Crosstalk with lung epithelial cells regulates Sfrp2-mediated latency in breast cancer dissemination

Marco Montagner, Rahul Bhome, Steven Hooper, Probir Chakravarty, Xiao Qin, Jahangir Sufi, Ajay Bhargava, Colin D. H. Ratcliffe, Yutaka Naito, Arianna Pocaterra, Christopher J. Tape and Erik Sahai

The process of metastasis is complex. In breast cancer, there are frequently long time intervals between cells leaving the primary tumour and growth of overt metastases. Reasons for disease indolence and subsequent transition back to aggressive growth include interactions with myeloid and fibroblastic cells in the tumour microenvironment and ongoing immune surveillance. However, the signals that cause actively growing cells to enter an indolent state, thereby enabling them to survive for extended periods of time, are not well understood. Here we reveal how the behaviour of indolent breast cancer cells in the lung is determined by their interactions with alveolar epithelial cells, in particular alveolar type 1 cells. This promotes the formation of fibronectin fibrils by indolent cells that drive integrin-dependent pro-survival signals. Combined in vivo RNA sequencing and drop-out screening identified secreted frizzled-related protein 2 (SFRP2) as a key mediator of this interaction. Sfrp2 is induced in breast cancer cells by signals from lung epithelial cells and promotes fibronectin fibril formation and survival, whereas blockade of Sfrp2 expression reduces the burden of indolent disease.

To analyse indolent breast cancer, we used the D2.0R (dormant) and D2A1 (proliferative) cell model of metastasis (Extended Data Fig. 1a). D2.0R cells persisted for many weeks in the lungs (Fig. 1a and Extended Data Fig. 1b), but did not form large colonies, whereas D2A1 cells grew aggressively (Extended Data Fig. 1b). The indolent behaviour of the D2.0R cells parallels that of oestrogen receptor 1 (ESR1)-positive breast cancer; consistent with this, D2.0R cells express ESR1 in vivo and respond to oestrogen. D2.0R cells were indolent in both Balb/C and Balb/C nude mice, indicating that their phenotype is not due to the adaptive immune system (Extended Data Fig. 1c). Closer examination revealed that D2.0R cells had extravasated into the alveolar space and contacted the lung parenchyma after 2 d (Fig. 1a). In this context, both D2.0R and D2A1 cells formed long extensions reminiscent of filopodia, and D2.0R cells had close contact with aquaporin 5 (AQP5)-positive and podoplanin (PDPN)-positive alveolar type 1 (AT1) cells (Fig. 1c, left). Frequent contacts were also observed with surfactant protein C (SFTPC)-positive alveolar type 2 (AT2) cells and mucin 1 (MUC1)-positive endothelial cells (Extended Data Fig. 1e,f). Pulse labelling with 5-ethyl-2′-deoxyuridine (EdU) revealed that AT1 cells, which are normally quiescent, were proliferating proximal to D2.0R cells 3 d and 14 d after arrival in the lungs. This suggests that the expansion of the lung parenchyma around indolent metastases at two weeks results mostly from proliferation of AT1 cells (Fig. 1ac and Extended Data Fig. 1g,h). Similar contacts with PDPN-positive AT-1 cells and an increase in EdU positivity were observed with human MCF7 cells (Fig. 1d and Extended Data Fig. 1i).

To study how breast cancer cells might interact in the lung environment, we established a co-culture system that recapitulated the indolent behaviour of D2.0R cells. We co-cultured lung cells that express the markers of AT1 and AT2 cells and fibroblasts on a gas-permeable substrate in mitogen-low glucose-low (MLNL) medium (Fig. 1e). The addition of low numbers of either D2.0R or D2A1 cells to these co-cultures recapitulated the indolent and aggressive growth of D2.0R and D2A1 cells observed in vivo (Fig. 1f,g; increased Ki67+ cells are shown in Extended Data Fig. 2b). These differences could not be attributed to intrinsic differences in growth between D2.0R and D2A1 cells in mitogen-high glucose-high (MHNH) medium or MLNL medium (Extended Data Fig. 2c). The indolent behaviour of D2.0R cells in the co-culture was reversible if cells were subsequently returned to conventional cell culture conditions (Extended Data Fig. 2d). We next explored the effect of individual cell types in the co-culture assay, which is not possible to do in mice. Co-cultures with individual cell types in MLNL medium indicated that AT1-like cells were able to boost numbers of D2.0R cells, with AT2-like cells having a smaller positive effect (Fig. 1h). Similar results were obtained with 4T1 cells and MCF7 cells (Extended Data Fig. 3a). Time-lapse imaging revealed that AT1-like cells both suppress apoptosis and increase the mitotic rate of D2.0R cells (Extended Data Fig. 3b). The omission of individual epithelial cell types from the ‘full’ co-culture revealed a more nuanced picture of the interaction between breast cancer cells and AT1-like cells (Extended Data Fig. 3c). Increased growth was observed in the absence of AT1-like cells, suggesting that as well as generating pro-survival signals in the more restrictive MLNL conditions (Fig. 1h), AT1-like cells can also generate growth suppressive cues that counteract proliferative cues, most probably emanating from the AT2-like cells. To test directly whether AT1-like cells could suppress growth in the face of strong proliferative cues,
we cultured D2.0R cells in MHNH medium for 7 d. Extended Data Fig. 3d shows that AT1-like cells were able to reduce the growth of D2.0R cells in favour-able conditions. These results suggest that prosurvival and growth-restrictive signals from AT1 cells probably coexist in vivo and in vitro. We reasoned that the greatest clinical benefit would result from being able to target the supportive signals, therefore we concentrated on the interaction between lung epithelial cells and breast cancer cells in MLNL medium.

In culture conditions in which AT1-like cells support D2.0R cells (MLNL medium), long protrusions were observed similar to those observed in vivo (Fig. 1lj). This change reduced the circular-ity of D2.0R cells from about 0.8 to 0.2–0.4 (Extended Data Fig. 3e); furthermore, the protrusions were positive for active phosphorylated (p)Src, which is implicated in pro-metastatic signals15,16 (Fig. 1k). The protrusions formed by D2.0R cells in the presence of AT1-like cells were associated with fibronectin fibris (Fig. 1l).

Similar increases in cell protrusion were obtained with 4T1 mouse mammary tumour, MCF7 human breast cancer and T47D-DBM dormant bone metastatic cell lines (Extended Data Fig. 3f–h), but not for the aggressive D2A1 mouse mammary cell line, which had a higher baseline of protrusions (Extended Data Fig. 3i). Blockade of integrins using cilegintide, which mimics the Arg-Gly-Asp (RGD) integrin-binding motif of fibronectin and other extracellular matrix (ECM) molecules, reduced both numbers and protrusions of D2.0R and MCF7 cells (Fig. 1ml and Extended Data Fig. 3j). Notably, treatment of mice with cilegintide, even after cells had already seeded the lungs, reduced the number of metastases (Fig. 1n).

These results suggest that persistence of indolent breast cancer cells and the induction of cellular protrusions by AT1 cells represent intertwined aspects of metastatic dissemination. Furthermore, they demonstrate that targeting this axis is a viable strategy for the elimination of indolent breast cancer cells.

To understand the signalling pathways involved in breast cancer–alveolar cell crosstalk, we undertook mass cytometry analysis of co-cultures17 (Extended Data Fig. 4a). Consistent with data in Extended Data Fig. 3b, there was an increase in phosphorylation events associated with proliferation (S807/811-pRb and T37/46-p4E-BP1) in D2.0R and MCF7 cells in co-culture (Fig. 2a). Changes in signalling were also observed, with prominent increases in activities of ERK, MKK4 (also known as SEK1), MKK3/6, PDPK1, β-catenin and NF-kB (Fig. 2a). Conversely, both D2.0R and MCF7 cells triggered proliferative responses in AT1-like cells (indicated by S-phase markers S807/811-pRb and T37/46-p4E-BP1, and mitotic marker pS28-H3), supporting the in vivo observation of EdU+ nuclei around micro-metastases (Extended Data Fig. 2b,c). Density resampled estimation of mutual information (DREMI) analysis revealed increased connectivity from PDPK1 to PKCα and AKT, and from AKT to ERK (Extended Data Fig. 4d, DREMI score in white). Several of these pathways have been linked to cancer dormancy20–24; we therefore investigated how inhibition of these and other prominent signalling pathways affected D2.0R behaviour in the presence of AT1-like cells.

Blockade of EGFR, MEK, JNK and Src-family kinase (SFK) signalling, but not of p38 MAPK or β-catenin signalling, reduced the number of D2.0R cells when co-cultured with AT1-like cells (Fig. 2c). EGFR, MEK and SFK inhibition both increased apoptosis and reduced mitotic events without greatly affecting D2.0R and AT1-like cells in monoculture (Fig. 2b and Extended Data Fig 4e). Combining inhibitors with pERK analysis indicated that EGFR inhibition, SFK inhibition and MEK inhibition all reduced pERK levels, supporting a role for EGFR and SFK signalling upstream of ERK/MAP kinase (Extended Data Fig. 4f). The importance of ERK/MAP kinase activation was confirmed by a reduction in metastatic colony size in vivo (Extended Data Fig. 4g). The formation of cell protrusions and fibronectin fibris described in Fig. 1 depends on EGFR and SFK signalling, but not on MEK signalling, indicating a bifurcation in the signalling cascade at a point downstream of SFK (Fig. 2c and Extended Data Fig. 5a–e). The reduced fibronectin staining was not correlated with reduced fibronectin gene transcription (Extended Data Fig. 5d).

These results reinforce the correlation between cell protrusions and signals that boost survival of indolent breast cancer cells.

In a parallel effort to understand indolent breast cancer cells, we investigated how the metastatic microenvironment alters cancer cell gene expression in vivo. We compared gene expression in indolent D2.0R cells isolated from the lungs with aggressively growing D2A1 cells and both cell types grown in culture (Fig. 3a). As expected, the expression of cell cycle and DNA-replication genes was reduced in indolent D2.0R cells and we observed upregulation of bone morphogenetic protein (BMP) signalling25–27 and the dormancy-associated factors Nr2f1 and Bhlhe41 (also known as Ddoc or Sharp1)28–31 (Fig. 3b and Extended Data Fig. 6a,b). We also noted an increase in expression of ECM genes, including several linked to aggressive metastatic behaviour (Postn and Tnc) and epithelial-to-mesenchymal transition (EMT) factors25–27 (validated by quantitative PCR with reverse transcription (RT–qPCR); Fig. 3c). We next investigated links between the transcriptome of indolent D2.0R cells and human breast cancer. A signature of genes highly expressed in indolent D2.0R cells in vivo compared with the other groups was clearly linked with improved distant metastasis free survival (DMFS) in human ER+ breast cancer, which shows long latency periods before relapse32 (Fig. 3d; genes listed in Supplementary Table 1 and
compared with other dormancy signatures29,30 in Extended Data Fig. 7a,b). The signature remained associated with favourable outcome in ER+ lymph node-negative disease ($P=0.00038$), indicating that it is not simply a metric of lymph-node status. Patients with high expression of the signature who received tamoxifen therapy had increased DMFS at 10 yr (confirmed using the GSE9515 dataset; Extended Data Fig. 7c). Conversely, patients with low expression of the indolence signature had a significantly increased hazard ratio of 2.5 (Fig. 3c). Furthermore, re-plotting the analysis from 2 yr onwards confirmed the ability of the signature to indicate lower likelihood of distant relapses at prolonged time points (Extended Data Fig. 7c). Genes specifically upregulated in D2A1
cells or in culture showed no link with outcomes (Extended Data Fig. 7d and Supplementary Table 1). This clear link to human outcomes further reinforces the relevance of our experimental analysis.

We next attempted to identify the factor in the lung environment that might be responsible for triggering the transcriptomic changes in indolent D2.0R cells. We hypothesized that AT1-like cells might trigger the upregulation of genes in D2.0R cells in the lung. Figure 3f and Extended Data Fig. 7e show that AT1-like cells could indeed induce the expression of genes that are highly expressed in indolent cells in vivo, including a wide range of ECM genes and EMT factors. Thus, interaction with the lung parenchyma can trigger the expression of indolence-associated genes in vitro and in vivo.

Among the genes upregulated in indolent cells, some may have a role in maintaining the cells in a non-aggressive state—hence the overall correlation with good outcomes—and others might be involved in supporting their continued survival in the lung microenvironment. Genes in this second category might be implicated in the survival signals emanating from AT1 cells and be desirable to target for therapeutic benefit. To identify these genes we performed an in vivo screen using a short hairpin RNA (shRNA) library targeting 59 genes highly expressed in indolent D2.0R cells (Supplementary Table 2). Sub-pools
of an shRNA library were transduced into D2.0R cells and injected into the tail vein of mice; reference genomic DNA was also prepared from the sub-pools before injection. After three weeks, D2.0R cells were isolated from the lungs, and the relative representation of each shRNA was compared with reference genomic DNA representing the initial composition of the library by sequencing (Fig. 4a). We
Sfrp2 is expressed at low levels in cell culture and primary tumours, but its levels markedly increase when in the lung environment (Extended Data Fig. 8e). Co-culture experiments demonstrated that AT1-like lung epithelial cells could induce Sfrp2 in D2.0R cells in a Src-dependent manner (Fig. 5a,b), thus providing a potential explanation for the effect of SFK inhibitor observed in Fig. 2. AT1-like cells also partially induced other SFRP family members in D2.0R cells and 4T07 cells (Extended Data Fig. 8f). SFRP2 has been widely reported as WNT-signalling regulator; however, we did not observe any consistent modulation of canonical WNT targets in cells depleted of Sfrp2 (Extended Data Fig. 9a). It has been reported that SFRP2 binds fibronectin and is incorporated into an insoluble ECM fraction19. Furthermore, heparin binds the C-terminus of SFRP family proteins, releasing them from the ECM, thereby leading to their inactivation19. We confirmed that heparin could increase the level of soluble inactive SFRP2 in the medium (Extended Data Fig. 9b). Notably, this treatment also reduced D2.0R cell numbers in vitro (Extended Data Fig. 9c). These results suggest that insoluble extracellular SFRP2 promotes cell numbers by

Fig. 5 | SFRP2 regulates Src-mediated fibrillogenesis, protrusions and survival. a, RT-qPCR for Sfrp2 in D2.0R cells expressing eGFP cells cultivated alone or with AT1-like cells in MLNL medium for 4 d. Mean normalized pooled samples (n = 24 for control, n = 27 for AT1-like group) from 7 independent experiments. Mann–Whitney test. b, As in a, but in addition, cells were treated with SFK inhibitor (50 nM) or DMSO. Mean normalized pooled samples (n = 10 for control, n = 12 for other groups) from 3 independent experiments. One-way ANOVA. c, Mean fibronectin expression per cell in control and SFRP2-overexpressing (OE) breast cancer cells. Mann–Whitney test. Pooled data from n = 2 independent experiments. d, Protrusions in control or eGFP-expressing, SFRP2-overexpressing D2.0R cells alone or with AT1-like cells with or without SFK inhibitor. Mean values of independent experiments. Mann–Whitney test. e, Heat maps of EMD values in control or SFRP2-overexpressing D2.0R cells. n = 1 experiment. f, Immunofluorescence for F-actin and fibronectin in control or SFRP2-overexpressing D2.0R cells expressing eGFP, alone or with AT1-like cells in MLNL medium with SFKi for 48 h. n = 3. Scale bar, 20 μm. g, eGFP-expressing, SFRP2-overexpressing D2.0R cells alone or in co-culture with AT1-like cells treated with SFK inhibitor for 48 h. Quantification as in Extended Data Fig. 5b. n = 3 experiments. h, Quantification of cell death in control or SFRP2-overexpressing D2.0R cells with or without SFK inhibition (48 h). Quantification as in Fig. 2b. Mean and s.e.m. are shown. Mann–Whitney test. i, Left, fluorescent in situ immunohistochemistry images of control or SFRP2-overexpressing D2.0R cells expressing eGFP cells in the lung alveolar space after tail-vein injection. Scale bar, 20 μm. Right, circularity of lung-disseminated control or SFRP2-overexpressing eGFP-expressing D2.0R cells (across four mice per group). Mann–Whitney test. j,k, Quantification of metastatic burden and metastatic colony area two weeks after intravenous injection of control or SFRP2-overexpressing 4T07 or D2.0R cells into wild-type or nude mice, respectively. Mann–Whitney test for metastatic burden. Unpaired t-test with Welch’s correction for colony area experiments. In e,i (right), j,k (left): whisker plots: midline, median; box, 25–75th percentile; whisker, minimum to maximum. In j (right), k (right): plots show mean with 95% CI.
increasing the deposition and organization of fibronectin (Fig. 5c). SFRP2 overexpression promoted the organization of fibronectin into fibrils (Fig. 5f, bottom) and increased numbers of cell protrusions (Fig. 5d and Extended Data Fig. 9d). Furthermore, cytometry by time of flight (CyTOF) analysis revealed an overlap in the activation of intracellular pathways between D2.0R cells co-cultured with AT1-like-cells and D2.0R cells overexpressing SFRP2 (Fig. 2a,5e), including pPDPK1, pMKK4, pMKK3/6 and pERK. These data, combined
with the effect of SFK inhibition on Sfrp2 induction, prompted us to perform epistasis experiments. SFRP2 overexpression reduced the ability of SFK inhibition to block the formation of protrusions and fibronectin fibrils (Fig. 5d,f,g; compare with Extended Data Fig. 5a,b). Additionally, apoptosis in the presence of SFK inhibition was reduced when SFRP2 was overexpressed (Fig. 5h; Extended Data Fig. 9e shows no effect on proliferation). These analyses suggest that SFRP2 supports D2.0R persistence through pro-survival integrin–fibronectin signalling, leading to enhanced output across a range of oncogenic signalling pathways (Extended Data Fig. 10d).

Finally, we tested sFRP2 overexpression in vivo. SFRP2-overexpressing cells had more protrusions than control cells in vivo (assessed using the circularity metric; Fig. 5i). SFRP2 overexpression also increased the metastatic burden in both human and mouse models, with a particularly pronounced increase in large metastases (area > 5 × 10^4 μm^2) in the 4T07 model (Fig. 5jk and Extended Data Fig. 10a). Proliferation in vitro in the absence of AT1-like cells was not affected (Extended Data Fig. 10b), nor was proximity to other stromal cells (Extended Data Fig. 10c).

Delayed recurrence of latent disseminated cells represents an unmet clinical need. Our work suggests that parenchymal epithelial cells constitute a critical and underappreciated component of the microenvironment in metastases to epithelial organs. One possible reason for the lack of attention paid to epithelial cells in the tumour microenvironment is that they are outcompeted by the malignant cells in growing tumours. However, in indolent micrometastases or during the first steps of colonization, they are abundant relative to the cancer cells and can exert a large influence on their behaviour. In future, it will be interesting to study the signals from the lung epithelium that induce Sfrp2 and determine why some highly aggressive cancers might be able to activate survival mechanisms upon arrival in the lung, while not being subject to growth-suppressive or -limiting signals. To conclude, our results indicate that carcinoma cells are highly responsive to signals coming from non-transformed epithelial cells at metastatic locations. We propose that this will prove to be a recurring theme in the metastatic spread of epithelial cancers to distant epithelial tissues and, crucially, we demonstrate that interference in this crosstalk reduces survival of disseminated indolent breast cancer cells.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-020-0474-3.

Received: 28 September 2018; Accepted: 20 January 2020; Published online: 24 February 2020

References
1. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging biological principles of metastasis. Cell 168, 670–691 (2017).
2. Hedley, B. D. & Chambers, A. F. Tumor dormancy and metastasis. Adv. Cancer Res. 102, 67–101 (2009).
3. Sosa, M. S., Bragado, P. & Aguirre-Ghiso, J. A. Mechanisms of disseminated cancer cell dormancy: an awakening field. Nat. Rev. Cancer 14, 611–622 (2014).
4. Obenauf, A. C. & Massagué, J. Surviving at a distance: organ-specific metastasis. Trends Cancer 1, 76–91 (2015).
5. Malladi, S. et al. Metastatic latency and immune evasion through autocrine inhibition of WNT. Cell 165, 45–60 (2016).
6. Avugustinova, A. et al. Tumour cell-derived Wnt7a recruits and activates fibroblasts to promote tumour aggressiveness. Nat. Commun. 7, 10305 (2016).
7. Naumov, G. N. et al. Persistence of solitary mammary carcinoma cells in a secondary site: a possible contributor to dormancy. Cancer Res. 62, 2162–2168 (2002).
8. Barkan, D. et al. Inhibition of metastatic outgrowth from single dormant tumor cells by targeting the cytoskeleton. Cancer Res. 68, 6241–6250 (2008).
9. Sosa, M. S. et al. NR2F1 controls tumour cell dormancy via SOX9- and RAR-dependent signalling. Nat. Cell Biol. 15, 807–817 (2013).
10. Shibue, T., Brooks, M. W. & Weinberg, R. A. An integrin-linked machinery of cytoskeletal regulation that enables experimental tumor initiation and metastatic colonization. Cancer Cell 24, 481–493 (2013).
11. Shibue, T., Brooks, M. W., Inan, M. F., Reinhardt, F. & Weinberg, R. A. The outgrowth of micrometastases is enabled by the formation of filopodium-like protrusions. Cancer Discov. 2, 706–712 (2012).
12. Ghajar, C. M. et al. The perivascular niche regulates breast tumour dormancy. Nat. Cell Biol. 15, 807–817 (2013).
13. van den Bogaard, E. H., Dailey, L. A., Thorley, A. J., Tetley, T. D. & Forbes, B. Inflammatory response and barrier properties of a new alveolar type 1-like cell line (TT1). Pharm. Res. 26, 1172–1180 (2009).
14. Seguin, L., Desgroesselier, J. S., Weis, S. M. & Cheresh, D. A. Integrins and cancer: regulators of cancer stemness, metastasis, and drug resistance. Trends Cell Biol. 25, 234–240 (2015).
15. Qin, X. et al. Single-cell signalling analysis of heterocellular organoids. Preprint at bioRxiv https://doi.org/10.1101/599696 (2019).
16. Li, D., Suny, L. H. et al. Combined SFK/MEK inhibition prevents metastatic outgrowth of dormant tumor cells. J. Clin. Invest. 124, 156–168 (2014).
17. Bragado, P. et al. TGF-β2 dictates disseminated tumour cell fate in target organs through TGF-β-βRII and p38βII signalling. Nat. Cell Biol. 15, 1351–1361 (2013).
18. Gawrzak, S. et al. MSK1 regulates luminal cell differentiation and metastatic dormancy in Er knockout breast cancer. Nat. Cell Biol. 20, 211–221 (2018).
19. Gao, H. et al. The BMP inhibitor Coc reactivates breast cancer cells at lung metastatic sites. Cell 150, 764–779 (2012).
20. Kobayashi, A. et al. Bone morphogenetic protein 7 in dormancy and metastasis of prostate cancer stem-like cells in bone. J. Exp. Med. 208, 2641–2655 (2011).
21. Cackowski, F. C. et al. Mer tyrosine kinase regulates disseminated prostate cancer cellular dormancy. J. Cell Biochem. 118, 891–902 (2017).
22. Fluegen, G. et al. Phenotypic heterogeneity of disseminated tumour cells is present by primary tumour hypoxic microenvironments. Nat. Cell Biol. 19, 120–132 (2017).
23. Sosa, M. S. et al. NR2F1 controls tumour cell dormancy via SOX9- and RARβ-driven quiescence programmes. Nat. Commun. 6, 6170 (2015).
24. Malanchi, I. et al. Interactions between cancer stem cells and their niche govern metastatic colonization. Nat. 481, 85–89 (2011).
25. Okarsen, T. et al. Breast cancer cells produce tenasin C as a metastatic niche component to colonize the lungs. Nat. Med. 17, 867–874 (2011).
26. Zhang, X. H., Giuliano, M., Trivedi, M. V., Schif, R. & Osborne, C. K. Metastasis dormancy in estrogen receptor-positive breast cancer. Clin. Cancer Res. 19, 6389–6397 (2013).
27. Kim, R. S. et al. Dormancy signatures and metastasis in estrogen receptor α-positive and negative breast cancer. PLoS One 7, e35569 (2012).
28. Cheng, Q. et al. A signature of epithelial-mesenchymal plasticity and stromal activation in primary tumor modulates late recurrence in breast cancer independent of disease subtype. Breast Cancer Res. 16, 407 (2014).
29. Lee, J. L., Lin, C. T., Chueh, L. I. & Chang, C. J. Autocrine/paracrine secreted Frizzled-related protein 2 induces cellular resistance to apoptosis: a possible mechanism of mammary tumorigenesis. J. Biol. Chem. 279, 14602–14609 (2004).
30. Bovolenta, P., Esteve, P., Ruiz, J. M., Cisneros, E. & Lopez-Rios, J. Beyond Wnt inhibition: new functions of secreted Frizzled-related proteins in development and disease. J. Cell. Sci. 121, 737–746 (2008).
31. Kaur, A. et al. sFRP2 in the aged microenvironment drives melanoma metastasis and therapy resistance. Nature 532, 250–254 (2016).
32. Ombrato, L. et al. Metastatic-niche labelling reveals parenchymal cells with stem features. Nature 572, 603–608 (2019).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. © The Author(s), under exclusive licence to Springer Nature Limited 2020.
Methods

Cell lines. Alveolar-type 1-like cells (TT1 cells) were a gift from J. Dowward (The Francis Crick Institute, London) and were originally provided by T. Tetley (Imperial College, London) as described in ref. 1. T47D-DBM cells were a gift from R. Gomis (Institute for Research in Biomedicine, Barcelona). Alveolar-type 2 cells (H441 cells) were purchased from ATCC (HTB-174). Human normal lung fibroblasts were derived from primary lung fibroblasts (Cancer Research UK Cell Service AG02603) immortalized with pBABE-hygro-iptERT. D2.0R, D2A1 and MCF7-GFP cells were a gift from D. Barkan (University of Haifa). The AT2 cells were provided by S. Piccolo (University of Padua). All cells were confirmed under standard culture conditions in DMEM with 10% FBS (Thermo Fisher Scientific, 41965-039) and routinely screened for mycoplasma at the Cell Services facility at The Francis Crick Institute or with Universal Mycoplasma Detection kit (ATCC, 30–1012 K).

Lung organotypic system and quantification. Lung cells and breast cancer cells were plated onto Lumox 24-multiwell plate (Sarstedt, 94.699.00.14) in MLNL medium (low-glucose DMEM/1% FCS, Thermo Fisher Scientific 21885025) or MHNH medium (high-glucose DMEM/10% FCS, Thermo Fisher Scientific, 41965-039) as indicated. In detail: AT1-like cells (12.5 × 10^4 cells per well) and AT2-like cells (2.5 × 10^4 cells per well) were plated at day 1, human normal lung fibroblasts were plated at day 2 (2.5 ×10^5 per well) and cancer cells were plated at day 3 (100 cells per well). Medium was replaced every 3 d and GFP+ cells were manually counted under an inverted fluorescence microscope after replacing the medium with HBSS. For experiments where relative number of cells per ml is shown, cells were trypsinized, counted through a 70 μm cell strainer and resuspended in 200 μl of FACS buffer (PBS, 2 mM EDTA and 3% BSA). The number of GFP+ cells per ml was then measured with a MACSQuant Analyzer (Miltenyi Biotec) with a 96-well plate module.

Drug treatments. Drugs and inhibitors were added in the medium together with cancer cells (unless stated otherwise) and replaced every other day together with fresh medium. Drugs and inhibitors included in the study are: MEK inhibitor (1 μM PD184352, Sigma-Aldrich P20181), INK inhibitor (10 μM SP600125, Tosco 1496), p38 inhibitor (10 μM SB203580, Tosco 1202), tankyrase inhibitor (5 μM XAV939, Sigma-Aldrich X3040), EGFR inhibitor (1 μM lapatinib, LCLabs.com L-4804), Sec-familie kinase inhibitor (250 μM dasatinib, LCLabs.com D-3307), PI3K inhibitor (1 μM pimonidazole, DCC-0941, Selleckchem S1065) and cilengitide (10 μM, MedChem Express, HY-16141).

Metastasis assays. The study is compliant with all relevant ethical regulations regarding animal research. All protocols were in accordance with UK Home Office regulations under project licence PPL80/2368 and subsequently PPL70/8380, which passed ethical review by the LRI Animal Welfare Ethical Review Board in 2014. In brief, mouse breast cancer cells were trypsinized, washed with PBS, and then resuspended at appropriate concentration before injecting into the tail vein of mice (100 μl per mouse) using a 25 G needle. Before analysis of the lung tissue, mice were killed by cervical dislocation through a 70 μm cell strainer and resuspended in 200 μl of FACS buffer (PBS, 2 mM EDTA and 3% BSA) and counted. The number of GFP+ cells per ml was then measured with a MACSQuant Analyzer (Miltenyi Biotec) with a 96-well plate module.

Quantification of disseminated cells and metastasis. For quantification of disseminated indolent cells after gene knockdown, 5 × 10^4 D2.0R-mCherry shControl cells (Sigma-Aldrich, SHC016) were injected into the tail vein of 6- to 8-week-old female nude athymic BALB/c mice (Charles River). After three weeks, lung tissue were removed, digested with papain and resuspended in HBSS and 2 mM EDTA and sorted according to GFP positivity (Bio Rad S3e Cell Sorter) directly into lysis buffer (1.5–3 × 10^6 cells per sample). Total RNA extraction was performed using Total RNA Purification Plus Kit (Norgen Biotek, 48400) according to the manufacturer’s protocol and the whole RNA eluate was reverse transcribed with SuperScript III (Thermo Fisher Scientific, 18080044) using oligo(dT) as primers. cDNA was further purified with QIAquick PCR Purification Kit (Qiagen, 28106) before qPCR analysis was carried out with triplicate samples of each sample cDNA on QuantStudio 6 Flex Real-Time PCR System (Thermo Fisher Scientific) with a FastStart SYBR Green Master Mix (Roche 48673492010).

For gene sequencing experiments of disseminated breast cancer cells in vivo, 1 × 10^6 D2A1-eGFP cells or D2.0R-eGFP cells were injected into the tail veins of 6- to 8-week-old female nude athymic BALB/c mice (Charles River). After three weeks, lungs were removed, digested into a single-cell suspension as described above and labelled with CD45-APC as indicated above, and CD45-eGFP cells were sorted with Flow Cytometry Facility at Cancer Research UK-LRI and The Francis Crick Institute) directly into lysis buffer and total RNA was extracted with RNeasy Plus Micro Kit (Qiagen) following the manufacturer’s protocol. RNA samples were assessed for quantity and integrity using the NanoDrop 8000 spectrophotometer v2.0 (Thermo Fisher Scientific) and Agilent 2100 Bioanalyzer (Agilent Technologies), respectively. Low level RNA samples were freeze-dried with RNA integrity numbers between 6.4 and 7.8. Full-length cDNA molecules were generated from 4 ng of total RNA per sample using the SMARTer kit for cDNA generation (Clontech). cDNA quantity was measured using the ddDNA High-sensitivity Qubit kit with the Qubit 2.0 Fluorometer (Thermo Fisher Scientific), and were checked for quality using a D1000 ScreenTape with the Agilent 2200 TapeStation (Agilent Technologies). Libraries were prepared using the Illumina Nextera XT Sample Preparation Kit (Illumina) with an input of 150 pg of cDNA per sample. The resulting libraries were checked for average fragment size using the Agilent D1000 ScreenTape, and were quantified using the Qubit dsDNA High-sensitivity reagent kit. Equimolar quantities of each sample library were pooled together and 75 bp paired-end reads were generated for each library using the Illumina NextSeq 500 Highoutput sequencing kit. For in vitro samples, breast cancer cells were grown in multiwell plates under standard culture conditions, trypsinized, sorted and processed in parallel with the in vivo samples.

For qPCR analysis of disseminated breast cancer cells in vivo, cells were isolated and total RNA was purified from homogenized tissue. To obtain enough DNA as template for RT-qPCR analysis, total RNA was amplified with Amplitaq RiboAmp HS PLUS kit before reverse transcription with with dT-primed M–MLV Reverse Transcriptase (Thermo Fisher Scientific, 28025013). qPCR analysis was carried out in a QuantStudio 6 Flex Real-Time PCR System (Thermo Fischer Scientific) with Fast SYBR Green Master Mix (Applied Biosystems 4385612).

For gene expression studies of orthotopic breast tumours, 1 × 10^6 D2A1-eGFP cells or D2.0R-eGFP cells were injected into mammary fat pads of 6- to 8-week-old female nude athymic BALB/c mice (Charles River). After 12 tumour masses were collected, processed and sorted as above. For in vitro samples, breast cancer cells were grown in multiwell plates, trypsinized, labelled and sorted in parallel with the in vivo samples. For gene expression studies of breast cancer cells treated with conditioned medium, 4 × 10^6 AT1-like cells were plated in 10 cm dishes with MLNL medium. After 48 h medium was collected, cleared of dead cells and debris by centrifugation (20 min at maximum speed) and added to breast cancer cells. After 12 h, cells background autofluorescence and the ‘analyse particles’ command was used to acquire the metrics for every contiguous patch of signal (that is, colonies). For experiments using MC7/4 and T47D cells, mice were implanted with a β-oestradiol pellet one week before the injection of cancer cells (0.72 mg per pellet, 90 day release).

Tissue dissociation. Lungs and primary tumours were collected from mice, immersed in PBS and promptly chopped up with scissors to small fragments. Minced lungs were then added to digestion solution (PBS buffer with 75 mg·μl−1 TM Liberase (Roche 0540115001), 75 μg·ml−1 TH Liberase (Roche 05041127001), 12.5 μg·ml−1 DNAase (Sigma-Aldrich DN23)) for 1 h at 37 °C on a rocker. Digested lung pieces were spun down for 5 min at 1,300 rpm, resuspended in calcium- and magnesium-free PBS containing 1 mM EDTA by vigorous pipetting until the solution was homogeneous and then filtered through a 70 μm cell strainer to remove undigested fibrous tissue. In the case of stiffer tissues, such as primary tumours, tissue fragments were also mechanically disrupted by passing them through needles of decreasing thickness. Cells were then pelleted and red blood cells were lysed with Red Blood Cells Lysis Solution (Miltenyi Biotec, 130-094-183) following the manufacturer’s protocol. After washing, cells were resuspended in FACS buffer (PBS, 2 mM EDTA and 3% BSA) and labeled with CD45–APC antibody for 30 min (eBioscience, 30-F11, 1:400) to avoid contamination from leukocytes during sorting. Samples were then washed repeatedly, filtered through a 70 μm cell strainer and kept on ice during FACS.

Gene-expression studies. For gene-expression studies of cancer cells co-cultured with lung stromal cells, 1.36 × 10^6 AT1-like cells per dish were plated onto 6 cm dishes on day 1 (in MLNL medium) followed by 6 × 10^5 cancer cells the next day, in renewed MLNL medium. On day 5, 5 × 10^4 cells were trypsinized, resuspended in 200 μl of FACS buffer (PBS, 2 mM EDTA and 3% BSA) and counted. The number of GFP+ cells per ml was then measured with a MACSQuant Analyzer (Miltenyi Biotec) with a 96-well plate module.

Background
were collected and total RNA was isolated using Total RNA Purification Plus Kit (Norgen Biotechnik, 48400) according to the manufacturer’s protocol. Total RNA was reverse transcribed with t³-primed M-MLV Reverse Transcriptase. qPCR analyses were carried out using the GeneAmp PCR System 9700 (Applied Biosystems) with FastStart SYBR Green Master Mix.

All expression levels were calculated relative to GAPDH. Oligonucleotide sequences used in this study are listed in Supplementary Table 3.

Time-lapse imaging. Twenty thousand AT1-like cells/well were plated onto Lumox 24-well plates (Sarstedt, 94.699.00.14) in MLNL or MHNH medium as indicated. The following day, 2,000 D2.0R cells were plated in the same medium. Three to four hours after plating, the indicated inhibitors were added and 2 h later, cells were imaged for 48 h using either a LSM510 or Nikon Eclipse Ti2. The movies were analyzed to determine the number of mitoses and the number of cell death events.

Library screening. A custom shRNA library was designed on the basis of our in vivo gene-expression data and synthesized by Sigma-Aldrich (custom MISSION shRNA library). All shRNAs are cloned inside pKO.1-based plasmids (TRC version as indicated in Supplementary Table 2) and were individually amplified to avoid representation biases of the clones. We generated 12 shRNA pools, or sets, by randomly combining 14–15 shRNA plasmid clones per set and including a non-targeting control shRNA in each pool (Sigma-Aldrich, SCCHO16) as a quality control of transduction (that is a single guide RNA or single hairpin, accessions) and not for normalization purposes. Plasmid DNA of each set was individually transfected in 293FT cells together with packaging plasmids (pMD2 and psPAX2), collected after 48 h and added to D2.0R-eGFP cells at a low concentration to ensure a single shRNA integration per cell. Successfully transfected cells were selected with puromycin and injected into the tail veins of 6–10-week-old female nude mammary nymth BALB/c mice (3 mice per pool, 3 × 10^6 cells per mouse). After three weeks, lungs were collected and CD45^+eGFP^ D2.0R cells were isolated as described above. Genomic DNA was purified from sorted cells, and from pre-injection samples, with QIAmp DNA Micro Kit (Qiagen) and used as template for two rounds of PCR prior to next generation sequencing. In the first round of PCR we used a forward primer with a unique barcode sequence for each pool, whereas in the second reaction we used primers containing adapter sequences for NGS. All primers and barcodes are listed in Supplementary Table 3. After PCR amplification, DNA fragments were purified and combined in order to obtain four sets, each one containing one sample per pool (one sample pre-injection, three samples after in vivo selection). Samples were sequenced on a paired-end 101 bp run (Illumina HiSeq 2500) and the representation of each shRNA post-injection relative to the representation pre-injection was calculated as described in ‘Bioinformatic analysis’ section.

Stable protein expression. Fluorescent proteins were stably expressed in cancer cells by transduction with retroviruses. pCX4-neo-eGFP or pCX4-blasti-mCherry plasmids were transfected into 293T cells together with packaging plasmids (pCMV-SPORT6 and psPAX2), collected after 48 h and added to D2.0R-eGFP cells at a low concentration to ensure a single shRNA integration per cell. Successfully transfected cells were selected with puromycin and injected into the tail veins of 6–10-week-old female nude mammary nymth BALB/c mice (3 mice per pool, 3 × 10^6 cells per mouse). After three weeks, lungs were collected and CD45^+eGFP^ D2.0R cells were isolated as described above. Genomic DNA was purified from sorted cells, and from pre-injection samples, with QIAmp DNA Micro Kit (Qiagen) and used as template for two rounds of PCR prior to next generation sequencing. In the first round of PCR we used a forward primer with a unique barcode sequence for each pool, whereas in the second reaction we used primers containing adapter sequences for NGS. All primers and barcodes are listed in Supplementary Table 3. After PCR amplification, DNA fragments were purified and combined in order to obtain four sets, each one containing one sample per pool (one sample pre-injection, three samples after in vivo selection). Samples were sequenced on a paired-end 101 bp run (Illumina HiSeq 2500) and the representation of each shRNA post-injection relative to the representation pre-injection was calculated as described in ‘Bioinformatic analysis’ section.

Proliferation assays. Breast cancer cells were plated on flat bottom 96-well plates (2,000 cells per well) and confluence was measured over time with Incucyte (Essen Bioscience) every 3–4 h for 100 h. Percentage of covered area was Log-transformed and plotted against time. The 95% confidence bands of the best-fit line were calculated and these total raw counts were normalized to the maximum total number of reads across all samples to enable direct comparisons between samples.

Single-cell mRNA sequencing. Total RNA was isolated using Tractron 100 automated RNA extraction system (Fluidigm) with FastPrep (MP Biomedicals) and DNase free water. RNA samples were then converted to cDNA using Superscript III Reverse Transcriptase (Invitrogen). cDNA samples were amplified using Superscript III Platinum One-Step cDNA Synthesis Kit (Invitrogen) with SYBR Green I and Fusion Red dyes. Amplified cDNA quality was confirmed using Low Preset v1.3 and sequencing was performed on 384-well plates with FastTrack Prime Flowcell on a NextSeq 500 (Illumina). Single-cell sequencing data was then processed using CellRanger (version 3.0, SingleCellGenomics) with the appropriate parameters to perform a GSEA analysis. Sequencing was performed on 384-well plates with FastTrack Prime Flowcell on a NextSeq 500 (Illumina). Single-cell sequencing data was then processed using CellRanger (version 3.0, SingleCellGenomics) with the appropriate parameters to perform a GSEA analysis.

For frozen sections, lungs were perfused with 4% PFA in PBS immediately following sacrifice. Whole lungs were rapidly frozen. Ten micrometre sections were cut before staining. Slides were fixed in 4% PFA for 15 min at room temperature and blocked with PBS, 0.2% Triton X-100, 0.5% BSA for 1 h at room temperature. After blocking, slides were incubated with primary antibodies overnight at 4°C and washed with PBS, 0.2% Triton X-100, 0.5% BSA for 2 × 10 min at room temperature. After washing, slides were incubated with secondary antibodies (diluted 1:100) for 1 h at room temperature and washed with PBS, 0.2% Triton X-100, 0.5% BSA for 2 × 10 min at room temperature. After washing, slides were incubated with secondary antibodies (diluted 1:100) for 1 h at room temperature and washed with PBS, 0.2% Triton X-100, 0.5% BSA for 2 × 10 min at room temperature. After washing, slides were incubated with secondary antibodies (diluted 1:100) for 1 h at room temperature and washed with PBS, 0.2% Triton X-100, 0.5% BSA for 2 × 10 min at room temperature. After washing, slides were incubated with secondary antibodies (diluted 1:100) for 1 h at room temperature and washed with PBS, 0.2% Triton X-100, 0.5% BSA for 2 × 10 min at room temperature.
acquired with a Zeiss LSM 780 using ZEN software. Antibodies used in this study are: PDPN (1:100, Acris DM3501), AQP5 (1:100, Abcam ab78486), SP-C (1:100, Abcam Ab90716), CD68 (1:100, Biologend 1370704), Vimentin (1:100, Abcam ab92547), α-smooth muscle actin (1:200, Sigma C6198). EDU incorporation was visualized with Click-IT Plus Edu Alexa Fluor 647 (Invitrogen C10640) in accordance with the manufacturer's instructions. For in situ staining, the same steps were performed (excluding the freezing in OCT and sectioning) with the modification that all blocking and antibody steps were performed for at least 24 h at 4 °C.

Immunofluorescence. Cells were fixed in 4% PFA for 15 min at room temperature. After washes, cells were permeabilized with PBS and 0.2% Triton X-100 for 5 min at room temperature and blocked with immunofluorescence buffer (PBS, 0.05% Tween-20 and 3% BSA for Ki67 or PBS and 3% BSA for other staining) for 1 h. Primary antibodies were incubated in immunofluorescence buffer overnight at 4 °C in a wet chamber. The day after, cells were washed several times with immunofluorescence buffer and incubated with secondary antibodies for at least 1 h at room temperature together with DAPI (1 μg/ml stock, 1:500; Sigma-Aldrich D9542) and Phalloidin (Phallolidin–Atto 633, 20 μM stock, 1:1,000; Sigma-Aldrich 68825) when indicated. Images were acquired with a Zeiss LSM 780 using ZEN software. Antibodies used in this study are: Ki67 (1:1,000, Abcam ab15580), fibronectin (1:500, Sigma F3648) and pSRC Y418 (1:100, Invitrogen, 44–660 G).

Western blotting of conditioned medium. To visualize soluble SFRP2 protein, confluent D2.0R cells were cultivated in DMEM without serum. After 5 d, conditioned medium was pooled from three 15 cm dishes per condition, spun for 20 min at maximum speed to remove debris and then concentrated by spinning the samples for 30 min at 4 °C at 3,000g (Amarion Ultra-15 Centrifugal Filter Devices 30,000 MWCO, Millipore). As loading control, remaining cells were collected and processed as in ref. 1. Western blotting was performed as in ref. 2. Antibodies: SFRP2 (1:1,000, Abcam, ab137560), GAPDH (1:25,000, Millipore, MAB374). The antibody for SFRP2 has been validated with recombinant mouse SFRP2 (R&D Systems, 1169-FR).

Cell morphology assessment. To calculate circularity, we used the ImageJ plug-in described in the following link: https://imagej.nih.gov/ij/plugins/circularity.html. This calculates circularity (4π(area/perimeter²)). When less than 50 cells were being measured, manual tracing of the cell outline was used to ensure that single cells were being analysed. When more than 50 cells were measured, automatic thresholding was used. This latter method precludes a definitive determination of whether a GFP patch contains a single cell or a small cluster of cells. Hence, we use the term cell/colony circularity to reflect that the measurement includes both isolated cell and micro-cluster values. Cell extensions of more than 15 μm in length were classified as protrusions in manual scoring.

Statistics and reproducibility. Statistical analyses used GraphPad Prism software. For experiments with samples sizes greater than ten, normality of data was tested with the Shapiro–Wilk test. For normally distributed samples, we performed Student’s two-tailed t-test for single comparisons (paired or unpaired) and one-way ANOVA for multiple comparisons. In case of different variances within samples to be compared, we applied Welch’s correction. For non-normal data, we performed two-tailed Mann–Whitney test for analysis of unpaired data and two-tailed Wilcoxon matched pairs rank test for paired data. For multiple comparisons of non-normal data, we applied Dunn’s test. For samples below ten in size, it is not easy to assess the underlying distribution of the data, and non-parametric tests were preferred, unless the sample size was below five, in which case we preferred parametric tests owing to the minimum possible P value becoming large in the non-parametric case. Data are plotted as the mean ± SEM of all independent experiments. In some experiments, the mean-normalized values from all independent experiments are plotted to provide information about assay variability. For animal experiments, each mouse was considered as a biologically independent sample. Linear regression P values are calculated from the observed t-statistic ratio of the parameter estimates to their standard errors. For survival plots (Kaplan–Meier analysis), data were analysed with GraphPad Prism software, GOBO (http://cob.mhc.lsu.edu/goibo/gabo.php) or KM Plotter (https://kmplot.com/analysis/) online tools which all calculate log-rank P value (Mantel–Cox method). For analysis with GraphPad Prism, P values calculated with Gehan–Breslow–Wilcoxon methods are provided. GSEA is generated from the GSEA online tool (https://software.broadinstitute.org/gsea/index.jsp), which also calculates the two principal statistics of the analysis: NES and false discovery rate (FDR). NES is calculated by normalizing enrichment score to gene-set size; false discovery rate represents an estimated likelihood that a gene set with a given NES represents a false positive. The threshold for significance was set at 0.05 for all experiments except for GSEA, where we considered a significant FDR as below 0.25. Data in histograms are presented as mean ± S.D. (if present) unless stated otherwise.

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

Data availability
RNAseq data have been deposited at the Gene Expression Omnibus with accession number GSE120628. Other data that support the findings of this study are available upon reasonable request from the corresponding authors.

References
35. Liberson, A. et al. Molecular signatures database (MSigDB) 3.0. Bioinformatics 27, 1739–1740 (2011).
36. Merico, D., Isslerin, R., Steuiker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS One 5, e13984 (2010).
37. Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics 26, 589–595 (2010).
38. Behebhani, G. K., Bendall, S. C., Clutter, M. R., Fandt, W. J. & Nolan, G. P. Single-cell mass cytometry adapted to measurements of the cell cycle. Cytometry A 81, 552–566 (2012).
39. Zunder, E. R. et al. Palladium-based mass tag cell barcoding with a doublet-filtering scheme and single-cell deconvolution algorithm. Nat. Protoc. 10, 316–333 (2015).
40. Finn, R. et al. Normalization of mass cytometry data with bead standards. Cytometry A 83, 483–494 (2013).
41. van Dijk, D. et al. Recovering gene interactions from single-cell data using data diffusion. Cell 174, 716–729 (2018). e727.
42. Poceterra, A. et al. F-actin dynamics regulates mammalian organ growth and cell fate maintenance. J. Hepatol. 71, 130–142 (2019).

Acknowledgements
We thank J. Downward, D. Barkan and R. Goris for gifts of cell lines; I. Malanchi, S. Piccolo, S. Dupont and G. Martello for thoughtful discussion and reagents; Flow Cytometry, Experimental Histopathology, Bioinformatics and Biostatistics (in particular S. Horswell), Biological Research, Cell Services and Advanced Sequencing facilities at the Crick Institute for exceptional scientific and technical support throughout the project; and C. Mein for support and advice for RNA sequencing. E.S. and M.M. were funded by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001144), the UK Medical Research Council (FC001144) and the Wellcome Trust (FC001144). M.M. also received funding from Marie Curie Actions—Intra-European Fellowships no. 625496 and BIRD Seed grant from Department of Molecular Medicine (University of Padua). C.J.T. and J.S. are supported by a Cancer Research UK Career Development Fellowship awarded to C.J.T.

Author contributions
M.M. and E.S. conceived, designed, and wrote the study. M.M. performed all the experiments with the exception of: the CyTOF analysis, which was performed by R.B. with assistance from X.Q. and J.S.; and supervision from C.J.T.; some of the in vitro co-cultures, which were performed by S.H. and E.S.; and the in vivo analysis of proliferation in the lungs and sFRP2 overexpression, which were performed by S.H. with assistance from A.B., Y.N. and E.S. C.D.H.R. and A.P. assisted with cell culture and analysis of gene expression. J. Downward, D. Barkan and R. Goris for gifts of cell lines; I. Malanchi, S. Piccolo, S. Dupont and G. Martello for thoughtful discussion and reagents; Flow Cytometry, Experimental Histopathology, Bioinformatics and Biostatistics (in particular S. Horswell), Biological Research, Cell Services and Advanced Sequencing facilities at the Crick Institute for exceptional scientific and technical support throughout the project; and C. Mein for support and advice for RNA sequencing. E.S. and M.M. were funded by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001144), the UK Medical Research Council (FC001144) and the Wellcome Trust (FC001144). M.M. also received funding from Marie Curie Actions—Intra-European Fellowships no. 625496 and BIRD Seed grant from Department of Molecular Medicine (University of Padua). C.J.T. and J.S. are supported by a Cancer Research UK Career Development Fellowship awarded to C.J.T.

Competing interests
The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41556-020-0474-3. Supplementary information is available for this paper at https://doi.org/10.1038/s41556-020-0474-3.

Correspondence and requests for materials should be addressed to M.M. or E.S.

Reprints and permissions information is available at www.nature.com/reprints.
Extended Data Fig. 1 | Response of lung stromal cells to dissemination of indolent breast cancer cells. a. Heatmap of Estrogen Receptor (ESR1) and HER2 (ERBB2) expression in D2 cells in vivo based on RNAseq presented in Fig. 3. Heatmap has been generated with ClustVis tool (https://biit.cs.ut.ee/clustvis/#pathways). b. Immunohistochemistry for GFP of lungs from nude mice injected with D2.0R- or D2A1-EGFP labelled cells. n=2. Scale bar is 100 μm. c. Representative images of lungs from wild-type BALB/c mice injected either with D2.0R-EGFP or with D2A1-EGFP. Lungs were collected and imaged on the GFP channel at the lung surface. Dashed box, lung area magnified in the middle image. n=4. Scale bars, 1 mm (low magnification) or 100 μm (high magnification). d. Circularity of D2.0R and D2A1 cells within the lung parenchyma at 4 days after injection (n=23 cells). Unpaired t-test. e. Fluorescent IHLC of D2.0R-EGFP cells in the lungs two weeks after intravenous injection Left: Magenta, AT2 cells (TTF1+). Middle and Right: Magenta, AT2 cells (SP-C+); Yellow: myeloid cells (CD68+); Blue, AT1 cells (PDPN+). Right panel shows a control lung without D2.0R-EGFP cells. n=3. Scale bar, 20 μm. f. Percentage of D2.0R-EGFP cells in contact with each lung stromal cell type at 3- or 14-days post-injection. n=3 mice. Black: in contact; white: not in contact. AT1: Alveolar Type 1 cells (PDPN+); F: Fibroblasts (VIM+); EC: Endothelial cells (MUC+); AT2: Alveolar Type 2 cells (SP-C+); Act-F: Activated Fibroblasts (aSMa+); M: Macrophages (CD68+). g. Colony area (left) and proliferating number (right) of disseminated D2A1-EGFP or D2.0R-EGFP cells upon tail vein injection at different time points. n=3 mice/group. h. Number of proliferating PDPN+ve cells surrounding metastatic lesions and disseminated cells in g. i. Number of proliferating PDPN+ve cells surrounding MCF7-GFP cells after tail vein injection. n=3 mice/group. d,g,h,i plots show data as whisker plots: midline, median; box, 25-75th percentile; whisker, minimum to maximum.
Extended Data Fig. 2 | Lung stromal cells regulate proliferation of breast cancer cells in vitro. a, Relative mRNA levels of stromal cell-type specific markers of the different cellular populations included in the lung coculture system. AT1, alveolar type 1 cells; AT2, alveolar type 2 cells. Dots are means from independent experiments (n = 3). Unpaired t-test. b, Percentage of Ki67+ -D2.0R-EGFP or -D2A1-EGFP cells cultivated together with lung stromal cell lines in Mitogen Low-Nutrients Low medium (MLNL) for 4 days. Mean normalized pooled samples (n = 7) from independent experiments (n = 2). Data presented as whisker plot: midline, median; box, 25–75th percentile; whisker, minimum to maximum. Unpaired t-test. c, Growth curves of D2.0R-EGFP and D2A1-EGFP cells in vitro with permissive (MHNH) or restrictive (MLNL) medium. d, D2.0R-EGFP cells were cocultured with lung stromal cells for 5 days (or on air-permeable surface only as control), isolated by fluorescence-activated cell sorting (FACS), and their growth kinetic in vitro in MLNL on standard plastic plates measured over time (lines are overlapped). n = 2 independent experiments. For c and d, Confluency values at indicated time points were log10-transformed and linear regression was calculated. Line was forced to go through the origin. n = 6 biological replicates/group, n = 2 independent experiments. Solid line, mean of best-fit line; dashed lines, 95% confidence bands.
Extended Data Fig. 3 | Alveolar type I cells promote formation of protrusions and survival of indolent breast cancer cells in vitro. a, Relative number of 4T07-EGFP or MCF7-EGFP cells cultivated alone or together with AT1-like cells in MLNL medium for 5 days. Mean normalized pooled samples, n = 12 from 3 independent experiments for 4T07 (Mann-Whitney test), n = 24 from 4 independent experiments for MCF7 (unpaired t-test). b, Plots show the relative frequency (number of events/starting number of D2.0R cells) of mitotic (left) and apoptotic (right) events in D2.0R cells cultured in MLNL media in the absence or presence of AT1-like cells. n = 7 independent experiments. Wilcoxon test. c, Relative number of D2.0R-EGFP cells cocultured with different combinations of lung stromal cells (as indicated) in MLNL medium for 5 days. Mean normalized pooled samples (n = 9) from independent experiments (n = 3). Mann-Whitney test. d, Quantification of D2.0R cells co-cultured with individual lung stromal cells in MhNh medium for 7 days. Pooled samples (n = 8) from independent experiments (n = 3). Dunn’s multiple comparisons test. e, Circularity of D2.0R cells alone or cocultured with AT1-like cells (n = 26–29 cells). Unpaired t-test. f, Percentage of indicated human breast cancer cells with protrusions alone or in coculture with AT1-like cells. n = 3 independent experiments. Paired t-test. g, Representative images of protrusions induced by coculturing of MCF7 cells with AT1-like cells. n = 3. Scale bar, 20 μm. h, Immunofluorescence for fibronectin (FN) of 4T07 cells cultured alone (left) or co-cultured with AT1-like cells (right). n = 3. Scale bar, 20 μm. i, Percentage of D2a1 cells with protrusions alone or in coculture with AT1-like cells. n = 3 independent experiments. Paired t-test. j, Percentage of MCF7 cells with protrusions, in coculture with AT1-like cells, upon treatment with cilengitide. n = 3 independent experiments. Paired t-test. a–e plots show data as whisker plots: midline, median; box, 25–75th percentile; whisker, minimum to maximum.
Extended Data Fig. 4 | Mass cytometry analysis reveals signaling pathways involved in the crosstalk between AT1 and indolent breast cancer cells.

a, Schematics representation of the experimental outline of mass cytometry assay. b, Heatmaps of EMD values showing the activation of relevant markers in AT1-like cells cocultured with D2.0R or MCF7. n = 1 experiment. c, Plot showing increase phospho-HistoneH3 (S28) signal in AT1-like cells co-cultured with D2.0R cells. Representative of two independent experiments. d, DREI plots showing the relationship between the indicated phospho-antibody signals in D2.0R monocultures or cocultures with AT1-like cells (DREMI score in upper left corner). Representative of two independent experiments.

e, Number of cells after the indicated treatment (for two days) relative to untreated cells. Mean values from n = 3 (for Control, EGFRi, SFKi) and 4 (for MEKi) independent experiments. One-way ANOVA test. f, Histogram of EMD values showing the inhibition of P-ERK abundance in D2.0R cells cocultured with AT1-like cells. Bars show the average of two technical replicates. g, Plot shows the area of D2.0R colonies ten days after intravenous delivery into either control Balb/C nude mice (n = 1147 across 5 mice) or Trametinib treated mice (n = 179 across 4 mice, MEKi). Mann-Whitney test. e, g plots show data as whisker plots: midline, median; box, 25–75th percentile; whisker, minimum to maximum.
Extended Data Fig. 5 | EGFR and SFK, but not MEK, regulate protrusions and fibrillogenesis induced by coculture of indolent breast cancer cells with AT1-like cells. a, Percentage of D2.0r cells with protrusions after treatment with indicated inhibitors for two days. Mean values of n = 3 independent experiments. One-way ANOVA test. b, D2.0R-EGFP cells have been treated for two days with indicated drugs. Fibronectin fibrils were quantified after immunostaining. Mean values from n = 3 (for EGFri and MEKi) and 5 (for Control and SFKi) experiments. One-way ANOVA test between “extensive fibrils” category. c, Percentage of MCF7 cells with protrusions after treatment with SFKi in monoculture or coculture with AT1-like cells. n = 2 (for Control alone) and 3 (for Control coculture and SFKi treated) independent experiments. One-way ANOVA test. d, Relative expression of fn mRNA in D2.0R-EGFP cells cultivated with AT1-like cells in MLNL medium +/- SFKi for 4 days. Mean normalized pooled samples (n = 12) from independent experiments (n = 3). Mann Whitney test. e, Images showing F-actin and activated phospho-Src (Y418) in D2.0R cells cocultured with AT1-like cells. n = 2. Scale bar is 20μm.
Extended Data Fig. 6 | Gene expression analysis of lung-disseminated indolent breast cancer cells *in vivo*. **a**, Representative GSEA analysis from the top up- and down-regulated gene sets in D2.0R cells *in vivo* compared to the other groups (relative to Fig. 3a). NES, normalized enrichment score. FDR, false discovery rate. **b**, Heatmap shows normalized expression values for two dormancy markers (*Bhlhe41* is also known as *Dec2* and *Sharp1*).
Extended Data Fig. 7 | Survival analysis of patients according to dormancy signature. a, Kaplan-Meier curves of ER+ breast cancer patients from publicly available datasets used in Kim et al., 2012, stratified according to the dormancy signature. b, Overlap between our dormancy signature and genes included in the dormancy score or 51-gene signature (Supplementary Table 1). c, Kaplan-Meier curves showing DMFS (Distant Metastasis Free Survival) of ER+ breast with high and low expression of dormancy signature. Data have been plotted starting from month 0 (left) or month 24 (right). Indicated statistical tests are two-sided. d, Kaplan-Meier curves showing DMSF of patients derived from http://co.bmc.lu.se/gobo/gsa.pl stratified according additional signatures generated from the other groups analysed with RNAseq in Fig. 3 (Supplementary Table 1). f, BMP and Wnt target genes expression as in Fig. 3f. Mean normalized pooled samples, n = 8 samples from 3 independent experiments. Mann-Whitney test.
Extended Data Fig. 8 | A loss-of-function screen in vivo identifies SFRP2 as survival regulator in lung disseminated indolent breast cancer cells.

a, Volcano plot of RNAseq expression data of D2.0R cells in vivo compared to the other groups (as in Fig. 3a). In blue, candidate genes selected for step 2 validation. b, Step 2 validation of candidate genes. Subpopulations of D2.0R-EGFP cells bearing a single shRNA for the indicated gene were individually generated (3 shRNA sequences/gene). Cells with shRNA for the same gene were mixed together in equal amount, injected in tail vein of BALB/c nude mice (n = 6 mice for Control and n = 3 for the other groups) and processed as in Fig. 4c. Unpaired t-test with Welch’s correction. c, Subpopulations of D2.0R-EGFP-shSfrp2 cells were mixed and injected in the tail vein with an equal amount of D2.0R-mCherry-shControl. After 3 days to allow seeding and extravasation in the lung parenchyma, lungs were collected and GFP+ and mCherry+ simultaneously quantified to rule out pre-dissemination role of SFrP2 (n = 4 mice). Scale bar, 1 mm. Unpaired t-test with Welch’s correction. d, In vitro growth curves of D2.0R-EGFP cells bearing the indicated shRNAs for Sfrp2. Confluency values at indicated time points were log10-transformed and linear regression was calculated (n = 2 independent experiments). Line was forced to go through the origin. Solid line, mean of best-fit line; dashed lines, 95% confidence bands. e, Relative expression levels of Sfrp2 in D2.0R-EGFP cells on plastic, isolated from mammary fat pad or lung-disseminated (n = 5 wells for Control group, n = 3 mice for the other groups). Unpaired t-test. f, Histogram showing the induction of SFRP family members by AT1-like conditioned media in both D2.0R and 4T07 cells. Mean normalized pooled samples (n = 9) from independent experiments (n = 3–4). Mann-Whitney test.
Extended Data Fig. 9 | SFRP2 regulates protrusions independently of Wnt signaling. a, Left, qPCR for canonical Wnt target genes of D2.0r-EGFP carrying interfering sequences for Sfrp2 cultivated with AT1-like cells in MLNL medium for 4 days. Right, qPCR for the Wnt target Axin2 in control and SFRP2-overexpressing cells. Mean normalized pooled samples from independent experiments (n = 3 for CCND2 and n = 4 for Axin2). Mann-Whitney test. b, Conditioned media from confluent D2.0R-EGFP-Control or SFRP2 OE cells plated in MLNL were concentrated and analyzed by Western Blotting. Cells have been treated or not with 50ug/mL of Heparin to allow SFRP2 solubilization in the medium. Representative of two independent experiments. Unprocessed blots are available as Source Data file. c, Plot shows the effect of heparin, which binds and inhibits SFRP family proteins, on D2.0R cell number when co-cultured with AT1-like cells. Mean normalized pooled samples (n = 18) from independent experiments (n = 3). Plots shows data as whisker plots: midline, median; box, 25–75th percentile; whisker, minimum to maximum. Unpaired t-test. d, Control or SFRP2 overexpressing MCF7 cells were plated alone or in presence of AT1-like cells. Plot shows the percentage of cells with protrusions in each experiment. n = 3 independent experiments. Paired t-test. e, Quantification of cell D2.0R cell proliferation (as judged by mitoses) in the indicated conditions: +/- AT1-like cells, +/- SFRP2 over-expression, and +/- SFKi treatment. Mean and S.E.M. are shown (n = 5 independent experiments). Unpaired t-test.
Extended Data Fig. 10 | SFRP2 regulates survival of disseminated indolent breast cancer cells. **a**, Quantification of metastatic burden two weeks after intravenous injection of human indolent breast cancer cell. n = 6 slices across 3 mice for T47D-DBM cells and n = 10 slices across 3 mice for MCF7 cells. Plots show data as whisker plots: midline, median; box, 25–75th percentile; whisker, minimum to maximum. Unpaired t-test. **b**, In vitro growth curves of Control and SFRP2 over-expressing D2.0R and 4T07 cells. Confluency values at indicated time points were log10-transformed and linear regression was calculated. Line was forced to go through the origin. Solid line, mean of best-fit line; dashed lines, 95% confidence bands. n = 2 independent experiments. **c**, Proximity of disseminated SFRP2-overexpressing D2.0R cells to indicated lung stromal cells at 3 or 14-days post-injection. Lung slices from 3 mice injected with D2.0R-EGFP cells have been stained with multiple markers for different stromal subpopulations. Graphs indicate the percentage of EGFP+ cells in contact with each stromal cells subtype (black: in contact; white: not in contact). Staining as in Extended Data Fig. If. **d**, Schematic illustration of the signaling between AT1 cells and breast cancer cells that supports metastatic persistence.
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed
---|---
☑️ | The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☑️ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☑️ | The statistical test(s) used AND whether they are one- or two-sided
☑️ | Only common tests should be described solely by name; describe more complex techniques in the Methods section.
☑️ | A description of all covariates tested
☑️ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☑️ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) and variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☑️ | For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
☑️ | Give P values as exact values whenever suitable.
☑️ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☑️ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☑️ | Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection

All confocal and time lapse images were acquired with a Zeiss LSM 780 using ZEN 2009 software (version 6.0,0.303). Genomic DNA was sequenced with Illumina HiSeq 2500. RNAseq data were collected with Illumina NextSeq 500. In vitro cellular confuency was measured with Incucyte S3 and Incucyte Zoom (Sartorius). qPCR data were acquired with QuantStudio 6 Flex Software 1.0. Mass cytometry data were collected with Helios Mass-Cytometer (Fluidigm 6.7.1014). BD FACS AriaIII and BD Influx used for FACS analysis run BD FACSDiva v8.0.1.

Data analysis

For RNAseq data: RSEM package (version 1.2.11) and Bowtie2 were used to align reads to the mouse mm10 transcriptome, taken from refGene reference table available at UCSC downloaded on May 2014 [https://genome.ucsc.edu/]. TMM (treated mean of M-values) normalisation and differential expression analysis using the negative binomial model was carried out with the R-Bioconductor package "Deseq2" (www.bioconductor.org R version 3. 1.0), Geneset enrichment Analysis was carried out with GSEA tool (version 2.2.3). Enrichment map was generated with Enrichment map plug-in for Cytoscape (3.6.0). For shRNA library screening: generated sequences were mapped against all annotated library sequences using Burrows-Wheeler Aligner v0.5.9. For mass cytometry: files from Helios Mass-Cytometer were analysed with Cytobank (Version 7.2.0) and publicly available R and python packages. Specifically: Python 3.6 with packages scprep 0.12.2 and scikit-learn 0.21.1. R 3.5.2 with packages ggplot2 3.0.0 and RColorBrewer 1.1-2. All the scripts used for data analysis are available at the repository of the Cell Communication Lab at UCL’s Cancer Institute (https://github.com/TAPE-Lab/CyTOF_DataAnalysis). BD FACS AriaIII and BD Influx used for FACS analysis run BD FACSDiva v8.0.1. Images were analyzed with Fiji (Image), version 1.8.0). All other data were analyzed with Microsoft Excel 16.16.2 for Mac or Prism 7 for Mac.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.
Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNAseq data have been deposited at GEO Database (GSE120628) and will be available concomitant with publication. Other data that support the findings are available upon reasonable request from the corresponding authors.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size
Sample size was determined based on previous experience, published literature, or to specific requirements of a given technique.

Data exclusions
No data were excluded.

Replication
All attempts at replication were successful, provided that internal controls for technical reliability were acceptable. All data in the paper was successfully replicated in at least two independent experiments performed under identical conditions (unless stated otherwise). Several steps were taken to ensure the reproducibility of experimental findings. Cell cultures were regularly checked for mycoplasma and cross contamination (using STR profiling). Key results were confirmed in several human and mouse cellular models.

Randomization
Randomization was used to assign D2.0R-shRNA clones to different pools for the shRNA screen in vivo. Standardized cell culture conditions (except for the intended experimental perturbation) were used to minimize variation across samples. Western blot and RNA analysis used bulk cell lysates and therefore cell selection methods are not applicable.

Blinding
RNAseq data analysis was double-blinded, and hierarchical clustering indicated significant differences and correct grouping before sample identity disclosure. shRNA screening were blinded to the experimenter by allocating numerical identifiers to each reagent. Other experiments in this work necessitated frequent interventions by the researcher and repeated microscopic intervention during the protocol, this precludes effectively blinding.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | Antibodies            |
| ☒   | Eukaryotic cell lines |
| ☒   | Palaeontology         |
| ☒   | Animals and other organisms |
| ☒   | Human research participants |
| ☒   | Clinical data         |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
| ☒   | ChIP-seq              |
| ☒   | Flow cytometry        |
| ☒   | MRI-based neuroimaging |

Antibodies

| Antibodies used |
|-----------------|
| Antibodies for immunohistochemistry: PDPN (1:100, Acris DM3501), AQPS (1:100, Abcam ab78486), SP-C (1:100, Abcam Ab50716), CD68 (1:100, Biolegend 137004), Vimentin (1:100, Abcam ab92547), αSMA (1:200, Sigma C6198). Antibodies for immunofluorescence: Ki-67 (1:1000, Abcam ab15580), Fibronectin (1:500, Sigma F3648), phospho-Src Y418 (1:100, Invitrogen, 44-660G). Antibodies for Western Blotting: SFRP2 1:1000 (Abcam, ab137560), GAPDH 1:25000 (Millipore, MAB374). Antibodies for Mass Cytometry. Metal Antibody Name Clone Supplier |
Validation

Antibodies used for immunofluorescence: Ki67 and EdU antibody shows correct cellular localization, correlation with growth promoting/suppressive signals and has been validated in knock-out cells by manufacturer. Fibronectin antibody was validated with fibronectin-coated coverslips. Phospho-Src antibody shows correct cellular localization and is inhibited by treatment with Src-inhibitor Dasatinib.

Antibodies used for immunohistochemistry show expected pattern of histological staining and colocalization with additional markers of same cell lineages.

Cell-type specific antibodies for mass cytometry were validated using cells known to and not-to express the antigen. All signalling antibodies were validated using stimulation of upstream pathways (e.g. using growth factors or serum) and inhibition of upstream kinases.

Antibody for SFRP2 used in Western blotting was validated with mouse recombinant sFRP2 protein (in Supplementary Information file).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s) Alveolar-Type1 cells [TT1 cells] were a kind gift of Prof. J. Downward (The Francis Crick Institute, London). Alveolar-Type2 cells (H441 cells) were purchased from ATCC (HTB-174). Human Normal Lung Fibroblasts (HNLf) were derived from primary lung fibroblasts (CRUK Cell Service AG02603). D2.OR, D2.A1 and MCF-EGFP cells were a kind gift of D. Barkan (University of Haifa). 4T07 were gently provided by Prof. Stefano Piccolo (University of Padua). T47D-DBM cells were a gift of Prof. R. Gomis (IRB, Barcelona). 293FT cells from Thermofisher. HEK293T cells from CRUK Cell Service.

Authentication TT1, H441, HNLf, 293FT, HEK293T have been STR profiled and species verified. D2.OR, D2.A1, 4T07, MCF7, T47D-DBM have been species verified and tested for the expected phenotype in vivo.

Mycoplasma contamination All cells are routinely screened for mycoplasma at Cell Services facility at The Francis Crick Institute and/or with Universal Mycoplasma Detection Kit (ATCC, 30-1012X) at University of Padua. All the cell lines tested were negative for mycoplasma contamination.

Commonly misidentified lines (See ICCLAC register) No commonly misidentified cell lines were used.
Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

| Laboratory animals | 6- to 8-weeks old female nude athymic or wild-type BALB/c mice |
|--------------------|---------------------------------------------------------------|
| Wild animals       | This study does not involve wild animals                      |
| Field-collected samples | This study does not involve samples collected from the field |
| Ethics oversight   | The study is compliant with all relevant ethical regulations regarding animal research. All protocols were in accordance with UK Home Office regulations under project licence PPL80/2368 and subsequently PPL70/8380, which passed ethical review by the LRI Animal Welfare Ethical Review Board in 2014. |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:
- The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).
- The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
- All plots are contour plots with outliers or pseudocolor plots.
- A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation | A detailed protocol for samples preparation is provided in Methods section. |
Instrument          | BD FACS AriaIII and BD Influx |
Software            | BD FACSDiva v8.0.1 |
Cell population abundance | GFP+ or mCherry+ disseminated breast cancer cells were on average 0.03-0.15% of the total processed lung cells, according to injection efficiency. |
Gating strategy     | Gates applied in every experiment: SSC-A/FSC-A gate was used to remove debris and dead cells, SSC-A/SSC-H and FSC-A/FSC-H were used to remove cell doublets, CD45-APC negative populations were selected with 670/30 filter. Were indicated GFP signal was detected with 530/30 filter and mCherry signal was detected with 610/20 filter. Boundaries were defined based on negative controls (uninjected/unlabelled lungs, uninjected/CD45-labelled lungs) and positive controls (GFP+ and mCherry+ parental cells). |

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.