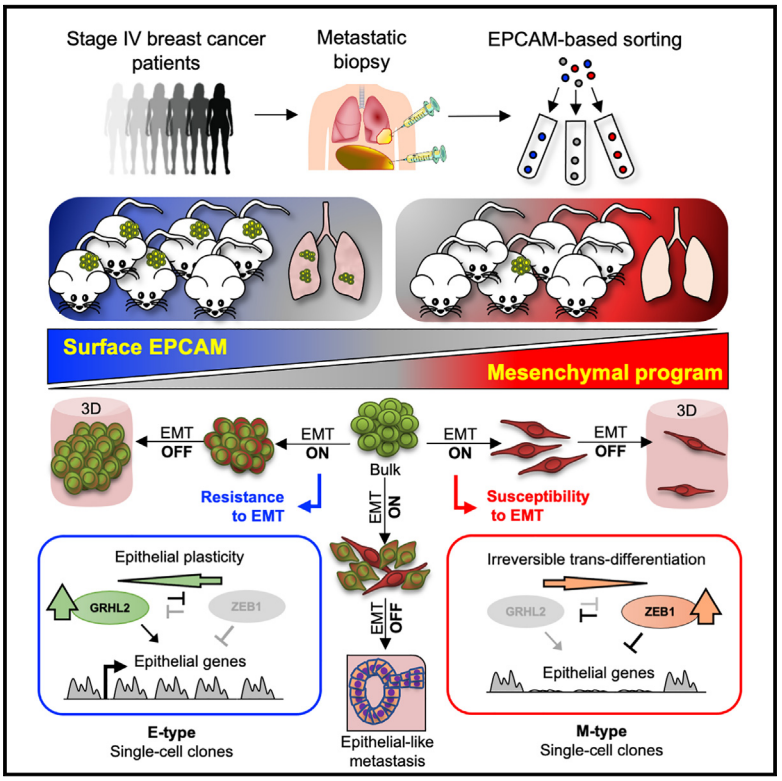
Cell Reports

Article

Resistance to mesenchymal reprogramming sustains clonal propagation in metastatic breast cancer

Graphical abstract



Highlights

- EPCAM(high) cells from metastatic biopsies can propagate breast cancer in xenografts
- Resistance to EMT sustains tumorigenesis and metastatic spread of the human disease
- Irreversible EMT results in restrained tumorigenic potential in clonal subsets
- ZEB1 and its target GRHL2 control distinct epigenetic paths of EMT susceptibility

Authors

Massimo Saini, Laura Schmidleitner, Helena Domínguez Moreno, ..., Martin R. Sprick, Andreas Trumpp, Christina H. Scheel

Correspondence

massimo.saini@biol.ethz.ch (M.S.), martin.sprick@hi-stem.de (M.R.S.), a.trumpp@dkfz-heidelberg.de (A.T.), christina.scheel@klinikum-bochum.de (C.H.S.)

In brief

Saini et al. show that human breast cancer cells resist EMT induction and maintain high EPCAM expression to propagate the metastatic disease. In contrast, loss of EPCAM defines clones arresting in a mesenchymal state with suppressed tumorigenic features. Clonal levels of ZEB1 and GRHL2 control global epigenetic programs determining EMT resistance.

Saini et al., 2023, Cell Reports *42*, 112533
June 27, 2023
https://doi.org/10.1016/j.celrep.2023.112533

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Resistance to mesenchymal reprogramming sustains clonal propagation in metastatic breast cancer

Massimo Saini,[1,2,3,4,24,*] Laura Schmidleitner,[3,5,6,7,8,24] Helena Domínguez Moreno,[9] Elisa Donato,[1,2] Mattia Falcone,[1,2,19] Johanna M. Bartsch,[3] Corinna Klein,[1,2] Vanessa Vogel,[1,2] Roberto Würth,[1,2] Nicole Pfarr,[10] Elisa Espinet,[1,2,20] Mareike Lehmann,[11,12] Melanie Königshoff,[13,21] Manuel Reitberger,[1,2] Simon Haas,[1,2,14,22] Elisabeth Graf,[15]

(Author list continued on next page)

[1]Division of Stem Cells and Cancer, German Cancer Research Center (DKFZ), and DKFZ-ZMBH Alliance, Heidelberg, Germany
[2]Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM gGmbH), Heidelberg, Germany
[3]Institute of Stem Cell Research, Helmholtz Center Munich, Neuherberg, Germany
[4]Department of Biology, Institute of Molecular Health Sciences, Swiss Federal Institute of Technology (ETH) Zurich, Zurich, Switzerland
[5]Translational Pancreatic Cancer Research Center, Klinik und Poliklinik für Innere Medizin II, Klinikum rechts der Isar, Technical University of Munich, Munich, Germany
[6]Center for Functional Protein Assemblies (CPA), Technical University of Munich (TUM), Garching, Germany
[7]Center for Organoid Systems (COS), Technical University of Munich (TUM), Garching, Germany
[8]Munich Institute of Biomedical Engineering (MIBE), Technical University of Munich (TUM), Garching, Germany
[9]Division of Molecular Biology, Biomedical Center, Faculty of Medicine, Ludwig-Maximilian University of Munich (LMU), Munich, Germany
[10]Institute of Pathology, Technical University of Munich (TUM), Munich, Germany
[11]Institute for Lung Health and Immunity (LHI) and Comprehensive Pneumology Center (CPC), Helmholtz Center Munich, Member of the German Center for Lung Research (DZL), Munich, Germany
[12]Institute for Lung Research, Philipps-University Marburg, Member of the German Center for Lung Research (DZL), Marburg, Germany
[13]Research Unit Lung Repair and Regeneration, Helmholtz Center Munich, Member of the German Center of Lung Research (DZL), Munich, Germany
[14]Berlin Institute for Medical Systems Biology, Max Delbrück Center for Molecular Medicine in the Helmholtz Association, Berlin, Germany
[15]Institute of Human Genetics, Helmholtz Center Munich, Neuherberg, Germany
[16]Department of Gynaecology and Obstetrics, University Women's Clinic, University Medical Centre Mannheim, Mannheim, Germany
[17]National Center for Tumor Diseases, University Hospital and German Cancer Research Center, Heidelberg, Germany
[18]German Cancer Consortium (DKTK), Germany

(Affiliations continued on next page)

SUMMARY

The acquisition of mesenchymal traits is considered a hallmark of breast cancer progression. However, the functional relevance of epithelial-to-mesenchymal transition (EMT) remains controversial and context dependent. Here, we isolate epithelial and mesenchymal populations from human breast cancer metastatic biopsies and assess their functional potential *in vivo*. Strikingly, progressively decreasing epithelial cell adhesion molecule (EPCAM) levels correlate with declining disease propagation. Mechanistically, we find that persistent EPCAM expression marks epithelial clones that resist EMT induction and propagate competitively. In contrast, loss of EPCAM defines clones arrested in a mesenchymal state, with concomitant suppression of tumorigenicity and metastatic potential. This dichotomy results from distinct clonal trajectories impacting global epigenetic programs that are determined by the interplay between human ZEB1 and its target GRHL2. Collectively, our results indicate that susceptibility to irreversible EMT restrains clonal propagation, whereas resistance to mesenchymal reprogramming sustains disease spread in multiple models of human metastatic breast cancer, including patient-derived cells *in vivo*.

INTRODUCTION

More than 97% of breast cancer deaths are secondary to the spread of the disease to vital organs.[1] The progression of a localized tumor to metastatic breast cancer (MBC) may span years to decades after the primary diagnosis.[2,3] In several models, breast cancers undergo paths of *trans*-differentiation during disease progression[4–6] such as the epithelial-to-mesenchymal transition (EMT).[7–10] In the mammary epithelium, diverse cellular states can be induced and maintained through paracrine EMT-induced stimuli.[11] This results in the acquisition of mesenchymal gene expression programs and the loss of epithelial markers (e.g., EPCAM) driven by specific transcription factors (EMT-TFs) (e.g., ZEB1, SNAI1/2, and TWIST1, among others).[7,12,13] In

Thomas Schwarzmayr,[15] Tim-Matthias Strom,[15] Saskia Spaich,[16] Marc Sütterlin,[16] Andreas Schneeweiss,[17] Wilko Weichert,[10,18] Gunnar Schotta,[9] Maximilian Reichert,[5,6,7,8,18] Nicola Aceto,[4] Martin R. Sprick,[1,2,18,25,*] Andreas Trumpp,[1,2,18,25,*] and Christina H. Scheel[3,23,25,26,*]

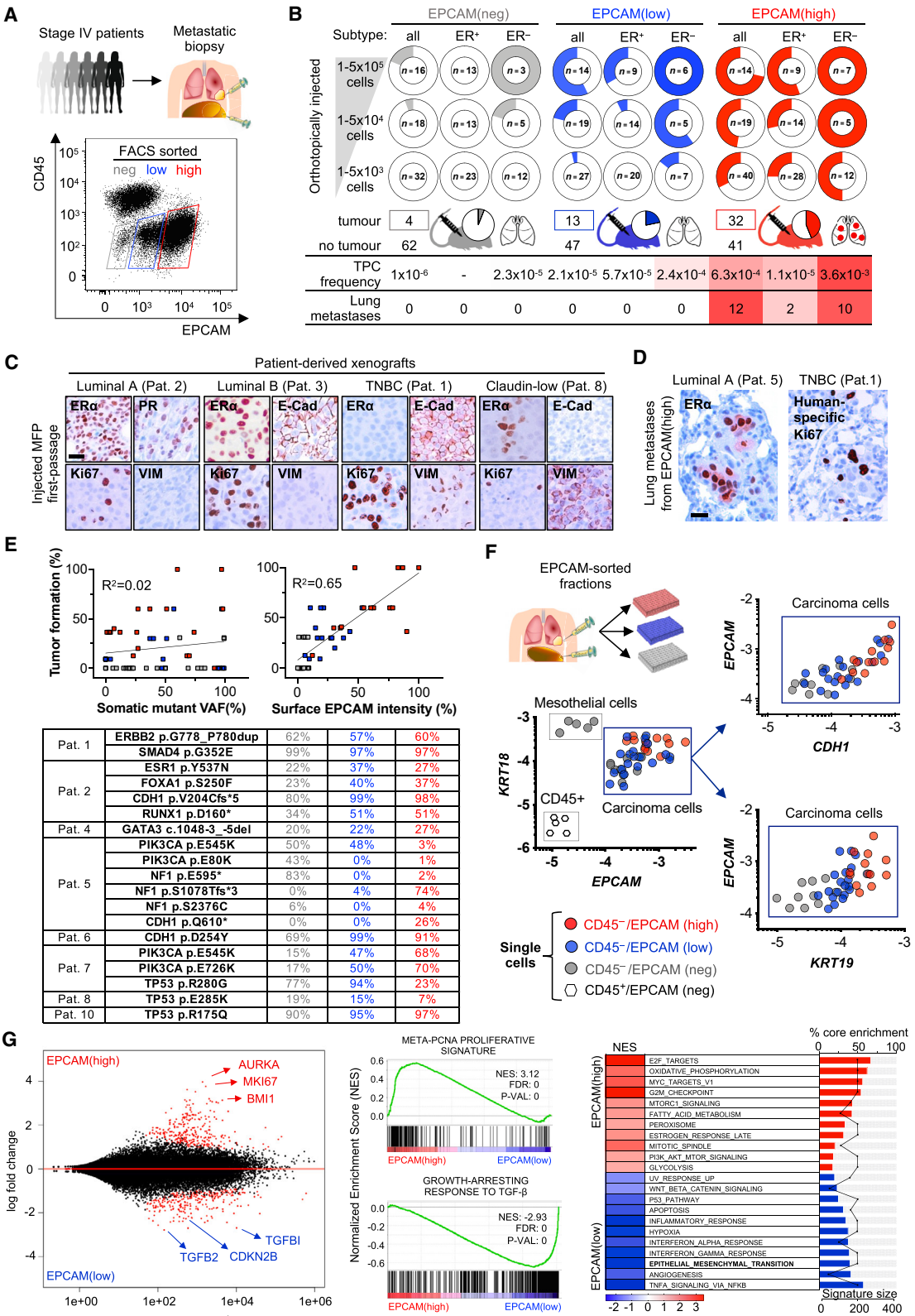
[19]Present address: Division of Oncogenomics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, the Netherlands
[20]Present address: Department of Pathology and Experimental Therapy, School of Medicine, University of Barcelona (UB), and Molecular Mechanisms and Experimental Therapy in Oncology Program (Oncobell), Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain
[21]Present address: Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh, Pittsburgh, PA, USA
[22]Present address: Berlin Institute of Health (BIH) at Charité – Berlin University of Medicine, Berlin, Germany
[23]Present address: Skin Cancer Center, Department of Dermatology, St. Josef Hospital, Ruhr-University Bochum, Bochum, Germany
[24]These authors contributed equally
[25]Senior author
[26]Lead contact
*Correspondence: massimo.saini@biol.ethz.ch (M.S.), martin.sprick@hi-stem.de (M.R.S.), a.trumpp@dkfz-heidelberg.de (A.T.), christina.scheel@klinikum-bochum.de (C.H.S.)
https://doi.org/10.1016/j.celrep.2023.112533

carcinomas, molecular changes induced by the EMT have been held accountable for the progression of localized breast cancers toward MBC.[8,9] However, experimental studies addressing the functional impact of the EMT in breast cancer and other carcinomas have yielded a wealth of controversial results.[14,15] For example, in genetically engineered mouse models, it was shown that the genetic loss of EMT-TFs (*Snai1* and *Twist1*) or microRNAs that are crucial for the EMT (*miR-200* family) does not impact metastatic progression.[16,17] By contrast, other studies have revealed that deletion of *Zeb1* decreases invasion and lung colonization without affecting primary tumor growth,[18] while conditional deletion of *Snai1* in other backgrounds can hamper the spread of mammary tumors,[19] suggesting that EMT-TFs can contribute to metastatic progression. Some of this apparent controversy might be explained by functional redundancy[20] and pleiotropy of EMT-inducing factors.[21] For instance, expression of the EMT inducer *Twist1* can decrease proliferation of tumor cells in some contexts[22] while increasing their abundance within the circulation.[23] More discrepancy arises from differences in the tissue or cell of origin of cancer[24] and/or incomplete or inappropriate modeling of the human disease by genetic mouse models.[25] Thus, it was suggested that the impact of EMT depends on the carcinoma type and reflects a reactivation of similar programs active in the corresponding normal tissue of origin during development.[8] Further complicating the scenario, an intermediate ("hybrid") epithelial-mesenchymal phenotype was shown to harbor the greatest metastatic potential in some contexts,[26,27] suggesting that partial EMT may have a different impact and functional role from complete EMT.[28,29] This finding involves a further layer of regulation by microenvironmental tumor-stromal interactions, tuning the intensity of the EMT and producing a heterogeneous spectrum of epithelial-mesenchymal states[10] that might also cooperate with each other.[30] Finally, multiple studies emphasize that a transient, rather than a permanent, expression of EMT-TFs promotes the outgrowth of metastases,[22,31,32] as macroscopic metastases generated by carcinomas display an epithelial profile, as typically observed in breast cancer histopathology.[33,34] Hence, a strict role for the EMT during breast cancer progression and metastatic spread is still contended, and it remains poorly understood how epithelial, mesenchymal, or intermediate phenotypes arise and how they contribute to metastatic outgrowth. To address this issue in the context of the human disease, we examined the *in vivo* tumor-propagating ability of cells directly isolated from metastatic patients. By extending the analysis across different breast cancer subtypes and models, we identified intra-tumor heterogeneity (ITH) for clonally restricted programs that determine epigenetic susceptibility to the EMT. We show that resistance to an irreversible EMT is key for disease propagation but is confined to a subset of clones, offering a mechanistic basis to integrate disparate findings regarding EMT events in breast cancer.

RESULTS

High EPCAM levels mark disease-propagating cells in MBC biopsies

To determine how epithelial or mesenchymal cell phenotypes contribute to MBC, we collected liquid biopsies from metastatic effusions of stage IV patients (n = 10) encompassing several subtypes of breast cancer (Figure S1A). By fluorescence-activated cell sorting (FACS), we excluded CD45-positive cells and sorted 3 distinct subpopulations based on their surface levels of the epithelial cell adhesion molecule (EPCAM) as EPCAM(high), -(low), and -(neg) (Figures 1A and S1B). We assessed their tumorigenicity and spontaneous metastatic potential *in vivo* by orthotopic xenotransplantation in immunodeficient mice, thus generating patient-matched xenografts (PDXs) as clinically relevant models of the metastatic disease.[35–38] First, the number of tumor-propagating cells (TPCs) was assayed for each subset. Limiting dilution analysis revealed that EPCAM(high) cells were almost 3 times more efficient in seeding tumors (TPC frequency: 1 in 6.3×10^4) compared with EPCAM(low) cells (1 in 2.1×10^5) and 16 times more than EPCAM(neg) cells (1 in 10^6) (Figures 1B, S1C, and S1D). In biopsies from estrogen receptor (ER)-positive (ER^+) patients, the EPCAM(high) population was 5 times more tumorigenic (TPC frequency: 1 in 10^5) compared with the EPCAM(low) population (1 in 5.7×10^5), while the EPCAM(neg) population lacked detectable TPCs (Figure 1B). In ER-negative (ER^-) samples, all populations were tumorigenic (Figures 1B, S1C, and S1D), consistent with a higher degree of plasticity observed in ER^- liquid biopsies.[25] Still, a higher TPC frequency was evident in EPCAM(high) (1 in 3.6×10^3) compared with matched



Pat. 1	ERBB2 p.G778_P780dup	62%	57%	60%
	SMAD4 p.G352E	99%	97%	97%
Pat. 2	ESR1 p.Y537N	22%	37%	27%
	FOXA1 p.S250F	23%	40%	37%
	CDH1 p.V204Cfs*5	80%	99%	98%
	RUNX1 p.D160*	34%	51%	51%
Pat. 4	GATA3 c.1048-3_-5del	20%	22%	27%
Pat. 5	PIK3CA p.E545K	50%	48%	3%
	PIK3CA p.E80K	43%	0%	1%
	NF1 p.E595*	83%	0%	2%
	NF1 p.S1078Tfs*3	0%	4%	74%
	NF1 p.S2376C	6%	0%	4%
	CDH1 p.Q610*	0%	0%	26%
Pat. 6	CDH1 p.D254Y	69%	99%	91%
Pat. 7	PIK3CA p.E545K	15%	47%	68%
	PIK3CA p.E726K	17%	50%	70%
	TP53 p.R280G	77%	94%	23%
Pat. 8	TP53 p.E285K	19%	15%	7%
Pat. 10	TP53 p.R175Q	90%	95%	97%

(legend on next page)

EPCAM(low) (1 in 2.4 × 10^4) and EPCAM(neg) cells (1 in 2.3 × 10^5). Together, these data indicated functional ITH in metastatic samples and show that EPCAM(high) breast cancer cells predominantly propagated the disease, suggesting a higher metastatic potential for more epithelial cells. Necroscopic inspection of tumor-bearing mice revealed human metastatic nodules in the lungs of 24 out of 73 mice injected with EPCAM(high) cells (Figures 1B and 1D) but none in the 126 mice injected with EPCAM(low) or EPCAM(neg) cells despite comparable tumor size at the endpoint. Histopathological analysis confirmed that mammary fat pad (MFP) xenografts and lung metastases recapitulated the original disease subtype (Figures 1C, S1E, and S1F) but also revealed a widespread positive staining for epithelial markers such as cytokeratins, EPCAM, and (in $CDH1^{WT/WT}$ samples) E-cadherin, with little or no expression of mesenchymal markers. For instance, Vimentin was co-expressed with epithelial markers and almost exclusively detected in a claudin-low sample plus in rare epithelial cells in ER^- xenografts (Figures S1E and S1F). To pursue a deeper characterization of the transplanted cells, we assessed the variant allele frequency (VAF) in each EPCAM-sorted subset by a panel-based targeted sequencing and detected 18 somatic mutations either common to all subsets (e.g., *TP53* and *SMAD4*) or distributed asymmetrically across EPCAM-defined subsets within individual samples (e.g., *PIK3CA* in Pat.5) (Figure 1E). Overall, the VAF of the detected mutations could not explain (R^2 = 0.02) the differences in tumor-propagating potential of each respective EPCAM-defined subset. By contrast, we observed a robust correlation (R^2 = 0.654) between EPCAM mean fluorescence intensity (MFI) and the tumorigenicity of the corresponding sorted fraction (Figure 1E). To further examine the composition of the biopsies, we performed single-cell gene expression analysis on individual cells from the original sample and classified them through specific markers. A prevalent cluster of *EPCAM*-expressing KRT19(high)/KRT18(high) cells could be discriminated from traces of non-carcinomatous cells. Within the former, a trending co-expression between EPCAM and epithelial markers like CDH1 and KRT19 was evident (Figure 1F). The (high) and (low) subsets, which stained overall positive for surface EPCAM, almost exclusively contained KRT19(high)/KRT18(high) carcinoma cells (95%), as did the majority (83%) of the EPCAM(neg) subset. The remaining EPCAM(neg) cells did not express *EPCAM* at the transcript level and could be classified as $KRT18^+$ mesothelial cells, a frequent contaminant in serous effusions[39] that lack *EPCAM* expression[40] (Figure 1F). Together, these results suggest that the observed differences in TPC frequency are best explained by differences between tumor-intrinsic characteristics that defined the three identified populations. In line with this hypothesis, bulk RNA sequencing of EPCAM(high) vs. EPCAM(low) subpopulations revealed consistent changes in cell cycle genes, with *AURKA* and *MKI67* on the top among differentially expressed genes (Figure 1G). By Gene Ontology (GO) and gene set enrichment analysis (GSEA), we found evidence of cytostatic signaling in the poorly tumorigenic EPCAM(-low) cells, while EPCAM(high) cells were enriched for features associated with an active cell cycle, oxidative phosphorylation, and self-renewal (Figures 1G and S1G). Globally, EPCAM reduction was paralleled by a decrease in the expression of epithelial markers (including *CDH1*, *ESRP1*, *SPDEF*, and *GRHL2*), which was mirrored by a gain in mesenchymal and stromal genes (Figure S1G). Indeed, the EMT process was among the largest and non-redundant gene sets enriched in EPCAM(low) cells compared with EPCAM(high) (Figure 1G). Taken together, these data revealed a correlation between the potential of discrete MBC cell subsets to spread the disease *in vivo* and their progressive loss in epithelial traits such as EPCAM. We therefore hypothesized that EMT-related cellular programs may negatively impact metastatic progression in human breast cancer, prompting further investigation of the link between EMT and tumorigenicity.

Divergent response to EMT stimulation generates heterogeneity for EPCAM

To investigate how EMT programs affect MBC outgrowth, we examined EPCAM-dependent phenotypes in human breast cancer lines known to spontaneously undergo continuous EMT events.[41–43] In the triple-negative SUM-149PT cell line and in the SUM-229PE and HCC1143 cell lines, we found evidence of EPCAM heterogeneity (Figure 2A) and confirmed EPCAM as a robust marker to prospectively isolate breast cancer cells based on epithelial or mesenchymal traits (Figures S2A and S2B). Similar to our findings in patient samples, EPCAM(neg) carcinoma cells that express the *EPCAM* mRNA but not the protein are enriched for stromal and mesenchymal markers and upregulate EMT-TFs such as *TWIST1/2*, *SNAI1/2*, and *ZEB1/2* (Figure S2B) compared with the EPCAM(pos) majority of cancer cells. To directly evaluate *in vivo* metastatic potential, EPCAM(pos) and EPCAM(neg) subpopulations from SUM-149PT were FACS-sorted and injected in mouse hind leg femurs (Figures 2A and 2B). Importantly, this assay was previously adopted to determine metastasis-initiating activity[44] and thus serves as a proxy for disease propagation in the bone, one of the most frequently affected

Figure 1. High EPCAM levels mark disease-propagating cells in MBC biopsies

(A) Schematic representation of liquid biopsy collection and FACS gates utilized.

(B) Xenotransplantation assay: (top) proportions and numbers of tumors formed in injected mice (n = x), (middle) frequency of tumor-propagating cells (TPCs), and (bottom) number of mice displaying lung metastasis; patient samples stratified per ER subtype.

(C) Representative immunohistochemistry (IHC) staining in tumor sections from mammary fat pad (MFP) xenografts. ERα, estrogen receptor alpha; PR, progesterone receptor; E-Cad, E-cadherin; Ki67, human-specific Ki67; VIM, Vimentin. Scale bar: 25 μm.

(D) Representative IHC staining of ERα and human-specific Ki67 on spontaneous lung metastases of mice injected with EPCAM(high) cells; scale bar: 25 μm.

(E) Charts relating (top left) mutational VAF or (top right) EPCAM expression to the proportion of xenograft tumors; n = 10 patients; (bottom) somatic mutations and relative VAF in subsets per patient; color code as in (A).

(F) Transcript abundance in cells from metastatic effusions assessed by single-cell (sc)-qRT-PCR; Log scale; color code as in (A).

(G) Summary of RNA-seq results from metastatic effusions; (left) volcano plot of poly-adenylated RNAs; colored dots: adjusted p (padj) < 0.01; (middle) hallmark GSEA; (right) signatures with false discovery rates (FDR) <0.25 ranked by the percentage of core enrichment; n = 4.

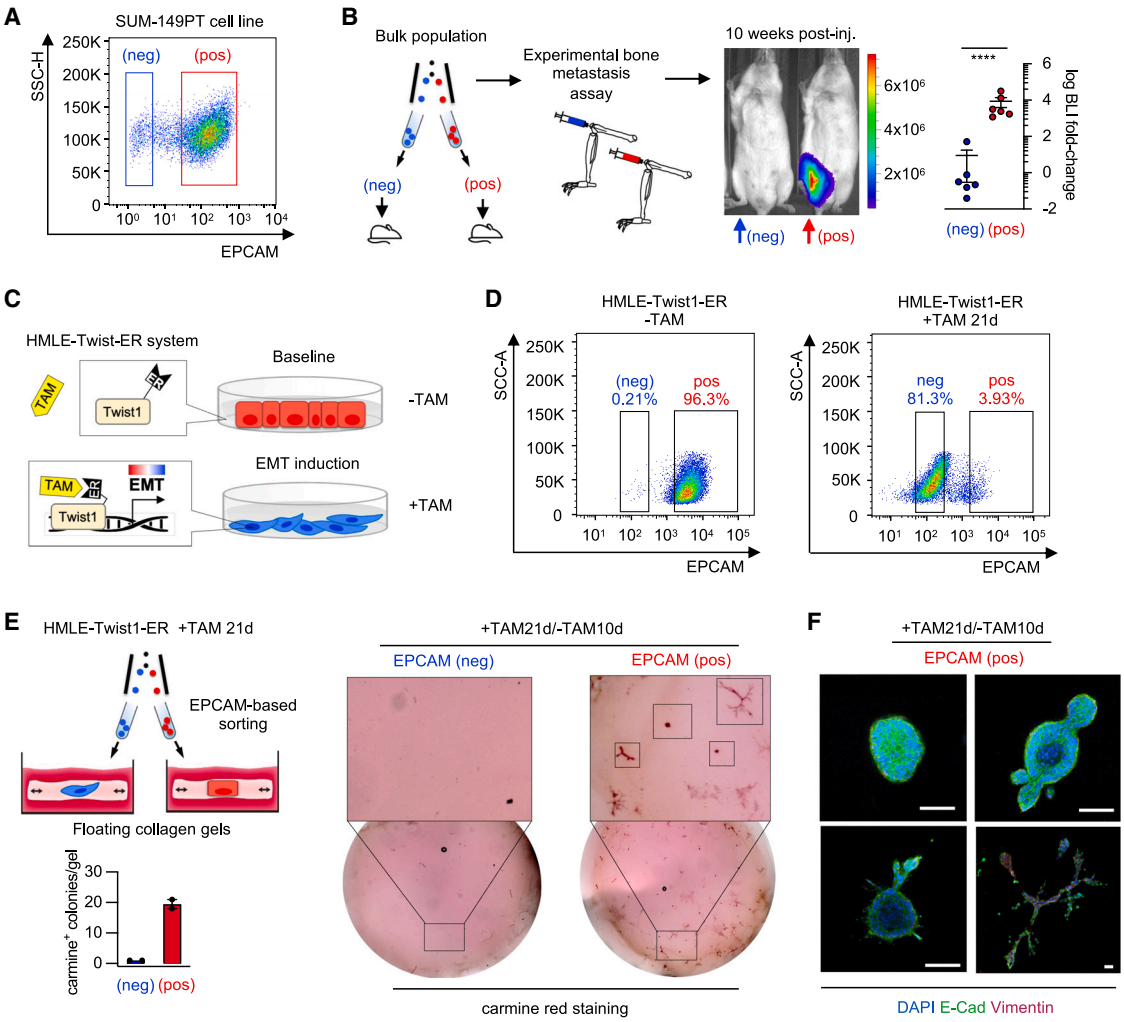


Figure 2. Divergent response to EMT stimulation generates heterogeneity for EPCAM

(A) FACS gates utilized for the SUM-149PT bulk population.

(B) Bone colonization assay; bioluminescence imaging (BLI) at 10 weeks after intra-femoral injection of cells sorted as in (A); rainbow scale: photons/s/cm^2/sr; dot plot graph shows fold change to baseline at 20 weeks (endpoint); n = 6 mice per group; mean ± SEM; unpaired Student's t test of log-transformed values with Welch's correction; ****$p < 0.0001$.

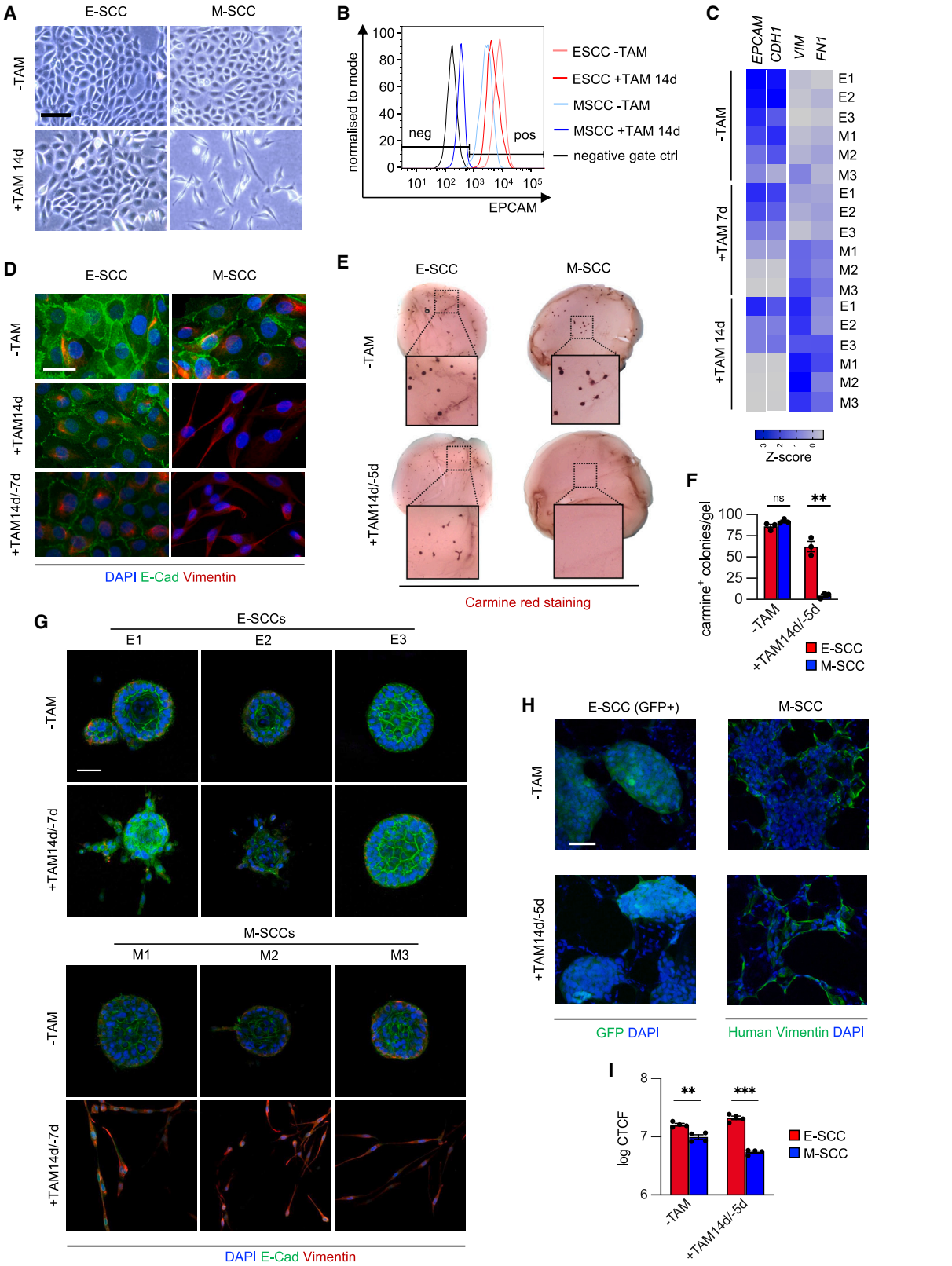
(C) Schematic representation of the HMLE-Twist1-ER experimental system.

(D) FACS gates utilized in HMLE-Twist1-ER cells in −TAM or +TAM 21 days conditions.

(E) Organotypic assay from HMLE-Twist1-ER cells +TAM 21 days; (left) number of organoids per gel; n = 2 (3 replicates each); mean ± SEM; (right) representative pictures of carmine red-stained collagen gels of EPCAM-sorted subsets (right panel); quadrants show focal 10× magnification.

(F) Immunofluorescence (IF) staining for EMT markers in collagen gel organoids produced by sorted EPCAM(pos) cells as in (E); scale bar: 50 μm.

organs in MBC. In line with the results obtained from MBC biopsies, EPCAM(pos) cells generated bone metastases after 20 weeks of observation, while EPCAM(neg) cells failed to seed any lesion during the remaining mouse lifespan (Figure 2B). Similarly, we found no detectable tumorigenic activity by *in vivo* transplantation of pure EPCAM(neg) cancer cells from the HCC1143 and SUM-229PE cell lines (Figure S2C), although injection of the EPCAM(pos) counterpart resulted in *in vivo* outgrowth. Together, all human models tested suggested that a lack of *in vivo* metastatic potential by EPCAM(neg) mesenchymal cells might be a shared feature for the most common breast cancer subtypes. Since *in-vitro*-cultivated cells preserve such heterogeneity, we tried to unravel how EPCAM heterogeneity originates during an EMT event. For this purpose, we used a tightly controlled system to induce EMT in human mammary epithelial (HMLE) cells by tamoxifen (TAM)-controllable *Twist1* (HMLE-Twist1-ER)[31,45–48] (Figure 2C). As previously described,[31] HMLE cells converted into a mesenchymal state after 21 days of persistent Twist1-activation (+TAM21d) phenocopying EPCAM(neg) carcinoma cells (Figure S2D). Analysis of cell-surface EPCAM expression after +TAM 21 days revealed that most cells had lost EPCAM cell surface expression. A minor subset of EPCAM(pos) cells, however, persisted in culture (Figure 2D). To address this heterogeneity, we isolated EPCAM(neg) and EPCAM(pos) populations



(legend on next page)

after induction of EMT as described above. Sorted cells were transferred to TAM-free conditions for 10 days (+TAM21d/−TAM10d) to discontinue EMT induction and thus reproduce the effect of a transient EMT stimulus, as described.[31] Based on morphology and expression of markers such as *CDH1* and *OVOL2*, we noticed that EPCAM(neg) sorted cells remained mesenchymal even after cessation of Twist1 activation and could not recover EPCAM expression. In contrast, corresponding EPCAM(pos) cells preserved an epithelial profile, thus appearing insensitive to the EMT-inducing stimulus (Figures 2D, S2E, and S2F). Since HMLE cells are not tumorigenic in mice, we tested the ability of each FACS-separated subset to generate organoids in a three-dimensional organotypic assay, performed in conditions that mimic the breast microenvironment.[49] Strikingly, only EPCAM(pos) cells gave rise to branching organoids, whereas EPCAM(neg) cells failed to generate multicellular structures in collagen gels (Figures 2E and 2F) despite their retained ability to proliferate indefinitely in two-dimensional culture conditions (Figure S2E). Together, these results demonstrated that induction of an EMT produces contrasting cellular responses, giving rise to a functionally relevant heterogeneity marked by EPCAM. We thus hypothesized that a subset of breast cancer cells may be able to resist the EMT and develop functional properties linked to epithelial tumorigenesis.

Single-cell clonal analysis reveals intrinsic resistance to EMT vs. stable mesenchymal reprogramming

Next, we set out to determine why EPCAM-defined populations appeared differentially susceptible to EMT. To do so, we isolated a series of single-cell clones (SCCs) from the HMLE-Twist1-ER bulk population and treated each clone with TAM for 7 days (+TAM7d) or 14 days (+TAM14d) to assess clonal responsiveness to EMT (Figure S3A). Only a subset of SCCs ("M-SCCs"; 72% of all SCCs) responded to TAM stimulation by a morphologically evident EMT (Figures 3A and S3A). This change was characterized by a loss in epithelial marker expression (EPCAM, CDH1) and a concurrent gain of mesenchymal markers (VIM, FN1) in all M-SCCs at mRNA and protein levels (Figures 3B, 3C, and S3B). The remaining SCCs ("E-SCCs"; 28% of all SCCs) remained epithelial at the morphological and molecular level, as assessed at the +TAM 7 days and +TAM 14 days time points, where E-SCCs conserved expression of EPCAM and E-cadherin but gained Vimentin expression in a time-dependent manner (Figures 3B, 3C, and S3B). We verified that TAM-dependent nuclear translocation of Twist1 occurred in the tested E-SCCs (E1–E3) and M-SCCs (M1–M3). Moreover, we quantified the induction of Twist1 together with its direct target gene *WNT5A*[50] at the mRNA level. Together, these results indicated comparable activation of the Twist1-ER inducible system in all SCCs (Figure S3D and S3E), suggesting that the divergent response to TAM was indicative of a different proclivity to the EMT process. To verify if these phenotypes were stable, we stimulated E-SCCs and M-SCCs by +TAM 14 days and then cultivated the cells in TAM-free conditions for at least 7 days (+TAM14d/−TAM7d). We observed that in all M-SCCs, epithelial markers were permanently downregulated, despite TAM discontinuation. Thus, M-SCCs preserved their mesenchymal profile, including morphology, gene expression markers, and fibroblast-like distribution of the Vimentin cytoskeleton (Figures 3D and S3F). By contrast, stimulated E-SCCs preserved their epithelial morphology as well as the typical epithelial distribution of E-cadherin and Vimentin after TAM discontinuation. Prolonged cultivation in TAM-free conditions (for 14 or 21 days) had no further impact on the epithelial or mesenchymal phenotype produced by transient EMT in each clonal type (Figure S3G), and prolonged Twist1 activation by 21 or 28 days of TAM treatment did not alter the epithelial traits of E-SCCs (Figures S3F–S3H). These results indicated a clonally divergent response to transient EMT, with E-SCCs resisting loss of epithelial state and M-SCCs *trans*-differentiating irreversibly to a mesenchymal state. Since only M-SCCs transitioned to an EPCAM(neg) state in response to TAM, while E-SCCs displayed a persistent EPCAM(-pos) profile during stimulation (Figure 3B), we hypothesized that the EMT-refractory E-SCC phenotype may produce the distinctive tumorigenic features emerging from bulk EPCAM-based sortings. In support of this hypothesis, TAM-stimulated E-SCCs were 3 times more clonogenic in mammosphere (MS) assays than without TAM, a finding in line with previous data on epithelial HMLE subsets.[31] By contrast, limiting dilution and replating assays revealed a dramatic loss in MS-forming potential for M-SCCs after +TAM 14 days/−TAM 7 days stimulation, even though we did not observe an impact on their proliferative ability in two dimensions (Figure S3I). Altogether, these data suggested that cell-intrinsic sensitivity to an irreversible EMT can reduce clonogenicity in three-dimensional environments. To verify if these observations were linked to the poor tumorigenicity of EPCAM(neg) cells, we seeded individual SCCs in organoid assays.[49] When −TAM cells were analyzed, both E-SCCs and M-SCCs produced epithelial organoids expressing E-cadherin with similar frequency. However, when +TAM 14 days/−TAM 7 days stimulated clones were assayed, only E-SCCs generated multicellular organoids, while M-SCCs remained dispersed in the

Figure 3. Single-cell clonal analysis reveals intrinsic resistance to EMT vs. stable mesenchymal reprogramming

(A) Bright-field images of a representative E-SCC and an M-SCC in two-dimensional culture; scale bar: 100 μm.

(B) Flow cytometry for EPCAM of one representative E-SCC and one M-SCC.

(C) Heatmap of relative mRNA expression of indicated genes in 3 E-SCCs (E1–E3) and 3 M-SCCs (M1–M3); n = 3 per clone; mean values.

(D) IF staining for EMT markers in two-dimensional cultures of one representative E-SCC and one M-SCC; scale bar: 20 μm.

(E) Carmine red-stained collagen gels one E-SCC and one M-SCC; quadrants show focal 10× magnification.

(F) Number of organoids per gel generated by a series of 3 E-SCCs (E) and 3 M-SCCs (M). n = 3 per clone; mean ± SEM; multiple t tests (Holm-Sidak correction); **p = 0.002, ns, not significant.

(G) IF staining of EMT markers in organotypic cultures of 3 E-SCCs (E1–E3) and 3 M-SCCs (M1–M3); n = 3 per clone; scale bar: 20 μm.

(H) Representative IF of one GFP$^+$ E-SCC (E) and one Vimentin-stained M-SCC (M) in a PCLS assay; scale bar: 50 μm.

(I) Corrected total cell fluorescence (CTCF) values for DAPI on green-stained areas in (H); n = 4; mean ± SEM; multiple t tests (Holm-Sidak correction); **p = 0.0022, ***p = 0.0001.

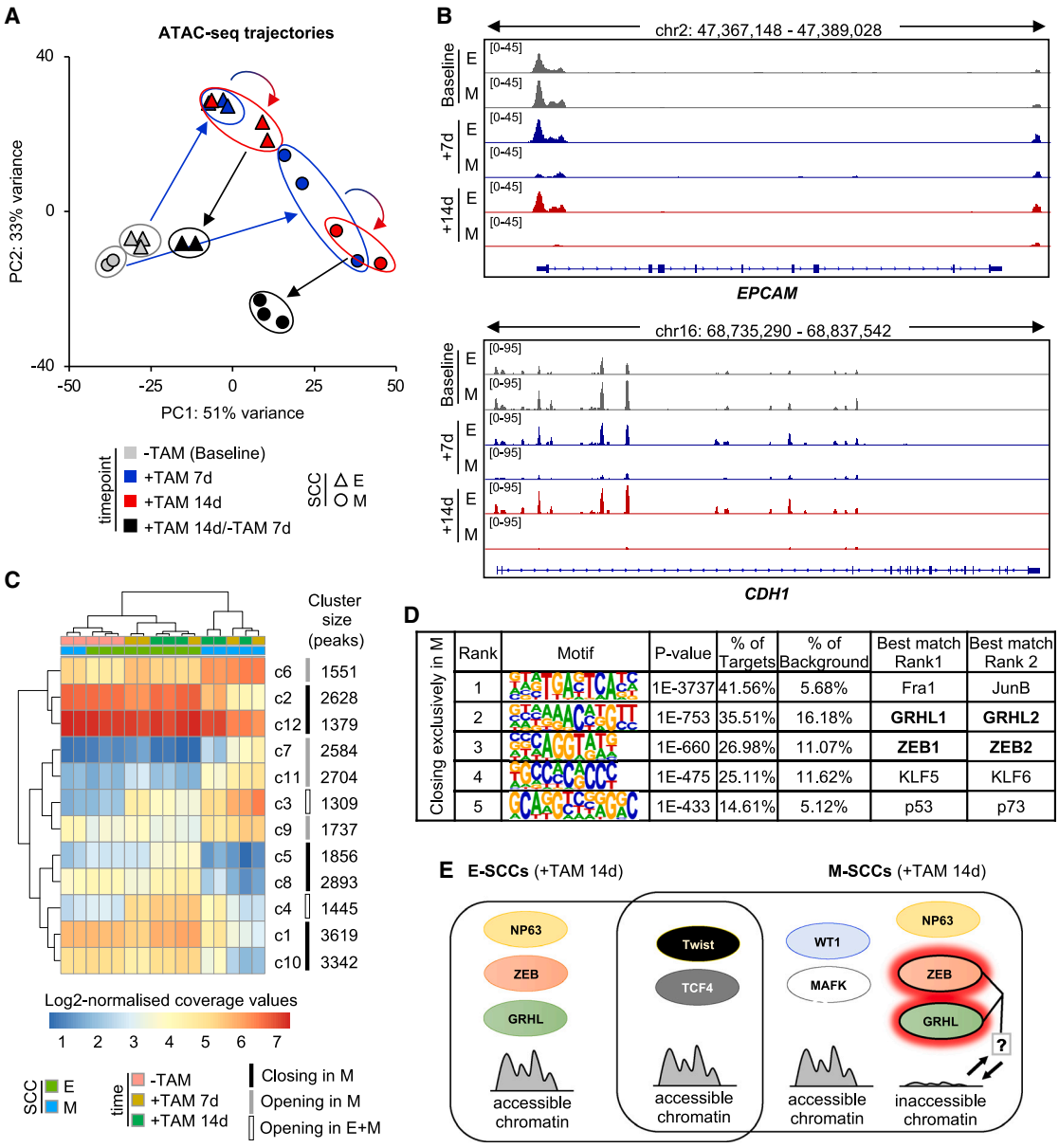
collagen gel as individual spindle-like cells (Figures 3E–3G). Intriguingly, the organoids generated by E-SCCs were epithelial in nature, highly proliferative, and branched, displaying signs of collective invasion (Figure 3G). Since HMLE cells do not grow in mice, we interrogated the SCCs for their lung-colonizing potential by reconstituting the three-dimensional architecture of lung tissue *ex vivo*.[51] For this purpose, precision-cut murine lung slices (PCLSs) were incubated for 5 days in TAM-free conditions with GFP-labeled E-SCCs or M-SCCs that were either untreated or treated with TAM for 14 days. Strikingly, untreated E-SCCs were 1.6-fold more efficient in colonizing the PCLS (Figure 3H). Moreover, after +TAM 14 days, the ability of M-SCCs to proliferate in the lung microenvironment further declined, while E-SCCs became almost 4-fold more efficient than M-SCCs in colonizing PCLSs (Figure 3H). These data suggested that irreversible mesenchymal differentiation impairs the clonal growth of tumorigenic or metastatic colonies in tissues. Moreover, the data suggested that EMT resistance promotes clonogenicity in relevant models, consistent with the enhanced metastatic potential of EPCAM(high) cells from human MBCs.

Global chromatin changes upon transient stimulation define EMT resistance vs. susceptibility

Next, we investigated the epigenetic basis for EMT resistance vs. susceptibility. To do so, we performed assay for transposase-accessible chromatin (ATAC) sequencing in conjunction with RNA sequencing in E-SCCs and M-SCCs before TAM stimulation (−TAM) at the +TAM7d and +TAM14d time points (EMT induction phase) and after TAM discontinuation at the +TAM14d/−TAM 7d time point (EMT deactivation phase) (Figures 4A and S4A). Principal-component analysis (PCA) and differential peak analysis suggested a similar landscape of chromatin accessibility in E-SCCs and M-SCCs at baseline, which began to diverge upon TAM stimulation (Figure S4B). Strikingly, M-SCCs harbored the greatest variation in chromatin accessibility compared with baseline, with changes associated with PC1 at +7 and +14 days of EMT induction (Figure 4A). By contrast, E-SCCs developed changes more along PC2 and less on the PC1 axis at the same time points. Remarkably, upon deactivation of the EMT stimulus (+TAM 14 days/−TAM 7 days), changes in chromatin accessibility were exclusively reversed in E-SCCs, as they returned near to the baseline, but not in M-SCCs, which landed instead in a distinct state (Figure 4A). This global observation was further corroborated by focusing on the loci of epithelial genes like *EPCAM* and *CDH1*. Here, both loci were similarly accessible at baseline; however, during the EMT stimulus, access to both loci was lost in M-SCCs, but not in E-SCCs (Figure 4B), leading to persistently remodeled chromatin beyond EMT induction (Figure S4C). Thus, the dynamics of chromatin accessibility in E-SCCs and M-SCCs mirrored our observations on their divergent response, leading to irreversible EMT or rebound to an epithelial state. Furthermore, ATAC-seq provided indication that the apparently EMT-resistant E-SCCs acquire a transient epigenetic state during EMT induction, coincident with the acquisition of a temporarily hybrid epithelial-mesenchymal phenotype (Figure 4A). Importantly, ATAC and RNA sequencing revealed comparable dynamics, suggesting that the chromatin accessibility results were reflecting specific differences in gene expression, with specific changes at the transcript level that could be traced back to chromatin peaks (Figure S4A and S4D). We then hypothesized that the EMT may convert M-SCCs to a novel and less tumorigenic transcriptional status by stabilizing specific chromatin rearrangements. Thus, we wished to pinpoint chromatin regions leadingly involved in key transitions by peak clustering of the ATAC sequencing (ATAC-seq) data. Unsupervised analysis returned peaks of chromatin regions that lost accessibility only in M-SCCs, encompassing repressed epithelial loci (Figure S4D; Table S1). Additionally, closed loci becoming accessible during EMT induction formed distinct peak clusters either only in M-SCCs or in both M- and E-SCCs (Figure 4C). Next, we sought binding sites of candidate TFs that better explained the divergent response to the EMT. Homer *de novo* TF motif analysis of combined clusters highlighted a set of binding motifs per each cluster class (Figures 4D and S4E), including motifs that were previously found enriched in cells undergoing EMT *in vivo*.[26] For example, *TP53/TP73* binding motifs were enriched in the cluster repressed in M-SCCs, an already reported pattern pointing toward the epithelial lineage-supportive ΔNp63 protein.[24] Predictably, binding motifs for "Twist" were enriched during EMT induction in loci commonly accessible to both E-SCCs and M-SCCs, together with binding motifs for interactors of TWIST1, such as TCF4,[52] supporting our observation that Twist1 is similarly active in both E-SCCs and M-SCCs. Next, we focused on TFs recruited downstream of Twist1, starting from motifs that persistently enriched in M-SCCs even after EMT stimulus cessation, i.e., marking irreversible EMT. Relevant hits included the tumor suppressor gene and mesenchymal gene activator *WT1*,[53] whose motifs enriched in the loci that became open after the EMT (Figure S4E), and Grainyhead-like factors (*GRHL1/2/3*), whose motifs were enriched in the permanently repressed loci, together with the binding motifs of zinc finger E-box-binding homeobox (ZEB) factors (Figure 4D), a class of well-known EMT-TFs. Importantly, ZEB motifs within closing chromatin regions were highly suggestive of ZEB1/2-mediated repression, a process known to occur genome wide during the EMT.[54] Moreover, GRHL factors such as GRHL2 are known to prime epithelial enhancers for transcriptional activation[55] in antagonism to ZEB factors (Figure 4E). By comparing ATAC-seq peaks with RNA-seq results, we focused on TFs that were consistently expressed when their matched motifs were enriched. This approach confirmed that ZEB and GRHL TFs as the most relevant candidates behind irreversible EMT (Figure S4F). Taken together, these data reveal that global changes in chromatin accessibility and gene expression upon EMT induction are largely reversible in E-SCCs but not in M-SCCs, and the genomic regions that become permanently repressed in M-SCCs are marked by specific EMT-TF binding motifs. This suggested that even a transient EMT induction resulted in the onset of a *de novo* chromatin status in M-SCCs.

ZEB1-mediated mesenchymal reprogramming is restrained by GRHL2 in E-SCCs

We then explored the functional involvement of the TFs pinpointed by ATAC-seq. *ZEB1* and *GRHL2* were the most differentially expressed TFs of their gene families. Before TAM stimulation, ZEB1 mRNA levels were already 40 times more



Closing exclusively in M	Rank	Motif	P-value	% of Targets	% of Background	Best match Rank1	Best match Rank 2
	1		1E-3737	41.56%	5.68%	Fra1	JunB
	2		1E-753	35.51%	16.18%	**GRHL1**	**GRHL2**
	3		1E-660	26.98%	11.07%	**ZEB1**	**ZEB2**
	4		1E-475	25.11%	11.62%	KLF5	KLF6
	5		1E-433	14.61%	5.12%	p53	p73

Figure 4. Global chromatin changes upon transient stimulation define EMT resistance vs. susceptibility

(A) PCA of ATAC-seq data of individual E-SCCs (Δ) and M-SCCs (◯) (−TAM: n = 3 E-SCCs, n = 3 M-SCCs; +TAM 7d: n = 3 E-SCCs, n = 3 M-SCCs; +TAM14d: n = 3 E-SCCs, n = 2 M-SCCs; +TAM14d/−TAM7d: n = 2 E-SCCs, n = 3 M-SCCs). Each data point represents one SCC at the indicated time point.

(B) ATAC-seq profile of the entire locus of *EPCAM* and *CDH1* in one representative M-SCC and one E-SCC.

(C) Heatmap of ATAC-seq peak clustering, highlighting differential chromatin accessibility for 12 clusters in E-SCCs and M-SCCs; color scale indicates log2-normalized coverage of the cluster center.

(D) Top 5 hits of Homer *de novo* transcription factor motif analysis in peak clusters exclusively closing in M-SCCs during TAM treatment.

(E) Graphical abstract of relevant TFs among statistically significant hits obtained from Homer motif analysis.

abundant in M-SCCs over E-SCCs, while *GRHL2* was higher in E-SCCs. During TAM stimulation, *ZEB1* was upregulated 3 times more in M-SCCs than in E-SCCs, while *GRHL2* was further downregulated in M-SCCs (200 to almost 600 times at +14 days) (Figure S5A). ZEB1 induction in M-SCCs was mirrored by the repression of direct targets such as *OVOL2*, *ESRP1*, *CDH1*, and *MIR200* family members, which include factors involved in positive and negative feedback loops with *ZEB1*,[56–58] such as *GRHL2* itself[59,60] (Figure 5A). Progressive accumulation of the ZEB1 protein was observable in M-SCCs during TAM stimulation (Figure 5B). After TAM discontinuation (+14d/−7d), stabilization of *ZEB1* occurred along with repression of targets such as OVOL2 (Figures S5B–S5D), accompanied by chromatin shutdown at the promoter region

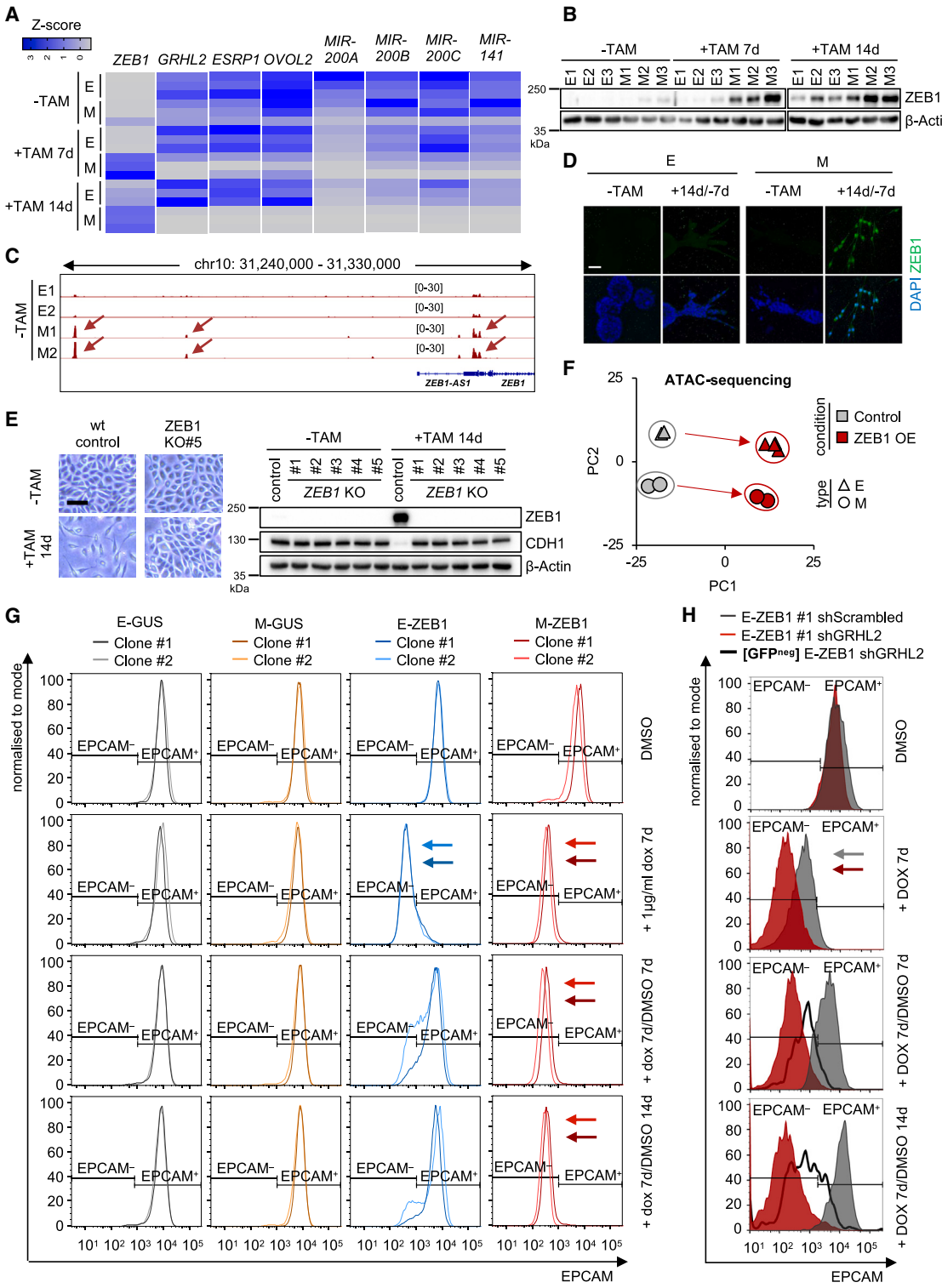


Figure 5. ZEB1-mediated mesenchymal reprogramming is restrained by GRHL2 in E-SCCs

(A) Heatmap of relative mRNA expression of the indicated genes in 3 individual E-SCCs (E) and 3 M-SCCs (M); n = 3 per clone; mean values.

(B) Immunoblot of ZEB1 in 3 individual E-SCCs (E1–E3) and 3 M-SCCs (M1–M3).

(C) ATAC-seq profile of the *ZEB1* promoter region (distal and proximal) in two representative E-SCCs (E1–E2) and M-SCCs (M1–M2) prior to TAM treatment. Arrows indicate peaks of differentially accessible DNA.

(legend continued on next page)

(Figure S5E). By contrast, TAM stimulation of E-SCCs produced limited and rapidly reversed induction of ZEB1, with no impact on expression or chromatin state of ZEB1 targets, suggesting key differences in the regulatory networks of ZEB1. Importantly, chromatin at the *ZEB1* promoter was in an open state in M-SCCs before induction but not in E-SCCs (Figure 5C). It was reported that poised chromatin at the *ZEB1* promoter can raise the proclivity of breast cancer cells to EMT induction.[61] Furthermore, accumulation of ZEB1 became noticeable in the nuclei of growth-arrested M-SCCs in 3D cultures (Figure 5D), suggesting a role in the irreversible EMT. However, prolonged TAM stimulation did not result in the repression of ZEB1 targets in E-SCCs (Figures S5B–S5D), so we tested if *ZEB1* regulation can determine EMT permissiveness. First, we generated M-SCC *ZEB1* knockout (KO) clones (Figure S5F and S5G). We observed that, in contrast to control M-SCCs, *ZEB1* KO M-SCCs failed to undergo EMT upon TAM induction, as evidenced by preservation of epithelial morphology (Figures 5E and S5H) and epithelial marker expression (Figures S5I and S5J). This demonstrated that *ZEB1* is necessary for the EMT and that its induction determines M-SCC susceptibility to EMT-inducing stimuli. To address if *ZEB1* induction was sufficient to induce the EMT, we ectopically expressed full-length *ZEB1* in E-SCCs and M-SCCs using a Tet-inducible system (Figures S5K and S5L). After 7 days of conditional overexpression (OE), ZEB1 protein accumulated in the nuclei of both E-SCCs and M-SCCs (+doxycycline [DOX] 7 days) (Figure S5M). We also observed EMT in M- and E-SCCs, with complete loss of EPCAM and downregulation of ZEB1 direct targets in contrast to controls (Figures S5N–S5P). Chromatin immunoprecipitation (ChIP) confirmed direct engagement of ZEB1 on the promoter of its target genes in both E-SCCs and M-SCCs at +DOX7d (Figure S5Q), consistent with the repression of their transcripts (Figure S5O). Next, we tested whether ZEB1 OE could establish distinct epigenetic programs at the global level. We thus performed ATAC-seq after introducing *ZEB1* OE compared with controls in E-SCCs and M-SCCs. PCA revealed a similar trajectory in both clone types and no substantial differences in chromatin accessibility between DOX-treated E-SCCs and M-SCCs (Figure 5F). To explore downstream effects of the program imposed by transient ZEB1 OE, we next subjected E- and M-SCC clones to +DOX 7 days, then stopped ZEB1 expression for 14 days (+DOX7d/+DMSO14d). As expected, all M-SCCs transitioned to an EPCAM(neg) state, acquiring a mesenchymal state that was not reversed by DOX withdrawal (Figure 5G). These results suggested that ZEB1 induction was sufficient to cause an irreversible EMT in M-SCCs. However, transient ZEB1 failed to produce durable EMT in E-SCCs, as a stepwise recovery in EPCAM positivity was observed in E-SCCs during DOX withdrawal, producing heterogeneity for EPCAM reminiscent of metastatic samples (Figure 5G). In the EPCAM(pos) population arising, such plastic rebound was marked by reacquisition of epithelial morphology and regained expression of GRHL2 along with other epithelial markers (Figure S5R; data not shown). ChIP analysis of E-SCCs at +DOX7d/+DMSO14d compared with +DOX7d revealed a sudden drop in ZEB1 signal from its target genes, which was particularly noticeable for the promoter of GRHL2 (Figure S5Q). Discriminative analysis of the ectopic vs. endogenous *ZEB1* transcript revealed persistent induction of endogenous *ZEB1* only in EPCAM(neg) cells. This was further confirmed by ChIP analysis, which revealed preservation of the ZEB1 protein signal on the promoter of *GRHL2* and other target genes at the +DOX7d/+DMSO14d time point (Figures S5Q and S5R). Mechanistically, these results indicated that *ZEB1* triggered a self-sustained mesenchymal program in M-SCCs, whereas in E-SCCs a plastic rebound mechanism protected epithelial traits. A role for GRHL2 in this mechanism is supported by its specific pioneering function, enabling it to bind and open closed chromatin in epithelial loci.[55,62] Next, we tested the effects of GRHL2 release from *ZEB1* repression because its binding motifs were co-enriched with ZEB1 in the chromatin regions closed during irreversible EMT. To do so, we knocked down *GRHL2* by a short hairpin RNA (shRNA; shGRHL2) prior to ZEB1 OE induction. Although *GRHL2* silencing alone had no impact on *EPCAM* levels prior to EMT induction (Figure S5S) and did not cause EMT per se, specific targets of GRHL2 such as HER3 were downregulated upon knockdown (Figures S5S and S5T). Importantly, after transient induction of ZEB1 in E-SCCs (+DOX7d/+DMSO14d), only cells able to express GRHL2 could reexpress EPCAM and E-cadherin, showing that GRHL2 activity is required for the plastic epithelial rebound that mediates EMT resistance (Figures 5H and S5U). At the functional level, ZEB1 OE compromised the MS-forming capacity of shGRHL2 E-SCCs, but not that of control (shScrambled) E-SCCs (Figure S5V), phenocopying the loss of MS replating potential that we observed in M-SCCs after irreversible EMT (Figures S3H and S3G). Overall, loss of GRHL2 in E-SCCs recapitulated major features of M-SCCs, like stabilization of endogenous *ZEB1* together with a gain of permanent mesenchymal features upon EMT induction. Together, these data showed that ZEB1 and GRHL2 are mechanistically involved in controlling resistance vs. susceptibility to the EMT. Prompt *ZEB1* induction and permanent repression of GRHL2 were favored in M-SCCs and, together, enabled ZEB1-directed cellular reprogramming toward a permanent mesenchymal state. Thus, the divergent response to the EMT depended on the proclivity to strong, self-sustained

(D) IF staining for ZEB1 in organotypic cultures of one representative E-SCC and one M-SCC. Scale bar: 100 μm.
(E) Bright-field images of one representative M-SCC *ZEB1* knockout clone and one M-SCC control clone in two-dimensional culture (left panel); scale bar: 100 μm; immunoblot of an M-SCC control clone and of a series of M-SCC *ZEB1* knockout clones (right panel).
(F) PCA of ATAC-seq data of E-SCCs (Δ) and M-SCCs (○) overexpressing a control *GUS* gene (n = 2 E-SCCs, n = 2 M-SCCs) or *ZEB1* (n = 4 E-SCCs, n = 2 M-SCCs) treated with DOX for 7 days. Each data point represents one SCC.
(G) Flow cytometry of EPCAM in individual E-SCC and M-SCC clones overexpressing a control *GUS* gene (representative of n = 2 E-SCCs and n = 2 M-SCCs) or *ZEB1* (representative of n = 4 E-SCCs and n = 2 M-SCCs); arrows indicate relevant differences.
(H) Flow cytometry of EPCAM in a representative E-SCC overexpressing *ZEB1* and transduced with shGRHL2 or shScrambled; representative of n = 2 E-SCCs.

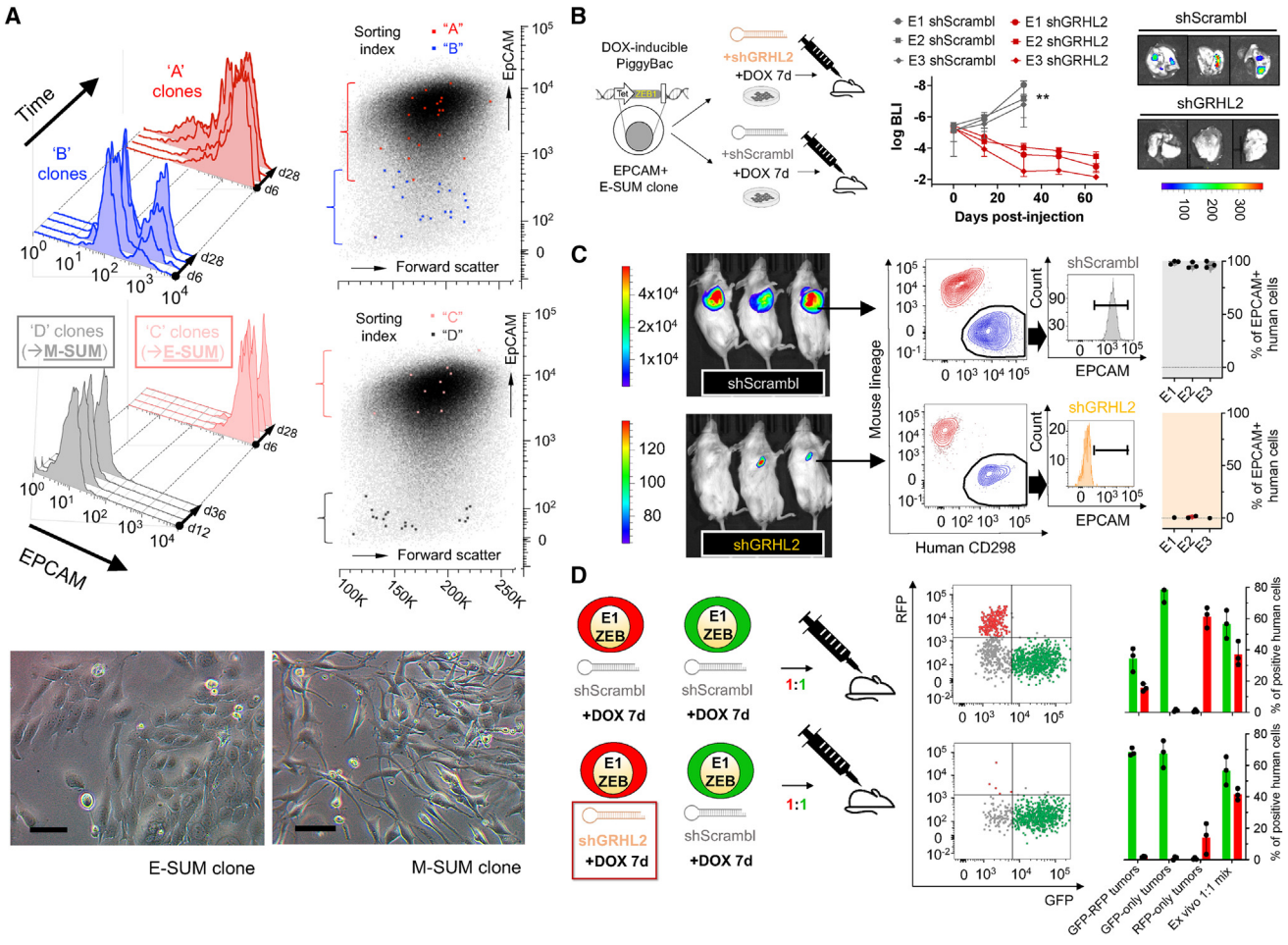


Figure 6. Clonal propagation *in vivo* requires resistance to EMT

(A) (Top) flow cytometry of EPCAM in individual single-cell clones isolated by index sorting from the SUM-149PT line; (bottom) bright-field images of two-dimensional cultures from one representative E-SUM (E1) and one M-SUM (M1) clone; scale bar: 50 μm.

(B) (Left) experimental outline for combined ZEB1/GRHL2 perturbation; (middle) BLI assessment over 9 weeks after MFP injection; (right) BLI visualization of spontaneous lung metastases in representative mice sacrificed after 9 weeks; n = 3 mice per group; mean ± SEM, unpaired Student's t test with Welch's correction; **p = 0.0097; rainbow scale: photons/s/cm^2/sr.

(C) Flow cytometry for EPCAM on human cells *ex vivo,* isolated from dissociated MFP tissue 9 weeks post-injection as referred to in (B); BLI visualization at the experimental endpoint.

(D) (Left) labeling strategy on clones from (B); (right) flow cytometry for GFP and RFP on human cells *ex vivo,* obtained as in (C) by the indicated 1:1 combinations and relative controls.

induction of endogenous *ZEB1* (as in M-SCCs) or the potential to maintain *GRHL2* expression over *ZEB1* (as in E-SCCs) in response to EMT-inducing stimuli.

Clonal propagation *in vivo* requires resistance to EMT

We then asked whether GRHL2 and ZEB1 are important for clonal propagation dynamics and human breast cancer ITH. For this, we used the SUM-149PT cell line, searching for individual clones with features of EMT resistance or mesenchymal reprogramming. By FACS, we sorted 40 individual EPCAM(neg) and 29 EPCAM(pos) cells, followed by clonal expansion *in vitro.* We recorded the original EPCAM level of each SCC and followed its changes throughout downstream events. Since all SUM-149PT cells are naturally subjected to spontaneous EMT events,[41] we assessed EPCAM in each clone over more than 2 weeks and thus identified recurring patterns of phenotypic transition (Figure 6A). The majority of clones (40 out of 66) progressively differentiated from their original EPCAM status. These included EPCAM(pos) cells that regenerated a bimodal EPCAM distribution, thus producing a stable progeny that resembled the parental population (pattern "A"). Clones originally mapping to the EPCAM(low) and EPCAM(neg) area progressively recovered full EPCAM expression over time once isolated (pattern "B"). Interestingly, 10 clones maintained full unimodal expression of EPCAM over 3 weeks (pattern "C"). We referred to the latter as E-SUM cells, as they preserved stable epithelial traits (Figure S6A). Conversely, and in line with the occurrence of an irreversible EMT, more than half of the sorted EPCAM(neg) clones

produced a highly mesenchymal progeny that preserved the phenotype for over 3 weeks (pattern "D"). We referred to them as M-SUM clones, and they were marked by persistent induction of ZEB1 (Figure 6A) as opposed to the parental cell line, where ZEB1 was undetectable (Figure S6A). These data show that clones following divergent EMT paths arise in polyclonal cancer populations, suggesting that sensitivity and resistance to the EMT may play a role in ITH. To test if such diversity reflects a different proclivity to EMT, we challenged E-SUM cells with ZEB1 OE and GRHL2 knockdown (Figure S6B). E-SUM showed resistance to mesenchymal reprogramming upon transient ZEB1 induction, as evidenced by full recovery of EPCAM expression after treatment with +DOX 7 days/+DMSO 14 days. As in HMLE cells, the EMT resistance of E-SUM cells could be overridden by concomitant GRHL2 knockdown, resulting in stable transition to an EPCAM(neg) state in response to ZEB1 OE (Figure S6C), thus phenocopying the irreversible EMT portrayed by M-SUM clones at the molecular level (Figure S6D). Next, we aimed to understand the relevance of irreversible EMT *in vivo*. We observed that E-SUM clones were tumorigenic and spontaneously metastatic in orthotopically transplanted mice. However, M-SUM failed to generate solid lesions *in vivo* and consistently failed to form multicellular proliferative structures in organoid assays (Figures S6E and S6F), while both clonal types could proliferate indefinitely in two dimensions (Figures S6G and S6H). Intriguingly, M-SUM cells survived for at least 2 weeks in the injected MFP as living, bioluminescent cells, without producing any visible solid lesion, suggesting growth arrest (Figure S6E). To further explore this possibility, we mimicked an irreversible EMT in E-SUM clones by combining ZEB1 OE together with GRHL2 silencing. Then, we orthotopically injected shScrambled and shGRHL2 E-SCCs after +DOX 7 days and monitored the bioluminescent signals in both groups over 9 weeks (Figure 6B). Strikingly, irreversible EMT prevented every E-SUM clone we tested from producing solid tumors *in vivo* and spreading metastasis; however, it did not prevent long-term survival of the cells in the injected MFP (Figure 6C) nor did it suppressed their proliferative potential in culture (Figure S6I). At the experimental endpoint, remaining human cells could be isolated, stained for EPCAM, and analyzed by flow cytometry, where they appeared "locked" in an EPCAM(neg) state (Figure 6C). In summary, our data confirm that mesenchymal reprogramming prevented clonal propagation in an *in vivo* environment, impeding proliferation but preserving long-term survival. To verify if the latter observation was reversible in polyclonal contexts, we tested if inter-clonal crosstalk between an EPCAM(neg) and an EPCAM(pos) clone could rescue the tumorigenic features of the former. To this aim, we co-injected mice orthotopically with 1:1 mixtures of shScrambled E-SUM cells combined with matched E-SUM cells that were forced in an EPCAM(neg) state or with another (shScrambled) control clone. We tracked each co-transplanted group by a dual color (GFP vs. RFP) labeling strategy (Figures 6D and S6J). In parallel, we co-cultivated in three-dimensional organotypic conditions control E-SUM cells together with E-SUM cells that had undergone irreversible EMT. Consistent with the results obtained from single clones, we found that control (shScrambled) GFP-expressing E cells predominantly contributed to the mixed tumors and three-dimensional organoid structures emerging (Figures 6D and S6K). No metastasis nor organoid was progeny of RFP-expressing EPCAM(neg) cells, as the shGRHL2 clones did not contribute to multicellular tumors and organoids in a polyclonal context, suggesting that these were entirely outcompeted by EPCAM(high) cells. Altogether, we conclude that irreversible EMT leads to loss of clonal propagation, which is not reverted by interactions with EMT-resistant clones *in vivo.* To validate these findings in patients, we assessed *ZEB1* induction in single cells from clinical biopsies. Single-cell transcriptional analysis of EPCAM(low) and EPCAM(neg) cells in metastatic effusions of two patients with MBC captured a group of individual carcinoma cells displaying 3- to 4-fold higher *ZEB1* expression, compared with EPCAM(high) cells, and lower *CDH1* levels (Figure S6L), suggesting that M-SCC-like, EMT-susceptible clones reside in the EPCAM(low) fraction, which might explain their reduced performance in tumorigenicity assays. In conclusion, our data reveal a novel layer of ITH in breast cancer, marked by differential resilience to mesenchymal reprogramming and determined by the capacity of plastic rebound to an epithelial state in response to EMT stimuli. Individual clones with a reduced *ZEB1* to *GRHL2* mRNA ratio are more prone to epithelial lineage-preserving plasticity in response to EMT induction. This mechanism sustains tumorigenic and metastatic outgrowth by preventing a mesenchymal chromatin remodeling program, whereas tumor clones unable to prevent it become growth arrested. In conclusion, we identified an epigenetic barrier to irreversible EMT and concomitant suppression of growth and epithelial features in breast cancer cells and show its regulation by a critical balance between TFs involved in chromatin reprogramming.

DISCUSSION

Here, we provide in-depth characterization of E and M phenotypes arising in human MBC, focusing on the effects of the EMT on tumorigenicity. We present results from patient cells and human models that align with common clinical histopathology observations, suggesting little evidence of EMT in solid MBC lesions.[33] In liquid biopsies from ER^+ and ER^- MBCs, we describe the EMT as a major cause of ITH, impacting the potential to propagate the metastatic disease in xenografts. At the mechanistic level, only EPCAM-expressing clones retain tumorigenicity and epithelial plasticity in three-dimensional assays. In line with others, we show that a transient or incomplete EMT can confer crucial advantages for cancer metastasis, such as the ability to collectively invade.[30,31] For the first time, we additionally show that such functional effects appear dependent on the retention of epithelial traits, while stable mesenchymal *trans*-differentiation leads to permanent functional impairment. Importantly, we do not exclude that disease-propagating epithelial clones may acquire the expression of additional mesenchymal traits at least in part, as shown previously.[26] Instead, our results indicate that sensitivity to an irreversible EMT results from the imbalance between ZEB1 and its target GRHL2 at the clonal level, producing global effects. Distinct chromatin and transcriptional landscapes of E and M clones explain how EMT stimuli can impose distinct trajectories, marked by differentially accessible chromatin regions enriched in binding sites

for both GRHL2 and ZEB1. These regions include *cis* elements governing well-known regulators of epithelial cell fate that are actively repressed by ZEB1.[56,57,63] Intriguingly, it was shown that ZEB1 can co-opt multiple genomic regions to maintain mesenchymal differentiation in human cells,[54] while GRHL2 functions as a pioneering factor to prime epithelial genes.[55] In line with our findings, regulation of GRHL2 has been described within a larger network of feedback loops crucial for the fine-tuning of oncogenic signaling and invasiveness in breast cancer cells undergoing an EMT.[59,60,64,65] We show that high GRHL2 expression over ZEB1 is crucial to preserve a global epigenetic phenotype enabling tumor propagation. Our evidence for co-enrichment of GRHL2 and ZEB1 motifs in the same chromatin regions is highly suggestive of TF competition for the epigenetic control of epithelial traits. Remarkably, co-transplantation of E and M clones could not rescue the cell-autonomous defect produced by an irreversible EMT, causing an enduring disadvantage to M clones for *in vivo* outgrowth. Moreover, it appears that metastatic outgrowth is directly connected to the strong epithelial identity of EPCAM(high) carcinoma cells, whereas this capacity appears suppressed in EPCAM(low) cells from liquid biopsies, even if these can still regenerate MFP tumors with low efficiency. Considering the comparable mutational profile we observed in the two subsets, we speculate that stronger sensitivity to the EMT of EPCAM(low) cells may trigger durable loss of metastatic propagation through epigenetic regulation. In support of this hypothesis, lung metastases could only be found in patient-derived tumors generated from epithelial EPCAM(high) cells, even though we cannot rule out that rare cells may have undergone an EMT event in the mouse, and marginally contributed to metastasis. Together, these data suggest that the prevalent propagation of EMT-resistant clones may be a physiological phenomenon that proceeds along with the natural spread of MBC. In line with this hypothesis, a direct role for epithelial cells in breast cancer metastasis has been highlighted by multiple independent discoveries in the last decade. Circulating tumor cells (CTCs) in the blood of patients with breast cancer expressing EPCAM can be transplanted to seed metastatic tumors *in vivo*.[44] Furthermore, clusters of CTCs held together by epithelial junctions are far more metastatic than individual CTCs,[66] and perturbations leading to the dissolution of such epithelial anchors can abrogate most of their metastatic potential.[67] Consistently, E-cadherin^{+} epithelial cells are efficient metastatic precursors in multiple models of breast cancer,[68] and the degree of the EMT, from partial to complete, correlates with a decreasing contribution to mammary cancer metastasis in mice.[29] Albeit EPCAM alone cannot capture the complex spectrum of intermediate EMT states, we show for the first time that its expression is sufficient to resolve functionally relevant heterogeneity and to discriminate EMT-resistant clones within liquid biopsies. Our results go beyond observations in support of individual epithelial genes, as we reveal that epithelial lineage identity is actively preserved by a global epigenetic program enabling disease propagation. Cells harboring such a program appear overrepresented in metastatic specimens, so EPCAM could be useful to detect the most competitive disease-propagating clones in liquid biopsies, although its role in solid tumors and in the non-metastatic disease warrants further investigation.

Limitations of the study

We have conducted a pilot study in a limited number of metastatic specimens and experimental models. Our findings need to be validated with a larger cohort of patient samples spanning more clinical subtypes, including samples collected at a less advanced disease stage. All patients within this study were previously treated with chemotherapy; as it was shown that mesenchymal cells can display chemoresistance,[16,17,69] the detection of EPCAM(low) and -(neg) cells in metastatic biopsies could be explained by their positive selection during chemotherapy, in line with their reduced proliferation. As a limitation of this study, clonal mechanisms of tumor evasion from irreversible EMT that are unseen in our models might emerge under clinical circumstances such as anti-neoplastic treatments. Focusing on this aspect is warranted to shed light on the diverse roles assumed by the EMT in human breast cancer. Of note, we were not able to find CTCs in the blood of the xenografted mice, a limitation that could be explained by the low disseminating potential of first-passage xenografts *in vivo*. Moreover, even if lung metastases were not observable in mice engrafted by EPCAM(low) and EPCAM(neg) cells, we did not search for dissemination to the bones and other metastatic organs where tumor dormancy could occur. Overall, the maintenance of elevated EPCAM expression appeared to be tightly linked to tumor growth and metastasis in liquid biopsies; however, EPCAM expression in solid lesions is not sufficient to confer metastatic traits, and the precise mechanisms by which loss of EPCAM through an EMT leads to the disappearance of metastatic traits remain to be determined in greater detail.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - MBC liquid biopsies
  - Cell cultures
- METHOD DETAILS
  - Xenograft assay for breast cancer cells
  - Immunohistochemical analysis of MBC xenografts
  - Targeted sequencing of somatic mutations
  - RNA-sequencing of MBC cells
  - Single-cell gene expression profiling of MBC cells
  - Bidimensional (2D) cell cultures
  - Single-cell cloning and MS assay
  - Organotypic cultures in 3D collagen gels
  - Colonizing assay on murine PCLS
  - CRISPR-Cas9 mediated knockout of ZEB1

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112533.

ACKNOWLEDGMENTS

We thank all members of the ETH Hönggerberg Flow Cytometry Core Facility and the DKFZ Heidelberg Flow Cytometry Core Facility for their kind assistance with cell sorting. We thank all members of the DKFZ Animal Laboratory Core Facility for excellent animal welfare and husbandry. We are grateful to Robert Weinberg (Whitehead Institute) for providing the HMLE-Twist1-ER cells and to Roland Rad (Technical University Munich) for providing the hyperactive transposase for the PiggyBac system. We also thank Christopher Breunig and Stefan Stricker (Helmholtz Center Munich) for sharing reagents and guidance for STAgR cloning and Sandy Lösecke for technical assistance during RNA-seq. We thank all members of the Scheel group at the Helmholtz Center Munich for sharing experimental expertise. This work was supported by the Transluminal-B and Integrate-TN consortia funded by the German Cancer Aid Organisation (Translational Oncology Program, Project 701134509), the Swiss Bridge Award, and the Dietmar Hopp Foundation (all to A.T.). This work was further supported by a Max Eder Grant of the German Cancer Aid Organisation (Deutsche Krebshilfe 110225) (to C.H.S.). M. Reichert acknowledges financial support by the German Cancer Aid Organisation (Max-Eder Program 111273 and Project 70114328) and the German Research Foundation (DFG, SFB1321 Modeling and Targeting Pancreatic Cancer, Projects S01 and TP12, Project ID329628492, and Project RE 3723/4-1). M. Reichert was supported by the Federal Ministry of Education and Research, SATURN3 "Spatial and Temporal Resolution of Intratumoral Heterogeneity in 3 hard-to-treat Cancers" (Project 01KD2206P). Work in the G.S. laboratory was funded by the Deutsche Forschungsgemeinschaft (DFG; German Research Foundation), Project-ID 213249687 – SFB 1064 and Project-ID 329628492 – SFB 1321 TP13. The laboratory of N.A. is supported by the European Research Council (101001652), the strategic focus area of Personalized Health and Related Technologies at ETH Zurich (PHRT-541), the Future and Emerging Technologies programme of the European Commission (801159-B2B), the Swiss National Science Foundation (212183), the Swiss Cancer League (KLS-4834-08-2019), the ETH Lymphoma Challenge (LC-02-22), and the ETH Zürich.

AUTHOR CONTRIBUTIONS

M. Saini, L.S., and C.H.S. conceived the project and designed experiments. C.H.S., A.T., M.R.S., M. Reichert, and N.A. supervised the work. M. Saini, L.S., H.D.M., E.D., C.K., J.M.B., M. Reitberger, V.V., S.H., N.P., and R.W. performed research and analyzed the data. M.F., E.E., R.W., E.D., and S.H. analyzed the data, provided crucial resources, and contributed significantly to data interpretation. H.D.M. and G.S. performed ATAC-seq and data analysis. E.G. and T.-M.S. performed RNA-seq. T.S. analyzed RNA-seq data. M.L. and M.K. provided lung slices and the PCLS assay. A.S., M. Sütterlin, and S.S. provided patient samples and supervised and managed clinical data.W.W. supervised histopathological analysis of patient-derived xenografts. M. Saini and L.S. assembled the figures. M. Saini, L.S., and C.H.S. wrote the manuscript. All authors read and approved the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: January 19, 2022
Revised: October 4, 2022
Accepted: May 3, 2023
Published: May 30, 2023

REFERENCES

1. ACS (2019). Breast Cancer Facts & Figures 2019-2020 (American Cancer Society, Inc).
2. Klein, C.A. (2020). Cancer progression and the invisible phase of metastatic colonization. Nat. Rev. Cancer *20*, 681–694. https://doi.org/10.1038/s41568-020-00300-6.
3. Riggio, A.I., Varley, K.E., and Welm, A.L. (2021). The lingering mysteries of metastatic recurrence in breast cancer. Br. J. Cancer *124*, 13–26. https://doi.org/10.1038/s41416-020-01161-4.
4. Ishay-Ronen, D., Diepenbruck, M., Kalathur, R.K.R., Sugiyama, N., Tiede, S., Ivanek, R., Bantug, G., Morini, M.F., Wang, J., Hess, C., and Christofori, G. (2019). Gain fat-lose metastasis: converting invasive breast cancer cells into adipocytes inhibits cancer metastasis. Cancer Cell *35*, 17–32.e6. https://doi.org/10.1016/j.ccell.2018.12.002.
5. Koren, S., Reavie, L., Couto, J.P., De Silva, D., Stadler, M.B., Roloff, T., Britschgi, A., Eichlisberger, T., Kohler, H., Aina, O., et al. (2015). PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. Nature *525*, 114–118. https://doi.org/10.1038/nature14669.
6. Van Keymeulen, A., Lee, M.Y., Ousset, M., Brohée, S., Rorive, S., Giraddi, R.R., Wuidart, A., Bouvencourt, G., Dubois, C., Salmon, I., et al. (2015). Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. Nature *525*, 119–123. https://doi.org/10.1038/nature14665.
7. Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G., et al.; EMT International Association TEMTIA (2020). Guidelines and definitions for research on epithelial-mesenchymal transition. Nat. Rev. Mol. Cell Biol. *21*, 341–352. https://doi.org/10.1038/s41580-020-0237-9.
8. Brabletz, T., Kalluri, R., Nieto, M.A., and Weinberg, R.A. (2018). EMT in cancer. Nat. Rev. Cancer *18*, 128–134. https://doi.org/10.1038/nrc.2017.118.
9. Derynck, R., and Weinberg, R.A. (2019). EMT and cancer: more than meets the eye. Dev. Cell *49*, 313–316. https://doi.org/10.1016/j.devcel.2019.04.026.
10. Pastushenko, I., and Blanpain, C. (2019). EMT transition states during tumor progression and metastasis. Trends Cell Biol. *29*, 212–226. https://doi.org/10.1016/j.tcb.2018.12.001.
11. Scheel, C., Eaton, E.N., Li, S.H.J., Chaffer, C.L., Reinhardt, F., Kah, K.J., Bell, G., Guo, W., Rubin, J., Richardson, A.L., and Weinberg, R.A. (2011). Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast. Cell *145*, 926–940. https://doi.org/10.1016/j.cell.2011.04.029.
12. Scheel, C., and Weinberg, R.A. (2011). Phenotypic plasticity and epithelial-mesenchymal transitions in cancer and normal stem cells? Int. J. Cancer *129*, 2310–2314. https://doi.org/10.1002/ijc.26311.
13. Chaffer, C.L., San Juan, B.P., Lim, E., and Weinberg, R.A. (2016). EMT, cell plasticity and metastasis. Cancer Metastasis Rev. *35*, 645–654. https://doi.org/10.1007/s10555-016-9648-7.
14. Diepenbruck, M., and Christofori, G. (2016). Epithelial-mesenchymal transition (EMT) and metastasis: yes, no, maybe? Curr. Opin. Cell Biol. *43*, 7–13. https://doi.org/10.1016/j.ceb.2016.06.002.

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

16. Zheng, X., Carstens, J.L., Kim, J., Scheible, M., Kaye, J., Sugimoto, H., Wu, C.C., LeBleu, V.S., and Kalluri, R. (2015). Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature *527*, 525–530. https://doi.org/10.1038/nature16064.

17. Fischer, K.R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S.T.C., Choi, H., El Rayes, T., Ryu, S., Troeger, J., et al. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature *527*, 472–476. https://doi.org/10.1038/nature15748.

18. Krebs, A.M., Mitschke, J., Lasierra Losada, M., Schmalhofer, O., Boerries, M., Busch, H., Boettcher, M., Mougiakakos, D., Reichardt, W., Bronsert, P., et al. (2017). The EMT-activator Zeb1 is a key factor for cell plasticity and promotes metastasis in pancreatic cancer. Nat. Cell Biol. *19*, 518–529. https://doi.org/10.1038/ncb3513.

19. Ni, T., Li, X.Y., Lu, N., An, T., Liu, Z.P., Fu, R., Lv, W.C., Zhang, Y.W., Xu, X.J., Grant Rowe, R., et al. (2016). Snail1-dependent p53 repression regulates expansion and activity of tumour-initiating cells in breast cancer. Nat. Cell Biol. *18*, 1221–1232. https://doi.org/10.1038/ncb3425.

20. Aiello, N.M., Brabletz, T., Kang, Y., Nieto, M.A., Weinberg, R.A., and Stanger, B.Z. (2017). Upholding a role for EMT in pancreatic cancer metastasis. Nature *547*, E7–E8. https://doi.org/10.1038/nature22963.

21. Caramel, J., Ligier, M., and Puisieux, A. (2018). Pleiotropic roles for ZEB1 in cancer. Cancer Res. *78*, 30–35. https://doi.org/10.1158/0008-5472.CAN-17-2476.

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

23. Xu, Y., Lee, D.K., Feng, Z., Xu, Y., Bu, W., Li, Y., Liao, L., and Xu, J. (2017 Oct 24). Breast tumor cell-specific knockout of *Twist1* inhibits cancer cell plasticity, dissemination, and lung metastasis in mice. Proc. Natl. Acad. Sci. USA *114*, 11494–11499. https://doi.org/10.1073/pnas.1618091114.

24. Latil, M., Nassar, D., Beck, B., Boumahdi, S., Wang, L., Brisebarre, A., Dubois, C., Nkusi, E., Lenglez, S., Checinska, A., et al. (2017). Cell-type-specific chromatin states differentially prime squamous cell carcinoma tumor-initiating cells for epithelial to mesenchymal transition. Cell Stem Cell *20*, 191–204.e5. https://doi.org/10.1016/j.stem.2016.10.018.

25. Sikandar, S.S., Kuo, A.H., Kalisky, T., Cai, S., Zabala, M., Hsieh, R.W., Lobo, N.A., Scheeren, F.A., Sim, S., Qian, D., et al. (2017). Role of epithelial to mesenchymal transition associated genes in mammary gland regeneration and breast tumorigenesis. Nat. Commun. *8*, 1669. https://doi.org/10.1038/s41467-017-01666-2.

26. Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., et al. (2018). Identification of the tumour transition states occurring during EMT. Nature *556*, 463–468. https://doi.org/10.1038/s41586-018-0040-3.

27. Kröger, C., Afeyan, A., Mraz, J., Eaton, E.N., Reinhardt, F., Khodor, Y.L., Thiru, P., Bierie, B., Ye, X., Burge, C.B., and Weinberg, R.A. (2019). Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. Proc. Natl. Acad. Sci. USA *116*, 7353–7362. https://doi.org/10.1073/pnas.1812876116.

28. Li, Y., Lv, Z., Zhang, S., Wang, Z., He, L., Tang, M., Pu, W., Zhao, H., Zhang, Z., Shi, Q., et al. (2020 Sep 14). Genetic fate mapping of transient cell fate reveals N-cadherin activity and function in tumor metastasis. Dev. Cell *54*, 593–607.e5. https://doi.org/10.1016/j.devcel.2020.06.021.

29. Lüönd, F., Sugiyama, N., Bill, R., Bornes, L., Hager, C., Tang, F., Santacroce, N., Beisel, C., Ivanek, R., Bürglin, T., et al. (2021 Dec 6). Distinct contributions of partial and full EMT to breast cancer malignancy. Dev. Cell *56*, 3203–3221.e11. https://doi.org/10.1016/j.devcel.2021.11.006.

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

31. Schmidt, J.M., Panzilius, E., Bartsch, H.S., Irmler, M., Beckers, J., Kari, V., Linnemann, J.R., Dragoi, D., Hirschi, B., Kloos, U.J., et al. (2015). Stem-cell-like properties and epithelial plasticity arise as stable traits after transient Twist1 activation. Cell Rep. *10*, 131–139. https://doi.org/10.1016/j.celrep.2014.12.032.

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

33. Chui, M.H. (2013). Insights into cancer metastasis from a clinicopathologic perspective: epithelial-Mesenchymal Transition is not a necessary step. Int. J. Cancer *132*, 1487–1495. https://doi.org/10.1002/ijc.27745.

34. Kowalski, P.J., Rubin, M.A., and Kleer, C.G. (2003). E-cadherin expression in primary carcinomas of the breast and its distant metastases. Breast Cancer Res. *5*, R217–R222. https://doi.org/10.1186/bcr651.

35. DeRose, Y.S., Wang, G., Lin, Y.C., Bernard, P.S., Buys, S.S., Ebbert, M.T.W., Factor, R., Matsen, C., Milash, B.A., Nelson, E., et al. (2011). Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. Nat. Med. *17*, 1514–1520. https://doi.org/10.1038/nm.2454.

36. Cassidy, J.W., Caldas, C., and Bruna, A. (2015). Maintaining tumor heterogeneity in patient-derived tumor xenografts. Cancer Res. *75*, 2963–2968. https://doi.org/10.1158/0008-5472.CAN-15-0727.

37. Byrne, A.T., Alférez, D.G., Amant, F., Annibali, D., Arribas, J., Biankin, A.V., Bruna, A., Budinská, E., Caldas, C., Chang, D.K., et al. (2017). Interrogating open issues in cancer precision medicine with patient-derived xenografts. Nat. Rev. Cancer *17*, 254–268. https://doi.org/10.1038/nrc.2016.140.

38. Scheidmann, M.C., Castro-Giner, F., Strittmatter, K., Krol, I., Paasinen-Sohns, A., Scherrer, R., Donato, C., Gkountela, S., Szczerba, B.M., Diamantopoulou, Z., et al. (2022 Feb 15). An in vivo CRISPR screen identifies stepwise genetic dependencies of metastatic progression. Cancer Res. *82*, 681–694. https://doi.org/10.1158/0008-5472.CAN-21-3908.

39. Lat, T., and Paul, M. (2021). Malignant effusion. In StatPearls (StatPearls Publishing Copyright © 2021, StatPearls Publishing LLC.).

40. Ordóñez, N.G. (1999). Role of immunohistochemistry in differentiating epithelial mesothelioma from adenocarcinoma. Review and update. Am. J. Clin. Pathol. *112*, 75–89. https://doi.org/10.1093/ajcp/112.1.75.

41. Jordan, N.V., Prat, A., Abell, A.N., Zawistowski, J.S., Sciaky, N., Karginova, O.A., Zhou, B., Golitz, B.T., Perou, C.M., and Johnson, G.L. (2013). SWI/SNF chromatin-remodeling factor Smarcd3/Baf60c controls epithelial-mesenchymal transition by inducing Wnt5a signaling. Mol. Cell Biol. *33*, 3011–3025. https://doi.org/10.1128/MCB.01443-12.

42. Prat, A., Karginova, O., Parker, J.S., Fan, C., He, X., Bixby, L., Harrell, J.C., Roman, E., Adamo, B., Troester, M., and Perou, C.M. (2013). Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. Breast Cancer Res. Treat. *142*, 237–255. https://doi.org/10.1007/s10549-013-2743-3.

43. Risom, T., Langer, E.M., Chapman, M.P., Rantala, J., Fields, A.J., Boniface, C., Alvarez, M.J., Kendsersky, N.D., Pelz, C.R., Johnson-Camacho, K., et al. (2018). Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. Nat. Commun. *9*, 3815. https://doi.org/10.1038/s41467-018-05729-w.

44. Baccelli, I., Schneeweiss, A., Riethdorf, S., Stenzinger, A., Schillert, A., Vogel, V., Klein, C., Saini, M., Bäuerle, T., Wallwiener, M., et al. (2013). Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. Nat. Biotechnol. *31*, 539–544. https://doi.org/10.1038/nbt.2576.

45. Casas, E., Kim, J., Bendesky, A., Ohno-Machado, L., Wolfe, C.J., and Yang, J. (2011 Jan 1). Snail2 is an essential mediator of Twist1-induced epithelial mesenchymal transition and metastasis. Cancer Res. *71*, 245–254. https://doi.org/10.1158/0008-5472.CAN-10-2330.

46. Lourenço, A.R., Roukens, M.G., Seinstra, D., Frederiks, C.L., Pals, C.E., Vervoort, S.J., Margarido, A.S., van Rheenen, J., and Coffer, P.J. (2020). C/EBPa is crucial determinant of epithelial maintenance by preventing epithelial-to-mesenchymal transition. Nat. Commun. *11*, 785. https://doi.org/10.1038/s41467-020-14556-x.

47. Wagner, J., Masek, M., Jacobs, A., Soneson, C., Sivapatham, S., Damond, N., de Souza, N., Robinson, M.D., and Bodenmiller, B. (2022 Feb 9). Mass cytometric and transcriptomic profiling of epithelial-mesenchymal transitions in human mammary cell lines. Sci. Data *9*, 44. https://doi.org/10.1038/s41597-022-01137-4.

48. Dragoi, D., Krattenmacher, A., Mishra, V.K., Schmidt, J.M., Kloos, U.J., Meixner, L.K., Hauck, S.M., Buggenthin, F., Schwartz, D., Marr, C., et al. (2016 May 24). Twist1 induces distinct cell states depending on TGFBR1-activation. Oncotarget *7*, 30396–30407. https://doi.org/10.18632/oncotarget.8878.

49. Linnemann, J.R., Miura, H., Meixner, L.K., Irmler, M., Kloos, U.J., Hirschi, B., Bartsch, H.S., Sass, S., Beckers, J., Theis, F.J., et al. (2015). Quantification of regenerative potential in primary human mammary epithelial cells. Development *142*, 3239–3251. https://doi.org/10.1242/dev.123554.

50. Shi, J., Wang, Y., Zeng, L., Wu, Y., Deng, J., Zhang, Q., Lin, Y., Li, J., Kang, T., Tao, M., et al. (2014). Disrupting the interaction of BRD4 with diacetylated Twist suppresses tumorigenesis in basal-like breast cancer. Cancer Cell *25*, 210–225. https://doi.org/10.1016/j.ccr.2014.01.028.

51. Uhl, F.E., Vierkotten, S., Wagner, D.E., Burgstaller, G., Costa, R., Koch, I., Lindner, M., Meiners, S., Eickelberg, O., and Königshoff, M. (2015). Preclinical validation and imaging of Wnt-induced repair in human 3D lung tissue cultures. Eur. Respir. J. *46*, 1150–1166. https://doi.org/10.1183/09031936.00183214.

52. Huttlin, E.L., Bruckner, R.J., Paulo, J.A., Cannon, J.R., Ting, L., Baltier, K., Colby, G., Gebreab, F., Gygi, M.P., Parzen, H., et al. (2017). Architecture of the human interactome defines protein communities and disease networks. Nature *545*, 505–509. https://doi.org/10.1038/nature22366.

53. Martínez-Estrada, O.M., Lettice, L.A., Essafi, A., Guadix, J.A., Slight, J., Velecela, V., Hall, E., Reichmann, J., Devenney, P.S., Hohenstein, P., et al. (2010). Wt1 is required for cardiovascular progenitor cell formation through transcriptional control of Snail and E-cadherin. Nat. Genet. *42*, 89–93. https://doi.org/10.1038/ng.494.

54. Balestrieri, C., Alfarano, G., Milan, M., Tosi, V., Prosperini, E., Nicoli, P., Palamidessi, A., Scita, G., Diaferia, G.R., and Natoli, G. (2018). Co-Optation of tandem DNA repeats for the maintenance of mesenchymal identity. Cell *173*, 1150–1164.e14. https://doi.org/10.1016/j.cell.2018.03.081.

55. Jacobs, J., Atkins, M., Davie, K., Imrichova, H., Romanelli, L., Christiaens, V., Hulselmans, G., Potier, D., Wouters, J., Taskiran, I.I., et al. (2018). The transcription factor Grainy head primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes. Nat. Genet. *50*, 1011–1020. https://doi.org/10.1038/s41588-018-0140-x.

56. Hong, T., Watanabe, K., Ta, C.H., Villarreal-Ponce, A., Nie, Q., and Dai, X. (2015). An ovol2-zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. PLoS Comput. Biol. *11*, e1004569. https://doi.org/10.1371/journal.pcbi.1004569.

57. Brabletz, S., Bajdak, K., Meidhof, S., Burk, U., Niedermann, G., Firat, E., Wellner, U., Dimmler, A., Faller, G., Schubert, J., and Brabletz, T. (2011). The ZEB1/miR-200 feedback loop controls Notch signalling in cancer cells. EMBO J. *30*, 770–782. https://doi.org/10.1038/emboj.2010.349.

58. Sánchez-Tilló, E., Lázaro, A., Torrent, R., Cuatrecasas, M., Vaquero, E.C., Castells, A., Engel, P., and Postigo, A. (2010). ZEB1 represses E-cadherin and induces an EMT by recruiting the SWI/SNF chromatin-remodeling protein BRG1. Oncogene *29*, 3490–3500. https://doi.org/10.1038/onc.2010.102.

59. Cieply, B., Farris, J., Denvir, J., Ford, H.L., and Frisch, S.M. (2013). Epithelial-mesenchymal transition and tumor suppression are controlled by a reciprocal feedback loop between ZEB1 and Grainyhead-like-2. Cancer Res. *73*, 6299–6309. https://doi.org/10.1158/0008-5472.CAN-12-4082.

60. Mooney, S.M., Talebian, V., Jolly, M.K., Jia, D., Gromala, M., Levine, H., and McConkey, B.J. (2017 Sep). The GRHL2/ZEB feedback loop-A key Axis in the regulation of EMT in breast cancer. J. Cell. Biochem. *118*, 2559–2570.

61. Chaffer, C.L., Marjanovic, N.D., Lee, T., Bell, G., Kleer, C.G., Reinhardt, F., D'Alessio, A.C., Young, R.A., and Weinberg, R.A. (2013). Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. Cell *154*, 61–74. https://doi.org/10.1016/j.cell.2013.06.005.

62. Chen, A.F., Liu, A.J., Krishnakumar, R., Freimer, J.W., DeVeale, B., and Blelloch, R. (2018). GRHL2-Dependent enhancer switching maintains a pluripotent stem cell transcriptional subnetwork after exit from naive pluripotency. Cell Stem Cell *23*, 226–238.e4. https://doi.org/10.1016/j.stem.2018.06.005.

63. Preca, B.T., Bajdak, K., Mock, K., Sundararajan, V., Pfannstiel, J., Maurer, J., Wellner, U., Hopt, U.T., Brummer, T., Brabletz, S., et al. (2015 Dec 1). A self-enforcing CD44s/ZEB1 feedback loop maintains EMT and stemness properties in cancer cells. Int. J. Cancer *137*, 2566–2577.

64. Wang, Z., Coban, B., Liao, C.Y., Chen, Y.J., Liu, Q., and Danen, E.H.J. (2023 Jan 28). GRHL2 regulation of growth/motility balance in luminal versus basal breast cancer. Int. J. Mol. Sci. *24*, 2512.

65. Werner, S., Frey, S., Riethdorf, S., Schulze, C., Alawi, M., Kling, L., Vafaizadeh, V., Sauter, G., Terracciano, L., Schumacher, U., et al. (2013). Dual roles of the transcription factor grainyhead-like 2 (GRHL2) in breast cancer. J. Biol. Chem. *288*, 22993–23008. https://doi.org/10.1074/jbc.M113.456293.

66. Aceto, N., Bardia, A., Miyamoto, D.T., Donaldson, M.C., Wittner, B.S., Spencer, J.A., Yu, M., Pely, A., Engstrom, A., Zhu, H., et al. (2014). Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. Cell *158*, 1110–1122. https://doi.org/10.1016/j.cell.2014.07.013.

67. Gkountela, S., Castro-Giner, F., Szczerba, B.M., Vetter, M., Landin, J., Scherrer, R., Krol, I., Scheidmann, M.C., Beisel, C., Stirnimann, C.U., et al. (2019). Circulating tumor cell clustering shapes DNA methylation to enable metastasis seeding. Cell *176*, 98–112.e14. https://doi.org/10.1016/j.cell.2018.11.046.

68. Padmanaban, V., Krol, I., Suhail, Y., Szczerba, B.M., Aceto, N., Bader, J.S., and Ewald, A.J. (2019). E-cadherin is required for metastasis in multiple models of breast cancer. Nature *573*, 439–444. https://doi.org/10.1038/s41586-019-1526-3.

69. Debaugnies, M., Rodríguez-Acebes, S., Blondeau, J., Parent, M.A., Zocco, M., Song, Y., de Maertelaer, V., Moers, V., Latil, M., Dubois, C., et al. (2023). RHOJ controls EMT-associated resistance to chemotherapy. Nature *616*, 168–175. https://doi.org/10.1038/s41586-023-05838-7.

70. Gazdar, A.F., Kurvari, V., Virmani, A., Gollahon, L., Sakaguchi, M., Westerfield, M., Kodagoda, D., Stasny, V., Cunningham, H.T., Wistuba, I.I., et al. (1998). Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. Int. J. Cancer *78*, 766–774. https://doi.org/10.1002/(sici)1097-0215(19981209)78:6<766::aid-ijc15>3.0.co;2-l.

71. Forozan, F., Veldman, R., Ammerman, C.A., Parsa, N.Z., Kallioniemi, A., Kallioniemi, O.P., and Ethier, S.P. (1999). Molecular cytogenetic analysis of 11 new breast cancer cell lines. Br. J. Cancer *81*, 1328–1334. https://doi.org/10.1038/sj.bjc.6695007.

72. Naldini, L., Blömer, U., Gage, F.H., Trono, D., and Verma, I.M. (1996 Oct 15). Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector. Proc. Natl. Acad. Sci. USA *93*, 11382–11388. https://doi.org/10.1073/pnas.93.21.11382.

73. Breunig, C.T., Durovic, T., Neuner, A.M., Baumann, V., Wiesbeck, M.F., Köferle, A., Götz, M., Ninkovic, J., and Stricker, S.H. (2018). One step generation of customizable gRNA vectors for multiplex CRISPR approaches through string assembly gRNA cloning (STAgR). PLoS One *13*, e0196015. https://doi.org/10.1371/journal.pone.0196015.

74. Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011 Jan 25). A hyperactive piggyBac transposase for mammalian applications. Proc. Natl. Acad. Sci. USA *108*, 1531–1536. https://doi.org/10.1073/pnas.1008322108.

75. Hu, Y., and Smyth, G.K. (2009 Aug 15). ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J. Immunol. Methods *347*, 70–78. https://doi.org/10.1016/j.jim.2009.06.008.

76. Gomis, R.R., Alarcón, C., Nadal, C., Van Poznak, C., and Massagué, J. (2006 Sep). C/EBPbeta at the core of the TGFbeta cytostatic response and its evasion in metastatic breast cancer cells. Cancer Cell *10*, 203–214. https://doi.org/10.1016/j.ccr.2006.07.019.

77. Castro, F., Dirks, W.G., Fähnrich, S., Hotz-Wagenblatt, A., Pawlita, M., and Schmitt, M. (2013 Jan 15). High-throughput SNP-based authentication of human cell lines. Int. J. Cancer *132*, 308–314. https://doi.org/10.1002/ijc.27675.

78. Schmitt, M., and Pawlita, M. (2009 Oct). High-throughput detection and multiplex identification of cell contaminations. Nucleic Acids Res. *37*, e119. https://doi.org/10.1093/nar/gkp581.

79. Kim, S.I., Oceguera-Yanez, F., Sakurai, C., Nakagawa, M., Yamanaka, S., and Woltjen, K. (2016). Inducible transgene expression in human iPS cells using versatile all-in-one piggyBac transposons. Methods Mol. Biol. *1357*, 111–131. https://doi.org/10.1007/7651_2015_251.

80. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

81. Anders, S., Pyl, P.T., and Huber, W. (2015 Jan 15). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

82. Wickham, H. (2016). ggplot2 (Springer International Publishing). https://doi.org/10.1007/978-3-319-24277-4.

83. Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017 Oct). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat. Methods *14*, 959–962. https://doi.org/10.1038/nmeth.4396.

84. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218. https://doi.org/10.1038/nmeth.2688.

85. Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr. Protoc. Bioinformatics *11*, 11.7, Chapter 11:Unit 11.7. https://doi.org/10.1002/0471250953.bi1107s32.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-alpha-tubulin	Sigma Aldrich	Clone B-5-1-2, Cat#T5168, RRID:AB_477579
anti-beta-actin	Abcam	Clone AC-15, Cat#ab6276, RRID:AB_2223210
anti-CD10	Leica Biosystems	Cat# NCL-CD10-270 RRID:AB_442048
anti-CD298-PE	Biolegend	Clone LNH-94, Cat# 341704, RRID:AB_2274458
anti-CD45-VioBlue	Miltenyi Biotec	Clone REA747, Cat# 130-110-637, RRID:AB_2658243
anti-CDH1	Agilent DAKO	Clone M3612, RRID:AB_2076672
anti-E-Cadherin (for WB)	GeneTex	Clone EP700Y, Cat# GTX61329, RRID:AB_10634595
anti-E-Cadherin-Alexa-488	Cell Signaling Technology	Clone 24E10, Cat#3199, RRID:AB_10691457
anti-EPCAM (for IHC)	Agilent DAKO	Clone Ber-EP4, Cat# F086001,RRID:AB_578689
anti-EPCAM-FITC (for FACS)	Miltenyi Biotec	Clone REA764, Cat# 130-110-998, RRID:AB_2657493
anti-EPCAM-FITC (for FACS)	GeneTex	Clone VU-1D9, Cat#GTX79849, RRID:AB_11172207
anti-estrogen receptor alpha (ER)	Thermo Fisher	Clone SP1, Cat# MA5-16362, RRID:AB_2537881
anti-GATA3	Becton Dickinson	Clone L50-823, Cat# 566642, RRID:AB_2813884
anti-HER2 polyclonal serum	Agilent DAKO	Cat# A0485, RRID:AB_2335701
anti-HER3-APC	R&D	Cat# FAB3481A, RRID:AB_416878
anti-Histone H3	Abcam	Cat#ab1791, RRID:AB_302613
anti-human Ki-67	Agilent DAKO	Clone Ki-67, Cat# F078801, RRID:AB_578672
anti-KRT18	Agilent DAKO	Cat# M7010, RRID:AB_2133299
anti-KRT19	Agilent DAKO	Clone RCK108, Cat# M0888, RRID:AB_2234418
anti-KRT5/6	Agilent DAKO	Clone D5/16 B4 Cat# M7237, RRID:AB_2281083
anti-progesterone receptor	Agilent DAKO	Clone PgR636, Cat# M3569, RRID:AB_2532076
anti-Twist1 (for WB) (for IF)	Santa Cruz	Clone Twist2C1a, Cat#sc-81417, RRID:AB_1130910
anti-vimentin (for WB) (for IF)	Abnova	Clone V9, Cat#MAB3578, RRID:AB_10677095
anti-vimentin (for IHC)	Agilent DAKO	Clone M0725, Cat# M0725, RRID:AB_10013485
anti-ZEB1 rabbit IgG H-102 (for WB) (for IF)	Santa Cruz	Cat#sc-25388, RRID:AB_2217979
anti-ZEB1 rabbit polyclonal antibody (for IHC)	Sigma Aldrich	Cat#HPA027524, RRID:AB_1844977
anti-ZEB1 rabbit polyclonal antibody (for ChIP)	Thermo Fisher	Cat#PA528221, RRID:AB_2545697
anti-ZEB1 rabbit polyclonal antibody (for ChIP)	Novus Biologicals	Cat#NBP1-05987, RRID:AB_1556166
donkey anti-mouse IgG Alexa Flour 488	Life Technologies	Cat#A21202, RRID:AB_141607
donkey anti-rabbit IgG Alexa Flour 594	Life Technologies	Cat#A21207, RRID:AB_141637
goat anti-mouse IgG Alexa Flour 594	Life Technologies	Cat#A11032, RRID:AB_2534091
goat anti-mouse IgG Alexa Fluor 546	Life Technologies	Cat#A11030, RRID:AB_2534089
goat anti-rabbit IgG Alexa Fluor 488	Life Technologies	Cat#A11034, RRID:AB_2576217
goat Mouse IgG HRP	Jackson ImmunoResearch	Cat#115-036-062, RRID:AB_2307346
goat Rabbit IgG HRP	Jackson ImmunoResearch	Cat#111-036-045, RRID:AB_2337943
normal rabbit IgG	Cell Signaling Technology	Cat#2729, RRID:AB_1031062
anti-mouse CD45 APC/Cyanine7	Biolegend	Clone 30-F11, Cat# 103116, RRID:AB_312981
anti-mouse CD11b APC/Cyanine7	Biolegend	Clone M1/70, Cat# 101226, RRID:AB_830642
anti-mouse F4/80 APC/Cyanine7	Biolegend	Clone BM8, BioLegend Cat# 123118, RRID:AB_893477
anti-mouse TER-119/Ly-76 APC/Cyanine7	Biolegend	Clone TER-119, Cat# 116223, RRID:AB_2137788
anti-mouse Ly-6G APC/Cyanine7	Biolegend	Clone 1A8, Cat# A25985, RRID:AB_2536046

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
anti-mouse Thy1.2 APC/Cyanine7	Biolegend	Clone 30-H12, BioLegend Cat# 105327, RRID:AB_10613280
anti-mouse H2-k^{d} APC/Cyanine7	Biolegend	Clone SF1-1.1, Cat# 116629, RRID:AB_2616847
Bacterial and virus strains		
MAX Efficiency™ Stbl2™ Competent Cells	Invitrogen	Cat#10268019
XL10-Gold Ultracompetent Cells	Agilent Technologies	Cat#200314
Biological samples		
Human metastatic breast cancer effusion samples	Heidelberg National Center for Tumor Disease and Mannheim Women's Clinic	Patient 1-10
Chemicals, peptides, and recombinant proteins		
(Z)-4-Hydroxytamoxifen	Sigma Aldrich	Cat#H7904
7-AAD (7-amino-actinomycin D)	Becton Dickinson	Cat# 559925, RRID:AB_2869266
BD Matrigel™ Matrix GFR reduced	BD Biosciences	Cat#356234
Collagen type I rat tail	Corning	Cat#354236
Beta-estradiol solid pellets	Innovative Research of America	Cat#SE-121
Mammary Epithelial Cell Growth Medium + Supplement Mix	PromoCell	Cat#C-21010+ Cat#C-39115
Propidium Iodide	Becton Dickinson	Cat# 556463, RRID:AB_2869075
Critical commercial assays		
MIR141 probe set	Qiagen	HS-miR-141_1
MIR200A probe set	Qiagen	HS-miR-200a_1
MIR200B probe set	Qiagen	HS-miR-200b_3
MIR200C probe set	Qiagen	HS-miR200c_1
RNU6-2 probe set	Qiagen	HS-RNU6-2_11
Deposited data		
ATAC-sequencing raw data	This paper	GSE156024
RNA-sequencing raw data	This paper	GSE138329
Experimental models: Cell lines		
HMLE-Twist1-ER immortalized cells	Robert Weinberg	Schmidt et al.[31]
HCC1143 cell line	American Type Culture Collection (ATCC)	Gazdar et al.[70] Cat#CRL-2321, RRID:CVCL_1245
SUM-149PT cell line	Asterand Bioscience	Forozan et al.[71] RRID: CVCL_3422
SUM-229PE cell line	Asterand Bioscience	Forozan et al.[71] RRID: CVCL_5594
Experimental models: Organisms/strains		
NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice	The Jackson Laboratory	Cat#005557 RRID:IMSR_JAX:005557Info
Oligonucleotides		
All oligonucleotides can be found in the Table S2.		
Recombinant DNA		
UBC-RFP-F2A-Luciferase lentivector	System Biosciences (SBI)	BLIV200PA-1-SBI
UBC-GFP-T2A-Luciferase lentivector	System Biosciences (SBI)	BLIV201PA-1-SBI
pcDNA3.1+/C-(K)DYK with subcloned ZEB1 (NM_001128128.2) cDNA ORF	GenScript Biotech	Clone OHu27170C
shGRHL2 (lentiviral SMARTvector system)	Horizon Discovery	Cat#V3SH7602-225673780
shScrambled (lentiviral SMARTvector system)	Horizon Discovery	Cat#VSC11709
pRRL-PPT-CMV-GFP-WPRE	Timm Schroeder	Naldini et al.[72]

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pSpCas9(BB)-2A-GFP	Stefan Stricker	Breunig et al.[73] http://n2t.net/addgene:48138, RRID:Addgene_48138
pENTR1A-no ccdb	Eric Campeau & Paul Kaufman	Plasmid #17398; http://n2t.net/addgene:17398, RRID:Addgene_17398
PB-TAC-ERP2	Knut Woltjen	Plasmid #80478 http://n2t.net/addgene:80478, RRID:Addgene_80478
pCMV-hyPBase	Roland Rad	Yusa et al.[74]
Software and algorithms		
Benchling	https://benchling.com/	RRID:SCR_013955
ELDA: Extreme Limiting Dilution Analysis	https://bioinf.wehi.edu.au/software/elda/	Hu and Smyth,[75]
MIT CRISPR design tool	http://crispr.mit.edu/	N/A

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Christina Scheel (christina.scheel@klinikum-bochum.de).

Materials availability

The PiggyBac plasmid constructs generated in this study will be shared by the lead contact upon request.

Data and code availability

- The datasets that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under the accession codes GSE138329 and GSE156024.
- This paper does not report original code.
- Any additional information is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

MBC liquid biopsies

Pleural and ascitic effusions were obtained from post-menopausal stage IV patients admitted to the Division of Gynecologic Oncology, National Center for Tumor Diseases and the Women's Clinic of the Mannheim University Hospital. The study was approved by the ethical committee of the University of Heidelberg (case number S295/2009) and University of Mannheim (2010-024238-46). All samples were transferred to the laboratory no later than 1 h after thoracentesis or laparocentesis and processed to extract cancer cells as previously described.[76] Patient's diagnostic information and clinical disease characteristics are described in detail in Figure S1A.

Cell cultures

HMLE-Twist1-ER were kindly gifted by Robert A. Weinberg (Whitehead Institute).The SUM-149PT, HCC1143 and SUM-229PE cell lines were purchased from Asterand Bioscience. The SUM-149PT was established by S. Ethier from a triple-negative inflammatory breast cancer. The HCC1143 and SUM229PE lines were established from a primary breast tumor[70] and from a pleural effusion,[71] respectively. All cell lines show negative staining for ER, PR and HER2 in culture, however no information is available on the original disease subtype. All cell lines were authenticated using Multiplex Cell Authentication (Multiplexion, Heidelberg, Germany) as described.[77] The SNP profiles matched known profiles or were unique. The cells were tested and were proven to be negative for contaminants by running the Multiplex cell Contamination Test (Multiplexion, Heidelberg) as described.[78]

METHOD DETAILS

Xenograft assay for breast cancer cells

Female NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ (NSG) mice were transplanted at 6 to 8 weeks of age for primary cells, or at 8 to 12 weeks for cell lines. A suspension of FACS-purified cells in sterile PBS was combined with growth factor-reduced Matrigel (Becton Dickinson) in a 1:1 ratio and then injected in the fourth or the fifth mammary fat pad for primary cells, or in the first or the second mammary fat pad for cell lines. For experimental bone metastasis formation, a 1:4 ratio of PBS:Matrigel was used, and no more than 10,000 cells were injected per femur. Mice transplanted with primary cells from liquid biopsies received subcutaneous implantation of beta-estradiol as solid pellets (Innovative Research of America) with modalities that were already described.[44] All mice subjected to limiting dilution assays for tumorigenesis were monitored for a maximum of 14 months post-transplantation. Animal care and procedures for primary cells transplantation experiments followed the German legal regulations and were previously approved by the governmental review board of the state of Baden-Württemberg, operated by the local Animal Welfare Office (Regierungspräsidium Karlsruhe) under the license number G240-11. Some transplantation experiments involving cell lines were carried out at ETH Zurich according to institutional and cantonal guidelines (approved mouse protocol number 33688, cantonal veterinary office of Zurich). Upon necropsy, fat pads and lungs containing xenografted cells were either fixed in 4% v/v paraformaldehyde for histologic preparations, or manually dissociated for downstream flow cytometric analysis. Metastasis development is scored as the absolute number of mice found to bear at least one detectable lung metastatic lesion, as assessed by histological examination.

Immunohistochemical analysis of MBC xenografts

Tumor lesions generated in mice and metastatic lungs were freshly collected upon necropsy and embedded in paraffin for downstream histological inspections. For immunohistochemistry, the following antibodies were used: anti-human Ki-67 (DAKO, clone Ki-67, concentration 35 mg/L), anti-estrogen receptor alpha (ER) (Thermo Fisher Scientific, clone SP1, diluted 1:10), anti-progesterone receptor (PR) (DAKO, clone PgR636, concentration 53.8 mg/L), anti-HER2 (Agilent, polyclonal serum A0485, concentration 320 mg/L), anti-CDH1 (DAKO, clone M3612, manufacturer's protocol), anti-Vimentin (DAKO, clone M0725, diluted 1:1,000), anti-EPCAM (DAKO, clone Ber-EP4, manufacturer's protocol), anti-KRT19 (DAKO, RCK108 clone, manufacturer's protocol), anti-KRT18 (DAKO, DC-10 clone, manufacturer's protocol), and anti-KRT5/6 (DAKO, D5/16 B4 clone, manufacturer's protocol), anti-CD10 (Leica Biosystems, diluted 1:10), anti-human ZEB1 (Sigma Aldrich, rabbit polyclonal #HPA027524, diluted 1:100) anti-GATA3 (Becton Dickinson, clone L50-823, concentration 0.2 mg/L).

Targeted sequencing of somatic mutations

Genomic DNA was extracted from FACS-purified cancer cells and matched germline controls (CD45$^+$ white blood cells purified by FACS sorting from the same metastatic effusion). The concentration of nucleic acids was assessed by fluorimetric measurement through QuBit 3.0 and the amount of amplifiable DNA (sequencing-grade quality) was determined using a quantitative assay (TaqMan RNAseP detection assay, Thermo Fisher Scientific) on a StepOnePlus device (Thermo Fisher Scientific). DNA samples were amplified using a custom-designed gene panel for breast cancer covering the most recurrent mutations. Library preparation and sequencing were performed using the multiplex PCR-based Ion Torrent AmpliSeq^{TM} technology (Thermo Fisher Scientific) and Ion S5XL technology.

RNA-sequencing of MBC cells

RNA extraction was carried out from 40,000 FACS-sorted cells using ARCTURUS PicoPure RNA isolation kit (Thermo Scientific, Cat. KIT0204). RNA quality was assessed using BioAnalyzer 2100 (Agilent). Libraries were prepared from 1 μg of total RNA using a TruSeq Stranded Total RNA Kit with Ribo-Zero Human/Mouse/Rat (Illumina) according to the manufacturer's instructions. The pooled RNA libraries were sequenced on an Illumina HiSeq2000 to obtain 100-bp paired-end reads. RNA-sequencing reads were aligned to the human reference genome hg19 and quantified using the stranded option of STAR version 2.4.2a30. The DESeq2 package was used to obtain normalized expression values. The GSEA desktop application was utilized for enrichment analysis of the indicated gene sets. Gene set databases downloaded from the Broad Institute website or custom-derived signatures were analyzed using default settings with 2000 permutations.

Single-cell gene expression profiling of MBC cells

RT-PCR primers were designed using the Roche Universal Probe Library Assay Design Center to obtain amplicons spanning between different exons. In silico validation of the primers was achieved using the MFE primer online tool (2.0 version) while the melting temperature effectively considered was calculated using the Applied Biosystem online tool. Defined panels of validated primers were generated by combining the primers in mini-pools for the interrogation of single-cells, which was carried out as follows. Briefly, single cells were sorted using a BD FACSAria II (Becton Dickinson) using the index-sort mode combined with single-cell precision mode. Cells were directly deposited in 96-well plates loaded with 5 μL of a lysis buffer plus RT/Taq-PCR master mix (1x CellsDirect reaction mix, Invitrogen and 0.1 μL RT/Taq enzyme mix, Invitrogen) which included a 0.1 μM dilution of the target-specific primer pool. Samples were immediately transferred to a PCR machine and processed for reverse transcription and subsequent pre-amplification. Pre-amplified samples were diluted 10-fold and 1 μL was used for qPCR analysis utilizing Power SYBR Green reagent (Applied Biosystems) on a ViiA 7 Real-Time PCR system (Applied Biosystems). For bulk-level gene expression profiling, 500 ng of total RNA were reverse transcribed using the

SuperScript VILO cDNA kit (Thermo Fisher Scientific) according to the manufacturer's instruction. Quantitative PCR analysis is presented as the multiplicative reciprocal of Ct-values.

Bidimensional (2D) cell cultures

HMLE-Twist1-ER cells cells were cultured in Mammary Epithelial Cell Basal Medium containing 0.004 mL/mL BPE, 10 ng/mL EGF, 5 μg/mL Insulin and 0.5 μg/mL hydrocortisone (PromoCell) supplemented with 1% Pen/Strep (Gibco) and Blasticidin S HCl (Thermo Fisher Scientific) at a final concentration of 10 μg/mL. To obtain HMLE-Twist1-ER CD24high single-cell clones (SCCs), cells were seeded to a 96-well plate (0.3 cells/well). Wells were checked by eye for single cells. Wells with a single cell were further passaged and expanded. For Twist1-activation, cells were treated with (Z)-4-Hydroxytamoxifen (TAM, Sigma-Aldrich) at a final concentration of 20 nM for the indicated number of days. The SUM-149PT, HCC1143 and SUM-229PE cell lines were cultivated in DMEM:F12 (Gibco) containing 5% FBS v/v. For downstream *in vivo* transplantations, SUM-149PT cells were transduced with an UBC-GFP-F2A-Luciferase lentivector (System Biosciences, SBI) or a UBC-RFP-T2A-Luciferase lentivector (System Biosciences, SBI) and FACS-sorted to select GFP- or RFP-positive cells.

Single-cell cloning and MS assay

To obtain HMLE-Twist1-ER CD24high single-cell clones (SCCs), cells were diluted to 3 cells/ml. The cell suspension was seeded into 96-well plate (100 μL/well ≙ 0.3 cells/well). For single-cell cloning of ZEB1 KO cells, 1 cell was seeded per 96-well. Wells were checked by eye for single cells. Wells with a single cell were further passaged and expanded. For single-cell cloning of ZEB1 OE clones, single-cell clones were isolated with cloning cylinders (Sigma) and expanded. The MS assay was performed as described,[31] by seeding 100 cells per well in 96-well ultra-low-adhesion plates (Corning) in Dulbecco's modified Eagle's medium/F12 medium (Life Technologies) containing 1.0% methylcellulose (R&D Systems) supplemented with 10 ng/mL epidermal growth factor (Millipore), 20 ng/mL basic fibroblast growth factor (Millipore), 0.5 μg/mL hydrocortisone (Sigma), 4 μg/mL heparin (Sigma), 10 μg/mL insulin (Sigma), and B27 (Life Technologies). For serial passaging experiments, primary spheres were dissociated into single cells by trypsinization and replated.

Organotypic cultures in 3D collagen gels

To produce organoids *in vitro*, floating 3D-collagen gels with a final concentration of 1.3 mg/mL collagen I (collagen type I from rat tail, Corning) were prepared as previously published.[49] Briefly, 300 HMLE-Twist1-ER cells were embedded per collagen gel. Cells were further cultivated in Mammary Epithelial Cell Basal Medium containing 0.004 mL/mL BPE, 10 ng/mL EGF, 5 μg/mL Insulin and 0.5 μg/mL hydrocortisone (PromoCell) supplemented with Pen/Strep (Gibco) and Blasticidin S HCl (Thermo Fisher Scientific) at a final concentration of 10 μg/mL. Medium was changed every two to three days. After 10 days in culture, floating 3D-collagen gels were fixed with 4% paraformaldehyde for 15 min. Immunofluorescent staining was performed as described below. To visualize colonies, floating 3D-collagen gels were incubated in a carmine alum staining solution (2 g/L carmine (Sigma-Aldrich), 5 g/L aluminum potassium sulfate (Sigma-Aldrich), two crystals of thymol) at 4°C overnight. Pictures of gels were taken with a Leica DFC450 C Stereomicroscope with a Plan 0.8× LWD objective and the Leica Application Suite V4 software. Pictures were stitched with Panorama Stitcher. Carmine positive colonies were counted using ImageJ.

Colonizing assay on murine PCLS

PCLS were generated using an established method.[51] Briefly, C57BL6/N mice of 8–12 weeks were intubated and after the dissection of the diaphragm, lungs were flushed through the heart with sterile sodium chloride solution. Using a syringe pump, lungs were filled with low gelling temperature agarose (2% by weight, A9414; Sigma; kept at 40°C) in sterile cultivation medium (DMEM/Ham's F12 (Gibco) supplemented with 100 U/ml penicillin, 100 μg/mL streptomycin and 2.5 μg/mL amphotericin B (Sigma)). Separated lobes were cut with a vibratome (Hyrax V55; Zeiss, Jena, Germany) to a thickness of 300 μm using a speed of 10–12 μm/s, a frequency of 80 Hz and an amplitude of 1 mm. The PCLS were cultivated in sterile cultivation medium containing 0.1% FCS. HMLE-Twist1-ER cells were labeled by lentiviral transduction using pLV-CMV-GFP (pRRL-PPT-CMV-GFP-WPRE).[72] Fluorescence-activated cell sorting (FACS) was used to select GFP-positive cells. 2x105 cells were plated per lung slice and cultured in Mammary Epithelial Cell Basal Medium containing 0.004 mL/mL BPE, 10 ng/mL EGF, 5 μg/mL Insulin and 0.5 μg/mL hydrocortisone (PromoCell) supplemented with Pen/Strep (Gibco) for 5 days, changing the medium every day.

CRISPR-Cas9 mediated knockout of ZEB1

Knock-out of *ZEB1* was carried out by CRISPR-Cas9 technology, using guide RNAs (gRNAs) targeting exons 5 and 7, as they are both included in all 9 transcript variants of *ZEB1*. Four non-overlapping gRNAs were designed with Benchling following the indications of the MIT CRISPR design tool (http://crispr.mit.edu/). All gRNAs were co-expressed by a single plasmid produced via Gibson Assembly using the string assembly gRNA cloning (STAgR) protocol.[73] The purified gRNA-containing STAgR-Neo plasmid was co-transfected into HMLE-Twist1-ER cells together with a Cas9-GFP-expressing plasmid (pSpCas9(BB)-2A-GFP) using the TransIT-X2 transfection reagent (Mirus Bio LLC) according to manufacturer's instructions. As a control, cells were transfected with the Cas9-GFP plasmid and a plasmid harboring the guide RNA scaffold and termination sequence. After 2.5 days cells were sorted for GFP expression on a FACSAriaIIIu (Becton Dickinson). To obtain single-cell clones, sorted cells were seeded into 96-well plates (1 cell/well). Wells were regularly checked by eye for single-cell outgrowth in the following days. Wells with a single cell clone were further

passaged and expanded. To screen for cells with successful deletion, cells were lysed in DNAzol reagent (Thermo Fisher Scientific) and genomic DNA (gDNA) of *ZEB1* knockout single-cell clones was isolated with the DNAzol Reagent (Invitrogen) according to the manufacturer's instructions for adherent cells (monolayer). PCRs of the specific loci were performed to screen for DNA modifications and *ZEB1* knockout was validated by immunoblotting. The following gRNAs were used:

gRNA ZEB1 NM128128 Exon 5:
5′-GCCTCTATCACAATATGGAC-3′ and 5′-ACAACTCAGCCCTCAATGGA-3′,
gRNA ZEB1 NM128128 Exon 7:
5′-AGTTCTGTCACAAGCATGCA-3′ and 5′-TTGCCGTATCTGTGGTCGTG-3′.

The following PCR primers were used:

ZEB1 NM128128 Exon 5:
Forward 5′-GCATAGGGACTCAGTGGAAACT-3′
Reverse 5′-AGGAGGCAACTCCCTTTACTAC-3′
ZEB1 NM128128 Exon 7:
Forward 5′-GGTCGGTGAAATGGGATAAGAAAAA-3′
Reverse 5′-ACCACCAGTGAAAACCCCATT-3′.

Overexpression of *ZEB1*

To overexpress *ZEB1*, we made use of the all-in-one PiggyBac system designed for inducible transgene expression.[79] In detail, the *ZEB1* (NM00128128.2) cDNA (pcDNA3.1+/C-(K)DYK with ZEB1 (NM_001128128.2) cDNA ORF clone OHu27170C GenScript) was cloned into the pENTR1A-no ccdb vector (#17398, Addgene) to generate the pENTR1A-*ZEB1* vector. Next, the Gateway LR Clonase II Enzyme mix (Invitrogen) was used to transfer the *ZEB1* cDNA from the pENTR1A-*ZEB1* vector into the PB-TAC-ERP2 (#80478, Addgene) to generate the PB-*ZEB1* vector. HMLE-Twist1-ER and SUM-149PT cells were transfected with the linearized PB-ZEB1 or PB-GUS plasmid and the pCMV-hyPBase[74] encoding vector (kind gift of Roland Rad, TU Munich) in a ratio of 8:1, respectively, using the TransIT-X2 transfection reagent (Mirus Bio LLC) according to manufacturer's instructions. HMLE-Twist1-ER or SUM-149PT *ZEB1* overexpression single-cell clones were isolated by seeding puromycin-resistant cells into a 96-well plate (0.5 cells/well) or by using cloning cylinders (Sigma-Aldrich) and propagated in their original culture medium further supplemented with Puromycin Dihydrochloride (Thermo Fisher Scientific) at a final concentration of 1 μg/mL. For experiments, individual clones were stimulated with 1 μg/mL Doxycycline Hydrochloride Hemiethanolate Hemihydrate (hyclate) (Sigma-Aldrich) (dissolved in DMSO at 1 mg/mL and then further diluted in water) or with an equivalent volume of solvent (DMSO).

Stable knock-down of *GRHL2*

To knock down *GRHL2*, we made use of a lentiviral SMARTvector system (hEF1a-TurboGFP) (Horizon Discovery) designed for constitutive shRNA expression. In detail, the V3SH7602-225673780 construct for human GRHL2 (indicated as shGRHL2) or a scrambled control hairpin (indicated as shScrambled) were chosen. HMLE-Twist1-ER cells were transduced to reach MOI equal to 1 and sorted for GFP expression on a FACSAria IIu (Becton Dickinson). For knock-down experiments in SUM-149PT cells, a Zeocin resistance cassette was utilized to replace the hEF1a-TurboGFP marker. Transduced cells were selected using 100 μg/mL Zeocin (Invivogen).

RNA-sequencing of HMLE-Twist1-ER cells

Library preparation was performed using the TruSeq Stranded Total RNA Library Prep Kit with Ribo-Zero (Illumina). Briefly, total RNA was isolated using the miRNeasy Mini Kit (Qiagen) and RNA integrity number (RIN) was determined with the Agilent 2100 BioAnalyzer (RNA 6000 Nano Kit, Agilent). For library preparation, 1 μg of RNA was depleted for cytoplasmatic rRNAs, fragmented, and reverse transcribed with the Elute, Prime, Fragment Mix. A-tailing, adaptor ligation, and library enrichment were performed as described in the High Throughput protocol of the TruSeq RNA Sample Prep Guide (Illumina). RNA libraries were assessed for quality and quantity with the Agilent 2100 BioAnalyzer and the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies). RNA libraries were sequenced as 150 bp paired-end runs on an Illumina HiSeq4000 platform. The STAR aligner (v 2.4.2a) with modified parameter settings (*–twopassMode = Basic*) is used for split-read alignment against the human genome assembly hg19 (GRCh37) and UCSC Gene annotation.[80]To quantify the number of reads mapping to annotated genes we use HTseq-count (v0.6.0).[81] FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) values are calculated using custom scripts. PCA plots were created with the R package *ggplot2*.[82]

Omni-ATAC-sequencing of HMLE-Twist1-ER cells

ATAC-sequencing was carried out exactly as described previously.[83] Briefly, 50,000 cells (viability >90%) were harvested by trypsinization and pelleted (500 rcf, 4°C, 5 min). If the cell viability of cell suspensions used for ATAC-seq was below 90%, dead cells were removed with the Dead Cell Removal Kit (Milteny Biotech) according to the manufacturer's instructions and then harvested as the others. Cell pellets were resuspended in ATAC Resuspension Buffer (10mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$)

supplemented with 0.1% NP40, 0.1% Tween 20 and 0.01% digitonin for lysis, incubated on ice for 3 min and then 1 mL of ATAC Resuspension Buffer supplemented only with 0.1% Tween 20 was added and spun (500 rcf, 4°C, 10 min) to collect nuclei. Nuclei were subsequently resuspended in 50 μL of Transposase Reaction containing 25 μL 2x Tagmentation Buffer (20 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 20% dimethyl Formamide, H_2O), 2.5 μL Tn5 Transposase (Illumina Nextera DNA Library Preparation Kit), 5.25 μL H_2O, 16.5 μL PBS, 0.25 μL of 2% digitonin (Promega) and 0.5 μL of 10% Tween 20. Reactions were incubated for 30 min at 37°C in a Thermomixer C (Eppendorf) shaking at 900 rpm and DNA was then purified using Qiagen PCR clean-up MinElute kit (Qiagen). Transposed DNA was subsequently amplified in 50 μL reactions with custom primers as described previously.[84] After 4 cycles of amplification, libraries were then monitored with qPCR: 5 μL PCR sample in a 15 μL reaction with the same primers. The additional number of cycles needed for the remaining volume of PCR reaction was determined by calculating the number of cycles that corresponds to ¼ of the maximal reached fluorescence. Amplified libraries were purified with the Qiagen PCR clean-up MinElute kit and size-selected for fragments less than 600 bp using the Agencourt AMPure XP beads (Beckman Coulter). Libraries were quality controlled by Qubit and Agilent DNA Bioanalyzer analysis. Deep sequencing was performed on a HiSeq 1500 system according to the standard Illumina protocol for 50bp single-end reads. ATAC-seq reads were aligned to the human genome hg38 using Bowtie with options "*-q -n 2 –best –chunkmbs 2000 -p 32 -S*". ATAC peaks over Input background were identified using Homer findPeaks.pl with option "*-style factor*".[85] Peaks from all samples were merged using *mergePeaks* resulting in a unified Peak set. The peak list was filtered to eliminate the promoter-associated peaks (distance to TSS <1000bp) using bedtools. Raw ATAC coverage counts were then calculated with annotatePeaks. PCA analysis was performed on ATAC peaks coverage data with the R function *prcomp*. The list of differential ATAC peaks was determined with the DESeq2 results function for all possible comparisons between the samples and filtered for adjusted p value <0.05 and log2 fold change >3. The cluster heatmap was generated based on normalized coverage data of all differential ATAC peaks using the R *pheatmap* package. In detail, differential ATAC peaks between epithelial and mesenchymal cells were calculated at baseline, d7 and d14 after TAM addition (filtered for log2fc > 3 and $p < 0.05$). Then, the log2 normalized coverage of all differential peaks was plotted using *pheatmap* in *k_means mode* with 12 clusters. Cluster-specific peaks were further analyzed for transcription factor motifs using homer *findMotifsGenome* function in *de novo* mode.

Flow cytometry and FACS

Breast cancer cells were stained in a 1% BSA solution containing 2 mM Ethylenediaminetetraacetic acid (EDTA) under the presence of FcR blocking reagent (Miltenyi). HMLE-Twist1-ER cell suspensions were stained in Mammary Epithelial Cell Basal Medium (PromoCell) while SUM-149PT cells were stained in ice-cold PBS. DAPI, Propidium Iodide (PI), or 7-AAD (BD Biosciences) diluted 1:125 were used to exclude dead cells, depending on the staining panel. Cells were sorted and analyzed on a FACSAria Illu, an LSR Fortessa, and a FACS Aria Fusion (all Becton Dickinson) and further analyzed with FlowJo. The following antibodies were used: anti-EPCAM-FITC or -APC (VU-1D9 clone, GeneTex, diluted 1:25, or REA764 clone, Miltenyi, diluted 1:11), anti-CD45-VioBlue (REA747 clone, Miltenyi, diluted 1:11) anti-CD298-PE (LNH-94 clone, Biolegend, diluted 1:500) anti-HER3-APC (FAB3481A clone R&D, diluted 1:25). The mouse lineage cocktail was prepared by pooling APC/Cyanine7-conjugated antibodies against the following mouse antigens: CD45, CD11b, F4/80, TER-119, Ly-6G, Thy1.2 and H2-kd (Biolegend).

Immunofluorescence staining and imaging

Cells were cultured on coverslips coated with poly-D-lysine (Sigma-Aldrich), in floating 3D-collagen gels, or on murine lung slices. Cells were fixed with 4% paraformaldehyde for 10 to 15 min and permeabilized with 0.2% Triton X-100. After blocking with 10% normal goat serum in 0.1% BSA (Sigma-Aldrich), coverslips, floating 3D-collagen gels, and murine lung slices were incubated overnight with primary antibodies in 0.1% BSA as follows: anti-E-Cadherin (clone EP700Y, GeneTex, diluted 1:250) or anti-E-Cadherin-Alexa-488 (clone 24E10, New England Biolabs, diluted 1:50), anti-Twist1 (clone Twist2C1a, Santa Cruz, 1:500), anti-Vimentin (clone V9, Abnova, diluted 1:100). Secondary antibodies were coupled to Alexa 488, -594, or −546 (Life Technologies, diluted 1:250). Cell nuclei were visualized with 40,6-diamidino-2-phenylindole-dihydrochloride (DAPI, Sigma-Aldrich). Slides were mounted with Aqua-Poly/Mount reagent (Polysciences). Corrected total cell fluorescence (CTCF) of DAPI was calculated using the following formula: CTCF = integrated density – (area of selected cells x mean fluorescence of background readings). Pictures were taken with a Zeiss Axio Imager.M2 using Zen software or with an Olympus Confocal using FV10-ASW software.

Quantitative RT-PCR for microRNAs and mRNAs

Total RNA was isolated using the miRNeasy Mini Kit (Qiagen) or the mRNeasy Mini Kit (Qiagen). Reverse transcription of mRNA was performed with the EasyScriptPlus Kit. For qRT-PCR, the Power SYBR Green-PCR Master Mix (Applied Biosystems) was used. *RPL32* was used as a control for normalization. Reverse transcription of miRNAs was performed with the miScript RT Kit (Qiagen) and qRT-PCR was performed with the Power SYBR Green-PCR Master Mix (Applied Biosystems). The following miScript Primer Assays (Qiagen) were used:

HS-RNU6-2_11 as loading control for normalization
HS-miR-141_1, HS-miR-200a_1.
HS-miR-200b_3, HSmiR200c_1.

Samples were run on a QuantStudio 12K Flex qPCR system (Life Technologies). Data were analyzed with the ΔCt method and relative expression levels are displayed as 1000 x 2-ΔCt.

For quantifying mRNAs, the following primer pairs were used:
CDH1:
Forward 5′-TGCCCAGAAAATGAAAAAGG-3′
Reverse 5'-GTGTATGTGGCAATGCGTTC-3′
EPCAM:
Forward 5′- ATAACCTGCTCTGAGCGAGTG-3′
Reverse 5′- TGCAGTCCGCAAACTTTTACTA-3′
FN1:
Forward 5′- CAGTGGGAGACCTCGAGAAG-3′
Reverse 5′- TCCCTCGGAACATCAGAAAC-3′
Twist1:
Forward 5′-GGACAAGCTGAGCAAGATTCA-3′
Reverse 5′-CGGAGAAGGCGTAGCTGAG-3′
OVOL2:
Forward 5′- ACAGGCATTCGTCCCTACAAA-3′
Reverse 5′- CGCTGCTTATAGGCATACTGC-3′
RPL32:
Forward 5′- CAGGGTTCGTAGAAGATTCAAGGG-3′
Reverse 5′- CTTGGAGGAAACATTGTGAGCGATC-3′
VIM:
Forward 5′- GAGAACTTTGCCGTTGAAGC-3′
Reverse 5′- GCTTCCTGTAGGTGGCAATC-3′
WNT5A:
Forward 5′- ATGGCTGGAAGTGCAATGTCT-3′
Reverse 5′- ATACCTAGCGACCACCAAGAA-3′
ZEB1:
Forward 5′- GCACAAGAAGAGCCACAAGTAG-3′
Reverse 5′- GCAAGACAAGTTCAAGGGTTC-3′.

Chromatin immunoprecipitation (ChIP)

For ChIP analysis, the equivalent of 50,000,000 HMLE cells was scaled up in 15 cm Petri dishes. Cells were fixed directly on the plates using 1% v/v final concentration of methanol-free paraformaldehyde (Pierce) for exactly 10 min, at room temperature. The cell monolayers were then quenched with 0.125 M glycine for exactly 10 min at room temperature, washed with PBS, harvested by scraping and flash-frozen. Nuclei were extracted resuspending each cell pellet in 5 mL of LB1 (50 mM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100) for 10 min on ice. The nuclei were spun down at 350 rcf for 5 min at 4°C and washed with 5 mL of LB2 (10 mM Tris-HCl pH 8, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA) shaking for 10 min at room temperature. Finally, the nuclei were resuspended in 1 mL of LB3 (10 mM Tris-HCl pH 8, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine) and sonicated with Covaris M220 (Peak Power: 75 w, Cycles/burst: 200, Duty factor: 10%, 150 s ON 10 s OFF). LB1, LB2, LB3 were freshly prepared and supplemented with protease (cOmplete Protease Inhibitor Cocktail, Sigma- Aldrich) and phosphatase inhibitors (PhosSTOP, Sigma- Aldrich). After sonication, 5 μL of each sample was decrosslinked at 65°C in TE+2% SDS for 30 min. DNA was extracted with MinElute PCR Purification Kit (QIAGEN) and checked at Bioanalyzer (Agilent) using High Sensitivity DNA Kit (Agilent). Samples were diluted to a final volume of 3 mL with LB3 and, after addition of 300 μL of 10% Triton X-100, were spun down at full speed for 10 min at 4°C to pellet debris. Supernatants were collected, aliquots of cell lysates were saved as whole cell DNA and 1 mL of cell lysate per sample was incubated overnight on a rotating wheel with a mix of 1:1 Protein A-Dynabeads (Thermo Fisher) and Protein B-Dynabeads (Thermo Fisher Scientific) previously incubated with either ZEB1 antibodies (cat# PA528221) (Thermo Fisher Scientific) or cat# NBP-05987 (Novus Biologicals) or normal Rabbit IgG (cat# 2729, Cell Signaling Technology). On the following day, samples were washed 6 times with Wash Buffer (50 mM HEPES pH 7.6, 500 mM LiCl, 1 mM EDTA, 1% NP-40, 0.7% Na-Deoxycholate) and once with TE+50 mM NaCl. Beads, as well as 10 μL of whole cell DNA, were resuspended in 150 μL of TE+2% SDS and decrosslinked overnight at 65°C in presence of Proteinase K (Thermo Fisher Scientific). The following day, DNA was extracted using MinElute PCR Purification Kit (QIAGEN) and qPCR was run in technical triplicates 3 μL of DNA, 2 μL of 20 μM primers (Forward+Reverse) and 5 μL of PowerUp SYBR Green Master Mix per reaction. Data were analyzed as percentage of enrichment over input.

For ChIP-qPCRs, the following primer pairs were used:

GRHL2:
Forward 5′- AACCAGTAGCCTCCACCTTC-3′
Reverse 5′- CCTTCTCCGAGTGCAGGTAA-3′
ESRP1:
Forward 5′- TGTCTCTTACCTGCACCACG-3′
Reverse 5′- GGCGTCTATGCAAAAAGCCT-3′
MIR200 A/B:
Forward 5′- GAGAAGCCCAGGAGCAAGTA-3′
Reverse 5′- TGGGTGGGGTGTGCTCAG-3′
Negative control locus:
Forward 5′- AATGTTGGGCCTTGAAACAG-3′
Reverse 5′- CCAGTGTGGTCCAAAGAGGT-3′

Immunoblotting

Adherent cells were harvested by scraping on ice after 1 wash with PBS. Collected cell pellets were resuspended in RIPA buffer (50 mM Tris pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% sodium dodecyl sulfate and 5 mM EDTA pH 8.0) containing a protease and a phosphatase inhibitor (Sigma-Aldrich). Protein concentration was measured using the DC Protein Assay (Bio-Rad Laboratories). 10 μg protein samples were separated on 10% SDS-PAGE gels and transferred on PVDF membranes by wet-blot transfer. Immunoblotting was performed using the following primary antibodies: E-Cadherin (clone EP700Y, GeneTex, diluted 1:25,000), Vimentin (clone V9, Abnova, diluted 1:1,000), Twist1 (clone Twist2C1a, Santa Cruz, diluted 1:200), ZEB1 (clone H-102, Santa Cruz Biotechnology, diluted 1:200) or ZEB1 (Sigma-Aldrich, diluted 1:5,000), Beta-actin (clone AC-15, Abcam, diluted 1:6,000) was used as a loading control. Histone H3 (rabbit polyclonal ab1791, Abcam, diluted 1:5,000) was used as loading control for isolated nuclear fractions. Alpha-tubulin (clone B-5-1-2, Sigma-Aldrich, diluted 1:5,000) was used as loading control for isolated cytoplasmatic fractions. Secondary antibodies (anti-mouse and anti-rabbit) were conjugated to horseradish peroxidase (Jackson Immuno Research, diluted 1:12,500). Chemiluminescence reaction was activated with ECL Prime Western Blotting Detection Reagent (GE Healthcare) and detection was carried out on a ChemiDoc Imaging System using Image Lab software (Bio-Rad Laboratories).

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis and data visualization were performed using GraphPad Prism software, except for ATAC-seq and RNA-seq analysis and visualization, which were carried out using R. The frequency of TPCs or MS-forming cells were both calculated using the ELDA (Extreme Limiting Dilution Analysis) method and assuming a confidence interval of 95%.[75] For comparison between two sample groups, statistical analysis was conducted using the two-tailed unpaired Student's t-test. A p value <0.05 was used as cut-off for significance. Data are generally presented as mean ± SEM of n = x experiments, with x indicating in figure legends the number of independent experiments performed, unless otherwise stated.