (red). Blood samples and lung specimens were isolated 5 weeks after primary tumor development. n = 5.

(G) Bar graphs showing the mean percentage of one-color versus rare multicolor CTC events captured by the ^{HB}CTC-Chip (left), the mean percentage of one-color versus multicolor CTC clusters (middle), as well as the mean percentage of one-color versus multicolor lung foci (right). n = 5.

See also Figures S1 and S2.

they appear to represent the aggregation of neighboring cells, most likely within the primary tumor mass (see below).

Metastatic deposits in the lungs were analyzed for both number and composition using anti-GFP and anti-mCherry antibodies, simultaneously with the CTC analyses (Figures 1B, 1C, and S2A). Given the distribution of GFP and mCherry staining in CTC clusters, we reasoned that metastatic tumors derived from a single CTC would be positive for a single marker, while those derived from CTC clusters would stain for both GFP and mCherry (Figures 1A and 1B). A mean of 323 lung foci were identified per mouse (n = 5 mice), of which 171 (53%) were multicolor, and therefore derived from CTC clusters, versus 152 (47%)

unicolor derivatives of single CTCs (Figures 1C and S2A). Normalizing the number and distribution of lung metastases with that of single CTCs and CTC clusters, we calculate that a CTC cluster is ~50 times more likely to give rise to a metastatic deposit than a single CTC (Figure 1D). Thus, while CTC clusters are much more rare than single CTCs in this orthotopic mouse model of breast cancer, they contribute equally to the metastatic burden in the lung.

To further validate (1) that oligoclonal CTC clusters arise from the fragmenting of primary tumor cell clumps into the vasculature and not from intravascular aggregation of single CTCs, and (2) that oligoclonal lung metastases arise from CTC clusters and not from the reseeding of a metastatic site by multiple single CTCs, we undertook a second series of orthotopic mouse xenograft experiments, injecting LM2-GFP cells into the right mammary fat pad and LM2-mCherry cells in the left fat pad of immunodeficient mice (Figure 1E). Five weeks after injection, mice harbored two independent and differentially tagged tumors, and we again simultaneously harvested the blood for analysis of CTCs and the lungs for enumeration of metastatic deposits. As expected, single CTCs in the circulation demonstrated equal contributions from the GFP and m-Cherry-tagged primary tumors. However, unlike the previous multitagged single tumor model, in mice with two independent individually-tagged tumors, the vast majority of CTC clusters (96%) were of a single color, with equal contributions from GFP- or mCherry-positive primary tumors (Figures 1F, 1G, and S2A). Thus, the vast majority of CTC clusters are derived from individual primary tumors, excluding intravascular aggregation of single CTCs as a significant source of CTC clusters.

A very small fraction of CTC clusters observed in the dual tumor-bearing mice were multicolor (4% of CTC clusters, corresponding to 0.12% of total CTC events) (Figures 1G and S2A). While extraordinarily rare, the presence of such CTC clusters derived from two independent tumors may originate either from the uncommon intravascular aggregation of single CTCs or from a mixing of cancer cells within the two primary tumors, due to the previously reported "tumor reseeding" phenomenon (Kim et al., 2009) (Figure S1B). Consistent with the latter hypothesis, we found that 3%–5% of cells within the GFP-tagged primary tumor were positive for mCherry and 3%–5% of cells within the mCherry-labeled tumor were positive for GFP (Figure S1B). In addition to rare multicolor CTC clusters, we observed a small fraction (8%) of multicolor tumors in the lung (Figures 1G and S2A). These metastatic lesions could result either from the rare multicolor CTC clusters or from the reseeding of metastatic lesions by multiple single CTCs.

We confirmed the findings derived from the two LM2 mouse xenograft experiments with a second, mouse-derived breast cancer cell line, 4T1 (Figures S2C and S2D). Consistent with the LM2 results, a 1:1 mixture of 4T1-GFP and 4T1-mCherry cells within an orthotopic mammary tumor generated CTC clusters that were overwhelmingly multicolored (90%), whereas two separate primary 4T1 tumors labeled either with GFP or mCherry produced CTC clusters that were of a single color (87%). These observations further support that CTC clusters arise as oligoclonal fragments derived from a single tumor (Figure S2C). Normalizing the number and color distribution of 4T1-derived lung metastases relative to the prevalence of single CTCs and CTC clusters, we calculated a 23-fold increase in metastatic competence for CTC clusters versus single CTCs (Figure S2D), an estimate that is comparable to the 50-fold increase derived from LM2 cell experiments. Taken together, these two mouse tumor models indicate that CTC clusters constitute only 2%–5% of all CTC events detected in the circulation, but their dramatically elevated metastatic potential (23–50 times that of single CTCs) contributes to approximately half of all metastatic lesions in orthotopic breast cancer models.

Clustered Cancer Cells Are More Resistant than Single Cells to Apoptosis following Dissemination to the Lung

We generated an in vitro assay that allowed us to obtain a suspension of either single cells or clustered cells (2–30 cells) from cultures of GFP-*Luciferase*-tagged LM2 cells (see Extended Experimental Procedures). We injected 200,000 LM2 cells prepared either as single cells (LM2-SC) or as clusters (LM2-CL) into the tail vein of immunodeficient mice and subjected them to serial luciferase-based imaging (Figures 2A and 2B). Both LM2-SC and LM2-CL cells reached the lungs with equal efficiency (day 0), as shown by both bioluminescence and GFP immunohistochemical (IHC) staining (Figure 2B). However, over the following days, the LM2-SC lung signal progressively diminished as the cells underwent massive apoptosis, demonstrated by staining for cleaved caspase 3 (Figure 2C). In contrast, the LM2-CL lung signal persisted following intravascular inoculation, with cells showing resistance to apoptosis and tumors expanding more rapidly (Figures 2B–2D). Lung tumors eventually grew in mice subjected to tail vein injection with either of the two LM2 derivatives, but injection of clustered cells resulted in reduced overall survival, with 12.7 weeks for LM2-CL versus 15.7 weeks for LM2-SC ($p < 0.016$) (Figure 2E). We confirmed the differential rate of apoptosis and metastatic growth in the lung for single versus clustered cancer cells using tail vein injection of two additional breast cancer cell lines, BT474 and 4T1 (Figures S3A–S3D).

Calculation of CTC Clusters and Single CTC Circulatory Clearance Rate Using In Vivo Flow Cytometry

Clusters of tumor cells may exhibit considerable flexibility as they navigate through narrow channels, and capillary beds themselves may have uneven vessel diameters or bypass tracts that allow transit of large multicellular structures. However overall, CTC clusters are more likely than single CTCs to be trapped in small capillaries of the lung and distal organs. Thus, the low steady-state level of CTC clusters in the circulation may reflect a considerably higher generation rate if their clearance rate is very high. To test if CTC clusters indeed have a faster clearance rate from the bloodstream than single CTCs, we used in vivo flow cytometry (IVFC) to monitor LM2-SC and LM2-CL cells labeled with the lipophilic carbocyanine membrane dye DiD, following tail vein injection in immunodeficient mice (Figure 3A). DiD was selected to achieve optimal detection of CTCs with our IVFC settings. Circulating DiD-labeled cells were detected in real time within the ear blood vessels for a total of 55 min in each mouse. Injected LM2-CL cells were cleared at least three times more rapidly than LM2-SC (half-life: 6–10 min for LM2-CL versus 25–30 min for LM2-SC) (Figure 3B). Together, these observations

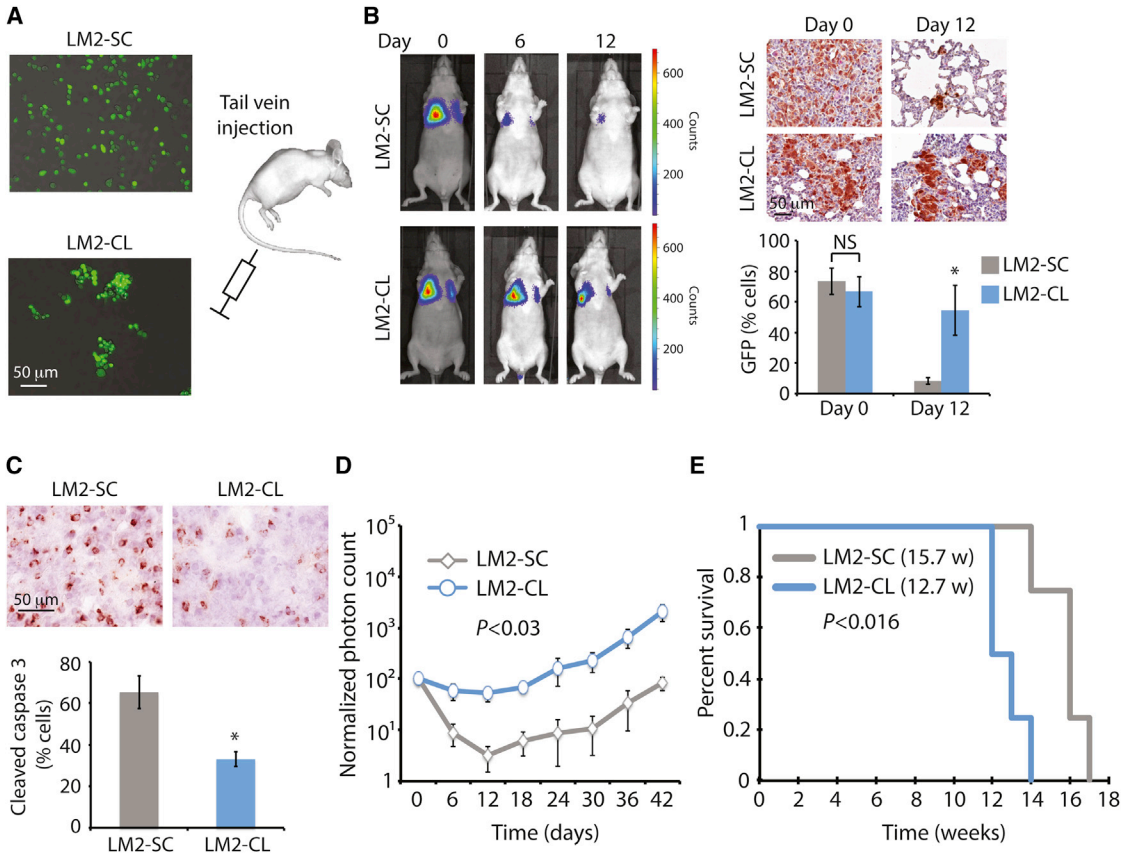

Figure 2. CTC Clusters Are More Resistant to Apoptosis at Distal Metastatic Sites

(A) Schematic showing MDA-MB-231-LM2-GFP-*Luciferase* (LM2) cells prepared as single cells (LM2-SC) or as clusters (LM2-CL) prior to injection into the tail vein of immunodeficient mice. Cells (2×10^5) were injected as LM-SC or LM2-CL per mouse.

(B) Representative bioluminescence images of mice at 0, 6, and 12 days after tail vein injection with LM2-SC or LM2-CL cells (left). n = 4. Representative images of GFP-stained sections of mouse lungs after injection with LM2-SC or LM2-CL cells (right). The bar graph shows the mean percentage of GFP-positive cells in lungs from LM2-SC- or LM2-CL-injected mice. Error bars represent SEM. n = 4; NS, not significant, *p = 0.03 by Student's t test.

(C) Representative images of cleaved caspase 3-stained sections of mouse lungs 24 hr after injection with LM2-SC or LM2-CL cells. The bar graph shows the mean percentage of cleaved caspase 3-positive cells in lungs from LM2-SC- or LM2-CL-injected mice. Error bars represent SEM. n = 4; *p < 0.02 by Student's t test.

(D) Lung metastasis growth curve from mice injected with LM2-SC or LM-CL. Error bars represent SEM. n = 4; p < 0.03 by Student's t test.

(E) Kaplan-Meier survival plot showing survival rates for mice injected with LM2-SC or LM2-CL. n = 4; p < 0.016 by log rank test.

See also Figure S3.

define a circulating time for CTCs in the bloodstream: the shorter circulation half-life of CTC clusters is consistent with their more rapid entrapment within capillaries of distal organs, where they may initiate metastatic growth (Liotta et al., 1976).

The Presence of CTC Clusters in Patients with Breast and Prostate Cancer Correlates with Poor Prognosis

Having characterized the origin and metastatic potential of CTC clusters in mouse models, we undertook to study their properties in patients with cancer. To first test the clinical significance of CTC clusters in the blood of patients with progressing metastatic breast cancer, we measured their presence in blood specimens from a total of 79 patients, drawn at multiple time points over a period of 19 months. Patients were recruited to an IRB-approved study at the Massachusetts General Hospital Cancer Center, including women with estrogen receptor-positive (n = 49), HER2-positive (n = 13), triple negative (n = 17) subtypes of breast cancer (total: 265 data points). For these experiments, we made use of the ^{HB}CTC-Chip, which is highly efficient in capturing both large and small CTC clusters (Stott et al., 2010). We coated the microfluidic chamber with a combination of antibodies, targeting the epithelial cell adhesion molecule (EpCAM), as well as the lineage markers epithelial growth factor receptor (EGFR) and human epithelial growth factor receptor 2 (HER2/ErbB2), which together efficiently capture both epithelial and mesenchymal breast CTCs (Yu et al., 2013). After processing 3 ml of whole blood from patients with breast cancer, the CTCs captured on the chip were stained with antibodies against wide spectrum cytokeratin (CK) to identify CTCs and against the leukocyte marker CD45 to assess white blood cell (WBC) contamination (Figure 4A). CTCs were identified in 54 out of 79 patients (68%). Among patients with CTCs, 3 (5.6%) had CTC clusters evident across more

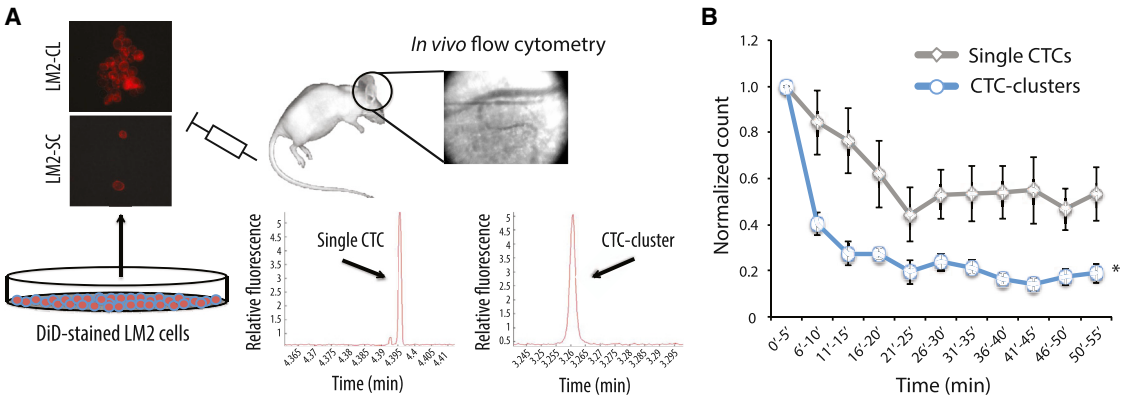

Figure 3. CTC Clusters Demonstrate a Faster Clearance Rate from the Bloodstream

(A) Schematic showing the experimental setup for measuring the clearance time of single CTCs and CTC clusters. Briefly, DiD-stained LM2 cells were prepared as LM2-SC or LM2-CL and injected into the tail vein of immunodeficient mice. In vivo flow cytometry was applied to the ear blood vessels to detect single CTCs and CTC clusters over a 55 min period after injection. Graphs show representative fluorescence peaks corresponding to the transit of a single CTC or CTC cluster through the ear blood vessel.

(B) Graph showing single CTCs and CTC clusters clearance curves. Error bars represent SEM. n = 5 for single CTCs and n = 4 for CTC clusters, *p < 0.01 by two-way ANOVA.

than three time points, while 16 (29.6%) had CTC clusters during one to three time points and 35 (64.8%) had no detectable clusters (Figure 4B). We correlated the presence of CTC clusters with progression-free survival (PFS) for all patients where such data were available (n = 30) (Table S1). Of note, PFS was calculated as time from initiation of therapy to discontinuation by the treating clinician (blinded to the CTC results), and PFS data analysis was performed only when clinical measurements bracketed the CTCs isolation time frame. Patients with CTC clusters across more than three time points had a mean progression-free survival time of 32.6 days, compared with 134.8 days for patients where CTC clusters were found during one to three time points and 160.5 days for patients with single CTCs only (p = 0.0002) (Figure 4C; Table S1). Thus, even among patients with advanced metastatic breast cancer, the continuous presence of CTC clusters is associated with an adverse clinical outcome.

Given the relatively short time to progression in patients with advanced breast cancer, we sought to test the correlation between CTC clusters and adverse prognosis in patients with a longer clinical course. We measured the number of CTCs in a total of 64 patients with prostate cancer using blood specimens drawn at multiple time points over a period of 53 months (total: 202 data points). CTCs in prostate cancer patients were visualized by staining with a cocktail of antibodies against prostate-specific antigen (PSA) and prostate-specific membrane antigen (PSMA); anti-CD45 staining was used to exclude white blood cells (Miyamoto et al., 2012). CTCs were detected in 48/64 patients (75%). CTC clusters were present in 6/48 samples (12.5%) (Figures 4D and 4E). In this cohort, the presence of CTC clusters during at least one time point strongly correlated with a dramatically shorter overall survival time (mean survival time was 115.8 days for patients with CTC clusters versus 930.1 days for patients with single CTCs; p = 0.00001) (Figure 4F; Table S1). While further studies will be required to ascertain the clinical utility of CTC clusters versus single CTCs as prognostic determinants in either breast or prostate cancer, these initial results point to the potential relevance of CTC clusters in the progression of human cancer.

Single-Cell Resolution RNA Sequencing of Matched CTC Clusters and Single CTCs Purified from Patients with Breast Cancer

The ability to capture both single CTCs and CTC clusters from the same blood specimen made it possible to undertake single-cell resolution RNA sequencing, searching for differences in expression profiles matched to individual patients. For these experiments, we applied the ^{neg}CTC-iChip, which enables isolation and single-cell manipulation of untagged CTCs, together with an optimized protocol for next generation RNA sequencing from minute amounts of template (Ozkumur et al., 2013; Tang et al., 2010). Blood specimens from ten patients with metastatic breast cancer were subjected to microfluidic depletion of RBCs and CD45- and CD66b-positive WBCs, leaving untagged single CTCs and small CTC clusters in the final product (Ozkumur et al., 2013). Unfixed tumor cells were stained for cell surface expression of EpCAM, HER2, and the mesenchymal marker CDH11 (Alexa488-conjugated), and counterstained with antibodies against CD45, CD14, and CD16 to identify contaminating leukocytes (TexasRed-conjugated) (Figure 5A). Individual CTC clusters (median of three cells per cluster) were isolated using a micromanipulator and compared with numerically matched pools of single CTCs from the same specimen, followed by next generation RNA sequencing (SOLiD 5500XL) (Figure 5A). We derived normalized expression profiles for a total of 29 samples (15 pools of single CTCs and 14 CTC clusters) isolated from ten breast cancer patients.

Unsupervised hierarchical clustering of RNA sequencing data showed no obvious distinctions at the global gene expression level between single CTCs and CTC clusters, with both of these clustering closely by patient of origin (Figure 5B). Consistent with the microscopic appearance of CTC clusters as primarily tumor cell-derived, we did not identify RNA signatures of other cell

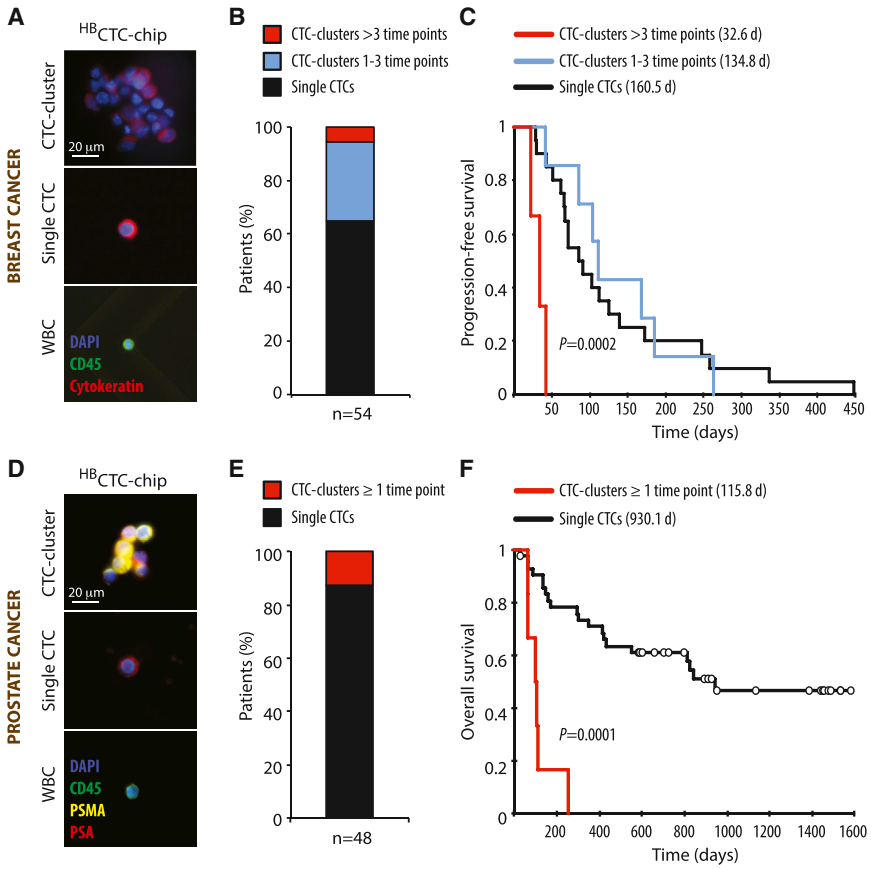

Figure 4. The Presence of CTC Clusters in Patients with Cancer Correlates with Poor Prognosis

(A) Representative images of a CTC cluster, a single CTC, and a white blood cell (WBC) isolated from a breast cancer patient using the HBCTC-Chip and stained with wide-spectrum cytokeratin (CK, red), CD45 (green), and DAPI (nuclei, blue).

(B) A total of 79 breast cancer patients (corresponding to 265 time points) were analyzed for the presence of CTCs, with 54 of the 79 patients scoring positive for CTCs. The bar graph shows the percentage of CTC-positive patients having CTC clusters during more than three time points (red), CTC clusters across one to three time points (blue) or single CTCs only (black).

(C) Kaplan-Meier progression-free survival plot showing progression rates for breast cancer patients having CTC clusters during more than three time points (red), CTC clusters across one to three time points (blue) or single CTCs only (black). The mean progression-free survival time for each group is given in parentheses. p = 0.0002 by log rank test.

(D) Representative images of a CTC cluster, a single CTC and a white blood cell (WBC) isolated from a prostate cancer patient using the HBCTC-Chip and stained with prostate-specific antigen (PSA, red), prostate-specific membrane antigen (PSMA, yellow), CD45 (green), and DAPI (nuclei, blue).

(E) A total of 64 prostate cancer patients (corresponding to 202 time points) were analyzed for the presence of CTCs, with 48 of the 64 patients scoring positive for CTCs. The bar graph shows the percentage of CTC-positive patients having CTC clusters during at least one time point (red) or single CTCs only (black).

(F) Kaplan-Meier overall survival plot showing progression rates for prostate cancer patients having CTC clusters during at least one time point (red) or single CTCs only (black). The mean overall survival time for each group is given in parentheses. p = 0.0001 by log rank test.

See also Table S1.

types, including T cells, B cells, dendritic cells, natural killer cells, hematopoietic stem cells, macrophages/monocytes, granulocytes, endothelial cells, or fibroblasts (Figure S4). Markers for platelets were present in both single CTCs and CTC clusters, consistent with their known adherence to cancer cells in the circulation. For each patient, we compared gene expression data of CTC clusters versus single-CTCs, generating a list of 31 CTC-cluster-associated genes shared across different patients ($q < 0.01$, $\log2FC > 1$, in more than 70% of all intrapatient comparisons) (Figures 5C and 5D; Table S2). To identify potential drivers of metastasis among CTC-cluster-enriched genes, we tested for correlation between their overexpression in primary tumor specimens and clinical outcomes in a cohort of 1,956 patients with ER-positive, HER2-positive, and triple-negative breast cancers. Among the candidate CTC cluster genes, plakoglobin was unique in its high level of overexpression in CTC clusters compared with single CTCs (219-fold) and the fact that its expression in primary tumors associated with a significantly reduced distant metastasis-free survival ($p = 0.008$) (Figures 5D, 6A, and S5). We therefore selected plakoglobin as a CTC-cluster-enriched transcript for more detailed analysis.

Plakoglobin (*JUP*) is a member of the Armadillo family of proteins and an important component of desmosomes and adherence junctions (Aktary and Pasdar, 2012), which has been reported to have both positive and negative roles in diverse malignancies (Hakimelahi et al., 2000; Kolligs et al., 2000; Shiina et al., 2005). Along with upregulation of plakoglobin RNA, multiple components of both desmosomes and adherence junctions were significantly enriched in CTC clusters (Figures S6A–S6E). Consistent with the RNA sequencing results, we confirmed plakoglobin protein expression in multiple CTC clusters, but not in matched single CTCs from a breast cancer patient (Figure 5E). While CTC clusters express epithelial cell junction components, including plakoglobin and E-cadherin, we have previously shown that some mesenchymal markers may also be upregulated in such clusters, an effect that may be associated with adherence in the bloodstream with TGFβ-rich platelets (Labelle et al., 2011; Yu et al., 2013). Matched primary and metastatic tumors biopsies were available from this patient: plakoglobin expression was remarkably heterogeneous in both the primary and metastatic breast tumors, with foci of high expression interspersed with regions without detectable protein (Figure 6B). Thus, while plakoglobin is a key component of intercellular junctions, its variable expression levels within primary tumors raises the possibility that it might demarcate tightly adherent groups of cells that may constitute precursors to CTC clusters.

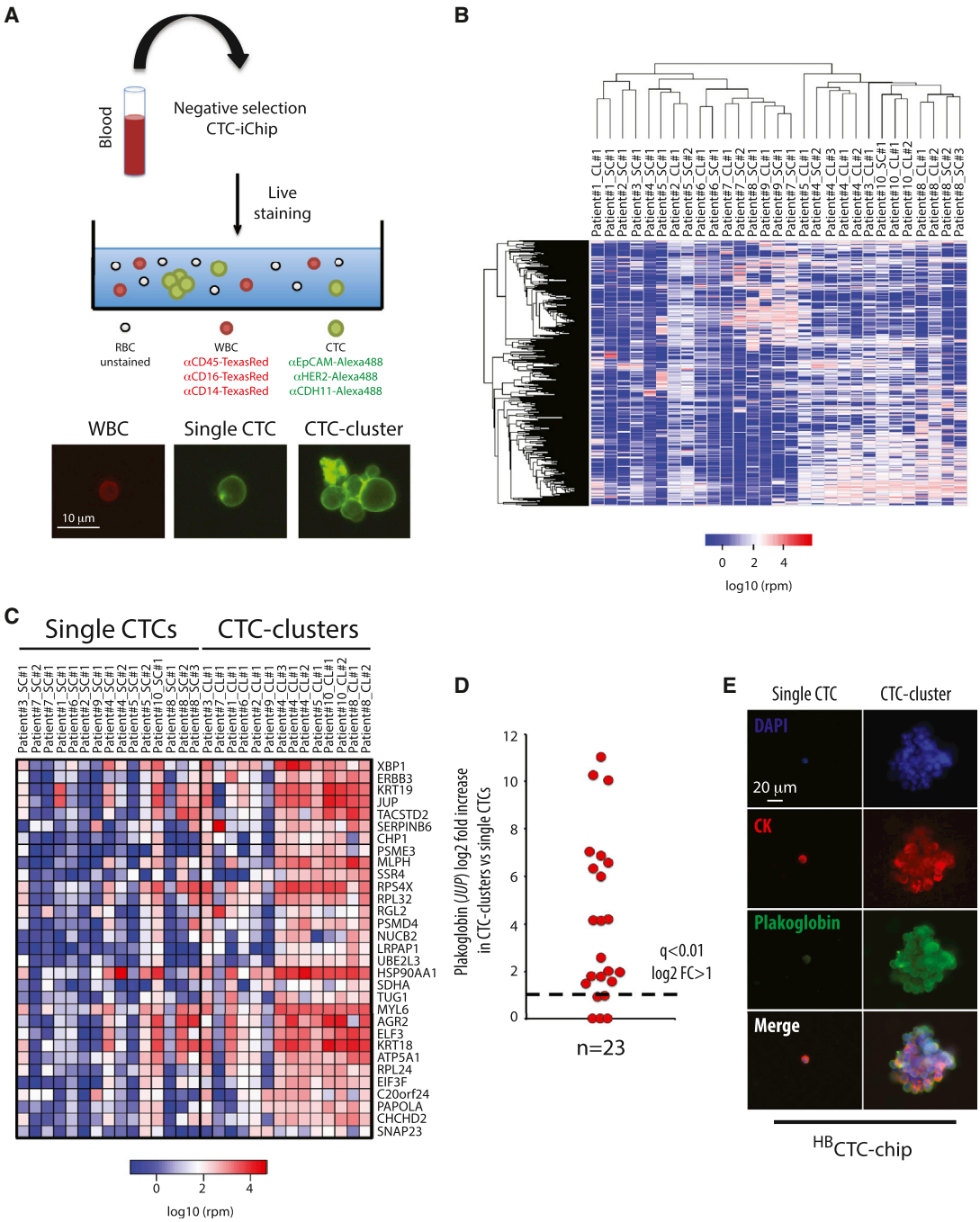

Figure 5. RNA Sequencing of CTC Clusters and Single CTCs Reveals a CTC-Clusters-Associated Gene Set

(A) Illustration of the experimental setup (top). Representative images of a labeled white blood cell (WBC, red), a single CTC and a CTC cluster (green) (bottom) are shown.

(B) Heatmap showing unsupervised hierarchical clustering of 15 single CTCs pools and matched 14 CTC clusters isolated from ten breast cancer patients (SC, single CTCs; CL, CTC cluster).

(C) Heatmap showing the top 31 transcripts upregulated in CTC clusters. n = 15 for single CTCs and n = 14 for CTC clusters; q < 0.01, log2 fold change (FC) >1 in more than 70% intrapatient comparisons (SC, single CTCs; CL, CTC cluster).

(D) Graph showing log2 fold increase in plakoglobin (*JUP*) for each comparison between matched CTC clusters versus single CTCs. The threshold line represents a q < 0.01 and log2 fold increase >1.

(E) Representative images of a single CTC and a CTC cluster captured on the ^{HB}CTC-Chip from a breast cancer patient and stained with wide-spectrum cytokeratin (CK, red), plakoglobin (green), and DAPI (nuclei, blue).

See also Figure S4 and Table S2.

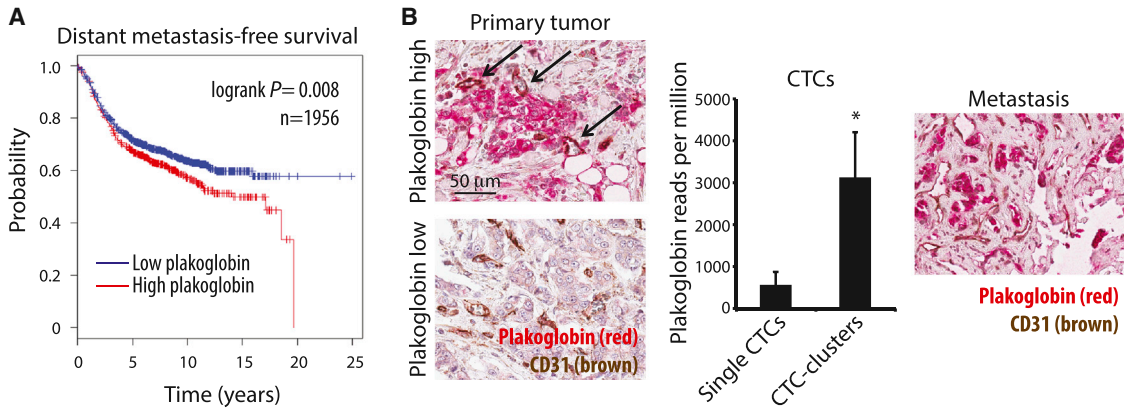

Figure 6. Plakoglobin Expression Correlates with Decreased Distant Metastasis-Free Survival

(A) Kaplan-Meier distant metastasis-free survival plot showing progression rates for patients whose primary tumor expressed either "low plakoglobin" or "high plakoglobin" transcript. n = 1,956; p = 0.008 by log rank test.

(B) Representative images of plakoglobin (red) and CD31 (blood vessels, brown)-stained tissue sections of matched primary tumor (left) and bone metastasis (right) from a hormone receptor-positive breast cancer patient with high CTC cluster counts. Arrows indicate blood vessels in "high plakoglobin" regions. Nuclei are stained with hematoxylin. The bar graph (middle) shows plakoglobin reads per million in matched single CTCs and CTC clusters isolated from the same patient. Error bars represent SEM. n = 3; *p = 0.031.

See also Figures S5 and S6.

Plakoglobin Is Required for CTC Cluster Formation and Contributes to Breast Cancer Metastasis

To define the functional consequences of plakoglobin expression in the context of CTC clusters, we first applied an in vitro assay (Vybrant), which utilizes a fluorogenic dye to measure cell-to-cell adhesion under a variety of culture conditions (El Khoury et al., 1996). We compared seven breast cancer cell lines (MDA-MB-231-LM2, BT474, MCF7, T47D, BT549, BT20, and ZR-75-1) with two nontransformed human mammary epithelial cells (HMEC and MCF10A), following stable lentiviral-mediated plakoglobin knockdown. shRNA-mediated plakoglobin suppression triggered disruption of cell-cell contacts in 6/7 breast cancer lines grown as a monolayer, while it had no detectable effect in either of the two nontransformed breast epithelial cells ($p < 0.04$) (Figures 7A, S7A, and S7B). Thus, breast cancer cells may be more dependent on plakoglobin-mediated cell junctions than normal epithelial cells, which may benefit from additional or alternative pathways in forming intercellular connections (Alford and Taylor-Papadimitriou, 1996; Cavallaro and Christofori, 2004).

To extend these observations in vivo, we introduced either plakoglobin shRNAs or nontarget controls into GFP-*Luciferase*-tagged LM2 and BT474 cells and prepared these as single cells (SC) or clusters (CL) for tail vein injection into immunosuppressed mice (Figure S7C). Consistent with our previous results, both LM2 and BT474 cells expressing control shRNAs showed dramatically increased persistence in the lung when prepared under CL versus SC conditions. In contrast, despite CL conditions, plakoglobin knockdown in both LM2 and BT474 cells dissociated clusters into single cells, consistent with the requirement for plakoglobin for intercellular adhesion in these cells. Following plakoglobin knockdown, tail vein inoculation of CL and SC preparations of both LM2 and BT474 were comparable in producing a reduced number of lung metastases (Figures 7B and S7C). Thus, plakoglobin knockdown abrogates intercellular interactions required to generate clustered cancer cells, thereby reducing their potential to produce lung foci after direct intravascular injection.

Finally, we generated orthotopic xenografts, injecting LM2-GFP-*Luciferase* cells expressing either control or plakoglobin shRNAs into the mammary fat pad of immunodeficient mice and measuring tumor growth as well as tumor-derived CTCs. Plakoglobin knockdown did not alter the primary tumor growth rate, measured for up to 30 days (Figures 7C and S7D), nor did it affect the total number of single CTCs derived from the primary tumor (Figure 7D). Remarkably, the number of tumor-derived CTC clusters was significantly reduced in mice bearing LM2 plakoglobin shRNA-expressing tumors compared to control mice (Figure 7D). In parallel, bioluminescence imaging of mouse lungs demonstrated a striking 80% reduction in lung nodules for mice bearing plakoglobin-suppressed primary tumors (Figure 7E).

Together, our data suggest a model whereby plakoglobin-expressing regions within a primary tumor produce aggregated tumor cells, i.e., CTC clusters, that are shed into the bloodstream, where they demonstrate rapid clearing at distant sites and enhanced metastatic potential (Figure 7F). This mechanism of metastatic spread and the possibility that CTC clusters may be targeted therapeutically through disruption of cell-cell junctions provide an opportunity for strategies to reduce the metastatic spread of breast cancer.

DISCUSSION

By applying microfluidic CTC isolation technologies to both patients with breast cancer and mouse models, we have characterized CTC clusters, a striking but poorly understood feature of bloodborne metastasis. CTC clusters have been observed in patients with cancers of different origin and using multiple

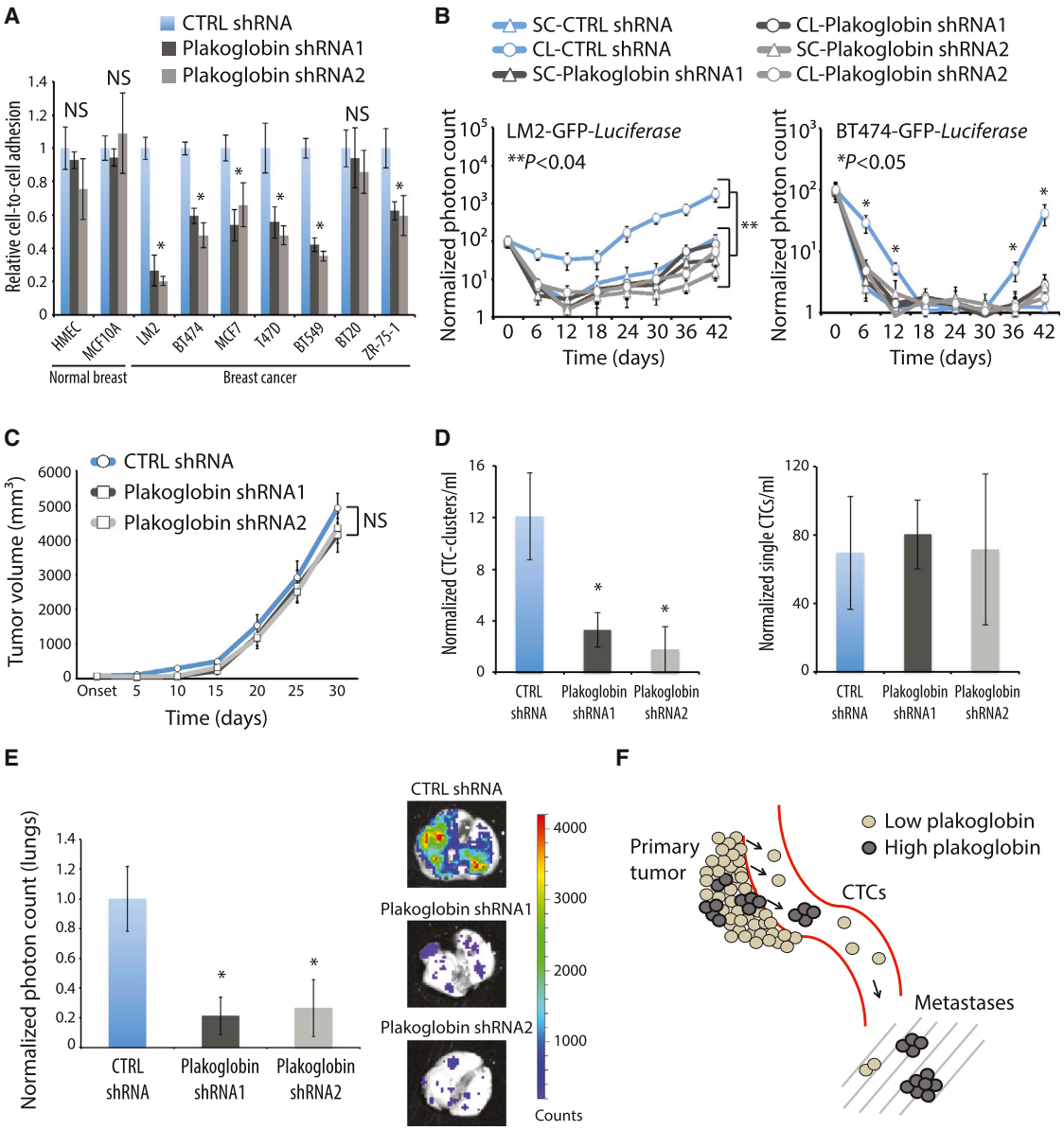

Figure 7. Plakoglobin Is Required for CTC Cluster Formation and Lung Metastasis

(A) Bar graph showing the relative cell-to-cell adhesion in a panel of mammary epithelial cells and breast cancer cell lines grown in the presence or absence of plakoglobin. Error bars represent SEM. n = 5; *p < 0.04.

(B) Lung metastasis growth curves from mice injected with LM2-GFP-*Luciferase* (left) or BT474-GFP-*Luciferase* (right) cells expressing control or plakoglobin shRNAs and prepared as single cells (SC) or clusters (CL). Error bars represent SEM. n = 4; *p < 0.05, **p < 0.04 by Student's t test.

(C) LM2-GFP-*Luciferase* tumor growth curves in the presence or absence of plakoglobin. n = 4; NS, not significant.

(D) Bar graphs showing the normalized number of CTC clusters (left) and single CTCs (right) per ml of blood. Blood samples were isolated 4 weeks after primary tumor development and processed with the ^{HB}CTC-Chip. Error bars represent SEM. n = 4; *p < 0.05 by Student's t test.

(E) Bar graph showing normalized lung photon counts from mice bearing a LM2-GFP-*Luciferase* control or plakoglobin knockdown primary tumor for 4 weeks. Error bars represent SEM. n = 4; *p < 0.045 by Student's t test.

(F) Schematic showing that "high plakoglobin" regions in the primary tumor are likely to generate CTC clusters with increased metastatic potential.

See also Figure S7.

technologies (Cho et al., 2012; Fidler, 1973; Liotta et al., 1976; Molnar et al., 2001; Stott et al., 2010; Yu et al., 2013). While most clusters are relatively small, some comprise dozens of tumor cells, raising the question of how they navigate through normal capillaries. Our in vivo flow cytometry studies indicate that clusters are more rapidly cleared from the circulation than single CTCs. Nonetheless, both the structural deformability of the aggregated cells within these clusters and the presence of

vascular shunts within the circulation may allow a subset of these to circulate. The rapid clearance of clusters within distal tissues, together with their potentially increased cellular viability may underlie their dramatically enhanced metastatic potential. The increased metastatic propensity of CTC clusters in reconstituted mouse models, together with the adverse prognosis of breast and prostate cancer patients with abundant CTC clusters, support an important role for these cellular aggregates in the bloodborne spread of cancer.

Based on cellular tagging and mixing studies in the mouse, almost all CTC clusters appear to be of oligoclonal origin, rather than being derived from the progeny of a single migratory cell. Our studies exclude intravascular aggregation of CTCs as a significant cause for CTC clusters, demonstrating instead that they originate from a single tumor. We cannot determine whether their entry into the vascular space results from "grouped migration," an active invasive process that has been described for epithelial cell masses (Friedl and Gilmour, 2009), or passive shedding into compromised tumor vasculature. Interestingly, the high expression of plakoglobin within foci of cells within the primary tumor raises the possibility that these demarcate the origin of clusters that ultimately enter the circulation. In mouse reconstitution models, plakoglobin knockdown in cells that constitute the primary tumor does not suppress tumorigenesis itself, but it abrogates the generation of CTC clusters in the circulation and greatly reduces the number of metastatic deposits in the lung.

The identification of specific transcripts that enhance the metastatic potential of tumor cells may enable therapeutic strategies to suppress the bloodborne spread of cancer, a critical although challenging goal. To date, candidate metastasis genes have been derived primarily from mouse tumor models. Some, like inducers of EMT, alter the migratory properties of breast cancer epithelial cells and confer stem-like properties (Mani et al., 2008). In human breast cancer CTCs, we recently documented marked enrichment for mesenchymal transcripts in CTCs, using quantitative RNA-in situ hybridization (Yu et al., 2013). In addition to generalized migratory properties associated with EMT, tissue-specific tropism studies in the mouse have identified subsets of genes involved in breast cancer metastases to lung (e.g., Epiregulin, CXCL1, SPARC, and MMP2) (Minn et al., 2005), brain (e.g., COX2, HBEGF, and ST6GALNAC5) (Bos et al., 2009), and bone (mainly driven by Src activation) (Zhang et al., 2009). A recent study interrogating candidate genes in breast CTCs derived from a patient with breast cancer has suggested that coexpression of EpCAM, CD44, CD47, and MET identifies a subset with increased metastatic capacity (Baccelli et al., 2013). While these candidate metastasis genes were not upregulated in CTC clusters compared with single CTCs, our study was not designed to compare CTCs with their matched primary tumors. Finally, we note that our RNA expression studies focused on clusters comprised of tumor cells, because the ^{neg}CTC-iChip favors isolation of small CTC clusters. We therefore cannot exclude additional contributions from nonmalignant cells within larger microemboli (Duda et al., 2010; Labelle et al., 2011; Stott et al., 2010), including potential stromal-derived tropism signals (Zhang et al., 2013).

Our study identifies mediators of metastasis by comparing two distinct populations of circulating tumor cells, one with very high metastatic potential (CTC clusters) compared with the other (single CTCs). The development of advanced microfluidic CTC isolation technology (Ozkumur et al., 2013) enabled us to undertake such a detailed study of human breast cancer cells as they transiently circulate in the bloodstream of patients with metastatic disease. Single-cell resolution RNA sequencing demonstrated a very high level of concordance in expression patterns between matched CTC clusters and single CTCs from individual breast cancer patients. We identified only a small number of candidate genes with significantly divergent expression (Table S2), including transcriptional regulators (XBP1), signaling molecules (AGR2 and HER3), and plakoglobin. While we focused this study on the functional characterization of plakoglobin due to the clinical association between high plakoglobin expression and adverse outcome in patients with breast cancer, additional CTC-cluster-associated genes may be involved in their generation and their metastatic potential. The striking consequences of plakoglobin knockdown, suppressing both CTC cluster generation and metastatic tumor formation in mouse models, point to this gene product being a major determinant of tumor dissemination. Plakoglobin contributes to both adherens junctions and desmosomes: in adherens junctions, the C-terminal intracellular domain of E-cadherin interacts in a mutually exclusive manner with either β-catenin or plakoglobin, which in turn associates with the actin-binding protein α-catenin (Harris and Tepass, 2010). At desmosomes, the intracellular domains of desmocolin and desmoglein interact with plakophilin and plakoglobin, which in turn binds the intermediate filament binding protein desmoplakin (Garrod and Chidgey, 2008). Thus, plakoglobin is a critical constituent of both adherens junctions and desmosomes, a role that may underlie its unique contribution to cell-to-cell adhesion in tumor cells. While plakoglobin has been implicated as both oncogene and tumor suppressor in different contexts (Hakimelahi et al., 2000; Kolligs et al., 2000; Shiina et al., 2005), it is neither in the model proposed here, functioning instead as an intercellular tether that confers added metastatic potential to tumor cells as they break off into the circulation. Interestingly, plakoglobin knockdown has far less impact on intercellular connections of nontransformed breast epithelial cells, which may benefit from additional adhesion mechanisms. This differential effect may offer an opportunity for therapeutic intervention.

In summary, our studies of CTCs in both breast and prostate cancer patients and mouse models point to CTC clusters as critical mediators of cancer metastasis. These coexist with single migratory CTCs, making a contribution to the metastatic burden that far exceeds their comparatively small numbers in the circulation. The ability of tumor cell aggregates to detach from a primary tumor and maintain their cohesion as they survive in the bloodstream may identify a novel and potentially targetable step in the bloodborne dissemination of cancer.

EXPERIMENTAL PROCEDURES

CTC Capture and Identification

Blood specimens for CTC analysis were obtained after informed patient consent, per institutional review board (IRB) protocol (05-300), at the Massachusetts General Hospital. A maximum of 20 ml of blood was drawn in EDTA vacutainers. Within 4 hr from blood draw, ~3 ml of blood was processed through the ^{HB}CTC-Chip or 6–12 ml of blood was processed through the ^{neg}CTC-iChip.

For mouse studies, blood was retrieved via cardiac puncture and ~1 ml of blood was processed through the ^{HB}CTC-Chip.

^{HB}CTC-Chips were manufactured on site at the Massachusetts General Hospital Cancer Center/BioMEMS Resource Facility. For patient samples and mouse xenografts, chips were functionalized as previously described (Yu et al., 2013) with a cocktail of 10 μg/ml each of biotinylated antibodies against EpCAM (R&D Systems), EGFR (Cetuximab, Lilly), and HER2 (R&D Systems). For 4T1 mouse mammary tumor cells, chips were functionalized with a cocktail of antibodies against mouse EpCAM (BioLegend) and EGFR (Cetuximab, Lilly). Samples from patients with prostate cancer were processed as described (Miyamoto et al., 2012). ^{neg}CTC-iChips were designed and fabricated as previously described (Ozkumur et al., 2013) (see Extended Experimental Procedures).

Tumorigenesis Assays

All mouse experiments were carried out in compliance with institutional guidelines. For tail vein experiments, NOD SCID Gamma (NSG) mice (Jackson Labs) were injected with 2×10^5 LM2 cells, 4×10^5 BT474 cells, or 2×10^5 4T1 cells and monitored with IVIS Lumina II (Caliper LifeSciences). For CTC clusters metastatic potential assessment, 2×10^6 LM2-GFP (or 4T1-GFP) and 2×10^6 LM2-mCherry (or 4T1-mCherry) cells were prepared separately or mixed 1:1, suspended in 100 μl of 50% Basement Membrane Matrix Phenol Red-free (BD Biosciences) in PBS and injected orthotopically in NSG mice. Blood draw for CTCs enumeration was performed 5 weeks after tumor onset. For plakoglobin knockdown experiments, 1×10^6 LM2-CTRL or LM2-Plakoglobin shRNA cells were suspended in 100 μl of 50% Basement Membrane Matrix Phenol Red-free in PBS and injected orthotopically in NSG mice. Blood draw for CTCs enumeration and lung metastasis analysis were performed 4 weeks after tumor onset.

Analysis of RNA Sequencing Data

Determination of reads-per-million (rpm): color space reads were aligned using tophat and bowtie1 with the no-novel-juncs argument set with human genome version hg19 and transcriptome defined by the hg19 knownGene table from http://genome.ucsc.edu. Reads that did not align or aligned to multiple locations in the genome were discarded. The hg19 table knownToLocusLink from http://genome.ucsc.edu was used to map, if possible, each aligned read to the gene whose exons the read had aligned to. The reads count for each gene was the number of reads that were so mapped to that gene. This count was divided by the total number of reads that were mapped to any gene and multiplied by one million to form the reads-per-million (rpm) count. We used rpm rather than rpkm because we noted a 3′ bias in the alignments. Clustering algorithms are described in the Extended Experimental Procedures.

ACCESSION NUMBERS

The Gene Expression Omnibus accession number for the sequencing data reported in this paper is GSE51827.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2014.07.013.

AUTHOR CONTRIBUTIONS

N.A., D.A.H., and S.M. designed and performed the experiments, analyzed the data, and wrote the manuscript. A.B. and D.T.M. provided clinical samples and analyzed clinical data. M.C.D., M.Y., A.E., H.Z., and B.W.B. performed immunofluorescence staining, CTC isolation from patient and mouse blood samples, RNA amplification, and library preparation. B.S.W. and S.R. analyzed the RNA sequencing data. J.A.S., A.P., and C.P.L. performed the in vivo flow cytometry. R.K., S.L.S., T.S., D.T.T., and M.T. provided the CTC isolation technology and the single-cell RNA sequencing platform.

ACKNOWLEDGMENTS

We express our gratitude to all the patients who participated in this study. We thank C. Hart, A. McGovern, K. Harrington, L.C. Davis, and the Massachusetts General Hospital (MGH) clinical research coordinators for help with the clinical studies; Drs. P.S. Spuhler and T.A. Barber for support with the CTC-iChip technology; L. Libby and J. Brockmann for excellent technical support. This work was supported by grants from the Breast Cancer Research Foundation (D.A.H.), Stand Up to Cancer (D.A.H., M.T., S.M.), National Foundation for Cancer Research (D.A.H.), Howard Hughes Medical Institute (D.A.H.), NIH CA129933 (D.A.H.), National Institute of Biomedical Imaging and Bioengineering (NIBIB) EB008047 (M.T., D.A.H.), Susan G. Komen for the Cure KG09042 (S.M.), ESSCO Breast Cancer Fund (S.M.), NCI Federal Share Program and Income (S.M.), NIH P41 E8015903-02S1 (C.P.L.), NIH P50 CA086355-12 (C.P.L.), and the MGH-Johnson & Johnson Center for Excellence in CTCs (M.T., S.M.). N.A. is a fellow of the Human Frontiers Science Program, the Swiss National Science Foundation, and the Swiss Foundation for Grants in Biology and Medicine.

Received: October 21, 2013
Revised: February 14, 2014
Accepted: July 9, 2014
Published: August 28, 2014

REFERENCES

Aktary, Z., and Pasdar, M. (2012). Plakoglobin: role in tumorigenesis and metastasis. Int. J. Cell Biol. *2012*, 189521.

Alford, D., and Taylor-Papadimitriou, J. (1996). Cell adhesion molecules in the normal and cancerous mammary gland. J. Mammary Gland Biol. Neoplasia *1*, 207–218.

Alix-Panabières, C., and Pantel, K. (2013). Circulating tumor cells: liquid biopsy of cancer. Clin. Chem. *59*, 110–118.

Baccelli, I., Schneeweiss, A., Riethdorf, S., Stenzinger, A., Schillert, A., Vogel, V., Klein, C., Saini, M., Bäuerle, T., Wallwiener, M., et al. (2013). 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.

Bos, P.D., Zhang, X.H., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., Foekens, J.A., and Massagué, J. (2009). Genes that mediate breast cancer metastasis to the brain. Nature *459*, 1005–1009.

Cavallaro, U., and Christofori, G. (2004). Cell adhesion and signalling by cadherins and Ig-CAMs in cancer. Nat. Rev. Cancer *4*, 118–132.

Cho, E.H., Wendel, M., Luttgen, M., Yoshioka, C., Marrinucci, D., Lazar, D., Schram, E., Nieva, J., Bazhenova, L., Morgan, A., et al. (2012). Characterization of circulating tumor cell aggregates identified in patients with epithelial tumors. Phys. Biol. *9*, 016001.

Duda, D.G., Duyverman, A.M., Kohno, M., Snuderl, M., Steller, E.J., Fukumura, D., and Jain, R.K. (2010). Malignant cells facilitate lung metastasis by bringing their own soil. Proc. Natl. Acad. Sci. USA *107*, 21677–21682.

El Khoury, J., Hickman, S.E., Thomas, C.A., Cao, L., Silverstein, S.C., and Loike, J.D. (1996). Scavenger receptor-mediated adhesion of microglia to beta-amyloid fibrils. Nature *382*, 716–719.

Fidler, I.J. (1973). The relationship of embolic homogeneity, number, size and viability to the incidence of experimental metastasis. Eur. J. Cancer *9*, 223–227.

Fidler, I.J. (2003). The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nat. Rev. Cancer *3*, 453–458.

Friedl, P., and Gilmour, D. (2009). Collective cell migration in morphogenesis, regeneration and cancer. Nat. Rev. Mol. Cell Biol. *10*, 445–457.

Garrod, D., and Chidgey, M. (2008). Desmosome structure, composition and function. Biochim. Biophys. Acta *1778*, 572–587.

Hakimelahi, S., Parker, H.R., Gilchrist, A.J., Barry, M., Li, Z., Bleackley, R.C., and Pasdar, M. (2000). Plakoglobin regulates the expression of the anti-apoptotic protein BCL-2. J. Biol. Chem. *275*, 10905–10911.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Harris, T.J., and Tepass, U. (2010). Adherens junctions: from molecules to morphogenesis. Nat. Rev. Mol. Cell Biol. *11*, 502–514.

Kim, M.Y., Oskarsson, T., Acharyya, S., Nguyen, D.X., Zhang, X.H., Norton, L., and Massagué, J. (2009). Tumor self-seeding by circulating cancer cells. Cell *139*, 1315–1326.

Kolligs, F.T., Kolligs, B., Hajra, K.M., Hu, G., Tani, M., Cho, K.R., and Fearon, E.R. (2000). gamma-catenin is regulated by the APC tumor suppressor and its oncogenic activity is distinct from that of beta-catenin. Genes Dev. *14*, 1319–1331.

Labelle, M., Begum, S., and Hynes, R.O. (2011). Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. Cancer Cell *20*, 576–590.

Ledford, H. (2011). Cancer theory faces doubts. Nature *472*, 273.

Liotta, L.A., Saidel, M.G., and Kleinerman, J. (1976). The significance of hematogenous tumor cell clumps in the metastatic process. Cancer Res. *36*, 889–894.

Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704–715.

Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L., and Massagué, J. (2005). Genes that mediate breast cancer metastasis to lung. Nature *436*, 518–524.

Miyamoto, D.T., Lee, R.J., Stott, S.L., Ting, D.T., Wittner, B.S., Ulman, M., Smas, M.E., Lord, J.B., Brannigan, B.W., Trautwein, J., et al. (2012). Androgen receptor signaling in circulating tumor cells as a marker of hormonally responsive prostate cancer. Cancer Discov *2*, 995–1003.

Molnar, B., Ladanyi, A., Tanko, L., Sréter, L., and Tulassay, Z. (2001). Circulating tumor cell clusters in the peripheral blood of colorectal cancer patients. Clin. Cancer Res. *7*, 4080–4085.

Nagrath, S., Sequist, L.V., Maheswaran, S., Bell, D.W., Irimia, D., Ulkus, L., Smith, M.R., Kwak, E.L., Digumarthy, S., Muzikansky, A., et al. (2007). Isolation of rare circulating tumour cells in cancer patients by microchip technology. Nature *450*, 1235–1239.

Nguyen, D.X., Bos, P.D., and Massagué, J. (2009). Metastasis: from dissemination to organ-specific colonization. Nat. Rev. Cancer *9*, 274–284.

Ozkumur, E., Shah, A.M., Ciciliano, J.C., Emmink, B.L., Miyamoto, D.T., Brachtel, E., Yu, M., Chen, P.I., Morgan, B., Trautwein, J., et al. (2013). Inertial focusing for tumor antigen-dependent and -independent sorting of rare circulating tumor cells. Sci. Transl. Med. *5*, 79ra47.

Robson, E.J., Khaled, W.T., Abell, K., and Watson, C.J. (2006). Epithelial-to-mesenchymal transition confers resistance to apoptosis in three murine mammary epithelial cell lines. Differentiation *74*, 254–264.

Shiina, H., Breault, J.E., Basset, W.W., Enokida, H., Urakami, S., Li, L.C., Okino, S.T., Deguchi, M., Kaneuchi, M., Terashima, M., et al. (2005). Functional Loss of the gamma-catenin gene through epigenetic and genetic pathways in human prostate cancer. Cancer Res. *65*, 2130–2138.

Stott, S.L., Hsu, C.H., Tsukrov, D.I., Yu, M., Miyamoto, D.T., Waltman, B.A., Rothenberg, S.M., Shah, A.M., Smas, M.E., Korir, G.K., et al. (2010). Isolation of circulating tumor cells using a microvortex-generating herringbone-chip. Proc. Natl. Acad. Sci. USA *107*, 18392–18397.

Tang, F., Barbacioru, C., Nordman, E., Li, B., Xu, N., Bashkirov, V.I., Lao, K., and Surani, M.A. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. Nat. Protoc. *5*, 516–535.

Tarin, D., Thompson, E.W., and Newgreen, D.F. (2005). The fallacy of epithelial mesenchymal transition in neoplasia. Cancer Res. *65*, 5996–6000, discussion 6000–6001.

Yu, M., Stott, S., Toner, M., Maheswaran, S., and Haber, D.A. (2011). Circulating tumor cells: approaches to isolation and characterization. J. Cell Biol. *192*, 373–382.

Yu, M., Ting, D.T., Stott, S.L., Wittner, B.S., Ozsolak, F., Paul, S., Ciciliano, J.C., Smas, M.E., Winokur, D., Gilman, A.J., et al. (2012). RNA sequencing of pancreatic circulating tumour cells implicates WNT signalling in metastasis. Nature *487*, 510–513.

Yu, M., Bardia, A., Wittner, B.S., Stott, S.L., Smas, M.E., Ting, D.T., Isakoff, S.J., Ciciliano, J.C., Wells, M.N., Shah, A.M., et al. (2013). Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. Science *339*, 580–584.

Zhang, X.H., Wang, Q., Gerald, W., Hudis, C.A., Norton, L., Smid, M., Foekens, J.A., and Massagué, J. (2009). Latent bone metastasis in breast cancer tied to Src-dependent survival signals. Cancer Cell *16*, 67–78.

Zhang, X.H., Jin, X., Malladi, S., Zou, Y., Wen, Y.H., Brogi, E., Smid, M., Foekens, J.A., and Massagué, J. (2013). Selection of bone metastasis seeds by mesenchymal signals in the primary tumor stroma. Cell *154*, 1060–1073.

CELL 7688

EXTENDED EXPERIMENTAL PROCEDURES

CTC Capture and Identification

Cells captured on the ^{HB}CTC-Chip were fixed with 4% paraformaldehyde and washed with PBS. Fixed cells were then permeabilized with 1% NP40 in PBS, blocked with 3% goat serum/2% BSA, and immunostained with antibodies against wide spectrum cytokeratin (Abcam), prostate specific antigen (DAKO), prostate-specific membrane antigen (obtained from N. Bander), CD45 (Abcam), plakoglobin (Sigma Aldrich) and DAPI. Alternatively, GFP- or mCherry-expressing cells captured on chip were washed with PBS and imaged directly. Stain-positive cells were detected using the BioView Ltd. automated imaging system (Billerica, MA). High-resolution images were obtained with an upright fluorescence microscope (Eclipse 90i, Nikon, Melville, NY).

^{neg}CTC-iChips were designed and fabricated as previously described (Ozkumur et al., 2013). Before processing, whole blood samples were exposed to biotinylated antibodies against CD45 (R&D Systems) and CD66b (AbD Serotec, biotinylated in house) and then incubated with Dynabeads MyOne Streptavidin T1 (Invitrogen) to achieve magnetic labeling and depletion of white blood cells (Ozkumur et al., 2013). The CTC-enriched product was stained in solution with Alexa488-conjugated antibodies against EpCAM (Cell Signaling Technology), Cadherin 11 (R&D Systems) and HER2 (Biolegend) to identify CTCs, and TexasRed-conjugated antibodies against CD45 (BD Biosciences), CD14 (BD Biosciences) and CD16 (BD Biosciences) to identify contaminating white blood cells.

Assessment of Metastasis-free Survival and Overall Survival

Kaplan-Meier survival curves based on clinical data from patients at Massachusetts General Hospital were generated with XLStat software (Addinsoft). For ''plakoglobin high'' versus ''plakoglobin low'' distant metastasis-free survival in breast cancer patients (as well as for the other CTC-clusters-associated genes) we identified publically available human primary breast cancer gene expression data sets and samples within them having the following characteristics: a) distant-metastasis-free survival information was available, b) there was no evidence of neo-adjuvant treatment, c) the platform used to measure gene expression measured at least 10,000 transcripts, d) if there were multiple samples for a patient, only one was used, e) there were at least 40 samples in the data set satisfying the preceding criteria. The following data sets were used (Bos et al., 2009; Chanrion et al., 2008; Chin et al., 2006; Desmedt et al., 2007; Li et al., 2010; Loi et al., 2008; Ma et al., 2004; Minn et al., 2005, 2007; Schmidt et al., 2008; Sotiriou et al., 2006; van 't Veer et al., 2002; van de Vijver et al., 2002; Wang et al., 2005). For each data set, we identified all probes or probesets for plakoglobin and used the one with greatest standard deviation across the samples of the data set. For each data set we characterized a sample as ''high plakoglobin'' if its plakoglobin expression was in the top third of plakoglobin expression for that data set and as ''low plakoglobin'' otherwise. We then created a Kaplan-Meier plot and calculated a logrank two-sided p value using the distant-metastasis-free survival information for the samples from all the data sets and the ''high plakoglobin'' versus ''low plakoglobin'' classification.

Single-Cell Micromanipulation

The CTC-enriched product was collected in a 35mm petri dish and viewed using a Nikon Eclipse Ti inverted fluorescent microscope. Single CTCs and CTC clusters were identified based on intact cellular morphology, Alexa488-positive staining and lack of TexasRed staining. Target cells were individually micromanipulated with a 10 μm transfer tip on an Eppendorf TransferMan NK 2 micromanipulator and ejected into PCR tubes containing RNA protective lysis buffer (10X PCR Buffer II, 25mM MgCl2, 10% NP40, 0.1 M DTT, SUPERase-In, Rnase Inhibitor, 0.5 uM UP1 Primer, 10mM dNTP and Nuclease-free water) and immediately flash frozen in liquid nitrogen.

Single-Cell RNA Amplification and Sequencing

RNA samples extracted from CTCs were thawed on ice and incubated at 70°C for 90 s. To generate cDNA, samples were treated with reverse transcription master mix (0.05 uL RNase inhibitor, 0.07uL T4 gene 32 protein, and 0.33uL SuperScript III Reverse Transcriptase per 1X volume) and incubated on thermocycler at 50°C for 30 min and 70°C for 15 min. To remove free primers, 1.0uL of EXOSAP mix was added to each sample, which was incubated at 37°C for 30 min and inactivated at 80°C for 25 min. Next, a 3′-poly-A tail was added to the cDNA in each sample by incubating in master mix (0.6uL 10X PCR Buffer II, 0.36uL 25mM $MgCl_2$, 0.18uL 100mM dATP, 0.3uL Terminal Transferase, 0.3uL RNase H, and 4.26uL H_2O per 1X volume) at 37°C for 15 min and inactivated at 70°C for 10 min. A second strand cDNA was synthesized by dividing each sample into 4 and incubating in master mix (2.2uL 10X High Fidelity PCR Buffer, 1.76uL 2.5mM each dNTP, 0.066uL UP2 Primer at 100uM, 0.88uL 50mM $MgSO_4$, 0.44uL Platinum Taq DNA Polymerase, and 13.654uL H_2O per 1X volume) at 95°C for 3 min, 50°C for 2 min, and 72°C for 10 min. PCR amplification (95°C for 3 min, 20 cycles of 95°C for 30 s, 67°C for 1 min, and 72°C for 6 min 6 s) was performed with master mix (4.1uL 10X High Fidelity PCR Buffer, 1.64uL 50mM $MgSO_4$, 4.1uL 2.5mM each dNTP, 0.82uL AUP1 Primer at 100uM, 0.82uL AUP2 Primer at 100uM, 0.82uL Platinum Taq DNA Polymerase, and 6.7uL H_2O per 1X volume). The 4 reactions of each sample were pooled and purified using the QIAGEN PCR Purification Kit (Cat. No 28106) and eluted in 50uL EB buffer. Samples were selected by testing for genes Gapdh, ActB, Ptprc (CD45), Krt8, Krt18 and Krt19 using qPCR. Each sample was again divided in 4 and a second round of PCR amplification (9 cycles of 98°C for 3 min, 67°C for 1 min, and 72°C for 6 min 6 s) was performed with master mix (9uL 10X High Fidelity PCR Buffer, 3.6uL 50mM $MgSO_4$, 13.5uL 2.5mM each dNTP, 0.9uL AUP1 Primer at 100uM, 0.9uL AUP2 Primer at 100uM, 1.8uL Platinum Taq DNA Polymerase, and 59.1uL H_2O per 1X volume). Samples were pooled and purified using Agencourt AMPure XP beads and eluted in 40uL 1X low TE buffer.

Sequencing Library Construction

To shear the DNA using the Covaris S2 System, 1X low TE buffer and 1.2uL shear buffer were added to each sample. Conditions of the shearing program include: 6 cycles, 5°C bath temperature, 15°C bath temperature limit, 10% duty cycle, intensity of 5, 100 cycles/burst, and 60 s. Then, samples were end-polished at room temperature for 30 min with a master mix (40uL 5X Reaction Buffer, 8uL 10mM dNTP, 8uL End Polish Enzyme1, 10uL End Polish Enzyme2, and 14uL H_2O per 1X volume). DNA fragments larger than 500bp were removed with 0.5X volumes of Agencourt AMPure XP beads. Supernatant was transferred to separate tubes. To size-select 200-500bp DNA products, 0.3X volumes of beads were added and samples were washed twice with 70% EtOH. The products were eluted in 36uL low TE buffer. A dA-tail was added to each size-selected DNA by treating with master mix (10uL 5X Reaction Buffer, 1uL 10mM dATP, and 5uL A-Tailing Enzyme I per 1X volume) and incubated at 68°C for 30 min and cooled to room temperature. To label and distinguish each DNA sample for sequencing, barcode adaptors (5500 SOLiD 4464405) were ligated to DNA using the 5500 SOLiD Fragment Library Enzyme Module (4464413). Following barcoding, samples were purified twice using the Agencourt AMPure XP beads and eluted in 22uL low TE buffer. Following a round of PCR Amplification (95°C for 5 min, 12 cycles of 95°C for 15 s, 62°C for 15 s, and 70°C for 1 min, and 70°C for 5 min), the libraries were purified with AMPure XP beads. Finally, to quantify the amount of ligated DNA, SOLiD Library TaqMan Quantitation Kit was used to perform qPCR. Completed barcoded libraries were then subjected to emulsion PCR with template beads preparation and sequenced on the ABI 5500XL.

Analysis of RNA Sequencing Data

Clustering: the minimum of 1 and the smallest positive value of the rpm matrix was added to the rpm matrix to eliminate zeros. The result was then log transformed and median polished. The rows (corresponding to genes) with the top 2000 standard deviations were retained and the rest of the rows discarded. The result was clustered using agglomerative hierarchical clustering with average linkage with distance metric equal to 1 minus the Pearson correlation coefficient.

Supervised differential gene expression: samples that showed high expression of contaminant WBC markers and no expression of CTC markers at the RNA level were excluded from the analysis. For each pair of single CTCs sample and CTC cluster sample from the same patient, we calculated a FDR q-value and a normalized fold change using the DEGexp function of version 1.10.0 of the Bioconductor DEGseq package (Wang et al., 2010) with method set to 'MARS' and q-values calculated using Benjamini-Hochberg. For each pair and direction (e.g., up in CTC clusters versus single CTCs) a gene was considered a hit if its q-value was less than 0.01 and its fold change was greater than 2. Then, for each direction, we considered the genes that were hits for 70% or more of the pairs. The desmosome (resp. adherence junction) metagene was defined to be the mean over the desmosome (resp. adherence junction) marker genes of the normalized log2 fold change between the CTC clusters and the single CTCs as determined by DEGseq.

In Vivo Flow Cytometry

DiD-labeled LM2 single or clustered cells were adoptively transferred intravenously and detected in the peripheral circulation (Novak et al., 2004). Of note, single and clustered LM2 cells were injected separately in different animals to avoid signal misinterpretation. DiD was excited by a 635 nm laser and detected with a 695 ± 27.5 nm bandpass filter using a photomultiplier tube. Circulation kinetics for LM2-SCs and LM2-CLs were quantified using MATLAB (Mathworks).

Immunohistochemistry

Formalin-fixed and paraffin embedded mouse xenografts primary tumors, lung metastases, as well as human primary tumors and matched metastatic lesions were sectioned and stained overnight at 4°C with antibodies against cleaved caspase 3 (Cell Signaling Technology), GFP (Cell signaling Technology), mCherry (Abcam), plakoglobin (Sigma Aldrich) and CD31 (Abcam). GFP/mCherry and Plakoglobin/CD31 double-stainings were performed with EnVision G/2 Doublestain System (Dako). All specimens were counterstained with Hematoxylin. Images of the whole tissue were taken with ScanScope (Aperio).

Cell Culture and Reagents

HMEC, MCF10A, BT474, MCF7, T47D, 4T1, BT549, BT20 and ZR-75-1 cells were purchased from the American Type Culture Collection (ATCC) and propagated according to the manufacturer's instructions. MDA-MB-231 LM2 cells were obtained from J. Massague (MSKCC, New York, NY, USA) and propagated in DMEM (Life Technologies) supplemented with 10% fetal bovine serum (Life Technologies). To generate BT474, 4T1 or LM2 single cells or clusters for tail vein injections, cells growing in monolayer at 80% confluence were incubated with trypsin (Life Technologies) for one minute to generate floating clusters. After trypsin neutralization with serum-containing medium, clusters were then distributed equally in two separate dishes. In one of the two dishes, clusters were mechanically dissociated by pipetting to generate a single-cell suspension.

Cell-to-cell adhesion assay was performed with the Vybrant Cell-to-Cell Adhesion Assay Kit (Invitrogen) according to the manufacturer's instructions.

The plasmid expressing GFP-Luciferase was obtained from C. Ponzetto (University of Torino, Italy). The plasmid expressing mCherry was purchased from Addgene. Plakoglobin TRC shRNAs were purchased from Thermo Scientific. Lentiviral packaging vectors (Addgene) were used to transfect 293T cells (ATCC) and produce lentiviral particles. Infections of target cells lines was performed overnight at a MOI = 10 in growth medium containing 8 μg/ml polybrene (Thermo Scientific).

SUPPLEMENTAL REFERENCES

Chanrion, M., Negre, V., Fontaine, H., Salvetat, N., Bibeau, F., Mac Grogan, G., Mauriac, L., Katsaros, D., Molina, F., Theillet, C., and Darbon, J.M. (2008). A gene expression signature that can predict the recurrence of tamoxifen-treated primary breast cancer. Clin. Cancer Res. *14*, 1744–1752.

Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T., et al. (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell *10*, 529–541.

Desmedt, C., Piette, F., Loi, S., Wang, Y., Lallemand, F., Haibe-Kains, B., Viale, G., Delorenzi, M., Zhang, Y., d'Assignies, M.S., et al.; TRANSBIG Consortium (2007). Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin. Cancer Res. *13*, 3207–3214.

Li, Y., Zou, L., Li, Q., Haibe-Kains, B., Tian, R., Li, Y., Desmedt, C., Sotiriou, C., Szallasi, Z., Iglehart, J.D., et al. (2010). Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. Nat. Med. *16*, 214–218.

Loi, S., Haibe-Kains, B., Desmedt, C., Wirapati, P., Lallemand, F., Tutt, A.M., Gillet, C., Ellis, P., Ryder, K., Reid, J.F., et al. (2008). Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. BMC Genomics *9*, 239.

Ma, X.J., Wang, Z., Ryan, P.D., Isakoff, S.J., Barmettler, A., Fuller, A., Muir, B., Mohapatra, G., Salunga, R., Tuggle, J.T., et al. (2004). A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. Cancer Cell *5*, 607–616.

Minn, A.J., Gupta, G.P., Padua, D., Bos, P., Nguyen, D.X., Nuyten, D., Kreike, B., Zhang, Y., Wang, Y., Ishwaran, H., et al. (2007). Lung metastasis genes couple breast tumor size and metastatic spread. Proc. Natl. Acad. Sci. USA *104*, 6740–6745.

Novak, J., Georgakoudi, I., Wei, X., Prossin, A., and Lin, C.P. (2004). In vivo flow cytometer for real-time detection and quantification of circulating cells. Opt. Lett. *29*, 77–79.

Schmidt, M., Böhm, D., von Törne, C., Steiner, E., Puhl, A., Pilch, H., Lehr, H.A., Hengstler, J.G., Kölbl, H., and Gehrmann, M. (2008). The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res. *68*, 5405–5413.

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B., et al. (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J. Natl. Cancer Inst. *98*, 262–272.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J., et al. (2002). A gene-expression signature as a predictor of survival in breast cancer. N. Engl. J. Med. *347*, 1999–2009.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J., et al. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet *365*, 671–679.

Wang, L., Feng, Z., Wang, X., Wang, X., and Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. Bioinformatics *26*, 136–138.

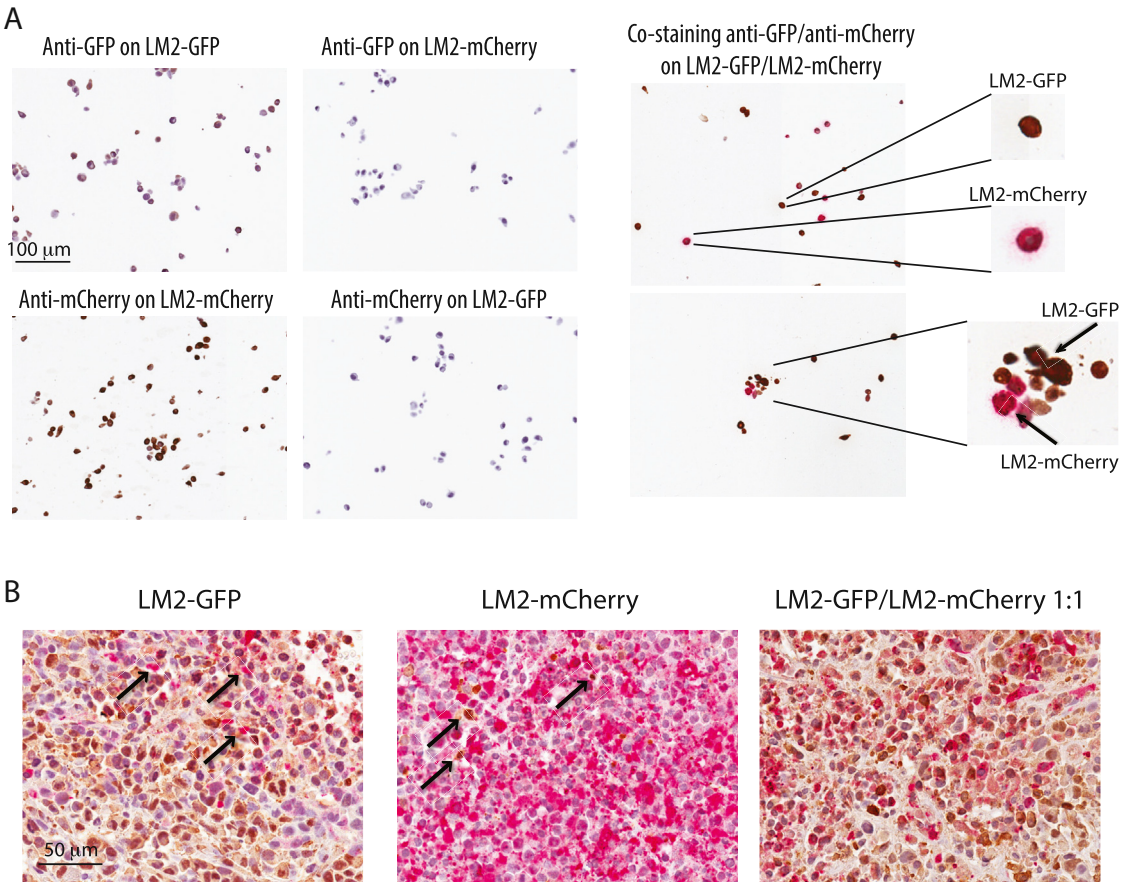

Figure S1. GFP and mCherry Staining of LM2 Mammary Tumors, Related to Figure 1

(A) Representative images of control sections of paraffin-embedded LM2-GFP and LM2-mCherry cells grown separately (left) or as coculture (right) and stained with anti-GFP (brown) and anti-mCherry (red) antibodies.

(B) Representative images of a LM2-GFP, a LM2-mCherry and a LM2-GFP/LM2-mCherry (1:1) primary tumor stained with antibodies anti-GFP (brown) and anti-mCherry (red). The LM2-GFP and LM2-mCherry tumors were present in the same animals. Arrows indicate mCherry-positive cells in the LM2-GFP tumor and GFP-positive cells in the LM2-mCherry tumor, respectively, as examples of tumor self-seeding.

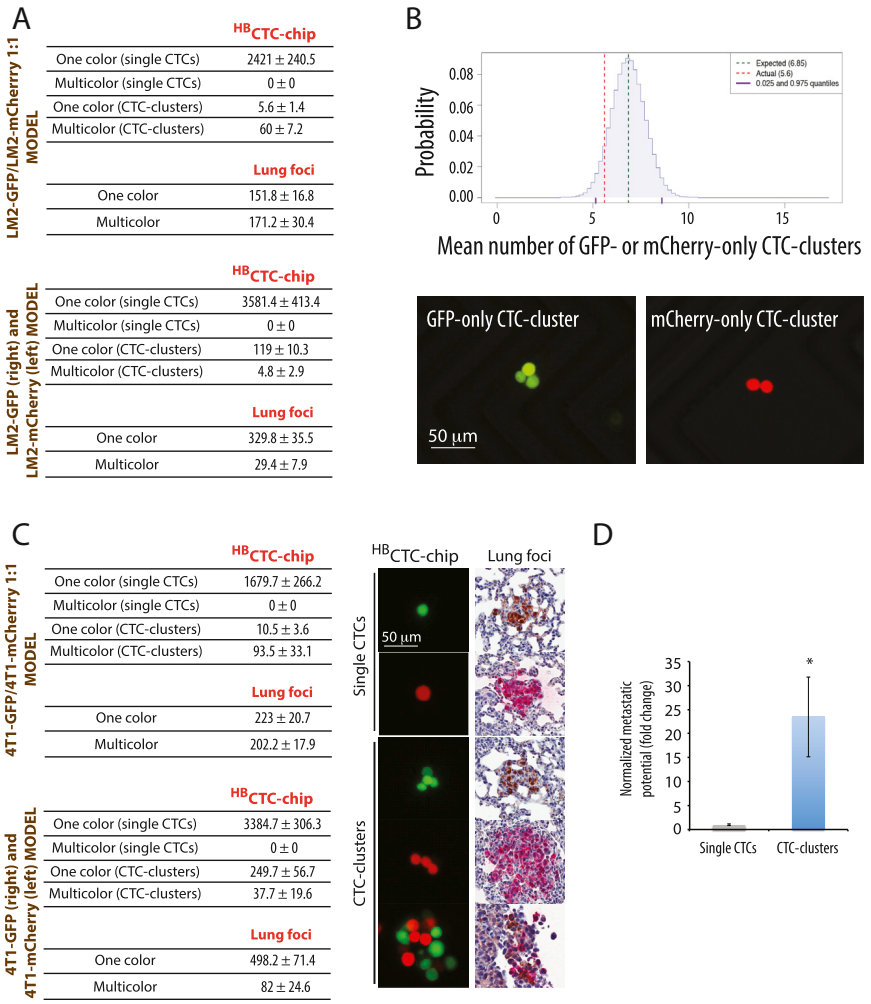

Figure S2. Counts of One-Color and Multicolor Events in the LM2 and 4T1 Xenografts, Related to Figure 1

(A) Table showing counts of one color versus multicolor events within CTCs and lung foci from both the "LM2-GFP/LM2-mCherry 1:1" and the "LM2-GFP (right) and LM2-mCherry (left)" models. Results represent means ± SEM.

(B) Distribution curve describing the expected numbers of GFP- or mCherry-only CTC clusters per mouse given our experimental setup, with the actual value shown as a red dashed line (top). Representative pictures of LM2-GFP-only and mCherry-only CTC clusters captured on the ^{HB}CTC-chip from the blood of mice belonging to the "LM2-GFP/LM2-mCherry 1:1" model (bottom). Blood samples were isolated 5 weeks after primary tumor development.

(C) Table showing counts of one color versus multicolor events within CTCs and lung foci from both the "4T1-GFP/4T1-mCherry 1:1" and the "4T1-GFP (right) and 4T1-mCherry (left)" models. Mice were sacrificed for CTCs and lungs isolation 3 weeks after primary tumor development. Results represent means ± SEM (n = 4) (left). Representative images of one color and multicolor 4T1 CTCs and lung foci (right).

(D) Bar graph showing the normalized metastatic potential of 4T1 single CTCs and CTC clusters. Error bars represent SEM. n = 4, *p < 0.036 by Student's t test.

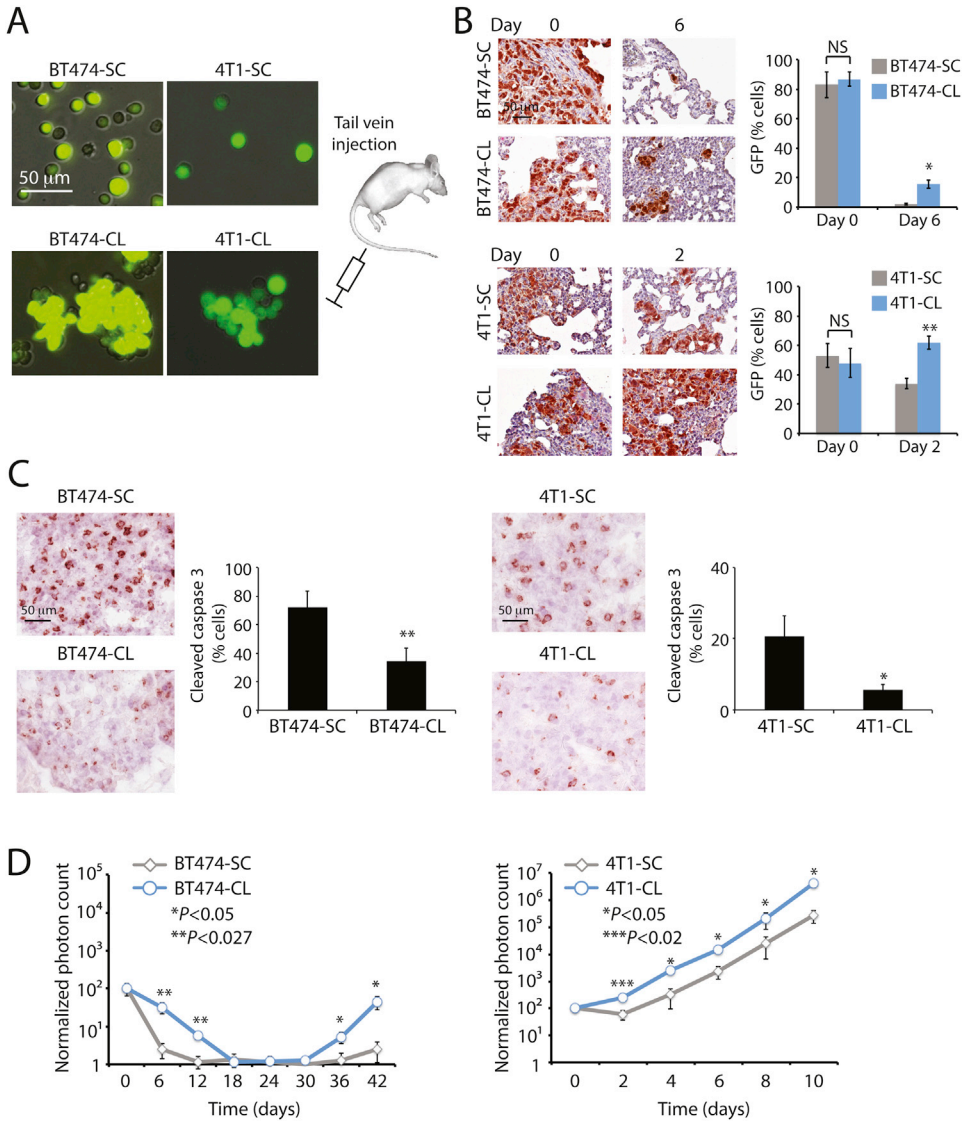

Figure S3. BT474 and 4T1 Clusters Are More Resistant to Apoptosis at Distal Metastatic Sites, Related to Figure 2

(A) Schematic showing BT474- or 4T1-GFP-Luciferase cells prepared as single cells (SC) or as clusters (CL) prior to injection into the tail vein of immunodeficient mice.

(B) Representative images of GFP-stained sections of mouse lungs after injection with BT474 or 4T1 SC versus CL (left). The bar graphs show the mean percentage of GFP-positive cells in lungs from mice injected with BT474 or 4T1 SC versus CL (right). Error bars represent SEM. n = 4; NS = not significant, *p = 0.003 **p = 0.002 by Student's t test.

(C) Representative pictures of cleaved caspase 3-stained sections of mouse lungs 24 hr after injection with BT474 or 4T1 SC versus CL. The bar graph shows the mean percentage of cleaved caspase 3-positive cells in lungs from mice injected with BT474 or 4T1 SC versus CL. Error bars represent SEM. n = 4; *p = 0.037 **p = 0.028 by Student's t test.

(D) Lung metastasis growth curves from mice injected with BT474 or 4T1 SC versus CL. Error bars represent SEM. n = 4; ***p < 0.02 *p < 0.027 *p < 0.05 by Student's t test.

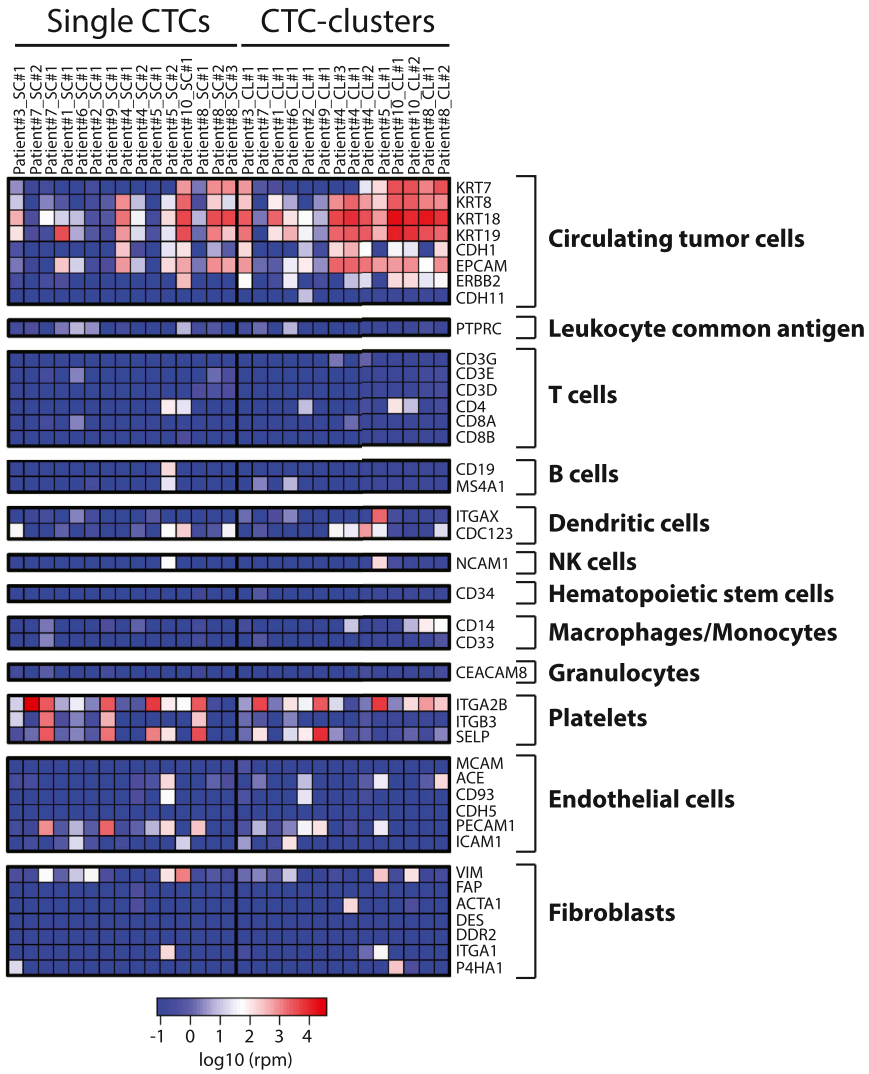

Figure S4. Analysis of the Cellular Composition of CTC Clusters, Related to Figure 5

Heatmap showing expression levels of CTCs-, leukocytes-, T cells-, B cells-, dendritic cells-, natural killer (NK) cells-, hematopoietic stem cells-, macrophages/monocytes-, granulocytes-, platelets-, endothelial cells- and fibroblasts-associated transcripts in the 15 single CTCs and 14 CTC clusters samples used to derive CTC clusters upregulated transcripts.

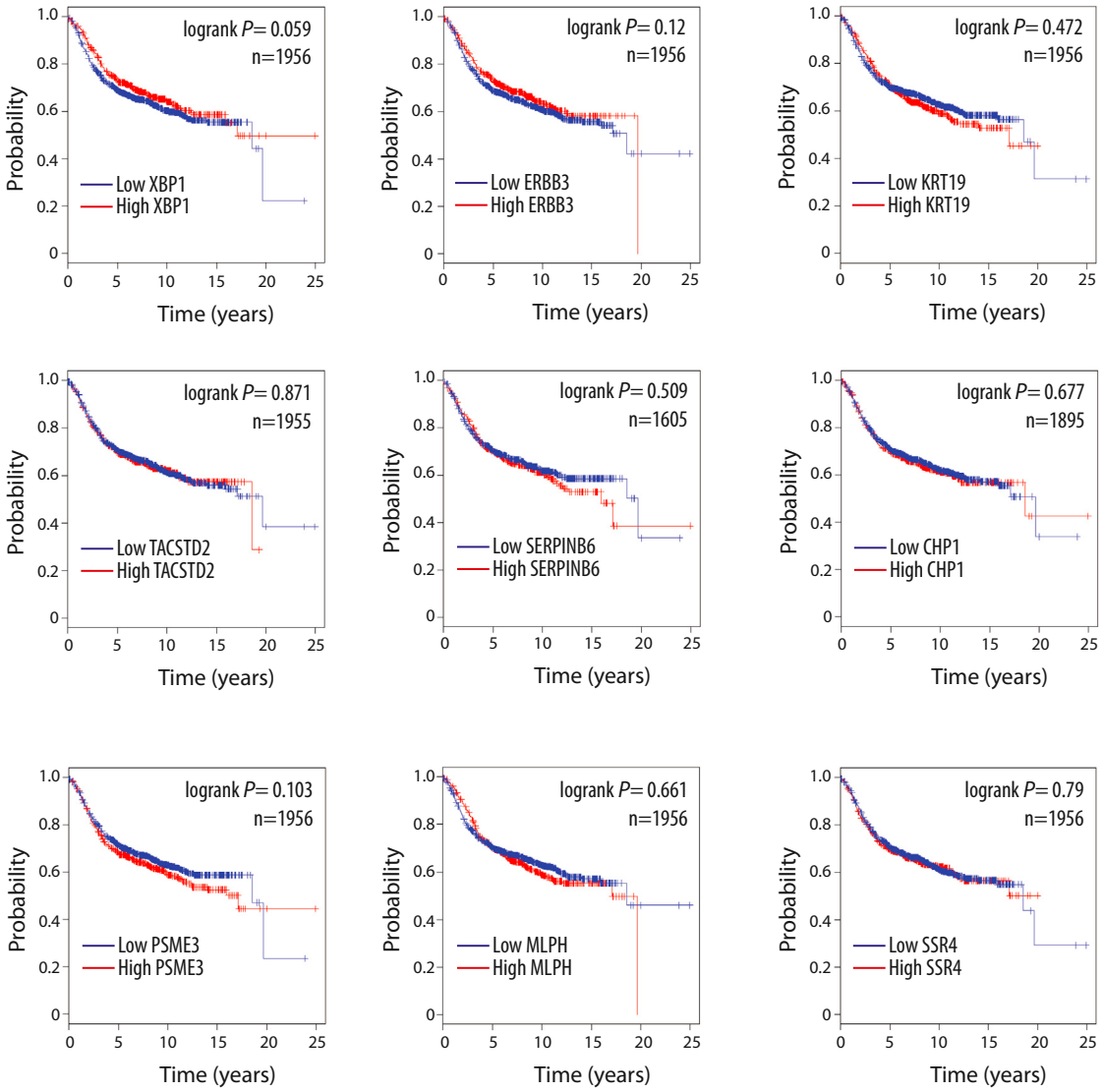

Figure S5. Kaplan-Meier Plots of CTC-Clusters-Associated Genes, Related to Figure 6

Kaplan-Meier distant metastasis-free survival plots showing progression rates for patients whose primary tumor expressed either "low" or "high" levels of the top CTC-clusters-associated marker genes.

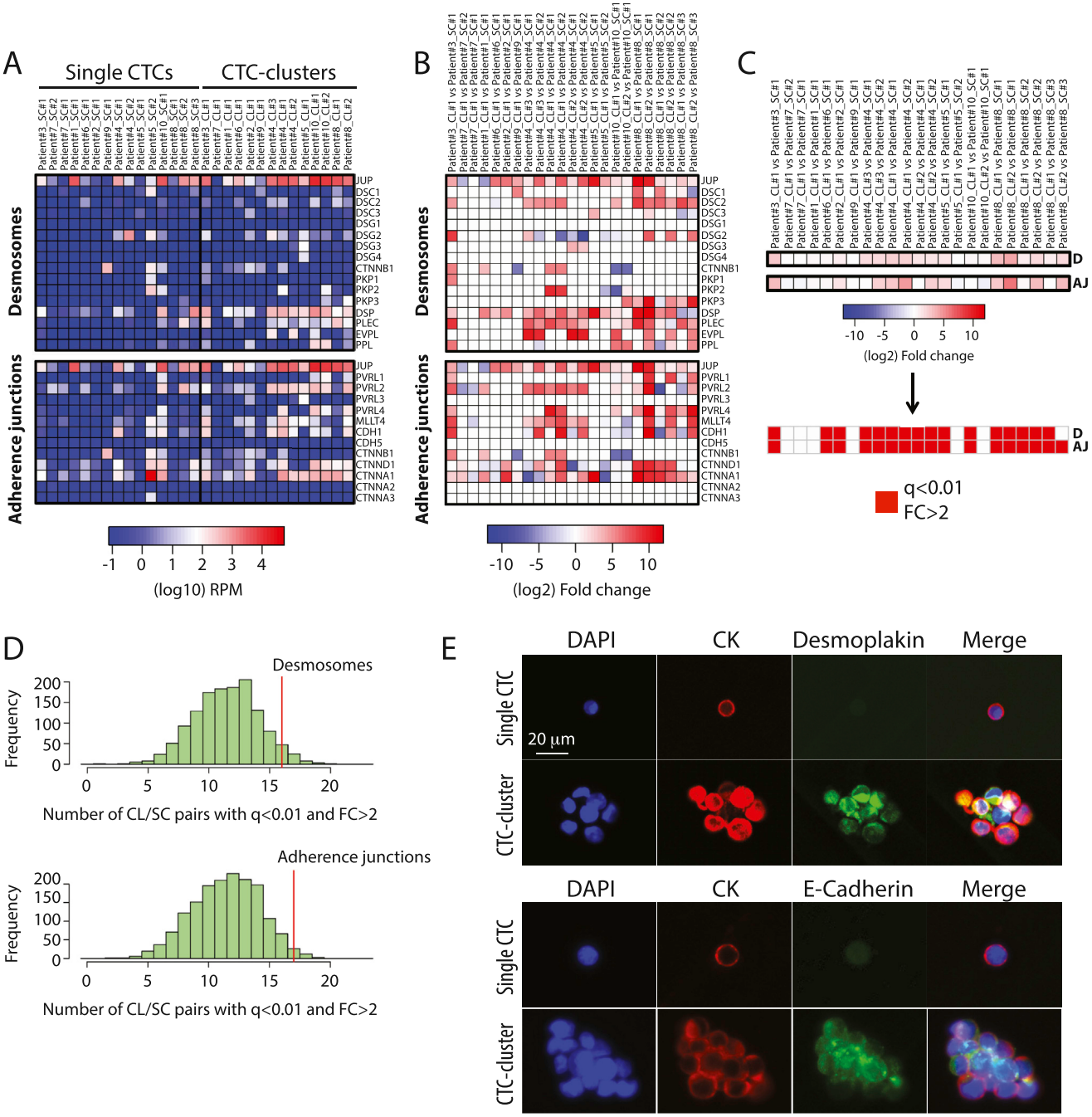

Figure S6. Desmosomes and Adherence Junctions Are Enriched in CTC Clusters, Related to Figure 6

(A) Heatmap showing expression levels of desmosomes (top) and adherence junctions (bottom) marker genes in the 15 single CTCs and 14 CTC clusters samples used to derive CTC clusters upregulated transcripts.

(B) Heatmap representing fold change of desmosomes (top) and adherence junctions (bottom) marker genes in all "CTC clusters vs single CTCs" intrapatient comparisons.

(C) Heatmap showing fold change of desmosome (D) and adherence junction (AJ) metagenes in all "CTC clusters vs single CTCs" intrapatient comparisons.

(D) Representation of the frequency of "CTC clusters vs single CTCs" pairs with $q < 0.01$ and fold change > 2 for randomly generated metagenes of the same size as desmosomes (top) and adherence junctions (bottom). Actual number of "CTC clusters vs single CTCs" pairs with $q < 0.01$ and fold change > 2 for desmosomes (top) and adherence junctions (bottom) metagenes is shown as a red line.

(E) Representative images of a single CTC and a CTC cluster captured on the ^{HB}CTC-Chip from a breast cancer patient and stained with wide-spectrum cytokeratin (CK, red), the desmosome marker desmoplakin (green, top), the adherence junctions marker E-cadherin (green, bottom) and DAPI (nuclei, blue).

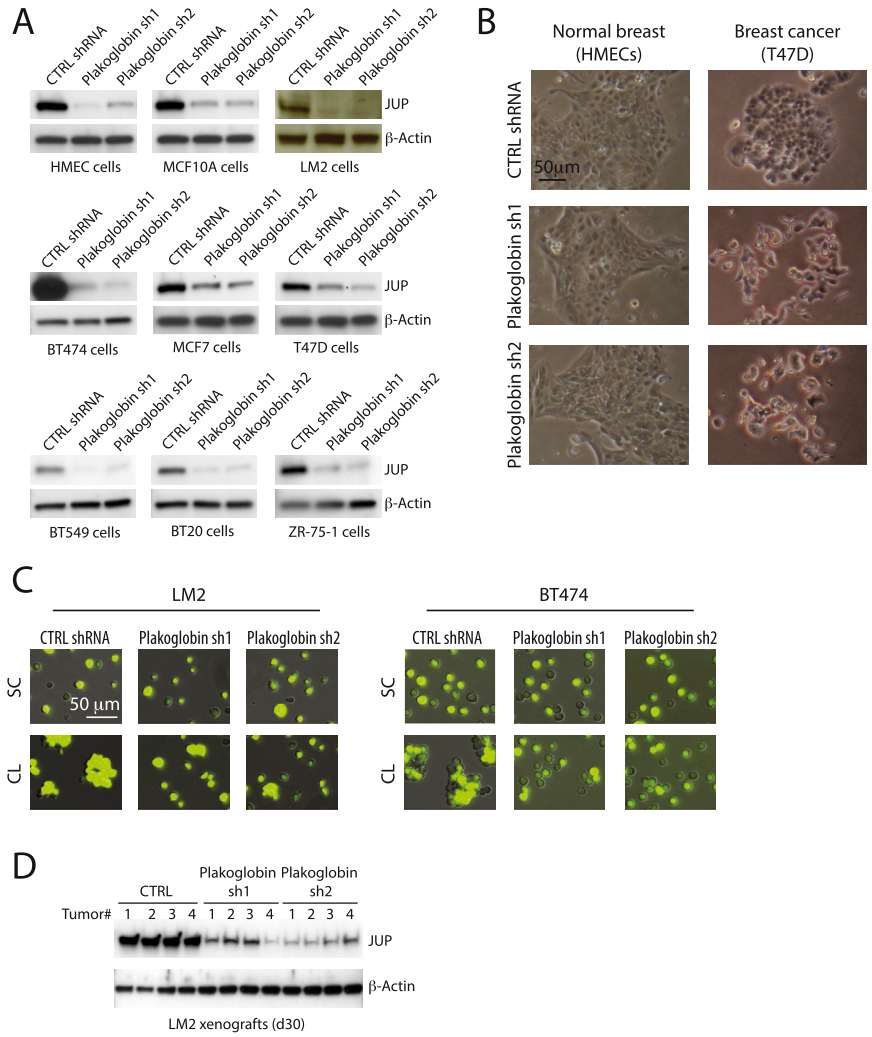

Figure S7. Plakoglobin Is Required for Cancer Cells Clustering, Related to Figure 7

(A) Immunoblot showing the expression levels of plakoglobin and β-Actin (loading control) in lysates from a panel of non-transformed mammary epithelial cells (HMEC, MCF10A) and breast cancer cell lines (LM2, BT474, MCF7, T47D, BT549, BT20, ZR-75-1) expressing control (CTRL) or plakoglobin shRNAs.

(B) Representative brightfield pictures of normal breast cells (HMECs) and breast cancer cells (T47D) grown in monolayer and expressing CTRL or plakoglobin shRNAs.

(C) Representative pictures of LM2-GFP-Luciferase (LM2) and BT474-GFP-Luciferase (BT474) cells expressing CTRL or plakoglobin shRNAs and prepared as single cells (SC) or as clusters (CL) prior to injection into the tail vein of immunodeficient mice. $2x10^5$ LM2 and $4x10^5$ BT474 cells were injected as SC or CL per mouse.

(D) Immunoblot showing the expression levels of plakoglobin and β-Actin (loading control) in lysates from LM2 xenografts expressing CTRL or plakoglobin shRNAs for 30 days.