Cell Reports, Volume *14*

Supplemental Information

Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity

Evelyne Beerling, Daniëlle Seinstra, Elzo de Wit, Lennart Kester, Daphne van der Velden, Carrie Maynard, Ronny Schäfer, Paul van Diest, Emile Voest, Alexander van Oudenaarden, Nienke Vrisekoop, and Jacco van Rheenen

Supplemental figures and information

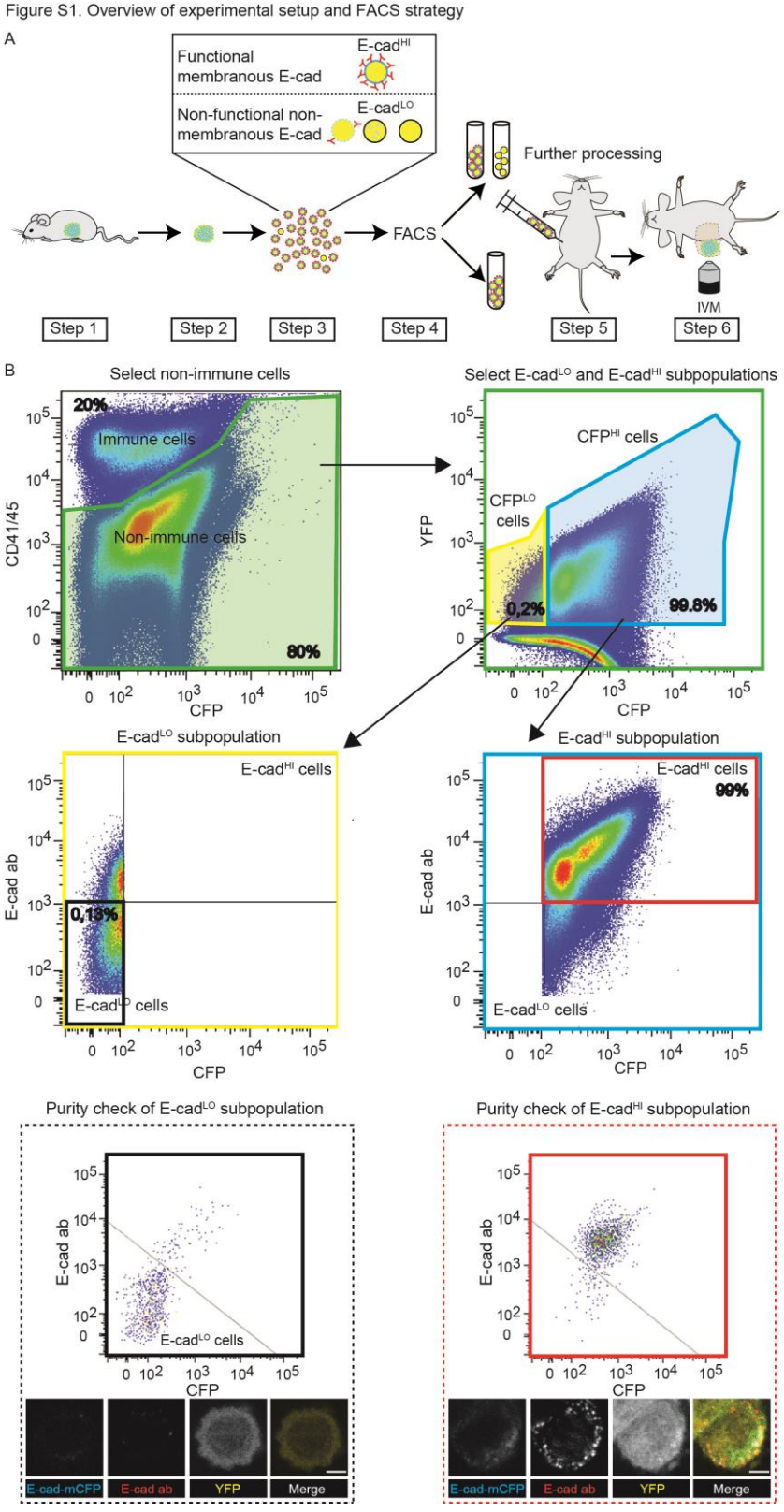

Figure S1. Overview of experimental setup and FACS strategy (related to Figure 2)

(A) Schematic overview of experimental setup: (Step 1) *MMTV-PyMT;MMTV-Cre;R26R-YFP;E-Cad-mCFP* mice with primary mammary tumor. (Step 2) Isolation of primary tumor. (Step 3) Preparation of single cell suspensions and labeling with E-cad antibody (red dotted lines). E-Cad^{HI} cells, cells with high membranous E-cad, will be labeled with the E-cad antibody. E-cad^{LO} cells, cells with low membranous E-cad due to low expression and/or intracellularly localized E-cad, will not be stained by the E-cad antibody. (Step 4) Single cell sorting of E-cad^{HI} and E-cad^{LO} tumor cells using FACS, based on E-cad-mCFP expression/E-cad antibody binding. (Step 5) Transplantation of E-cad^{HI} tumor cells and (Step 6) subsequent IVM.
(B) FACS strategy for E-cad^{LO} and E-cad^{HI} tumor cell sorting. A broad FSC SSC gate was followed by a gate excluding doublets, after which immune cells and megakaryocytes were excluded from the single cell population based on the expression of CD41/45 (upper left plot). YFP^{+} tumor cells (upper right plot) were subdivided in E-cad-mCFP^{HI} (blue frame) and E-cad-mCFP^{LO} (yellow frame) using very stringent gating, percentages are indicated in the frame. Finally, YFP^{+}mCFP^{LO} tumor cells were additionally stringently gated for absence of E-cad antibody expression (black frame; E-cad^{LO}) and YFP^{+}mCFP^{HI} tumor cells for high E-cad antibody expression (red frame; E-cad^{HI}). The sorted cell pools are then checked for purity (lower plots) and imaged using a confocal microscope. Separate channels and merged image are shown for both populations of tumor cells. Scale bar, 2 μm.

Figure S2. The relative mRNA expression of E-cad^{LO} gene set

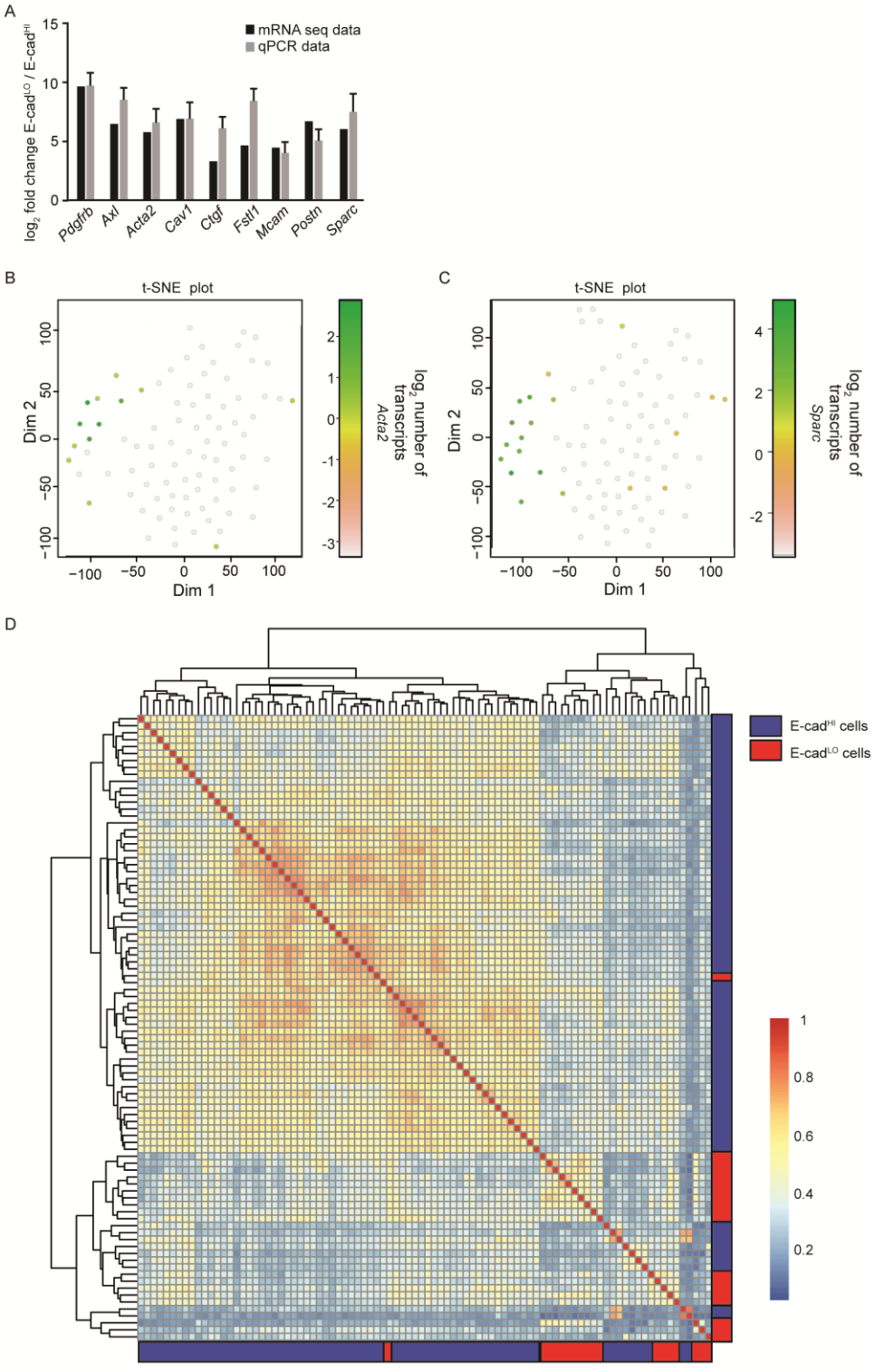

Figure S2. The relative mRNA expression of E-cad^{LO} gene set (related to Figure 2)

(A) Shown are the relative expressions of genes as determined by mRNAseq (black bars) and confirmed by reverse transcriptase (RT) qPCR (grey bars). Mean $\log_2$ fold change is shown of 3 mice, ±SEM.
(B) Single cell sequencing data from E-cad^{HI} and E-cad^{LO} cells represented in a T-distributed stochastic neighbor embedding (t-SNE) intensity plot for Acta2. The scale bar for t-SNE plots is $\log_2$.
(C) Single cell sequencing data from E-cad^{HI} and E-cad^{LO} cells represented in a t-SNE intensity plot for Sparc. The scale bar for t-SNE plots is $\log_2$.
(D) Heat map was made based on Pearson's correlation between individual cells after filtering out genes that had less than 5 transcripts in at least 1 cell. The expression profiles of E-cad^{HI} cells show a Pearson correlation of 0.45 to each other and a Pearson correlation of 0.35 to E-cad^{LO} cells (Wilcoxon 2-sided test = $p<10^{-15}$).

Figure S3. E-cad^{LO} tumor cells show plasticity

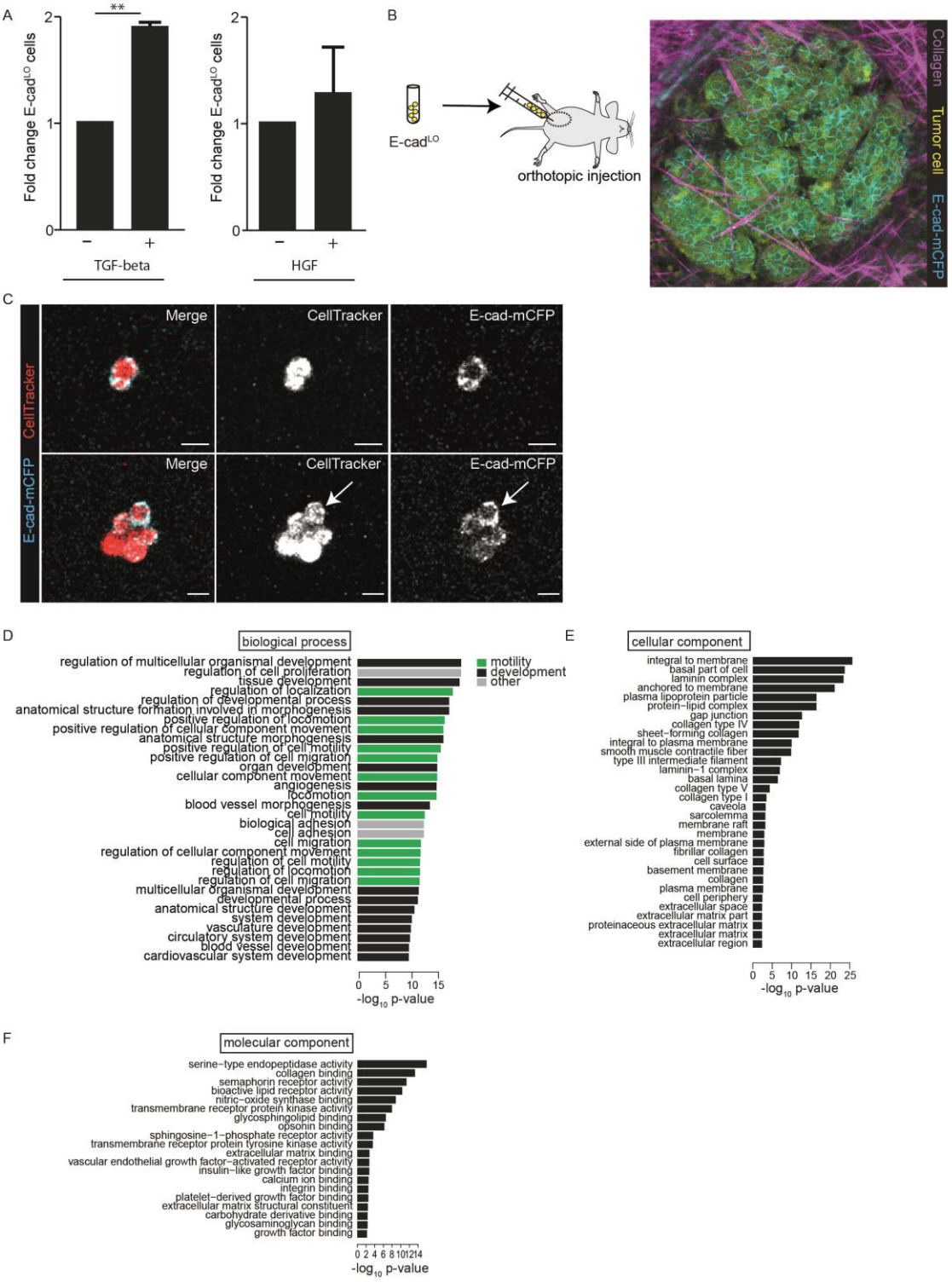

Figure S3. E-cad^{LO} tumor cells show plasticity (related to Figure 2)

(A) Organoids were cultured in collagen and stimulated with either HGF or TGF-beta. Fold change in the percentage of E-cad^{LO} cells compared to unstimulated organoids. Asterisks indicate a p value < 0,01 and n=3 independent replicates. Values are shown as SEM.

(B) An image of a PyMT tumor that developed upon orthotopic injection of E-cad^{LO} cells. Scale bar, 30 μm.

(C) Representative images of E-cad^{LO} tumor cells plated in collagen one day after sorting. Shown are merged images (left panel) and individual channels in which the middle panel represents the tumor cells stains with CellTracker red and the right panel represents E-cad-mCFP. The white arrow indicates a cell that has divided. Scale bar, 20 μm.

(D) Gene ontology list of biological processes of the E-cad^{LO} gene set divided into motility (green bars), developmental (black bars), and other (grey bars) categories. Barplots show Benjamini-Hochberg corrected p-values from WebGestalt (Zhang et al., 2005).

(E) Gene ontology of list cellular components of the E-cad^{LO} gene set.

(F) Gene ontology of list molecular components of the E-cad^{LO} gene set.

Figure S4. *In vivo* behavioral characterization of E-cad^{HI} and E-cad^{LO} tumor cells

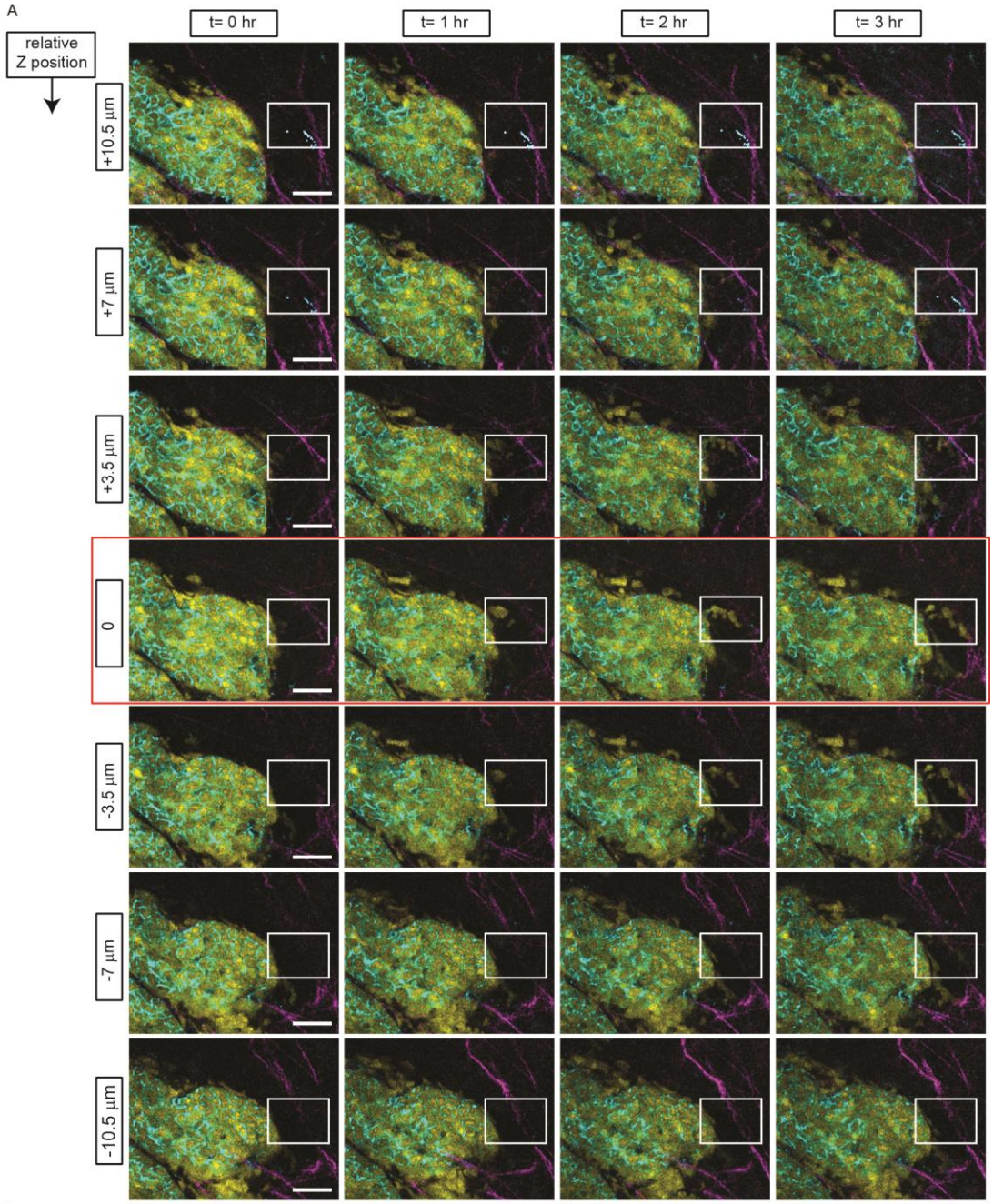

Figure S4. *In vivo* behavioral characterization of E-cad^{HI} and E-cad^{LO} tumor cells (related to Figure 4)
(A) Shown are the Z-stacks of the images in Figure 4C. The red box illustrates the Z-position shown in Figure 4C (Z-level 0). In the Z-levels 3.5 μm above and below, the cells are still visible, whereas the Z-levels 7 and 10.5 μm above and below show no cells in the circles. This observation excludes Z-drift as the source for cell movement and confirms true tumor cell movement. Scale bars, 50 μm.
(B) A representative image of an organoid stained with CK14 antibody. Left panel shows a merged image in which yellow represent tumor cells, cyan represent E-cad-mCFP and red is CK14 antibody staining. The right panels show a zoom of the protrusion in the organoids with in the left top a merged image and the individual channels shown in gray. Each channel is indicated in the figure. Shown is a single Z-level of 2.5 μm. Scale bar, 20 μm.
(C) Quantification of the E-cadherin intensity per pixel in CK14 positive and negative areas. SEM, n=8 organoids, 2 z-levels per organoid.

Figure S5. Circulation and proliferation of tumor cells

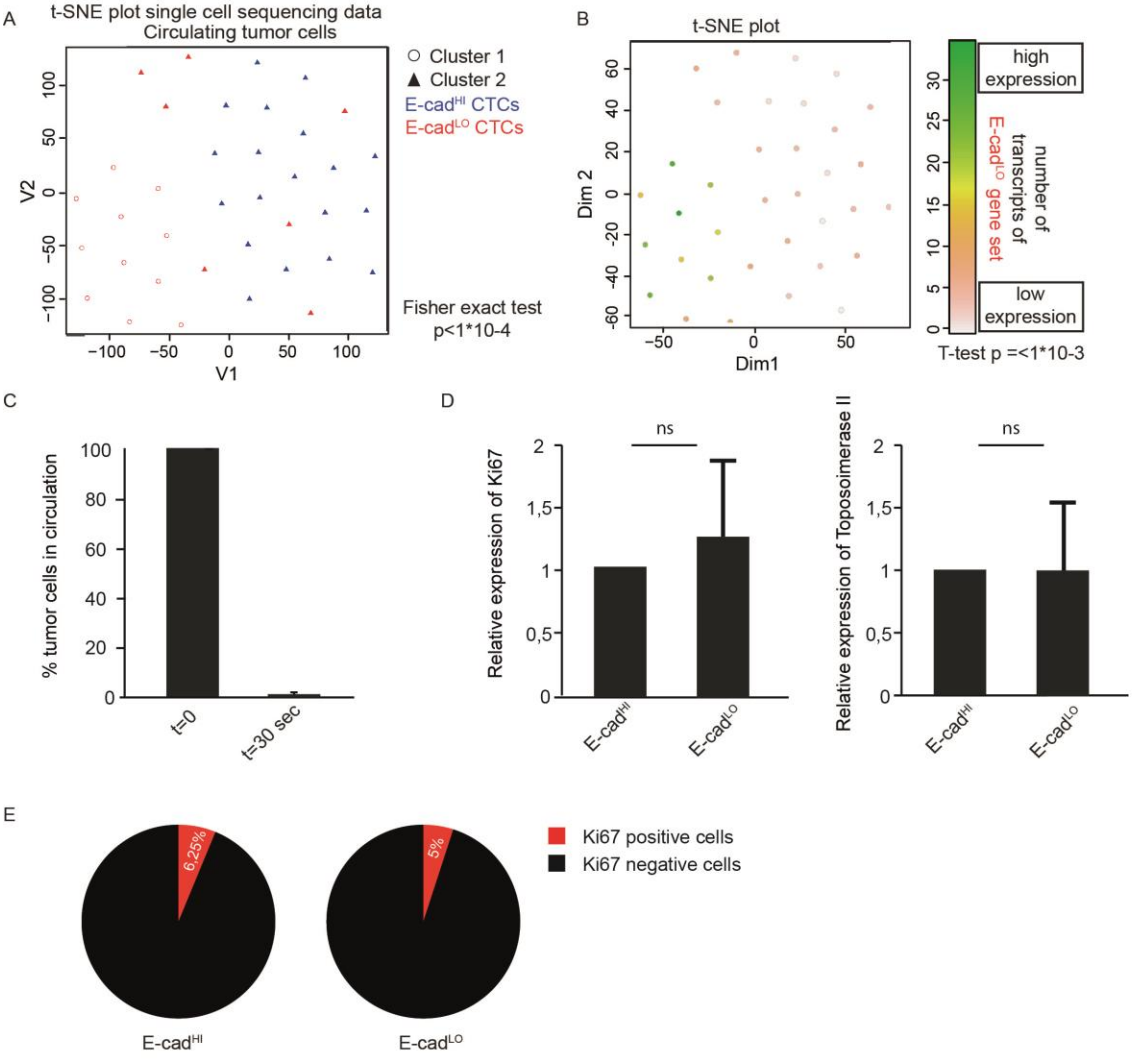

Figure S5. Circulation and proliferation of tumor cells (related to Figure 5)

(A) t-SNE plot of circulating tumor cells (CTCs). Using unsupervised K-medoids clustering, two separate clusters of CTCs were identified indicated as squares and triangles that overlap with E-cad^{HI} (blue) and E-cad^{LO} (red) tumor cells.

(B) t-SNE intensity plot of CTCs for the E-cad^{LO} gene set of Figure 2B. (C) Percentage of tumor cells found back in the circulation 30 seconds after tail vein injection. SEM, n=3.

(D) Relative expression of Ki67 and toposoimerase II in E-cad^{LO} cells by qPCR. SEM, n=4.

(E) Percentage of Ki67 positive tumor cells with immunohistochemistry staining after sorting by flow cytometry (n=120 cells).

Table S1. Significantly upregulated genes between E-cad^{HI} and E-cad^{LO} cells (related to Figure 2)

See separate Excel sheet.

Supporting Videos

Video S1. **A three hour intravital timelapse of a lobe of a PyMT-induced tumor with non-motile cells (related to Figure 4).** All tumor cells express YFP and fusion protein of mCFP to the endogenous E-cad. Shown are the separate channels for CFP (E-cad-mCFP, left), YFP (YFP, middle) and the merged images (Merge, right). The still images of this video are shown in Figure 4B. Scale bars, 50 μm.

Video S2. **A three hour intravital timelapse of a lobe of a PyMT-induced tumor where the tumor edge is broken and cells migrate into the surrounding stroma (related to Figure 4).** All tumor cells express YFP and fusion protein of mCFP to the endogenous E-cad. Shown are the separate channels for CFP (E-cad-mCFP, left), YFP (YFP, middle) and the merged images (Merge, right). The still images of this video are shown in Figure 4C. Scale bars, 50 μm.

Supplemental Materials and Methods

Human material

Four µm thick sections were cut from the paraffin block of an invasive ductal carcinoma and stained for E-cad using the Ventana BenchMark Ultra autostainer according to the manufacturer's instructions. Appropriate negative and positive controls were used throughout. Human tissues were obtained in compliance with Dutch law that does not require informed consent when leftover materials are used anonymously.

Mice

MMTV-PyMT and *MMTV-Cre* mice were purchased from Jackson Laboratory, Sacramento, CA, USA. *E-cad-mCFP* mice were a gift from Hans Clevers and *R26-loxP-stop-loxP-YFP (R26R-YFP)* mice a gift from Jacqueline Deschamps. Experiments were performed with *MMTV-PyMT;MMTV-Cre;R26R-YFP;E-Cad-mCFP* mice on a mixed and a pure FVB genetic background. Immunocompetent mice were housed under standard laboratory conditions and non-obese diabetic SCID IL-2 receptor gamma chain knockout (NSG) mice (own colony) were housed under IVC conditions. Mice received food and water ad libitum. All experiments were carried out in accordance with the guidelines of the Animal Welfare Committee of the Royal Netherlands Academy of Arts and Sciences, The Netherlands.

Flow cytometry on mouse material

From all mammary glands of *MMTV-PyMT;MMTV-Cre;R26R-YFP;E-Cad-mCFP* mice, tumors were collected and minced manually on ice using sterile scalpels. The tumor mass was digested in PBS supplemented with 25 µg/ml DNase I (Roche) and 5 Wünsch units TH Liberase /ml (Roche) at 37 ^{0}C for 35 minutes, followed by mashing through a 70 µm filter (BD Falcon) while adding DMEM/F12 + GlutaMAX (GIBCO, Invitrogen Life Technologies) supplemented with 5% (v/v) fetal bovine serum (Sigma), 100 µg/ml streptomycin and 100 U/ml penicillin (Invitrogen Life Technologies), 5 ng/ml insulin (I0516-5ML Sigma, St. Louis, MO, USA), 5 ng/ml EGF (Invitrogen) and 25 µg/ml DNase. After spin down (4 minutes at 500 RCF at RT) the pellet was resuspended in 6 ml 5mM EDTA/PBS, after which a Ficoll gradient (Histopaque-1077, Sigma) was used to select for live cells (30 minutes at 400 RCF at RT, break 1). Cells were washed once in 5mM EDTA/PBS and centrifuged (4 minutes at 500 RCF at RT) before proceeding with antibody labeling.

Blood was collected via intracardiac injection, the red blood cells were depleted by NH4Cl treatment. The remaining circulating tumor cells and immune cells were spun down (4 minutes 500 RCF at RT). Tumor cells and blood cells were blocked in 80% FACS buffer (5 mM EDTA in PBS supplemented with 5% fetal calf serum) / 20% serum mix (50/50 normal goat serum (monx10961, Monosan) and FcγII/III receptor blocking serum 2.4G2 (kind gift from Kiki Tesselaar, UMCU, The Netherlands)) for 10 minutes on ice before labeling with the following antibodies: rat anti-E-cad (DECMA-1 U 3254, Sigma) or E-cad-eFluor660 (DECMA-1, eBioscience), rabbit-anti-rat linker antibody (Southern Biotech), biotin-conjugated anti-mouse CD41 (clone eBioMWReg30, eBiocience) and anti-CD45 (clone 30-F11, eBioscience). Secondary labeling was performed using goat-anti-rabbit AF647 (Invitrogen) and Streptavadin-PerCP (Biolegend). After putting cells through a 70 µm strainer cap (BD Falcon) cells were sorted on a FACS AriaII Special Ordered Reseach Product (BD Biosciences). The sort strategy is illustrated in Figure S1B. A broad FSC SSC gate was followed by a gate excluding doublets, after which immune cells and megakaryocytes were excluded in a dump channel. YFP^{+} tumor cells were subdivided in E-cad-$mCFP^{HI}$ and E-cad-$mCFP^{LO}$ using very stringent gating. Finally, $YFP^{+}mCFP^{LO}$ tumor cells were additionally stringently gated for absence of E-cad antibody expression (E-cad^{LO}) and $YFP^{+}mCFP^{HI}$ tumor cells for high E-cad antibody expression (E-cad^{HI}).

Flow cytometry on human material

Female breast cancer patients diagnosed with invasive ductal carcinoma who underwent mastectomy in the Netherlands Cancer Institute were included in the study.

The tumor material was received after pathological examination and kept on ice from that point onwards. Tumors were minced using sterile scalpels and digested in PBS supplemented with 25 µg/ml DNase I (Roche) and 15 Wünsch units TH Liberase /ml (Roche) at 37 ^{0}C for 35 minutes, followed by mashing through a 70 µm filter (BD Falcon) while adding DMEM/F12 + GlutaMAX (GIBCO, Invitrogen Life Technologies), 1% Hepes, 1% Pen-Strep. Tumor cells were blocked in 80% FACS buffer (5 mM EDTA in PBS supplemented with 5% fetal calf serum) / 20% serum mix (50/50 normal goat serum (monx10961, Monosan) and FcγII/III receptor blocking serum 2.4G2 (kind gift from Kiki Tesselaar, UMCU, The Netherlands)) for 10 minutes on ice before labeling with E-cad-eFluor660 (DECMA-1, eBioscience) and EpCAM PE (1B7, eBioscience) labeled antibodies for 40 minutes on ice. DAPI was added as a life/death marker, cells were sorted on a FACS AriaII Special Ordered Research Product (BD

Biosciences) or on a FACS Jazz (BD biosciences). A broad FSC SSC gate was followed by a gate excluding doublets, after which the DAPI negative cells were selected. EpCAM positive tumor cells were subdivided in E-cad antibody positive and E-cad antibody negative cells.

Imaging sorted E-cad^{LO} cells

Cells were spun down at 0.2 RCF for 4 minutes. The cells were dissolved in medium with CellTracker Red CMPTX (Life technologies) as described by the manufacturer's instructions. After incubation cells were spun down and dissolved in rat-tail collagen I (1mg/ml) (Gibco, 3mg/ml) and plated in a glass bottom wilko dish. The cells were maintained with medium described below for organoids supplemented with RhoKinase Inhibitor (Abmole). Images were acquired using a Leica SP5 confocal microscope (Mannheim, Germany). All images were collected in 12 bit with a 25x (HCX IRAPO N.A. 0.95 WD 2.5 mm) water objective. CFP was excited at 458nm and emission was collected at 455-495 nm. CellTracker Red was excited at 633 nm and emission was collected at 640-700nm.

Ki67 staining sorted E-cad^{HI} and E-cad^{LO} cells

Cells were spun down after the FACS and fixated and permeabalized with BD cytofix/cytoperm (BD biosciences) according to manufacturer's instructions. Incubated with Ki67-Alexafluor 647 (clone B57, BD biosciences) for 30 minutes at 4 degrees. After washing cells were put on a glass slide and imaged at a Leica SP5 confocal microscope (Mannheim, Germany). All images were collected in 12 bit with a 25x (HCX IRAPO N.A. 0.95 WD 2.5 mm) water objective. Cells were imaged with DIC and Alexa647 was excited at 633 nm and emission was collected at 640-700nm. Cells were manually scored for positive staining.

Intravital imaging

For the intravital imaging experiments, YFP-expressing E-cad^{HI} tumor cells were isolated by flow cytometry and orthotopically transplanted in non-obese diabetic SCID IL-2 receptor gamma chain knockout mice to exclude potential YFP-labeling of non-epithelial lineages. For this, 1 x 10^5 sorted E-cad^{HI} tumor cells were injected in the right inguinal mammary gland of female non-obese diabetic SCID IL-2 receptor gamma chain knockout mice at 10-20 weeks of age. Mice bearing tumors of $\leq$ 500 mm^3 were used for intravital imaging. Mice were sedated using isoflurane inhalation anesthesia (1.5% to 2% isoflurane/O_2 mixture). The imaging site was surgically exposed, and the mouse was placed with its head in a facemask within a custom designed imaging box. The isoflurane was introduced through the facemask, and ventilated by an outlet on the other side of the box. The temperature of the imaging box and microscope were constantly adjusted to keep the mice between 36 and 37°C by a climate chamber that surrounds the whole stage of the microscope including the objectives. Imaging was performed on an inverted Leica TCS SP5 AOBS multi-photon microscope (Mannheim, Germany) with a chameleon Ti:Sapphire pumped Optical Parametric Oscillator (Coherent Inc. Santa Clare, CA, USA). The microscope is equipped with four non-descanned detectors: NDD1 (<455 nm), NDD2 (455-505 nm), NDD3 (505-550 nm), and NDD4 (560-650 nm). CFP was excited at 820 nm and detected in NND 2 and 3. YFP and Texas Red were excited at 960 nm. Second harmonic generation was detected in NDD1, YFP in NDD3, and Texas Red in NDD4. All images were collected in 12 bit and acquired with a 25x (HCX IRAPO N.A. 0.95 WD 2.5 mm) water objective. All images were processed using ImageJ software; pictures were converted to RGB, corrected for blead through (if necessary), smoothed (if necessary), cropped (if necessary), rotated (if necessary) and contrasted linearly. Videos were corrected for XY and Z drift using custom-written software (codes on request available from J.v.R.).

Imaging analysis and quantification

Imaging positions showing migrating tumor cells were selected and single migrating and protruding tumor cells were counted manually. For every tumor cell, the E-cad-mCFP status (level and localization) was determined by 3 independent researchers, after which the cells were counted as mCFP^{HI} or mCFP^{LO}.

PyMT tumor organoids

PyMT tumor organoids were established as previously described (Nguyen-Ngoc et al., 2012)**.** In short, tumors were harvested and enzymatically digested using trypsin (from bovine pancreas, Sigma) and collagenase A (Roche). The digested tumors were spun down in several steps until only the cell fragments of 200 to 1000 cells were left. These organoids were embedded in BME (RGF BME type 2 pathClear). Organoids were maintained in medium consisting of DMEM/F12 Glutamax supplemented with Hepes (1M Gibco), Penicillin-streptomycin, FGF (Life technologies) and B27 (50x Gibco). For experiments the organoids were plated in rat-tail collagen 1 (Gibco, 3mg/ml) in a concentration of 1mg/ml and the medium was further supplemented with 5 ng/ml HGF(RnD) or 100ng/ml TGF-beta protein (eBioscience).

Flow cytometry organoids

Organoids were harvested, spun down at 0.8 RCF for 3 minutes, digested in PBS supplemented with 25 µg/ml DNase I (Roche) and 5 Wünsch units TH Liberase /ml (Roche) at 37 ^{0}C for 20 minutes. Organoid cells were blocked in 80% FACS buffer (5 mM EDTA in PBS supplemented with 5% fetal calf serum) / 20% serum mix (50/50 normal goat serum (monx10961, Monosan) and FcγII/III receptor blocking serum 2.4G2 (kind gift from Kiki Tesselaar, UMCU, The Netherlands)) for 10 minutes on ice before labeling with E-cad-eFluor660 (DECMA-1, eBioscience). The cells were analyzed on FACS Jazz (BD biosciences) using the same strategy as the tumor cells.

Immunohistochemistry on organoids

Organoids were fixed in periodate-lysine-paraformaldehyde (PLP) buffer (2.5 ml 4% PFA in PBS + 0.0212 g $NaIO_4$ + 3.75 ml L-Lysine + 3.75 ml P-buffer (81 ml of 0.2 M Na_2HPO_4 and 19 ml of 0.2 M NaH_2PO_4 added to 100 ml demi water (pH 7.4)) for 1,5 hours at 4°C in a 4 compartment glass wilko dish, washed with P-buffer and incubated in 30% sucrose in P-buffer for 2 hours at 4°C. Samples were washed with PBS and blocked and permeabilized with PBS containing 1% Triton-X (Sigma), 10% FBS (Life technologies) and 1% BSA (Roche) for 3 hours at 4°C. CK14 Antibody (Covance, PRB-155P) was diluted in permeabilization buffer O/N at 4°C. Samples were washed with PBS and secondary anti-rabbit Alexa647 diluted in 1% BSA in PBS was incubated for 4 hours at 4°C and washed with PBS.

Images were acquired using a Leica SP5 confocal microscope (Mannheim, Germany). All images were collected in 12 bit with a 25x (HCX IRAPO N.A. 0.95 WD 2.5 mm) water objective. CFP was excited at 458nm and emission was collected at 455-495 nm. YFP was excited at 514nm and emission was collected at 520-560 nm. Alexa-647 was excited at 633 nm and emission was collected at 640-700nm. At least ten organoids with protrusions were analyzed using ImageJ software.

Western blot

Cells were lysed in 1% SDS buffer and equal amounts of protein were loaded onto an 8% SDS/PAGE gel. Antibodies against the following proteins were used: mouse-anti E-cad (clone 36/E-Cadherin, BD Biosciences), guinea pig-anti Vimentin (Fitzgerald), mouse-anti ZO-1 (Invitrogen), rabbit-anti N-cadherin (clone GC-4, Takara), mouse-anti Twist (Santa Cruz Biotechnology), mouse-anti β-actin (clone AC-15, Sigma). Immunoreactive bands were detected by HRP-conjugated secondary antibody incubation (mouse NA931V; rabbit NA934V, GE Healthcare, guinea pig ab 97155 Abcam) and ECL treatment (Thermo Scientific) according to the manufacturer's instructions.

RNA isolation and mRNA-sequencing

RNA was isolated using Trizol reagent (Invitrogen Life Technologies) according to the manufacturer's protocol and stored at -80°C. The amount and purity of isolated RNA was analyzed by the Nanodrop spectrophotometer (Wilmington, DE, USA). RNAseq libraries were prepared using Clontech's SMARTer Ultra Low RNA Kit and sequenced on an Illumina HiSeq.

mRNA-sequencing analysis

Paired end sequencing reads were mapped to mouse mm9 using TopHat2/Bowtie2 (Kim et al., 2013). Differential expression was calculated using CuffDiff2 (Trapnell et al., 2013) using UCSC genome annotation. A complete list of significantly upregulated genes between E-cad^{HI} and E-cad^{LO} tumor cells can be found in Table S1 and the data is deposited in the European Nucleotide Archive (accession number is: PRJEB5939). To compare the gene expression profile from our mouse model with human datasets we used a human-mouse ortholog database from the UCSC genome browser.

HMLE expression data was taken from (Onder et al., 2008) (GEO accession: GSE9691). Affymetrix data was normalized using RMA (Irizarry et al., 2003) from the affy Bioconductor package. Both the E-cad shRNA and the E-cad dominant negative data were compared to the wild-type data. Differential expression in Figure 3A was defined as the deviation from the median over times 3 replicate experimental and 3 replicate control experiments. Single molecule RNAseq data was taken from GEO accession GSE41245 (Yu et al., 2013). We performed quantile normalization on the raw expression values and calculated the differential expression values between EpCAM$^+$ cells and the corresponding IgG$^+$ cells. Expression values were averaged per gene over multiple experiments belonging to the same class (i.e. "blood specimen healthy donor", "epithelial CTC" and "mesenchymal CTC"). The set of upregulated genes was analyzed using the WebGestalt tool (Zhang et al., 2005) for functional classification. The set of upregulated RNAs was used as the input set; as a background set we used all the genes in the mouse genome.

Single cell mRNA sequencing

Single E-cad^{HI} or E-cad^{LO} cells were sorted into trizol. The RNA was extracted using chloroform and precipitated with iso-propanol. The RNA-pellet was then processed using the CEL-seq protocol as described in (Hashimshony et al., 2012; Grun et al., 2014) and sequenced on an Illumina Nextseq using 75bp paired end sequencing. After sequencing, read 1 was aligned to the mm10 RefSeq mouse transcriptome downloaded from the UCSC genome browser using with default parameters. Read 2 contains a barcode identifying the sample from which the read originated. CEL-seq only sequences the most 3' prime end of a transcript and generates one read per transcript. Data analysis was done in R. Samples with less than 1000 transcripts were discarded after which the remaining samples were downsampled to 1000 transcripts. T-distributed stochastic neighbor embedding (t-SNE) was performed on the Euclidean distances between cells. Heat map was made based on Pearson's correlation between cells after filtering out genes that had less than 5 transcripts in at least 1 cell.

cDNA preparation and qPCR

Complementary DNA was synthesized using the Avian Myeloblastosis Virus Reverse Transcriptase kit (Promega, Madison, WI, USA) according to the manufacturer's instructions. Sequences of used primers can be found below. qPCR was performed using Power SYBR Green PCR Master Mix (Applied Biosystems). Thermal cycle conditions used for all qPCR reactions were as follows: 5 min at 95°C, followed by 40 cycles consisting of denaturation for 30 sec at 95°C, annealing for 30 sec at 58°C, and extension for 1 min at 72°C. PCR reactions were concluded with incubation for 10 min at 72°C to complete the extension of all synthesized products. Relative quantification values were calculated using the ddCt method and mean values were plotted with SD using GraphPad Prism v5.0 (GraphPad Software, Inc.).

The following primers were used for qPCR: E-cadherin (Fwd: 5'- GCT TCA GTT CCG AGG TCT AC -3', Rev: 5'- GCC AGT GCA TCC TTC AAA TC -3'), Twist-1 (Fwd: 5'- CGG AGA CCT AGA TGT CAT TGT TT -3', Rev: 5'- CGC CCT GAT TCT TGT GAA TTT G -3'), Vimentin (Fwd: 5'- GAG GAG ATG CTC CAG AGA GA -3', Rev: 5'- TCC TGC AAG GAT TCC ACT TT -3'), N-Cadherin (Fwd: 5'- GTG GAG GCT TCT GGT GAA AT -3', Rev: 5'- GGC TCG CTG CTT TCA TAC T -3'), Snail1 (Fwd: 5'- CTG CAC GAC CTG TGG AAA -3', Rev: 5'- GGC ACT GGT ATC TCT TCA CAT C -3'), Fibronectin (Fwd: 5'- GGT GTC CGA TAC CAG TGT TAC -3', Rev: 5'- TCT CCG TGA TAA TTA CTT GGA CAG -3'), ZO-1 (Fwd: 5'- GTC ACG ATC TCC TGA CCA AC -3', Rev: 5'- CCA GGT TTA GAC ATT CGC TCT -3'), Pdgfrb (Fwd: 5'- AGC GAG AAG CAA GCC TTA AT – 3', Rev: 5'- GAT CAC CGT ATC GGC AGT ATT C – 3'), Axl (Fwd: 5'-AAA CTC CAG GCC TGA ACA AG- 3', Rev: 5'- TTG GCA TTG TGG GCT TCA – 3'), Acta2 (Fwd: 5'-TAA GGC CAA CCG GGA GAA-3', Rev: 5'-GGG ACA TTG AAG GTC TCA AAC A- 3'), Cav1 (Fwd: 5'- CAA CAT CTA CAA GCC CAA CAA C – 3', Rev: 5'- TCC CTT CTG GTT CTG CAA TC – 3'), Ctgf (Fwd: 5'- AAC CGC AAG ATC GGA GTG – 3', Rev: 5'- TGC TTT GGA AGG ACT CAC C – 3'), Fstl1 (Fwd: 5'- GGC TGG AAG CTG AGA TCA TT -3', Rev: 5'- CAG GTG AGA GTC GCC ATT AT -3'), Mcam (Fwd: 5'- CCC ATT CCT CAA GTC CTA TGG – 3', Rev: 5'- CCC ATT CCT CAA GTC CTA TGG – 3'), Postn (Fwd: 5'- CAG CTC CTG TAA GAA CTG GTA TC -3', Rev: 5'- ATA TAG CCA GGG CAG CAT TC – 3'), Sparc (Fwd: 5'- GCT GGA TCA GCA CCC TAT T -3', Rev5'- TGT CTA GGT CAC AGG TCT CA -3'); Toposoimerase II (Fwd: 5'- AAC GAG AGA CAC ATC ATT GTC AG-3', Rev5'- TCA CCT TCC CTA TCA CAG TCC-3'); Ki67 (Fwd: 5'- ACC GTG GAG TAG TTT ATC TGG-3', Rev5'- TGT TTC CAG TCC GCT TAC TTC T -3')

Mouse tumor and tissue processing for histology

Tumors and other tissues were isolated from the mice at the end of the experiment and fixed in periodate-lysine-paraformaldehyde (PLP) buffer (2.5 ml 4% PFA + 0.0212 g $NaIO_4$ + 3.75 ml L-Lysine + 3.75 ml P-buffer (pH 7.4)) O/N at 4°C. The following day, the fixed tumors and tissues were washed twice with P-buffer and placed for at least 6 hours in 30% sucrose at 4°C. The tumors and tissues were then embedded in tissue freezing medium (Leica Microsystems, Nussloch, Germany) and stored at -80°C before cryosectioning.

Human immunohistochemistry

E-cad staining on 4 µm paraffin-embedded tumor slices was performed using mouse anti-E-cad antibody (clone 36/E-Cadherin, BD Biosciences) overnight, followed by an anti-mouse horseradish peroxidase labeled secondary antibody (Dako, Envision+ System- HRP labeled Polymer anti-mouse).

Analyzing metastatic outgrowth in the lungs

Tumor or tissue cryosections (20 µm thick tumor sections and 100-150 µm thick lungs sections) were rehydrated for 10 min in PBS and embedded in Vectashield mounting medium (hard set; Vector Labs, Burlington, Ontario, Canada). When indicated, cryosections were counterstained with TO-PRO3 (Invitrogen molecular probes, Paisley, UK) to visualize the nuclei. Images were acquired using a Leica SP5 or SP8 confocal microscope (Mannheim, Germany) equipped with 10x NA 0.3 and 20x NA 0.7 dry objectives. CFP was excited at 405 nm or at 458nm and emission was collected at 455-495 nm for CFP. YFP was excited at 514nm and emission was collected at 520-560 nm. TO-PRO-3 was excited at 633 nm and emission was collected at 650-670nm. At least 4 frozen sections per mouse were analyzed.

Determination of the time in circulation for tumor cells

PyMT tumor organoids were cultured as described above and processed until single cells as described with flow cytometry for organoids. The cells were counted on a hemocytometer and 10.000 cells were diluted in the appropriate amount of PBS for a tail vein injection. The mice were anesthetized using isoflurane. The tumor cells were injected into the tail vein of the mouse and after 30 seconds blood was harvested via cardiac injection of the right heart chamber. The blood was processed for flow cytometry as described above.

Intrahepatic injection of tumor cells

From all mammary glands of *MMTV-PyMT;MMTV-Cre;R26R-YFP;E-Cad-mCFP* mice, tumors were collected and processed to single cell suspension as described for flow cytometry (see above). After spin down (4 minutes at 500 RCF at RT) the pellet was suspended in PBS. For the formation of hepatic tumors $5x10^5$ tumor cells were injected. The appropriate amount of cells was spun down (4 minutes at 500RCF at RT) and suspended in Matrigel (Corning Matrigel Basement Membrane Matrix Growth Factor Reduced, Phenol Red Free). NSG mice were sedated using isoflurane inhalation anesthesia (1.5% to 2% isoflurane/O_2 mixture) and received subcutaneous analgetica of 3mg/100ul buprenorphine (Temgesic). After a midline incision, the liver was mobilized and the cells were injected superficially into the liver. After coagulation of potential bleeding caused by the injection, the liver was placed back and the abdomen was closed. Tumor cells were injected into the liver of different recipient mice for five subsequent rounds to prime the tumor material for growing out in the liver. After the final round, tumors were harvested and processed for FACS as described above.

Cells obtained from cell sorting were spun down (4 minutes at 500 RCF at RT) and suspended in Matrigel (Corning Matrigel Basement Membrane Matrix Growth Factor Reduced, Phenol Red Free) and injected into the liver of three NSG mice. Tumors developed within 4-8 weeks and the sets of mice were sacrificed by cervical dislocation when one of the three showed signs of tumor development. The tumors and organs were then processed as described above.

Statistical analyses

Statistical analyses were performed using Graphpad software, We used a unpaired t-test with Welsh correction for the E-cadherin intensity in CK14 positive and negative areas, a one-sample t-test for the induction of EMT with TGF-beta or HGF in organoids. The proliferation difference in E-cad^{LO} cells compared to E-cad^{HI} cells was performed with a one-sample t-test compared to hypothetical value one.

Tumor-initiating cell frequency was tested by Elda-limiting dilution test on this website http://bioinf.wehi.edu.au/software/elda/ (Hu et al. 2009).

Supplemental references:

Hashimshony, T., Wagner, F., Sher, N., and Yanai, I. (2012). CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. Cell reports *2*, 666-673.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics *4*, 249-264.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome biology *14*, R36.

Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nature biotechnology *31*, 46-53.

Grun, D., Kester, L., and van Oudenaarden, A. (2014). Validation of noise models for single-cell transcriptomics. Nature methods *11*, 637-640

Nguyen-Ngoc, K.V., Cheung, K.J., Brenot, A., Shamir, E.R., Gray, R.S., Hines, W.C., Yaswen, P., Werb, Z., and Ewald, A.J. (2012). ECM microenvironment regulates collective migration and local dissemination in normal and malignant mammary epithelium. Proceedings of the National Academy of Sciences of the United States of America *109*, E2595-2604.

Zhang, B., Kirov, S., and Snoddy, J. (2005). WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic acids research *33*, W741-748.

Hu, Y, and Smyth, GK (2009). ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. Journal of Immunological Methods 347, 70-78.