Inter- and intra-tumor heterogeneity in EMT states confers functional diversity in cells that lead collective invasion.

Apsra Nasir#, Sharon Camacho#, Alec T. McIntosh, Garrett T. Graham, Raneen Rahhal, Molly E. Huysman, Anna T. Riegel and Gray W. Pearson*

Lombardi Comprehensive Cancer Center and Department of Oncology,

Georgetown University, Washington, DC 20057, USA.

# Equal contribution

* To whom correspondence should be addressed.

Gray W. Pearson

Lombardi Comprehensive Cancer Center

Georgetown University

3970 Reservoir Road NW,

Washington, DC 20057

Phone- 202-687-0607

E-mail: **gp507@georgetown.edu**

Running title: There is extensive heterogeneity in cells that lead invasion.

Key Words: Heterogeneity; Epithelial-to-Mesenchymal Transition; Collective Invasion; TGFβ; EGFR

The authors declare no potential conflicts of interest.

ABSTRACT

The collective invasion of cohesive tumor cells promotes dissemination and metastasis. Here, we investigated how tumor cells are conferred with the ability to collectively invade in genetically engineered mouse mammary tumor models and breast cancer patient tumors. We discovered that the Epithelial-Mesenchymal Transition (EMT) program that induced collective invasion in basal-type tumors was phenotypically and mechanistically distinct from the K14-dependent EMT necessary for invasion in luminal B type tumors. In basal-type tumors, TGFβ induced multiple distinct states defined by different degrees of EMT reprogramming and rates of collective invasion. Notably, the EMT states with an enhanced ability to lead invasion in basal-type tumors were dependent on a complex alliance between parallel TGFβ-Smad2/3-Zeb1/2 and Egfr-Erk1/2-Fra1 pathways that involved a new mode of signal integration. Together, our results reveal inter- and intra-tumor heterogeneity in cells that lead collective invasion, suggesting breast tumor subtype specific mechanisms of metastasis.

Significance. Our results indicate that defining how cells invade and metastasize requires an understanding of how EMT functions differently in distinct subtypes of breast tumors.

INTRODUCTION

Epithelial-to-mesenchymal transition (EMT) is a form of reprogramming in which cells lose epithelial traits, such as apical-basal polarity and durable cell-cell adhesions, and gain mesenchymal properties that promote motility, including changes to the composition and organization of the cytoskeleton and the expression of ECM degrading proteases (1). EMT promotes migration during embryonic development, tissue morphogenesis and wound healing (2). Importantly, EMT is also an established feature of primary tumors that promotes progression and metastasis by inducing invasion, tumor initiating activity , suppression of the immune system and resistance to therapeutics (3). EMT was originally considered a singular process that induced a binary conversion between distinct epithelial and mesenchymal states (4). However, it has recently become clear that EMT embodies multiple different states with distinct invasive, proliferative and metastatic abilities (5), and that these distinct EMT states can be found within the same tumor population (6). Moreover, the transcriptional regulatory programs that induce EMT vary between tumors, and it is unclear how these different EMT activation mechanisms influence EMT induced phenotypes (7). Thus, precisely how EMT is induced and the factors that discriminate between different EMT states and phenotypes are unclear (5).

We focused on the regulation of a trailblazer state that promotes the collective invasion of coherent groups of tumor cells to define principles that contribute to EMT state specification. Collective invasion is the most frequent type of invasion detected in primary tumors and can promote the dissemination of tumor clusters that seed metastases (8,9). Cells in a trailblazer state have an enhanced ability to extend cellular protrusions that reorganize the ECM into microtracks that parallel the direction of invasion using force and proteolytic mechanisms (10). Notably, a trailblazer state is not required for trailing tumor cells to invade through micro tracks created by trailblazer cells (11). Thus, small subpopulations of trailblazer cells can promote the widespread invasion of the bulk tumor population through a non-cell autonomous mechanism. A trailblazer state is not a general property of all types of EMTs (12). Instead, the trailblazer state is characterized by the increased expression of genes that are specifically required to extend actin-rich cellular protrusions that promote micro track formation through the ECM, but are dispensable for other types of cell movement (11). How EMT induces the expression of genes that confer a

trailblazer state is poorly understood. Therefore, understanding how EMT induces a trailblazer state is necessary to understand the regulation of a common type of local dissemination and a potential mode of metastasis.

Here, we defined the requirements for inducing a trailblazer state through the analysis of tumor organoid invasion dynamics, an unbiased data driven functional dissection of cellular re-programming and quantitative imaging of fixed tumor sections, in genetically engineered mouse models (GEMMs) and patient tumor samples. Our analysis revealed that mammary tumor subtype dictates how EMT programs induce trailblazer states. We further determined that there was heterogeneity in the regulation and functional attributes of trailblazer cells within basal-like mammary tumors. Together, our results reveal how the integration of cell intrinsic traits and extrinsic stimuli confer inter- and intratumor heterogeneity in trailblazer populations.

MATERIALS AND METHODS

Mice. The C3-TAg [FVB-Tg(C3-1-TAg)cJeg/Jeg] were a gift from Anna T. Riegel (Georgetown University) and purchased from The Jackson Laboratory (013591). REAR [FVB/N-Tg(C3-1-TAg)cJeg/2JegJ] mice and PyMT [FVB/N-Tg(MMTV-PyVT)634Mul/J] mice were purchased from The Jackson Laboratory (REAR, 030386 and PyMT, 002374) and (**Table S1**). For orthotopic tumor experiments, 1339-org organoids were dissociated using TryPLE (Gibco, 12605-010) and 500,000 cells were bilaterally injected in the 4th fat pads of 8-12 weeks old female REAR mice. Mice were euthanized and tissue collection was performed on day 27 post injection. Mice were housed and bred in accordance with a protocol approved by the Institutional Animal Use and Care Committee at Georgetown University and in compliance with the NIH Guide for the Care and Use of Laboratory animals. Female mice were used for all experiments.

Tumor organoid derivation. The largest tumors from female C3-TAg and PyMT mice were minced and dissociated for up to 120 min at 37°C in a mixture of 1 mg/ml Collagenase (Worthington, LS004188), 20 U/ml DNase (Invitrogen, 18047-019), 5% Fetal Bovine Serum (FBS) (Peak Serum, PS-FB2), ITS (Lonza, 17-838Z), non-essential amino acids (Sigma, M7145) and 10 ng/ml FGF-2 (Peprotech, 100-18C) in DMEM/F-12 (Corning, 10-092-CV). Dissociated tumors were pelleted at 80 x *g* for 2 min and the supernatant was discarded. Tumor organoids were then rinsed up to 5 times in 5% FBS in DMEM/F-12 followed by filtering through a 250 µm tissue strainer. Organoids were then tested in invasion assays or to establish organoid lines and cell lines (**Table S1**).

Cell and organoid culture. Cell and organoid lines were generated from C3-TAg tumors **(Table S1)** and tested for mycoplasma (Lonza, LT07-703) prior to use and the creation of frozen stocks. Cells and organoids were routinely used within 25 passages. Culture conditions for all cell and organoid lines are summarized in **Table S1.** Organoids were passaged at least once per week by dissociating tumor organoids into single cell suspensions with Dispase (Sigma, SCM133) and TryPLE (Gibco, 12605-010). The cell suspensions were plated at a density of 200,000-500,000 cells per well in a 24-well ultra-low adhe-

sion plate (Sigma, CLS3474) to form multicellular aggregates for at least 16 h. The aggregates were then embedded in Matrigel (Corning, 354230) or Cultrex (Biotechne, 3533-005-02) and overlaid with Organoid Media (**Table S1**) in 6-well ultra-low adhesion plates (Sigma, CLS3471). H2B:GFP retrovirus (13) and H2B:mCherry and GFP lentivirus were produced and cells were infected as described previously (14).

Organoid invasion. Organoids in 30 µl of ECM were plated onto a 20 µl base layer of ECM in 8 well chamber slides (Falcon, 354108) and allowed to invade for 48 h unless otherwise indicated. The ECM was a mixture of 2.4 mg/ml rat tail collagen I (Corning, CB-40236) and 2 mg/ml of reconstituted basement membrane (Matrigel or Cultrex) unless otherwise indicated. Tumor organoids were analyzed in a base media of DMEM/F-12 and ITS in Fig. 1A,F, 5G and 6G, supplemented with 10 ng/ml FGF2 in 1E and 1% FBS + 10 ng/ml FGF2 in 1B, 1D. A base media of DMEM/F-12 and ITS was used for the analysis of 1339-org, 1863-org and 1788-org lines. Growth factors (2 ng/ml TGFβ1; 50 ng/ml EGF) and inhibitors were added to the base media and tested as indicated in the text and **Tables S1**. Organoids from established organoid lines **(Table S1)** were tested after dissociation and aggregation for at least 16 h as described for organoid passaging. Organoids were fixed and stained with Hoechst and Phalloidin. Images were acquired with a Zeiss LSM800 laser scanning confocal microscope using 10x/0.45 (Zeiss, 1 420640-9900-000) or 20x/0.8 (Zeiss, 1 420650-9902-000) objectives. To determine the area of invasion and circularity, an image mask for each organoid was generated using ImageJ (NIH) based on the Hoechst (invasion area) or Phalloidin (circularity) signal. The total area of the invading nuclei was determined using the "Measure" function in ImageJ. Circularity was quantified using the "Analyze Particle" function.

siRNA experiments. Cells were transfected with 50 nM of siRNA using RNAiMax transfection reagent (Invitrogen, 13778-150) for 48 h before testing. Cells in all conditions designated as "Control" were transfected with a pool of siRNAs that does not target mouse genes. The catalog numbers for each siRNA are located in **Table S1**.

Vertical invasion assay. Vertical invasion assays were performed and quantified as described previously using a Collagen/Basement Membrane (BM) mix (2.4 mg/ml Collagen I and 2 mg/ml Matrigel or Cultrex) (12). Cells expressing H2B:GFP (10,000-18,000/well) were transfected in 96-well plates (Thermo, 165305) for 48 h. Media was then replaced with 50 µl Collagen/BM) ECM and overlaid with fresh media (DMEM-F12, ITS, 1% FBS, 10 ng/ml FGF2) for 48 h. Cells were imaged with a 10X/0.30 objective (Zeiss, 1 420340-9901-000) on a Zeiss LSM800 laser scanning confocal microscope. Twenty-one z-slices at 10 µm intervals over a total span of 200 µm were acquired for 9 tiled x, y positions per well. At least 3 wells (27 x,y positions total) were imaged per condition in each experiment. ImageJ software (NIH) was used to process images and quantify invasion using a custom macro as described (11). Invasion was normalized by dividing the total number of cells that had migrated ≥ 40 µm above the monolayer by the number of cells in the monolayer. The normalized invasion value for each condition was divided by the normalized invasion for the control wells transfected with an siRNA pool that does not target any mouse genes to provide a relative invasion value.

Spheroid invasion. Cells (10,000-18,000/well) were transfected in 96-well plates (Corning, 3904). After 24 h, the transfected cells were trypsinized and clustered in 96-well Nunclon Sphera U-bottom low adhesion plates (Thermo Scientific, 174925) overnight (1000 cells/well). Homogeneous and mosaic spheroids were also generated by plating 1000 cells/well in U-bottom Nunclon Sphera low adhesion plates. Cells then aggregated into spheroids for least 16 h before embedding in ECM. Spheroids were resuspended in 30 µl Collagen I/BM mix (2.4 mg/ml Collagen I and 2 mg/ml Matrigel or Cultrex), plated on 20 µl of a base layer of Matrigel/Collagen I, overlaid with media containing 1% FBS and 2 ng/ml TGFβ1 and allowed to invade for 24 h-48 h unless stated otherwise. The spheroids were fixed and stained with Hoechst and Phalloidin and imaged with a 10x/0.45 objective (Zeiss, 1 420640-9900-000) using a Zeiss LSM800 laser scanning confocal microscope. To determine the area of invasion, Hoechst staining was used to identify and mask nuclei invading into the ECM using ImageJ. The total area of the invading nuclei was then determined using the “Measure” function in ImageJ. For circularity analysis, whole well imaging was performed and TIFFs were extracted and loaded onto ImageJ. Using the Phalloidin or Hoechst stain-

ing (which highlight the organoid shape and protrusions), masked images were generated and circularity was quantified using the "Analyze Particle" function.

Live cell imaging. Imaging was performed using a Zeiss LSM800 laser scanning confocal microscope enclosed in a 37˚C chamber supplemented with humidified CO_2. Images were acquired every 30 min for 14-18 h with a 10X/0.30 objective (Zeiss, 1 420340-9901-000). At least 15 different x,y coordinates with 5 to 7 z-slices over 50-100 µm span for each condition were acquired in parallel.

Immunofluorescence and immunoblotting. Experiments and analysis were performed as described (12) using antibodies detailed in **Table S1.** Immunoblots were imaged using an Odyssey scanner (Licor, 9120). Immunofluorescence images were acquired using 10x/0.45 (Zeiss, 1 420640-9900-000) and 20x/0.8 (Zeiss, 1 420650-9902-000) objectives. For analysis of the relationship between K14 expression and invasion, organoids were classified as invasive if they had at least two protrusions with each protrusion having at least two cells invading outward from the center mass. Images were exported as TIFFs and analyzed with ImageJ. The freehand selection tool was used to define ROIs containing individual organoids for analysis. The K14 fluorescence for each organoid was then defined as the "Mean Gray Value", as determined using the "Measure" function. To define K14 expression in cells leading invasion, >10 organoids stained with Phalloidin, Hoechst, K14 and K8 were analyzed. Seven Z slices at 10 µm intervals over a total span of 100 µm were acquired, and the leading cells of each protrusion were designated as K14- or K14+. Vimentin expression was analyzed in 5 µm histology sections. Organoids in 8-well chamber slides (Falcon, 354108) were washed 3x with distilled water. After detaching the chamber walls from the slides, the organoid/ECM gels were gently picked up using a single edge blade (Personna, 94-120-71) and laid out on a strip of parafilm (Bemis, PM-999). The 8-well chamber walls were placed around the gels and 150-200 µl of hydroxyethyl agarose processing gel (Thermo Fisher, HG 4000-012) was poured into the well. After the hydroxyethyl agarose processing gel solidified, it was transferred into histosettes (Simport, M498-3) and stored in 70% EtOH until embedding in paraffin by the Georgetown Histology and Tissue Shared Resource. The 5 µm sections were deparaffinized and immunostained as described (11) to detect

Vimentin and E-cadherin expression (**Table S1)**. Using ImageJ, Vimentin expression was defined as the "Mean Gray Value" for individual organoids using the "Measure" function.

Immunohistochemistry. Sample processing and staining was performed as described (12). Images were acquired using 10x/0.45 (Zeiss, 1 420640-9900-000) and 20x/0.8 (Zeiss, 1 420650-9902-000) objectives. At least 10 regions of interest (ROIs) were randomly selected for each tumor. Cell Profiler (Broad Institute) was used to define image masks based on K8, K14 or Vimentin signal. To define their expression in tumor cells, masks based on the area of overlap between the K14 and K8 and the Vimentin and K8 signals were determined for each ROI. The area of individual channel and overlap masks created in CellProfiler were then loaded onto ImageJ and quantified using the "Measure" function.

Quantitative real-time PCR. Total RNA was isolated using the RNeasy Kit (Qiagen, 74104) or the PureLink™ RNA Mini Kit (Invitrogen, 12183025). Reverse transcription was performed using iScript cDNA Synthesis Kit (Bio-Rad, 1708891) following the manufacturer's protocol. Real time PCR was performed with iQ SYBR Green Supermix (Bio-Rad, 1708882) using a RealPlex 2 thermocycler (Eppendorf, 2894). All primers were designed and purchased from Integrated DNA Technologies and their sequences are listed in **Table S1**.

Flow Cytometry. Cells were dissociated with TrypLe (Gibco, 12605-010) into single cell suspensions, and then washed 2x with FACs buffer (PBS + 2% FBS). Cells were then incubated with anti-CD326 (EpCAM)-PE/Dazzle 594 (Biolegend, 118236) at 4°C for 30 minutes. After staining, cells were washed with FACs buffer 2x, and stained with Helix NP Blue (Biolegend, 425305) in order to exclude dead cells. Cells were then subjected to flow cytometry analysis using BD LSR Fortessa. Results were analyzed using FCSExpress 7 (De Novo Software).

RNA sequencing. For the analysis of C3-TAg tumor organoids and noninvasive spheroids in **Fig. 2 and S2**, at least 1000 C3-TAg tumor organoids or 60,000-80,000 noninvasive cells were embedded in a 500 µl ECM mixture (2.4 mg/ml Collagen I and 2 mg/ml Matrigel or Cultrex) and plated on top of 500 uL base

layer of ECM in 6 cm dishes (Falcon, 353004). C3-TAg tumor organoids were allowed to invade for 48 h. Noninvasive cells were allowed to grow and form spheroids for 6 days. Experiments were performed twice using frozen stocks of tumors and cell lines. For the analysis of C3-TAg organoid responses to growth factors and inhibitors in **Fig. 3, 6 and S6**, 250,000 1339-org organoids were embedded in a 100 µl ECM mixture (2.4 mg/ml Collagen I and 2 mg/ml Matrigel or Cultrex) of and plated on top of a 200 µl ECM base layer in 6 cm dishes (Falcon, 353004). The gels were overlaid with control or treatment media. Organoid/ECM gels were collected in 15 mL conical tubes (Fisher, 339651), snap frozen and stored in -80^0 C freezer. Experiments were performed twice. For RNA isolation in all experiments, the gels were thawed on ice and disintegrated in 1 mL Trizol (Invitrogen, 15596026) by repeatedly passing the mixture through a 23G needle (BD, 305143). After the gels were homogenized, the Trizol-gel mixture was allowed to sit at RT for 5 min. Then 200 uL of chloroform (Sigma-Aldrich, C2432-25ML) was added. Following vigorous shaking for 15 sec, the mix was left for 2 min at RT to allow for phase separation. The mix was then centrifuged at 12,000 x g for 15 m at 4^0 C. The upper aqueous phase was then transferred to a new tube and RNA was then further purified using the RNeasy Kit (Qiagen, 74104) and or PureLink™ RNA Mini Kit (Invitrogen, 12183025). For **Fig. 2 and S2**, RNA sequencing, was performed by the Genomics and Epigenomics Shared Resource at Georgetown University. The cDNA libraries were generated using the Illumina TruSeq Stranded Total RNA Library Preparation Kit (Illumina, 20020597). Sequencing was performed using an Illumina NextSeq 500 at a depth of 12 million paired-end 75 base pair reads per sample. Data is available from GEO at GSE188312. Reads were quantified using Salmon with mouse reference genome (GRCm38/mm10). Counts were normalized and differential expression was calculated using DESeq2. For **Fig. 3, 6 and S6**, RNA sequencing was performed by Novogene. RNA quality analysis was assessed by Bioanalyzer using the RNA Nano 6000 Assay Kit (Agilent, 5067-1511). The cDNA libraries were prepared and indexed using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB, E7530L). Libraries were sequenced with the NovaSeq 6000 at a depth of 20 million paired-end 150 base pair reads per sample. Sequencing data available from GEO at GSE188300 and GSE188323. Reads were mapped to the mouse reference genome (GRCm38/mm10) using HISAT2 and reads per gene determined using HTSeq. Counts were normalized and differential expression was calculated using DESeq2.

Polar plot analysis. Expression data was normalized and differentially expressed genes with a ≥2-fold change in expression with $p < 0.05$ after Benjamini-Hochberg (BH) correction were identified using the *DESeq2* package in R (15,16). Normalized expression data was collapsed for all genes as the average between condition replicates and were subset to only include genes that were differentially expressed between at least one of the three conditions (control, growth factor, growth factor plus inhibitor) assessed. To define the integration between the pathways, we modified a previously described approach for generating polar plots that has been used to identify dynamic trends in gene expression (17,18). Relative mean signal changes over successive conditions, of control, growth factor, and growth factor plus inhibitor treatment (T^0, T^1, and T^2, respectively), were transformed to an angle between -180 and +180 degrees (θ). This angle was further transformed to a scale from 0 to 360 degrees and then interpolated into a color based on the angle. Differentially expressed genes were plotted using polar coordinates (r,θ) with the radius (r), or distance from the focus, being equal to the value of largest absolute $\log_2$ fold change between conditions and theta (θ) being equal to the angle corresponding to the trend of change and calculated using the following equations:

$$I = \max(T^0, T^1, T^2)$$

$$R = 1 - \frac{\min(T^0, T^1, T^2)}{I}$$

$$\theta = \left(\frac{T^0 + T^1 - T^2 - I}{I \times R} + 2 \right) \times sign\left(T^1 - T^0\right) \times 60$$

The intensity (I) is equivalent to the maximum expression for a gene across conditions and the ratio (R) between the minimum expression of a gene across conditions over the intensity, are used to calculate the angle of expression trend (θ). As such, genes are plotted as a point with an angle (θ) corresponding to relative changes in a gene's expression with subsequent treatments at a distance from the focus cor-

responding to the largest magnitude of change between any pair of conditions in the treatment series. Three-point profile annotations of expression change (corresponding from left to right as T^0, T^1, and T^2, respectively) are indicated on the outer edge of the polar scatter plots every 30 degrees in the color assigned to that particular θ. These profiles are accompanied by a radial histogram denoting the density of differentially changed genes across organoid conditions at particular θ of the polar scatterplot.

To summarize the meaning of genes with particular profiles in the polar scatterplots, genes were grouped by profiles at 30-degree intervals starting at 0 degrees, which denotes genes that were unchanged in expression from T^0 to T^1 but demonstrated a significant increase in expression following growth factor plus inhibitor treatment of the organoids (T^2). Genes from each group were assessed for enrichment of Gene Ontology biological processes (GP) using the EnrichGo function of the *clusterProfiler* package in R (19). Processes with $p < 0.01$ after BH correction are included in **Tables S4, 5**. Values were transformed ($-\log_{10}$) and displayed alongside polar scatterplots. Genes within each profile group of the polar scatterplot and utilized for GO enrichment were scaled and displayed as a line plot. A dotted line was added to these linear plots to denote the mean scaled expression pattern across organoid treatment conditions for all genes within a particular interval.

Breast Cancer Tissue Microarray. The construction of the TNBC tissue microarray (TMA) containing duplicate cores from 50 patients was performed by the Histopathology and Tissue Shared Resource (HTSR) at Georgetown University. Details of the methodology, patient consent, de-identification and demographics have been previously described (20). Multiplex staining of deparaffinized and rehydrated cores was performed through sequential rounds of staining, stripping and re-probing with antibodies recognizing K14, Vimentin and E-cadherin and OPAL fluorescent secondary antibodies **(Table S1)**. The TMA cores were imaged with a Vectra3 Multi-Spectral Imaging microscope (Perkin Elmer) and analyzed with inForm software (Perkin Elmer). Tumor and stroma were segmented based on E-cadherin expression (tumor). Cores with folded sections or no clear tumor-stomal interface by manual inspection were excluded from analysis. Cores were classified “high invasion” if invasive strands were detected in >50% of the tumor-

stomal interface by manual inspection. The K14 and Vimentin signals were not visible when classifying invasion status.

Statistical Methods. Data was analyzed with Prism 9.0.2 (Graphpad). Data with a normal distribution, as determined by Shapiro-Wilk test, was analyzed by two tailed Student's t-test with. Data that did not pass a normality test was analyzed by Mann-Whitney U test.

Data Availability. The RNA sequencing data generated in this study are publicly available in Gene Expression Omnibus (GEO) at GSE188300, GSE188312 and GSE188323. All other data are available within the article and its supplementary data files.

RESULTS

Mammary tumor subtypes have distinct trailblazer states that lead collective invasion.

To define factors that promote a trailblazer state, we investigated the invasive properties of tumor organoids isolated from the C3(1)/SV40 T-antigen (C3-TAg) GEMM of breast cancer (21) and **(Table S1)**. C3-TAg tumors are most similar to basal-type triple-negative breast cancer (TNBC) tumors, which lack estrogen receptor, progesterone receptor and HER2 expression (22,23). C3-TAg organoid invasion was heterogeneous within and between tumors **(Fig. 1A, Video S1)**. Spheroids formed by dissociated cells were also relatively less invasive than the larger organoids from which they were derived **(Fig. S1A,B, Video S2)**. This indicated that individual C3-TAg organoids were comprised of cells with different invasive properties. In addition, clonal trailblazer (D6 and C7) and noninvasive "opportunist" (B6 and G2) cells were also derived from a single C3-TAg tumor **(Fig. S1C-E, Table S1)**, consistent with the heterogeneity detected in tumor organoids and our prior analysis of breast cancer cell lines (11). Notably, C3-TAg tumor cells and trailblazer D6 cells led the invasion of opportunist B6 cells when clustered together in heterogeneous spheroids **(Fig. 1B, S1E, Videos S3-6)**. These combined results indicated that there were trailblazer subpopulations in the C3-TAg tumors with an enhanced ability to lead collective invasion.

The activation of a gene expression program exemplified by K14 expression promotes a trailblazer state that is required for collective invasion in PyMT GEMM mammary tumors (24,25). Based on these results, K14 has been suggested as a marker of collectively invading breast cancer cells (24,26). K14 was expressed in a greater proportion of cells in C3-TAg tumors compared to PyMT tumors **(Fig. 1C)**. However, K14 expression did not correlate with the C3-TAg invasion in our analysis of primary tumors, organoids or trailblazer clones **(Fig 1C-E, S1C,D)**. In addition K14 depletion only modestly reduced the invasion of C3-TAg derived 1863 cells **(Fig. S1F, Table S1)**. C3-TAg organoids were relatively more invasive than PyMT organoids, demonstrating that the K14 independent invasion of C3-TAg cells was more aggressive than K14 dependent invasion of PyMT cells **(Fig. 1F, S1G, Videos S7-8)**. K14 expression also did not correlate with invasion in TNBC tumor sections, indicating that the C3-TAg invasive traits reflected

the properties of TNBC tumors **(Fig. 1G, S1H)**. Together, our results indicated that there were tumor subtype specific trailblazer states with distinct functional requirements and invasive properties **(Fig. 1H)**.

Mammary tumor subtype specific EMT traits distinguish the properties of trailblazer cells.

To understand how EMT promotes a trailblazer state independent of K14 expression, we analyzed the mRNA content of organoids from 4 C3-TAg tumors and the 3 C3-TAg opportunist spheroid populations **(Fig. 2A, S2A)**. Gene set enrichment analysis (GSEA) showed that genes highly expressed in C3-TAg organoids were associated with biological processes that contribute to invasion (ECM Receptor Interaction, Focal Adhesions and Regulation of the Actin Cytoskeleton) **(Fig. S2B and Table S2)**. The expression of Hallmark EMT attributes and canonical mesenchymal genes were also elevated in the C3-TAg organoids (**Fig. 2B-D, Table S2)**, consistent with EMT inducing a trailblazer state. Epithelial gene expression was retained in the C3-TAg organoids, indicating that cells were in a hybrid EMT state **(Fig. 2E)**. Among mammary epithelial keratins, K14 was specifically suppressed in the C3-TAg tumors, consistent with our earlier analysis (**Fig. 2F)**. The EMT inducing transcription factors (EMT-TFs) Zeb1, Zeb2 and Fra1 were more highly expressed in C3-TAg organoids **(Fig. 2G)**. However, the EMT-TFs Snail, Slug and Twist1 were expressed at similar or higher levels in the noninvasive clones **(Fig. 2G)**. These results indicated that the specific features of EMT states determined the invasive phenotypes of C3-TAg cells, rather than simply the presence or absence of EMT activity. The inability of EMT to confer a trailblazer state in the opportunist clones was potentially due to the expression of the transcription factor ΔNp63 **(Fig. S2C)**, which suppresses trailblazer protrusion formation, Zeb1 and 2 expression, and is only detected in cells that opportunistically invade in C3-TAg tumors and breast cancer cell lines (12,27,28). Vimentin expression was more frequently detected in C3-TAg tumors compared to PyMT tumors, indicating that EMT functioned differently in each tumor type **(Fig. 2H, S2D)**. The inter-and intra-tumor heterogeneity in Vimentin expression was also consistent with the invasive heterogeneity of C3-TAg organoids **(Fig. 1A)**. In addition, Vimentin expression was heterogeneous in TNBC tumors, with a trend towards more extensive Vimentin expression in regions of invasion **(Fig. 2I, S2E)**. Our combined data indicated that the differences in the trailblazer states detected in the PyMT and C3-TAg tumors were caused by distinctions in EMT program characteristics **(Fig. 2J)**.

TGFβ induces multiple trailblazer EMT states through distinct phases of re-programming.

We next investigated how EMT was regulated in C3-TAg tumors. GSEA indicated that TGFβ signaling was more active in C3-TAg organoids than the noninvasive clones **(Fig. S3A, Table S2)**. Thus, we evaluated the contribution of TGFβ towards C3-TAg EMT and invasion. The Tgfbr1 inhibitor A83-01 suppressed tumor organoid invasion **(Fig. 3A)**. Organoid lines cultured in the presence of A83-01 for at least 4 weeks were also less invasive than organoids tested directly after isolation from primary tumors **(Fig. 3B, Table S1)**. In addition, 1863 cells derived in the presence of A83-01 were less invasive than the 1863T cells derived from the same tumor in the absence of A83-01 **(Fig. S3B, Table S1)**. A83-01 and a second Tgfbr1 inhibitor, SB431542, suppressed the intrinsic invasion of C3-TAg organoid lines, further indicating a requirement of TGFβ signaling for collective invasion **(Fig. S3C-E, Videos S9,10)**. Exogenous TGFβ1 was sufficient to enhance organoid invasion **(Fig. S3C-E, Videos S9-11)**. In addition, tumors that developed from 1339-org cells were more invasive than the 1339-org cells themselves, consistent with trailblazer ability being conferred by TGFβ from the tumor microenvironment **(Fig. S3F)**. Depletion of the transcription factor SMAD3 reduced 1863 and 1863T cell invasion **(Fig. S3G-I)**, indicating that TGFβ regulated gene expression contributed to the observed phenotypes.

Persistent exposure of 1339-org cells to exogenous TGFβ for up to 22 days progressively increased the rate of collective invasion, indicating that TGFβ induced multiple different trailblazer states **(Fig. 3C, Videos S9-12)**. Single cell invasion was also induced **(Fig. 3C, Videos S9-12)**. Indeed, hierarchical clustering revealed 4 different groups of genes regulated by TGFβ, with different timing in their changes in expression **(Fig. 3D, Table S3).** Gene ontology analysis revealed that these groups were associated with distinct combinations of biological processes **(Fig. 3D, Table S3)**. Group 1 gene suppression was detected 22 days after TGFβ treatment and was not associated with processes related to invasion. Instead, Group 1 genes were connected with immune responses and viral defense, demonstrating the extent of re-programming induced by TGFβ **(Fig. 3D, Table S3)**. Group 2 genes were suppressed as early as 2 days after TGFβ treatment and were associated with promoting cell-cell adhesion and cell differentiation, consistent with a loss of epithelial character promoting invasion **(Fig. 3D, Table S3)**. K14 and other basal ep-

ithelial keratins were part of Group 2, further indicating that invasion was not reliant upon K14 **(Fig. 3D,E, Table S3)**. However, key epithelial genes were unchanged or suppressed in smaller populations **(Fig. 3E, Table S3)**, consistent with our prior analysis showing C3-TAg tumor trailblazer cells were in a hybrid state **(Fig. 2E)**. Group 3 genes were closer to maximal expression by 7 days whereas the Group 4 genes showed the greatest degree of increase between d7 and d22 **(Fig. 3D)**. Group 3 and 4 genes were both associated with the regulation of ECM, indicating that there were different mechanisms for regulating genes associated with the same biological functions in response to TGFβ **(Fig. 3D, Table S3)**. Group 3 and Group 4 genes were also associated with unique functions, further indicating that the nature of reprogramming changed over time and was controlled by distinct mechanisms. For example, Group 3 genes were associated with calcium sensing and Group 4 genes were classified with additional processes that modulated the microenvironment and developmental programs of other cell lineages, indicating that increased lineage plasticity was a later stage event in response to TGFβ **(Fig. 3D, Table S3)**. Notably, "trailblazer" genes that were previously shown to confer the ability to lead collective invasion were included in Group 4 **(Fig. 3D,F, Table S3)**. Interestingly, genes commonly associated with EMT were included in Group 3 (Vcam1, Itgb3, Fn1) and Group 4 (Vimentin, N-cadherin, Periostin) **(Table S3)**. EMT-TFs were also split between Group 3 (Zeb1), Group 4 (Zeb2, Twist), or not included in any group (Snail) **(Fig. 3E)**. The distinct expression profile of these EMT-TFs and EMT effectors further indicated that TGFβ induced EMT involved cells adopting multiple different trailblazer states in response to evolving regulatory mechanisms.

To understand TGFβ induced molecular changes in individual cells, we immunostained organoids treated with TGFβ. We found that TGFβ induced a shift in the distribution of Vimentin expression in organoids at day 7 and day 22, as opposed to a switch between two distinct Vimentin-low and Vimentin-high states **(Fig. 3G)**. Similarly, TGFβ induced a shift in the distribution of EpCAM expression rather than a switch between two unique EpCAM-high and EpCAM-low states **(Fig. S3J)**. The Vimentin expressing cells predominantly retained E-cadherin expression, with a subset of cells showing E-cadherin suppression, further indicating that TGFβ induced multiple different cell states **(Fig. S3K)**. Together, these results indicated that there were multiple different trailblazer states with distinct invasive abilities and patterns of

gene expression induced by TGFβ, not in a single distinct trailblazer state with limited phenotypic variability **(Fig. 3H)**. The ability of cells to adopt multiple different trailblazer states was also consistent with the invasive heterogeneity detected in C3-TAg organoids **(Fig. 1A).**

The transcription factors Zeb1, Zeb2 and Fra1 confer a trailblazer EMT state.

We next investigated how EMT-TF expression influenced the trailblazer state. To prioritize our investigation, we determined how siRNAs targeting EMT-TFs influenced the invasion of 1863T cells. The siRNAs targeting Fra1, Zeb1 and Zeb2 suppressed invasion to the greatest extent whereas siRNAs targeting Twist, Slug and Snail failed to suppress invasion **(Fig. 4A, S4A)**. Zeb1 depletion also suppressed TGFβ stimulated 1863 invasion, whereas transfection with Zeb2 siRNAs had a negligible effect (**Fig. 4B,C)**. The requirement of Zeb1, but not Zeb2, for the invasion of TGFβ stimulated 1863 cells was consistent with our results showing that Zeb1 expression rapidly increased within 7 days while increased Zeb2 expression occurred after day 7 of extrinsic TGFβ exposure **(Fig. 3E)**. A miR200a-3p mimic, which suppressed Zeb1 and 2 expression, also inhibited 1863 and 1863T invasion **(Fig. 4A, S4B-D)**. Further investigation of Fra1, showed that Fra1 depletion reduced the invasion of 1863T and TGFβ stimulated 1863 and 1339 spheroids, as well as D6 trailblazer cell invasion **(Fig. 4D,E, S4D-I)**. Exogenous Fra1 expression promotes Zeb1 and Zeb2 expression (29). However, depletion of Fra1 did not alter Zeb1 or 2 expression, indicating that Fra1 regulated the trailblazer EMT state independent of controlling Zeb1 and 2 expression **(Fig. 4F)**. Fra1 interacts with Jun to regulate gene expression as an AP-1 complex (30). Indeed, Jun was required for 1863 and 1863T cell invasion, suggesting Fra1 containing AP-1 complexes regulate the expression of genes necessary for the induction of a trailblazer state **(Fig. 4G,H, S4J)**. Zeb1 and 2 expression were regulated by autocrine TGFβ signaling in C3-TAg organoids, along with being stimulated by exogenous TGFβ (**Fig. 3E, 4I)**. However, Fra1 expression was not influenced by autocrine TGFβ signaling and Fra1 was not included in Group 3 or 4 genes detected as part of a TGFβ induced gene regulatory program **(Fig. 4I and Table S3)**. Thus, the trailblazer state required the activity of at least one signaling pathway acting in parallel with TGFβ signaling to induce the expression of Fra1 **(Fig. 4J)**.

Egfr pathway activity promotes Fra1 expression and a trailblazer EMT state.

We next determined how Fra1 expression was regulated in C3-TAg tumor cells. Fra1 mRNA was more highly expressed in C3-TAg tumor organoids compared to the noninvasive clones, indicating that the signaling pathway responsible for the control of Fra1 was more active in the C3-TAg organoids **(Fig. 2G)**. GSEA identified enrichment of multiple active KRAS regulated expression signatures in the C3-TAg organoids **(Fig. 2B, S5A,B and Table S2)**. The Kras gene can be amplified, but is not mutated in C3-TAg tumors (31), indicating that pathway activation occurred upstream of Kras. Analysis of the genes in the active KRAS signature showed that the Egfr ligand Hbegf was more highly expressed in C3-TAg organoids than opportunist clones **(Fig. 5A, Table S2)**. Further inspection found that the Egfr ligands Areg and Epgn were expressed at a higher level in C3-TAg organoids as well **(Fig. 5A)**. These combined results prompted us to investigate how the Egfr-Ras-Mek1/2-Erk1/2 signaling pathway contributed to Fra1 expression and the induction of a trailblazer state. Erlotinib (Egfr inhibitor) and Trametinib (Mek1/2 inhibitor) suppressed Fra1 mRNA and protein expression in C3-TAg cell lines and organoids **(Fig. 5B,C, S5C)**. By comparison, Jun expression was unchanged by treatment with either inhibitor **(Fig. S5D)**. In addition, Erlotinib and Trametinib did not inhibit the activating phosphorylation of Smad2 or Zeb1 expression, indicating that Fra1 did not influence cell state through autocrine TGFβ signaling **(Fig. 5C, S5E)**. A83-01 suppressed Smad2 phosphorylation, but did not influence Fra1 expression and phosphorylation **(Fig. 5C)**. Erlotinib and Trametinib reduced the invasion of TGFβ stimulated 1339-org organoids **(Fig. 5D,E)**. EGF also enhanced the invasion of 1339-org, 1863-org and 1788-org organoids and this enhancement was blocked by A83-01, further indicating that the integration of Tgfbr1 and Egfr pathways promoted a trailblazer state **(Fig. 5F, S5F,G)**. Trametinib inhibited the invasion of C3-TAg primary tumor organoids and phosphorylated Erk1/2 was detected in collectively invading cells in C3-TAg primary tumors **(Fig. 5G,H)**. Erk1/2 activity is pulsatile in response to EGF stimulation, and these pulses promote Fra1 expression (32). Thus, persistent Erk1/2 phosphorylation was not necessarily expected in all collectively invading cells or to promote Fra1 expression. Together, these results indicate that the regulation of Fra1 by Egfr and Erk1/2 promotes the induction of a trailblazer EMT **(Fig. 5I)**.

A compromise between Tgbr1 and Egfr regulated signaling programs confers the trailblazer state.

To understand the contribution of the Tgfbr1 and Egfr regulated signaling pathways towards the induction of the trailblazer state, we performed RNA-seq on organoids treated with diluent, TGFβ , TGFβ +Erlotinib, EGF and EGF+A83-01 **(Fig. 6A, S6A)**. Polar plots were used to cluster genes based on the nature and the degree of integration between the Tgfbr1 and Egfr signaling pathways (17). This analysis revealed multiple modes of interaction between the two pathways in response to either TGFβ or EGF and that the type of interaction (positive or negative) was associated with different biological processes **(Fig. 6B,C, S6B-E, Tables S4,5)**. Thus, the relationship was far more complex than the simple positive signal integration previously suggested by phenotypic analysis and the expression of specific subsets of EMT-TFs and canonical EMT program genes (33,34). In addition, the gene expression changes induced by extrinsic TGFβ and EGF were associated with unique combinations of biological processes, indicating that TGFβ and EGF induced distinct trailblazer states **(Fig. 6C, S6F, Tables S4,5)**. Nevertheless, there were conserved features of pathway integration that influenced the regulation of processes associated with invasion regardless of the external stimulus. Genes involved in controlling the organization and composition of the ECM, including previously defined trailblazer genes, were predominantly induced by Tgfbr1 and not positively impacted by Egfr signaling **(Fig. 6C,D, S6F,G, Tables S4,5).** Notably, autocrine and extrinsic Egfr activation suppressed a number of Tgfbr1 regulated ECM reorganization genes and Hallmark EMT genes, including Postn, Bmp2, Itgb3, Orai2, Mmp11 and Erg **(Fig. 6C,D, S6F,G Tables S4,5)**, which contribute to invasion and metastasis (35-40). By comparison, autocrine or extrinsic EGF activity alone, or in combination with Tgfbr1 signaling, induced the expression of genes that promote cell motility **(Fig. 6C,E, S6F,H, Tables S4,5)**. Thus, a compromise between the Tgfbr1 and Egfr signaling programs was necessary to induce a set of pro-motility genes. To determine how the Egfr dependent motility genes influenced the induction of the trailblazer state, we tested the function of the receptor tyrosine kinase Axl. Tgbr1 and Egfr signaling was required to promote Axl expression **(Fig. 6F)**, consistent with Axl expression being directly regulated by Fra1 (41). Axl also promotes cell migration in other contexts (42). The Axl inhibitor BGB324 suppressed C3-TAg organoid invasion, demonstrating that a gene positively regulated by combined Tgfbr1 and Egfr activity was necessary for trailblazer invasion **(Fig. 6G)**. Together, these results suggested that the induction of the trailblazer state involved a compromise between the Tgfbr1 and Egfr

regulated signaling programs that limited the maximal extent of ECM reorganization to promote cell motility that was essential for trailblazer cell invasion **(6H)**.

DISCUSSION.

Our findings indicate that defining how cells invade requires an understanding of how EMT functions differently in distinct subtypes of breast tumors. **(Fig. 7)**. PyMT tumors contain small subpopulations of Snail and K14 expressing cells that lead tumor organoid invasion and are found in cells at the tumor-stromal interface (24,43). In contrast, C3-TAg collective invasion was induced by a trailblazer state characterized by hallmark EMT attributes and the expression of genes specifically required for trailblazer TNBC cells to initiate collective invasion. Importantly, K14 expression was dispensable for C3-TAg collective invasion and K14 low populations led collective invasion in C3-TAg tumors and patient primary tumors, consistent with K14-low cells leading collective invasion in breast cancer TNBC cell lines (11). The prominent K14 expression in the C3-TAg tumors likely reflected the early acquisition of a mixed luminal-basal state that occurs during tumor initiation rather than specifying cells with an enhanced ability to collectively invade (44). The distinction between trailblazer cells in PyMT and C3-TAg tumors may reflect different mechanisms for inducing EMT and collective invasion. TGFβ induced C3-TAg invasion whereas p63 has been reported to induce the K14 expression that is necessary for PyMT invasion (24). C3-TAg tumors do contain p63 expressing cells (12). However p63 expressing cells were relatively less invasive than p63-low trailblazer cells in primary tumors in our prior analysis (12) and in our investigation here. Moreover, TGFβ suppressed K14 and p63 expression in C3-TAg tumors, in other breast cancer models and in squamous tumors (6,45). These results, and the lack of Vimentin expression in PyMT tumors here and in previous studies (46), suggests that either TGFβ promotes collective invasion through tumor subtype specific EMT mechanisms or that the cues for conferring EMT and trailblazer activity are distinct between subtypes. Notably, subtype specific mechanisms for inducing EMT and a trailblazer state are potentially a general property of cancer progression, as molecular subtypes are a feature of carcinomas (47-50). Thus, further investigation is required to determine the extent to which distinctions in EMT mechanisms and trailblazer states align with existing tumor subtypes or define new categories of inter-tumor heterogeneity.

The heterogeneity in C3-TAg organoid invasion indicated that the trailblazer population itself could be heterogeneous within tumors with the same primary driving genetic factor and molecular sub-

type **(Fig. 7)**, as opposed to a strictly single uniform state within tumor types, as suggested by the analysis of other models (11,24,51). Notably, our analysis of C3-TAg organoids revealed that this invasive variability could be a consequence of differences in cellular responses to a single stimulus, in this case extrinsic TGFβ, and therefore did not necessarily reflect EMT program differences related to different activating signals, as may by the case with PyMT tumors and other contexts (1). Gene expression changes in response to TGFβ are heterogeneous and temporally regulated in other systems as well (52). However, the influence of this heterogeneity on cell phenotype has not been clear, and has often been modeled as a progressive accumulation of new functions (53). The phased changes in gene expression we defined in C3-TAg organoids suggested how trailblazer phenotypes could evolve over time in response to TGFβ activity, with early onset gene expression changes inducing trailblazer states that were further modulated by later stage reprogramming events. Interestingly, the late stage expression changes included genes associated with other cell lineages, suggesting that enhanced trailblazer states were associated with more general lineage plasticity and not simply a re-balancing of epithelial and mesenchymal gene expression programs. Together, our results revealed that EMT state heterogeneity could induce differences in the properties of cells with the same function, in this case trailblazer cells leading invasion. Thus, our investigation of the regulation of trailblazer states has revealed new concepts related to EMT and suggests further investigation of the heterogeneity induced by EMT at the single-cell level to understand the contribution of EMT to biological traits.

Our results revealed how the balance of activity in the Tgfbr1 and Egfr regulated pathways influenced the induction of a trailblazer state and provided new insight into how cells are conferred with the ability to lead invasion. The integration of parallel signaling pathways is frequently required to confer specific cell states and is typically classified from the standpoint of being a positive or negative interaction (34,54,55). We found that the combined activity of the Tgfbr1 and Egfr signaling pathways was necessary to induce a trailblazer state, yet there were both positive and negative elements in the interactions between the signaling programs. The integration of Tgbr1 and Egfr signaling was essential to maximize Axl expression. Additional genes regulated by combined Tgfbr1 and Egfr activity, including Sema7a, Ncam1 and Itga2, have also been implicated in the control of cell motility (56-58), further indicating the

dependence of the integration of the two pathways on the control of genes that confer a trailblazer state. However, TGFβ signaling alone also regulated the expression of trailblazer genes, ECM reorganization genes, EMT hallmark genes and EMT-TF expression. In addition, extrinsic EGF stimulation and autocrine Egfr activity limited the expression of canonical EMT attributes, such as Postn expression. Thus, key elements of EMT and the conversion to a trailblazer state were not dependent on the integration of Tgbr1 and Egfr signaling, with some attributes potentially suppressed by Egfr activity. Therefore, our results revealed how the induction of a trailblazer state reflected a compromise between distinct signaling programs regulated by different activating signals. Notably, the compromise mode of integration between the Tgfbr1 and Egfr pathways is distinct from prior studies suggesting that TGFβ and Ras regulated pathways positively integrate at the point of EMT-TF expression to initiate EMT programs (33,34,59,60). Because their integration was not based on coordinating the activation of a single program in C3-TAg cells, the nature of the interaction between the Tgfbr1 and Egfr pathways also promoted heterogeneity in trailblazer states. Extrinsic TGFβ predominantly promoted ECM reorganization while extrinsic EGF induced genes involved in migration. Thus, a trailblazer state could be induced by amplifying different traits that are required to initiate collective invasion, with the other process functioning in an essential complementary capacity **(Fig. 7)**.

Our results provide new insight into approaches used to stratify patients into outcome groups and challenges used when treating patients with targeted inhibitors. Genes induced as part of the basal-type trailblazer EMT program defined in C3-TAg cells are more frequently detected at a high level in TNBC patients. Increased basal trailblazer EMT gene expression also correlates with reduced odds of survival in triple negative breast cancer patients, but not patients classified as ER-positive or HER2-positive. Our findings indicate that the unique expression and correlation with patient outcome of this gene set reflects a basal subtype specific trailblazer EMT, as opposed to the presence or absence of a trailblazer EMT in tumors. Similarly, the expression of K14 that specifies luminal B trailblazer cells is unlikely to identify basal type trailblazer cells, leaving the basis of any prognostic value related to K14 expression in basal-type tumors unclear. Notably, the heterogeneity of the trailblazer EMT states detected in the C3-TAg tumor populations indicates that improving on prognostic measures related to trailblazer cells will

likely require detection modalities with single-cell resolution and a precise understanding of how these different trailblazer states influence metastasis. The heterogeneity of trailblazer populations also creates challenges in defining treatments to ablate trailblazer populations or assessing the impact of treatments designed to kill tumor cells. Egfr and Mek1/2 inhibitors have been investigated as breast cancer treatments and are approved for treating certain types of other cancers based on their ability to suppress tumor growth (61-63). TGFβ inhibition has been tested as a means to augment immunotherapy (64). Our results suggest how partially inhibiting either pathway can shift, as opposed to ablate, the trailblazer cell phenotype. Thus, understanding the risk associated with altering the balance of activity between TGFβ and EGFR in future investigations is needed to inform the best implementation of inhibitors of either pathway. Together, our results indicate that an improved understanding of how EMT is regulated and functions in different breast cancer subtypes has the potential to improve the accuracy of patient diagnosis and reveal new treatment options.

DISCLOSURE OF POTENTIAL CONFLICTS OF INTEREST

There are no potential conflicts of interest.

AUTHORS' CONTRIBUTIONS

A. Nasir: Conceptualization, investigation, writing-review and editing. **S. Camacho:** Conceptualization, investigation, writing-review and editing. **A.T. McIntosh:** Investigation, writing-review and editing. **G. Graham:** Investigation, writing-review and editing. **R. Rahhal:** Investigation, writing-review and editing. **M.E. Huysman:** Investigation. **A.T. Riegel:** Resources, funding acquisition. **G.W. Pearson:** Conceptualization, investigation, writing-review and editing, funding acquisition.

ACKNOWLEDGEMENTS

Work was supported by NIH R01CA218670, Georgetown Women and Wine (to G.W. Pearson), NIH T32-CA009686 (to S. Camacho, G. Graham) and NIH P30CA051008.

REFERENCES

1. Katsuno Y, Derynck R. Epithelial plasticity, epithelial-mesenchymal transition, and the TGF-β family. Developmental Cell **2021**;56:726-46
2. Nieto MA, Huang Ruby Y-J, Jackson Rebecca A, Thiery Jean P. EMT: 2016. Cell **2016**;166:21-45
3. Brabletz S, Schuhwerk H, Brabletz T, Stemmler MP. Dynamic EMT: a multi-tool for tumor progression. Embo j **2021**;40:e108647
4. Nieto MA. Epithelial plasticity: a common theme in embryonic and cancer cells. Science **2013**;342:1234850
5. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M*, et al.* Guidelines and definitions for research on epithelial-mesenchymal transition. Nat Rev Mol Cell Biol **2020**;21:341-52
6. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S*, et al.* Identification of the tumour transition states occurring during EMT. Nature **2018**;556:463-8
7. Stemmler MP, Eccles RL, Brabletz S, Brabletz T. Non-redundant functions of EMT transcription factors. Nat Cell Biol **2019**;21:102-12
8. Cheung KJ, Padmanaban V, Silvestri V, Schipper K, Cohen JD, Fairchild AN*, et al.* Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. Proc Natl Acad Sci U S A **2016**;113:E854-63
9. Pearson GW. Control of Invasion by Epithelial-to-Mesenchymal Transition Programs during Metastasis. J Clin Med **2019**;8
10. Vilchez Mercedes SA, Bocci F, Levine H, Onuchic JN, Jolly MK, Wong PK. Decoding leader cells in collective cancer invasion. Nat Rev Cancer **2021**;21:592-604
11. Westcott JM, Prechtl AM, Maine EA, Dang TT, Esparza MA, Sun H*, et al.* An epigenetically distinct breast cancer cell subpopulation promotes collective invasion. J Clin Invest **2015**;125:1927-43
12. Westcott JM, Camacho S, Nasir A, Huysman ME, Rahhal R, Dang TT*, et al.* ΔNp63-Regulated Epithelial-to-Mesenchymal Transition State Heterogeneity Confers a Leader–Follower Relationship That Drives Collective Invasion. Cancer Research **2020**;80:3933-44
13. Kanda T, Sullivan KF, Wahl GM. Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. Curr Biol **1998**;8:377-85
14. Dang TT, Prechtl AM, Pearson GW. Breast cancer subtype-specific interactions with the microenvironment dictate mechanisms of invasion. Cancer Res **2011**;71:6857-66
15. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological) **1995**;57:289-300
16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology **2014**;15:550
17. Siersbæk R, Madsen JGS, Javierre BM, Nielsen R, Bagge EK, Cairns J*, et al.* Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. Molecular Cell **2017**;66:420-35.e5
18. Rauch A, Haakonsson AK, Madsen JGS, Larsen M, Forss I, Madsen MR*, et al.* Osteogenesis depends on commissioning of a network of stem cell transcription factors that act as repressors of adipogenesis. Nature Genetics **2019**;51:716-27
19. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS: A Journal of Integrative Biology **2012**;16:284-7
20. Fernandez AI, Geng X, Chaldekas K, Harris B, Duttargi A, Berry VL*, et al.* The orphan nuclear receptor estrogen-related receptor beta (ERRβ) in triple-negative breast cancer. Breast Cancer Research and Treatment **2020**;179:585-604
21. Green JE, Shibata MA, Yoshidome K, Liu ML, Jorcyk C, Anver MR*, et al.* The C3(1)/SV40 T-antigen transgenic mouse model of mammary cancer: ductal epithelial cell targeting with multistage progression to carcinoma. Oncogene **2000**;19:1020-7

22. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z*, et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biol **2007**;8:R76
23. Pfefferle AD, Herschkowitz JI, Usary J, Harrell JC, Spike BT, Adams JR*, et al.* Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. Genome Biol **2013**;14:R125
24. Cheung KJ, Gabrielson E, Werb Z, Ewald AJ. Collective invasion in breast cancer requires a conserved basal epithelial program. Cell **2013**;155:1639-51
25. Hwang PY, Brenot A, King AC, Longmore GD, George SC. Randomly Distributed K14(+) Breast Tumor Cells Polarize to the Leading Edge and Guide Collective Migration in Response to Chemical and Mechanical Environmental Cues. Cancer Res **2019**;79:1899-912
26. Padmanaban V, Tsehay Y, Cheung KJ, Ewald AJ, Bader JS. Between-tumor and within-tumor heterogeneity in invasive potential. PLoS Comput Biol **2020**;16:e1007464
27. Tucci P, Agostini M, Grespi F, Markert EK, Terrinoni A, Vousden KH*, et al.* Loss of p63 and its microRNA-205 target results in enhanced cell migration and metastasis in prostate cancer. Proceedings of the National Academy of Sciences **2012**;109:15312
28. Dang TT, Esparza MA, Maine EA, Westcott JM, Pearson GW. DeltaNp63alpha Promotes Breast Cancer Cell Motility through the Selective Activation of Components of the Epithelial-to-Mesenchymal Transition Program. Cancer Res **2015**;75:3925-35
29. Bakiri L, Macho-Maschler S, Custic I, Niemiec J, Guío-Carrión A, Hasenfuss SC*, et al.* Fra-1/AP-1 induces EMT in mammary epithelial cells by modulating Zeb1/2 and TGFβ expression. Cell Death Differ **2015**;22:336-50
30. Eferl R, Wagner EF. AP-1: a double-edged sword in tumorigenesis. Nature Reviews Cancer **2003**;3:859-68
31. Liu ML, Von Lintig FC, Liyanage M, Shibata MA, Jorcyk CL, Ried T*, et al.* Amplification of Ki-ras and elevation of MAP kinase activity during mammary tumor progression in C3(1)/SV40 Tag transgenic mice. Oncogene **1998**;17:2403-11
32. Albeck John G, Mills Gordon B, Brugge Joan S. Frequency-Modulated Pulses of ERK Activity Transmit Quantitative Proliferation Signals. Molecular Cell **2013**;49:249-61
33. Oft M, Peli J, Rudaz C, Schwarz H, Beug H, Reichmann E. TGF-beta1 and Ha-Ras collaborate in modulating the phenotypic plasticity and invasiveness of epithelial tumor cells. Genes Dev **1996**;10:2462-77
34. Derynck R, Muthusamy BP, Saeteurn KY. Signaling pathway cooperation in TGF-β-induced epithelial–mesenchymal transition. Current Opinion in Cell Biology **2014**;31:56-66
35. Parvani JG, Galliher-Beckley AJ, Schiemann BJ, Schiemann WP. Targeted inactivation of β1 integrin induces β3 integrin switching, which drives breast cancer metastasis by TGF-β. Mol Biol Cell **2013**;24:3449-59
36. Blee AM, Huang H. ERG-Mediated Cell Invasion: A Link between Development and Tumorigenesis. Medical Epigenetics **2015**;3:19-29
37. Huang P, Chen A, He W, Li Z, Zhang G, Liu Z*, et al.* BMP-2 induces EMT and breast cancer stemness through Rb and CD44. Cell Death Discovery **2017**;3:17039
38. González-González L, Alonso J. Periostin: A Matricellular Protein With Multiple Functions in Cancer Development and Progression. Front Oncol **2018**;8:225
39. Ma B, Ran R, Liao H-Y, Zhang H-H. The paradoxical role of matrix metalloproteinase-11 in cancer. Biomedicine & Pharmacotherapy **2021**;141:111899
40. Wu S, Chen M, Huang J, Zhang F, Lv Z, Jia Y*, et al.* ORAI2 Promotes Gastric Cancer Tumorigenicity and Metastasis through PI3K/Akt Signaling and MAPK-Dependent Focal Adhesion Disassembly. Cancer Research **2021**;81:986

41. Sayan AE, Stanford R, Vickery R, Grigorenko E, Diesch J, Kulbicki K*, et al.* Fra-1 controls motility of bladder cancer cells via transcriptional upregulation of the receptor tyrosine kinase AXL. Oncogene **2012**;31:1493-503
42. Holland SJ, Powell MJ, Franci C, Chan EW, Friera AM, Atchison RE*, et al.* Multiple roles for the receptor tyrosine kinase axl in tumor formation. Cancer Res **2005**;65:9294-303
43. Ye X, Tam WL, Shibue T, Kaygusuz Y, Reinhardt F, Ng Eaton E*, et al.* Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. Nature **2015**;525:256-60
44. Christin JR, Wang C, Chung CY, Liu Y, Dravis C, Tang W*, et al.* Stem Cell Determinant SOX9 Promotes Lineage Plasticity and Progression in Basal-like Breast Cancer. Cell Rep **2020**;31:107742
45. Bui NHB, Napoli M, Davis AJ, Abbas HA, Rajapakshe K, Coarfa C*, et al.* Spatiotemporal Regulation of ΔNp63 by TGFβ-Regulated miRNAs Is Essential for Cancer Metastasis. Cancer Res **2020**;80:2833-47
46. Li Y, Lv Z, Zhang S, Wang Z, He L, Tang M*, et al.* Genetic Fate Mapping of Transient Cell Fate Reveals N-Cadherin Activity and Function in Tumor Metastasis. Developmental Cell **2020**;54:593-607.e5
47. Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer. Cell **2015**;163:1011-25
48. Chen F, Zhang Y, Parra E, Rodriguez J, Behrens C, Akbani R*, et al.* Multiplatform-based molecular subtypes of non-small-cell lung cancer. Oncogene **2017**;36:1384-93
49. McConkey DJ, Choi W. Molecular Subtypes of Bladder Cancer. Current Oncology Reports **2018**;20:77
50. Collisson EA, Bailey P, Chang DK, Biankin AV. Molecular subtypes of pancreatic cancer. Nature Reviews Gastroenterology & Hepatology **2019**;16:207-20
51. Konen J, Summerbell E, Dwivedi B, Galior K, Hou Y, Rusnak L*, et al.* Image-guided genomics of phenotypically heterogeneous populations reveals vascular signalling during symbiotic collective cancer invasion. Nat Commun **2017**;8:15078
52. van Dijk D, Sharma R, Nainys J, Yim K, Kathail P, Carr AJ*, et al.* Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell **2018**;174:716-29.e27
53. Lambert AW, Weinberg RA. Linking EMT programmes to normal and neoplastic epithelial stem cells. Nat Rev Cancer **2021**;21:325-38
54. Chakrabarti R, Hwang J, Andres Blanco M, Wei Y, Lukačišin M, Romano RA*, et al.* Elf5 inhibits the epithelial-mesenchymal transition in mammary gland development and breast cancer metastasis by transcriptionally repressing Snail2. Nat Cell Biol **2012**;14:1212-22
55. Lehmann W, Mossmann D, Kleemann J, Mock K, Meisinger C, Brummer T*, et al.* ZEB1 turns into a transcriptional activator by interacting with YAP1 in aggressive cancer types. Nature Communications **2016**;7:10498
56. Zecchini S, Bombardelli L, Decio A, Bianchi M, Mazzarol G, Sanguineti F*, et al.* The adhesion molecule NCAM promotes ovarian cancer progression via FGFR signalling. EMBO Molecular Medicine **2011**;3:480-94
57. Black SA, Nelson AC, Gurule NJ, Futscher BW, Lyons TR. Semaphorin 7a exerts pleiotropic effects to promote breast tumor progression. Oncogene **2016**;35:5170-8
58. Adorno-Cruz V, Hoffmann AD, Liu X, Dashzeveg NK, Taftaf R, Wray B*, et al.* ITGA2 promotes expression of ACLY and CCND1 in enhancing breast cancer stemness and metastasis. Genes & Diseases **2021**;8:493-508
59. Doehn U, Hauge C, Frank SR, Jensen CJ, Duda K, Nielsen JV*, et al.* RSK Is a Principal Effector of the RAS-ERK Pathway for Eliciting a Coordinate Promotile/Invasive Gene Program and Phenotype in Epithelial Cells. Molecular Cell **2009**;35:511-22
60. Shin S, Dimitri CA, Yoon S-O, Dowdle W, Blenis J. ERK2 but Not ERK1 Induces Epithelial-to-Mesenchymal Transformation via DEF Motif-Dependent Signaling Events. Molecular Cell **2010**;38:114-27
61. Roberts PJ, Usary JE, Darr DB, Dillon PM, Pfefferle AD, Whittle MC*, et al.* Combined PI3K/mTOR and MEK inhibition provides broad antitumor activity in faithful murine cancer models. Clin Cancer Res **2012**;18:5290-303

62. Usary J, Zhao W, Darr D, Roberts PJ, Liu M, Balletta L*, et al.* Predicting drug responsiveness in human cancers using genetically engineered mice. Clin Cancer Res **2013**;19:4889-99
63. Reguart N, Remon J. Common EGFR-mutated subgroups (Del19/L858R) in advanced non-small-cell lung cancer: chasing better outcomes with tyrosine kinase inhibitors. Future Oncology **2015**;11:1245-57
64. Li S, Liu M, Do MH, Chou C, Stamatiades EG, Nixon BG*, et al.* Cancer immunotherapy via targeted TGF-β signalling blockade in TH cells. Nature **2020**;587:121-5

Figure 1

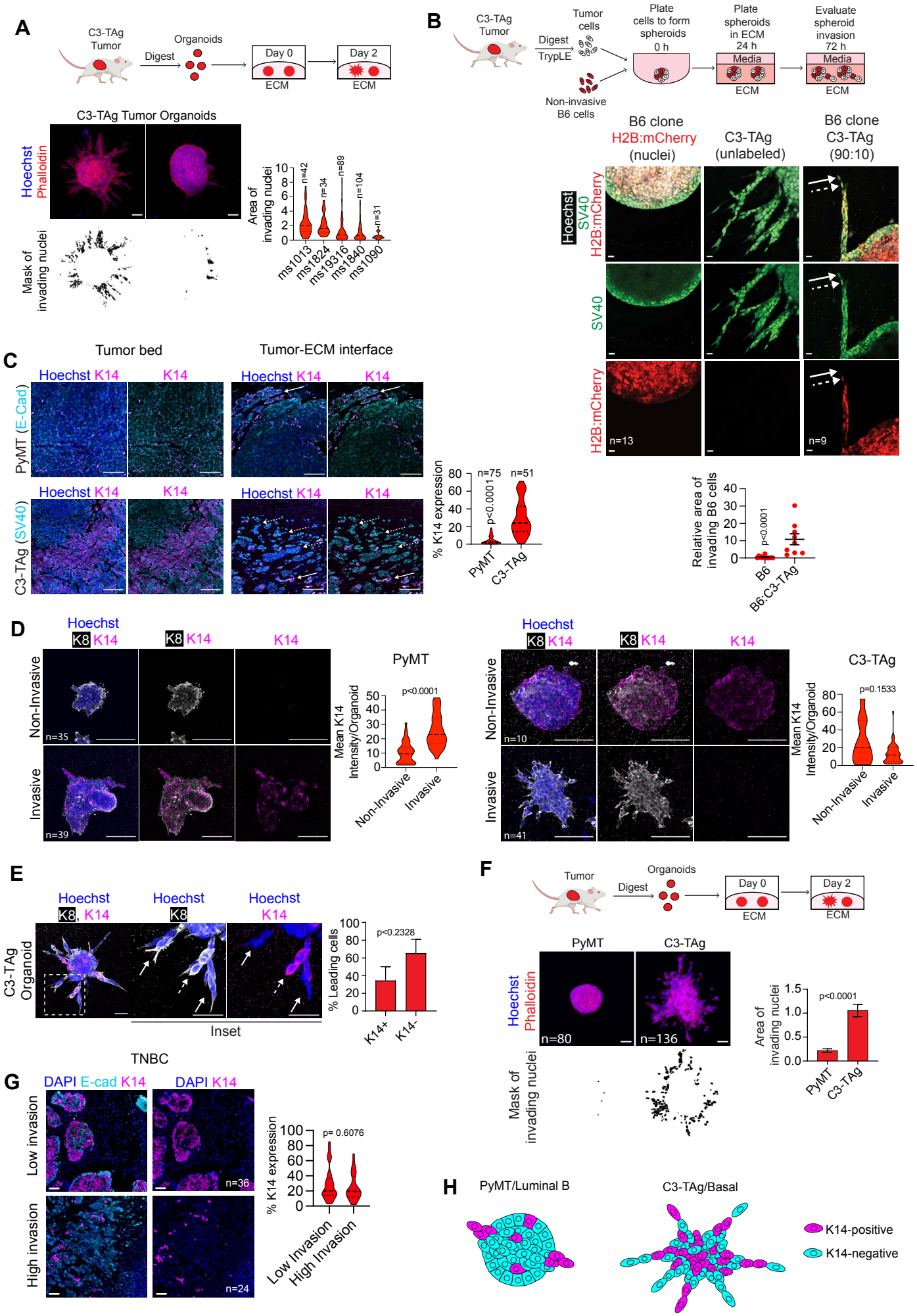
A
C3-TAg Tumor
Digest
Organoids
Day 0
Day 2
ECM
C3-TAg Tumor Organoids
Hoechst
Phalloidin
Mask of invading nuclei
Area of invading nuclei
n=42
n=34
n=89
n=104
n=31
ms1013
ms1824
ms19316
ms1840
ms1090
B
C3-TAg Tumor
Digest
TrypLE
Tumor cells
Non-invasive B6 cells
Plate cells to form spheroids
0 h
Plate spheroids in ECM
24 h
Media
Evaluate spheroid invasion
72 h
B6 clone H2B:mCherry (nuclei)
C3-TAg (unlabeled)
B6 clone C3-TAg (90:10)
Hoechst
SV40
H2B:mCherry
n=13
n=9
Relative area of invading B6 cells
p<0.0001
B6
B6:C3-TAg
C
Tumor bed
Tumor-ECM interface
Hoechst K14
K14
PyMT (E-Cad)
C3-TAg (SV40)
% K14 expression
n=75
n=51
p<0.0001
PyMT
C3-TAg
D
Hoechst
K8 K14
K14
Non-Invasive
Invasive
n=35
n=39
n=10
n=41
PyMT
C3-TAg
Mean K14 Intensity/Organoid
p<0.0001
p=0.1533
E
Hoechst
K8, K14
K8
C3-TAg Organoid
Inset
% Leading cells
p<0.2328
K14+
K14-
F
Tumor
Digest
Organoids
Day 0
Day 2
ECM
PyMT
C3-TAg
Hoechst
Phalloidin
n=80
n=136
Mask of invading nuclei
Area of invading nuclei
p<0.0001
G
TNBC
DAPI E-cad K14
DAPI K14
Low invasion
High invasion
n=36
n=24
% K14 expression
p= 0.6076
Low Invasion
High Invasion
H
PyMT/Luminal B
C3-TAg/Basal
K14-positive
K14-negative

Figure 1. Mammary tumor subtypes have distinct trailblazer states that lead collective invasion.
A. There is intra- and inter-tumor invasive heterogeneity in C3-TAg organoids. Graphic shows the process for the derivation and analysis of tumor organoids. Representative images show the invasive heterogeneity of C3-TAg organoids derived from a single tumor. Image mask defines invasive cell nuclei. Violin plot shows the quantification of organoid invasion from 5 C3-TAg tumors. (Mann Whitney test, n= organoids). Scale bars, 50 µm. **B.** C3-TAg tumors contain trailblazer cells that can lead the collective invasion of intrinsically less invasive cells. Graphic shows the process for generating homogeneous and heterogeneous spheroids. Images show spheroids immunostained with anti-SV40 antibody (detects TAg in all cells), counterstained with Hoechst (nuclei) and H2B:mCherry fluorescence (B6 cells only). C3-TAg tumor cells (unlabeled) promoted the invasion of B6 cells (H2B:mCherry). Solid arrow indicates a C3-TAg tumor cell leading invasion. Dashed arrow indicates a B6 cell invading behind the C3-TAg cell. Graph shows the relative area of invading B6 nuclei from 3 independent experiments (mean±SEM, Mann Whitney test, n= spheroids,). Scale bars, 20 µm. **C.** K14 expression does not correlate with C3-TAg tumor invasion. K14 SV40 (C3-TAg) and E-Cad (PyMT) expression in the tumor core and tumor-ECM interface of C3-TAg and PyMT tumors. Solid and dashed arrows indicate K14 high and low cells at the leading front of collective invasion respectively. Violin plot shows total area of K14 expressing tumor cells. (Mann Whitney Test, n=ROIs). Scale bars, 100 µm. **D.** K14 expression does not correlate with C3-TAg organoid invasion. Violin plot shows mean K14 fluorescence in invasive and noninvasive PyMT and C3-TAg organoids. (Mann Whitney Test, n=organoids). Scale bars, 100 µm. **E.** K14 expression does not correlate with the ability to lead collective invasion in C3-TAg organoids. Bar graph shows keratin expression in leading cells of organoids from 3 tumors (mean±SEM, unpaired Student's t test, n=organoids) Scale bars, 50 µm. **F.** C3-TAg organoids were more invasive than PyMT organoids. Bar graph shows the area of invading nuclei (mean±SEM, Mann Whitney test, n=spheroids, r=2) Scale bars, 50 µm. **G.** K14 expression in TNBC patient tumors did not correlate with invasion. Violin plot shows % K14 positive cells in regions of low and high invasion (Mann Whitney test, n=tumor sections). Scale bars, 50 µm. **H.** Model depicting the tumor subtype relationship between K14 expression and trailblazer cell activity.

Figure 2

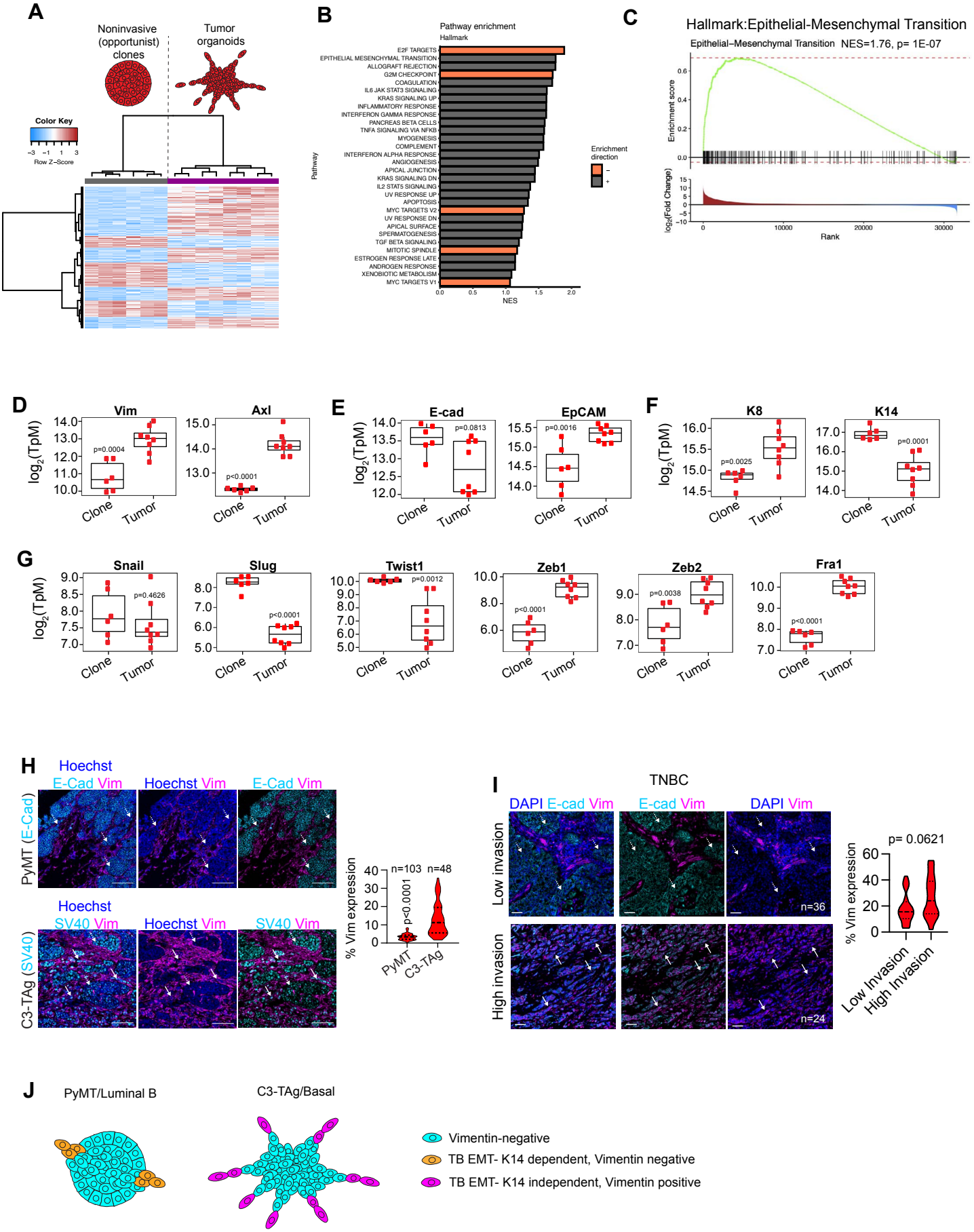
A
Noninvasive (opportunist) clones
Tumor organoids
Color Key
Row Z-Score
B
Pathway enrichment
Hallmark
Pathway
NES
Enrichment direction
C
Hallmark:Epithelial-Mesenchymal Transition
Epithelial–Mesenchymal Transition NES=1.76, p= 1E-07
Enrichment score
log2(Fold Change)
Rank
D
Vim
Axl
E
E-cad
EpCAM
F
K8
K14
G
Snail
Slug
Twist1
Zeb1
Zeb2
Fra1
log2(TpM)
Clone
Tumor
H
Hoechst
E-Cad Vim
Hoechst Vim
PyMT (E-Cad)
SV40 Vim
C3-TAg (SV40)
% Vim expression
n=103
n=48
p<0.0001
PyMT
C3-TAg
I
TNBC
DAPI E-cad Vim
E-cad Vim
DAPI Vim
Low invasion
High invasion
n=36
n=24
p= 0.0621
Low Invasion
High Invasion
J
PyMT/Luminal B
C3-TAg/Basal
Vimentin-negative
TB EMT- K14 dependent, Vimentin negative
TB EMT- K14 independent, Vimentin positive

Figure 2. Mammary tumor subtype specific EMT traits distinguish the properties of trailblazer cells.
A. Unsupervised clustering and differentially expressed genes in non-invasive C3-TAg clones (B6, C6, E8) and organoids from 4 C3-TAg tumors. **B-C.** There was increased expression of Hallmark EMT genes in C3-TAg tumor organoids compared to noninvasive C3-TAg clones. GSEA was performed on RNA-seq data from C3-TAg organoids and noninvasive C3-TAg clones to identify pathway enrichment in C3-TAg organoids. Graph shows Hallmark pathways with NES ≥1. Plot shows the details of the Hallmark EMT GSEA. **D-G.** Boxplots showing the expression of canonical mesenchymal (Vimentin, Axl), epithelial (E-cadherin, EpCAM), luminal (K8), basal (K14), and EMT-TF (Snai1, Slug, Twist1, Zeb1, Zeb2, Fra1) genes in C3-TAg organoids compared to noninvasive clones from RNA-seq data (Mann Whitney test for E-cad, Student's t-test for all others, n=6 for clones, (3 biological replicates); n=8 for tumor (4 biological replicates)). **H.** Vimentin is more highly expressed in C3-TAg primary tumors compared to PyMT tumors. Solid arrows indicate collectively invading Vimentin positive tumor cells. Dashed arrows indicate vimentin negative tumor cells. Violin plot shows total area of Vimentin expressing tumor cells (Mann Whitney Test, n=ROIs from 5 tumors). Scale bars, 100 µm. **I.** Invasive regions in TNBC patient tumors trend towards higher Vimentin expression compared to regions of low invasion. Solid arrows indicate collectively invading Vimentin positive tumor cells. Dashed arrows indicate vimentin negative tumor cells. Violin plot shows total area of Vimentin expressing tumor cells (Mann Whitney test, n=tumor sections). Scale bars, 50 µm. **J.** Model showing the tumor subtype relationship between an EMT exemplified by Vimentin expression and trailblazer activity (TB=trailblazer).

Figure 3

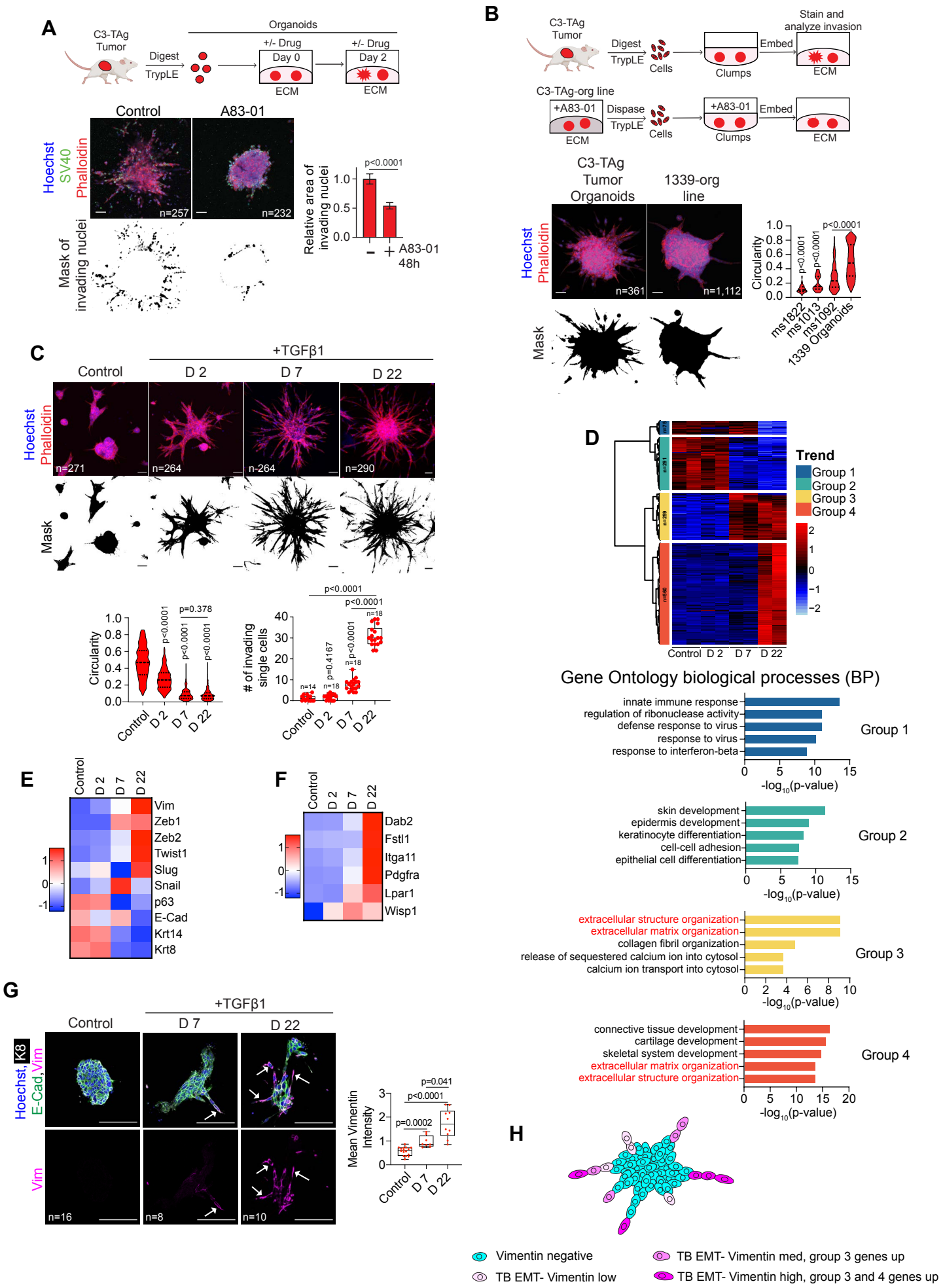
A
C3-TAg Tumor
Digest TrypLE
Organoids
+/- Drug
Day 0
Day 2
ECM
Control
A83-01
Hoechst
SV40
Phalloidin
n=257
n=232
Mask of invading nuclei
Relative area of invading nuclei
p<0.0001
A83-01 48h
B
C3-TAg Tumor
Digest TrypLE
Cells
Clumps
Embed
Stain and analyze invasion
C3-TAg-org line
+A83-01
Dispase TrypLE
C3-TAg Tumor Organoids
1339-org line
n=361
n=1,112
Mask
Circularity
ms1822
ms1013
ms1092
1339 Organoids
C
+TGFβ1
Control
D 2
D 7
D 22
n=271
n=264
n-264
n=290
Circularity
p=0.378
p<0.0001
# of invading single cells
p=0.4167
n=14
n=18
D
Trend
Group 1
Group 2
Group 3
Group 4
Gene Ontology biological processes (BP)
innate immune response
regulation of ribonuclease activity
defense response to virus
response to virus
response to interferon-beta
skin development
epidermis development
keratinocyte differentiation
cell-cell adhesion
epithelial cell differentiation
extracellular structure organization
extracellular matrix organization
collagen fibril organization
release of sequestered calcium ion into cytosol
calcium ion transport into cytosol
connective tissue development
cartilage development
skeletal system development
-log10(p-value)
E
Vim
Zeb1
Zeb2
Twist1
Slug
Snail
p63
E-Cad
Krt14
Krt8
F
Dab2
Fstl1
Itga11
Pdgfra
Lpar1
Wisp1
G
Hoechst, K8
E-Cad, Vim
Vim
n=16
n=8
n=10
Mean Vimentin Intensity
p=0.041
p=0.0002
H
Vimentin negative
TB EMT- Vimentin low
TB EMT- Vimentin med, group 3 genes up
TB EMT- Vimentin high, group 3 and 4 genes up

Figure 3. TGFβ induces multiple trailblazer EMT states through distinct phases of re-programming.
A. The Tgfbr1 inhibitor A83-01 suppresses C3-TAg organoid invasion. Graph shows the quantification of invasion (mean±SEM, Mann Whitney test, n=organoids). Scale bars, 50µm. **B.** C3-TAg organoid lines cultured in A83-01 are less invasive than organoids tested immediately after isolation from primary tumors. Graph shows organoid circularity. Reduced circularity indicates increased invasion (mean±SEM, Mann Whitney test, n=organoids). Scale bars, 50 µm. **C.** Persistent exposure to exogenous TGFβ1 progressively enhances the invasive phenotype in C3-TAg organoid lines. Graphs show the quantification of organoid circularity and the number of invading single cells (Mann Whitney test, n=organoids, r=3) Scale bars, 50 µm. **D.** TGFβ1 induces 4 distinct patterns of gene expression changes in 1339-org organoids. Heatmap shows unsupervised clustering and 4 groups of DEGs. Graphs show the top 5 Gene Ontology terms associated with each group. **E.** Heatmap shows changes in expression of mesenchymal (Vim), basal (K14), luminal (K8), epithelial (E-Cad, p63) markers and EMT-TFs (Zeb1-2, Twist1, Snail, Slug) upon TGFβ1 exposure. **F.** Heatmap shows persistent exposure to TGFβ1 increases expression of trailblazer genes previously reported in human cell lines. **G.** TGFβ1 progressively increases Vimentin expression in 1339-org line organoids. Arrows indicate representative Vimentin expressing cells leading invasion. Graph shows vimentin expression in organoids (Student's t-test, n=images analyzed). Scale bars, 100 µm. **H.** Model showing the heterogeneity in trailblazer states with unique invasive traits induced by TGFβ (TB=trailblazer).

Figure 4

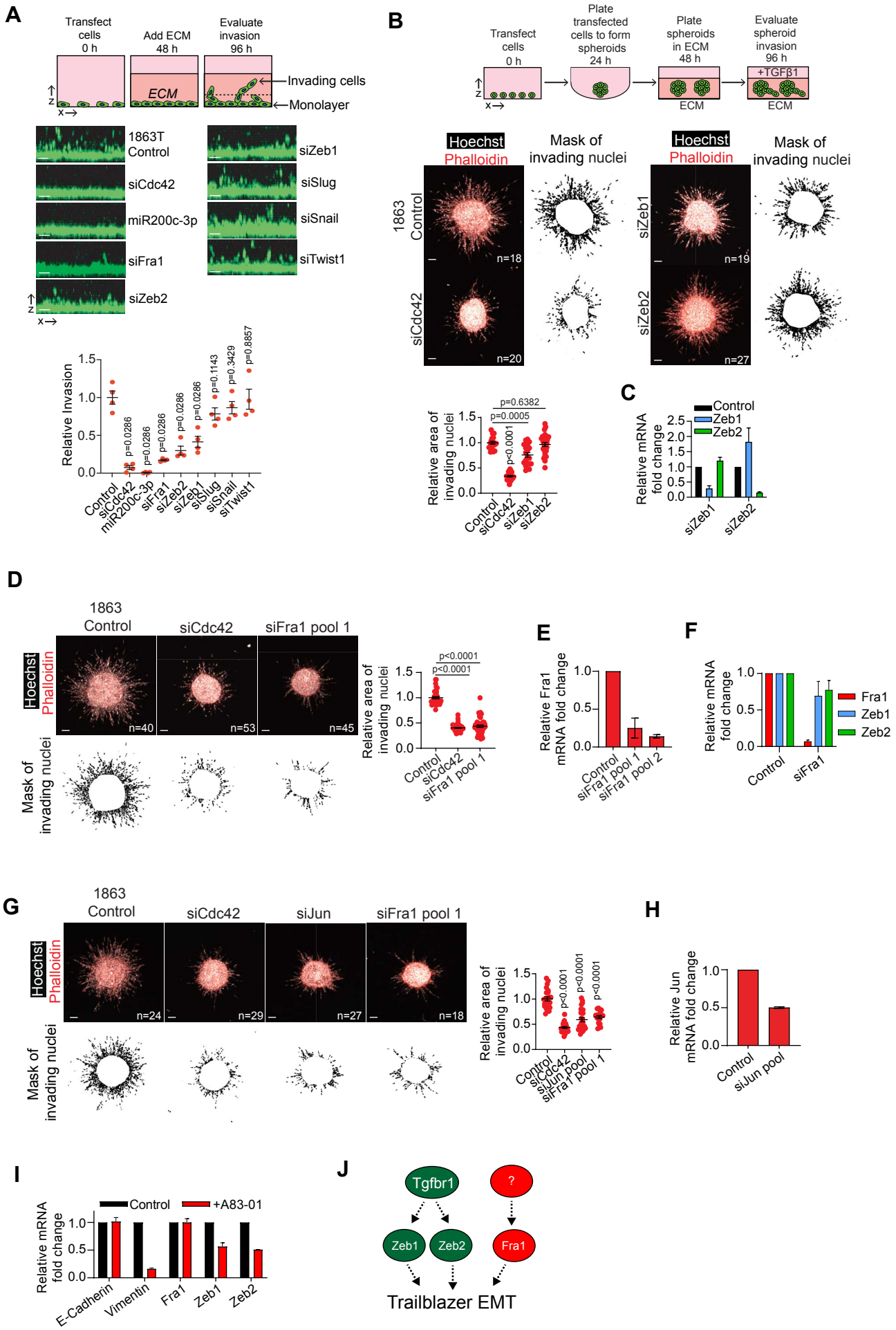
A
Transfect cells 0 h
Add ECM 48 h
Evaluate invasion 96 h
ECM
Invading cells
Monolayer
1863T Control
siCdc42
miR200c-3p
siFra1
siZeb2
siZeb1
siSlug
siSnail
siTwist1
Relative Invasion
p=0.0286
p=0.0286
p=0.0286
p=0.0286
p=0.1143
p=0.3429
p=0.8857
Control
siCdc42
miR200c-3p
siFra1
siZeb2
siZeb1
siSlug
siSnail
siTwist1
B
Transfect cells 0 h
Plate transfected cells to form spheroids 24 h
Plate spheroids in ECM 48 h
Evaluate spheroid invasion 96 h
+TGFβ1
ECM
Hoechst
Phalloidin
Mask of invading nuclei
1863 Control
siCdc42
siZeb1
siZeb2
n=18
n=20
n=19
n=27
Relative area of invading nuclei
p=0.6382
p=0.0005
p<0.0001
Control
siCdc42
siZeb1
siZeb2
C
Relative mRNA fold change
Control
Zeb1
Zeb2
siZeb1
siZeb2
D
1863 Control
siCdc42
siFra1 pool 1
Hoechst
Phalloidin
Mask of invading nuclei
n=40
n=53
n=45
Relative area of invading nuclei
p<0.0001
p<0.0001
Control
siCdc42
siFra1 pool 1
E
Relative Fra1 mRNA fold change
Control
siFra1 pool 1
siFra1 pool 2
F
Relative mRNA fold change
Fra1
Zeb1
Zeb2
Control
siFra1
G
1863 Control
siCdc42
siJun
siFra1 pool 1
Hoechst
Phalloidin
Mask of invading nuclei
n=24
n=29
n=27
n=18
Relative area of invading nuclei
p<0.0001
p<0.0001
p<0.0001
Control
siCdc42
siJun pool
siFra1 pool 1
H
Relative Jun mRNA fold change
Control
siJun pool
I
Control
+A83-01
Relative mRNA fold change
E-Cadherin
Vimentin
Fra1
Zeb1
Zeb2
J
Tgfbr1
?
Zeb1
Zeb2
Fra1
Trailblazer EMT

Figure 4. The transcription factors Zeb1, Zeb2 and Fra1 confer a trailblazer EMT state.
A. Transfection with siRNAs targeting Zeb1, Zeb2 and Fra1 suppresses invasion of 1863T cells into the ECM. Graph shows the quantification of cells invading ≥40 µm into the ECM (mean±SEM, Mann Whitney test, n=4 wells from 2 independent experiments). Scale bars, 50 µm **B.** Zeb1 depletion suppresses TGFβ1 induced 1863 spheroid invasion (mean±SEM, Mann Whitney test, n=spheroids, r=2). Scale bars, 50µm **C.** Graphs showing the depletion of Zeb1 and Zeb2 by siRNAs (mean±SEM, n=2) **D.** Fra1 depletion decreases TGFβ1 induced 1863 spheroid invasion (mean±SEM, Mann Whitney test n=spheroids, r=5) Scale bars, 50 µm. **E.** Graph shows the depletion of Fra1 expression by two different pools of Fra1 siRNA (mean±SEM, n=3 for pool 1 and n=2 for pool 2). **F.** Graph shows that Fra1 depletion does not suppress Zeb1 and Zeb2 expression (mean±SEM, n=2) **G.** Depletion of cJun suppresses TGFβ1 induced 1863 spheroid invasion (mean±SEM, Mann Whitney test, n=spheroids, r=2) Scale bars, 50 µm. **H.** Graph showing the depletion of Jun by siRNAs (mean±SEM, n=2). **I.** E-cadherin, Vimentin, Fra1, Zeb1 and Zeb2 expression after 48h treatment with Tgfbr1 inhibitor A83-01 (mean±SEM, n=2). **J.** Model showing the parallel regulation of the trailblazer EMT by the TGFβ dependent expression of Zeb1 and 2 and second pathway that induces Fra1.

Figure 5

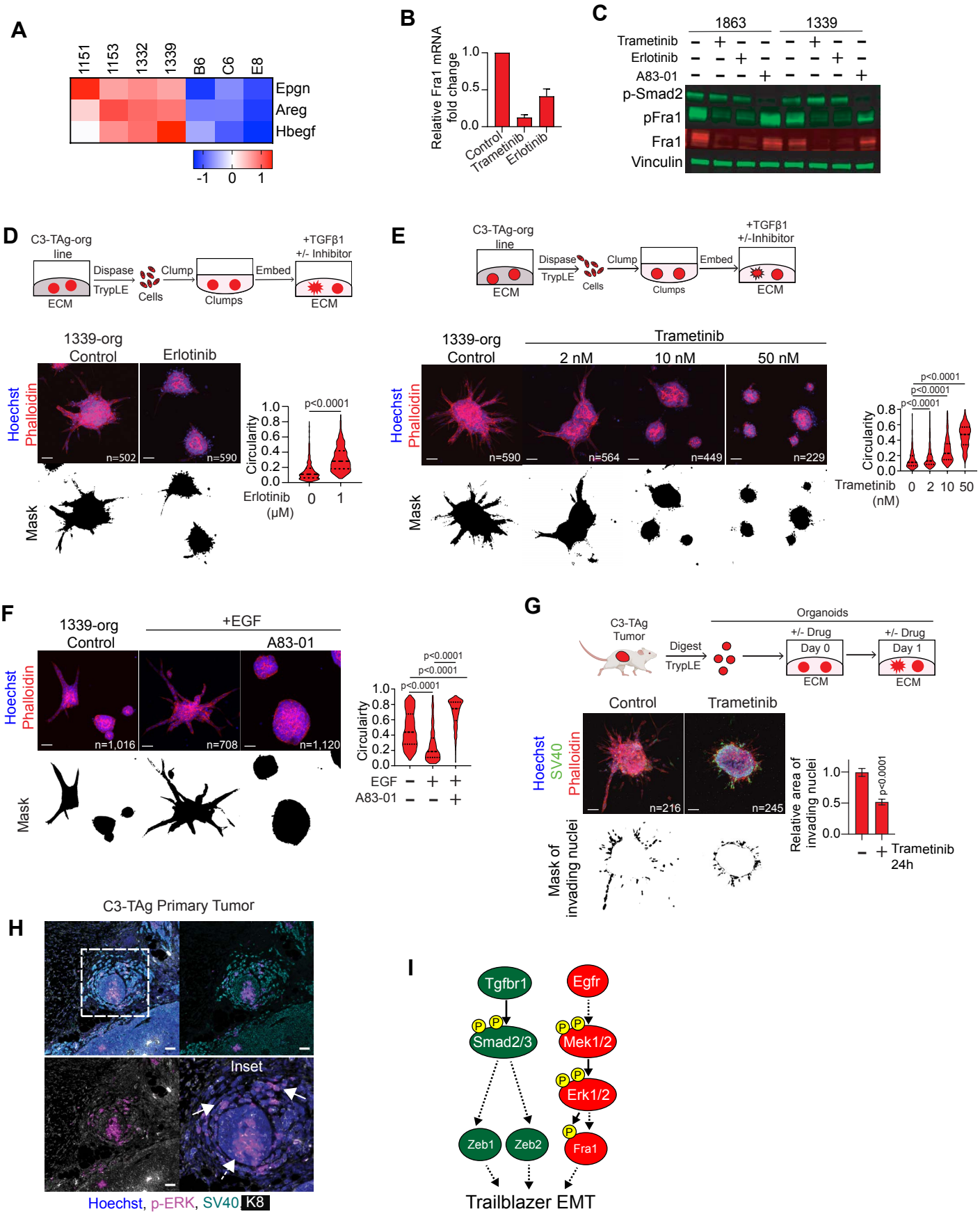
A
1151
1153
1332
1339
B6
C6
E8
Epgn
Areg
Hbegf
-1 0 1
B
Relative Fra1 mRNA fold change
1.0
0.5
0.0
Control
Trametinib
Erlotinib
C
1863
1339
Trametinib
Erlotinib
A83-01
p-Smad2
pFra1
Fra1
Vinculin
D
C3-TAg-org line
Dispase
TrypLE
Cells
Clump
Clumps
Embed
+TGFβ1
+/- Inhibitor
ECM
1339-org
Control
Erlotinib
Hoechst
Phalloidin
n=502
n=590
Mask
Circularity
p<0.0001
Erlotinib (μM)
0
1
E
C3-TAg-org line
Trametinib
2 nM
10 nM
50 nM
n=590
n=564
n=449
n=229
p<0.0001
Trametinib (nM)
0 2 10 50
F
+EGF
A83-01
n=1,016
n=708
n=1,120
EGF
G
C3-TAg Tumor
Digest
Organoids
+/- Drug
Day 0
Day 1
n=216
n=245
Relative area of invading nuclei
Mask of invading nuclei
Trametinib 24h
H
C3-TAg Primary Tumor
Inset
Hoechst, p-ERK, SV40, K8
I
Tgfbr1
Egfr
Smad2/3
Mek1/2
Erk1/2
Zeb1
Zeb2
Fra1
Trailblazer EMT

Figure 5. Egfr pathway activity promotes Fra1 expression and a trailblazer EMT state.
A. Expression of the Egfr ligands Epigen (Epgn), Amphiregulin (Areg) and Heparin-binding epidermal growth factor (Hbegf) in C3-TAg organoids and noninvasive clones. **B.** The Mek1/2 inhibitor Trametinib (10 nM) and Egfr inhibitor Erlotinib (1 µM) reduces Fra1 expression in 1339-org organoids (mean±SEM, n=2). **C.** Immunoblot showing the response of 1863 and 1339 cells to Trametinib, Erlotinib and A83-01. **D.** Erlotinib (1 µM) decreases the TGFβ1 induced invasion 1339-org organoids. Violin plot shows the quantification of circularity (Mann Whitney test, n=organoids, r=3) Scale bars, 50 µm. **E.** Trametinib suppresses the TGFβ1 induced invasion in 1339-org organoids in a dose dependent manner. Violin plot shows the quantification of circularity (Mann Whitney test, n=organoids, r=3). Scale bars, 50 µm **F.** A83-01 suppresses EGF induced invasion in the 1339-org line organoids. Violin plot shows the quantification of circularity (Mann Whitney test, n=organoids, r=4) Scale bars, 50 µm. **G.** Trametinib suppresses C3-TAg tumor organoid invasion. Graph shows the quantification of invading nuclei from 4 different tumors (mean±SEM, Mann Whitney test, n=organoids) Scale bars, 50 µm. **H.** Images show p-ERK expression in regions of invasion (solid arrows) and no invasion (dotted arrow) in primary C3-TAg tumors. Scale bars, 50 **I.** Model showing the parallel regulation of the trailblazer EMT by signaling pathways coordinated by Tgfbr1 and Egfr.

Figure 6

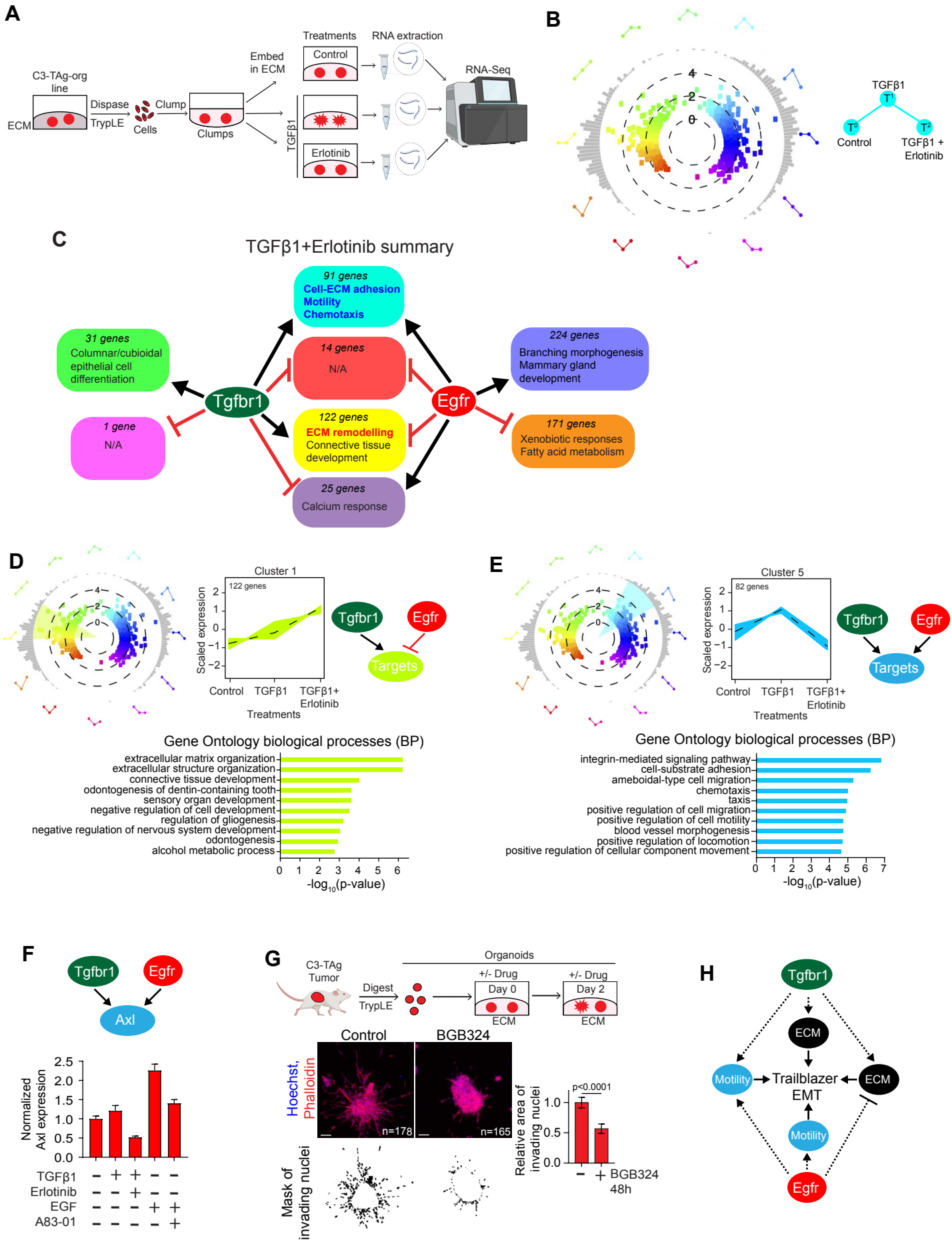
A
C3-TAg-org line
ECM
Dispase
TrypLE
Cells
Clump
Clumps
Embed in ECM
Treatments
Control
TGFβ1
Erlotinib
RNA extraction
RNA-Seq
B
TGFβ1
Control
TGFβ1 + Erlotinib
C
TGFβ1+Erlotinib summary
91 genes
Cell-ECM adhesion
Motility
Chemotaxis
31 genes
Columnar/cubioidal epithelial cell differentiation
14 genes
N/A
224 genes
Branching morphogenesis
Mammary gland development
Tgfbr1
Egfr
1 gene
N/A
122 genes
ECM remodelling
Connective tissue development
171 genes
Xenobiotic responses
Fatty acid metabolism
25 genes
Calcium response
D
Cluster 1
122 genes
Scaled expression
Control
TGFβ1
TGFβ1+ Erlotinib
Treatments
Tgfbr1
Egfr
Targets
Gene Ontology biological processes (BP)
extracellular matrix organization
extracellular structure organization
connective tissue development
odontogenesis of dentin-containing tooth
sensory organ development
negative regulation of cell development
regulation of gliogenesis
negative regulation of nervous system development
odontogenesis
alcohol metabolic process
-log10(p-value)
E
Cluster 5
82 genes
Scaled expression
Control
TGFβ1
TGFβ1+ Erlotinib
Treatments
Tgfbr1
Egfr
Targets
Gene Ontology biological processes (BP)
integrin-mediated signaling pathway
cell-substrate adhesion
ameboidal-type cell migration
chemotaxis
taxis
positive regulation of cell migration
positive regulation of cell motility
blood vessel morphogenesis
positive regulation of locomotion
positive regulation of cellular component movement
-log10(p-value)
F
Tgfbr1
Egfr
Axl
Normalized Axl expression
TGFβ1
Erlotinib
EGF
A83-01
G
C3-TAg Tumor
Digest
TrypLE
Organoids
+/- Drug
Day 0
Day 2
ECM
Control
BGB324
Hoechst, Phalloidin
n=178
n=165
Mask of invading nuclei
Relative area of invading nuclei
p<0.0001
BGB324 48h
H
Tgfbr1
ECM
Motility
Trailblazer EMT
ECM
Motility
Egfr

Figure 6. A compromise between the Egfr and TGFβ signaling programs confers the trailblazer EMT state.

A. Overview of the approach to test the interactions between the Egfr and TGFβ signaling pathways. **B.** Polar plot organizing genes represented as individual dots based on their change in expression in response to treatment with TGFβ1 and co-treatment with TGFβ1 and Erlotinib. Genes that underwent ≥2 fold-change with a $p < 0.05$ after Benjamini-Hochberg correction in at least one condition relative to one of the other conditions are shown (3 total comparisons). The distance from the center of the plot represents the $\Delta\log_2$ value of each gene (increased difference in expression, further from the center). The bar plots on the surface of the circle show the number of genes at the degree point in the plot. Genes are grouped into 12 color-coded groups along 30 degree increments over the 360-degree plot. As an example, genes in group 5 are increased in expression by TGFβ1 and this increase in expression is inhibited by Erlotinib. In contrast, genes in group 1 are increased in expression by TGFβ1, and this induction is further enhanced by treatment with Erlotinib. **C.** Summary of the number of genes and represented biological processes associated with different modes of interaction between the Tgfbr1 and Egfr pathways in cells treated with TGFβ. Also see **Fig. S6C,D** and **Table S4** for the assignment of polar plot clusters and detailed biological processes used to define the summary integration groups. **D.** Egfr suppressed the expression of a subset of genes induced by extrinsic TGFβ. Line plots indicate the expression of genes in Cluster 1. Dashed line indicates the average scaled expression of all Cluster 1 genes. Lower bar graph shows the top 10 biological processes (GO-BP) associated with Cluster 1 genes (Fisher's exact test). **E.** Egfr was necessary for the expression of a subset of genes induced by extrinsic TGFβ**.** Line plots show the expression of genes in Cluster 5. Dashed line indicates the average scaled expression of all Cluster 5 genes. Lower bar graph shows the top 10 biological processes (GO-BP) associated with Cluster 5 genes (Fisher's exact test). **F.** Graph of Axl expression (n=2). **G.** The Axl inhibitor BGB324 (1 μM) suppressed C3-TAg tumor organoid invasion. Graph shows the quantification of invading cells (mean±SEM, Mann Whitney test, n=organoids, r=2). **H.** Model showing how the induction of the trailblazer EMT state requires a compromise between the Tgfbr1 and Egfr regulatory programs. The EGFR dependent induction of genes, such as Axl, that are required for trailblazer cell motility has the cost of restricting the expression ECM remodeling genes known to promote invasion and metastasis.

Figure 7

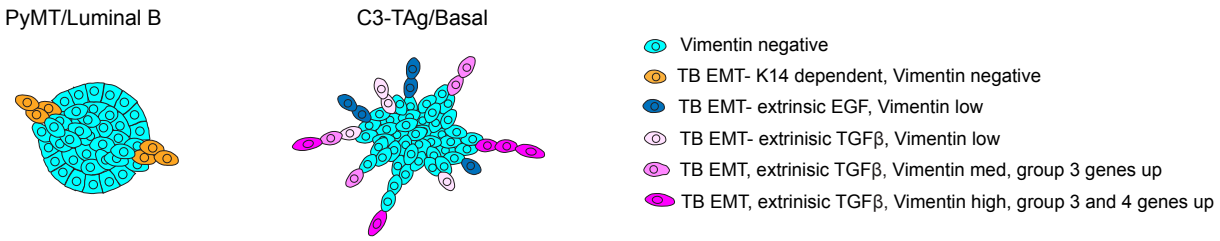
PyMT/Luminal B
C3-TAg/Basal
Vimentin negative
TB EMT- K14 dependent, Vimentin negative
TB EMT- extrinsic EGF, Vimentin low
TB EMT- extrinisic TGFβ, Vimentin low
TB EMT, extrinisic TGFβ, Vimentin med, group 3 genes up
TB EMT, extrinisic TGFβ, Vimentin high, group 3 and 4 genes up

Figure 7. Model of inter- and intra-tumor heterogeneity in trailblazer EMT states. The characteristic features and functional requirements of the trailblazer EMTs induced by extrinsic TGFβ and EGF in C3-TAg/Basal tumors are fundamentally distinct from the trailblazer EMT in PyMT/Luminal B tumors.