ISOLATION AND CHARACTERIZATION OF TWO BASAL BREAST CANCER MODEL CELL LINES

by
Melissa L Gentz

A thesis submitted to Johns Hopkins University in conformity with the requirements for the degree of Master of Science

Baltimore, Maryland
May 2021

© 2021 Melissa Gentz
All Rights Reserved

Abstract

Epithelial-to-mesenchymal transition (EMT) and its reversed process, mesenchymal-to-epithelial transition (MET) remain a topic of debate among the cancer research community. Here, we sought to isolate two novel cell lines from the basal-like mouse models ($ROSA^{mT/mG}$ C3(1)-Tag and C3(1)-Tag) of triple-negative breast cancer (TNBC) to investigate the effects of 2D and 3D culture methods on the expression of epithelial and mesenchymal phenotypes. We demonstrated that co-expression of E-cadherin and vimentin is present in both cell lines when cultured in 2D. Our data also suggests that when embedded in 3D matrices including Rat tail collagen 1 and Matrigel, TGFβ signaling induces spheroid invasion. However, it may play a role in inhibiting colony formation in Matrigel. Taken together, my data reveal that C3(1)-TagM and C3(1)-Tag cell lines may provide researchers with an alternate platform to research hybrid EMT.

Keywords: C3(1)-Tag, triple negative breast cancer, TNBC, basal breast cancer, EMT, MET, vimentin, TGFβ

Thesis Advisor: Andrew J. Ewald, PhD
Professor and Director
Department of Cell Biology
Johns Hopkins University, School of Medicine

Thesis Reader: Jennifer M. Kavran, PhD
Assistant Professor
Johns Hopkins Bloomberg School of Public Health.

Acknowledgements

I would like to begin by expressing my gratitude towards my thesis advisor and academic mentor, Dr. Andrew J. Ewald. Without him, the completion of my Masters's degree would not have been possible. Thank you for providing me with your unwavering support, guidance and endless patience. I would also like to extend a special thanks to my lab supervisor, Dr. Eloïse Grasset, who accepted me as her mentee and trained me with necessary skills to become a young researcher. I would also like to thank my secondary thesis reader, Dr. Jennifer Kavran, who provided me with valuable insights. My gratitude extends to all members of the Ewald Lab for their support and encouragement.

I would also like to acknowledge the entire faculty and staff of the Biochemistry and Molecular Biology department at the Johns Hopkins University Bloomberg School of Public Health, and the Department of Cell Biology of the Johns Hopkins University School of Medicine.

Finally, I would like to say thank you to my family and friends for their endless support throughout my journey as a Master's student.

Table of Contents

List of Figures

List of Abbreviations

BC	Breast cancer
CSC	Cancer stem cells
CTC	Circulating tumor cells
DMEM/F-12	DMEM/F-12 without FBS
DMEM/F-12/FBS	DMEM/F-12 with 10% FBS
E-cad	E-cadherin
E-cad+/Vim-	E-cadherin-positive/Vim-negative
E-cad+/Vim+	E-cadherin-positive/Vim-positive
ECM	Extracellular matrix
EMP	Epithelial-mesenchymal plasticity
EMT	Epithelial-mesenchymal transition
EMT-TFs	EMT transcription factors
ER	Estrogen receptor
ER-/PR-/HER+	ER-negative/PR-negative/HER2-positive
ER-/PR-/HER2-	ER-negative/PR-negative/HER2-negative
ER+/PR+/HER2+	ER-positive/PR-positive/HER2-positive
ER+/PR+/HER2-	ER-positive/PR-positive/HER2-negative
ERBB2	Receptor tyrosine-protein kinase
GEMMs	Genetically engineered mouse models
HER2	Human growth factor receptor 2
HER2+	HER2-postiive
IHC	Immunohistochemistry
K14	Cytokeratin 14
K17	Cytokeratin 17
K5/6	Cytokeratin 5/6
Ki-67	Cell proliferation marker Ki-67
MET	Mesenchymal-epithelial transition
MMPs	Matrix metalloproteinases
MMTV	Mouse mammary tumor virus
p63	tumor protein 63
PR	Progesterone receptor
PyMT	Polyoma middle T antigen
TGFβ	Transforming growth factor beta
TNBC	Triple negative breast cancer
TNBC	Triple negative breast cancer
Vim	Vimentin

Introduction

Breast cancer (BC) is the most common type of cancer among women worldwide (Ferlay et al, 2015). It is a complex disease that exhibits heterogeneity in genomic events, levels of gene expression, metastatic potential, and chemoresistance. In the clinical setting, classical breast cancer subtyping has relied on the status of three biomarkers including the estrogen receptor (ER), progesterone receptor (PR), and human growth factor receptor 2 (HER2) (Spitale et al, 2009). Combined immunohistochemistry (IHC) analysis of ER, PR, and HER2 has provided clinicians with significant insights into breast tumor behavior and response to cancer therapies. Four basic groups have been defined by IHC nomenclature: ER-positive/PR-positive/HER2-negative (ER+/PR+HER2-), ER-positive/PR-positive/HER2-positive (ER+/PR+/HER2+), ER-negative/PR-negative/HER2-postive (ER-/PR-/HER+), and ER-negative/PR-negative/HER2-negative (ER-/PR-/HER2-) (Vallejos et al, 2009).

More recently, however, BCs have been further divided into groups according to their intrinsic molecular features. Currently, there are at least six subtypes that have been characterized, including normal-like, Luminal A and B, HER2-postitive, basal-like, and claudin-low (Fedele et al, 2017). Normal-like and Luminal A tumors are generally (ER+/PR+HER2-). Luminal B tumors, on the other hand, are typically (ER+/PR+/HER2+). However, there are exceptions to this (e.g. Luminal B tumors that are HER2-) and nomenclature equivalence for BC subtyping has not yet been established (Dai et al, 2015). Therefore, additional markers such as the cell proliferation marker Ki-67 have been utilized to provide additional information for clinical management. Normal-like and Luminal A tumors tend to test negative or express low levels of the cell proliferation marker Ki-67, while Luminal B tumors upregulate its expression (Kanyılmaz et al, 2019). Furthermore, basal-like and claudin-low BCs are considered subdivisions of triple-negative

breast cancer (TNBC), which is categorized by tumors that lack of ER, PR and HER2 expression (ER-/PR-/HER2-) (Kast et al., 2015). Unfortunately, TNBC is not only the most aggressive form of all BCs, but it is also the most challenging subtype to treat (Garrido-Castro et al 2019; Dunnwald et al, 2007). Although there is heterogeneity, the majority of TNBCs express basal cell markers including cytokeratins 5/6 (K5/6), 14 (K14), (K17), and tumor protein 63 (p63) Dunnwald et al, 2007; Liu et al, 2019). Additionally, claudin low TNBCs have been characterized by the downregulation of epithelial biomarkers, including claudins 3, 4, 7, occludin and E-cadherin (E-cad), and by the upregulation of epithelial-to-mesenchymal transition (EMT) associated with cell stemness (Garrido-Castro et al 2019; Herschkowitz et al, 2007).

Epithelial-mesenchymal transition (EMT) is a biological process in which polarized apico-basal epithelial cells switch toward a more mesenchymal phenotype (Kalluri & Weinberg, 2009). During EMT, adhesion molecules of epithelial cells are modified, leading to fewer intercellular contacts and more cell-ECM contacts. Coordinated upregulation of matrix metalloproteinases (MMPs) then cause the degradation of the basement membrane. As a result, cells are able to adopt increased migratory and invasive behavior, typically seen in fibroblasts, mesenchymal cells, and cancer stem cells (CSCs) (Kalluri & Weinberg, 2009; Mani et al, 2008; Moreno-Bueno et al, 2008). EMT is carried out in response to cell signaling factors which induce EMT transcription factors (EMT-TFs) (e.g. SNAIL1/2, ZEB1/2, TWIST1/2, FOXC1/2), miRNAs, epigenetic, and post-translational regulators (Moreno-Bueno et al, 2008; Kim et al, 2017). Many of these signaling factors are also present during embryonic development, wound healing, fibrosis, and cancer metastasis, giving EMT a role in each of these biological processes (Kalluri & Weinberg, 2009). The reverse process to EMT, mesenchymal-epithelial transition (MET), is associated with the loss of migratory behavior and the reestablishment of apico-basal polarity in cells (Kim et al, 2017).

Early on, EMT-MET was first conveyed as full oscillation from one cell state to the other, as is seen during development (Thiery et al, 2009). The word "transformation" was often used by researchers to describe the complete switch between epithelial and mesenchymal cell states (Hay 1995, Kalluri & Weinberg, 2009). In the most recent decade, however, studies revealed that cell populations may actually display a spectrum of intermediate EMT-MET phenotypes, known as epithelial-mesenchymal plasticity (EMP), hybrid EMT, or partial EMT. Research groups have demonstrated that MET is required to form metastases at distant sites (Tsai et al, 2012; Ocanã et al, 2012). Growing evidence also points towards a hybrid EMT program being present in circulating tumor cells (CTCs) of TNBC patients (Alix-Panabières et al, 2017; Francart et al, 2018). Nevertheless, the role of hybrid EMT remains a topic of controversy among the cancer research community partly due to the limitations of current model systems used to study this phenomenon. Assessments of the partial EMT phenotype should be made taking into consideration the type of model used to represent EMT plasticity, invasion, or metastasis.

The most popular mouse models used to study spontaneous breast cancer metastasis *in vivo* includes MMTV-PyMT and MMTV-Neu. In these genetically engineered mouse models (GEMMS), expression of the polyoma middle T antigen oncogene (PyMT) or the mouse receptor tyrosine-protein kinase ERBB2 (Neu) oncogene is driven by the Mouse Mammary Tumor Virus long terminal repeat (MMTV) (Fantozzi et al, 2006). As a result, the mice develop mammary tumors that are highly sensitive to hormones or pregnancy. Although MMTV GEMMs have shown to be a valuable model for breast cancer research, they do present a couple of limitations when it comes to mimicking the TBNC phenotype. For example, the transcriptomic analysis demonstrated that these models clustered with Luminal B and HER2+ human BC (Pfefferle et al, 2013). Previous research has also indicated that both MMTV-PyMT and MMTV-Neu models display a

very small percentage of cells that express EMT biomarkers, some of which may be important for the characterization of TNBC (Beerling et al, 2016; Fischer et al, 2015; Ye et al, 2015; Li et al, 2020). On the other hand, in the C3(1)-Tag mouse model, all female mice that carry the Large T oncogene (Tag) and develop mammary tumors that are more similar to basal -like BCs. C3(1)-Tag mice also develop primary tumors without the requirement of hormone administration or pregnancy (Maroulakou et al, 1994; Aprelikova et al, 2016).

In our recent manuscript written primarily by Eloïse Grasset, a post-doctoral fellow in the Ewald Lab, we sought to investigate the role of hybrid EMT in both invasion and metastasis (Appendix 1). We utilized *in vitro* and *in vivo* model systems to assess EMT gene expression among GEMMs (MMTV-PyMT and C3[1]-Tag), patient-derived xenografts (PDXs), and TNBC patient tumor samples. Our results showed that the C3(1)-Tag mouse model not only expressed higher levels of EMT biomarkers than MMTV-PyMT, but also that the C3(1)-Tag displayed a hybrid EMT phenotype more similar to TNBC samples from patients (Appendix, Figure 1). Furthermore, we demonstrated that while the EMT biomarker vimentin (Vim) is required for C3(1)-Tag cells to invade and disseminate to distant sites*,* loss of Vim actually increases metastatic outgrowth both in 3D culture assays and *in vivo* (Appendix 1; Figures 6 and 7). By assessing lung metastases in the C3(1)-Tag mice, we demonstrated that the metastases composed of hybrid-EMT cells that were also positive for Ki-67 (Appendix 1; Figure 7). Knowing that transforming factor beta (TGFβ) regulates EMT from previous studies, our results further confirmed that its presence induced cell proliferation, migration, and invasion in the C3(1)-Tag mouse line (Appendix 1; Figure S6). Our results also demonstrated that lung metastases in the C3(1)-Tag mouse line displayed a hybrid EMT phenotype and expressed the cell proliferation marker Ki-67, indicating that that EMT-MET cells promote lung metastatic colonization *in vivo* (Appendix 1; Figure 7).

The main objective of this thesis project was to isolate novel cells lines from the basal TNBC mouse models (ROSA$^{mT/mG}$ C3(1)-Tag and C3(1)-Tag) to enable further investigation of TNBC cell invasion and metastasis. We knew from previous research that three-dimensional (3D) cell culture is more physiologically relevant than two-dimensional (2D) cell culture to model cell invasion and metastasis (Cree et al, 2011). However, we wanted to investigate if isolated cell lines cultured in 2D could be used to generate multicellular 3D spheroids that retain the hybrid-EMT phenotype that we observed when C3(1)-Tag primary organoids are embedded in 3D matrices (Appendix 1; Figures 1, 3, 6, 7, S6, S7). The difference among spheroids and organoids is that the prior are generated using 2D cell lines and the latter are isolated directly from a tumor. Although organoids have shown to be incredibly useful for BC research, some research groups have switched to culturing 3D spheroids derived from cell lines to minimize the use of *in vivo* studies that are more time consuming and expensive (Rosenbluth et al, 2020; De Meulenaere et al, 2020; Schmidt et al, 2016; Kadletz et al, 2015; Melissaridou et al, 2019). I successfully isolated two cell lines from primary organoids: C3(1)-TagM (expression of mTmG) and C3(1)-tag (or whatever name you like). To characterize these cell lines, I performed immunofluorescence staining for E-cad and Vim, to check for the presence of cells that were double-positive for E-cad and Vim (E-cad+/Vim+), the two most common markers to assess epithelial and mesenchymal phenotypes, respectively (Lu and Kang, 2019). Then, I investigated the invasion of these cell lines in 3D collagen matrix and their ability to form colonies in 3D Matrigel matrix, similar to our work done with primary organoids (Appendix 1). Taken together, my data reveal that C3(1)-TagM and C3(1)-Tag primary cell lines may provide researchers with an alternate platform to research hybrid EMT. This thesis project highlights the physiological relevance exhibited by spheroids cultured using 3D

cell technology and reports the development and characterization of novel cell lines that will accelerate TNBC metastasis and EMT research.

Materials and Methods

Reagents

10× DMEM (low glucose)	Sigma, D2429
4% PFA	Electron Microscopy Sciences, 15714S
Bambanker Freezing Media	Bulldog Bio, 101974-112
BSA	Sigma, A9576
Collagenase from *Clostridium histolyticum*	Sigma, C2139
DMEM/F-12	Gibco, 10565-018
DNase	Sigma, D4263
DPBS	Sigma, D8662
FBS	Sigma, F0926
Gentamycin	Gibco, 15750-060
Insulin	Sigma, I9278
Insulin-transferrin selenium,	Gibco, 51500056
Matrigel (growth factor reduced)	Corning,354230
Penicillin-streptomycin	Sigma, P4333
Rat-tail collagen I	Corning, 354236
Sodium hydroxide	Sigma, S2770
TGFβ receptor 1 inhibitor SB525334	APExBIO, A5602
TGFβ-1	Peprotek, 100-21 (2nM)
TritonX-100	Sigma, X100
Trypan Blue Solution 0.4%	Gibco, 15250061
Trypsin	Gibco, 27250-018
Trypsin-EDA 0.25%	Gibco, 25200056

Immunohistochemistry

DAPI (1:3000)	Invitrogen, D57
E-cadherin (anti-rat)	Abcam, ab11512, 1:100
Vimentin (anti-rabbit)	Abcam, ab92547, 1:250
Phalloidin	Invitrogen, A22283 1:200

Reagent Stock Solutions

BSA Coating Solution

Make 2.5% (vol/vol) BSA solution in DPBS. Filter-sterilize (0.22 μm). Store at 4 °C. The solution can be reused for ~6–8 weeks if kept sterile. To coat pipette tips with BSA, pipette desired volume and release. To coat the conical, transfer 1ml of BSA to tube, tighten cap and invert tube, allowing the interior surface of the tube to be covered. Remove excess BSA by aspiration

Collagenase Digesting Solution

For ~1g murine tumors: 2 mg ml−1 collagenase (Sigma, C2139), 2 mg ml−1 trypsin (Gibco, 27250- 018), 5% FBS (Sigma, F0926), 5 μg ml−1 insulin (Sigma, I9278), and 50 μg ml−1 gentamycin (Gibco, 15750) in 30 ml DMEM-F-12 (Gibco, 10565-018).

General Equipment

15 ml Conical Centrifuge Tubes	Falcon, 352095
50 ml Conical Centrifuge Tubes	Corning, 352070
10 cm Petri dishes	Corning, cat. no. 430293
T75 adherent cell culture flask	Sarstedt Inc, 83.3911.002
Hemacytometer	Z359629-1EA, Sigma
Cryogenic Tubes	Thermo Scientific, 5000-1012
Cell Freezing Container	Thermo Scientific, 5100-0001
Biological safety cabinet	LabGard ES NU-540 Class II
Cell culture incubator 5% CO2, 37 °C	Thermo Fisher Scientific Heracell Vios 160i
High-speed centrifuges	Thermo Fisher Scientific Sorvall Legend XIR
Laboratory-grade refrigerator (4 °C)	VWR, GDM 49
Laboratory-grade freezer (−20 °C)	Fisher Scientific, 13-986-428F
Laboratory-grade freezer (−80 °C)	Thermo Scientific, Revco UxF
0.22-μm sterile filter	Millipore, SCGP00525
Serological pipettes: 5 ml and 10 ml	Genemate, P2837-5;Falcon, 357551
Filter tips: 0.1–10 μl, 10–100 μl, 20–200 μl, 100–1,000 μl	-
Heating block	Benchmark Scientific, BSH1002
24-well black coverslip-bottomed plate	Greiner Bio-One, 662892
96-well black coverslip-bottomed plate	Greiner Bio-One, 655892
Ice	-
Ice Bucket	-

Mice

All mouse procedures were based on animal care guidelines disapproved by the Johns Hopkins Medical Institute Animal Care and Use Committee (ACUC). All mice were female and maintained on an FVB/n background. (C3(1)-Tag) (Maroulakou et al, 1994) and B6.129(Cg)-Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo/J (mT/mG) (Muzumdar et al, 2007) mice were obtained from the Jackson Laboratories. ROSA$^{mT/mG}$ C3(1)-Tag, (here, referred to as C3(1)-TagM mice were derived by crossing C3(1)-Tag with ROSA$^{mT/mG}$ mice. Mouse euthanasia was performed in conformity with ethically approved guidelines. Mammary tumor organoids were isolated as previously described (Padmanaban et al, 2020).

Isolation of Mammary Tumors
For reagent stock solutions, see Padmanaban et.al, Nature Protocols, 2020.

Sterilize work bench and dissecting tools with 70% ethanol. Lie mouse face-up and sterilize with 70% ethanol spray. Use standard forceps to lift skin above groin and Spencer ligature scissors to make first incision. Extend the incision along the mid-line up to the chin of the mouse. Make lateral incisions towards each of the mouse limbs. Pull away skin to expose the mammary tumors. Use the dorsal side of Graefe forceps to detach tumors from peritoneum. Next, dissect out mammary tumors using Iris scissors, avoiding the surrounding tissue. Transfer dissected tumors to a sterile Petri dish.

Tissue Digestion

In a sterile cell culture cabinet, use a #10 scalpel to mince tumors to small pieces (~1 mm × 1 mm). Transfer minced tissue to a 50 ml conical tube containing collagenase digesting solution . Shake collagenase solution for 1 hour at 37°C at 180 rpm in a benchtop shaking incubator. Pellet the digested tissue by spinning the 50 ml conical at 1500 rpm for 10 minutes. Aspirate the supernatant. Add 8 ml DMEM/F-12 and gently resuspend the pellet 5-10 times using a BSA-coated pipette. If pipette tips and conicals are not coated with BSA prior to coming in contact with the cell solution, this can significantly impact organoid yield. Add 2 U μl^{-1} DNase and carefully invert the tube for 3-5 minutes. Add 20 ml DMEM/F-12 and gently mix by inverting the tube. Spin down at 1500 rpm for 10 minutes. Aspirate the supernatant and add 25 ml DMEM/F-12 to the pellet. Use a BSA-coated pipet to gently mix solution. Next, wait 5-10 seconds to allow large particles to settle at the bottom of the tube. Using a BSA-coated pipette, transfer the supernatant to a new BSA-coated conical.

Differential Centrifugation

The following steps are essential to deplete single cells, stromal cells, and enrich for epithelial organoids. Pellet the organoids by spinning at 1500 rpm for 3 seconds. Aspirate the supernatant leaving ~1 ml in the tube to prevent losing organoids. Resuspend the loose pellet using a BSA-coated pipette and 10 ml of DMEM/F-12. Repeat this process 3 more times. The pellet will become increasingly clear or white with each differential spin, indicating the presence of epithelial cells. Resuspend the pellet in 10 ml of DMEM/F-12/FBS. Determine organoid density by transferring 20 µl of the organoid suspension into a small Petri dish using a BSA-coated pipette. Count organoids using an inverted phase contrast microscope at 20X magnification.

Derivation of C3(1)-Tag and C3(1)-TagM primary cell lines

Transfer organoid suspension to a sterile Petri dish and incubate at 37°C, 5% CO_2. Due to the absence of a 3D matrix (e.g. collagen or Matrigel), the cells should spread and begin to form a monolayer of cells over the next few days. Check cells under an inverted phase microscope using low (4X) and high power (10X and 20X) objectives microscope to observe proliferation and morphology on a daily basis. Change media when necessary or if dead cells are in suspension. Perform a cell passage with regular trypsinization once cells reach 80% confluency in the Petri dish. This may take ~4-7 days depending on the cell type.

Cell Passage

These methods were adapted from previously a published protocol (Masters et al, 2007). Pre-warm DMEM/F-12/FBS and Trypsin-EDA 0.25% to 37°C. In a biosafety cabinet, aspirate cell medium from the Petri dish. Transfer 5 ml of sterile DPBS to remove serum from cells. Aspirate

DPBS and add 4 ml of trypsin to the cells. Make sure that the trypsin covers the entire surface of the Petri dish by carefully tilting the flask. Cover the Petri dish and incubate at 37°C for 5 minutes. Under an inverted microscope, inspect cells and observe if they have rounded up and detached. Sharply tap the bottom edge of the Petri dish to help detach cells if necessary. If cell detachment is incomplete, return Petri dish to the incubator for 2-3 more minutes. Once detachment is complete, return cells to the biosafety cabinet and add 6 ml of DMEM/F-12/FBS to neutralize the trypsin. Transfer cell suspension to a new 15 ml conical and centrifuge the cells at 1500rpm for 5 minutes. Aspirate the supernatant and resuspend the pellet in 10 ml of fresh cell media. Pipette the cell suspension ~5 times to separate clumps of cells in the mixture. Next, count the cells using a hemocytometer. Pipette 10-15 ml of DMEM/F-12/FBS into a new adherent T-75 cell culture flask. Transfer the appropriate concentration of cell suspension to a new the T-75 flask and gently pipette to provide a more even distribution of cells in the flask. Incubate the T-75 flask at 37°C, 5% CO_2. Check cells daily under an inverted phase microscope and replace media if it begins to turn yellow (acidic) or if many dead cells are present in the suspension. For all subsequent experiments, utilize C3(1)-TagM and C3(1)-Tag cell lines at a low cell passage (<10) for reliable and reproducible results. For immunofluorescence staining of cell lines in 2D, $1x10^5$ cells/well were plated in 6-well plates. On the next day, the cells were washed with DPBS once then fixed in 4% PFA in DPBS for 10 min at room temperature then washed 3 times for 5 min with DPBS.

Freezing cells

After passaging cells to a new T75 flask, use the remaining cell suspension for freezing. Calculate the volume of cell suspension required for a concentration of $2x10^6$ cells/ ml per cryotube. Spin cell suspension for 5 minutes at 1500rpm. Resuspend the pellet in Bambanker

Freezing Media (250 µl freezing media per $2x10^6$ cells) and transfer to a sterile cryotube. Place the cryotube in cell freezing container and transfer to -80°C freezer for 24h. Cells can be kept in -80°C freezer or can be transferred to liquid nitrogen for storage.

Thawing Cells

Prewarm cell culture media (DMEM/F-12/FBS) to 37°C. In a biosafety cabinet, transfer 10 ml of media to a 50 ml sterile conical. Quickly transfer the cryotube with frozen cells from the -80°C freezer to a water bath at 37°C. Check cells every 15s as they thaw in less than 2 minutes on average. Once the cells have thawed, quickly pipette the thawed cells into the 50 ml conical containing cell culture media. Next, carefully transfer the cell suspension to a sterile T-75 cell culture flask and incubate the cells at 37° at 5% CO_2. It is not necessary to pipette the cell suspension once transferred to the T-75 flask as this could cause additional stress to the cells. Change the cell media on the next day to remove dead cells or as required.

Spheroid Formation Assay

In a biosafety cabinet, transfer 10ml DPBS to a Petri dish. The DPBS provides moisture so that the hanging droplets of cell suspension do not dry out during the incubation period that may range from 24-72 hours. Coat a 50ml conical with BSA. Aspirate BSA entirely and transfer 10ml of fresh cell medium. Pipette the $2,5 \times 10^4$ or $3,75 \times 10^4$ cells/ml into the 50 ml conical to generate spheroids that contain 500 or 750 cells, respectively. Invert the tube 3-5 times to mix contents. Take Petri dish lid and invert it so that the interior side facing up. Coat each new pipette tip with BSA and carefully pipette 60-100 20 µl droplets onto the interior surface of the Petri dish lid. Invert the conical throughout this process to make sure that cells do not begin to settle at the bottom

of the tube. Leave a small distance between droplets so that they do not collide with one another when the Petri dished is moved. Next, quickly invert the Petri dish lid onto the Petri dish. This will leave the droplets "hanging" from the Petri dish. Gravity and the shape of the droplet allows the cells to cluster at the bottom of each drop. Each 20 µl droplet yields one spheroid composed of 500 or 750 cells depending on the concentration of cells used. Incubate Petri dish at 37°C, 5% CO_2 for 24-72h. Wait at least 24 hours prior to checking for spheroid formation. Inverting the droplets over too early may disrupt spheroids that still composed of loose cell clusters. After 72 hours, check to see if spheroids have formed using an inverted phase microscope at 20X magnification. A round spheroid or cluster of cells should be present in each droplet. If a spheroid cannot be seen in the center of the droplet, be sure to check the entire surface area including the edges of each drop, as spheroids may move within the droplets when the Petri dish lid is inverted to expose the droplets.

Collecting Spheroids

Prewarm DMEM/F-12/FBS to 37°C. In a sterile culture hood, coat a 50 ml conical with BSA. Transfer the Petri dishes containing "hanging" droplets to the culture hood. Quickly invert the Petri dish lid to expose the droplets. Use the bottom of a sterile Petri dish as a lid to maintain the droplets sterile and check to see if spheroids have formed using an inverted phase-contrast microscope. Transfer the spheroids back to the culture hood and invert the lid of the Petri dish once again to expose the droplets. With the droplets still facing up, place the lid at an angle using the bottom of the Petri dish for support. Using a P1000, pipette 1 ml of DMEM onto the spheroids and allow the droplets to pool together at the edge of the lid. Coat a new pipette tip with BSA and gently pipette the media directly over the spheroids to detach them from the Petri dish lid. Make

sure the entire surface of the Petri dish is covered in cell media to prevent the loss of spheroids. It may be necessary to repeat this step 1-2 more times to detach all spheroids from the lid. Collect the spheroid suspension using a p1000 pipette and transfer to a 50 ml conical coated with BSA.

Collagen Invasion Assay

For the collagen invasion assay, we plated a minimum of 45 spheroids per well in a 24-well plate with a glass-bottom. The following methods are adapted from a previously established protocol (Padmanaban 2020). Rat tail collagen was prepared at a 3mg/ ml concentration and allowed to slowly polymerize at 4°C. Perform all steps to prepare the collagen mixture on ice and in a biosafety cabinet.

For a 3.5 ml collagen I solution, combine 375 µl of 10X DMEM and 100 µl of NaOH in a 15 ml conical tube. Pipette the solution to mix contents. Once homogenized, the solution should be dark pink in color. Slowly transfer 3 ml of rat tail collagen I to the conical using a 5 ml serological pipette. Be careful while pipetting to avoid forming air bubbles in the collagen solution. Collagen I is highly viscous and will require constant and careful pipetting for 1-2 minutes. If the resulting solution is pink or dark pink, this indicates that it is basic and needs to be titrated with collagen I (20-40 µl). If the resulting solution is yellow, it is acidic and needs to be titrated with small volumes (1-10 µl) of NaOH. The goal is to attain a stable light pink or salmon color, which corresponds to a pH of 7. If an optimal pH of 7 is not achieved, both the collagen polymerization time and fiber density will be affected. The pH can also be confirmed with the use of pH strips. Place a 24-well coverslip-bottomed imaging plate on a heat block set to 37°C. Plate 15 µl underlays in each well of the plate. Incubate the underlays at 37 °C for 30–60 min before plating. Next, transfer the 15 ml conical containing collagen to a 4°C fridge for polymerization. Check the conical

every 30 minutes to see if polymerization has begun. The collagen solution will become cloudy and translucent when collagen fibers begin to form. The fibers can be visualized under an inverted microscope for confirmation, if necessary. Once the collagen solution starts to polymerize, transfer the 15 ml conical to a biosafety cabinet and keep it on ice. Take the 24-well plate with collagen underlays and place it on the heating block for 5 minutes. Spin down the conical with the spheroids at 1500 rpm for 5 minutes. Carefully aspirate the supernatant without disrupting the pelleted spheroids. Resuspend the pellet in the collagen solution. For a 24-well plate, 100-12 µl of collagen is recommended per well. Transfer 100-120 µl spheroids resuspended in collagen to the bottom of the each well containing a collagen underlay. Pipette gently to avoid the creation of bubbles. Incubate the 24-well plate at 37°C for 1 hour prior to adding media. Add 1 ml of DMEM/F-12 media to each well with the corresponding treatments listed in Table X. Maintain the plate in the incubator at 37°C 5% CO_2 for the duration of the experiment. Change the media every 2 days. On day 5, rinse the gels once with DPBS, then fix the spheroids in 4% PFA for 10 minutes, then wash in DPBS 3x for 10 minutes. Next, perform immunofluorescence staining.

Immunofluorescence of 2D cell lines and 3D embedded spheroids

On a laboratory bench, aspirate the DPBS from each well containing previously fixed cells. Permeabilize the cells using 300 µl 0.5% Triton-X dissolved in DPBS (Table 1). Incubate the plate for 1 hour at room temperature with gentle shaking. Remove the 0.5% Triton-X and block each well with 300 µl Blocking Buffer for 2-3 hours. Aspirate the blocking solution and add 300 µL primary antibody dissolved in Antibody Dilution Buffer. Wrap the plate in aluminum foil and incubate overnight at 4° with gentle shaking. Remove the primary antibody diluent and add 300 µl of secondary antibody dissolved in Antibody Dilution Buffer. Incubate for 2-3 hours at room

temperature with gentle shaking. Remove the secondary antibody and perform 3 x 10-minutes washes using 1 ml DPBS in each well. Wrap the stained plate in aluminum foil. Stained plates can be stored in DPBS at 4 °C. Image the plate within 48h. Primary antibodies used include rabbit Vimentin (Abcam, ab92547, 1:250) and rat E-cadherin (Abcam, ab11512, 1:100). All secondary antibodies used were Alexa Fluor-conjugates (Invitrogen, 1:200). Confocal images were taken using a Zeiss 780 laser-scanning confocal microscope. Fiji was used to adjust brightness/contrast levels and reduce noise background, add scale bars, and export the images. Schematic diagrams were created using Procreate.

Table 1: Summary of immunofluorescence staining of cell lines

Step	**Reagents for 50 ml stock solution**	**Time in gel**
0.5%Triton-X Permeabilization Buffer	2.5 ml Triton X 47.5 ml DPBS	1 hour
Blocking Buffer	500 µl Triton X 5 ml FBS 500 µl BSA 44 ml DPBS	2-3 hours
Antibody Dilution Buffer	500 µl Triton-X 500 µl FBS 500 µl BSA 48.5 ml DPBS	Primary Antibody: Overnight at 4°C Secondary Antibody: 2-3 hours at RT

Colony formation assay

In a biosafety cabinet, perform a cell passage with shorter trypsinization time (about 4 minutes) to avoid dissociating the C3(1)-TagM and C3(1)-Tag cells into single cells. After trypsinization, resuspend the cells in a 15ml conical using 10ml DMEM/F-12/FBS. Pipette the cell suspension no more than 2-5 times to avoid dissociating the cell clusters. Transfer the tube to a centrifuge and spin down at 1500rpm for 30 seconds. Next, aspirate the supernatant carefully

without disrupting the loose cell pellet to get rid of the single cells in suspension. Next, resuspend the pelleted cell clusters in 10ml DMEM/F-12/FBS once again and perform a cell count using a hemacytometer. Verify that all cell clusters contain at least two cells. It is important to minimize the presence of single cells, if any transfer the cell suspension to a centrifuge and spin again at 1500 rpm for 30 seconds. Aspirate the supernatant and plate the conical on ice. Resuspend the pellet in the appropriate volume of Matrigel. We used 6000 clusters of cells/well. Next, move a heat block into the biosafety cabinet set at 37°degrees. Pre-incubate a 96-well plate for at least 5 mins. Plate 15 µl of cell clusters resuspended in Matrigel into the center of each well. Incubate the 96-well plate for at least 45 minutes prior to adding media. Add 100 µl of media to each well with the corresponding treatments Table 2. Maintain the plate in the incubator at 37°C 5% CO2 for the duration of the experiment. Change the media every 2 days. On day 5, fix the cell clusters in 1 % PFA with 0.25 % glutaraldehyde. Since C3(1)-TagM cells express TdTomato, they can readily be visualized without requiring staining. To visualize C31T cell clusters in Matrigel, stain using Phalloidin diluted in Antibody Dilution Buffer (Table X) for 2-3 hours. It is not necessary to permeabilize the cells or to perform blocking. All wells containing cell clusters in Matrigel were imaged in 3D using a Molecular Devices ImageXpress Micro High Content Analysis System.

Table 2: Conditions defined for 3D Invasion Assays using C3(1)-TagM and **C31T cell lines**

Condition 1	No FBS	DMEM/F-12, 1% (vol/vol) Penn-step, 1% (vol/vol) Insulin-transferrin selenium
Condition 2	1% FBS	DMEM/F-12, 1% (vol/vol) Penn-step, 1% (vol/vol) Insulin-transferrin selenium, 1% (vol/vol) FBS
Condition 3	TGFβ receptor 1 inhibitor	DMEM/F-12, 1% (vol/vol) Penn-step, 1% (vol/vol) Insulin-transferrin selenium, 1% (vol/vol) FBS, 1% (vol/vol) TGFβ receptor 1 inhibitor (SB-525334, 1uM)
Condition 4	TGFβ 1	DMEM/F-12, 1% (vol/vol) Penn-step, 1% (vol/vol) Insulin-transferrin selenium, 1% (vol/vol) FBS, 1% (vol/vol) TGFβ 1 (Peprotek 2nM)

Statistical analysis

Statistical analysis and graphs were performed using Prism 8 (GraphPad). We used the Kruskal–Wallis test to analyze data without normal distribution. Statistical significance was considered starting with $P < 0.05$.

Results

Derivation of C3(1)-TagM and C3(1)-Tag cell lines

To derive the primary cell lines, mammary tumors were isolated directly from C3(1)-TagM and C31T mice. The mouse tumors were minced and enzymatically digested to enrich for epithelial organoids (Figure 1A). The organoids were then plated on 10 cm Petri dishes and transferred to a cell culture incubator at 37°C to evaluate whether the cells would survive and proliferate on the Petri dishes. Extensive cell death was observed in both C3(1)-TagM and C3(1)-Tag Petri dishes over the next couple of days. Nonetheless, subpopulations of C3(1)-TagM and C3(1)-Tag cells survived in their new environment, proliferated, and began to form monolayer a of cells in each Petri dish.

Figure 1B shows C3(1)-TagM cells 3 days after being plated on the Petri dish. The area outlined in red shows a C3(1)-TagM organoid that seemed to have partially adhered to the bottom of the Petri dish; the cells surrounding the organoid were tightly packed compared to cells visualized in other areas of the Petri dish (Figure 1B). In some areas of the Petri dish, we noted a subset C3(1)-TagM cells that displayed a cobble stone-like phenotype closely resembling epithelial cells (Figure 1B; area traced in yellow). However, C3(1)-TagM cells that exhibited a mesenchymal-like phenotype were also observed (Figure 1B; blue arrows). C3(1)-TagM cells

required 4 days to reach >80% confluence after being plated. In comparison, C3(1)-Tag cells required 5 days reach >80% confluence in the Petri dish.

Following the first cell passage (PI), both cell lines were cultured in T-75 flasks. Three days after the third cell passage (P3), C3(1)-TagM cells presented a more mesenchymal-like phenotype with cells for the first time (Figure 1C). The C3(1)-Tag cell line adopted a similar phenotype within a couple of days although the cells proliferated more slowly compared to C3(1)-TagM. A minimum of 5 cell passages (P5) were performed before either one of the cell lines adopted a standard pattern of growth in culture reaching a concentration of $1x10^6$ cells/ ml after 4-5 days. C3(1)-TagM and C3(1)-Tag cell lines were cultured for 2.5 months with regular trypsinization prior to being utilized in experiments.

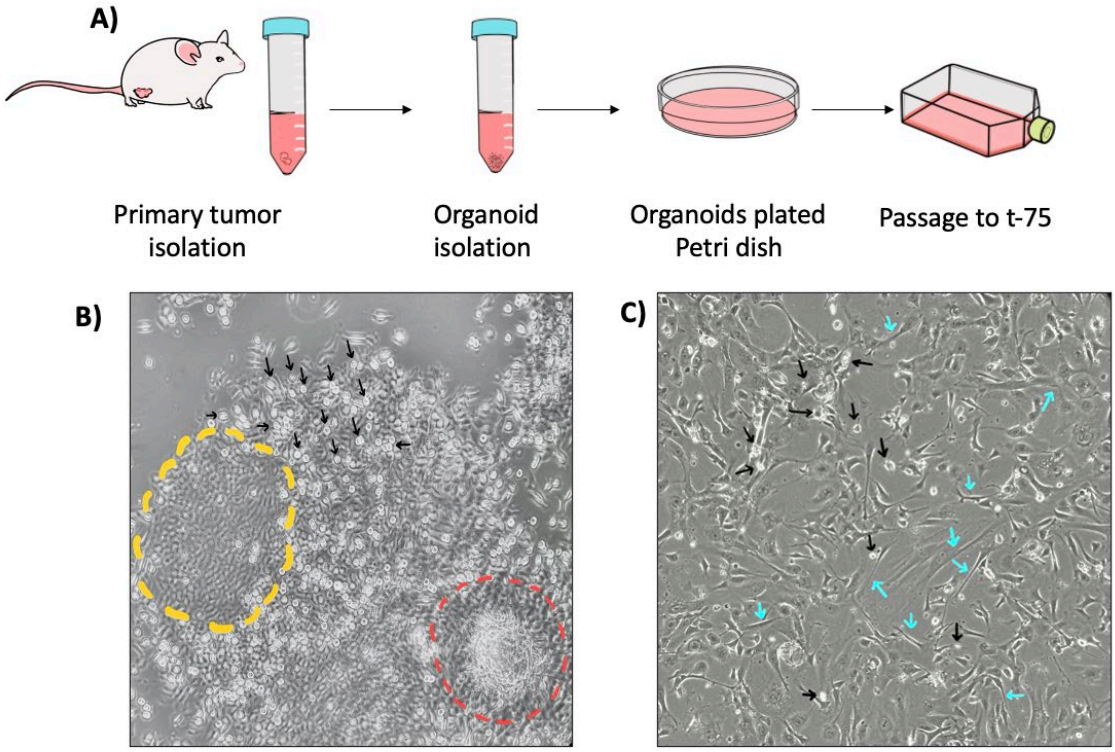


Figure 1: Derivation of C3(1)-TagM and C3(1)-Tag cell lines. A) Schema for derivation of cell lines using mouse mammary tumor organoids. Primary tumors were isolated from the mouse and enzymatically digested to generate epithelial organoids. The organoids were then plated in 10 cm Petri dishes and transferred to a cell culture incubator at 37°C, 5% CO_2. B) iPhone 10 picture of C3(1)-TagM cells 3 days after being plated in the Petri dish. Cells were visualized under an inverted phase-contrast microscope at 20X magnification. Black arrows point to dead cells in suspension. The area traced in yellow highlights C3(1)-TagM cells with epithelial-like characteristics. The area traced in red shows an organoid adhered to the bottom of the Petri dish plate. C) iPhone 10 picture of C3(1)-TagM cells visualized at 20X magnification after first cell passage (P1). Black arrows point to dead cells in suspension. Blue arrows point to cells displaying long and thin cell processes similar to mesenchymal cells.

C3(1)-TagM and C3(1)-Tag cell lines express EMT markers in 2D cell culture

To confirm our observation that C3(1)-TagM and C3(1)-Tag cell lines display hybrid-EMT characteristics in 2D cell culture, C3(1)-TagM and C3(1)-Tag cells were plated in 6-well plates. After one day, both cell lines were stained with antibodies against E-cadherin (E-Cad) and vimentin (Vim), the most common markets to assess epithelial and mesenchymal phenotypes, respectively (Lu and Kang, 2019). We observed that both cell lines contained cells that expressed E-Cad and/or Vim (Figure 2A and 2B). C3(1)-TagM and C3(1)-Tag cells were manually counted to determine the percentage of cells that were double-positive for E-cad and Vim (Ecad+/Vim+). Our quantification analysis shows that C3(1)-TagM cells were 5.9% Ecad+/Vim+ and 94% E-Cad+/Vim- (Figure 2B). In comparison, the C3(1)-TagT cell line contained a greater proportion of cells that displayed a partial EMT phenotype, with 23.7% of the cells being E-Cad+/Vim+ and 76.2% being Ecad+/Vim-. Similar to our work in vivo and in primary C3(1)-Tag cells, neither of the cell lines seemed to consist of cells Vim+/E-Cad- as this phenotype was not observed in image analysis. Taken together, our data suggests C3(1)-TagM and C3(1)-Tag cell lines maintain hybrid EMT characteristics in 2D cell culture.

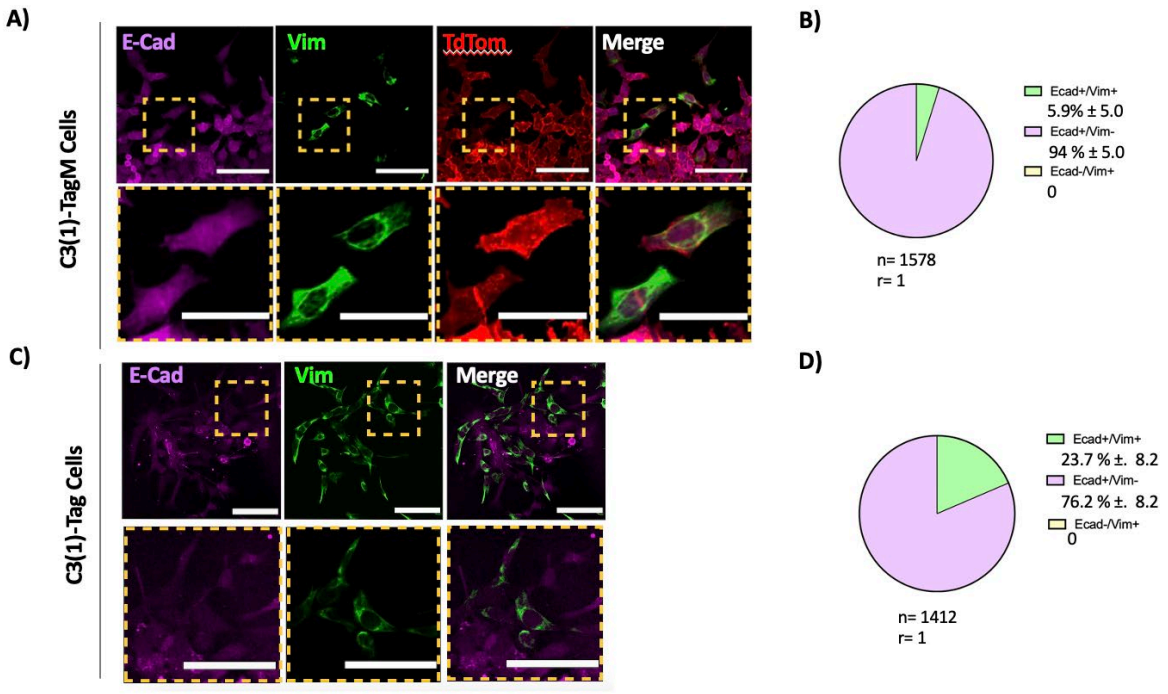


Figure 2: C3(1)-TagM and C3(1)-Tag primary cell lines express epithelial and mesenchymal markers in 2D cell culture. Representative confocal image of C3(1)-TagM (a) and C3(1)-Tag (b) isolated cell lines stained with anti-Ecad and anti-Vim antibodies. The areas outlined in yellow (top) show cells expressing both E-Cad and Vimentin in vitro. Bottom images show the zoomed-in areas. Scale bars are 100 μm. (c) and depict the distribution of the different cell states presented by each cell line. Numbers represent mean percentage of cells per condition ± SD. N= total number of cells counted per cell line. R= number of replicates per experiment.

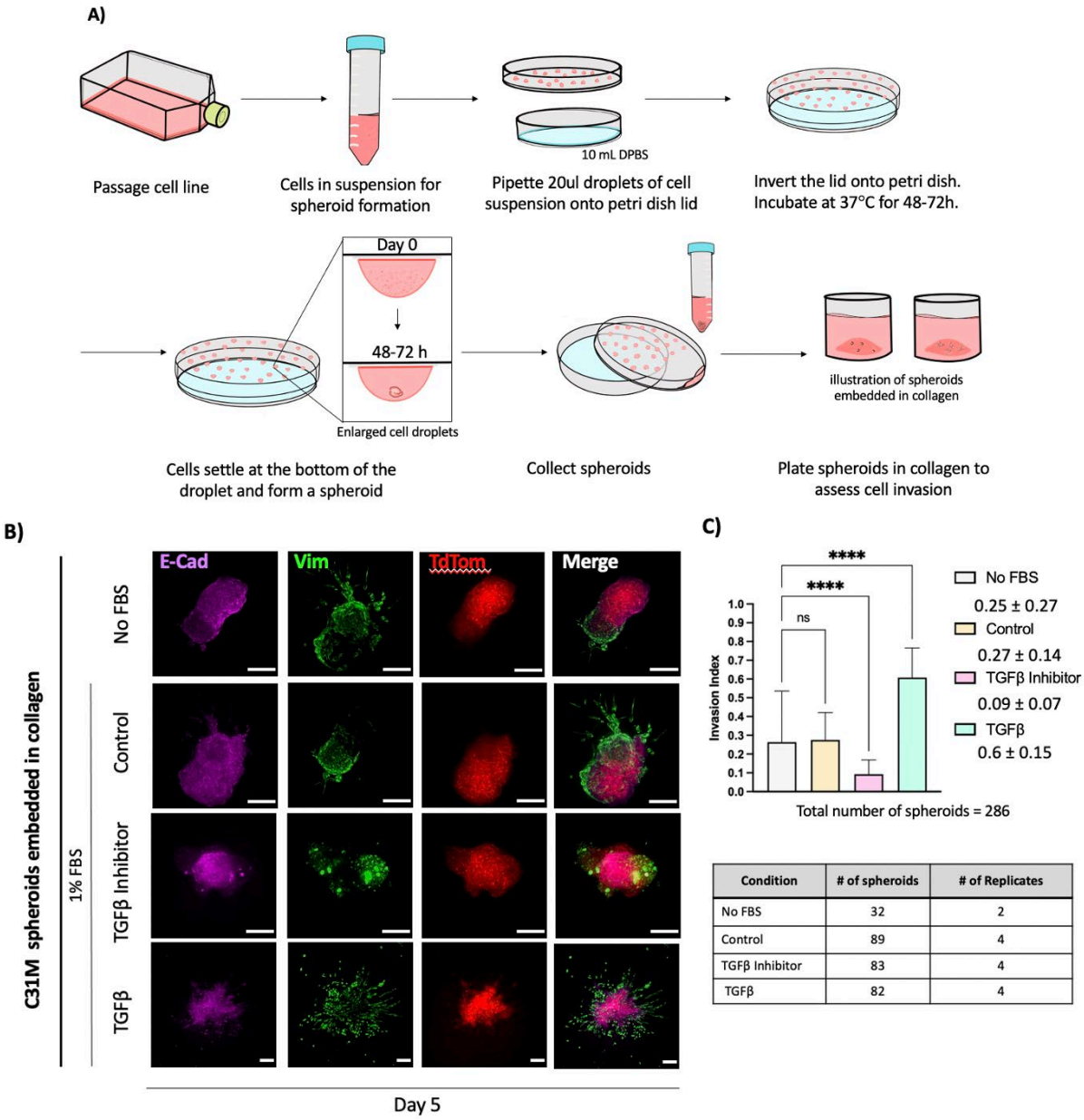
TGFβ induces C3(1)-TagM and C3(1)-Tag invasion in 3D collagen gels

To assess whether C3(1)-TagM and C3(1)-Tag cell lines could model 3D cell invasion *in vitro*, we sought to validate if the cell lines could successfully form spheroids, or 3D spherical aggregates of cells. To form spheroids using cells from each cell line, we utilized a variation of the "hanging drop" method which relies on plating droplets of cell suspension and subsequently inverting them. Gravity causes the cells to settle and aggregate at the bottom of the droplet, meanwhile forces of cell-cell adhesion and cell growth contribute to spheroid formation (Timmins & Nielsen, 2007). We plated 20 μl droplets consisting of 500-750 cells from each cell line onto interior surface of a 10 cm Petri dish lid, carefully inverted the lid, and placed it on top of a Petri dish containing DPBS (Figure 3A). The Petri dishes were then incubated at 37° C for 3 days. Following incubation, we observed the droplets under a microscope and confirmed the presence of spheroids.

In our recent manuscript, we demonstrated that the TGFβ1 promotes EMT and breast cancer invasion both *in vitro* and *in vivo* using the C(3(1)-Tag mouse model. We noticed that TGFβ is correlated with vimentin expression during invasion and that vimentin expression is required for invasion in TNBC models. Here, we sought to investigate whether C3(1)-TagM and C3(1)-Tag spheroids displayed similar patterns of invasion to C(3(1)-Tag organoids cultured *in vitro*. C3(1)-TagM and C3(1)-Tag spheroids were cultured in 1% FBS, No FBS, TGFβ receptor 1 inhibitor, or TGFβ1 for 5 days (Figures 3B and 4B). On day 5, C3(1)-TagM and C3(1)-Tag spheroids were fixed with 4% PFA and stained with antibodies against E-Cad and Vim (Figure 3B and 4B). Most C3(1)-TagM and C3(1)-Tag spheroids contained cells that exhibited hybrid EMT and were double-positive for E-Cad+/Vim+ (Figures 3B and 34). We also observed a large portion of C3(1)-TagM

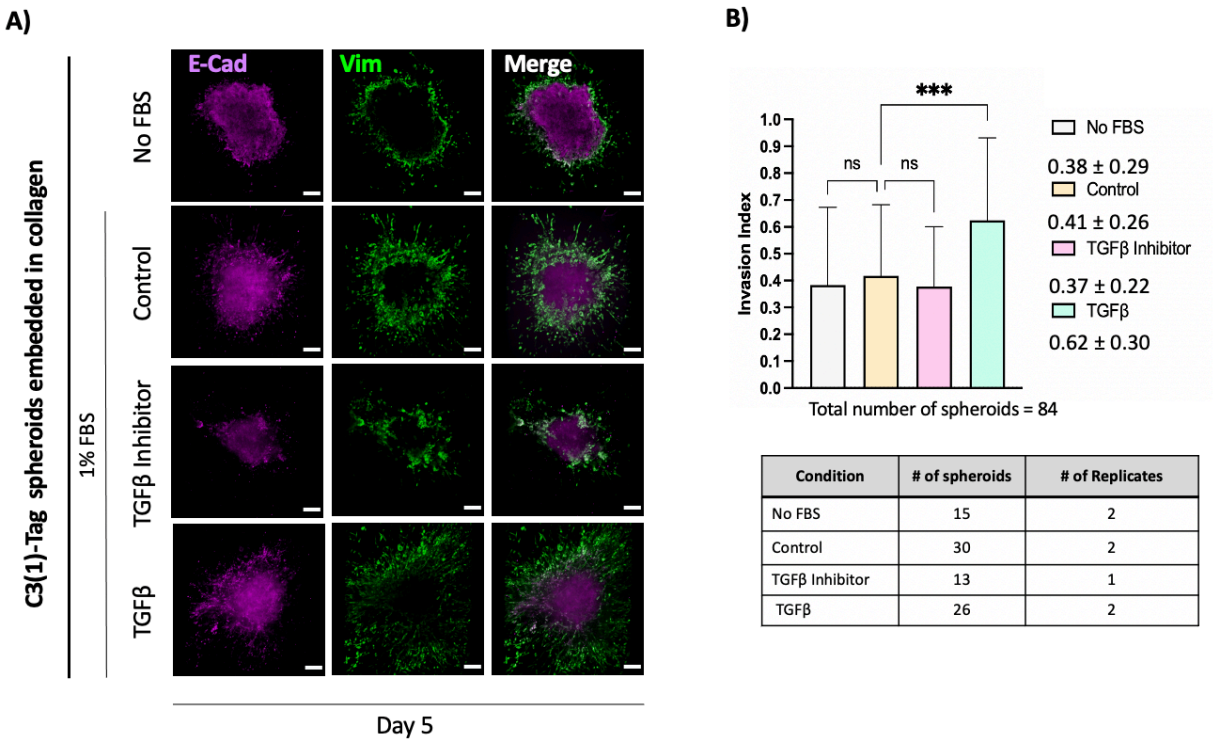
and C3(1)-Tag spheroids were E-Cad+/Vim+ in cells that lead the invasive. However, spheroids remained mainly E-Cad+/Vim- in the core (Figures 2B and 3B; Appendix 1, Figure 1B).

To assess the invasion index of the spheroids, we used Fiji to measure the non-invasive and the invasive area for each spheroid analyzed. The invasive index was calculated using (1- non-invasive area/invasive area). Our statistical analysis showed that there is no significant difference in invasion among C3(1)-TagM and C3(1)-Tag spheroids cultured without FBS or in the presence of 1% FBS (Figures 3C and 4C). However, our data suggests that C3(1)-TagM spheroid invasion may be suppressed in the presence of TGFβ inhibitor compared to C3(1)-TagM spheroids cultured in 0% FBS (Figures 3B and 4C). Our results also showed that TGFβ signaling induced invasion in both C3(1)-TagM and C3(1)-Tag spheroids compared to spheroids cultured in 0% FBS or 1% FBS (Figures 3B, 3C, 4A, 4B). Our data also demonstrates that the presence of TGFβ inhibitor significantly reduced invasion in C3(1)-TagM spheroids (Figures 3B and 3C). Taken together, these results suggest that C3(1)-TagM and C3(1)-Tag cell lines can be utilized to derive spheroids that model EMT invasion in collagen.



Condition	# of spheroids	# of Replicates
No FBS	32	2
Control	89	4
TGFβ Inhibitor	83	4
TGFβ	82	4

Figure 3: TGFβ signaling induces C3(1)-TagM Spheroid Invasion in Collagen. A) Schema of the spheroid formation assay. A) 3D Invasion Assay of C3(1)-TagM spheroids embedded in collagen for 5 days with 1% FBS, No FBS, TGFB Inhibitor, or TGFB. Representative confocal images of maximum intensity projection of C3(1)-TagM spheroids immunostained with anti-E-Cad and anti-vim antibodies. Scale bars are 100 μm (C) Histogram of C3(1)-TagM spheroids invasion index. Columns correspond to the mean ± SD per condition. The table below indicates the number of spheroids and replicates per condition. ****P<0.0001 (Kruskall-Wallis).



Condition	# of spheroids	# of Replicates
No FBS	15	2
Control	30	2
TGFβ Inhibitor	13	1
TGFβ	26	2

Figure 4: TGFβ signaling induces C3(1)-Tag Spheroid Invasion in collagen I. A) 3D Invasion Assay of C3(1)-Tag spheroids embedded in collagen for 5 days with 1% FBS, No FBS, TGFβ inhibitor, or TGFβ. Representative confocal images of maximum intensity projection of C3(1)-Tag spheroids immunostained with anti-E-Cad and anti-Vim antibodies. Scale bars are 100 μm (C) Histogram of C3(1)-Tag spheroids invasion index. Columns correspond to the mean ± SD per condition. The table below indicates the number of spheroids and replicates per condition. ****P<0.0001 (Kruskall-Wallis).

TGFβ signaling may regulate C3(1)-TagM and C3(1)-Tag colony formation in Matrigel

Previously, we demonstrated that our *ex vivo* colony formation assay models cancer cell proliferation and metastatic outgrowth. In our recent manuscript, we showed that contrary to the collagen invasion assay, C3(1)-Tag cells embedded in Matrigel exhibited a decrease in vimentin expression and an increase in E-Cad expression. Here, we sought to investigate and compare the effects of TGFβ on C3(1)-TagM and C3(1)-Tag colony formation. For this, we cultured cell clusters of each cell line embedded in Matrigel in 1% FBS, No FBS, TGFβ receptor 1 inhibitor, or TGFβ1 for 5 days (Figure 5A and 5B) Our data suggests that there was no significant difference in the number of colonies formed after 5 days when C3(1)-TagM and C3(1)-Tag cells are cultured without FBS or in the presence of 1% FBS. Our statistical analysis showed that treatment with TGFβ receptor 1 inhibitor significantly reduced the number of colonies formed after 5 days in both C31M and C31T cell. We also observed a significant decreased in colony formation in C3(1)-TagM cells cultured in the presence of TGFβ. Taken together, these data suggest that TGFβ signaling may play a role in inhibiting colony formation in C3(1)-TagM and C3(1)-Tag cell lines.

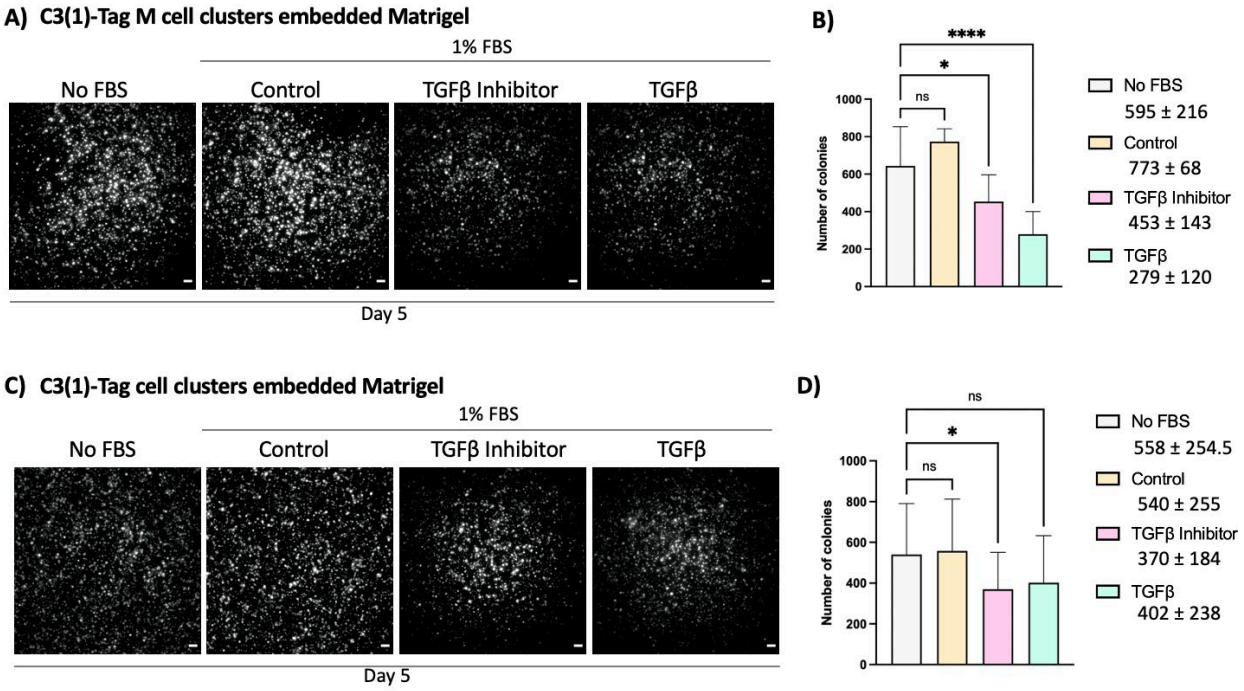


Figure 5: TGFβ signaling may regulate C3(1)-TagM and C3(1)-Tag colony formation in Matrigel. A and C) 3D Colony Formation Assay of C3(1)-TagM and C3(1)-Tag cell clusters embedded in Matrigel for 5 days with 1% FBS, No FBS, TGFβ inhibitor, or TGFβ. Cell clusters were imaged at 4X. Scale bars are 100 μm. B and D) Numbers indicate the mean number of cell colonies ± SD per condition. ***P<0.0001 *P= 0.0474 (Kruskall-Wallis)

Discussion

Over the few past decades, our knowledge of breast cancer has increased significantly. Research has provided us with key insights into better understanding the processes that drive tumor cell invasion and metastasis. Nevertheless, metastasis remains responsible for the majority of deaths among BC patients. For patients diagnosed with TNBC, the prevalence of disease relapse and metastasis is even more significant compared to all other BC subtypes. Recent studies have shown that TNBC cells exhibit hybrid EMT characteristics. Although EMT is a controversial topic among researchers, recent studies have confirmed the presence of a hybrid EMT phenotype among TNBC cells from patients (Alix-Panabières et al, 2017; Francart et al, 2018). In addition, we recently demonstrated across murine and human models that TNBC tumors exhibit hybrid EMT phenotypes. This was shown using the basal-like C3(1)-Tag mouse model, multiple PDXs, and TNBC patient samples (Appendix 1). This novel finding is what led to the development of my thesis project, as we sought to further investigate whether cell lines isolated from ROSA$^{mT/mG}$ C3(1)-Tag and C3(1)-Tag mice would maintain hybrid EMT characteristics when cultured in 2D and 3D.

The difference between the two cell lines is that ROSA$^{mT/mG}$ C3(1)-Tag cells express constitutive levels of the fluorescent membrane-targeted tandem dimer, TdTomato. Cre-mediated excision can be used to drive the expression of membrane-targeted green fluorescent protein (GPF) (Muzumdar et al, 2007). Although we did not utilize Cre-lox recombinant technology in this study, the C3(1)-TagM cell line could easily be distinguished from the C3(1)-Tag cell line without requiring additional IHC staining. In addition, the mT/mG, double-fluorescent construct enables cell labeling at single cell resolution, which could be a useful tool for future live cell imaging.

For this thesis project, I utilized 2D and 3D cell culture techniques to characterize C3(1)-TagM and C3(1)-Tag isolated cell lines and compare them to our previous work done in vivo and ex vivo using freshly isolated organoids (Appendix 1). I demonstrated that the isolated C3(1)-TagM and C3(1)-TagM cell lines maintain subpopulations of cells displaying hybrid EMT phenotypes in vitro. To further characterize C3(1)-TagM and C3(1)-Tag cell lines, additional replicates of each experiment are required. I was unable to achieve 3 experimental replicates of each experiment due to limitations imposed by Covid-19. This affected the statistical analysis of our results because we were only able to use the non-parametric Kruskall-Wallis test to analyze my data as the assumptions of one-way ANOVA were not met. Both one-way ANOVA and Kruskall-Wallis are used to assess for significant differences on a continuous and dependent variable by two or more groups. However, one-way ANOVA assumes that the dependent variable displays equal variance or a normal distribution among all groups tested, which Kruskall-Wallis does not. For this reason, statistical analysis using Kruskall-Wallis is not as powerful as one-way ANOVA. Nonetheless, this preliminary data suggest that cell lines cultured in 2D monolayers exhibit differences in the expression of EMT genes compared to 3D culture methods (consistent with De Meulenaere et al, 2020; Melissaridou et. al, 2019). Indeed, while we detected only a small percentage of hybrid EM cells in 2D culture, we observed an enrichment for hybrid EM cells in collagen invasion assay with hybrid EM cells leading invasion in 3D.

Moreover, the data revealed that spheroids derived from C3(1)-TagM and C3(1)-Tag cell lines exhibited similar features to primary C3(1)-Tag organoids embedded in 3D matrices, suggesting that these cell lines could be a physiologically relevant model of basal breast cancer. I confirmed that co-expression of E-cadherin and vimentin is present in both cell lines when cultured in 2D and in 3D (Figure 1). My results also suggest that while TGFβ signaling induces C3(1)-

TagM and C3(1)-Tag spheroid invasion in collagen, it may play a role in inhibiting colony formation in Matrigel (Figures 3, 4 and 5). TGFβ has been shown to induce EMT. In accordance with this, I observed an increase in invasion and vimentin positive cells in spheroids exposed to TGFβ in collagen matrix. Our colony formation assay is used to model metastatic outgrowth. The potential inhibition of colony formation in presence of TGFβ could indicate that mesenchymal gene expression represses colony formation suggesting a role of MET for metastatic outgrowth in these cells. In our previous work, we demonstrated using shRNA that vimentin represses colony formation. It would be interesting to reproduce this experiment in these cell lines. My results also demonstrated that although C3(1)-TagM and C3(1)-Tag cell lines displayed hybrid EMT phenotypes in 2D and 3D cell culture, the mesenchymal phenotype was more prevalent when cells were embedded in collagen (Figures 3 and 4). Compared to 2D culture, C3(1)-TagM and C3(1)-Tag spheroids embedded in collagen consisted of a majority of cells that were positive for vimentin. Similar to what we observed in our recent manuscript, the leading cells located at the tip of invasive strands of spheroids were all positive for vimentin.

To further validate hybrid EMT and characterize C3(1)-TagM and C3(1)-Tag cell lines, it would be interesting to assess the effects of TGFβ when the cell lines are cultured in 2D. Although this was attempted, both TGFβ 1 and its inhibitor counterpart showed to be toxic to cells plated at $1{,}5 \times 10^5$ when added at the same concentration that was utilized for the collagen invasion and colony formation assays. It is likely that the dose of TGFβ-1 and TGFβ-1 receptor inhibitor would have to be decreased to assess their effects on the cell lines in 2D. Nonetheless, it would be interesting to compare the effects of TGFβ or anti-cancer drugs in both cell lines cultured in 2D vs 3D. As a next experimental step, I would perform qPCR on both C3(1)-TagM and C3(1)- cell lines treated in the presence of TGFβ to assess for quantitative changes in EMT gene expression.

Future Considerations

In the most recent decade, there has been a shift towards organotypic culture methods, especially in the field of cancer research. It has been observed that when cells are cultured in 2D, the lack of cell-cell interaction and cell-ECM interactions make it difficult to assess cell behavior and response to anti-cancer therapies (Cree et al, 2011. The use of multicellular 3D spheroids will continue to grow and become ever more popular among the research community, for they mimic primary tumor cell behavior similar to what has been observed in 3D organoids. By isolating cell lines from the C3(1)-Tag mouse model of TNBC, we were able to demonstrate that hybrid EMT phenotypes are maintained across 2D and 3D culture systems. However, the 3D spheroids generated using both C3(1)-TagM and C3(1)-Tag cell lines showed greater similarity to organoids isolated from the same mouse model, futher suggesting that 3D culture technology most likely represents the future of breast cancer and metastasis research.

References

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D. and Bray, F. (2015), Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int. J. Cancer, 136: E359-E386. https://doi.org/10.1002/ijc.29210

Spitale, A., Mazzola, P., Soldini, D., Mazzucchelli, L., & Bordoni, A. (2009). Breast cancer classification according to immunohistochemical markers: clinicopathologic features and short-term survival analysis in a pop μlation-based study from the South of Switzerland. *Annals of oncology : official journal of the European Society for Medical Oncology*, *20*(4), 628–635. https://doi.org/10.1093/annonc/mdn675

Vallejos, C. S., Gómez, H. L., Cruz, W. R., Pinto, J. A., Dyer, R. R., Velarde, R., Suazo, J. F., Neciosup, S. P., León, M., de la Cruz, M. A., & Vigil, C. E. (2010). Breast cancer classification according to immunohistochemistry markers: subtypes and association with clinicopathologic variables in a peruvian hospital database. *Clinical breast cancer*, *10*(4), 294–300. https://doi.org/10.3816/CBC.2010.n.038

Fedele, M., Cerchia, L., & Chiappetta, G. (2017). The Epithelial-to-Mesenchymal Transition in Breast Cancer: Focus on Basal-Like Carcinomas. *Cancers*, *9*(10), 134. https://doi.org/10.3390/cancers9100134

Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, Shi B. Breast cancer intrinsic subtype classification, clinical use and future trends. Am J Cancer Res. 2015 Sep 15;5(10):2929-43. PMID: 26693050; PMCID: PMC4656721.

Kanyılmaz, G., Yavuz, B. B., Aktan, M., Karaağaç, M., Uyar, M., & Fındık, S. (2019). Prognostic Importance of Ki-67 in Breast Cancer and Its Relationship with Other Prognostic Factors. *European journal of breast health*, *15*(4), 256–261. https://doi.org/10.5152/ejbh.2019.4778

Kast, K., Link, T., Friedrich, K., Petzold, A., Niedostatek, A., Schoffer, O., Werner, C., Klug, S. J., Werner, A., Gatzweiler, A., Richter, B., Baretton, G., & Wimberger, P. (2015). Impact of breast cancer subtypes and patterns of metastasis on outcome. *Breast cancer research and treatment*, *150*(3), 621–629. https://doi.org/10.1007/s10549-015-3341-3

Garrido-Castro, A. C., Lin, N. U., & Polyak, K. (2019). Insights into MolecularClassifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer discovery*, *9*(2), 176–198. https://doi.org/10.1158/2159-8290.CD-18-1177

Dunnwald, L. K., Rossing, M. A., & Li, C. I. (2007). Hormone receptor status, tumor characteristics, and prognosis: a prospective cohort of breast cancer patients. *Breast cancer research : BCR*, *9*(1), R6. https://doi.org/10.1186/bcr1639

Liu, Y., Xin, T., Jiang, Q. Y., Huang, D. Y., Shen, W. X., Li, L., Lv, Y. J., Jin, Y. H., Song, X. W., & Teng, C. (2013). CD147, MMP9 expression and clinical significance of basal-like breast cancer. *Medical oncology (Northwood, London, England)*, *30*(1), 366. https://doi.org/10.1007/s12032-012-0366-x

Herschkowitz, J. I., Simin, K., Weigman, V. J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K. E., Jones, L. P., Assefnia, S., Chandrasekharan, S., Backlund, M. G., Yin, Y., Khramtsov, A. I., Bastein, R., Quackenbush, J., Glazer, R. I., Brown, P. H., Green, J. E., Kopelovich, L., Furth, P. A., … Perou, C. M. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome biology*, *8*(5), R76. https://doi.org/10.1186/gb-2007-8-5-r76

Kalluri, R., & Weinberg, R. A. (2009). The basics of epithelial-mesenchymal transition. *The Journal of clinical investigation*, *119*(6), 1420–1428. https://doi.org/10.1172/JCI39104

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., Campbell, L. L., Polyak, K., Brisken, C., Yang, J., & Weinberg, R. A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, *133*(4), 704–715. https://doi.org/10.1016/j.cell.2008.03.027

Moreno-Bueno, G., Portillo, F., & Cano, A. (2008). Transcriptional reg µlation of cell polarity in EMT and cancer. Oncogene, 27(55), 6958–6969. https://doi.org/10.1038/onc.2008.346

Kim, D. H., Xing, T., Yang, Z., Dudek, R., Lu, Q., & Chen, Y. H. (2017). Epithelial Mesenchymal Transition in Embryonic Development, Tissue Repair and Cancer: A Comprehensive Overview. *Journal of clinical medicine*, *7*(1), 1. https://doi.org/10.3390/jcm7010001

Thiery, J. P., Acloque, H., Huang, R. Y., & Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell*, *139*(5), 871–890. https://doi.org/10.1016/j.cell.2009.11.007

Hay E. D. (1995). An overview of epithelio-mesenchymal transformation. *Acta anatomica*, *154*(1), 8–20. https://doi.org/10.1159/000147748

Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S., & Yang, J. (2012). Spatiotemporal reg µlation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer cell*, *22*(6), 725–736. https://doi.org/10.1016/j.ccr.2012.09.022

Ocaña, O. H., Córcoles, R., Fabra, A., Moreno-Bueno, G., Acloque, H., Vega, S., Barrallo-Gimeno, A., Cano, A., & Nieto, M. A. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer cell*, *22*(6), 709–724. https://doi.org/10.1016/j.ccr.2012.10.012

Beerling, E., Seinstra, D., de Wit, E., Kester, L., van der Velden, D., Maynard, C., Schäfer, R., van Diest, P., Voest, E., van Oudenaarden, A., Vrisekoop, N., & van Rheenen, J. (2016). Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell reports*, *14*(10), 2281–2288. https://doi.org/10.1016/j.celrep.2016.02.034

Alix-Panabières, C., Mader, S., & Pantel, K. (2017). Epithelial-mesenchymal plasticity in circ µlating tumor cells. *Journal of molecularmedicine (Berlin, Germany)*, *95*(2), 133–142. https://doi.org/10.1007/s00109-016-1500-6

Francart, M. E., Lambert, J., Vanwynsberghe, A. M., Thompson, E. W., Bourcy, M., Polette, M., & Gilles, C. (2018). Epithelial-mesenchymal plasticity and circ µlating tumor cells: Travel companions to metastases. *Developmental dynamics : an official publication of the American Association of Anatomists*, *247*(3), 432–450. https://doi.org/10.1002/dvdy.24506

Fantozzi, A., & Christofori, G. (2006). Mouse models of breast cancer metastasis. *Breast cancer research : BCR*, *8*(4), 212. https://doi.org/10.1186/bcr1530

Fischer, K., Durrans, A., Lee, S. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527,** 472–476 (2015). https://doi.org/10.1038/nature15748

Ye, X., Tam, W. L., Shibue, T., Kaygusuz, Y., Reinhardt, F., Ng Eaton, E., & Weinberg, R. A. (2015). Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature*, *525*(7568), 256–260. https://doi.org/10.1038/nature14897

Pfefferle, A. D., Herschkowitz, J. I., Usary, J., Harrell, J. C., Spike, B. T., Adams, J. R., Torres-Arzayus, M. I., Brown, M., Egan, S. E., Wahl, G. M., Rosen, J. M., & Perou, C. M. (2013). Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome biology*, *14*(11), R125. https://doi.org/10.1186/gb-2013-14-11-r125

Maro µlakou, I. G., Anver, M., Garrett, L., & Green, J. E. (1994). Prostate and mammary adenocarcinoma in transgenic mice carrying a rat C3(1) simian virus 40 large tumor antigen fusion gene. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(23), 11236–11240. https://doi.org/10.1073/pnas.91.23.11236

Aprelikova, O., Tomlinson, C. C., Hoenerhoff, M., Hixon, J. A., Durum, S. K., Qiu, T. H., He, S., Burkett, S., Liu, Z. Y., Swanson, S. M., & Green, J. E. (2016). Development and Preclinical Application of an Immunocompetent Transplant Model of Basal Breast Cancer with Lung, Liver and Brain Metastases. *PloS one*, *11*(5), e0155262. https://doi.org/10.1371/journal.pone.0155262

Li, Y., Lv, Z., Zhang, S., Wang, Z., He, L., Tang, M., Pu, W., Zhao, H., Zhang, Z., Shi, Q., Cai, D., Wu, M., Hu, G., Lui, K. O., Feng, J., Nieto, M. A., & Zhou, B. (2020). Genetic Fate Mapping of Transient Cell Fate Reveals N-Cadherin Activity and Function in Tumor Metastasis. *Developmental cell*, *54*(5), 593–607.e5. https://doi.org/10.1016/j.devcel.2020.06.021

Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L., & Luo, L. (2007). A global double-fluorescent Cre reporter mouse. *Genesis (New York, N.Y. : 2000)*, *45*(9), 593–605. https://doi.org/10.1002/dvg.20335

Cree I.A. (2011) Principles of Cancer Cell Culture. In: Cree I. (eds) Cancer Cell Culture. Methods in Molecular Biology (Methods and Protocols), vol 731. Humana Press. https://doi-org.proxy1.library.jhu.edu/10.1007/978-1-61779-080-5_2

Rosenbluth, J.M., Schackmann, R.C.J., Gray, G.K. *et al.* Organoid cultures from normal and cancer-prone human breast tissues preserve complex epithelial lineages. *Nat Commun* **11,** 1711 (2020). https://doi.org/10.1038/s41467-020-15548-7

De Me µlenaere V, Descamps B, De Wever O, Vanhove C, Deblaere K (2020) In vivo selection of the MDA-MB-231br/eGFP cancer cell line to obtain a clinically relevant rat model for triple negative breast cancer brain metastasis. PLOS ONE 15(12): e0243156. https://doi.org/10.1371/journal.pone.0243156

Schmidt, M., Scholz, C. J., Polednik, C., & Roller, J. (2016). Spheroid-based 3-dimensional culture models: Gene expression and functionality in head and neck cancer. *Oncology reports*, *35*(4), 2431–2440. https://doi.org/10.3892/or.2016.4581

Kadletz, L., Heiduschka, G., Domayer, J., Schmid, R., Enzenhofer, E., & Thurnher, D. (2015). Evaluation of spheroid head and neck squamous cell carcinoma cell models in comparison to monolayer cultures. *Oncology letters*, *10*(3), 1281–1286. https://doi.org/10.3892/ol.2015.3487

Melissaridou, S., Wiechec, E., Magan, M., Jain, M. V., Chung, M. K., Farnebo, L., & Roberg, K. (2019). The effect of 2D and 3D cell c utures on treatment response, EMT profile and stem cell features in head and neck cancer. *Cancer cell international*, *19*, 16. https://doi.org/10.1186/s12935-019-0733-1

Lu, W., & Kang, Y. (2019). Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis. *Developmental cell*, *49*(3), 361–374. https://doi.org/10.1016/j.devcel.2019.04.010

Padmanaban, V., Grasset, E.M., Neumann, N.M. *et al.* Organotypic culture assays for murine and human primary and metastatic-site tumors. *Nat Protoc* **15,** 2413–2442 (2020). https://doi.org/10.1038/s41596-020-0335-3

Masters, J. R., & Stacey, G. N. (2007). Changing medium and passaging cell lines. *Nature protocols*, *2*(9), 2276–2284. https://doi.org/10.1038/nprot.2007.319

Timmins, N. E., & Nielsen, L. K. (2007). Generation of multicellular tumor spheroids by the hanging-drop method. *Methods in molecular medicine*, *140*, 141–151. https://doi.org/10.1007/978-1-59745-443-8_8

Title:

Metastasis in triple negative breast cancer requires hybrid epithelial to mesenchymal transition

Authors:
Eloise M. Grasset[1*], Matthew Dunworth[1], Gaurav Sharma[2], Melissa Gentz[1], Sydney Bracht[1], Meagan Haynes[1], Elana J. Fertig[2,3,4], Andrew J. Ewald[1,2,3*]

[1] Department of Cell Biology, School of Medicine, Johns Hopkins University, Baltimore, MD 21205, USA
[2] Department of Biomedical Engineering, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[3]Department of Oncology, Sidney Kimmel Comprehensive Cancer Center, Johns Hopkins University School of Medicine, Baltimore, MD 21205, USA
[4] Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD 21205, USA

Correspondence:
Egrasse1@jhmi.edu
Andrew.ewald@jhmi.edu

Lead Contact: Andrew J. Ewald

Summary:

The role of the epithelial to mesenchymal transition (EMT) in metastasis remains controversial due to conflicting evidence from spontaneous *in vivo* models. We sought to identify clinical contexts in which EMT was essential, focusing on triple-negative breast cancer (TNBC) based on its known expression of mesenchymal markers, with luminal breast cancer as a comparison group. We demonstrate across murine and human models that TNBC tumors frequently exhibit a hybrid EMT, while luminal tumors typically retain epithelial differentiation. Using single-cell RNA-seq we identify a continuous transition in TNBC cells from epithelial to mesenchymal states during invasion and back towards an epithelial state during metastatic outgrowth. We next identified transcriptional regulators of these transitions, showing genetically that *Zeb1*, *Zeb2*, and *Foxc2* promote mesenchymal gene expression without regulating epithelial gene expression. In contrast, *Grhl2* suppresses mesenchymal gene expression while promoting epithelial identity. We then demonstrated genetically that the EMT marker vimentin promotes invasion and represses metastatic outgrowth in culture and *in vivo*. Finally, hybrid EMT cell clusters seed distant organs more efficiently, then metastasize by dissociate to single cells, proliferate and migrate throughout the lungs, before eventually fusing to form large multiclonal metastases. Taken together, our results provide strong evidence for an essential role for a hybrid EMT in physiologically realistic models of TNBC metastasis, dissect the molecular mechanisms regulating hybrid EMT, and report a novel cellular strategy for breast cancer metastasis. We speculate that this novel strategy explains both the chemotherapy responsiveness and rapid metastatic recurrence that is observed in TNBC patients.

Keywords: C3(1)-Tag, triple negative breast cancer, TNBC, basal breast cancer, EMP, EMT, MET, metastasis, vimentin

Introduction:

The epithelial-mesenchymal transition (EMT) is the dominant conceptual model to emerge from *in vitro* studies of cancer cell migration and invasion but its role during breast cancer invasion and metastasis *in vivo* remains controversial[1,2]. EMT is a developmental process whereby epithelial cells depolarize, lose their cell-cell contacts and acquire molecular features and cell behaviors normally associated with mesenchymal cells, such as fibroblasts[3,4]. The EMT model postulates that cancer cells acquire mesenchymal features to invade, access the circulation, and survive in transit to ultimately seed and grow in distant sites. The EMT model has split over whether this mesenchymal state can be maintained in distant sites or whether the reverse process of mesenchymal-epithelial transition (MET) is required to form clinically significant metastases[5-7]. The main challenges to the EMT model are the frequent retention of epithelial features in the cancer cells of murine and human tumors and the relatively limited *in vivo* evidence supporting an essential role for EMT in spontaneous models of metastasis.

The main experimental platform for testing the role of EMT during spontaneous metastasis *in vivo* has been genetically engineered mouse models (GEMMs). In breast cancer, the most frequently studied GEMMs utilize the mouse mammary tumor virus (MMTV) LTR to drive expression of the polyoma middle T oncogene (MMTV-PyMT) or ERBB2 (MMTV-Neu), which respectively model metastatic luminal breast cancer and HER2-amplified breast cancer. Analyses in both models suggest that only a small percentage (<10%) of cancer cells express EMT markers in these GEMM tumors[8-11]. However, the role of EMT in metastasis in these models is still debated. Lineage tracing strategies in the MMTV-PyMT model demonstrated that cancer cells acquire the mesenchymal marker N-cadherin but not fibroblast specific protein 1 or vimentin to metastasize[9,11,12]. Using a knock-down strategy, it was shown that the EMT transcription factor (EMT-TF) Snail but not Slug is required for metastasis[10]. On the other hand, we demonstrated that the epithelial markers keratin-14 and E-cadherin are required for breast cancer metastasis in the MMTV-PyMT model[13-15]. Taken together, these studies motivate further examination of the role of EMT in breast cancer metastasis but leave its requirement and molecular mechanisms incompletely understood.

We sought to understand the clinical context in which EMT is most likely to regulate metastasis, building on the knowledge that distinct breast cancer subtypes present different molecular characteristics[16]. Specifically, multiple groups have noted mesenchymal gene expression in the aggressive triple-negative breast cancer (TNBC) subtype, which is characterized by the lack of estrogen receptor (ER), progesterone receptor (PR), and HER2[17,18]. Notably, the presence of cancer cells expressing both epithelial and mesenchymal markers (hybrid EMT) correlates with poor outcomes in TNBC patients[19-23]. However, it remains unclear whether hybrid EMT cells play an essential role in metastasis. To our knowledge, the role of EMT and MET in *in vivo* spontaneous TNBC models has not been reported. Deeper insight into the molecular mechanisms regulating TNBC metastasis are urgently needed as distant recurrence occurs frequently and rapidly in this subtype and there are few options for targeted therapy.

In this study, we use genetically engineered mouse models (GEMMs), patient-derived xenografts (PDXs), and patient tumor samples to demonstrate that TNBC tumors typically contain abundant

hybrid EMT cancer cells both in culture and *in vivo*. We show that hybrid EMT in TNBC models cells exhibit mesenchymal molecular features, mesenchymal cell behaviors, and increased chemoresistance. Furthermore, these cancer cells transition along a spectrum towards mesenchymal during invasion and towards epithelial during formation of metastatic colonies. Single cell gene expression analysis led us to identify candidate regulators of these transitions and genetic techniques enabled us to elucidate the roles of different transcription factors in regulating distinct aspects of the hybrid EMT phenotype. Using integrated *ex vivo* and *in vivo* analyses, we demonstrate crucial roles for the mesenchymal marker vimentin in different stages of the metastatic cascade. Finally, we reveal that hybrid EMT cells initially seed distant organs as clusters that then dissociate to highly proliferative single cells that disperse throughout the lungs, eventually growing together into large metastases. Our data together reveal an essential role for hybrid EMT in TNBC metastasis across three spontaneous and physiologic models, uncover novel cellular strategies for metastasis, and reveal essential transcription factor regulation of hybrid EMT programs.

Results:

TNBC cells present mesenchymal characteristics related to EMT.

To test the frequency of EMT phenotype cells in models of different breast cancer subtypes, we evaluated the expression of epithelial and mesenchymal genes in the basal phenotype C3(1)-Tag GEMM (Basal GEMM) and in the luminal phenotype MMTV-PyMT GEMM (Luminal GEMM). C3(1)-Tag mice express a transgene containing the 5' flanking region of the C3(1) component of the rat prostate steroid binding protein (PSBP) fused to the simian virus 40 large T-antigen (Tag)[24]. These mice essentially all develop invasive carcinomas, with sporadic metastasis to the lungs[25]. MMTV-PyMT mice express the polyoma virus middle T oncogene under the mouse mammary tumor virus long terminal repeat promoter (MMTV-PyMT)[26] and develop invasive mammary tumors that metastasize efficiently to the lungs[27]. Previous gene expression profiling studies of these models have demonstrated that MMTV-PyMT tumors cluster with the luminal B subtype of human breast cancer, while C3(1)-Tag tumors cluster with the more aggressive basal subgroup, which is part of TNBC subtype[24,28].

Briefly, we isolated fresh organoids from these models and used differential centrifugations to remove the stromal cell population, as previously described[13-15,29]. Organoids were embedded in collagen I to induce invasion (Figure 1A). After 5 days, the organoids were stained with antibodies recognizing E-cadherin (Ecad) and vimentin (Vim), the most commonly used markers for epithelial and mesenchymal phenotypes, respectively[30]. Luminal GEMM organoids had few $Ecad^+$ Vim^+ cells and the leader cells were typically Vim^-. In contrast, Basal GEMM organoids contained a large fraction of $Ecad^+$ Vim^+ cells that led the invasive strands, while the core remained largely $Ecad^+$ Vim^- (Figure 1B). In addition, we observed an enrichment for broader mesenchymal gene expression in the Basal GEMM compared to Luminal GEMM organoids, such as *Twist1*, *Zeb1*, *Zeb2*, *Snail* and a decrease in epithelial gene expression such as *Krt5*, *Krt14*, *Dsp*, *Ocln* and *Cdh1* (Ecad) (Figure 1C). Increased resistance to chemotherapy is frequently observed in cells undergoing EMT[9,31]. To compare the chemosensitivity of cancer cells from these two models, we dissociated tumors into small clusters and embedded them in a Matrigel based colony formation assay (Figure 1D). Escalating doses of doxorubicin, epirubicin, paclitaxel, or daunorubicin revealed that Basal GEMM cancer cells are more chemoresistant than Luminal GEMM cancer cells (Figure 1E). Taken together, these data demonstrate that Luminal GEMM cells conserve epithelial characteristics during metastasis whereas Basal GEMM cancer cells display key molecular markers, cell behaviors, and chemoresistance features of a hybrid EMT.

We then validated these results in two TNBC patient-derived xenografts (PDXs) (Jackson Laboratory TM00096 and TM00089). Similar to the Basal GEMM organoids, TNBC PDX organoids are mainly $Ecad^+$ Vim^- at the core but 93% and 94% of leader cells were $Ecad^+$ Vim^+ in TM00096 and TM00089, respectively (Figure 1F). Next, we analyzed the expression of Ecad and Vim in freshly isolated TNBC patient organoids (Figure S1A). The three TNBC patients analyzed presented hybrid EMT cells that typically localized in the lead position of invasive strands. The presence of hybrid EMT cells at the invasive strand of organoids across these different independent TNBC models suggest that hybrid EMT cells drive the invasive phenotype in TNBC.

Hybrid EMT cells are abundant in TNBC GEMM, PDXs, and patient primary tumors.
To test for the presence of hybrid EMT cells *in vivo*, we performed immunofluorescence and quantified the percentage of Ecad$^+$ Vim$^+$ areas within whole sections from 23 Luminal GEMM and Basal GEMM tumors. Our results showed that both Luminal GEMM and Basal GEMM cancer cells express Ecad and that large areas of Basal GEMM primary tumors stained double positive for Ecad and Vim (Figure 2A,B). In Luminal GEMM tumors, the double positive areas varied between 0% to 24% of tumor area, with a median at 12%. In Basal GEMM tumors, double positive areas vary between 0% to 60% with a median at 28.6%. It is important to note that this area-based measurement will overestimate the representation of double positive cells as an area can be positive due to the presence of, for example, adjacent Ecad$^-$ Vim$^+$ stromal fibroblasts and Ecad$^+$ Vim$^-$ epithelial phenotype cancer cells. To define the percentage of Ecad$^+$ Vim$^+$ cells within these models, we used a Fluorescence-Activated Cell Sorting (FACS) strategy. Briefly, we isolated fresh organoids from our GEMM models, dissociated them into single cells, and fixed and double stained them with Ecad and Vim (Figure 2C). We observed that >90% of Luminal GEMM tumor cells are Ecad$^+$ Vim$^-$, and only 3.3% (± 1) are Ecad$^+$ Vim$^+$, consistent with previous studies[10,12]. We noticed that 0% of the cells analyzed were Ecad$^-$ Vim$^+$ which suggests that our organoid preparation efficiently depleted stromal fibroblasts (Figure 2C and 2D). Interestingly, 92.3% (± 2) of basal cancer cells were double positive for Ecad and Vim (Figure 2C and 2D). We noted two different populations of Ecad$^+$ Vim$^+$ cells with a small population (13.7%) expressing higher level of Vim. We also observed that 5% of the basal cancer cell population is Ecad$^-$ Vim$^+$ which suggests that a subpopulation of C3(1)-Tag cancer cells might be fully mesenchymal.

We next validated these results in TNBC PDX tumors and revealed that 65.67% and 30.40% of the primary tumor areas were double positive for Ecad and Vim in TM00096 and TM00089, respectively (Figure 2E and F). To verify the presence of hybrid EMT cells in a larger human cohort, we used a tissue microarray enriched for TNBC. Immunofluorescence with Ecad and Vim revealed that TNBC tumors present between 14% to 90% of Ecad$^+$ Vim$^+$ areas with a median of 51 % compared to 18 % for the other breast cancer subgroups (Figure 1G and 1H). Details by breast cancer subgroups are presented in Figure S1B. Altogether, these results demonstrated that cancer cells in luminal models typically retain an epithelial differentiation state, while those from TNBC models display a hybrid EMT state *in vivo.*

Single-cell RNA sequencing reveals a temporal progression towards EMT during invasion.
The presence of hybrid EMT cells in TNBC models suggest that cancer cells are able to transition from epithelial states to more mesenchymal states during invasion. To test this theory, we performed a time course analysis of our invasion assay *ex vivo*. Basal GEMM organoids were isolated and plated into collagen I matrix. After fixation at day 0 (12h) and day 5, immunostaining with Ecad and Vim was performed (Figure 3A). We observed that while 22% of cancer cells expressed Vim at day 0, the number of Vim$^+$ cells increased to 56% at day 5 (Figure 3B and 3C), suggesting an EMT. Similar results were obtained in TNBC PDX (Figure S2A).

We sought to identify the molecular alterations associated with this observed shift between Ecad and Vim in the immunostaining analysis, leading us to perform further single cell RNA sequencing of the organoid cells over time. The main advantage of this assay compared to an *in vivo* analysis

is that we are able to capture the transcriptional change in each of these cancer cell states during the specific metastatic processes (e.g. invasion). Although single cell sequencing is limited to sampling tumor cells at discrete time points, computational tools for trajectory inference have been shown to accurately model continuous transitions in cells. Therefore, we can adapt these approaches to identify the molecular changes that correspond to transition of hybrid EMT observed in our organoid models. To enable this analysis, single cell profiling was performed at days 0 (12 h in matrix), 3, and 5. A challenge for time course experiments with single cell RNA sequencing is that the variation in library preparation between days can introduce a technical bias in the data, leading us to profile an additional cell line control (4T1) that will have a consistent transcriptional profile over time for the libraries generated at each day. This additional sample allows us to perform batch correction of all the data, ensuring the accuracy of the molecular changes this assay associates with the transitions in the hybrid EMT state (Figure 3D). Uniform Manifold Approximation and Projection (UMAP) of the uncorrected data showed that the 4T1 cells cluster together for day 0 and day 5 but not for day 3, consistent with this anticipated batch effect. We then corrected for the batch effects between libraries using the Batchelor[32] algorithm, and observed that we retain transcriptional heterogeneity in the organoid culture while successfully removing the technical sources of variation that cause the 4T1 cells to cluster separately between time points (Figure S2B). To further delineate the temporal dynamics of the state transitions in the organoid culture, we performed trajectory analysis with pseudotime[33] of the batch corrected single cell data for the cancer cells in the organoid culture. This pseudotime trajectory analysis provides a continuous measure of the changes between transcriptional profiles in the cells, which we found corresponds to the days at which cells are profiled (Figure 3E). We further identified that the cell state transitions inferred with pseudotime corresponds to a decrease in Ecad expression and increase in Vim expression, thereby indicating that the transcriptional shifts in the single cell data track the modifications of the cancer cells during invasion (Figure 3F). Examples of additional EMT-related gene expression that also vary over pseudotime are presented in Figure S2C.

Vimentin is required for invasion across TNBC models.

To investigate the functional relevance of EMT during invasion we next used shRNA to test whether Vim regulates invasion in Luminal GEMM, Basal GEMM, and TNBC PDX organoids. We tested five different short hairpin RNAs (shRNAs) targeting Vim and selected the two hairpins with the highest knockdown efficiency (Figures S3A-F). As a control, we used a shRNA that has been validated to not target any mammalian genes (NT-shRNA). To evaluate the requirement of Vim in invasion, we cultured NT-shRNA, shVim #1, and shVim #2 infected tumor organoids in 3D collagen I and observed them by differential interference contrast (DIC) microscopy (Figure 3G). Knocking down Vim did not affect invasion in Luminal GEMM organoids, while it eliminated invasion in Basal GEMM and TNBC PDX organoids (Figure 3H-I, movie 1). Taken together, these results in combination with the trajectory analysis of the single cell RNA-sequencing data demonstrate at single cell resolution that there is a temporal progression associated with EMT during invasion and that the acquisition of vimentin expression is required for invasion in TNBC models.

Single-cell RNA sequencing reveals a temporal progression towards MET during metastatic outgrowth.

We previously demonstrated that our *ex vivo* colony formation assay models the survival and proliferative processes involved in metastatic seeding and outgrowth[13,15]. We therefore used this assay to assess the heterogeneity of EMT in basal cancer cells during metastatic outgrowth *ex vivo*. Similar to our previous analysis in collagen, Basal GEMM clusters were freshly isolated and embedded in Matrigel. After fixation at day 0 (12h) and day 5, immunostaining with Ecad and Vim was performed (Figure 4A). In an inverse of the marker dynamics we observed during invasion, 77% of cells were Vim^+ at day 0 and only 10% at day 5, suggesting that this assay induces MET (Figure 4B and 4C). We validated these results in a TNBC PDX (Figure S4A).

To assess the molecular changes occurring during metastatic outgrowth, we next performed single-cell RNA sequencing at days 0 (12 h in matrix), 3, and 5 using this colony formation assay (Figure 4D). We conducted the same analysis that was performed for cells plated in collagen to model invasion, including batch correction with the control 4T1 cells. Contrary to the collagen invasion assay, progression through pseudotime during metastatic outgrowth was associated with a decrease in Vim expression and an increase in Ecad expression (Figure 4E and 4F). Examples of EMT-related gene expression over pseudotime are presented in Figure S4B.

As we observed a down-regulation of Vim during colony formation, we hypothesized that knocking-down its expression would increase colony formation. To test this hypothesis, we cultured NT-shRNA, shVim #1, and shVim #2 infected clusters in 3D Matrigel and assessed their capacity to form colonies (Figure 4G). Vim knock-down did not affect colony formation of Luminal GEMM clusters. However, Vim knock-down increased colony formation of Basal GEMM and TNBC PDX clusters (Figure 4H and 4I). These data reveal, at single cell resolution, that basal cancer cells regain epithelial gene expression during colony formation and that mesenchymal gene expression suppresses colony formation *ex vivo*.

The molecular regulators of hybrid EMT.

We next investigated the correlation of epithelial and mesenchymal genes with Vim and Ecad expression in the scRNA-seq data (Figure 5A). We noticed that the EMT regulators *Zeb1*, *Zeb2*, *Foxc2,* and *Tgfb1* are correlated with vimentin expression during invasion and inversely correlated with Ecad expression during metastatic outgrowth assays. On the contrary, *Ovol1* and *Grhl2* are inversely correlated with vimentin expression during invasion and are instead correlated with Ecad expression during metastatic outgrowth assays. Consequently, we hypothesized that these regulators control hybrid EMT in this basal breast cancer model. To test this hypothesis, we knocked-down the expression of these transcription factors and used a small inhibitor of TGFß activity (Figure 5B and S6).

We showed that *Grhl2* represses vimentin expression and cancer invasion, while promoting Ecad expression (Figure 5C-E). We investigated the impact of *Grhl2* knock-down on EMT-related gene expression and demonstrated that *Grhl2* promotes the expression of multiple epithelial genes, such as *Krt14*, *Ocln*, and *Dsp,* and represses expression of mesenchymal genes, including *Zeb1* (Figure S5A). This result indicates that *Grhl2* regulates the hybrid EMT state by modulating both

epithelial and mesenchymal gene expression. On the contrary, we demonstrated that *Zeb2* and *Foxc2* promote vimentin expression and invasion (Figure 5C-E). We observed that these EMT-TFs did not impact epithelial gene expression but promote expression of multiple mesenchymal genes (Figure 5E and S6B-C). We then explored the role of these EMT regulators in metastatic outgrowth *ex vivo*. We demonstrated that while *Zeb1* and *Ovol1* did not impact invasion (Figure 5D) they repressed metastatic outgrowth (Figure 5F) and promoted vimentin expression without affecting epithelial gene expression (Figure 5G and S5D-E). The discrepancy between EMT-TFs impacting invasion versus metastatic outgrowth demonstrates the importance of the context in EMT regulation.

To investigate the role of TGFß during invasion in this model, we cultured Basal GEMM organoids in the presence of a TGFß receptor I inhibitor. We observed a significant reduction of invasion in the presence of the inhibitor (Figure S6A and S6B) as well as a down-regulation of mesenchymal genes. However, similar to the previous EMT regulators identified, we did not observe an increase in epithelial gene expression (Figure S6C). Moreover, TGFß inhibition did not impact colony formation or chemoresistance (Figure S6D). Altogether, these results highlight a complex transcriptional network regulating EMT in this model. We showed that EMT-TFs regulate mesenchymal gene expression without impacting epithelial gene expression and revealed that *Grhl2* maintains epithelial gene expression during this hybrid EMT.

Hybrid EMT is required for invasion and metastasis formation *in vivo* in the Basal GEMM.
Metastasis is a systemic phenotype and therefore the requirement for hybrid EMT cells in metastasis must be tested *in vivo*. We previously demonstrated that Ecad is required for metastasis formation *in vivo* in both this Luminal GEMM and this Basal GEMM model[13]. Due to the numerous EMT regulators identified in this model, and the *ex vivo* results obtain with Vim knock-down, we decided to first test the requirement for EMT by evaluating the consequences of Vim knock-down in orthotopic transplant. Briefly, Basal GEMM tumor organoids were isolated from fluorescently labeled advanced carcinomas ($ROSA^{mT/mG}$ C3(1)-Tag), transduced with NT shRNA, shVim #1, or shVim #2 lentivirus, and then orthotopically transplanted into the cleared mammary fat pads of nonfluorescent nod scid gamma (NSG) mice (Figure 6A). Immunofluorescence of the whole primary tumors showed a significant reduction of $Ecad^{+}$ Vim^{+} areas with a median of 72% in control compare to 20% and 30% for shVim #1 and shVim #2, respectively (Figure 6B and 6C). We also noticed a change in cancer cell morphology with elongated mesenchymal-like cells in NT-shRNA condition versus epithelial-like cells in Vim knock-down conditions (Figure 6B and S7A). Primary tumor growth tended to increase in Vim knock-down conditions, with some variation across mice, possibly due to varying efficiency of Vim knock-down *in vivo* (Figure S7B-C).

To test the requirement for hybrid EMT in metastasis, we analyzed the lungs of transplant recipient mice. We observed a significant reduction in macrometastasis (>100 cells) number in the Vim knock-down groups (Figure 6D-E, S7D). We next quantified the number of micrometastasis (<100 cells) in the lungs. We also observed a strong reduction of micrometastasis number in Vim knock-down groups (Figure 6F and 6G). These results reveal that

the acquisition a key marker of the hybrid EMT state, vimentin, is required for metastasis formation *in vivo* in this Basal GEMM.

MET is advantageous for metastasis formation *in vivo.*

Our *in vitro* experiments suggest a requirement for Vim in invasion and a complete block of invasion would be expected to prevent metastasis. Hence, our orthotopic metastasis experiments do not necessarily evaluate the role of Vim or MET later in the metastatic cascade. We therefore next utilized a tail vein-based experimental metastasis assay to specifically interrogate the later stages of cell seeding and metastatic outgrowth. Briefly, fluorescently labeled organoids (ROSA$^{mT/mG}$ C3(1)-Tag) were isolated and infected with lentivirus containing NT-shRNA, shVim #1, or shVim #2. After selection, organoids were dissociated into clusters and injected into the tail vein of NSG mice (Figure 6H). Consistent with our *ex vivo* colony formation assay results, we observed a strong increase in the number of macrometastasis in Vim knock-down groups with only 1 macrometastasis in control versus 16 and 14 per mouse in shVim #1 and shVim #2 groups, respectively (Figure 6I and 6J). These results demonstrate that Vim suppresses metastatic outgrowth *in vivo*, supporting an advantageous role for MET at later stage of metastasis in this model.

Hybrid EMT clusters promotes lung metastatic colonization *in vivo*.

To determine the occurrence of MET *in vivo*, we performed a time-course analysis of metastatic outgrowth. Briefly, fluorescently labeled organoids (ROSA$^{mT/mG}$ C3(1)-Tag) were isolated and dissociated into a mixture of single cells and clusters and injected into the tail vein of NSG mice. Lungs were collected and analyzed at 72h, 1 week, 2 weeks, and 4 weeks (Figure 7A). To investigate whether single cells or clusters of cells have an advantage to seed the lungs, we compared their distribution within the cell suspension injected into the mice and their presence at 72h into the lungs (Figure 7B and 7C). We observed that while the majority of cells injected were single cells (~60%), the cells that persist into the lungs of the mice were mainly clusters (~80%), suggesting that clusters have an advantage to seed distant organs, which has previously been reported in other models[15]. To validate this result, we then injected an equal number of cells either as single cells or clusters into tail vein of NSG mice. After 4 weeks, analysis of the lungs confirmed that clusters are more efficient to form metastasis than single cells (Figure S8A-C). Next, we examined the state of the cells and observed that even if the majority of the injected cells were epithelial (E: Ecad^{+}, Vim^{-}), the clusters detected in the lungs of the mice expressed a hybrid EMT phenotype (EM: Ecad+, Vim+), indicating that hybrid EMT clusters are responsible for secondary organ seeding and colonization.

Metastases can retain mesenchymal character.

To determine the occurrence of MET during metastatic outgrowth *in vivo*, we then analyzed the later time points. We expected to see larger cluster of cells overtime, but instead observed an exponential increase of single cells and small clusters of cells (Figure 7B, 7E and 7F). As these mice lack primary tumors, we hypothesized that these single cells came from the proliferation and dissemination of the hybrid EMT clusters that seed into the lungs. To verify this hypothesis, we stained these cells for the proliferation marker Ki67 (Figure 7H and 7I). Approximately 80% of the tumor cells present in the lungs at 72h and 1 week, and 40% of cells at 2 and 4 weeks, were

positive for Ki67, indicating that these cells were actively proliferating. Finally, we analyzed the cell state of the macrometastases at 4 weeks. Similar to the primary tumors they came from, we observed metastatic heterogeneity. 43% of the metastases analyzed were composed of hybrid EMT cells. Another 43% of metastases were composed of mesenchymal (M: Ecad-,Vim+) and hybrid EMT cells (Figure 7B and 7G), suggesting that some cancer cells may continue to transition through more mesenchymal states at the secondary site. The remaining 14% of metastases were epithelial in identity (Ecad+, Vim-). These epithelial metastases could result from a MET or from the seeding of epithelial cells at the distant site. All together, these results demonstrate that TNBC metastasis require a hybrid EMT program through multiple steps of metastasis and that cancer cells can maintain their hybrid EMT features during metastatic outgrowth (Figure 7J).

Discussion:

We sought to understand the role of EMT during metastasis in both luminal and TNBC subtypes of breast cancer. We first demonstrated that spontaneous TNBC tumors from a GEMM, multiple PDXs, and archival patient samples exhibit large populations of hybrid EMT cells *in vivo*. We then used 3D culture of GEMM and PDX tumors to demonstrate that these hybrid EMT cancer cells led invasion, displayed classically mesenchymal cell behaviors, exhibited broadly elevated chemoresistance, and both expressed and required the mesenchymal marker vimentin during invasion in culture and *in vivo*. Conversely, vimentin knockdown enhanced colony formation in 3D culture and in the tail vein experimental metastasis assay. To visualize the transitions of the cells during these two key steps of the metastatic cascade we performed single-cell RNA sequencing over time in our 3D culture assays. We observed cancer cells progressing on a spectrum towards mesenchymal states during invasion and back towards epithelial states during colony formation. This EMT continuum concept finds support in diverse cancer types: including skin cancer[34], lung cancer[35], and head and neck cancer[36].

To identify the molecular regulators of hybrid EMT during TNBC metastasis, we analyzed our single-cell RNA sequencing data for genes whose expression increased during invasion and decreased during colony formation. The classic EMT regulators *Foxc2*, *Zeb1*, *Zeb2*, *Ovol1, Grhl2,* and *Tgfb1* each correlated with E-cadherin and vimentin dynamics, which led us to functionally perturb each pathway. Inhibition of either TGFß activity or knockdown of any of *Foxc2*, *Zeb1*, *Zeb2*, or *Ovol1* promotes mesenchymal gene expression without affecting epithelial gene expression. These results demonstrate that mesenchymal genes expression is upregulated during metastasis by a complex transcriptional network. We next dissected the distinct requirements for different EMT transcription factors in different metastatic processes. For example, knock-down of either *Foxc2* or *Zeb2* inhibited invasion without affecting colony formation whereas knock-down of *Zeb1* or *Ovol1* did not affect invasion but increased colony formation. These results suggest different EMT modules being deployed to accomplish different cellular tasks, consistent with recent work in lung cancer[35]. Importantly, we revealed that *Grhl2* regulates the hybrid EMT states of breast cancer cells by promoting epithelial gene expression and selectively promoting mesenchymal gene expression. This role for *Grhl2* is consistent with work in breast cancer cell lines in which *Grhl2* suppressed EMT, in a reciprocal feedback loop with *Zeb1*[37,38]. Moreover, *Grhl2* expression is frequently lost in claudin-low breast cancers[38] which exhibit a strongly mesenchymal gene expression program, further supporting the concept of *Grhl2* as a positive regulator of epithelial gene expression.

We next tested the requirement for hybrid EMT genetically. We previously demonstrated that E-cadherin suppresses invasion and promotes metastasis in both the Luminal GEMM and the Basal GEMM used here[13]. In the present study, we demonstrated that vimentin promotes invasion and represses metastatic outgrowth, both *ex vivo* and *in vivo*. Therefore, both the epithelial and mesenchymal aspects of the hybrid EMT observed in this Basal GEMM are required for metastasis, a strict demonstration of the requirement for hybrid EMT in this model. In contrast, vimentin knockdown did not affect invasion or colony formation in the Luminal GEMM, revealing subtype specific differences in metastatic requirements that were consistent with the observed marker expression in both fresh and archival patient specimens that we examined. Vimentin has

previously been implicated in cell migration and intracellular signaling[39], however the majority of these studies rely on cell lines cultivated in 2D. It will be important in future studies to gain further insight into the role of vimentin in TNBC metastasis.

We next sought to understand the cellular strategies employed by hybrid EMT cells to form metastases in distant organs. Consistent with our prior work in luminal models, we demonstrated that clusters of cancer cells from the Basal GEMM are more efficient in seeding the lungs than single cells[15,40]. The Basal GEMM clusters that persisted in the lungs exhibit hybrid EMT characteristics, revealing a selective advantage for this phenotype in this model, analogous to the results observed in our colony formation model. Our data is consistent with the concept that intermediate points on the EMT spectrum are highly efficient in metastasis and grounds this concept in spontaneous *in vivo* models[34,41]. However, our data suggest that metastases can form without an evident MET. We demonstrated that 85% of Basal GEMM metastases were composed of hybrid EMT cells (Ecad+, Vim+) or mesenchymal cells (Ecad-, Vim+), implying that hybrid EMT cancer cells can continue to transition towards more mesenchymal states at the secondary site. The remaining 15% of the metastases analyzed were epithelial (Ecad+, Vim-). These epithelial metastases could result from a MET or the seeding of epithelial cells at the distant site. In breast cancer patients, both types of metastases have been observed, including those with high E-cadherin expression[42,43] and those that display the mesenchymal markers vimentin and *fsp1*[44], thereby providing clinical support for our conclusion that there are both epithelial and hybrid EMT paths to metastasis.

Unexpectedly, time-course analysis in the experimental metastasis assay revealed that the hybrid EMT clusters next dissociate to single cells, then migrate throughout the lungs while proliferating at high levels, before eventually fusing into large multiclonal metastases. The breakup of the clusters and dissemination as single cells would be exceptionally challenging to detect by radiologic imaging in a clinical setting. Indeed, despite the large number of cells, we were not able to detect metastases at 2 weeks post-injection using a dissecting scope; they were only evident at high magnification. At 4 weeks post-injection, macrometastases would "suddenly" appear based on the fusion of many small islands of cancer cells. We speculate that the high fraction of proliferating cancer cells at early stages of metastatic outgrowth could explain the chemosensitivity of some TNBC tumors[45] and the scattering and fusion could explain the early lethal metastatic recurrence that is characteristic of this subtype[46-48].

To our knowledge, our study presents the first evidence of the requirement for hybrid EMT in a spontaneous breast cancer model, with consistent regulation of the classic molecular markers, cell behaviors, and chemoresistance properties predicted for this cell state. The GEMM and PDX models characterized here that displayed strong evidence of hybrid EMT are all commercially available and should provide a powerful resource for identifying the targetable effectors of EMT driven metastasis. Finally, our data argue that triple negative breast cancer is likely to be an important context in which to evaluate anti-EMT therapeutic strategies.

Acknowledgments
We thank R. Weinberg for insightful discussion of the modern EMT model and the TEMTIA community for critical feedback on our research. We thank all members of the Ewald Laboratory for critical discussions and specifically A. Fraser for his help with the single-cell RNA sequencing experiment and M. Haynes for technical assistance with mouse colony maintenance. We thank H. Hao and L.D. Orzolek from the Microarray and Deep Sequencing Core Facility for technical assistance and J. Bader for statistical advice. We thank C.S. McGinnis and Z.J. Gartner for the MULTI-seq single-cell RNA sequencing protocol and their help with the analysis. We thank J. Zhu for his advice on flow sorting. We thank N. Aceto, M. Saini and I. Krol for their help with preliminary experiments related to this study. We thank G. Stein-O'Brien for her critical comments on the manuscript. A.J.E. received support for this project through grants from: the Breast Cancer Research Foundation (BCRF-19-048), Twisted Pink Foundation, Hope Scarves, the Jayne Koskinas Ted Giovanis Foundation, and the National Institutes of Health / National Cancer Institute (U01CA217846, U54CA2101732, 3P30CA006973).

Author Contributions
E.M.G and A.J.E conceptualized the project and designed most experiments with the help of E.J.F. for the single-cell RNA sequencing. E.M.G and M.D. performed the experiments. G.S. and E.J.F. performed single-cell RNA sequencing analysis. E.M.G., G.S., E.J.F. and A.J.E. contributed to the interpretation of the sequencing data. S.B. and M.G. helped with immunofluorescence staining and sectioning. M.H. helped with tumor measurements. E.M.G. and A.J.E. wrote the manuscript with useful input from all authors.

Declaration of Interests
AJE has unlicensed patents related to the use of K14 as a biomarker in breast cancer and to the use of antibodies as cancer treatments. AJE's spouse is an employee of Immunocore. EJF is a consultant for Champions Oncology.

Methods

Mice

All mouse husbandry, procedures performed, and assay endpoints were in accordance with protocols approved by the Johns Hopkins Medical Institute Animal Care and Use Committee (ACUC). All mice were female maintained on the FVB/n background in a specific pathogen-free facility. FVB/N-Tg(MMTV-PyVT)634Mul/J (MMTV-PyMT)[26], FVB-Tg(C3-1-TAg)cJeg/JegJ (C3(1)-Tag)[49] and B6.129(Cg)-Gt(ROSA)26Sortm4(ACTB-tdTomato,-EGFP)Luo/J (mT/mG)[50] mice were obtained from the Jackson Laboratories. The triple negative PDX model of invasive ductal carcinomas used in this study were obtained from Jackson Laboratories (TNM00089, TM00096) and were passaged subcutaneously in *Nod scid gamma* mice. In all experiments involving tumor mice, mice were euthanized when the maximum tumor diameter was 20 mm or less, consistent with our ACUC-approved protocol limit. For *in vivo* experiments, littermates were randomized into different treatment groups. The number of mice used has been estimated based on our *ex vivo* assays to have a p-value inferior to 0.05 with a power of 80% to detect a 50% effect on metastases. Lung metastasis counts for transplant experiments were blinded for treatment group.

Cell line

4T1 (ATCC® CRL-2539™) cells were obtained from ATCC. Experiments were performed between passage 2 to 5 in RMPI-1640, 10% vol/vol FBS, 1% vol/vol Penicilin-Streptomycin. Cells were always maintained in 37°C incubators, 5% C02.

Tissue microarray

BR1202 breast cancer tissue array was obtained from Biomax, Inc.

Human breast cancer patient samples.

Tissue samples were provided by the NCI Cooperative Human Tissue Network (CHTN).

Isolation of primary mammary tumor organoids

We isolated fresh organoids before every experiment, as previously described[14,29,51] 15-20 mm mammary tumors from TM00089, TM0096, 10-to-12-week-old MMTV-PyMT mice, or 25-to-30-week-old C3(1)-tag mice were collected. All tumors were mechanically minced with a scalpel and enzymatically digested on a shaker at 180 r.p.m. at 37°C during 60 min in the presence of digesting solution. Digesting solutions were:

- Murine tumors: 2 mg ml^{-1} collagenase (Sigma, C2139), 2 mg ml^{-1} trypsin (Gibco, 27250-018), 5% FBS (Sigma, F0926), 5 µg ml^{-1} insulin (Sigma, I9278), and 50 µg ml^{-1} gentamycin (Gibco, 15750) in 30 ml DMEM-F12 (Gibco, 10565-018).
- PDXs: 4 mg ml^{-1} collagenase (Sigma, C2139), 5% FBS (Sigma, F0926), 1% penicillin–streptomycin (Sigma, P4333), and HEPES (Gibco, 15630-080) in 20 ml RPMI-1640 (Gibco, 11875-085).
- Patient samples: 2 mg ml^{-1} collagenase (Sigma, C2139), 5% FBS (Sigma, F0926), 5 µg ml^{-1} insulin (Sigma, I9278), 1% (vol/vol) GlutaMAX (Life Technologies, 35050-061) and 1% (vol/vol) Pen-strep (Sigma, P4333) in 20 ml DMEM (Sigma, D6546).

Following the digestion, tumors were centrifuged at 400g for 10 min. After supernatant removal, 2 U μl^{-1} DNase (Sigma, D4263) were added in DMEM-F12 and incubated for 5 min at room temperature. Next, tumors were centrifuged at 400g for 10 min. This was followed by three differential centrifugations (400g, 3 s) to separate the epithelial cell clusters from stromal cells. Small and large organoids (100–200 cells versus 250-500 cells, respectively) were then separated using differential centrifugation at a lower speed (150g, 3 s).

Murine organoids were cultures in DMEM-F12, 1% (vol/vol) insulin-transferrin-selenium-ethanolamine (Life Technologies, 51500-056), 1% (vol/vol) Pen-strep (Sigma, P4333), and 2.4 nM FGF2 (Sigma, F0291). PDX tumor organoids were cultured in human organoid growth medium with 2% FBS (DMEM (Sigma, D6546), 2 mM GlutaMAX (ThermoFisher, 35050-061), 1% penicillin-streptomycin (Sigma, P4333), HEPES (Gibco, 15630-080), 0.1% BSA (Sigma, A9576), 10 ng ml^{-1} cholera toxin (Sigma, C8052), 0.5 µg ml^{-1} hydrocortisone (Sigma, H4001), 5 µg ml^{-1} insulin (Sigma, I9278) and 5 µg ml^{-1} EGF (Sigma, E9644)). Patients organoids were culture in human organoid growth medium without FBS.

Fluorescence-Activated Cell Sorting

Isolated organoids were washed using PBS without Ca^{2+} or Mg^{2+} and trypsinized into single cells by incubating with 10X TrypLE (Thermo Fisher, A1217701) for ~10–15 min at 37 °C. After centrifugation at 400g for 10 min, 2 U μl^{-1} DNase (Sigma, D4263) were added in DMEM-F12 and incubated for 5 min at room temperature. Cells were resuspended in organoid medium and filtered through a 40-µm filter to remove the undigested organoids. The cells were next fixed with 80% ice-cold methanol (5 min) in PBS and then permeabilized with 0.1% PBS-Triton X-100 for 15 min. The cells were then incubated in PBS/10% FBS to block non-specific protein-protein interactions followed by the antibodies for 30 min at 22°C. The cells were then washed two times in PBS and analyzed using the Attune NxT Flow Cytometer. Negative population were determined using IgG control.

Antibodies used in this study were rabbit Vimentin-Alexa Fluor 555 (Abcam, ab203428, 1/100), rat E-cadherin-Alexa Fluor 488 (Thermo Fisher, 53-3249-82, 1/100), rat IgG-Alexa Fluor 488 (Thermo Fisher, 53-4301-80, 1/100) and rabbit IgG-Alexa Fluor 555 (Abcam, ab208569, 1/100).

Invasion assay

Neutralized, fibrillar rat-tail collagen I (Corning, 354236) was prepared as previously described[29,51]. Isolated large organoids were embedded at a density of 1.5 organoids per µl into the collagen solution and 100 µl of suspension per well was plated into a 24-well glass bottomed plates on a 37°C heating block. After 1-hour incubation at 37°C to allow polymerization, 1 ml of medium was added to the wells. Tumor organoids invade into collagen I over a period of 3 to 5 days before fixation using 4% paraformaldehyde (PFA; Electron Microscopy Sciences, 15714S). Vehicle (DMSO) and TGFß RI kinase inhibitor II (Millipore-Sigma, 616452, 1 µM) were added directly into the media at D0.

Invasion of organoids was quantified using ImageJ to outline the organoid non-invasive area (core of the organoid) and the total organoid area (invasive strand and core). Invasion index was calculated using the formula: invasion index = 1-(non-invading area/total area).

Colony formation assay

Organoids were washed using PBS without Ca^{2+} or Mg^{2+} and trypsinized into small cell clusters (2-10 cells) by incubating with TrypLE (Thermo Fisher, 12604013) for ~10–15 min at 37 °C. After centrifugation at 400g for 10 min, 2 U μl^{-1} DNase (Sigma, D4263) were added in DMEM-F12 and incubated for 5 min at room temperature. Cells were resuspended in organoid medium and filtered through a 40-µm filter to remove the undigested organoids. 125 MMTV-PyMT clusters μl^{-1} or 300 C3(1)-Tag clusters μl^{-1} were then resuspended in Matrigel and 15 µl was plated into 96 glass bottomed plates on a 37°C heating block. After 1-hour incubation at 37°C, organoid medium was added and the plate was incubated at 37°C, 5% C02 for 96-hour. Next, Matrigel gels were fixed in 1 % PFA with 0.25 % glutaradehyde and the entire gel was imaged in 3D using a Molecular Devices ImageXpress Micro High Content Analysis System. Gels were imaged at 4X and resulting colonies were identified, segmented, and quantified using ImageJ. Percent colony formation is calculated as the number of colonies in a given experimental condition normalized to the average of the control wells. Small molecule used in this study were a gift from the NCI DCTD DTP (http://dtp.cancer.gov) and the final concentration of DMSO was kept constant at 0.5 % in each well. Cells were treated after 12 hours.

Lentiviral knockdown of vimentin in tumor organoids

Approximately 1,000 small tumor organoids were resuspended in 50 µl of organoid media, or human organoid growth medium without FBS for PDX, and plated to a single well of a non-adherent 96-well microplate. Virus particles at a multiplicity of infection = 15 along with 3 µl of ViroMag R/L (OZ Bioscience, RL40200) nanoparticles were mixed up to a total volume of 50 µl into the corresponding media (organoid media or human growth medium without FBS) and incubated at room temperature for 30 min. Then, the virus-containing solution was added on top of the organoids and the plate was placed on a magnetic plate (OZ Biosciences, MF10000) for 90 min at 37 °C. The next day, ~70 µl of medium was replaced with 150 µl of fresh media (organoid media containing FGF2 or human growth media with FBS). After 48 h, ~150 µl of medium was replaced with 5 µg ml^{-1} puromycin-containing (Gibco, A1113803) medium. After selection for three days, organoids were collected, and downstream assays were performed.

Mouse vimentin shRNA (MISSION pLKO.1-puro, Sigma-Aldrich):

- Non-targeting shRNA (SHC216V):
  CCGGGCGCGATAGCGCTAATAATTTCTCGAGAAATTATTAGCGCTATCGCGCTTTTT
- A (TRCN0000089829):
  CCGGCGACGCCATCAACACTGAGTTCTCGAGAACTCAGTGTTGATGGCGTCGTTTTTG
- B (TRCN0000089832):
  CCGGGTGGAATCCTTGCAGGAAGAACTCGAGTTCTTCCTGCAAGGATTCCACTTTTTG
  #1 (TRCN0000317751):
  CCGGGTGGAATCCTTGCAGGAAGAACTCGAGTTCTTCCTGCAAGGATTCCACTTTTTG
- #2 (TRCN0000317675):
  CCGGGCATCACGATGACCTTGAATACTCGAGTATTCAAGGTCATCGTGATGCTTTTTG
- E (TRCN0000317676):
  CCGGGCGCAAGATAGATTTGGAATACTCGAGTATTCCAAATCTATCTTGCGCTTTTTG

Mouse EMT transcription factors shRNA (pZIP-mCMV-ZsGreen-Puro - V28, pooled of three sequences, Transomic):

Gene Symbol	***Sequence***
Foxc2	TGCTGTTGACAGTGAGCGACAGCCCCCTACTCTTACGACATAGTGAAGCCACAGATGTAT GTCGTAAGAGTAGGGGGCTGCTGCCTACTGCCTCGGA
	TGCTGTTGACAGTGAGCGAGGCCACACGTTTGCAACCCAATAGTGAAGCCACAGATGTAT TGGGTTGCAAACGTGTGGCCGTGCCTACTGCCTCGGA
	TGCTGTTGACAGTGAGCGCTGCAACCCAACAGCAAACTTATAGTGAAGCCACAGATGTAT AAGTTTGCTGTTGGGTTGCAATGCCTACTGCCTCGGA
Grhl2	TGCTGTTGACAGTGAGCGAAGACAAGAGACTTCTGTCTGATAGTGAAGCCACAGATGTAT CAGACAGAAGTCTCTTGTCTCTGCCTACTGCCTCGGA
	TGCTGTTGACAGTGAGCGATCCCGGTGAACCTCTGTCTAATAGTGAAGCCACAGATGTATT AGACAGAGGTTCACCGGGACTGCCTACTGCCTCGGA
	TGCTGTTGACAGTGAGCGATGGAGAAAATCACAAAGCTTATAGTGAAGCCACAGATGTAT AAGCTTTGTGATTTTCTCCACTGCCTACTGCCTCGGA
Ovol1	TGCTGTTGACAGTGAGCGAAGGCACCTCTGCACTTACTGATAGTGAAGCCACAGATGTAT CAGTAAGTGCAGAGGTGCCTCTGCCTACTGCCTCGGA
	TGCTGTTGACAGTGAGCGACGTCTCCACGTGCAAGAGGAATAGTGAAGCCACAGATGTAT TCCTCTTGCACGTGGAGACGCTGCCTACTGCCTCGGA
	TGCTGTTGACAGTGAGCGAGCATGCTGAATCGACACATGATAGTGAAGCCACAGATGTAT CATGTGTCGATTCAGCATGCGTGCCTACTGCCTCGGA
Zeb1	TGCTGTTGACAGTGAGCGCCCAGAACAGTGTTTATTCTGATAGTGAAGCCACAGATGTAT CAGAATAAACACTGTTCTGGTTGCCTACTGCCTCGGA
	TGCTGTTGACAGTGAGCGCCCAGAAGACCTTACTGTCAAATAGTGAAGCCACAGATGTAT TTGACAGTAAGGTCTTCTGGTTGCCTACTGCCTCGGA
	TGCTGTTGACAGTGAGCGCCGGACGAGAGAGAAAGTCTGATAGTGAAGCCACAGATGTA TCAGACTTTCTCTCTCGTCCGATGCCTACTGCCTCGGA
Zeb2	TGCTGTTGACAGTGAGCGACCTAATTCTGTTTCTTCTTCATAGTGAAGCCACAGATGTATG AAGAAGAAACAGAATTAGGGTGCCTACTGCCTCGGA
	TGCTGTTGACAGTGAGCGAGATACGGATCCCGAAACGATATAGTGAAGCCACAGATGTAT ATCGTTTCGGGATCCGTATCCTGCCTACTGCCTCGGA
	TGCTGTTGACAGTGAGCGAACGAGAAGAATGAAGAGAACATAGTGAAGCCACAGATGTA TGTTCTCTTCATTCTTCTCGTGTGCCTACTGCCTCGGA

Human vimentin shRNA (MISSION pLKO.1-puro, Sigma-Aldrich):

- Non-targeting (SHC202V) :
  CCGGCAACAAGATGAAGAGCACCAACTCGAGTTGGTGCTCTTCATCTTGTTGTTTTT
- A (TRCN0000278460):
  CCGGCGCCATCAACACCGAGTTCAACTCGAGTTGAACTCGGTGTTGATGGCGTTTTTG
- B (TRCN0000297469):
  CCGGCTCTGGTTGATACCCACTCAACTCGAGTTGAGTGGGTATCAACCAGAGTTTTTG
- C (TRCN0000297191):
  CCGGGCTAACTACCAAGACACTATTCTCGAGAATAGTGTCTTGGTAGTTAGCTTTTTG
- #2 (TRCN0000297192):
  CCGGGCAGGATGAGATTCAGAATATCTCGAGATATTCTGAATCTCATCCTGCTTTTTG
- #1 (TRCN0000278461):
  CCGGGACAGGTTATCAACGAAACTTCTCGAGAAGTTTCGTTGATAACCTGTCTTTTTG

Protein isolation and western blot

Lysis buffer was made fresh, using 142.8 µl of 7× protease inhibitor stock (two CompleteMini tablets (Sigma, 11836153001), 2 PhosphoSTOP tablets (Sigma, 4906837001), and 2.856 ml dH2O), 100 µl RIPA buffer (Millipore, 20-188), 50 µl glycerol (Sigma, G5516), 10 µl 10% SDS (Sigma, L3771), and 697.2 µl of dH2O. ~200 organoids were washed with ice-cold PBS and centrifuged for 5 min at 400g at 4°C before being resuspended in 30 µl of ice-cold lysis buffer, vortexed, and left on ice for 45 min. The samples were then centrifuged for 10 min at 18,000g at 4°C. The supernatants containing the proteins were then transferred into new tubes and the amount of protein was quantified using a BCA assay kit (Thermo Fisher Scientific, 23225) according to the manufacturer instructions. Equal protein amounts were mixed with Laemmli sample buffer (BioRad, 1610747) and β-mercaptoethanol and boiled 5 min at 95°C. Then, the protein suspensions were loaded in 4–15% Mini-PROTEAN pre-cast gels (BioRad, 456-1084). SDS–PAGE was performed at 100 V for 1 h–1.5 h, or until the dye front had just run off the gel. Gels were transferred onto PVDF membranes (Millipore, IPFL07820) at 100 V for 1.5 h at 4 °C. Membranes were then blocked with Odyssey blocking buffer (Licor, 92750000) for 1 h at room temperature. Primary antibodies were diluted into a 1:1 mixture of blocking buffer and TBS-T (TBS + 0.2% Tween (Millipore, P7949)). Primary antibodies used in the study included rabbit Vimentin (Abcam, ab92547, 1:5000), rat E-cadherin (mouse samples, Life Technologies, 13-1900, 1:1000), mouse E-cadherin (human samples, Abcam, ab1416, 1:50) and mouse β-actin (Sigma, A5441, 1:1000). After over-night incubation at 4°C with primary antibodies, membranes were washed three times, 5 min each with 0.1% TBS-T. Secondary antibodies diluted at 1:10,000 were then added for 1 h at room temperature. Secondary antibodies used included anti-rat IRDye 800CW (Licor, 925-32219), anti-rat IRDye 680RD (Licor, 926-68076), anti-mouse IRDye 680RD (Licor, 925-68070), anti-mouse IRDye 800CW (Licor, 926-32212), anti-rabbit IRDye 800CW (Licor, 925-32211) and anti-rabbit IRDye 680RD (Licor, 926-68073). Then, membranes were washed three times, 5 min each, in 0.1% TBS-T. Finally, membranes were scanned using the Odyssey CLx imaging system and band intensities were quantified using ImageJ.

Immunofluorescence

For organoids in gel, after fixation (see corresponding assay for details), gels were permeabilized using 0.5% TritonX-100 (Sigma, X100) for 1 h at room temperature. Samples were then blocked for 3-4 h with 10% FBS, 1% BSA in PBS at room temperature. Primary antibodies diluted in 1% FBS, 1% BSA, 0.2% TritonX in PBS were incubated overnight at 4 °C. After three washes of 10 min each using PBS, secondary antibodies were added diluted in 1% FBS, 1% BSA, 0.2% TritonX in PBS and incubated for 3-4 h at room temperature. Samples were washed three times with PBS and stored at 4 °C until they were ready to be imaged.

Primary tumors collected in this study were fixed in 4% PFA overnight at 4°C. Fixed tumors were then embedded in paraffin and sectioned at 5 µm by the Reference Histology Laboratory at Johns Hopkins Medical Institutions. Sections were deparaffinized and high temperature antigen retrieval was performed using citrate buffer, pH 6.0 (Sigma, C9999). Lungs were fixed in 1% PFA for 4 h, at 4 °C, transferred into 25% sucrose in PBS overnight at 4 °C, embedded into optimal cutting temperature (OCT) compound, and frozen at −80 °C. Sections (20-µm thick) were cut onto Superfrost Plus Gold Microscope slides (Thermo Fisher Scientific, 15-188-48) at −20 °C using a cryostat. The OCT was removed from these slides by washing in PBS three times, 10 min each. For immunostaining of cell suspension (tail vein assay), 200 µL of the cell suspension in PBS was spread onto slides with a proprietary adhesion coating (Thermo Fisher Scientific, 154852). Cells were left to settle on the slide for 5 min. The slide was then incubated at 37 °C for 60 min. Slide supernatant was discarded and adherent cells were fixed with 4% PFA at room temperature for 10 min. Slides were then washed with PBS. Sections of lungs or primary tumors and fixed cell suspension were next permeabilized for 1 h with 0.5% Triton-X in PBS, then blocked for 2 h with 10% FBS, 1% BSA in PBS at room temperature. Primary antibodies diluted in 1% FBS, 1% BSA, 0.2% TritonX in PBS were incubated overnight at 4 °C. After three washes of 10 min each using PBS, secondary antibodies and DAPI (Invitrogen, D57) were added diluted in 1% FBS, 1% BSA, 0.2% TritonX in PBS and incubated for 2 h at room temperature. Samples were washed three times with PBS and mounted with Fluoromount Aqueous Mounted Medium (Sigma; F4680) and covered with precision coverslips (Marienfeld, 0107222).

Immunofluorescence for all antibodies was performed for at least three independent experiments, each consisting of at least 15–20 organoids, 5–20 metastases or 3–30 tumor sections. Primary antibodies used in the studies include rabbit Vimentin (Abcam, ab92547, 1:250), rat E-cadherin (Abcam, ab11512, 1:100), and rabbit tdTomato (Takara, 632496, 1:200). For human samples specifically, mouse E-cadherin (Thermo Fisher Scientific, 33-4000, 1:50) was used. All secondary antibodies used were Alexa Fluor-conjugates (Invitrogen, 1:200).

Image acquisition

DIC imaging was performed using a LD Plan-Neofluar 20×/0.4 Korr Ph2 objective lens, a Zeiss AxioObserver Z1, and an AxioCam MRM camera. DIC time-lapse movie was collected with 20-min acquisition intervals for ~5 days, maintaining temperature at 37 °C and 5% CO2. Confocal imaging was conducted on either a Zeiss 780 laser-scanning confocal microscope, or on a spinning-disc microscope with an EM-CCD camera (Solamere Technology Group).

Epifluorescence images were obtained using the Axio Scan.Z1 microscope (Zeiss) using Zen Blue 2.1 with the following configurations: Fluor 5x/0.27NA M27 objective for coarse focus mapping, Plan-Apochromat 20x/0.80NA M27 objective for fine focus mapping, Hamamatsu Orca Flash 4.0 (fluorescence) camera for image capture, and Colibri 7 VIS-LED fluorescent light source.
Zen (Zeiss) and ImageJ were used to adjust brightness/contrast levels to maximize image clarity, reduce noise background (human organoids), place scale bars, and export.

Quantification of percentage of Ecad$^+$ Vim$^+$ area

Whole tumor images were exported in TIF by reducing the size of the image to 35% (in order to open the file with ImageJ) using Zen Blue. Images were exported as separated channel. In ImageJ, the images of Ecad and Vim channel were opened, and a ROI was drawn around the tumor on the Ecad channel image. We then calculated the percentage of Ecad$^+$ areas that are Vim$^+$ within the ROI using the *Colocalization Threshold* plugin. The percentages correspond to the thresholded Mander's split colocalisation coefficients multiply by 100, where 100 is 100% of Ecad$^+$ areas that are Vim$^+$. The pixel size of the images used in ImageJ was ~1 µm.

RNA Extraction and Quantitative RT-PCR

RNA was extracted with the TRIzol RNA extraction kit (Thermo Fisher, 15596026) following the manufacturer's protocol. cDNA was synthesized from 1 µg total RNA using the SuperScript IV VILO Master Mix (Thermo Fisher, 11766050) with a gDNA digestion step. Synthesized cDNA was diluted in RNase-free water prior to RT-qPCR. Quantitative PCR was performed using the SsoAdvanced Universal SYBR Green Supermix (BioRad, 1725271) with 5 ng cDNA per reaction and 500 nM primers. Reactions were run in triplicate on a CFX96 Touch Real-Time PCR Detection System (BioRad). Target gene expression values were normalized to RPLP0 expression and fold change was calculated as $2^{-\Delta\Delta Ct}$. RT-qPCR primers are listed in supplemental table 1.

Supplemental Table 1: Primers used for quantitative PCR.

Gene	Forward sequence	Reverse sequence
Twist1	GGACAAGCTGAGCAAGATTCA	CGGAGAAGGCGTAGCTGAG
Foxc2	AACCCAACAGCAAACTTTCCC	GCGTAGCTCGATAGGGCAG
Itga5	CTTCTCCGTGGAGTTTTACCG	GCTGTCAAATTGAATGGTGGTG
Timp1	GCAACTCGGACCTGGTCATAA	CGGCCCGTGATGAGAAACT
Zeb2	ATTGCACATCAGACTTTGAGGAA	ATAATGGCCGTGTCGCTTCG
Zeb1	GCTGGCAAGACAACGTGAAAG	GCCTCAGGATAAATGACGGC
Snai3	GGTCCCCAACTACGGGAAAC	CTGTAGGGGGTCACTGGGATT
Snail	CACACGCTGCCTTGTGTCT	GGTCAGCAAAAGCACGGTT
Mmp9	CTGGACAGCCAGACACTAAAG	CTCGCGGCAAGTCTTCAGAG
Mmp3	ACATGGAGACTTTGTCCCTTTTG	TTGGCTGAGTGGTAGAGTCCC

Camk2n1	TACGGCGACGAGAAGCTGA	AGCCCCGAAGAAGTTGTTGG
Smad2	ATGTCGTCCATCTTGCCATTC	AACCGTCCTGTTTTCTTTAGCTT
Gsc	CAGATGCTGCCCTACATGAAC	TCTGGGTACTTCGTCTCCTGG
Col1a1	GCTCCTCTTAGGGGCCACT	CCACGTCTCACCATTGGGG
Col1a2	GTAACTTCGTGCCTAGCAACA	CCTTTGTCAGAATACTGAGCAGC
Col5a2	TTGGAAACCTTCTCCATGTCAGA	TCCCCAGTGGGTGTTATAGGA
Col3a1	CTGTAACATGGAAACTGGGGAAA	CCATAGCTGAACTGAAAACCACC
Tgfb1	CTCCCGTGGCTTCTAGTGC	GCCTTAGTTTGGACAGGATCTG
Serpine1	TTCAGCCCTTGCTTGCCTC	ACACTTTTACTCCGAAGTCG
Cdh2	AGCGCAGTCTTACCGAAGG	TCGCTGCTTTCATACTGAACTTT
Itgav	CCGTGGACTTCTTCGAGCC	CTGTTGAATCAAACTCAATGGGC
Sparc	GTGGAAATGGGAGAATTTGAGGA	CTCACACACCTTGCCATGTTT
Vim	CGTCCACACGCACCTACAG	GGGGGATGAGGAATAGAGGCT
Igfbp4	AGAAGCCCCTGCGTACATTG	TGTCCCCACGATCTTCATCTT
Fn1	GCTCAGCAAATCGTGCAGC	CTAGGTAGGTCCGTTCCCACT
Tgfb2	CTTCGACGTGACAGACGCT	GCAGGGGCAGTGTAAACTTATT
Slug	TGGTCAAGAAACATTTCAACGCC	GGTGAGGATCTCTGGTTTTGGTA
Ocln	TTGAAAGTCCACCTCCTTACAGA	CCGGATAAAAAGAGTACGCTGG
Zo-1	GCCGCTAAGAGCACAGCAA	TCCCCACTCTGAAAATGAGGA
Cdh1	CTCCAGTCATAGGGAGCTGTC	TCTTCTGAGACCTGGGTACAC
Dsp	GGATTCTTCTAGGGAGACTCAGT	TCCACTCGTATTCCGTCTGGG
Krt14	AGCGGCAAGAGTGAGATTTCT	CCTCCAGGTTATTCTCCAGGG
Krt5	TCTGCCATCACCCCATCTGT	CCTCCGCCAGAACTGTAGGA
RPLP0	AGATTCGGGATATGCTGTTGGC	TCGGGTCCTAGACCAGTGTTC
Grhl2	AAGCCCAGTGCAACAACTCC	TGTCCGGTCCTCTGTAGGTTT
Ovol1	CTACCGGGAACCAGAGGCA	GGCAGGTAAAAAGGTCTCCATT

Orthotopic transplantation of mammary tumor organoids

Tumor organoids were harvested and either directly transplanted or infected and selected following lentiviral transduction. Orthotopic transplantations were performed on 3-4-week-old, female, *Nod scid gamma* mice for tumor organoids. Organoids were resuspended in 50:50 DMEM/F12:Matrigel (Corning, 354230) and kept on ice during the experiment. Host mice were

anaesthetized with 2.5% isofluorane and immobilized on a sterile surface. After ethanol wiping, an incision of ~0.5 cm was made along the mid-line. One of the #4 mammary gland was exposed by an additional incision at a 60° angle towards the lower limbs of the mouse. The region of the mammary gland proximal to the lymph node was excised and 40 organoids were injected into the residual gland using a microsyringe (Hamilton, 702RN-7636-01, 26-gauge). This procedure was repeated for the contralateral mammary gland when injecting organoids. Incisions were closed using 9-mm autoclips (BD Biosciences, 427631), and triple antibiotic ointment was applied. Primary tumors and lungs were collected when the maximum tumor diameter was 20 mm (~5-8 weeks~). In our experience, C3(1)-Tag tumors are heterogenous and not all donor mouse organoids are able to metastasize before the recipient mice need to be euthanized due to the primary tumor burden.

Counting macro- and micrometastases

Macrometastases were counted based on the expression of mT/mG fluorophores under the dissection microscope. Based on their size, we estimated these metastases to have >100 cells. Whole lung images were acquired using an iPhone8.

Micrometastases were counted on whole lung sections stained with rabbit tdTomato (Takara, 632496) and DAPI (1:1,000; Invitrogen, D571) that were scanned using the Axio Scan.Z1 microscope (Zeiss). We noticed a decrease in tdTomato fluorescence intensity for some C3(1)-Tag cancer cells which can be overcome by staining the cells with anti-tdTomato. Any metastasis with more that 100 cells was excluded from micrometastasis counting to limit double scoring in the macro- and micrometastasis categories.

Tail vein assay

Tumor organoids were harvested and either directly dissociated into clusters by incubating with TrypLE (Thermo Fisher, 12604013) for ~10–15 min at 37 °C, or infected and selected following lentiviral transduction. Clusters were resuspended in PBS without Ca^{2+} and Mg^{2+} at a concentration of 10^6 cells ml^{-1} (Figure 7) or 500,000 cells ml^{-1} (Figure 6H). To separate clusters from single cells (Figure S9A-C), we used differential centrifugations at 1500 rpm 25sec. The tail veins of 8-12-week-old female *Nod scid gamma* mice were dilated by exposure to a heat lamp for 1–2 min. 200 µl of cells were injected via the tail vein of the mouse using a 26.5 gauge needle. Lungs were collected after 72 hours, 1 week, 2 weeks, and/or four weeks as indicated in the result section, and examined under the dissection microscope using mT fluorescence to quantify the macrometastases.

Single cell RNA sequencing

Organoids were isolated from 4 different mice and either plated directly into collagen I matrix or further dissociated into clusters and plated into Matrigel. Cells were removed from the matrix at day 0 (12h in matrix), day 3 and day 5. Cells were removed from collagen matrix by a 10 min incubation on a benchtop shaker at 180 r.p.m. (37 °C) with 2 mg ml^{-1} collagenase in PBS. For Matrigel matrix, gels were washed once with ice-cold PBS without Ca^{2+} and Mg^{2+} followed by a 30 min incubation on ice with 10 mM EDTA (Sigma, EDS). After two washes in PBS, organoids were dissociated into single cells by incubating with 10X TryplE (Thermo Fischer, A1217701), 10 min on a benchtop shaker at 180 r.p.m. (37 °C). After centrifugation at 400g for 10 min, 2 U $µl^{-1}$

DNase (Sigma, D4263) were added in DMEM-F12 and incubated for 5 min at room temperature. Cells were resuspended in PBS and filtered through a 40-µm filter to remove the undigested organoids. 4T1 cell line was used as a control for batch effect during library construction for the different time points. 4T1 cells were trypsinized using trypsin-EDTA 0.25% (Thermo Fisher, 25200056). Detached cells were resuspended in culture media and centrifuged at 400g, 5 min at room temperature. Cells were then barcoded by condition with MultiSeq[52] and flow-sorted for live single-cells using Propidium iodide (Thermo Fisher, P3566).

Multi-seq library was prepared following 10X single cell controller protocol and Multi-Seq specific protocol. Library construction from the single cell suspensions were performed using the 10X Genomics Chromium Single Cell 3' Reagent Kit v2, following manufacturer's instructions. Briefly, Multi-Seq barcoded and pooled single cell suspension were counted using Countess II (Thermal Fisher) to determine cell concentration and viability. The cell volume to use was calculated based on the cell concentration and the targeted cell number. The cell suspension was brought to 46.6 µl with molecular biology water and then mixed with 33.4 µl premade single cell master mix containing RT reagent, template switch oligo, reducing reagent, and RT enzyme. 75 µl of the resulting cell/master mix solution were loaded to 10X single cell A Chip row 1. The gel beads were vortexed for 30 seconds and 40 µl were loaded to 10X single cell A Chip row 2. 280 µl partition oil were loaded to 10X single cell A Chip row 3. The unused lanes were filled on the Chip with 50% Glycerol. 10X provided gasket was attached to the chip, then the chip was loaded into 10X single cell controller and the cell partitioning program was run. After the program completes, 100 µl of Gel Bead-In EMulsions (GEM) were removed from the recovery well and slowly dispensed into pre-chilled PCR plate. The plate was sealed and incubated at 53°C for 45 min, 80°C for 5 min and then chill to 4°C. The recovery reagent was then added to the GEM-RT reaction to break up the single cell emulsion and the cDNA was purified using Thermal fisher Dynabeads MyOne Saline magnetic beads and amplified by PCR using 10X provided reagents and Multi-Seq primer. The amplified cDNA was purified using Beckman Coulter SPRI magnetic beads and the endogenous transcript cDNA and Multi-Seq barcode cDNA were separated. The purified cDNA was quantitated using Thermal Fisher Qubit high sensitivity dsDNA assay and ran on Agilent Bioanalyzer High sensitivity DNA chip for quality control. The transcript cDNA library was prepared following standard Illumina library preparation using 10X provided reagents, i.e., fragmentation, end repair and dA tailing, adapter ligation, and PCR amplification. The Multi-Seq barcode cDNA library was prepared by PCR amplification using universal i5 primer and RPI index primer. After library qc and quantitation, the transcript library and the Muti-Seq barcode library were pooled in 95:5 ration and sequenced together on NovaSeq. The transcript library sequencing data was first analyzed using 10X Genomics CellRanger pipeline and 10X Genomics provided reference (version 3.0.0). The analysis was performed on Maryland Advanced Research Computing Center high performance computer.

The fastq files were preprocessed to generate a gene count matrix with Cell Ranger (v3.0.2) using GRCm38 reference (v3.0.0). To remove empty cells, cells were filtered based on a knee plot of the UMI counts vs barcodes. The cell to sample barcode matrix was generated using kite workflow of kb_python (v0.24.4)[53] from the barcode fastq files. The matrix was subsetted using filtered cell barcodes from Cell Ranger and demultiplexed using demultiplex [50] (v1.2.0) to obtain

cell counts for each condition as summarized in Supplemental Table 2. Scanpy[54] (v1.4.4) was used to filter cells that had less than 200 genes expressed and genes that were detected in less than 3 cells. Further filtering for high quality cells was done by selecting cells that had less than 10 percent mitochondrial gene content. Analysis of the pre-processed single cell data was performed using Monocle3[55] (v0.2.1). A two dimensional UMAP embedding[56] was generated with default parameters in Monocole3, correcting the embedding for batch effects between days using the Batchelor method[32] on all the data and assessing the correction based upon clustering of the data for the control 4T1 cells across days. Gene imputation was done on the entire gene count matrix using SAVER (v1.2.2) to visualize gene expression on UMAP. Psuedotime analysis was performed to order cells independently from collagen and Matrigel experiments. Because Batchelor only corrects the UMAP embedding for batch effects, the ComBat[57] function in sva[58] (v3.32.1) was also used to perform batch correction of the log transformed expression values from SAVER for gene-level visualization. The ComBat model used day as batch and pseudotime as the model to preserve cellular state transitions associated with the hybrid EMT state and thereby mitigate confounding from day as a technical variable. Heatmaps are shown for these normalized and imputed gene expression values. Statistical analysis was performed on the unimputed expression values of EMT genes with Monocole3 with a linear model containing the batch corrected, imputed value of Vimentin expression and day as covariates. Genes with q-values below 0.05 for the Vimentin term were called statistically significant.

Supplemental Table 2: Total cell counts for each experimental condition

	Day 0		Day 3		Day 5	
	Collagen	Matrigel	Collagen	Matrigel	Collagen	Matrigel
Mouse 1	483	464	499	72	128	29
Mouse 2	351	440	1381	207	442	79
Mouse 3	288	507	890	134	799	52
Mouse 4	381	731	134	417	0	0

Data availability

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE149299 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149299)

Code availability

All analysis scripts are available from https://github.com/FertigLab/breast_cancer_EMT.

Statistical analysis

Statistical analysis and graphs were performed using Prism 8 (GraphPad). We utilized the D'Agostino–Pearson omnibus test to assess normality of the samples. For normal sample, two-tailed T Test was used to compare two datasets, or ANOVA to compare three or more datasets. When data were not normally distributed, we used the two-tailed Mann–Whitney test to compare two datasets, or the Kruskal–Wallis test (with Dunn correction for multiple comparisons) to compare three or more datasets. Statistical significance was considered starting from $P < 0.05$. Figure legends describe the statistical test used for each experiment.

References:

1 Williams, E. D., Gao, D., Redfern, A. & Thompson, E. W. Controversies around epithelial-mesenchymal plasticity in cancer metastasis. *Nat Rev Cancer* **19**, 716-732, doi:10.1038/s41568-019-0213-x (2019).

2 Bakir, B., Chiarella, A. M., Pitarresi, J. R. & Rustgi, A. K. EMT, MET, Plasticity, and Tumor Metastasis. *Trends in Cell Biology* **30**, 764-776, doi:10.1016/j.tcb.2020.07.003 (2020).

3 Greenburg, G. & Hay, E. D. Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells. *J Cell Biol* **95**, 333-339, doi:10.1083/jcb.95.1.333 (1982).

4 Hay, E. D. An Overview of Epithelio-Mesenchymal Transformation. *Cells Tissues Organs* **154**, 8-20, doi:10.1159/000147748 (1995).

5 Kalluri, R. & Weinberg, R. A. The basics of epithelial-mesenchymal transition. *J Clin Invest* **119**, 1420-1428, doi:10.1172/JCI39104 (2009).

6 Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S. & Yang, J. Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell* **22**, 725-736, doi:10.1016/j.ccr.2012.09.022 (2012).

7 Ocaña, O. H. *et al.* Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* **22**, 709-724, doi:10.1016/j.ccr.2012.10.012 (2012).

8 Beerling, E. *et al.* Plasticity between Epithelial and Mesenchymal States Unlinks EMT from Metastasis-Enhancing Stem Cell Capacity. *Cell Rep* **14**, 2281-2288, doi:10.1016/j.celrep.2016.02.034 (2016).

9 Fischer, K. R. *et al.* Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472-476, doi:10.1038/nature15748 (2015).

10 Ye, X. *et al.* Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256-260, doi:10.1038/nature14897 (2015).

11 Li, Y. *et al.* Genetic Fate Mapping of Transient Cell Fate Reveals N-Cadherin Activity and Function in Tumor Metastasis. *Developmental Cell* **54**, 593-607.e595, doi:10.1016/j.devcel.2020.06.021 (2020).

12 Fischer, K. R., Altorki, N. K., Mittal, V. & Gao, D. Fischer et al. reply. *Nature* **547**, E5-E6, doi:10.1038/nature22817 (2017).

13 Padmanaban, V. *et al.* E-cadherin is required for metastasis in multiple models of breast cancer. *Nature* **573**, 439-444, doi:10.1038/s41586-019-1526-3 (2019).

14 Cheung, K. J., Gabrielson, E., Werb, Z. & Ewald, A. J. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* **155**, 1639-1651, doi:10.1016/j.cell.2013.11.029 (2013).

15 Cheung, K. J. *et al.* Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proc Natl Acad Sci U S A* **113**, E854-863, doi:10.1073/pnas.1508541113 (2016).

16 Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747-752, doi:10.1038/35021093 (2000).

17 Foulkes, W. D., Smith, I. E. & Reis-Filho, J. S. Triple-Negative Breast Cancer. *New England Journal of Medicine* **363**, 1938-1948, doi:10.1056/NEJMra1001389 (2010).

18 Garrido-Castro, A. C., Lin, N. U. & Polyak, K. Insights into Molecular Classifications of Triple-Negative Breast Cancer: Improving Patient Selection for Treatment. *Cancer Discovery* **9**, 176-198, doi:10.1158/2159-8290.Cd-18-1177 (2019).

19 Cakir, A., Gonul, I. I. & Uluoglu, O. A comprehensive morphological study for basal-like breast carcinomas with comparison to nonbasal-like carcinomas. *Diagnostic Pathology* **7**, 145, doi:10.1186/1746-1596-7-145 (2012).

20 Jeong, H., Ryu, Y. J., An, J., Lee, Y. & Kim, A. Epithelial-mesenchymal transition in breast cancer correlates with high histological grade and triple-negative phenotype. *Histopathology* **60**, E87-95, doi:10.1111/j.1365-2559.2012.04195.x (2012).

21 Yamashita, N. *et al.* Epithelial Paradox: Clinical Significance of Coexpression of E-cadherin and Vimentin With Regard to Invasion and Metastasis of Breast Cancer. *Clin Breast Cancer* **18**, e1003-e1009, doi:10.1016/j.clbc.2018.02.002 (2018).

22 Thomas, P. A. *et al.* Association between keratin and vimentin expression, malignant phenotype, and survival in postmenopausal breast cancer patients. *Clin Cancer Res* **5**, 2698-2703 (1999).

23 Sarrió, D. *et al.* Epithelial-Mesenchymal Transition in Breast Cancer Relates to the Basal-like Phenotype. *Cancer Research* **68**, 989, doi:10.1158/0008-5472.CAN-07-2017 (2008).

24 Herschkowitz, J. I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* **8**, R76, doi:10.1186/gb-2007-8-5-r76 (2007).

25 Green, J. E. *et al.* The C3(1)/SV40 T-antigen transgenic mouse model of mammary cancer: ductal epithelial cell targeting with multistage progression to carcinoma. *Oncogene* **19**, 1020-1027, doi:10.1038/sj.onc.1203280 (2000).

26 Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol* **12**, 954-961, doi:10.1128/mcb.12.3.954 (1992).

27 Lin, E. Y. *et al.* Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *Am J Pathol* **163**, 2113-2126, doi:10.1016/s0002-9440(10)63568-7 (2003).

28 Pfefferle, A. D. *et al.* Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biol* **14**, R125, doi:10.1186/gb-2013-14-11-r125 (2013).

29 Padmanaban, V. *et al.* Organotypic culture assays for murine and human primary and metastatic-site tumors. *Nature Protocols*, doi:10.1038/s41596-020-0335-3 (2020).

30 Lu, W. & Kang, Y. Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis. *Dev Cell* **49**, 361-374, doi:10.1016/j.devcel.2019.04.010 (2019).

31 Zheng, X. *et al.* Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525-530, doi:10.1038/nature16064 (2015).

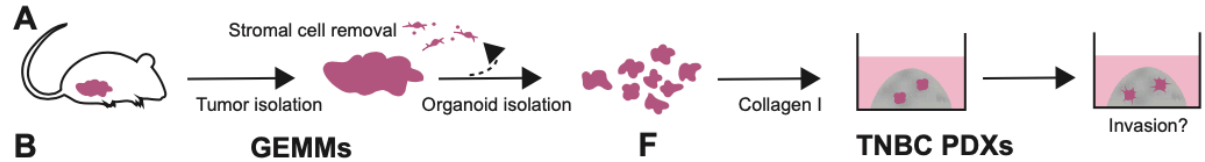
32 Haghverdi, L., Lun, A. T. L., Morgan, M. D. & Marioni, J. C. Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. *Nature Biotechnology* **36**, 421-427, doi:10.1038/nbt.4091 (2018).

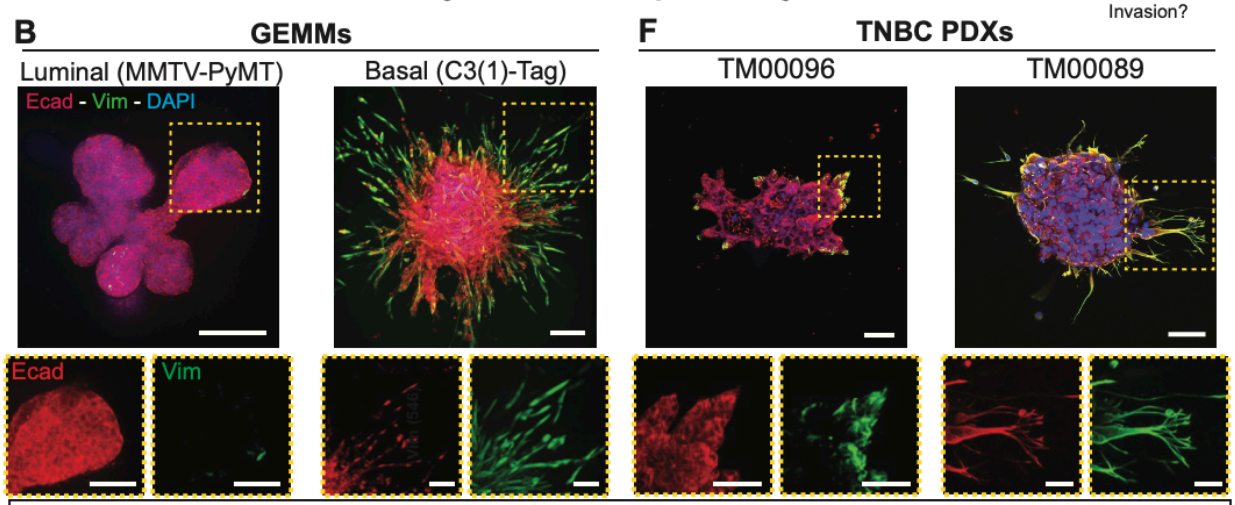
33 Qiu, X. *et al.* Reversed graph embedding resolves complex single-cell trajectories. *Nature methods* **14**, 979-982, doi:10.1038/nmeth.4402 (2017).

34 Pastushenko, I. *et al.* Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463-468, doi:10.1038/s41586-018-0040-3 (2018).
35 Karacosta, L. G. *et al.* Mapping lung cancer epithelial-mesenchymal transition states and trajectories with single-cell resolution. *Nature Communications* **10**, 5587, doi:10.1038/s41467-019-13441-6 (2019).
36 Puram, S. V. *et al.* Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624.e1624, doi:10.1016/j.cell.2017.10.044 (2017).
37 Cieply, B., Farris, J., Denvir, J., Ford, H. L. & Frisch, S. M. Epithelial–Mesenchymal Transition and Tumor Suppression Are Controlled by a Reciprocal Feedback Loop between ZEB1 and Grainyhead-like-2. *Cancer Research* **73**, 6299, doi:10.1158/0008-5472.CAN-12-4082 (2013).
38 Cieply, B. *et al.* Suppression of the Epithelial–Mesenchymal Transition by Grainyhead-like-2. *Cancer Research* **72**, 2440, doi:10.1158/0008-5472.CAN-11-4038 (2012).
39 Satelli, A. & Li, S. Vimentin in cancer and its potential as a molecular target for cancer therapy. *Cell Mol Life Sci* **68**, 3033-3046, doi:10.1007/s00018-011-0735-1 (2011).
40 Aceto, N. *et al.* Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110-1122, doi:10.1016/j.cell.2014.07.013 (2014).
41 Liu, X. *et al.* Epithelial-type systemic breast carcinoma cells with a restricted mesenchymal transition are a major source of metastasis. *Sci Adv* **5**, eaav4275-eaav4275, doi:10.1126/sciadv.aav4275 (2019).
42 Bukholm, I. K., Nesland, J. M. & Børresen-Dale, A. L. Re-expression of E-cadherin, α-catenin and β-catenin, but not of γ-catenin, in metastatic tissue from breast cancer patients. *The Journal of Pathology* **190**, 15-19, doi:https://doi.org/10.1002/(SICI)1096-9896(200001)190:1<15::AID-PATH489>3.0.CO;2-L (2000).
43 Kowalski, P. J., Rubin, M. A. & Kleer, C. G. E-cadherin expression in primary carcinomas of the breast and its distant metastases. *Breast Cancer Research* **5**, R217, doi:10.1186/bcr651 (2003).
44 Chao, Y., Wu, Q., Acquafondata, M., Dhir, R. & Wells, A. Partial Mesenchymal to Epithelial Reverting Transition in Breast and Prostate Cancer Metastases. *Cancer Microenvironment* **5**, 19-28, doi:10.1007/s12307-011-0085-4 (2012).
45 Carey, L. A. *et al.* The triple negative paradox: primary tumor chemosensitivity of breast cancer subtypes. *Clin Cancer Res* **13**, 2329-2334, doi:10.1158/1078-0432.Ccr-06-1109 (2007).
46 Kast, K. *et al.* Impact of breast cancer subtypes and patterns of metastasis on outcome. *Breast Cancer Res Treat* **150**, 621-629, doi:10.1007/s10549-015-3341-3 (2015).
47 Kennecke, H. *et al.* Metastatic behavior of breast cancer subtypes. *J Clin Oncol* **28**, 3271-3277, doi:10.1200/jco.2009.25.9820 (2010).
48 Reis-Filho, J. S. & Tutt, A. N. Triple negative tumours: a critical review. *Histopathology* **52**, 108-118, doi:10.1111/j.1365-2559.2007.02889.x (2008).
49 Maroulakou, I. G., Anver, M., Garrett, L. & Green, J. E. Prostate and mammary adenocarcinoma in transgenic mice carrying a rat C3(1) simian virus 40 large tumor antigen fusion gene. *Proc Natl Acad Sci U S A* **91**, 11236-11240, doi:10.1073/pnas.91.23.11236 (1994).

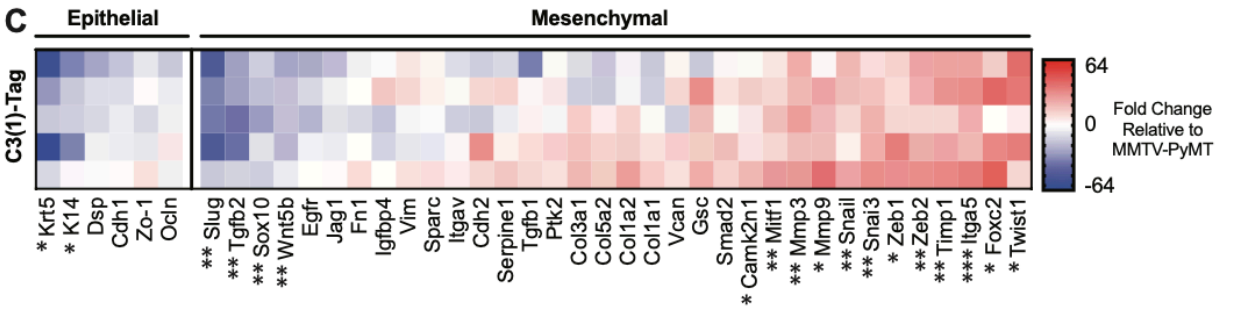
50 Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *genesis* **45**, 593-605, doi:10.1002/dvg.20335 (2007).
51 Nguyen-Ngoc, K. V. *et al.* ECM microenvironment regulates collective migration and local dissemination in normal and malignant mammary epithelium. *Proc Natl Acad Sci U S A* **109**, E2595-2604, doi:10.1073/pnas.1212834109 (2012).
52 McGinnis, C. S. *et al.* MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nature Methods* **16**, 619-626, doi:10.1038/s41592-019-0433-8 (2019).
53 Melsted, P. *et al.* Modular and efficient pre-processing of single-cell RNA-seq. *bioRxiv*, 673285, doi:10.1101/673285 (2019).
54 Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology* **19**, 15, doi:10.1186/s13059-017-1382-0 (2018).
55 Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502, doi:10.1038/s41586-019-0969-x (2019).
56 Dorrity, M. W., Saunders, L. M., Queitsch, C., Fields, S. & Trapnell, C. Dimensionality reduction by UMAP to visualize physical and genetic interactions. *Nat Commun* **11**, 1537, doi:10.1038/s41467-020-15351-4 (2020).
57 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127, doi:10.1093/biostatistics/kxj037 (2007).
58 Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882-883, doi:10.1093/bioinformatics/bts034 (2012).

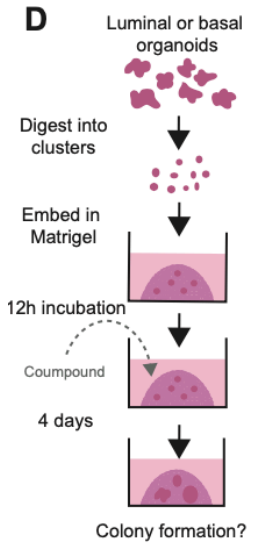
Figure 1: TNBC cells present mesenchymal characteristics related to EMT





% of leader cells Ecad⁺ Vim⁺			
Luminal (MMTV-PyMT)	Basal (C3(1)-Tag)	TM00096	TM00089
0 ± 3 (n=63, r=3)	**100** ± 3 (n=63, r=3)	**93** ± 2 (n=21, r=3)	**94** ± 1 (n=21, r=3)





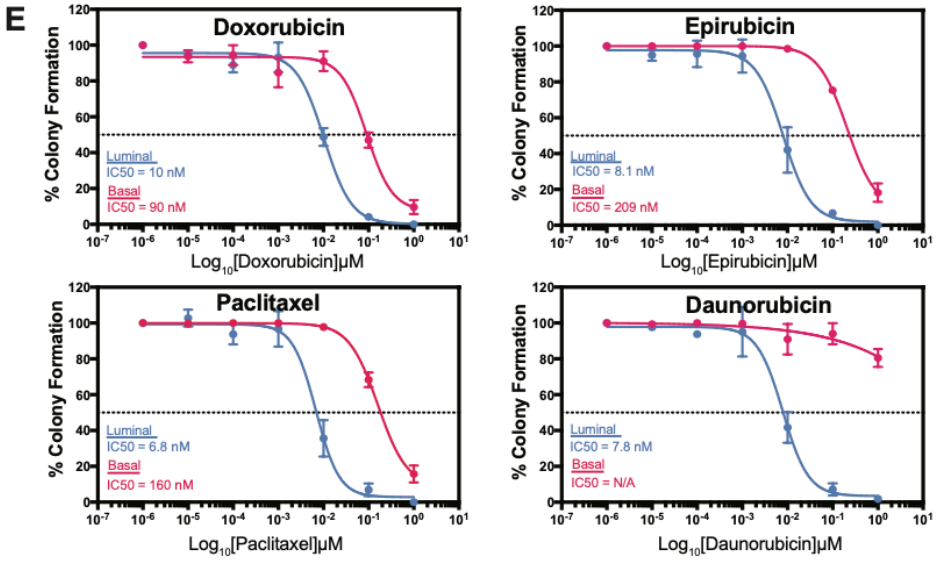
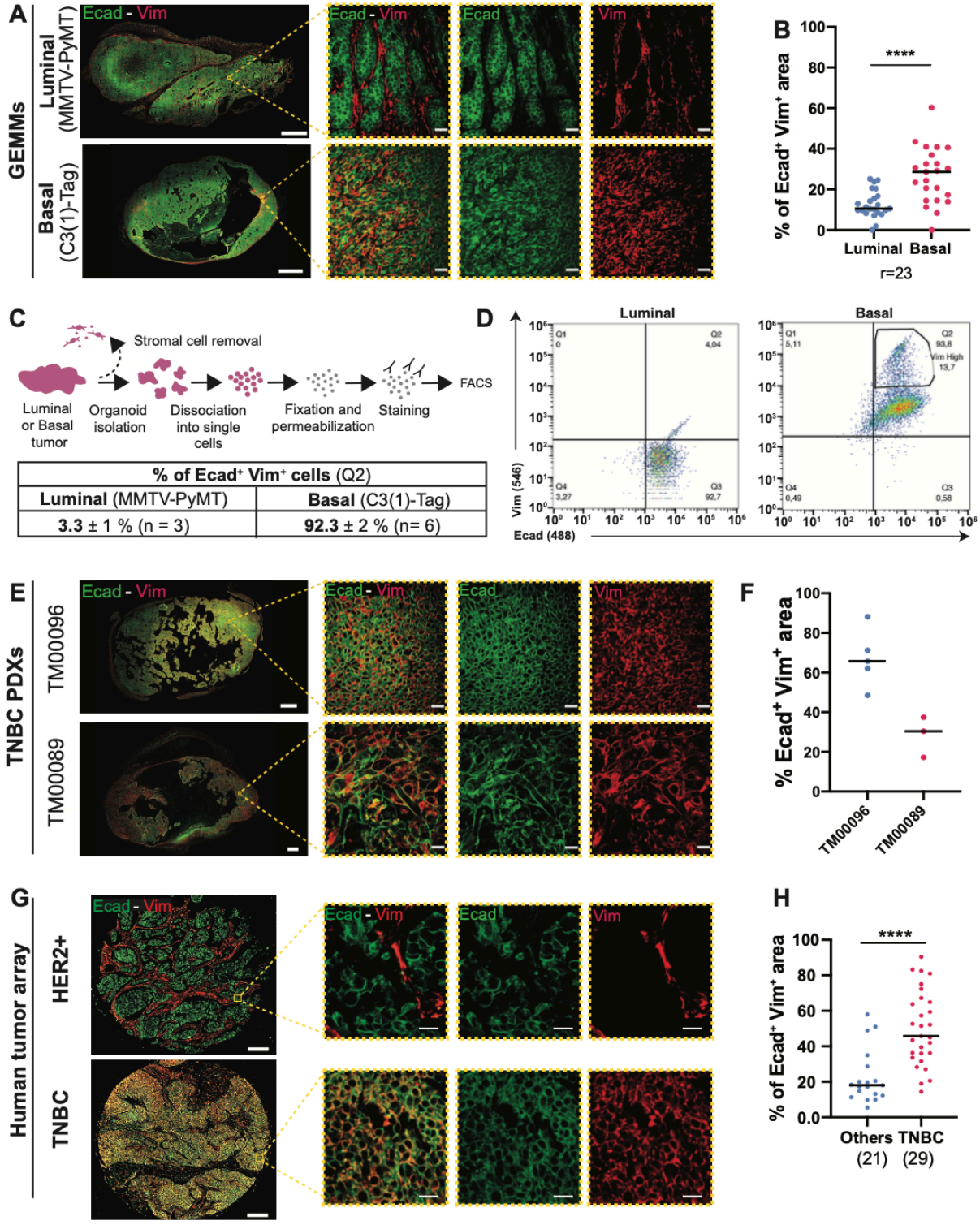


Figure 1: TNBC present mesenchymal characteristics related to EMT.
A) Schema of organoid isolation and invasion assay. The primary tumor is digested into tumor organoids and the stromal cells are removed by differential centrifugation. Each organoid is composed of 200 to 500 adherent cancer cells and embedded in 3D collagen I matrix. After 5 days, the organoids are fixed for immunostaining.
B,F) Representative confocal maximum intensity projections of whole organoids stained with anti-E-cadherin (Ecad), anti-Vimentin (Vim) antibodies and DAPI in GEMMS (B) and TNBC PDXs (F). Scale bars are 100 µm and 50 µm (zoom area). Frequency of leader cells expressing Ecad and vim is indicated in the table below the images with the median of the percentage and the standard error of mean (SEM).
C) Heatmap representing the fold change of Basal GEMM organoid gene expression, assayed by qRT-PCR, compared to the mean gene expression of Luminal GEMM organoids embedded in collagen for 5 days. Each row corresponds to a different C3(1)-Tag mouse. n=5 mice, * $P<0.05$, ** $P<0.01$, ***$P<0.005$ and ****$P<0.0001$ (unpaired T-test, two sided on the deltaCT value between C3(1)-Tag and MMTV-PyMT mice).
D) Colony formation assay scheme. Epithelial organoids are dissociated into clusters and embedded into Matrigel. After 12 hours, DMSO or compounds are added directly into the media. After 4 days, the cultures are fixed for imaging.
E) Dose response of the frequency of colony formation relative to DMSO condition in the presence of chemotherapy in the Basal GEMM and the Luminal GEMM. n = 3 per model.

Figure 2: Hybrid EMT cells are abundant in TNBC GEMM, PDXs, and patient primary tumors.



% of Ecad⁺ Vim⁺ cells (Q2)	
Luminal (MMTV-PyMT)	**Basal** (C3(1)-Tag)
3.3 ± 1 % (n = 3)	**92.3** ± 2 % (n= 6)

Figure 2: Hybrid EMT cells are abundant in TNBC GEMM, PDXs and patient primary tumors.
A,E) Representative epifluorescence (whole tumor, left) and confocal image (zoom, right) of MMTV-PyMT, C3(1)-Tag (A) and TNBC PDXs TM00089 and TM00096 (E) primary tumors stained with Ecad and Vim. Scale bars are 2000 µm for the whole tumor and 25 µm for the zoom.
B, F) Percentage of $Ecad^+$ Vim^+ area in whole MMTV-PyMT, C3(1)-Tag (B) and PDXs (F) primary tumors. Percentage calculated by pixel colocalization analysis. Each dot corresponds to a tumor, bars show the medians. ****P<0.0001 (two side T-Test)
C) Schema of FACS analysis of tumor organoids. After isolation, organoids are dissociated into single cells, fixed and permeabilized, stained with Ecad and Vim and analyzed on a flow sorter. The mean percentage of $Ecad^+$ Vim^+ cells with the standard error of mean (SEM) are indicated in the table. n=3 mice for MMTV-PyMT and n=6 mice for C3(1)-Tag.
D) Representative dot-plots of Ecad and Vim expression in tumor cells. 92.7% of MMTV-PyMT cells are $Ecad^+$ Vim^- (n= 2749 cells), 93.8% of C3(1)-Tag cancer cells are $Ecad^+$ Vim^+ (n=7927 cells).
G) Representative epifluorescence images of human primary breast tumors stained with Ecad and Vim. Scale bar is 200 µm and 20 µm for the zoom area.
H) Percentage of $Ecad^+$ Vim^+ double positive area in human breast tumors (n=21 tumor from other subgroups, n= 29 TNBC tumors). Percentage calculated by pixel colocalization analysis. Each dot corresponds to a tumor, bars are medians. ****P<0.0001 (Mann-Whitney).

Figure 3: Single-cell RNA sequencing reveals a temporal progression towards EMT during invasion.

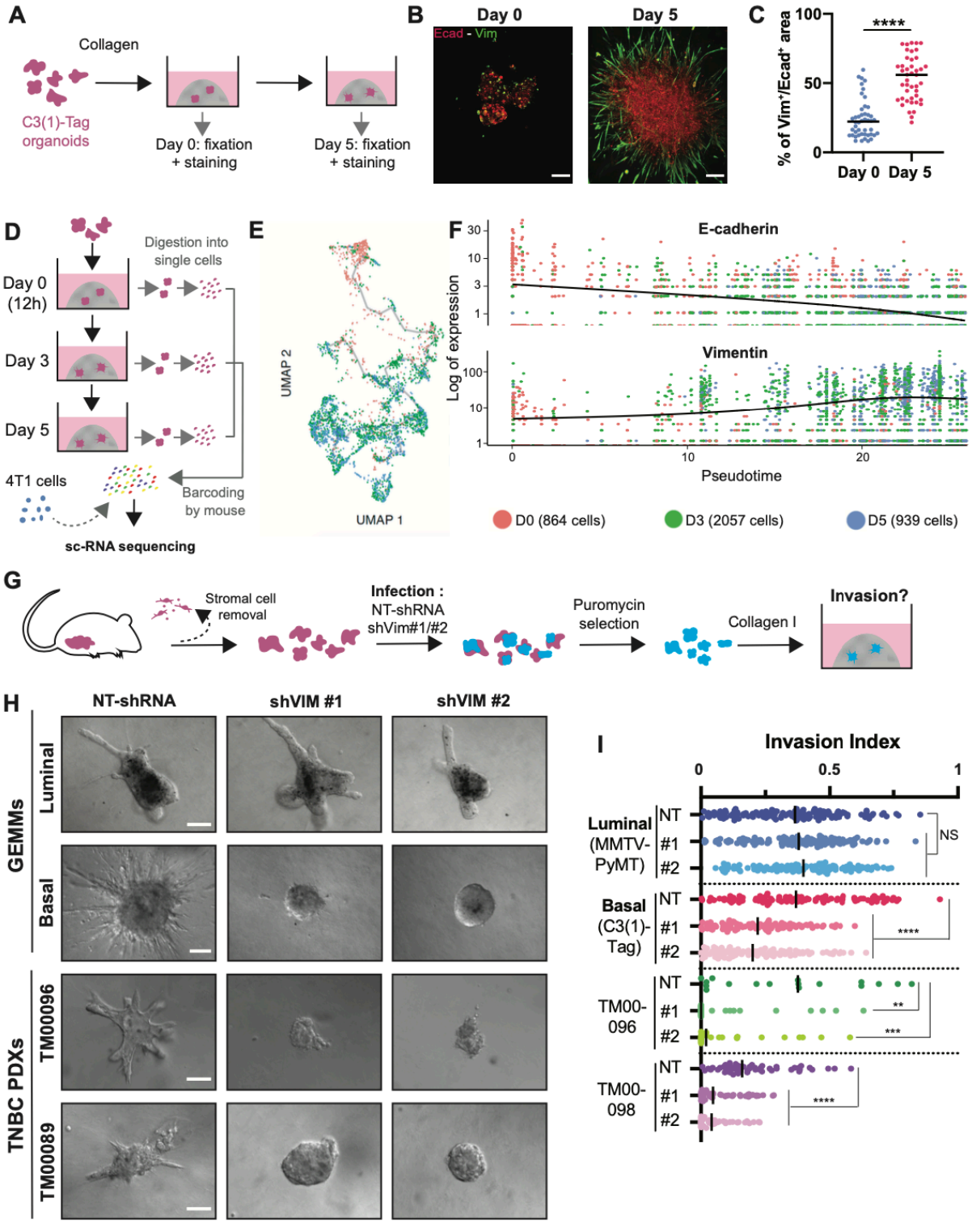


Figure 3: Single-cell RNA sequencing reveals a temporal progression towards EMT during invasion.

A) Invasion assay scheme. C3(1)-Tag organoids are plated into a 3D collagen I matrix and fixed either at Day 0 or Day 5.

B) Representative confocal images of maximum intensity projection of whole C3(1)-Tag organoids stained with Ecad and Vim. Scale bars are 100 µm.

C) Percentage of Vim^+ area over $Ecad^+$ area. Each dot corresponds to an organoid. Bars are medians. ****P<0.0001 (Mann-Whitney). n=91, r=3.

D) Scheme of the single-cell RNA sequencing experiment. C3(1)-tag organoids are embedded in collagen I. After 12h, 3 days and 5 days, organoids are extracted from the matrix and dissociated into single cells. These single cells and 4T1 cells, used as a reference for batch correction, are barcoded per condition before flow sorting for live cells using propidium iodide. Live cells are next used for 10X Genomics barcoding and library preparation. All conditions are then sequenced together.

E) UMAP and pseudotime analysis of the sequenced cells from collagen. Each dot corresponds to single cell and the colors indicate the day the cells were extracted from the collagen I. The grey line corresponds to the cell trajectories obtained with pseudotime.

F) Ecad and Vim expression in single cells according to pseudotime. Dots represent single cells; color scale indicate the day the cells were extracted from the collagen I.

G) Scheme of collagen I invasion assay using shRNA. After isolation, organoids are infected with either non-targeting shRNA (NT-shRNA) as control, or two shRNA sequences against vimentin (shVim #1 and shVim #2). After puromycin selection to remove the non-infected cells, organoids are plated into collagen I.

H) Representative DIC images of organoids infected with NT shRNA, shVim #1 or shVim #2 after 5 days in collagen I matrix. Scale bars are 100 µm.

I) Quantification of organoid invasion represented in H. Each dot corresponds to one organoid, bars show the medians. NS: not significant, ****P<0.0001, **P=0.0065 and ***P=0.0002 (Kruskal-Wallis). n=926, r=3.

Figure 4: Single-cell RNA sequencing reveals a temporal progression towards MET during metastatic outgrowth.

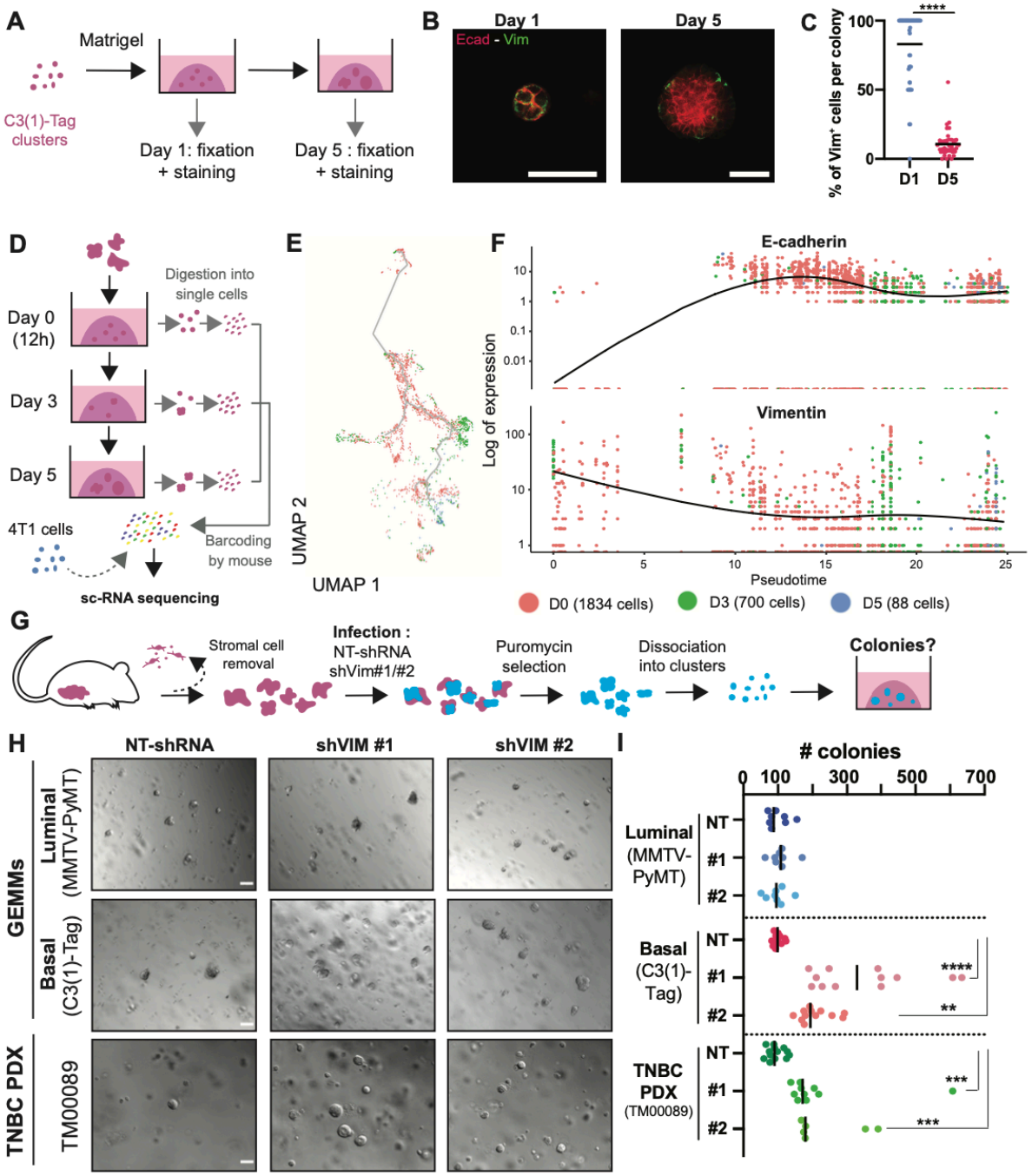


Figure 4: Single-cell RNA sequencing reveals a temporal progression towards MET during metastatic outgrowth.

A) Schema of colony formation assay. C3(1)-Tag organoids are dissociated into clusters and plated into Matrigel matrix. Cultures are fixed at Day 1 and Day 5.

B) Representative confocal images of C3(1)-Tag clusters stained with Ecad and Vim at Day 1 and Day 5. Scale bars are 50 µm.

C) Percentage of Vim^+ cells per colony. Each dot corresponds to a colony or cluster. Bars are medians. ****P<0.0001 (Mann-Whitney). n=99, r=3.

D) Scheme of the single-cell RNA sequencing experiment. C3(1)-tag clusters are embedded in Matrigel. After 12h (day 0), 3 days and 5 days, colonies are extracted from the matrix and dissociated into single cells. These single cells and 4T1 cells, used as a reference, are barcoded per condition before flow sorting for live cells using propidium iodide. Live cells are next used for 10X Genomics barcoding and library preparation. Finally, all the conditions are sequenced at the same time.

E) UMAP of pseudotime analysis of the sequenced cells from the colony formation assay. Each dot corresponds to a single cell and the colors indicate the day the cells were extracted from the Matrigel. The grey line corresponds to the cell trajectories obtained with pseudotime.

F) Ecad and Vim expression in single cells according to pseudotime. Dots represent single cells; color scale indicate the day the cells were extracted from the Matrigel.

G) Scheme of colony formation assay using shRNA. After isolation, organoids are infected with either NT shRNA as control, or two shRNA sequences against vimentin (shVim #1 and shVim #2). After puromycin selection to remove the non-infected cells, organoids are dissociated into clusters and embed in Matrigel.

H) Representative DIC images of colonies infected with NT shRNA, shVim #1 or shVim #2 after 5 days (GEMM) or 10 days (PDX) in Matrigel. Scale bars are 100 µm.

I) Number of colonies per well. Each dot represents the number of colonies in one well, bars show the medians. NS = not significant, **P=0.008, ***P<0.001 ****P<0.0001 (one-way ANOVA). n=89, r=3.

Figure 5: The molecular regulators of invasion and metastatic outgrowth.

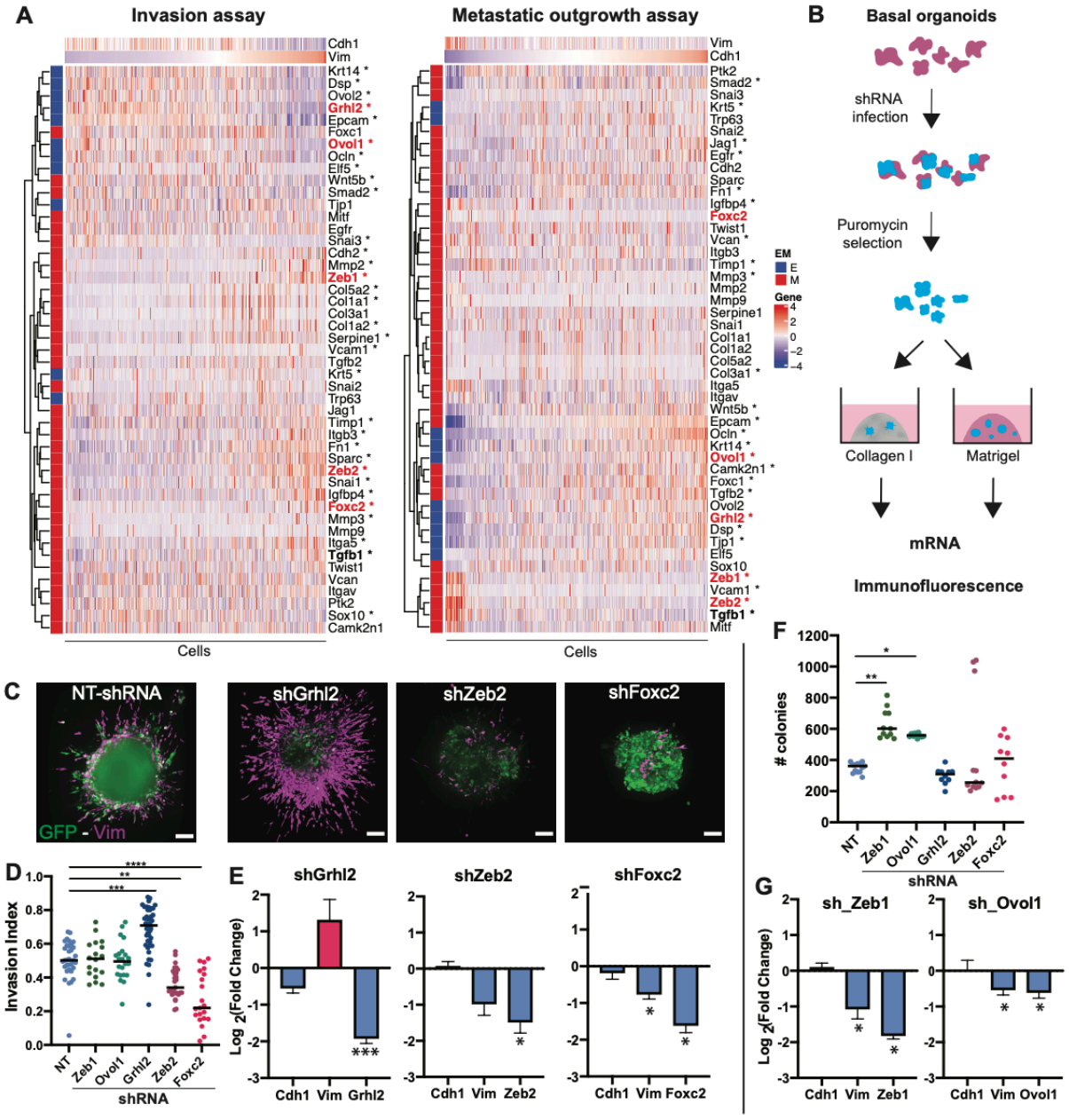


Figure 5: The molecular regulators of invasion and metastatic outgrowth.
A) Heatmap of the SAVER imputed and batch corrected log transformed expression values expression of EMT genes in the invasion assay or metastatic outgrowth assay. Each row corresponds to an individual cell ordered by its *Vim* or *Ecad* expression. Expression values for each gene are correlated with *Vim* or *Ecad* expression using a linear model from Monocole3, where * indicates a q-value < 0.05.
B) Schema for shRNA-based gene knock-down in the invasion and colony formation assays. After isolation, organoids are infected with either NT shRNA as control, or a pooled of 3 shRNA sequences against a specific gene. After puromycin selection to remove the non-infected cells, organoids are directly embedded into collagen I or dissociated into clusters and embedded in Matrigel. After 5 days, organoids are either fixed and immunostained for vimentin or RNA is extracted for qPCR.
C) Representative confocal images of maximum intensity projection of whole Basal GEMM organoids stained with Vim. GFP indicated infected cells. Scale bars are 100 μm.
D) Quantification of organoid invasion represented in C. Each dot corresponds to one organoid, bars show the medians. **P=0.0016, ***P=0.0003, ****P<0.0001 (Kruskal-Wallis), n=163, r=3.
E,G) Fold change of Basal GEMM organoids gene expression compare to NT-shRNA condition. * P<0.05, and ***P<0.001 (paired T-test, two sided on the deltaCT value between NT shRNA and shRNA against specific gene), r≥3.
F) Number of colonies per well. Each dot represents the number of colonies in one well, bars show the medians. *P=0.0267, **P=0.034 (Kruskal-Wallis), n=67, r=3.

Figure 6: Vimentin promotes invasion and represses metastatic outgrowth *in vivo*.

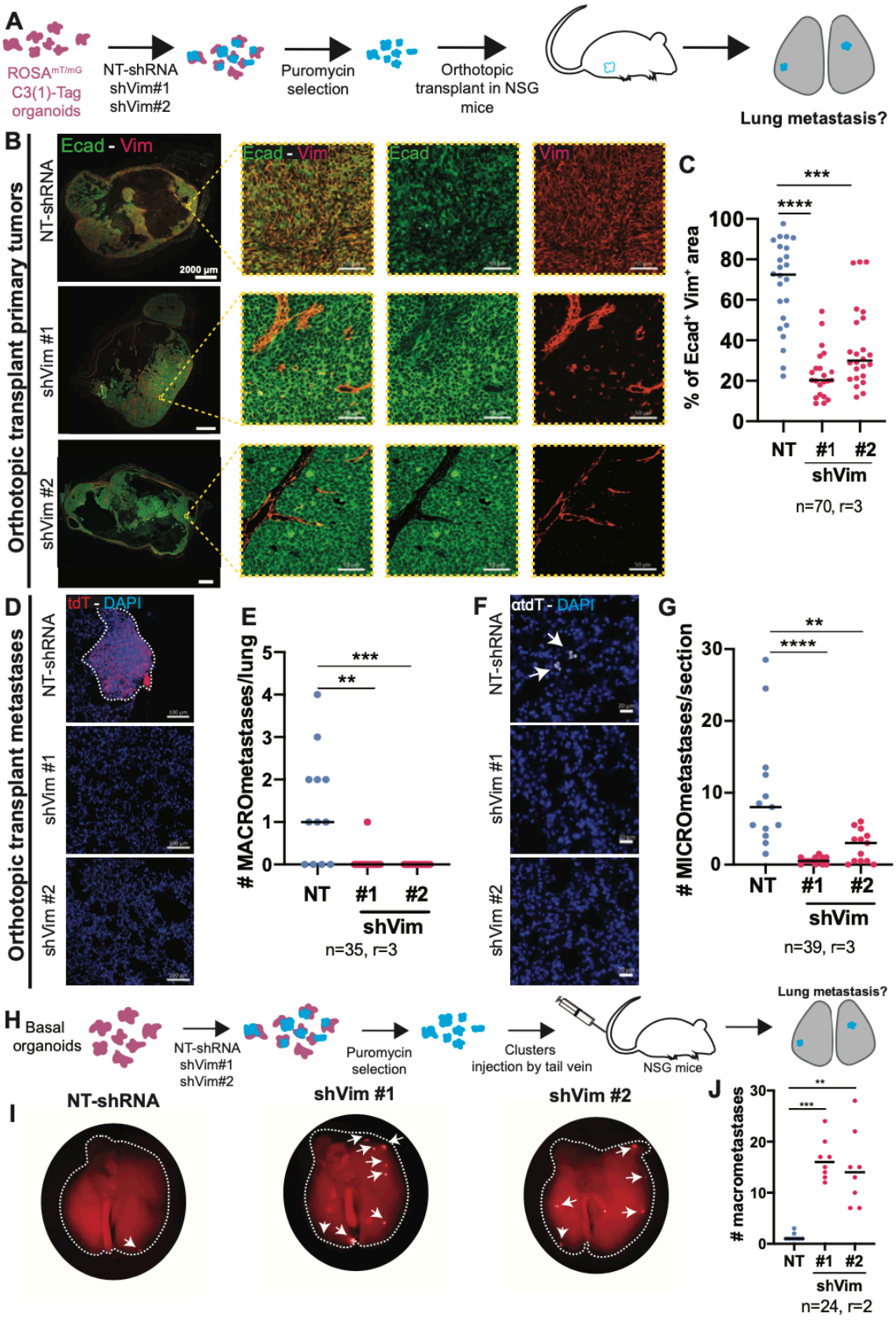


Figure 6: Vimentin promotes invasion and represses metastatic outgrowth *in vivo*.
A) Scheme of transplant assay to study metastasis. Fluorescent organoids expressing membrane tdTomato are isolated and infected with lentivirus expression NT shRNA or shVim #1 or shVim#2. After puromycin selection organoids were orthotopically transplanted into the cleared mammary fat pad of NSG mice. After 6-8 weeks, when the primary tumors are 20mm, lungs are collected.
B) Representative epifluorescence images of whole primary tumors stained with Ecad and Vim. Scale bars are 2000 µm and 50 µm.
C) Percentage of $Ecad^{+}$ Vim^{+} double positive area in whole C3(1)-Tag primary tumors developed from NT shRNA, shVim #1 or shVim #2 infected organoids transplanted into the cleared mammary fat pad of NSG mice. Percentage calculated by pixel colocalization analysis. Each dot corresponds to a tumor, bars show the medians. ***P=0.0006, ****P<0.0001 (Kruskal-Wallis).
D) Representative confocal images of lung sections stained with DAPI. Metastasis are detected by membrane tomato (tdT) signal. Scale bars are 100 µm.
E) Quantification of the number of MACROmetastases (>200 cells). Each dot corresponds to the number of macrometastases in whole lungs. Bars show the medians. **P=0.0019, ***P=0.0006 (Kruskal-Wallis).
F) Representative epifluorescence of MICROmetastases in the lungs. Micrometastases were detected by anti-tdT immunostaining and DAPI. Scale bars are 20µm. Arrows indicate single cells micrometastases.
G) Number of MICROmetastases per lung section (20 µm thickness). Each dot corresponds to the number of micrometastases in a lung section. Bars show medians. **P=0.0082, ****P<0.0001 (Kruskal-Wallis).
H) Tail vein assay scheme to study the late steps of the metastatic cascade. Fluorescent organoids are infected with NT shRNA or shVim #1 or shVim #2. After selection, organoids are dissociated into clusters and injected into the tail vein of NSG mice. After 4 weeks, the number of metastases in the lungs is assessed.
I) Representative epifluorescence images of whole lungs. Macrometastases indicated by arrows are detected by tdT signal.
J) Number of macrometastases per lung. Each dot corresponds to a lung. Bars show medians. **P=0.0036, ***P=0.0004 (Kruskal-Wallis).

Figure 7: Cellular strategies of metastasis by hybrid EMT cells.

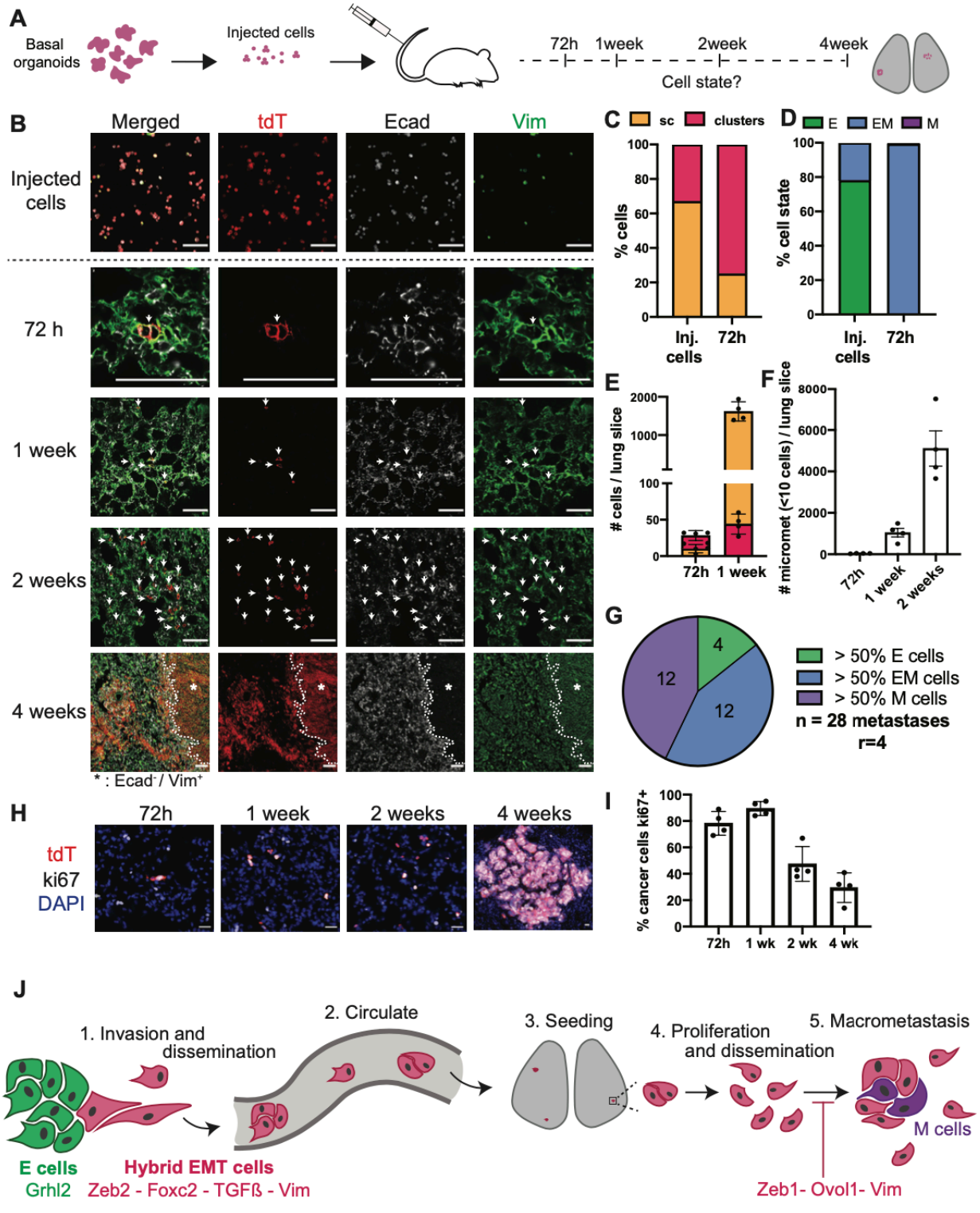
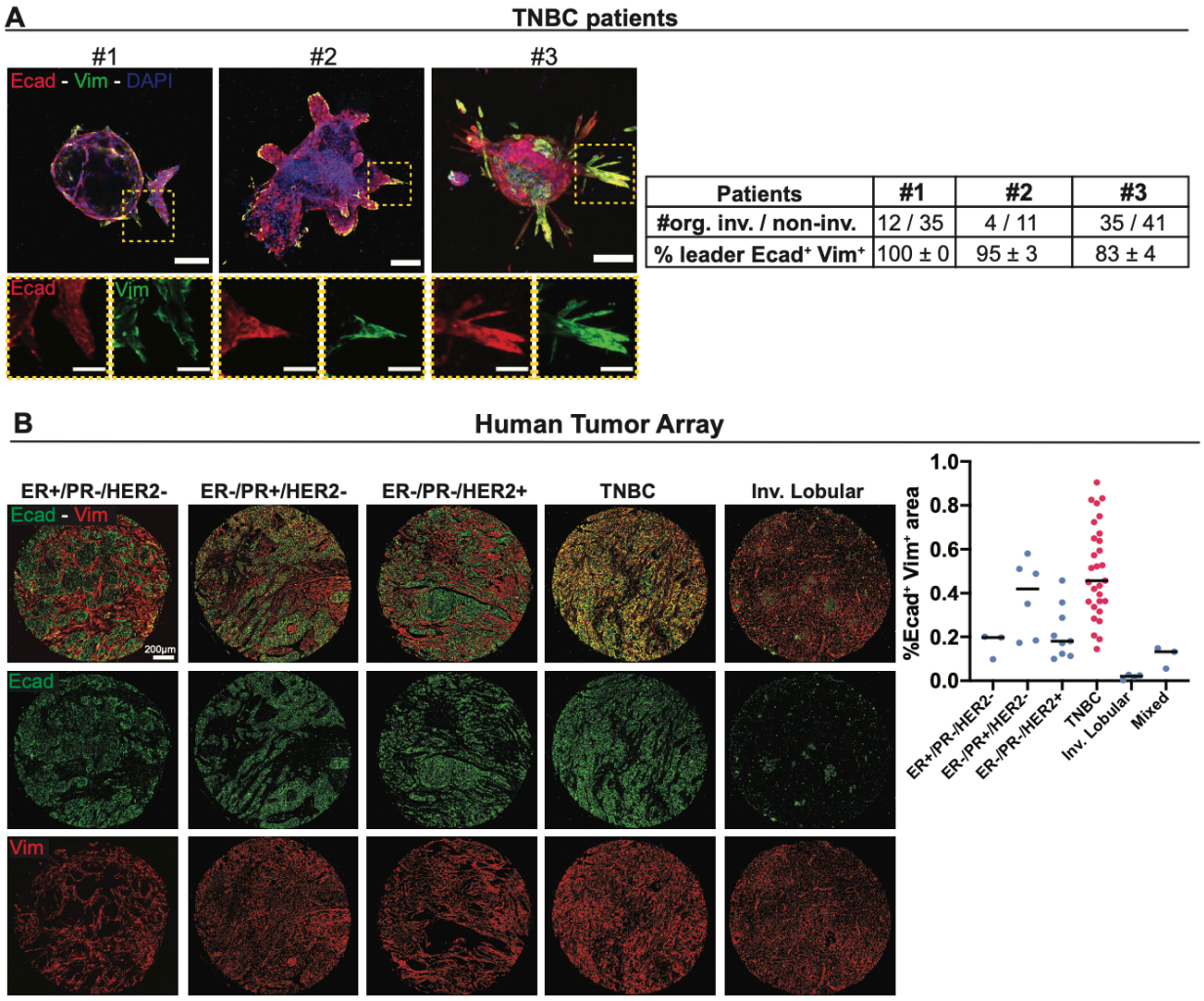


Figure 7: Cellular strategies of metastasis by hybrid EMT cells.
A) Scheme of time course assay to study metastatic outgrowth *in vivo*. Fluorescent organoids expressing membrane tdTomato are isolated, digested into single cells and clusters and injected into the tail vein of NSG mice. Lungs are collected at 72h, 1 week, 2 weeks and 4 weeks.
B) Representative confocal images of the injected cell suspension and mice lung sections at 72h, 1 week, 2 weeks and 4 weeks stained with Ecad and Vim. Scale bars are 50µm (injected cells, 72h, 1 week, 2 weeks) and 100µm (4 weeks). Arrows indicate cancer cells. * area of cancer cells negatives for Ecad.
C) Percentage of single versus clusters (2-10 cells) of cancer cells within injected cells and cells detected into the lungs at 72h. n=4.
D) Percentage of epithelial (E=Ecad^{+} Vim^{-}), hybrid (EM= Ecad^{+} Vim^{+}) and mesenchymal (M=Ecad^{-} Vim^{+}) cancer cells within injected cells and cells detected into the lungs at 72h. n=4.
E) Number of single and clusters of cancer cells per lung slice at 72h and 1 week. Scatter dot plot representing the mean with SEM. n=4.
F) Number of micrometastasis (single cells and clusters combined) per lung slice at 72h, 1 week and 2 weeks. Scatter dot plot representing the mean with SEM. n=4.
G) Proportion of macrometastases composed of more than 50% of epithelial cells (E=Ecad^{+} Vim^{-}), more than 50% of hybrid cells (EM= Ecad^{+} Vim^{+}) and more than 50% of mesenchymal cells (M=Ecad^{-} Vim^{+}). n=28, r=4.
H) Representative epifluorescence images of lung sections stained for ki67. Scale bars are 20µm.
I) Percentage of cancer cells positive for ki67 at 72h, 1 week, 2 weeks and 4 weeks. Scatter dot plot representing the mean with SEM. n=4.
J) Summary scheme. Basal or triple negative cancer cells acquire hybrid EMT characteristics during invasion. *Grhl2* represses this transition while *Zeb2*, *Foxc2* and *TGFß* promote it. Clusters of hybrid EMT cells travel in the circulation and seed distant organs. After seeding, these clusters proliferate and disseminate as single cells until they form macrometastases often composed of hybrid EMT cells and mesenchymal cells (M= Ecad^{-} Vim^{+}).

Figure S1: Hybrid EMT cell characteristics in TNBC patient primary tumors



Patients	#1	#2	#3
#org. inv. / non-inv.	12 / 35	4 / 11	35 / 41
% leader Ecad$^+$ Vim$^+$	100 ± 0	95 ± 3	83 ± 4

Figure S1: Hybrid EMT characteristics in TNBC patient primary tumors.
A) Representative confocal maximum intensity projections of whole organoids extracted from TNBC patients, stained with anti-E-cadherin (Ecad), anti-Vimentin (Vim) antibodies and DAPI. Scale bars are 100 µm and 50 µm (zoom area). Frequency of leader cells expressing Ecad and vim is indicated in the table with the median of the percentage and the standard error of mean (SEM).
B) Representative epifluorescence images of human primary breast tumors stained with Ecad and Vim (scale bar is 200 µm) and percentage of $Ecad^{+}$ Vim^{+} double positive area in human breast tumors. Each dot corresponds to a tumor, bars are medians.

Figure S2: Basal GEMM cancer cells undergo EMT during invasion.

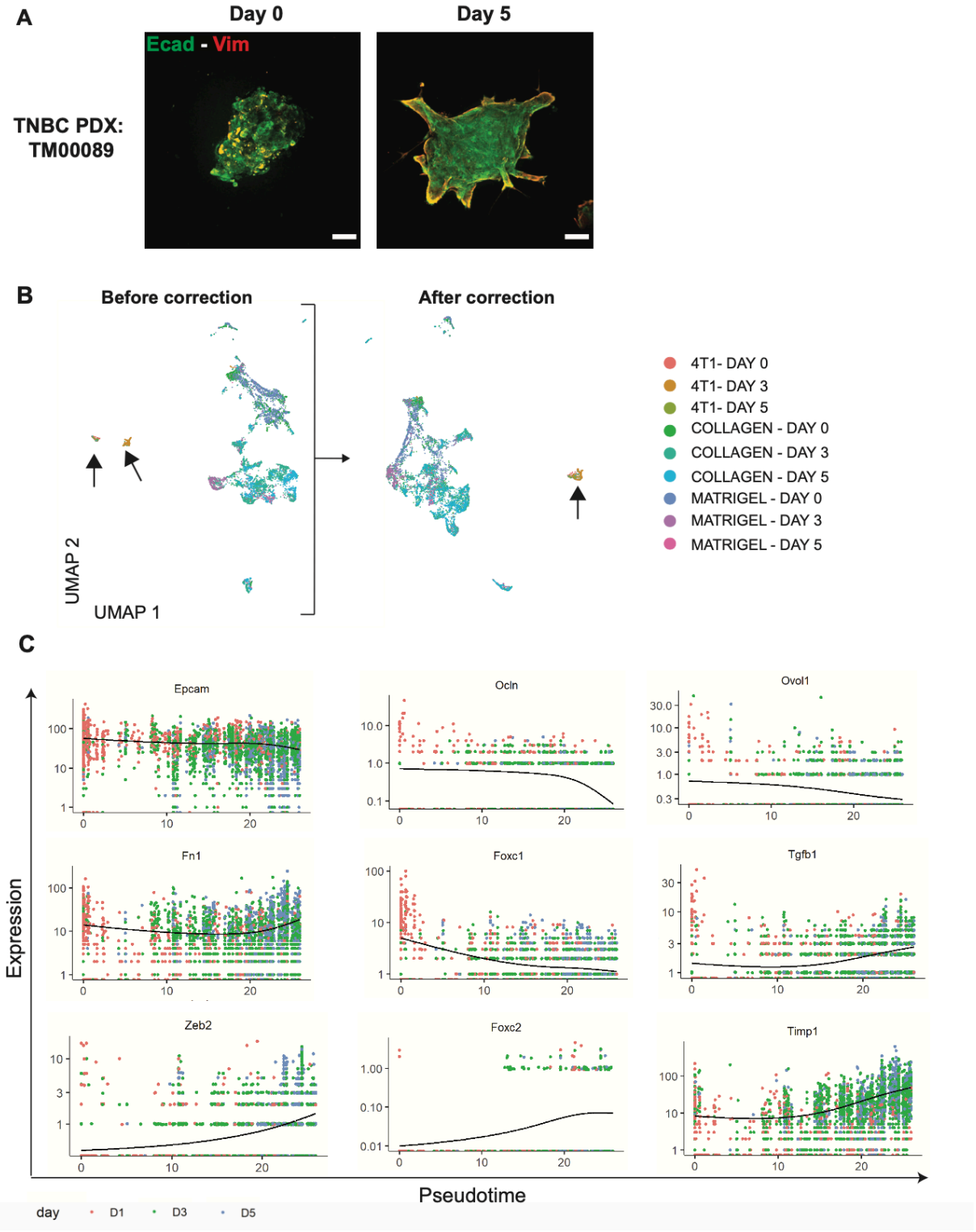


Figure S2: Basal GEMM cancer cells undergo EMT during invasion

A) Representative confocal images of maximum intensity projection of whole TNBC PDX organoids stained with Ecad and Vim. Scale bars are 100 µm.

B) UMAP of scRNA-seq experiment with invasion and colony formation assays before and after correction for batch effect with Batchelor. Before correction, 4T1 cells, indicated by arrows, do not cluster together while after correction they cluster together. Each dot corresponds to a single cell and the colors indicate the condition the cells were cultured.

C) Epithelial and mesenchymal gene expression in single cells according to pseudotime. Dots represent single cells; color scale indicate the day the cells were extracted from the collagen I.

Figure S3: Vimentin is required for Basal GEMM and TNBC PDX organoid invasion.

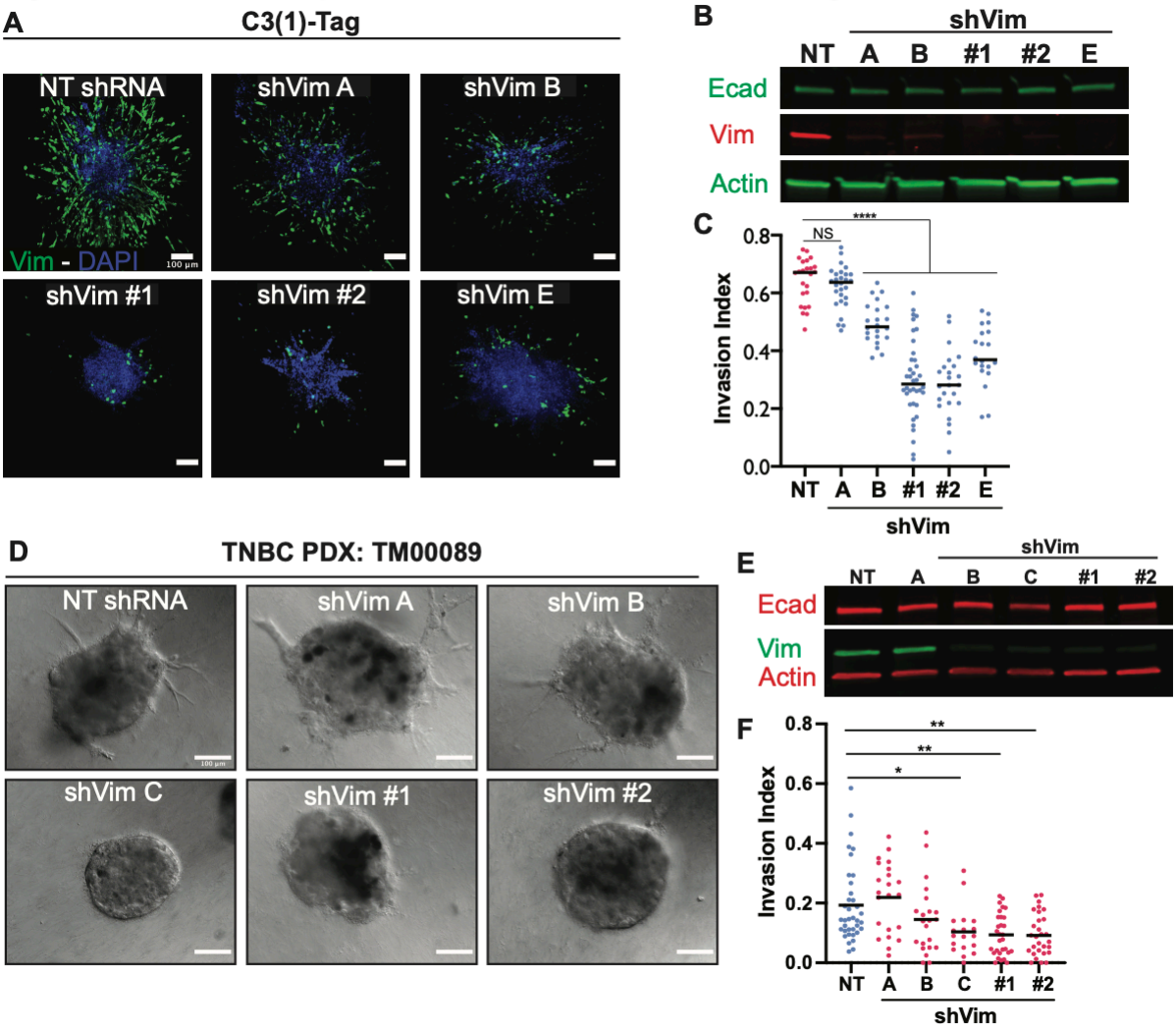


Figure S3: Vimentin is required for Basal GEMM and TNBC PDX organoid invasion
A) Representative confocal images of maximum intensity projections of whole Basal GEMM organoids stained with vimentin (Vim) and DAPI. Scale bars are 100 μm.
B) Representative western blot depicting protein levels of Ecad, Vim and actin as a loading control.
C) Quantification of organoid invasion. Each dot corresponds to an organoid, bars show the medians. ****P<0.0001 (ANOVA). n=156, r=1.
D) Representative DIC images of the TNBC PDX TM00089 organoids infected with NT shRNA as control or different shRNA sequences targeted against vim after 5 days in collagen I matrix. Scale bars are 100 μm.
E) Representative western blot depicting protein levels of E-cad, Vim and actin as a loading control from TNBC PDX TM00089 organoids infected with NT shRNA as control or different shRNA sequences targeted vim.
F) Quantification of organoid invasion. Each dot corresponds to an organoid, bars show the medians. NS= not significant, *P=0.0275, **P<0.01 (Kruskal-Wallis). n=159, r=1.

Figure S4: *Ex vivo* metastatic outgrowth is associated with MET gene signature.

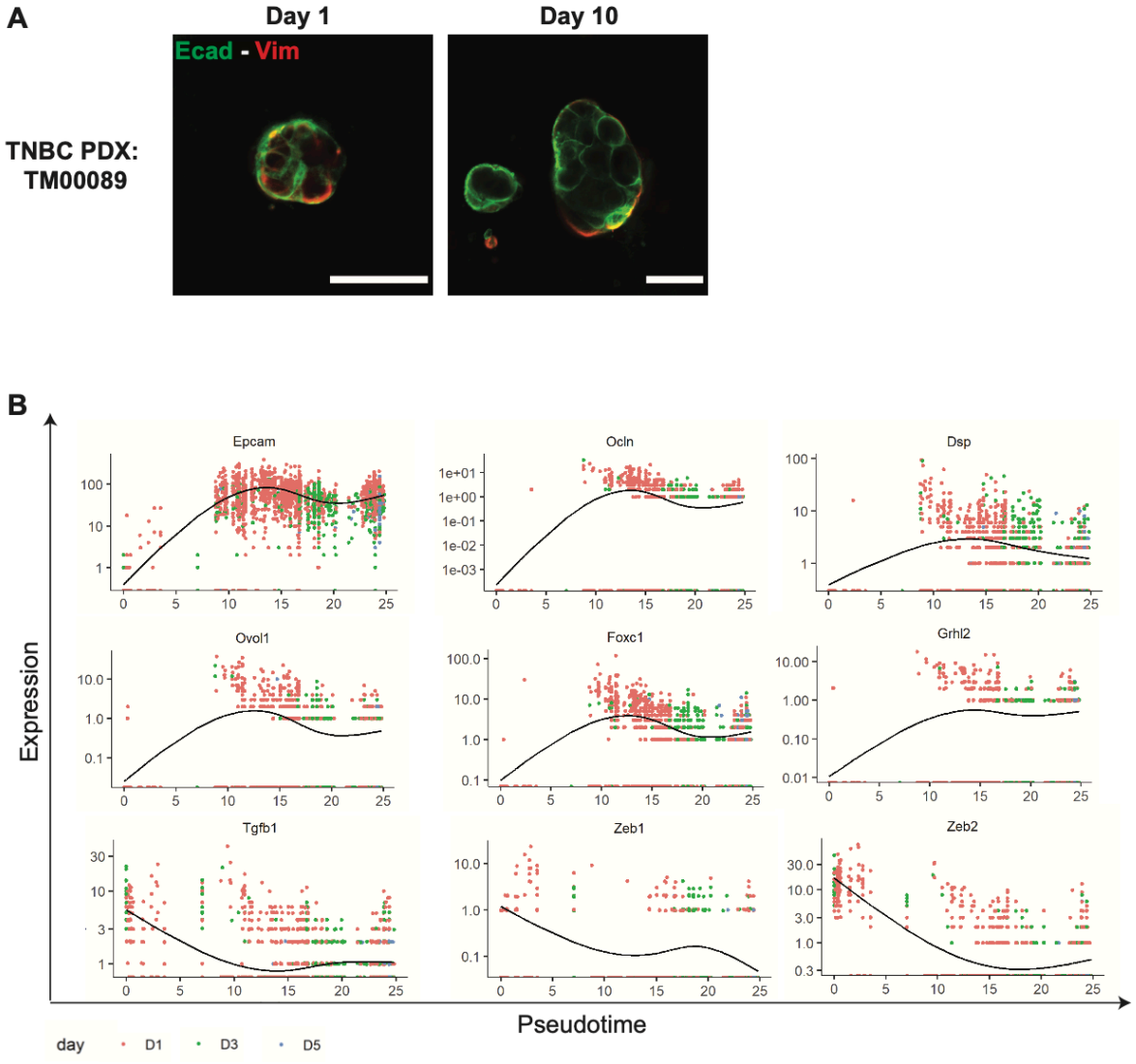


Figure S4*: Ex vivo* metastatic outgrowth is associated with MET gene signature
A) Representative confocal images of TNBC PDX clusters stained with Ecad and Vim. Scale bars are 50 μm.
B) Epithelial and mesenchymal gene expression in single cells according to pseudotime. Dots represent single cells; color scale indicate the day the cells were extracted from the Matrigel.

Figure S5: Impact of EMT regulators knock-down on EMT-related gene expression.

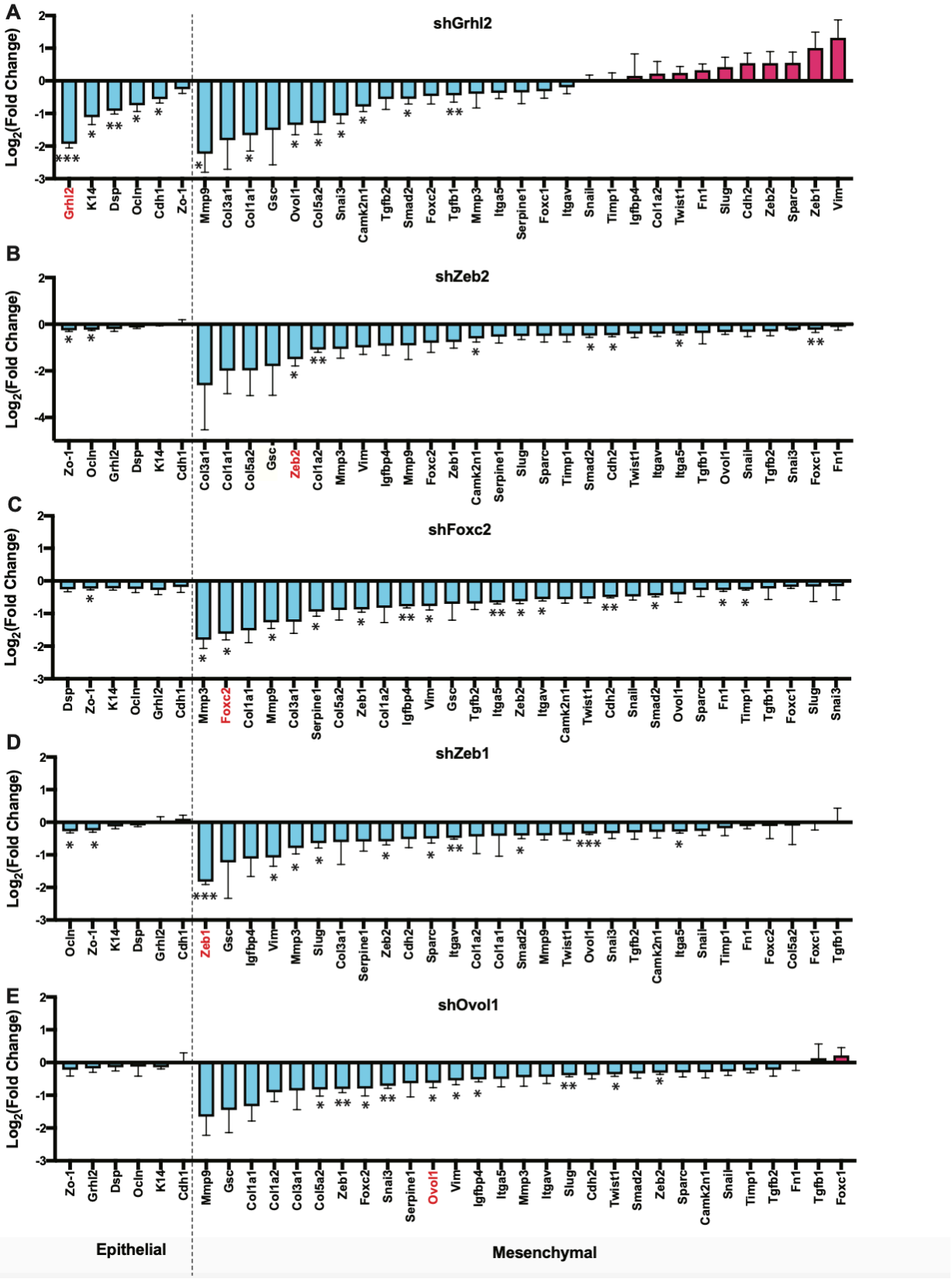


Figure S5: Impact of EMT regulator knock-down on EMT-related gene expression.
A,B,C,D, E) mRNA fold change, assayed by qRT-PCR, of epithelial and mesenchymal gene expression in Basal GEMM organoids knocked-down for *Grhl2* (A), *Zeb2* (B), *Foxc2* (C), *Zeb1* (D) and *Ovol1* (E). Histogram of mean with SEM. r ≥3. * P<0.05, ** P<0.01, ***P<0.005 and ****P<0.0001 (paired T-test, two sided on the deltaCT value between NT shRNA and shRNA against specific gene).

Figure S6: TGFβ promotes Basal GEMM breast cancer invasion without affecting metastatic outgrowth and chemoresistance.

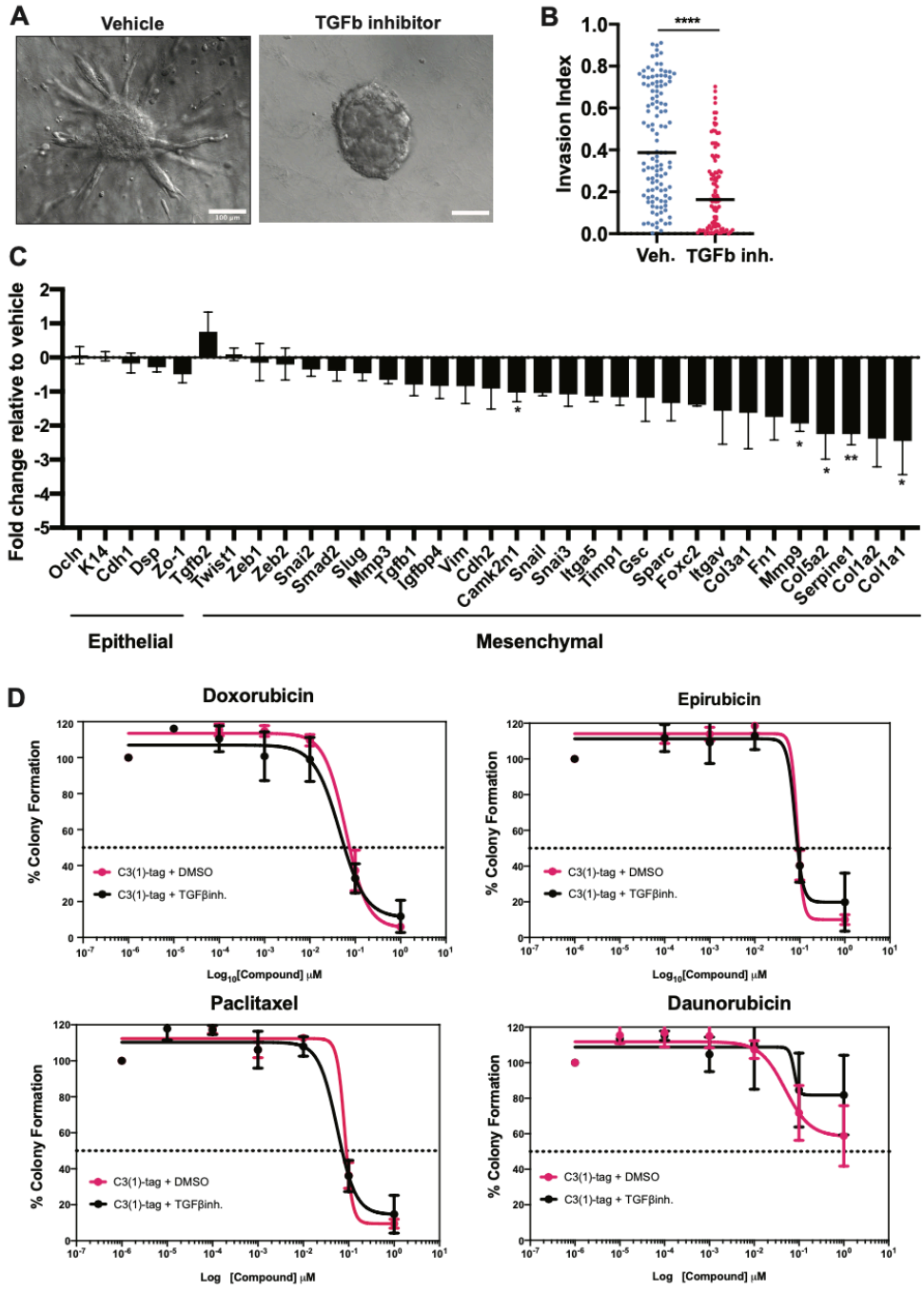


Figure S6: TGFß promotes Basal GEMM breast cancer invasion without affecting metastatic outgrowth and chemoresistance.
A) Representative DIC images of Basal GEMM organoids treated with vehicle (DMSO) or TGFß RI kinase inhibitor (1 µM) after 5 days in collagen I matrix. Scale bars are 100 µm.
B) Quantification of Basal GEMM organoid invasion in the presence of vehicle (veh.) control or TGFß RI kinase inhibitor. Dots represent organoids, bars are medians. ****P<0.0001 (Kruskal-Wallis). n=331, r=3.
C) Fold change of Basal GEMM organoids gene expression, assayed by qRT-PCR, in the presence of TGFß RI kinase inhibitor compared to the mean gene expression of DMSO-treated organoids embedded in collagen for 5 days. Histogram of mean with SEM. * P<0.05, ** P<0.01, ***P<0.005 and ****P<0.0001 (paired T-test, two sided on the deltaCT value between DMSO and TGFß RI kinase inhibitor). r=3.
D) Dose response of the frequency of colony formation relative to vehicle condition in the presence of chemotherapy in Basal GEMM organoids treated with DMSO or TGFß RI kinase inhibitor. r=3.

Figure S7: Vimentin is required for metastasis in the Basal GEMM model.

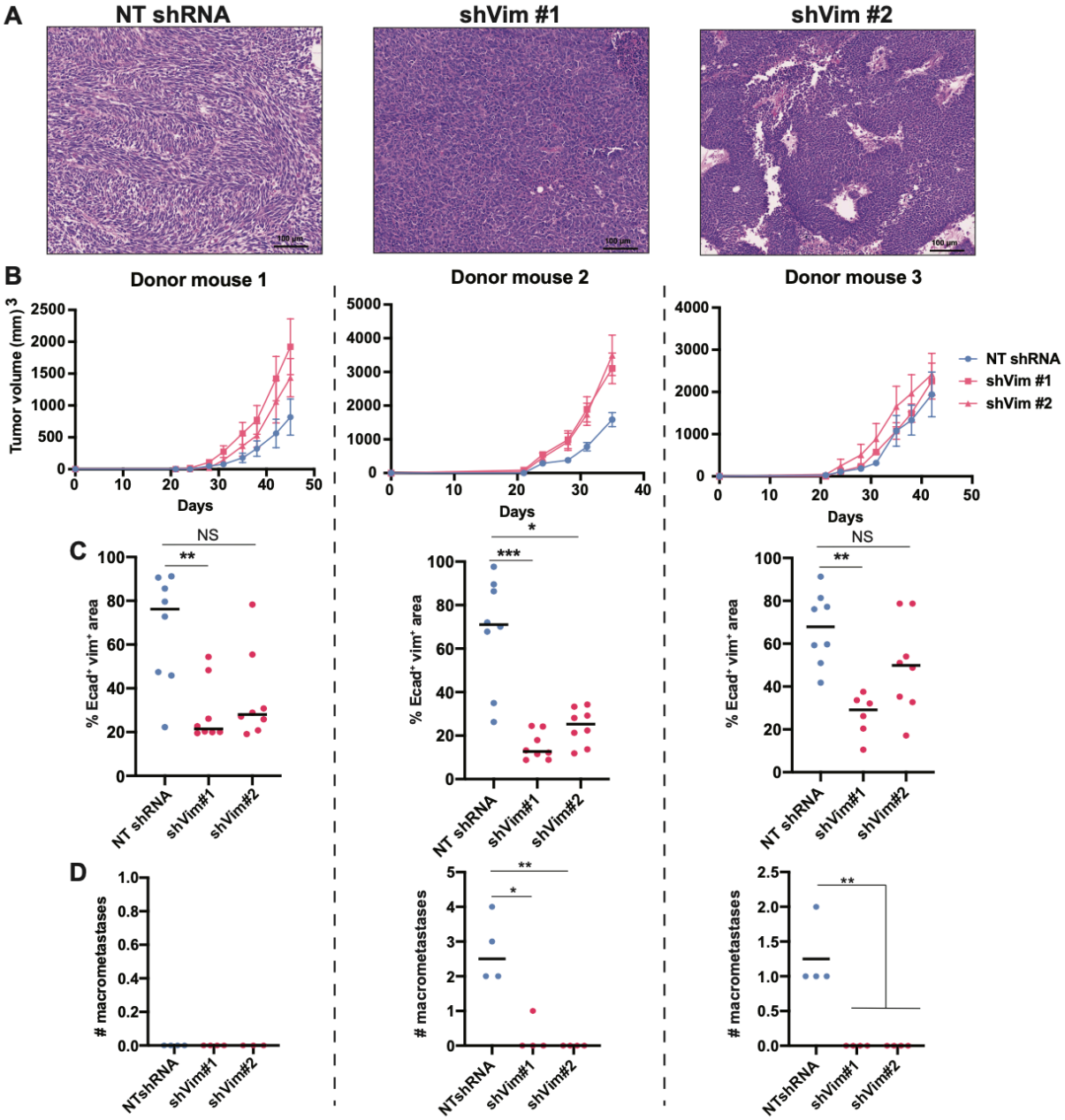


Figure S7: Vimentin is required for metastasis in the Basal GEMM model.
A) Hematoxylin and eosin staining of Basal GEMM primary tumors developed from NT shRNA, shVim #1 or shVim #2 infected organoids transplanted into the cleared mammary fat pad of NSG mice. Scale bars are 100 μm.
B) Primary tumor growth of NT shRNA, shVim #1 or shVim #2 infected organoids transplanted into the cleared mammary fat pad of NSG mice. Growth curves were stopped when one of the mice was euthanized for reaching the maximum allowed primary tumor burden. n= 24 tumors (2 tumors per mouse) for mouse donor 1 and 2 and n=22 for mouse donor 3. Error bars correspond to the standard error of the mean (SEM).
C) Percentage of $Ecad^{+}$ Vim^{+} double positive area in whole C3(1)-Tag primary tumors developed from NT shRNA, shVim #1 or shVim #2 infected organoids transplanted into the cleared mammary fat pad of NSG mice. Each dot corresponds to a tumor, bars show the medians. NS= not significant, * P<0.05, **P<0.01, ***P<0.005 (Kruskal-Wallis).
D) Number of macrometastases per lung. Each dot corresponds to a lung. Bars show medians. * P<0.05, **P<0.01 (Kruskal-Wallis).

Figure S8: Clusters are more efficient than single cells to form metastasis

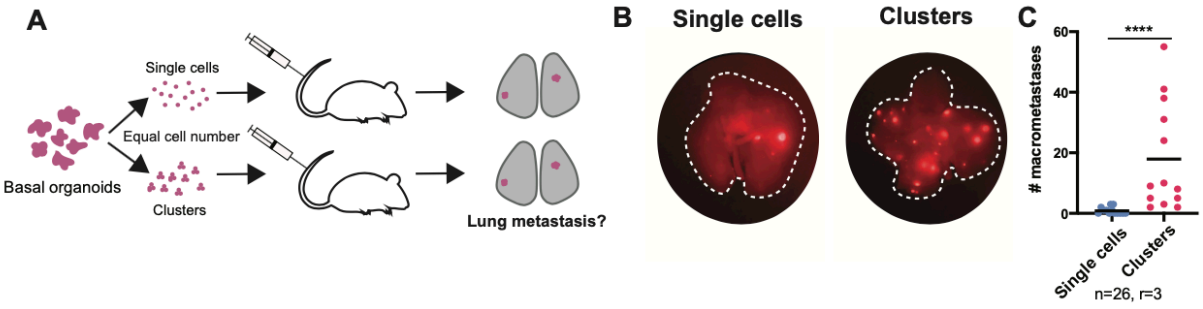


Figure S8: Clusters are more efficient than single cells to form metastasis.
A) Tail vein assay scheme. Fluorescent organoids are dissociated into clusters or single cells and the same number of cells is injected into the tail vein of NSG mice. After 4 weeks, the number of metastases in the lungs is assessed using tdT fluorescence under dissecting scope.
B) Representative epifluorescence images of whole lungs. Macrometastases are detected by red fluorescence.
C) Number of macrometastases per lung. Each dot corresponds to a lung. Bars show medians. ****P<0.0001 (Mann-Whitney).

Movie 1: Vimentin suppresses C3(1)-Tag cancer cell invasion ex vivo.
This movie shows a time-lapse sequence of representative C3(1)-Tag organoids infected with NT shRNA (left organoid), shVim #1 (middle organoid) or shVim #2 (right organoid) embedded in 3D collagen I. Scale bar 100 μm.