Epithelial-mesenchymal plasticity determines estrogen receptor positive breast cancer dormancy and epithelial reconversion drives recurrence

Patrick Aouad1, Yueyun Zhang1, Fabio De Martino1, Céline Stibolt1, Simak Ali2, Giovanna Ambrosini3, Sendurai A. Mani3, Kelly Maggs4, Hazel M. Quinn1, George Siflomos5 & Cathrin Brisken1,5

More than 70% of human breast cancers (BCs) are estrogen receptor α-positive (ER+). A clinical challenge of ER+ BC is that they can recur decades after initial treatments. Mechanisms governing latent disease remain elusive due to lack of adequate in vivo models. We compare intraductal xenografts of ER+ and triple-negative (TN) BC cells and demonstrate that disseminated TNBC cells proliferate similarly as TNBC cells at the primary site whereas disseminated ER+ BC cells proliferate slower, they decrease CDH1 and increase ZEB1,2 expressions, and exhibit characteristics of epithelial-mesenchymal plasticity (EMP) and dormancy. Forced E-cadherin expression overcomes ER+ BC dormancy. Cytokine signalings are enriched in more active versus inactive disseminated tumour cells, suggesting microenvironmental triggers for awakening. We conclude that intraductal xenografts model ER+ BC dormancy and reveal that EMP is essential for the generation of a dormant cell state and that targeting exit from EMP has therapeutic potential.

Breast cancer (BC) is the most commonly diagnosed malignancy worldwide1. Clinical management of BC relies on grade, stage and tumor subtype, which can be defined by histology and immunohistochemistry (IHC) as well as by gene expression profiling. More than 70% of BC cases are estrogen receptor α-positive (ER+). ER+ BCs tend to be of lower grade and lower proliferative indices than ER−, including HER2 and ER/PR HER2− (triple-negative, TN), BCs. Accordingly, patients with ER+ BC have better 5-year survival rates than patients with ER− BCs. However, while ER+ BC patients who do not relapse within the first five years after treatments are generally considered disease-free, ER+ BC patients remain lifelong at risk for relapses despite benefiting from endocrine therapy initially2.

Disseminated tumor cells (DTCs) that leave primary tumors early during tumorigenesis and remain dormant at distant sites contribute to delayed recurrence3. The existence of dormant DTCs is supported by clinical observations that recipients of organs transplanted from individuals with undiagnosed malignancies developed the corresponding types of tumor4-5. It is critical to understand the mechanisms governing dormancy for preventative intervention; however, the issue is understudied due to the lack of suitable in vivo models6-9. Most studies of the mechanisms underlying metastasis in vivo used genetically engineered mouse models (GEMMs) or xenograft models. Regarding mammary tumorigenesis, most GEMMs develop ER− tumors10-12 and most xenograft studies have relied on subcutaneous...
grafting or injecting BC cells directly into the bloodstream or at distant sites but these approaches poorly reflect the clinical situation.

We recently reported that grafting human ER+ BC cell lines and patient-derived xenografts (PDXs) into the milk ducts of immune-compromised mice (MIND) substantially improves take rates over the traditional subcutaneous engraftment. MIND models recapitulate human ER+ BC progression, from in situ stage to spontaneous dissemination to clinically relevant distant organs and reflect specific histopathological subtypes, including the invasive lobular carcinoma. Bioluminescence emanating from ER+ DTCs can be detected in different organs when tumors still appear to be confined within the milk ducts, suggesting early dissemination. While DTC load increases with prolonged tumor growth, no macro-metastasis is detected in up to seven months after engrafting MCF-7 cells and -1 year after engrafting ER+ PDXs. Here, we compare ER+ and TN BC progression in vivo using MIND models and show that dormancy is specific for ER+ DTCs and is driven by Epithelial-Mesenchymal Plasticity (EMP). Forced exit from EMP can overcome dormancy, providing a potential target against latent disease.

**Results**

**Tumor growth and progression in ER+ versus TN MIND models**

In ER+ BC MIND models, DTCs detected in clinically relevant distant organs failed to progress to macro-metastases. To determine whether this observation reflects ER+ BC tumor dormancy or is an experimental artifact, we compared the metastatic behavior of ER+ and ER- BC cells by the MIND approach. We modeled ER+ BC using TNBC cell lines BT20 and HCC1806 (Table 1) as well as a PDX derived from an untreated primary TNBC, T70 (Table 2). To model ER- BC, we used the established ER+ cell lines MCF-7 and T47D (Tables 1) and 2 ER+PR+/HER2- DTXs: T99 derived from an untreated primary tumor and METS15 derived from the ascites of a patient with advanced disease (Table 2). All tumor cells were infected with lentiviruses expressing GFP-Luc2 or RFP-Luc2 and injected into the milk ducts of NOD.Cg-Prkdc<sup>-/-</sup>Il2rg<sup>+/null</sup>SzJ (NSG) females (Fig. 1a). To take rates, defined as the percentage of glands showing in vivo tumor cell growth relative to the total number of glands injected, were 100% for TNBC and ≥90% for ER+ BC xenografts (Fig. 1b). Within five weeks, mice xenografted with BT20, HCC1806, and T70 developed palpable mammary tumors, increased bioluminescence 702, 592, and 1915 folds (Fig. 1c, d). At this stage, ex vivo bioluminescence of the lungs was comparable to those of primary tumors (Fig. 1c, d). At the respective experimental endpoints, primary tumor burdens were comparable, but bioluminescence in the lungs relative to that of the mammary glands was on average 100-fold higher in TNBC models than in ER+ BC models (Fig. 1h). Thus, metastases in mice with TNBC xenografts progressed faster than in mice with ER+ grafts, reflecting the clinically observed subtype difference in BC metastasis.

We then examined early tumor cell seeding by fluorescence stereo microscopy of mammary glands xenografted with BT20 and HCC1806 cells. One week after intraductal injection, the GFP signal was confined to the ducts (Supplementary Fig. 1a, b) and histological examination confirmed that tumor cells were within the ducts as is characteristic of the in situ stage of the human disease (Supplementary Fig. 1c, d). No disruption of the epithelial-stromal border was detected but the fibrous extracellular matrix (ECM) around ducts filled with tumor cells was thicker than around those without human cells (Supplementary Fig. 1c, d). At this stage, ex vivo bioluminescence of the lungs was just above background levels in 3 out of 10 mice (Fig. 1g), consistent with previous observations. The at the respective experimental endpoints, primary tumor burdens were comparable, but bioluminescence in the lungs relative to that of the mammary glands was on average 100-fold higher in TNBC models than in ER+ BC models (Fig. 1h). Thus, metastases in mice with TNBC xenografts progressed faster than in mice with ER+ grafts, reflecting the clinically observed subtype difference in BC metastasis.

To test for dormancy, we co-immuno-fluorescence labeled lungs for p27, Ki67, and CK8. p27 is a marker of dormancy whereas CK8 was used to distinguish DTCs (CK8+) from the surrounding mouse alveolar cells (CK8<sup>-/null</sup>). Less than 1% of CK8+ TNBC cells were p27<sup>+</sup> Ki67<sup>-</sup> whereas more than 10% of ER+ BC cells were p27<sup>+</sup> Ki67<sup>-</sup>.

**Table 1 | Cell lines used in this study**

| Cell line | Age (years) | Source | ER | PR | HER2 | Number of passages after obtaining from ATCC |
|-----------|-------------|--------|----|----|------|--------------------------------------------|
| MCF-7     | 69          | Pleural Effusion | + | + | -   | 7-30                                       |
| T47D      | 54          | Pleural Effusion | + | +, Amp | - | 6-8                                        |
| BT20      | 74          | Primary Tumor | - | - | -   | 6-11                                       |
| HCC1806   | 60          | Primary Tumor | - | - | -   | 20-23                                      |

Amp amplification, + positive, - negative.

---

Nature Communications | (2022) 13:4975

[Article](https://doi.org/10.1038/s41467-022-32523-6)
Fig. 1 | ER+ and TN BC cells show distinct growth and metastatic behavior. a Scheme illustrating the intraductal xenografting approach used in this study. H&E: Haematoxylin & Eosin, IF: Immunofluorescence. b Bar graph showing take rates for TN (blue) and ER+ (red) BC cells injected intraductally, 13–19 mammary glands of 5–9 mice injected in each group. The vertical dashed line indicates 90%. c Graph showing the fold-change of bioluminescence over time for all intraductal xenografts. Data represent mean ± SEM of 13 xenografts. d Fold-change of bioluminescence at 5 weeks after intraductal injection. Data represent mean ± SEM of 13–19 mammary glands from 5 to 9 mice in each group. Blue and red dashed lines represent the average change bioluminescence of TN (1148 fold) and ER+ BC (32 fold), respectively. One-way ANOVA, Kruskal-Wallis test relative to MCF-7 (control). e Box plot showing the fold-change of bioluminescence at endpoint for TN (5 weeks) and ER+ (5–6 months) BC cells from Fig. 1c. Boxes span the 25th to 75th percentile, whiskers 1.5 times the inter-quartile range. Boxplot whiskers show minimum and maximum values. Two-tailed Mann-Whitney test. f, g Bar plot showing ex vivo bioluminescence of resected organs from 9, 7, and 6 mice bearing BT20, HCC1806, and T70 xenografts, respectively, f and 20, 9, 10, and 14 mice bearing MCF-7, T47D, T99, METS15 xenografts, respectively, g Data represent mean ± SEM. h Dot plot showing the ratio of bioluminescence in lungs over primary tumor. Data represent mean ± SEM of n = 9 (BT20), 7 (HCC1806), 12 (MCF-7), 8 (T47D), 3 (T99) and 12 (METS15) mice. Blue and red dashed lines represent the average ratio for TN and ER+ BC cells, respectively.

(Fig. 2m, n and Supplementary Fig. 3a, b). Thus, a substantial fraction of ER+ DTCs are arrested in the G0/G1 phase of the cell cycle as characteristic of a quiescent and/or dormant state.

DTCs have features of epithelial-mesenchymal plasticity (EMP) Individual ER+ lung DTCs appeared mesenchymal (Fig. 3a), prompting us to quantify the cellular aspect ratio (CAR), as major axis: minor axis21, of tumor cells. A CAR > 1.7 characteristic of mesenchymal cells characterized 10–15% of tumor cells at the primary site Fig. 3b and 40–50% of the DTCs in the lungs (Fig. 3c). Next, we analyzed the E-cad expression by IF. The E-cad protein was readily detected in primary ER+ BC cells but signal intensity was reduced to <10% in matched lung DTCs, which were identified based on CK8 expression (Fig. 3d and Supplementary Fig. 4a–c). We micro-dissected fluorescent lesions from distant sites and compared their mRNA levels to those of the respective primary tumors by real-time PCR (RT-PCR) analysis using human-specific primers. MKI67 and CDH1 transcript levels were significantly lower in distant lesions than in primary tumors (Fig. 3e, f,

Table 2 | PDXs and corresponding patient tumor characteristics

| PDX   | Age at Surgery | Tumor Type | ER% | PR% | HER2% | Ki67% | Treatment | Generation in Mice |
|-------|----------------|------------|-----|-----|-------|-------|-----------|-------------------|
| T70   | 39             | NST, Primary | negative | negative | negative | >90 | untreated | 6 to 8            |
| T99   | 57             | NST, Primary | 57   | 70  | negative | 20   | untreated | 6 to 8            |
| METS15| 59             | NST, Ascites | 90   | 100 | negative | N/A  | chemo, AI, fulvestrant | 2 to 4          |

NST no special type, N/A not applicable, chemo chemotherapy, AI aromatase inhibitors.
**Fig. 2 | ER+ metastatic lesions are dormant.**

**a** Representative fluorescence stereo micrographs of lungs from ≥3 mice with BT20, HCC1806, MCF-7, or T47D intraductal xenografts, arrows point to DTCs. Scale bar, 1 mm.

**b** Bar plot showing percentage of cycling and non-cycling cells in primary tumors and in lung micro-metastases in MCF-7 intraductal xenografts-bearing mice. Data represent mean ± SD from 3 host mice. Paired t-test.

**c** Representative fluorescence stereo micrographs of the liver (d) and lungs (e) from ≥3 mice bearing T70 and METS15 MINO xenografts. Scale bars, 1 mm.

**d** Representative fluorescence stereo micrographs of H&E stained lung sections from ≥3 mice bearing MCF-7 intraductal xenografts. Scale bars, 50 μm.

**e** Percentage of Ki67+ cells in matched primary and lung sections from mice 5 weeks after intraductal injection of BT20 and HCC1806 cells. n = 14 sections.

**f** Bar graph showing the percentage of p27+Ki67- cells over total human cells in the lung. Data represent mean ± SD from n ≥ 3 mice, and dots represent ≥40 cells analyzed. One-way ANOVA.

**g** Representative immuno fluorescence micrographs for CK8 (blue), p27 (magenta) and Ki67 (cyan), counterstained with DAPI (gray) on lung section from MCF-7-bearing mice. Scale bar, 50 μm; inlet, 20 μm. *, ***, ****, and n.s represent P < 0.05, 0.001, 0.0001, and not significant, respectively.
**Fig. 3 | Dormant DTCs from ER+ intraductal xenografts have an EMP signature.**

- **a** CK8 staining on optically-cleared sections of METS15 cells in host's lungs. Scale bars, 10 µm.
- **b** Representative masks representing different cellular aspect-ratio (CAR) with respect to the morphology.
- **c** Percentage of cells with CAR > 1.7 in matched primary tumors and in the lung from intraductal ER+ BC xenograft-bearing mice. Data represent mean ± SD, each data point represents at least 100 primary cells and 10 lung DTCs from 3 mice. Paired Student's t-test.
- **d** Relative E-cad intensity (Int.) in matched primary tumor cells and lung DTCs in mice bearing indicated intraductal xenografts. Data represent n≥8 images, mean±SD from 3 mice. Student's t-test.
- **e** Relative MKI67 mRNA levels in MCF-7 cells in the primary tumor (n = 8 mice) or in the lung (n = 8), brain (n = 4), and liver (n = 3) DTCs. Data represent mean ± SD. One-way ANOVA.
- **f** Relative levels of indicated mRNAs in matched BT20 or HCC1806 primary tumors and lung metastases from 3 mice. Paired t-test.
- **g** Representative fluorescence stereo micrographs of lungs from at least 3 mice 3 weeks after intraductal injection of BT20 or HCC1806 cells. Arrows point to micro-metastases. Scale bar, 1 mm.
- **h** Relative mRNA levels of the selected genes in 3 versus 5-6 weeks lung DTCs retrieved from at least 4 mice bearing BT20 (left) or HCC1806 (right), n ≥ 4. Data represent mean ± SD, Wilcoxon test. Gene expression was normalized to the geometric mean of GAPDH and HPRT in panels e–j, and l. *, **, ***, ****, and n.s. represent P < 0.05, 0.01, 0.001, 0.0001, and not significant, respectively.
Supplementary Fig. 4d). Transcript levels of the EMT transcription factors (EMT-TFs), ZEB1 and ZEB2, and the mesenchymal intermediate filament VIM were significantly higher in lung lesions than in primary tumors (Fig. 3g, Supplementary Fig. 4d). SNAI2 and TWIST1 mRNA levels did not differ between primary tumors and metastatic lesions whereas SNAI1 expression was down-modulated in metastatic lesions in the MCF-7 model (Fig. 3g). CDH2, SNAI1, SNAI2, and TWIST1 transcripts were not detected in either primary tumors or lung lesions in the METS15 model. ZEB1, ZEB2, and VIM transcript levels were also up-regulated in brain DTCs (Fig. 3h). In the 3 liver lesions that we isolated, VIM transcripts were significantly up-regulated (P < 0.05). ZEB1 and ZEB2 transcript levels showed a trend to increase (P < 0.12 and P < 0.25, respectively) (Fig. 3i). As an alternative approach to microdissection, we performed qRT-PCR using primary tumours and lung tissues from METS15 xenograft-bearing mice. Similar alterations in expression levels of human-specific genes in lung DTCs compared to their matched primary tumors were observed (Supplementary Fig. 4e).

In BT20 and HCC1S06 models, MKRβ7, CDH1, CDH2, ZEB1, SNAI1, SNAI2, VIM, and TWIST1 transcript levels were comparable between matched primary site and lungs with macro-metastases (Fig. 3j), suggesting that dormancy was not a feature of TNBC metastases. To determine whether TN DTCs pass through a dormant stage prior to becoming micro-metastases, we compared expression levels of epithelial and EMP-metastatic genes in the lungs at 3 weeks, when micrometastases prevailed (Fig. 3k), versus 5-6 weeks, when macro-metastases were detected, after intraductal injection with BT20 and HCC1S06 cells. At both time points, similar expression levels of MKRβ7, CDH1, EPCAM, and EMT-TFs, and VIM were observed in lungs (Fig. 3l). In addition, the proliferative TNBC cells were in comparable EMP states in the primary (Supplementary Fig. 4o) and distant sites. Thus, mesenchymal morphology and expression of multiple EMP markers are features specific to ER+ DTCs.

Role of EMP in ER+ tumor progression
Accumulating evidence supports the notion that tumor cells assume different EMP states that are important for their invasion and metastasis24-27. To assess whether EMP occurred during the transition from in situ to invasive stage in ER+ BC, we measured CDH1 and EMT-TFs expression 1 month after intraductal injection of MCF-7 cells, when primary tumors were in situ, and at 5 months, when they were invasive (Fig. 4a). CDH1, ZEB1, and VIM transcript levels did not change whereas the ZEB2 expression decreased and TWIST1 and SNAI2 expressions increased at the invasive stage (Fig. 4b). If showed that MCF-7 cells expressed E-cad and CK8 proteins in situ and invasive lesions (Fig. 4c) as well as in areas of tumor budding, at the leading invasive edge, and in proximity to blood vessels (Supplementary Fig. 5a-c). Thus, EMP features characteristic of ER+ DTCs are not readily detected at the primary site in our models.

To test whether EMP favored the metastatic spread of ER+ BC cells early in tumor progression, we induced EMP by reducing CDH1 expression or by overexpressing ZEB1. Given the heterogeneity of the lactively transduced polyclonal cell populations, we expected that a whole spectrum of EMP states would be produced. Control cells formed epithelial islands with cobblestone morphology whereas MCF-7:shCDH1 cells showed dis cohesive growth (Supplementary Fig. 5d, e), decreased cell proliferation (Supplementary Fig. 5f) and increased the ZEB1 levels. Expression of ESR1, PGR or AR was unaffected (Supplementary Fig. 5g). ZEB1 overexpression (Supplementary Fig. 3h) had similar effects on cell morphology and proliferation (Supplementary Fig. 5h-j), reduced CDH1, ESR1, and PGR expression, and increased CDH2 expression (Supplementary Fig. 5k). Thus, both MCF-7:shCDH1 and MCF-7:ZEB1 cells acquire distinct EMP features defined by recent guidelines9,28.

When grafted intraductally, MCF-7:shCDH1 cells grew less than MCF-7:shSCR controls (Fig. 4d). Mammary gland weight and E-cad expression at end-point were reduced (Supplementary Fig. 6a-c). Analysis of H&E stained sections showed that MCF-7:shSCR cells invaded the stroma whereas MCF-7:shCDH1 cells remained mostly in situ (Fig. 4e). Image analysis of picrosirius red staining revealed increased fibrillar collagen deposition in MCF-7:shCDH1 xenografts, validating that experimental CDH1 down-modulation induced a functional EMT in vivo29-31 (Fig. 4f-g). Molecular analyses confirmed reduced CDH1 and increased ZEB1 transcripts, while the expression of other EMP-TFs was unaffected (Fig. 4h).

CDH1 knockdown reduced p120 and β-catenin proteins at cell junctions (Supplementary Fig. 6d). K667 and pH43 indices were decreased whereas the apoptotic index was increased in MCF-7:shCDH1 compared to MCF-7:shSCR grafts (Supplementary Fig. 6e, f). The micro-metastatic load of mice engrafted with MCF-7:shCDH1 cells was around 10% of that of mice with MCF-7:shSCR cells (Fig. 4i and Supplementary Fig. 6g). To exclude the possibility that the reduced metastatic burden merely reflected decreased primary tumor growth, we analyzed the micro-metastatic burden in mice with comparable primary tumor burden. In mice paired accordingly, MCF-7:shCDH1 cells caused smaller micro-metastatic loads (Supplementary Fig. 6h) indicating that decreased CDH1 expression and the resulting EMP features do not increase the metastatic propensity of ER+ BC cells.

To exclude any potential systemic effects of genetically different tumor cells on the host, we grafted MCF-7:shSCR:GFP and MCF-7:shCDH1:RFP cells contralaterally (Supplementary Fig. 6i). The signal emanating from glands engrafted with MCF-7:shCDH1:RFP cells at endpoint was lower than that from contralateral glands engrafted with MCF-7:shSCR:GFP cells (Supplementary Fig. 6j). The number of shCDH1:RFP+ lung foci was less than that of GFP+ loci and the total GFP+ area was smaller than the GFP+ area (Supplementary Fig. 6k,l). These results again support the hypothesis that E-cad down-modulation with associated EMP features does not favor metastasis in ER+ BC.

Similarly, MCF-7:ZEB1 xenografts grew less and the engrafted glands weighed lower than MCF-7:Ctrl xenografts at four months (Fig. 4j, Supplementary Fig. 6m). Histology revealed that control cells invaded the muscle tissue adjacent to the thoracic mammary gland whereas ZEB1-overexpressing cells remained in situ with occasional microinvasion foci (Fig. 4k). The fibrillar collagen deposits were larger in MCF-7:ZEB1 than in MCF-7:Ctrl xenografts (Fig. 4l, m). A trend for an increased apoptotic index was observed upon ZEB1 overexpression (Supplementary Fig. 6n). ZEB1 overexpression decreased levels of E-cad, p120, and β-Catenin proteins and occasional Vim+ cells were detected (Supplementary Fig. 6o). Again, the forced EMP state reduced the micro-metastatic burden (Fig. 4n, Supplementary Fig. 6p).

Overexpression of other EMP-TFs such as ZEB2, TWIST1, and SNAI1 similarly reduced tumor growth and metastatic load (Supplementary Fig. 6q, r) arguing that different EMP states induced by distinct means in MCF-7 cells neither favor tumor progression nor metastasis.

EMP in ER+ PDxs
To assess the relevance of these findings in MCF-7 cells to ER+ BC in general, we down-modulated E-cad expression in T99 and METS15 cells. In vivo growth (Fig. 5a) and mammary gland weight at end-point were reduced (Supplementary Fig. 6s). We confirmed reduced CDH1 mRNA levels in the xenografted glands at endpoint (Fig. 5b). Furthermore, E-cad down-modulation reduced tumor invasion (Fig. 5c) and the micro-metastatic burden (Fig. 5d and Supplementary Fig. 6c). Similarly, ectopic ZEB1 expression decreased T99 intraductal growth (Fig. 5e), the weight of xenografted glands (Supplementary Fig. 6u), and the micro-metastatic load in the lungs and bones (Fig 5f). Hence, an EMP per se, induced by genetic manipulation of individual genes, favors neither tumor progression nor metastasis of ER+ BC cells. This suggests that ER+ BC cells are plastic and the mesenchymal phenotype may be induced or stabilized after cells leave the primary tumor, possibly at the distant sites.
E-cadherin is sufficient to awaken DTCs from dormancy

Our finding that dormant DTCs are in, at least partially, mesenchymal states, prompted us to assess whether returning to the epithelial state may reactivate proliferation. To this aim, we dissociated mammary glands and lungs from NSG-EGFP mice engrafted with MCF-7:RFP cells to single cells, plated them in 2D and applied drug selection to avoid overgrowth of mouse cells. RFP+ MCF-7 cells derived from primary tumors proliferated within a few days and were confluent by 1-2 weeks (Supplementary Fig. 7a). The lung-derived DTCs resumed proliferation after 2 months and formed epithelial islets (Supplementary Fig. 7b). The brain DTCs, which had lower CDH1 transcript levels than the lung DTCs (Fig. 3f), took longer to emerge from quiescence and formed epithelial islets at 4 months (Supplementary Fig. 7c). CDH1 transcript levels were ultimately restored, and EMT-TFs, ZEB1, ZEB2 and VIM transcripts decreased after serial passages in culture (Supplementary Fig. 7d–g), consistent with the hypothesis that reacquisition of an epithelial state enables cell proliferation.

To address whether restoring an epithelial state is sufficient to awaken dormant DTCs in vivo, we overexpressed E-cad in MCF-7 cells (Supplementary Fig. 7h). Bioluminescence one day after intraductal
engraftment was 2-fold higher in glands engrafted with MCF-7:CDH1 cells compared to controls, suggesting that E-cad favors tumor cell survival and/or engraftment in the mouse milk ducts (Fig. 6a). Six months after injection, however, bioluminescence in MCF-7:CDH1 grafts was 2-fold lower than in the controls (Fig. 6b). Similar to the 10-fold increase initially observed in vitro (Supplementary Fig. 7h), an 8-fold increase in CDH1 transcript levels was observed (Fig. 6c). Ex vivo bioluminescence in the lungs from mice with MCF-7:CDH1 xenografts was 10% of that of MCF-7:Ctrl xenograft bearing mice (Fig. 6d). Yet, a single lung lesion > 500 µm in diameter was detected in 1 of the 4 mice engrafted with MCF-7:CDH1 (Fig. 6e) when such a large lesion was never observed in over 60 mice engrafted with parental MCF-7 cells suggesting that E-cad expression may overcome dormancy.

To circumvent potential confounding effects of E-cad overexpression on engraftment and in vivo growth of MCF-7 cells at the primary site, we used a doxycycline (DOX) inducible E-cad expressing T99 intraductal xenografts. Data represent mean ± SEM of 15 and 18 xenograft glands, respectively, for T99 and METS15 intraductal xenografts-bearing mice. Dashed line indicates the median, and dotted lines indicate the lower and upper quartiles. Each dot represents a single organ, the dashed line indicates the median and dotted lines indicate the lower and upper quartiles. Unpaired Student’s t-test. e Growth curve of vector control and ZEB1-overexpressing T99 intraductal xenografts. Data represent mean ± SEM of 15 and 18 xenograft glands, respectively. Two-way ANOVA, multiple comparisons. f Violin plot of the relative micro-metastatic burden in lungs and bones from 4 control and 5 ZEB1-overexpressing T99 xenografts-bearing mice. The dashed line indicates the median and dotted lines indicate the lower and upper quartiles. Unpaired Student’s t-test. *, **, ***, ****, and n.s. represent P < 0.05, 0.01, 0.001, 0.0001, and not significant, respectively.
chow for 2-3 months (Fig. 6f). Bioluminescence and weight of mammary glands at sacrifice were significantly reduced in the DOX-induced group (P < 0.01) (Fig. 6g, h) but lung bioluminescence and the overall bioluminescence in distant organs were significantly increased (Fig. 6i, j). DOX treatment per se did not affect the weight of mice (Supplementary Fig. 7j), tumor growth, nor the weight of the engrafted glands nor the micro-metastatic burden at endpoint (Supplementary Fig. 7k-m). Thus, ectopic E-cad expression is sufficient to drive DTCs out of dormancy.

To assess whether the increased lung bioluminescence was due to more seeding and/or increased size of individual lesions, we measured micro-metastases by fluorescence stereo microscopy and histology. CDH1 induction increased the overall fluorescent area (Fig. 6k, l) as well as the area and length of lung lesions (Fig. 6m, n). The anti-E-cad IF
signal intensity of lung micro-metastases was 5.5-fold higher in +DOX than in -DOX mice (Fig. 6o). Semi-quantitative RT-PCR on lungs confirmed the induction of CDH1 and revealed increased MKI67 transcript levels in the +DOX group (Fig. 6p). CDH1 expression correlated positively with MKI67 expression (P < 0.001) (Fig. 6q), supporting that E-cad restoration drove cell cycle entry and proliferation. Finally, we asked whether the increase in proliferation in DOX-treated mice involved loss of mesenchymal phenotype. We noted a significant decrease in CDH2 and VIM transcript levels (Fig. 6r), a trend to decrease ZEB2 and SNAI2 transcript levels, and unaltered SNAI1 transcript levels (Fig. 6s). ZEB2 was not reliably detected by the semi-quantitative RT-PCR approach (cycle threshold >37).

Similar experiments using T47D/E-cadKO cells (Fig. 6t) showed that overexpressing E-cad did not significantly affect primary tumor growth nor mammary gland weight (Supplementary Fig. 7n, o) but increased the metastatic load 3-fold (Supplementary Fig. 7p). Analysis of fluorescence stereographs showed a 10-fold increase in area and fluorescence stereographs showed a 10-fold increase in area and fluorescence intensity of lesions upon E-cad restoration (Supplementary Fig. 9a). The active clusters had positive GSEA scores for proliferation-related pathways, ER response, MTORC1 signaling, and oxidative phosphorylation (Supplementary Fig. 9c). IHC ER labeling index was reduced in lung DTCs compared to primary tumor cells (Supplementary Fig. 9d,e) suggesting decreased ER expression may result in reduced ER target gene expression. Cluster 2 contained some MKI67+ and MCM2+ cells while negatively enriched for an EMT signature (Fig. 7g). The inactive cluster 3 had increased expression of genes previously implicated in dormancy, including ELN, COL1A1, COL3A1, FNL, BMP92, and CXCL12 (Supplementary Fig. 9f). Together these findings suggest that DTCs have different degrees of EMP and dormancy, with cluster 3 representing the most dormant one.

These findings extended to the lung DTCs of METS15, in which CXCL12, BMP92, TGFB12, FNL, and COL3A1 transcripts were upregulated, and COL3A1 to a lesser extent (Supplementary Fig. 9g). Experimentally induced EMP by CDH1 knockdown or ZEB1 overexpression upregulated COL3A1 and CXCL12 while ZEB2 overexpression increased FNL (Supplementary Fig. 9h).

To test the hypothesis that mesenchymal-like DTCs transit to an epithelial state during awakening, we assessed dynamic changes in mRNA expression with RNA Velocity, which models the instantaneous rate-of-change of RNA expression based on the balance of spliced and unspliced RNA molecules. At the primary site the directionality of cells projected in the UMAP represented by velocity vectors varies between the two subpopulations with the inactive cluster (7) not showing any directionality (Supplementary Fig. 10a). In lung DTCs, the velocity vectors of cells between the inactive and active clusters, at the EMP bridge, point from the mesenchymal towards the epithelial clusters (Supplementary Fig. 10b). In addition, the phase plots of epithelial genes, CDH1, EPCAM, ESRI, and KRT18 indicate transcription induction (Supplementary Fig. 10c) whereas those of the mesenchymal markers, VIM, S100A4, and TCF4 show repression (Supplementary Fig. 10d). Lending further support to the hypothesis that ER+ dormant DTC awakening requires a transition from a mesenchymal to an epithelial state.

To gain insights into the signaling pathways that control awakening, we compared global gene expression between the active and inactive clusters. As expected, GSEA for hallmarks showed that the active populations were enriched for DNA repair, oxidative phosphorylation, E2F targets, mTOR and estrogen responses whereas the inactive populations were enriched for EMT (Supplementary Table 4). II.6JAK/STAT3 signaling and TNFα signaling via NFκB were positively
enriched in the active clusters (Fig. 7h), suggesting that cytokine signaling, likely from the microenvironment, triggers tumor cell awakening. Indeed, IL6-induced STAT3 activation increases metastatic burden in T47D intraductally xenografted mice. Analysis of pSTAT3 ChIP-Seq data from MCF-7 and T47D cells shows IL6-induced pSTAT3 binding to CDH1 and ESR1 regulatory regions (Fig. 7i and Supplementary Fig. 11a, b), in line with E-cad being a direct target of STAT3 activation that can determine awakening (Fig. 8).

Discussion
Clinically, dormancy and late recurrence represent important challenges in ER+ BC. Our understanding of the mechanisms controlling these phenomena is limited largely due to the lack of reliable pre-clinical models. We show that MIND xenografts address this need; they faithfully model the metastatic process of different BC subtypes; whereas TN DTCs keep characteristics of their primary tumors, ER+ DTCs exhibit different states of dormancy.
Fig. 7 | Heterogeneity of ER+ primary tumor cells and lung DTCs. a, b UMAP plots representing FACS-sorted MCF-7:RFP cells isolated from primary tumors (a) and lungs (b). Colors code for different cell clusters. c Violin plot showing “HALL-MARK EMT” gene set enrichment score in cell clusters of the primary tumors. Boxplots inside each violin describe the interquartile range, bold dots indicate median. Boxplot whiskers show minimum and maximum values. Statistical significance was assessed by fitting a generalized linear mixed model and computing the estimated marginal means across all the identified cell clusters. Tukey’s method was used to assess adjusted p-values. d Integrated UMAP plot showing MCF-7:RFP cells isolated from primary tumors (grey) and lung tissues (color-coded according to clusters from b). e–g UMAP plots representing levels of selected transcripts in primary MCF-7:RFP (e) and lung DTCs (f, g). Numbers describe centroid of clusters identified in Fig. 7a and b. Violet gradient represents low, yellow gradient high gene expression. h Ridge plot showing normalize enrichment scores (NES) in the active versus dormant clusters identified in lung DTCs for the “IL6_JAK_STAT3 SIGNALING” and “TNFA_SIGNALING_VIA_NFKB” hallmark gene sets. Statistical significance was assessed by Welch’s t-test. i UCSC genome browser screenshot showing IL6-induced pSTAT3 binding at CDHI promoter and enhancer regions. The browser window shows regions, highlighted in yellow, which are bound by phosphorylated STAT3 upon IL6 stimulation in MCF-7 and T47D cells. *, **, ***, ****, and n.s represent P < 0.05, 0.01, 0.001, 0.0001, and not significant, respectively.

Fig. 8 | Working Model for ER+ BC progression. ER+ tumor cells progress from in situ to invasive stage in an epithelial state characterized by high expression of E-cad and low expression of EMT-TFs. At distant sites, DTCs show EMP features and enter dormancy. Awakening from dormancy involves restoration of E-cad expression that may be triggered by cytokines resulting in STAT3 activation and suppression of the EMT-TFs. Drugs promoting EMP or inhibiting the reacquisition of an epithelial state may prevent disease recurrence. EMT-TFs Epithelial-Mesenchymal Transition-activating transcription factors, EMP Epithelial-Mesenchymal Plasticity, DTCs Disseminated Tumor Cells.

We observed differences in the biology of ER+ versus TN DTCs which provide a possible explanation for different utilities of adjuvant chemotherapy between ER+ BC and TNBC. Adjuvant chemotherapy is in general successful in reducing TNBC recurrence because it can kill proliferating DTCs, and hence incipient metastases, whereas it provides limited benefit to ER+ BC patients because the target cells have low proliferative indices.

EMP, shown here to be characteristic of ER+ BC dormant cells, has many important roles in tumor progression. It can drive invasion, migration, and metastasis. Our finding that reacquisition of an epithelial differentiation state is critical for the recurrence in ER+ BC (Fig. 8) is in line with a number of reports that the epithelial phenotype can instead be important for tumor cell dissemination, CTC survival, and metastatic outgrowth in different GEMMs and supports the notion that EMP may not always drive tumor progression but can have tumor type-dependent roles. Indeed, clinically, a partial EMT signature correlates with lymph node metastasis, reduced survival, and therapy resistance in head and neck carcinomas but no such correlation is detected in other cancer types. Likely a partial EMT exists in many contexts but the specific genes and the strength of the signature as well as the frequency and its role in metastasis vary.

The upregulation of SNAI2 and TWIST1 transcripts in the invasive versus in situ disease is likely a consequence not a cause of invasion; we observed it also when ER+ BC cells were directly injected into the fat pad. In line with cell type-specific EMP states, our findings suggest that in ER+ human BC cells, the critical EMT-TFs are ZEB1 and ZEB2 whereas SNAI2 and TWIST1, critical for EMP in other tissues, control a subset of EMP-driven genes and the strength of the signature as well as the frequency and its role in metastasis vary. While we find ZEB1/ZEB2 transcripts increased in all organs sites, further experiments are needed to determine organ-specific aspects of the EMP.

Our finding that both E-cad upregulation and down-modulation reduced the growth of primary tumors suggests that the E-cad levels need to be tightly regulated. The E-cad down-modulation observed in dormant DTCs is likely due to EMT-TFs directly binding the CDHI promoter and repressing its activity. The observation that ER+ PDXs, whether established from a primary tumor (T99) or a metastatic lesion (METSI5), disseminate and enter dormancy suggests that these
processes are not genetically but epigenetically controlled. Distinct epigenetic states may also account for the a priori surprising finding that cytokine signaling pathways previously implicated in dormancy induction may activate awakening. The transcriptional outcomes, and hence pro- versus anti-EMT effects, are likely chromatin context dependent. We have previously demonstrated that ER signaling is growth promoting during puberty but inhibitory during pregnancy in normal mouse mammary epithelial cells suggesting that epigenetic plasticity is an intrinsic property of ER$^+$ breast epithelial cells, be they normal or transformed.

A limitation of the MIND models used in the present study is the lack of an intact immune system in the host mice. The awakening process has been shown to involve multiple cell types$^{32,33}$ and secreted factors. Although NSG mice have some innate immune cells, functions of these cells might not be normal in the absence of B and T cells. Whether active immune cell signalings regulate the expression of CDK4 inhibitors or other effects also needs to be determined. This can ultimately be addressed by restoring a functional human immune system in host mice$^{34}$.

We failed to detect different EMP states in primary tumor cells. However, a minor cell cluster has a positive EMT enrichment score, with increased VIM, COLIA1, COLIA2, and PNI, and decreased CDH1 and proliferation markers. We were unable to unequivocally conclude whether these rare EMPs may be at the origin of DTCs because of our inability to isolate them based on specific cell surface markers for FACS-based purification. Given that the experimental induction of EMP did not increase metastatic dissemination and progression, we favor the hypothesis that the mesenchymal state is induced at the distant site. Barcoding experiments will help to distinguish between the two scenarios.

We present evidence that mesenchymal lung DTCs transit into an epithelial cell state using RNA velocity. This approach is intended for predicting cell states within a number of hours, likely a much shorter time-scale than the in vivo plasticity observed in patients. Yet, we are studying these DTCs at extremely late stages, where a number of cells are already transitioning from M to E. In addition to this, we cannot exclude that the majority of mesenchymal DTCs may have been lost during tissue dissociation. In order to distinguish between the two scenarios further studies would be required for example repeating these experiments using a lineage tracing barcoded library in combination with scRNA-seq.

The present findings suggest that stabilizing a mesenchymal cell state is critical to prevent recurrence. EMT-TFs, ZEB1/ZEB2, SNAIL, and TWIST, as well as EMT-inducing factors, such as TGF-β, have all been shown to halt the cell cycle by decreasing Cyclin D1 protein levels and increasing the phosphorylation of Rb$^{35-36}$. The tight link between EMT and cell cycle control suggests the intriguing possibility that underlying the success of CDK4 inhibitors may be a block of exit from dormancy. This and other hypotheses are now readily amenable to experimental testing with intraductal xenograft models.

**Methods**

**Clinical samples**

The Commission cantonale d’éthique de la recherche sur l’être humain (CER-VD 38/15, PB 2016-01185) approved this study, informed consent was obtained from all subjects. Tumor tissue was mechanically and enzymatically digested using parallel razor blades and collagenase, as previously described$^{13,14}$. Effusion samples (pleura or ascites) were enzymatically digested using parallel razor blades and collagenase, as described$^{39,60}$. The tight link between EMT and cell cycle control suggests the intriguing possibility that underlying the success of CDK4 inhibitors may be a block of exit from dormancy. This and other hypotheses are now readily amenable to experimental testing with intraductal xenograft models.

**Animal Experiments**

All mice were maintained and handled according to Swiss guidelines for animal safety with a 12-h-light-12-h-dark cycle, controlled temperature (22 ± 2°C) and humidity (55 ± 10%), and food and water ad libitum. Experiments were performed in accordance with protocol VD186.5 approved by the Service de la Consommation et des Affaires Vétérinaires, Canton de Vaud, Switzerland. NOD.Cg-Ptkcdc31ΔI22mutyny Sj (NSG) and NOD.Cg-Ptkcdc31ΔI22mutyny Tg(CAG-EGFP)1Osb/Sj (NSG-EGFP) mice were purchased from Charles River and The Jackson Laboratory. For MCF-7 intraductal xenografts, the maximal tumor radius was 2 × 10^3 psec/cm^2/sr, while for the rest of the intraductal xenografts, the maximal tumor radius was 1 × 10^0 psec/cm^2/sr or 1 cm. The maximal tumour size/burden was not exceeded. Most experiments were conducted using the NSG strain, unless stated otherwise in text. For the induction of E-cad, doxycycline (0.62 g/kg of food, SAFE 150 SP-25 www.safe-diets.com) was administered in the diet.

**Intraductal injections and re-transplantation**

Mice were anesthetized by intraperitoneal injection of 200 µl of 10 mg/kg xylazine and 90 mg/kg ketamine (Graeub). Intraductal injections of single-cell suspensions were performed as previously detailed$^{34,35}$. In all mice experiments, intraductal xenografts of cell lines and patient-derived cells were generated by injecting 1 × 10^4 and 2 × 10^4 cells, respectively, into the teats of 6–16-week-old NSG or NSG-EGFP female mice, and grown for 4–6 months. For re-transplantation of patient-derived cells, mammary glands were collected on ice-cold IX PBS, dissociated using parallel razor blades, and enzymatically digested using the tumor dissociation kit (Miltenyi Biotech) to generate single cells (human and mouse). To enrich for human cells the mouse depletion kit (Miltenyi Biotech) was used according to manufacturer’s instructions. Cells were counted and 2 × 10^5 cells were intraductally injected.

**Cell culture, cloning, and cell growth**

ER$^+$ BC cell lines MCF-7 and T47D as well as TN BC cell lines BT20 and HCC1806 were purchased from American Type Culture Collection (ATCC). MCF-7, T47D, BT20, and HCC1806 were maintained in Dulbecco’s modified Eagle’s medium (DMEM) medium (cat# 31966, Gibco) supplemented with 10% FCS (cat# 10270-106, Thermo Fisher Scientific Inc.) and penicillin/streptomycin (P/S, cat# 15140-063, Thermo Fisher Scientific Inc.). Parental and ESR1$^{-/-}$ (Y335S and D358G) MCF-7 cells were transduced with either ffLuc2/Turbo-GFP or ffLuc2/Turbo-RFP, and selected for the brightest fluorescent subpopulation by FACS sorting. shF3 and shZEB1, PsBT-GFP and PsBT-Zeb1 from Sendurul Mani (MD Anderson). Inducible E-cad vector was generated by annealing the ORF-CDH1 (Clone IDXIOH46767, Invitrogen) cassette into an inducible backbone (pLIX403, cat# 41395, Addgene, https://www.addgene.org/41395/) using the Gateway cloning strategy (cat# 12355, Invitrogen). Control eGFP, ZEB2, TWIST-1, and SNAI overexpression vectors from VectorBuilder (https://en.vectorbuilder.com/). pLKO.1 puro shRNA E-cad was a gift from R.A. Weinberg (Addgene plasmid # 18801) pLKO.1 puro, pLKO.1 shCDH1 were purchased from Addgene. pL3.7m-Clover-Geminin(1-110)-IRES-mKO2-Cdt(30-120) was a gift from Michael Lin (Addgene_83841; http://n2l.net/addgene:83841; RRID:Addgene_83841). Lentiviruses were produced as described$^{34}$, lenti-ONE GFP or RFP:Luc were designed by and lentiviruses were produced as described$^{34}$, lenti-ONE GFP or RFP:Luc were designed by and lentiviruses were produced as described$^{34}$.

**Tumor growth and metastasis analysis**

Tumor growth was monitored every other week by in vivo imaging system (IVIS, Caliper Life Sciences). Briefly, 12 min after intraperitoneal administration of 150 mg/kg luciferin (cat# L-8220, Biosynth AG), mice were anesthetized in an induction chamber (O2 and 2% isoflurane) and placed inside the machine. Images were acquired and analyzed using
Living Image Software version 4.4 (Caliper Life Sciences, Inc.). For ex vivo bioluminescence measurements, mice were first injected with 300 mg/kg luciferin for 7 min, then injected with 1 ml of 10 mg/kg xylazine and 90 mg/kg ketamine. Resected organs were then imaged 20 min after luciferin injection. Mammary glands were fixed in 4% paraformaldehyde (cat# 0335.3, Carl Roth) overnight at 4 °C or flash-frozen in liquid nitrogen for RNA or protein extraction.

**Immunohistochemistry and passive clarity**

Histological staining was performed as detailed previously. For passive clarity, tumor-bearing mammary glands and lungs were embedded in 5% agarose (w/v, PBS, cat# 16500500, Invitrogen), left at room temperature to solidify and agarose cubes then were removed from the plastic container and mounted for Vibratome sectioning (Leica VT1200 S) using glue (cat# 14460, Ted Pella, Inc.). The buffer tray was then filled with cold PBS. Blade travel speed was 0.8 mm/sec and minimum thickness of sections was 0.5 mm. To preserve endogenous tissue fluorescence, Rapiclear 1.52 solution (RC, SunJin Lab, https://www.sunjinalab.com/) was used according to the manufacturer’s instructions with some modifications. Tissues were permeabilized in 2% v/v Triton X100 (Sigma T8787) PBS overnight on a gentle rotor shaker at room temperature. Rapiclear 1.52 was pre-warmed at 37 °C and 2 ml were added on top of the tissues for 1 hr. Tissues were mounted between 2 coverslips using RC 1.52 and iSpacers for image acquisition and long-term storage. Z-stacks were acquired using a Zeiss equipped with a Leica DFC 340FX camera. Measurements were generated by ImageJ and CTCF was calculated by manually drawing an outline around the border of the tissue with no cells to serve as a measure of background noise. Images were processed with watershed segmentation to split adjacent cells, followed by cellular aspect ratio (CAR) quantification using the “analyse particles” function of Fiji (Size 50-infinity pixel², circularity 0-1). Mask for the blue channel was saved and the cell quantifications were output as a table for further data analysis. For analysis of mesenchymal-like cells in lungs, major and minor axis were drawn manually on each tumor cell for CAR quantification, >300 DTCs in lungs from each mouse, n = 3. Mesenchymal-like cells were defined as those with major axis:minor axis ratio (CAR) > 1.6. For analysis of fluorescence intensity, corrected total cell fluorescence (CTCF) was calculated by manually drawing an outline around the border of the cells using the drawing/selection tools in ImageJ, and a close-by region with no cells to serve as a measure of background noise. Measurements were generated by ImageJ and CTCF was calculated as: CTCF = integrated density (area x mean fluorescence background). For analysis of fibrillar collagen deposition, area of collagen was automatically quantified using ImageJ using size filter to exclude cells within the ducts.

Fluorescence images of whole mammary glands and lungs were acquired with a LEICA M205FA fluorescence stereo microscope equipped with a Leica DFC 340FX camera.

**Immunohistochemistry analysis**

For analysis of primary tumor staining, outlines of human cells were drawn manually with “Freehand selections” and a built-in function “waitForUser” for each image to exclude mouse cells, followed by ROIs creation. Channels were then binarized using “Huang” thresholding algorithm and images were processed with watershed segmentation to split closely touching cells and then denoised by a Minimum Filter with radius of 1 pixel. Cell numbers were quantified using the “analyse particles” function of Fiji (Size 0.01-infinity pixel², circularity 0-1), channels were saved, cell quantifications were output as a table for further data analysis. For analysis of metastatic BC cells in lungs, human cell nucleus and Ki67+ cells were counted manually.

**Immunoblotting**

Cell pellets and tumor-bearing mammary glands, and microdissected lung, liver, and brain microcortas were homogenized with TRizol reagent (Invitrogen), total RNA was isolated with miRNeasy Mini Kit (Qiagen), cDNA was synthesized with random p(dN), primers (Roche) and MMLV reverse transcriptase (Invitrogen). RT-PCR analysis in triplicates was performed with SYBR Green FastMix (Quanta) reaction mix and analyzed on QuantStudio 6 and 7 Flex Real-Time PCR System Software. Supplementary Table 3 provides the list of primers.

**RNA extraction and RT-PCR**

Cell pellets, tumor-bearing mammary glands, and microdissected lung, liver, and brain microcortas were homogenized in 1% SDS solution containing protease (cat# 118361S3001, Roche AG) and phosphatase inhibitor cocktail (cat# 04906845001, PhosStop, Roche, AG), followed by 20 min centrifugation at 4 °C (14,000 x g). Protein-containing supernatant was quantified and normalized to reference protein using the Image Studio Lite version 5.2 (LI-COR®). SDS-polyacrylamide and transfer as previously detailed, membranes were visualized with ECL or WesternSure PREMIUM Chemiluminescent Substrate (cat# 926-95010, LI-COR®). Antibodies are listed in Supplementary Table 2.

**Single cell RNAseq and RNA velocity**

Intratracheal xenografts of MCF-7-RFP-Luc were grown for 6 months. Mammary glands and lungs from 2 NSG-EGFP mice were collected and were mechanically and enzymatically digested using parallel razor blades and 1.5% Collagenase A/1% hyaluronidase, respectively. Samples were centrifuged at 380 x g at 25 °C for 10 min, and resuspended in red blood cell lysis buffer (Sigma, R7757) for 3-5 min, then diluted in PBS 2% CS, and centrifuged. Pellet was resuspended in trypsin for 3 min at 37 °C, centrifuged and resuspended in 0.1 mg/ml DNase (1284932, Roche AG). After centrifugation, pellet was resuspended in 10 ml PBS 2% CS, filtered through a 40 μm pore size filter (cat#352360, BD Falcon) for FACS. GFP RFP DRAQ7 Hoechst 33342 cells were sorted.

For analysis of mesenchymal-like cells in primary tumors, the blue channel of interest (DAPI) was binarized using the “Percentile” thresholding algorithm for MCF-7 and T47D and “Moments” for METSIS tumors based on the visual inspection of output images. Images were processed with watershed segmentation to split adjacent cells, followed by cellular aspect ratio (CAR) quantification using the “analyse particles” function of Fiji (Size 50-infinity pixel², circularity 0-1). Mask for the blue channel was saved and the cell quantifications were output as a table for further data analysis. For analysis of mesenchymal-like cells in lungs, major and minor axis were drawn manually on each tumor cell for CAR quantification, >300 DTCs in lungs from each mouse, n = 3. Mesenchymal-like cells were defined as those with major axis:minor axis ratio (CAR) > 1.6. For analysis of fluorescence intensity, corrected total cell fluorescence (CTCF) was calculated by manually drawing an outline around the border of the cells using the drawing/selection tools in ImageJ, and a close-by region with no cells to serve as a measure of background noise. Measurements were generated by ImageJ and CTCF was calculated as: CTCF = integrated density (area x mean fluorescence background). For analysis of fibrillar collagen deposition, area of collagen was automatically quantified using ImageJ using size filter to exclude cells within the ducts.

Fluorescence images of whole mammary glands and lungs were acquired with a LEICA M205FA fluorescence stereo microscope equipped with a Leica DFC 340FX camera.

**RNA extraction and RT-PCR**

Cell pellets, tumor-bearing mammary glands, and microdissected lung, liver, and brain microcortas were homogenized in 1% SDS solution containing protease (cat #118361S3001, Roche AG) and phosphatase inhibitor cocktail (cat# 04906845001, PhosStop, Roche, AG), followed by 20 min centrifugation at 4 °C (14,000 x g). Protein-containing supernatant was quantified and normalized to reference protein using the Image Studio Lite version 5.2 (LI-COR®). SDS-polyacrylamide and transfer as previously detailed, membranes were visualized with ECL or WesternSure PREMIUM Chemiluminescent Substrate (cat #926-95010, LI-COR®). Antibodies are listed in Supplementary Table 2.

**Single cell RNAseq and RNA velocity**

Intratracheal xenografts of MCF-7-RFP-Luc were grown for 6 months. Mammary glands and lungs from 2 NSG-EGFP mice were collected and were mechanically and enzymatically digested using parallel razor blades and 1.5% Collagenase A/1% hyaluronidase, respectively. Samples were centrifuged at 380 x g at 25 °C for 10 min, and resuspended in red blood cell lysis buffer (Sigma, R7757) for 3-5 min, then diluted in PBS 2% CS, and centrifuged. Pellet was resuspended in trypsin for 3 min at 37 °C, centrifuged and resuspended in 0.1 mg/ml DNase (1284932, Roche AG). After centrifugation, pellet was resuspended in 10 ml PBS 2% CS, filtered through a 40 μm pore size filter (cat#352360, BD Falcon) for FACS. GFP RFP DRAQ7 Hoechst 33342 cells were sorted.
using BD FACSAria-II SORP and analyzed on the BD FACSDiva software version 9.0 (Supplementary Fig. 12a, b).

Sorted cells were washed and resuspended to 1000 cells/μL in PBS. Single-cell RNA sequencing was performed using 10x Genomics Chromium V3.1 kit, and Illumina HiSeq 4000. After demultiplexing sequencing libraries to individual FASTQ files, Illumina output sequencing reads from 10x Genomics Chromium were processed and aligned to the GRCh38 genome using “count” function of Cell Ranger v5.0.1. Quality controls were performed >500 genes detected per cell and <20% mitochondrial reads. Seurat software v4.0 default settings were used to ensure conservative clusters. The number of principal components chosen to capture the majority of the variation in the data was based on elbow plots. GSEA were performed using the “enrichric” function of “escape” version 1.6.0 in R. The hallmark gene collection MSigDB v6.2 was used.

The scvelo algorithm version 0.2.4 was used to compute the steady-state model of RNA velocity using the tool sc.tl.velocity35. To fit the slope of the steady-state model 92nd instead of 95th percentile was used to better capture the steady-state. RNA velocity vectors were projected onto UMAP plot using default settings, single genes were interrogated using phase plots available in the package. For primary tumour cells, the scvelo package was used to calculate velocities on the deterministic setting with default parameters and to generate UMAP plots and the corresponding low-dimensional velocity embeddings.

ChIP-seq track visualization
The genome browser screenshots shown in Fig. 7i and Supplementary Fig. 11a, b were generated by accessing the UCSC genome browser via the UC epigenomics track (BC) Hub, (https://cbhub.epfl.ch) (bioRxiv 2022.05.01.490187).

Statistics and illustrations
Statistical analysis was performed using GraphPad Prism (version 8) (San Diego, California, USA, www.graphpad.com). Statistical tests are indicated in the figure legends. For growth curves, two-way ANOVA with multiple comparisons was used, for contralateral intraductal grafts, paired Student’s t-test was performed, and for individual grafts, unpaired Student’s t-test were performed, while testing for normality (Shapiro-Wilk test). Non-parametric statistical tests were used when Kolmogorov-Smirnov and Shapiro-Wilk normality tests failed to show a normal distribution. All statistical tests are two-tailed. * = ** = *** = **** and n.s. represent P < 0.05, 0.01, 0.001, 0.0001, and not significant, respectively. Graphic rendition of stereotype in Fig. 6f was created with BioRender.com.

Reporting summary
Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The raw and processed scRNA-seq generated in Fig. 7 and Supplementary Figure 9 have been deposited in the Gene Expression Omnibus (GEO) database under accession code GSE196936. All data for the main and supplementary figures are available within the article, in the source data file section. Source data are provided with this paper.

Code availability
The code employed for the analyses is open-source and available through the above mentioned packages in R.

References
1. Sung, H. et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA. Cancer J. Clin. https://doi.org/10.3322/caac.21660 (2021).
2. Pan, H. et al. 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. N. Engl. J. Med. 377, 1836–1846 (2017).
3. Klein, C. A. Framework models of tumor dormancy from patient-derived observations. Curr. Opin. Genet. Dev. 21, 42–49 (2011).
4. Buell, J. F. et al. Donor transmitted malignancies. Ann. Transplant. 9, 53–56 (2004).
5. Matser, Y. A. H. et al. Transmission of breast cancer by a single multiorgan donor to 4 transplant recipients. Am. J. Transplant. 18, 1810–1814 (2018).
6. Klein, C. A. Cancer progression and the invisible phase of metastatic colonization. Nat. Rev. Cancer 20, 681–694 (2020).
7. Richman, J. & Dowsett, M. Beyond 5 years: enduring risk of recurrence in oestrogen receptor-positive breast cancer. Nat. Rev. Clin. Oncol. 16, 296–311 (2019).
8. Risson, E., Nobre, A. R., Maguer-Satta, V. & Aguirre-Ghiso, J. A. The current paradigm and challenges ahead for the dormancy of disseminated tumour cells. Nat. Cancer 1, 672–680 (2020).
9. Zhang, X. H.-F., Giuliani, M., Trivedi, M. V., Schiff, R. & Osborne, C. K. Metastasis Dormancy in Estrogen Receptor–Positive Breast Cancer. Clin. Cancer Res. 19, 6389–6397 (2013).
10. Green, J. E. et al. The C3(3)/SV40 T-antigen transgenic mouse model of mammary cancer: ductal epithelial cell targeting with multistage progression to carcinoma. Oncogene 19, 1020–1027 (2000).
11. Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. Mol. Cell. Biol. 12, 954–961 (1992).
12. Sinn, E. et al. Coexpression of MMTV/v-Ha-ras and MMTV/c-myc transgenic mouse model for metastatic disease. Mol. Cell. Biol. 12, 422 (2016).
13. Scabia, V. et al. Estrogen receptor positive breast cancers have distinct spectrum and metastatic potential of human lesions. Cancer Cell 13, 287–292 (2019).
14. Sflomos, G. et al. A Preclinical Model for ERα-Positive Breast Cancer Points to the Epithelial Microenvironment as Determinant of Luminal Phenotype and Hormone Response. Cancer Cell 25, 407–422 (2016).
15. Kozma, K. J., Done, S. J. & Egan, S. E. The tumor cell-derived matrix of lobular breast cancer: a new vulnerability. EMBO Mol. Med. 13, e13807 (2021).
16. Sflomos, G. et al. Intraductal xenografts show lobular carcinoma cells rely on their own extracellular matrix and LOXL1. EMBO Mol. Med. 13, e13180 (2021).
17. Scabia, V. et al. Estrogen receptor positive breast cancers have patient specific hormone sensitivities and rely on progesterone receptor. Nat. Commun. 13, 3127 (2022).
18. Goddard, E. T. et al. The Rodent Liver Undergoes Weaning-Induced Involution and Supports Breast Cancer Metastasis. Cancer Discov. 7, 177–187 (2017).
19. Husemann, Y. et al. Systemic Spread Is an Early Step in Breast Cancer. Cancer Cell 13, 58–68 (2008).
20. Sänger, N. et al. Disseminated tumor cells in the bone marrow of patients with ductal carcinoma in situ. Int. J. Cancer 129, 2522–2526 (2011).
21. Bajar, T. et al. Fluorescent indicators for simultaneous reporting of all four cell cycle phases. Nat. Methods 13, 993–996 (2016).
22. Bragado, P. et al. TOFB2 dictates disseminated tumour cell fate in target organs through TGFβ-RIII and p38β signalling. Nat. Cell Biol. 15, 1351–1361 (2013).
23. O’Connor, J. W. & Gomez, E. W. Cell Adhesion and Shape Regulate TGF-Beta1-Induced Epithelial-Myofibroblast Transition via MRTF-A Signaling. PLOS ONE 8, e83188 (2013).
24. Aiello, N. M. et al. EMT Subtype Influences Epithelial Plasticity and Mode of Cell Migration. Dev. Cell 45, 681–695.e4 (2018).

25. Pastushenko, I. et al. Identification of the tumour transition states occurring during EMT. Nature 556, 463–468 (2018).

26. Rios, A. C. et al. Intracranial Plasticity in Mammary Tumors Revealed through Large-Scale Single-Cell Resolution 3D Imaging. Cancer Cell 35, 618–632.e9 (2019).

27. Simeonov, K. P. et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell 39, 1150–1162.e9 (2021).

28. Yang, J. et al. Guidelines and definitions for research on epithelial–mesenchymal transition. Nat. Rev. Mol. Cell Biol. 21, 341–352 (2020).

29. Kalluri, R. & Neilson, E. G. Epithelial-mesenchymal transition and its implications for fibrosis. J. Clin. Invest. 112, 1776–1784 (2003).

30. Jeselsohn, R. et al. Emergence of constitutively active estrogen receptor-alpha mutations in pretreated advanced estrogen receptor-positive breast cancer. Clin. Cancer Res 20, 1757–1767 (2014).

31. Harrod, A. et al. Genomic modelling of the ESR1 Y537S mutation for evaluating function and new therapeutic approaches for metastatic breast cancer. Oncogene 36, 2286–2296 (2017).

32. Di Martino, J. S. et al. A tumor-derived type III collagen-rich ECM niche regulates tumor cell dormancy. Nat. Cancer 3, 90–107 (2022).

33. Montagner, M. et al. Crosstalk with lung epithelial cells regulates Sfrp2-mediated latency in breast cancer dissemination. Nat. Cell Biol. 22, 289–296 (2020).

34. Bergen, V., Lange, M., Feidt, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. Nat. Biotechnol. 38, 1408–1414 (2020).

35. La Manno, G. et al. RNA velocity of single cells. Nature 560, 494–498 (2018).

36. Siersbæk, R. et al. IL6/STAT3 Signaling Hijacks Estrogen Receptor α Enhancers to Drive Breast Cancer Metastasis. Cancer Cell 38, 412–423.e9 (2020).

37. Criscitiello, C. et al. High Ki-67 score is indicative of a greater benefit from adjuvant chemotherapy when added to endocrine therapy in Luminal B HER2 negative and node-positive breast cancer. Breast 23, 69–75 (2014).

38. Brabletz, S., Schuhwerk, H., Brabletz, T. & Stemmler, M. P. Dynamic EMT: a multi-tool for tumor progression. EMBO J. 40, e108647 (2021).

39. Dong, H.-M. et al. Dominant-negative E-cadherin inhibits the invasiveness of inflammatory breast cancer cells in vitro. J. Cancer Res. Clin. Oncol. 133, 83–92 (2007).

40. Hugo, H. J. et al. Epithelial requirement for in vitro proliferation and xenograft growth and metastasis of MDA-MB-468 human breast cancer cells: oncogenic rather than tumor-suppressive role of E-cadherin. Breast Cancer Res. 19, 86 (2017).

41. Kleer, C. G., van Golen, K. L., Braun, T. & Merajver, S. D. Persistent E-Cadherin Expression in Inflammatory Breast Cancer. Mod. Pathol. 14, 458–464 (2001).

42. Kowalski, P. J., Rubin, M. A. & Kleer, C. G. E-cadherin expression in primary carcinomas of the breast and its distant metastases. Breast Cancer Res. 5, R217 (2003).

43. Padmanaban, V. et al. E-cadherin is required for metastasis in multiple models of breast cancer. Nature 573, 439–444 (2019).

44. Puram, S. V., Parikh, A. S. & Tirosch, I. Single cell RNA-seq highlights a role for a partial EMT in head and neck cancer. Mol. Cell. Oncol. 5, e1448244 (2018).

45. Tyler, M. & Tirosch, L. Decoupling epithelial-mesenchymal transitions from stromal profiles by integrative expression analysis. Nat. Commun. 12, 2592 (2021).

46. Battle, E. et al. The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nat. Cell Biol. 2, 84–89 (2000).

47. Comijn, J. et al. The Two-Handed E Box Binding Zinc Finger Protein SIP1 Downregulates E-Cadherin and Induces Invasion. Mol. Cell 7, 1267–1278 (2001).

48. Cagnet, S. et al. Oestrogen receptor α AF-1 and AF-2 domains have cell population-specific functions in the mammary epithelium. Nat. Commun. 9, 4723 (2018).

49. Nobre, A. R. et al. NC2G+nestin+ mesenchymal stem cells dictate DTC dormancy in the bone marrow through TGFβ2. bioRxiv 2020.10.22.349514 https://doi.org/10.1101/2020.10.22.349514 (2020).

50. Ombrato, L. et al. Metastatic niche labelling reveals tissue par- enchyma stem cell features. Nature 572, 603–608 (2019).

51. Werner-Klein, M. et al. Interleukin-6 trans-signaling is a candidate mechanism to drive progression of human DCCs during clinical latency. Nat. Commun. 11, 4877 (2020).

52. Herndler-Brandstetter, D. et al. Humanized mouse model supports development, function, and tissue residency of human natural killer cells. Proc. Natl Acad. Sci. U. S. A. 114, E9626–E9634 (2017).

53. Lovisa, S. et al. Epithelial-to-mesenchymal transition induces cell cycle arrest and parenchymal damage in renal fibrosis. Nat. Med. 21, 998–1009 (2015).

54. Mejlvang, J. et al. Direct repression of cyclin D1 by SIP1 attenuates cell cycle progression in cells undergoing an epithelial-mesenchymal transition. Mol. Biol. Cell 18, 4615–4624 (2007).

55. Vega, S. et al. Snail blocks the cell cycle and confers resistance to cell death. Genes Dev. 18, 1131–1143 (2004).

56. Yang, Y., Pan, X., Lei, W., Wang, J. & Song, J. Transforming growth factor-beta1 induces epithelial-to-mesenchymal transition and apoptosis via a cell cycle-dependent mechanism. Oncogene 25, 7235–7244 (2006).

57. Behbod, F. et al. An intraducal human-in-mouse transplantation model mimics the subtypes of ductal carcinoma in situ. Breast Cancer Res. 11, R66 (2009).

58. Onder, T. T. et al. Loss of E-Cadherin Promotes Metastasis via Multiple Downstream Transcriptional Pathways. Cancer Res. 68, 3645–3654 (2008).

59. Barde, I., Salmon, P. & Trono, D. Production and Titration of Recombinant Human E460 for in Vitro and in Vivo Applications. Current Protocols in Protein Science, 54.21.1-54.21.21 (2010).

60. Ataca, D. et al. The secreted protease A1 domain of F13 links hormone action to activation of the mammary stem cell niche. Nat. Commun. 11, 1571 (2020).

Acknowledgements
We thank S. Egan for critical comments and P. den Hollanader for careful reading of the manuscript, J. Aguirre-Ghiso, R. Siersbæk, I. Tirosch, G. LaManno and N. Aizaran for helpful discussions, L. Battista and Y. Liu for technical assistance, J. Dessimoz, EPFL histology core facility, T. Laroche, EPFL bioimaging and optics platform (BIOP), M. Garcia at the EPFL and D. Labes at AGORA flow cytometry facilities and B. Mangeat at the EPFL gene expression core facility (GEFC) for technical assistance. We thank students G. Pittet, J. Fan, M. Lasfargues, M. Levorato, J. Brune, A. Huguenin-Dumittan, and A.-M. Curat and the patients who participated in our study. P.A., H. O. and F.D.M were supported by SNF 310030_179163/1 Exploring key steps of the metastatic cascade in ER+ BC in vivo, Y.Z. by KFS-4758-02-2019-R Different facets of estrogen receptor alpha (ER) signalling during ER+ breast carcinogenesis, and G.S. by Biltema Stiftelsen Stefan Paulssons Cancerfond. H. Q. is supported by H2020-MSCA-ITN (ITN-2019-859860-CANCERPREV). This manuscript was edited at Life Science Editors.
Author contributions
Conceptualization P.A., C.B.; Formal Analysis P.A., Y.Z., F.D.M., G.A., K.M., C.S.; Investigation P.A., Y.Z., C.S., H.Q., G.S.; Resources S.M., S.A.; Writing P.A., C.B; Funding Acquisition C.B.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-022-32523-6.

Correspondence and requests for materials should be addressed to Cathrin Brisken.

Peer review information Nature Communications thanks Qing Nie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at http://www.nature.com/reprints