A novel labelling strategy enables spatial resolution of the lung metastatic niche and uncovers tissue stem cell-like features proximal to breast cancer cells

Luigi Ombrato[1], Emma Nolan[1], Ivana Kurelac[1,3], Antranik Mavousian[2], Victoria Bridgeman[1], Probir Chakravarty[4], Stuart Horswell[4], Ivonne Heinze[5], Estela Gonzales-Gualda[1], Giulia Matacchione[1], Anne Weston[6], Joanna Kirkpatrick[5], Ehab Husain[7], Valerie Speirs[8,] Lucy Collinson[6], Alessandro Ori[5], Joo-Hyeon Lee[2*], Ilaria Malanchi[1*]

1 Tumour Host Interaction laboratory, The Francis Crick Institute, 1 Midland Road, NW1 1AT London; 2 Wellcome Trust/Medical Research Council Stem Cell Institute, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK; 3 Dipartimento di Scienze Mediche e Chirurgiche, University of Bologna, Via Massarenti 9, 40138 Bologna, Italy; 4 Bioinformatics & Biostatistics Unit, The Francis Crick Institute, 1 Midland Road, NW1 1AT London; 6 Electron Microscopy Unit, The Francis Crick Institute, 1 Midland Road, NW1 1AT London; 5 Proteomics of aging, Leibniz Institute on Aging, Fritz Lipmann Institute (FLI) Beutenbergstrasse 11, 07745 Jena, Germany; 7 Department of Pathology, Aberdeen Royal Infirmary, Foresterhill, Aberdeen, AB25 2ZN, Scotland, UK; 8 Institute of Medical Sciences, University of Aberdeen, Foresterhill, Aberdeen, AB25 2ZD, Scotland, UK.

**Corresponding authors: Ilaria Malanchi (Ilaria.Malanchi@crick.ac.uk) and Joo-Hyeon Lee (jhl62@cam.ac.uk).*

Cancer cell behaviour is strongly influenced by the surrounding cellular environment, making the characterisation of the local tumour microenvironment (or niche) a fundamental question in tumour biology. To date, a direct investigation of the early cellular changes induced by metastatic cells within the surrounding tissue is difficult to achieve, especially at early micro-metastatic stages and for low frequency populations present in the niche. Here we present the strategy whereby metastatic cancer cells release a cell-penetrating fluorescent protein that is efficiently taken up by neighbouring cells, allowing spatial identification of the local metastatic cellular environment within the whole tissue. Notably, this strategy can be used to follow metastatic niches from early micro-metastasis to late macro-metastasis, allowing temporal resolution. Moreover, the presence of low represented niche cells can be detected and characterised among the bulk tissue. To highlight the potential of this niche-labelling strategy, we have applied this system to the study the lung metastatic environment of breast cancer cells. We report the unprecedented presence of cancer parenchymal associated cells (CAPs) within the lung metastatic niche, where lung epithelial cells show stem cell-like features with expression of lung progenitor markers, multilineage differentiation potential and self-renewal activity. Moreover, lung epithelial cells are directly perturbed by cancer cells in *ex vivo* co-culture assays and support their growth. In summary, here we describe a novel labelling system that enables spatial resolution of the metastatic microenvironment and provide evidence that the tissue cellular environment surrounding metastatic growth is characterised

by undifferentiated features. The data highlight the significant potential of this method as a platform for new discoveries.

During the early phase of metastatic growth, cancer cells generate a local tissue microenvironment (metastatic niche), which is very distinct from the normal tissue structure and key to support their survival and growth[1]. However, a detailed analysis of the cellular composition of the metastatic niche, especially at early stages, is significantly constrained by the difficulty to spatially discriminate the cells in the metastatic niche from the bulk of the tissue. This hampers identification of tissue cells that might respond to the early cancer infiltration but remain less represented when metastases grow bigger.

To overcome these limitations, we developed a system whereby metastatic cancer cells directly mark their neighbouring cells by releasing a modified version of a secreted monomeric Cherry red fluorescent protein (mCherry) containing a lipo-permeable TATk peptide[2,3] (sLP-mCherry) (Figure 1a and Extended Data Figure 1a). We generated 4T1 breast cancer cells expressing the sLP-mCherry protein alongside a canonical cell-retained GFP, which we refer to as Labelling-4T1. *In vitro,* sLP-mCherry protein released by Labelling-4T1 is re-up-taken within producing cells as observed by changes in the intracellular localisation of the red fluorescence (Extended Data Figure 1b, c). Importantly, sLP-mCherry protein is also taken up by unlabelled cells both in co-culture (Figure 1b-d) and when cultured with Labelling-4T1 conditioned medium (LCM) (Extended Data Figure 1d, e). Upon uptake, sLP-mCherry fluorescence has an intracellular half-life of 43h (Extended Data Figure 1f) and is localized in intracellular vesicles (IVs) which are $CD63^+$ multi-lamellar bodies (lysosomal-like structures). Due to its high photostability[4], mCherry retains high fluorescent intensity (Extended Data Figure 1g, h). LCM fractionation shows that only the soluble fraction displays labelling activity, while the extracellular vesicles (EVs), a portion of which contains sLP-mCherry, do not show *in vitro* labelling activity (Extended Data Figure 1i-k). Critically, *in vivo*, Labelling-4T1 cells (GFP/Cherry double positive) intravenously injected into syngeneic BALB/c mice efficiently label their surrounding host tissue cells (Cherry single positive), with a penetration of approximately five cell layers (Figure 1e, f and Extended Data Figure 2a, b). This allows metastatic niche cells to be specifically discriminated from the distal lung (GFP/Cherry double negative) using fluorescent activated cell sorting (FACS) (Figure 1g). Notably, when metastases form, the number of $mCherry^+$ niche cells in the tissue remains proportional to the growing metastatic cells (Extended Data Figure 2c). We detected no immunogenicity against sLP-mCherry, since the local increase of $CD45^+$ immune cells within the mCherry population was independent from the adaptive immune system and was observed specifically as a response to cancer cells (Extended Data Figure 2d-f). Hence, the *Cherry-niche* marking system (Cherry-niche) enables the spatial reconstitution of the local metastatic niche within

the whole tissue. This allows functional identification of labelled cells (Cherry-niche cells) and their direct comparison with the remaining unlabelled tissue cells within the same lung.

We first aimed to validate Cherry-niche by interrogating the local changes of known cells involved in niche formation. We focused on the micro-metastatic stage, where metastases are small, but established and therefore likely to have settled a successful metastatic niche. $CD45^+$ immune cells are the most abundant component within Cherry-niche, which are nearly exclusively from the myeloid lineage ($CD11b^+$) (Extended Data Figure 2d and 3a). Lung neutrophils are widely reported to enhance metastatic growth of cancer cells[5,6], and were indeed detected within the Cherry-niche (Extended Data Figure 3b). We isolated Cherry-niche neutrophils ($Ly6G^+$) and compared their proteome to unlabelled neutrophils from the same lungs (Figure 2a). We found the sub-pool of niche neutrophils to have distinct features, with an increase in translational and oxidative phosphorylation activity (Figure 2b, Extended Data Figure 3c, d and Supplementary File 1). Higher levels of intracellular ROS in Cherry-niche neutrophils confirmed the increase in respiratory activity (Extended Data Figure 3e, f). To validate the functional relevance of specific niche cells identified using the labelling tool for cancer growth, we employed a co-culture system mimicking complex tissue-like cell-cell interactions whereby primary cells are seeded on a three-dimensional (3D) scaffold (Alvetex). When monitoring the growth of primary cancer cells isolated from the actin-GFP Mouse Mammary Tumour Virus (MMTV) Polyoma virus Middle T antigen (PyMT) breast tumour model, we found lung neutrophils to boost cancer cell growth in a ROS dependent manner (Figure 2c-e and Extended Data Figure 3g, h). Collectively, these data highlight the potential of Cherry-niche to detect *in vivo* changes spatially restricted to the cancer metastatic environment.

Cherry-niche labelling during different stages of metastatic progression can additionally provide spatio-temporal information. To gain unbiased insights into the niche signals, we generated the gene expression profile of non-immune (CD45-ve) Cherry-niche cells at the time point directly preceding micro-metastases as well as at an advanced metastatic stage (Figure 2f, g). The majority of alterations were detected at the early stage, but additional changes subsequently discriminated the niche of macro-metastases (Figure 2h and Extended Data Figure 4a, b), confirming the evolution of the metastatic environment over time. MetaCore dataset enrichment highlighted changes in pathways related to proliferation, inflammation and tissue remodelling (Extended Data Figure 4c). We next focused on the upregulated (>2) genes encoding for soluble factors in the niche versus the unlabelled cells from the same lungs at both time points (Figure 2i). Again, as validation of the ability of our labelling system to faithfully capture the *in vivo* niche, we could find many of previously reported tumour promoting factors (Figure 2i)[7-14]. We also found Wnt1 induced protein (Wisp1), previously suggested to act as oncogene in breast cancer[15], to be an abundant niche

factor (Figure 2i). Indeed, upregulation of Wisp1 in both cancer and niche cells was detected and its pro-metastatic activity was confirmed by exogenous inhibition *in vivo* (Figure 2j and Extended Data Figure 5).

We next probed for the presence of previously uncharacterized niche cells, which are difficult to resolve by standard techniques due to a lower frequency. Interestingly, we found pathways associated with lung epithelial cells in the metastatic niche signature (Figure 2k). Micro-metastases grow embedded within the alveolar compartment of the lung, and alveolar type II cells (AT2) expressing Surfactant protein C (SP-C) were found in the metastatic environment (Figure 3a). Using the epithelial cell adhesion molecule (Epcam) marker, we found Cherry-niche epithelial cells predominantly as low/mid-Epcam epithelial cells, showing higher proliferative activity (Figure 3b-d). Intriguingly, at the metastatic borders of human breast cancer lung metastases we found the presence of alveolar cell clusters with increased proliferative activity, suggesting that a lung parenchyma response to metastatic growth may occur in both human and mouse (Extended Data Figure 6). Cancer cells profit from this response, as freshly isolated Epcam$^+$ cells from naïve lungs supported the growth of MMTV-PyMT/GFP$^+$ tumour cells in our 3D scaffold co-culture system (Figure 3e-g). Furthermore, in line with previous data, the presence of both lung neutrophils and epithelial cells further enhanced tumour growth (Extended Data Figure 7a-d).

We then focused on characterization of the perturbed lung epithelial cells in the niche. Firstly, despite metastases growing within the alveolar region, we found a reduction in expression of alveolar lineage markers in Cherry-niche epithelial cells (Figure 3a and 3h). To contextualize their presence among the other niche cellular components, we performed single cell RNA sequencing of CD45-ve cells. tSNE analysis of Cherry-niche cells identified the presence of a large stromal cluster, where different cancer associated fibroblasts (CAFs) subsets[16] can be identified (Figure 3i and Extended Data Figure 8). Notably, specifically in the niche, Epcam expressing cells are distributed in two clusters distinguished by the expression of E-Cadherin (Cdh1) (Fig 3i, j). We found that only niche Epcam$^+$Cdh1$^+$ cells share the expression of alveolar genes[17] with unlabelled lung Epcam$^+$ cells (Fig 3k). Conversely, niche Epcam$^+$Cdh1$^-$ cells express the Sca1 (Ly6a) and Tm4sf1 progenitor markers[18-20] (Fig 3k). The enrichment of Epcam$^+$Sca1$^+$ cells was confirmed by FACS in the lung Cherry-niche of different metastatic cell types (Figure 3l and Extended Data Figure 9a-c). Similarly, the presence of epithelial cells expressing another lung progenitor marker, integrin β4 (CD104)[21], was increased in the niche as well as in *ex vivo* co-cultures (Figure 3e and Extended Data Figure 9d-i).

Parenchymal cells have been previously described to trigger a tissue-wide inflammatory response to systemic primary tumour signals, thereby increasing susceptibility to metastasis[22,23]. Conversely, here we provide evidence of a direct local parenchymal response

to invading cancer cells, whereby tissue epithelial cells become part of the metastatic environment. Thus, we define them as cancer associated parenchymal cells (CAPs).

To functionally characterize CAP cells, we tested their lineage differentiation potential *ex vivo* using a 3D Matrigel-based organoid co-culture system[19] (Figure 4a). Unlabelled resident lung Epcam$^+$ cells largely contain alveolar cells, as previously shown[19], and formed mainly alveolar organoids in co-culture with CD31$^+$ cells (Figure 4b-d). Strikingly, Cherry-niche Epcam$^+$ cells favoured the bronchiolar lineage and showed a remarkable capacity to generate multilineage bronchioalveolar organoids (Figure 4b-d). However, there were no visible signs of Cherry labelling in bronchial lineages *in vivo* in micro-metastasis (Extended Data Figure 10a). CAPs also retained high self-renewal capacity over multiple passages (Figure 4e).

Next, we tested whether tumour cells could directly induce CAPs phenotype. Epcam$^+$ cells from either unlabelled distal micro-metastatic lungs or naïve lungs were co-cultured with metastatic cells, generating a higher proportion of bronchiolar and bronchioalveolar organoids (Figure 4f-h and Extended Data Figure 10b, c). Similar alterations were induced when co-cultured with mouse lung fibroblasts (MLg) (Extended Data Figure 10b, c).

Certainly, lung Epcam$^+$ cells are predominantly alveolar, but they contain other epithelial progenitors that could be enriched by cancer cells to generate an increased plasticity[19,24]. Therefore, we performed organoid cultures using lineage-labelled AT2 cells (Sftpc-lineage). Remarkably, AT2 lineaged cells, which show no *ex vivo* plasticity in co-culture with CD31$^+$ cells, when exposed to cancer cells generated a remarkable amount of multilineage bronchioalveolar organoids, supporting the idea of a reprogramming activity of cancer cell-derived factors in this *ex vivo* assay (Figure 4i ,j). Importantly, metastases growing *in vivo* in Sftpc-lineage mice demonstrated the alveolar origin of the epithelial cells in the niche (Figure 4k). Conversely, despite the potential of cancer cells to modulate the organoid formation ability of lineage-labelled club cells (Scgb1a1-lineage), only rare single Scgb1a1-lineage cells were found in proximity to lung metastases (Extended Data Figure 10d, e).

In summary, this study proposes a novel labelling system that enables cellular resolution of the metastatic microenvironment. We provide examples of its potential to characterise host tissue cell perturbations which are spatially restricted to regions surrounding cancer cells. Remarkably, we report the presence of lung epithelial cells (CAPs) within the metastatic niche showing striking tissue stem cell-like features, multilineage differentiation potential and increased self-renewal activity. Taken together, these results show the power of Cherry-niche to identify, isolate and functionally test tissue cells in the metastatic niche with unprecedented spatial resolution.

ACKNOLEDGMENTS

We thank E. Sahai, P. Scaffidi (The Francis Crick Institute, London) and V. Sanz-Moreno (Barts Cancer Institute, London) for scientific discussions, critical reading of the manuscript and sharing cell lines and mouse strains. We thank M. Izquierdo (CSIC, Madrid) for sharing the CD63-GFP plasmid. We are grateful to E. Nye from the Experimental Histopathology Unit and J. Bee from the Biological Resources Unit at the Francis Crick Institute for technical support with mice and mouse tissues. We are also grateful to Robert Goldstone and Amelia Edwards from the Advanced Sequencing Facility at the Francis Crick Institute for technical support. We thank Miriam Llorian-Sopena from the Bioinformatics & Biostatistics Unit at the Francis Crick Institute for helping with the RNA sequencing analysis. We also thank the Flow Cytometry Unit at the Francis Crick Institute, particularly S. Purewal and J. Cerveira, for invaluable technical help, and Cell Services Unit at the Francis Crick Institute. We are also grateful to Irina Pshenichnaya, Peter Humphreys, Simon McCallum and Cambridge Stem Cell Institute core facilities for technical assistance. This work was supported by the Francis Crick Institute which receives its core funding from Cancer Research UK (FC001112), the UK Medical Research Council (FC001112), and the Wellcome Trust (FC001112) and the European Research Council grant (ERC CoG-H2020-725492); and by the Wellcome Trust – MRC Stem Cell Institute which receives funding from the Sir Henry Dale Fellowship from Wellcome and the Royal Society (107633/Z/15/Z) and the European Research Council Starting Grant (679411).

AUTHOR CONTRIBUTIONS

L.O. designed and performed most of the experiments, analysed and interpreted the data and contributed to the manuscript preparation. E.N. assisted with data collection, performed all the 3D-scaffold co-culture experiments, the *in vivo* Wisp1 experiments, the scRNA sequencing, interpreted and analysed the data and contributed to the manuscript preparation. I.K. performed the qPCR analysis, some of the tissue IF staining and analysed the data. A.M. and J.H.L. performed some of the tissue IF staining, lung organoid experiments, interpreted and analysed the data. V.B. performed some of the tissue IF staining. P.C. and S. H. performed bioinformatics analysis. I.H., J.K. and A.O. performed the proteomic and analysed the data. E.G.G. helped with the collection of $Ly6G^+$ cells for proteomics. G.M. performed the 3D-scaffold co-culture to analyse $CD104^+$ cells. A.W. and L.C. performed the electron microscopy experiments. E. H. and V. S. provided human samples. L.O., E.N., I.K., V.B. and J.H.L., critically reviewed the manuscript. J.H.L., supervised the lung organoid experiments. I.M. designed and supervised the study, interpreted the data and wrote the manuscript.

Competing interests

The authors declare no competing financial interests.

Materials & Correspondence

Correspondence to Ilaria Malanchi and Joo-Hyeon Lee.

Supplementary material files

1. **Supplementary file 1**: entire list of differentially detected proteins (AVG Log2 Ratio) in Cherry-niche versus unlabelled Ly6G^{+} neutrophils. Differences were considered when changes were >0.58 and <-0.58.
2. **Supplementary file 2**: upregulated genes (>2 fold) in non-immune-Cherry-niche compared to unlabelled lung tissue at both 5 and 10 days post cancer cells seeding.
3. **Supplementary file 3**: FACS gating strategy examples.

References

1. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nature Medicine* **19,** 1423–1437 (2013).
2. Barash, S., Wang, W. & Shi, Y. Human secretory signal peptide description by hidden Markov model and generation of a strong artificial signal peptide for secreted protein expression. *Biochemical and Biophysical Research Communications* **294,** 835–842 (2002).
3. Flinterman, M. *et al.* Delivery of Therapeutic Proteins as Secretable TAT Fusion Products. *Mol Ther* **17,** 334–342 (2009).
4. Shaner, N. C., Steinbach, P. A. & Tsien, R. Y. A guide to choosing fluorescent proteins. *Nat Meth* **2,** 905–909 (2005).
5. Wculek, S. K. & Malanchi, I. Neutrophils support lung colonization of metastasis-initiating breast cancer cells. *Nature* **17,** 413–417 (2015).
6. Coffelt, S. B., Wellenstein, M. D. & de Visser, K. E. Neutrophils in cancer: neutral no more. *Nature review cancer* **16,** 431–446 (2016).
7. Kessenbrock, K., Plaks, V. & Werb, Z. Matrix Metalloproteinases: Regulators of the Tumor Microenvironment. *Cell* **141,** 52–67 (2010).
8. Kowanetz, M. *et al.* Granulocyte-colony stimulating factor promotes lung metastasis through mobilization of Ly6G+Ly6C+ granulocytes. *Proc. Natl. Acad. Sci. U.S.A.* **107,** 21248–21255 (2010).
9. Qian, B.-Z. *et al.* CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. *Nature review cancer* **475,** 222–225 (2011).
10. Acharyya, S. *et al.* A CXCL1 Paracrine Network Links Cancer Chemoresistance and Metastasis. *Cell* **150,** 165–178 (2012).
11. Oskarsson, T. *et al.* Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nature Medicine* **17,** 867–874 (2011).
12. Erez, N. Opening LOX to metastasis. *Nature* **522,** 41–42 (2015).
13. Onnis, B., Fer, N., Rapisarda, A., Perez, V. S. & Melillo, G. Autocrine production of IL-11 mediates tumorigenicity in hypoxic cancer cells. *J. Clin. Invest.* **123,** 1615–1629 (2013).
14. Malanchi, I. *et al.* Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* **481,** 85–89 (2011).
15. Su, F., Overholtzer, M., Besser, D. & Levine, A. J. WISP-1 attenuates p53-mediated apoptosis in response to DNA damage through activation of the Akt kinase. *Genes Dev.* **16,** 46–57 (2002).
16. Costa, A. *et al.* Fibroblast Heterogeneity and Immunosuppressive Environment in Human Breast Cancer. *CCELL* **33,** 463–479.e10 (2018).
17. Treutlein, B. *et al.* Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. *Nature* **509,** 371–375 (2014).
18. Kim, C. F. B. *et al.* Identification of Bronchioalveolar Stem Cells in Normal Lung and Lung Cancer. *Cell* **121,** 823–835 (2005).
19. Lee, J.-H. *et al.* Lung Stem Cell Differentiation in Mice Directed by Endothelial Cells via a BMP4-NFATc1-Thrombospondin-1 Axis. *Cell* **156,** 440–455 (2014).
20. Zacharias, W. J. *et al.* Regeneration of the lung alveolus by an evolutionarily conserved epithelial progenitor. *Nature* **555,** 251–255 (2018).
21. Chapman, H. A. *et al.* Integrin α6β4 identifies an adult distal lung epithelial population with regenerative potential in mice. *J. Clin. Invest.* **121,** 2855–2862 (2011).
22. Liu, Y. *et al.* Tumor Exosomal RNAs Promote Lung Pre-metastatic Niche Formation by Activating Alveolar Epithelial TLR3 to Recruit Neutrophils. *CCELL* **30,** 243–256 (2016).
23. Lee, J. W. *et al.* Hepatocytes direct the formation of a pro-metastatic niche in the liver. *Nature* **567,** 249–252 (2019).
24. McQualter, J. L., Yuen, K., Williams, B. & Bertoncello, I. Evidence of an epithelial stem/progenitor cell hierarchy in the adult mouse lung. *Proc Natl Acad Sci USA* **107,** 1414–1419 (2010).

Figures and Legends

Figure 1

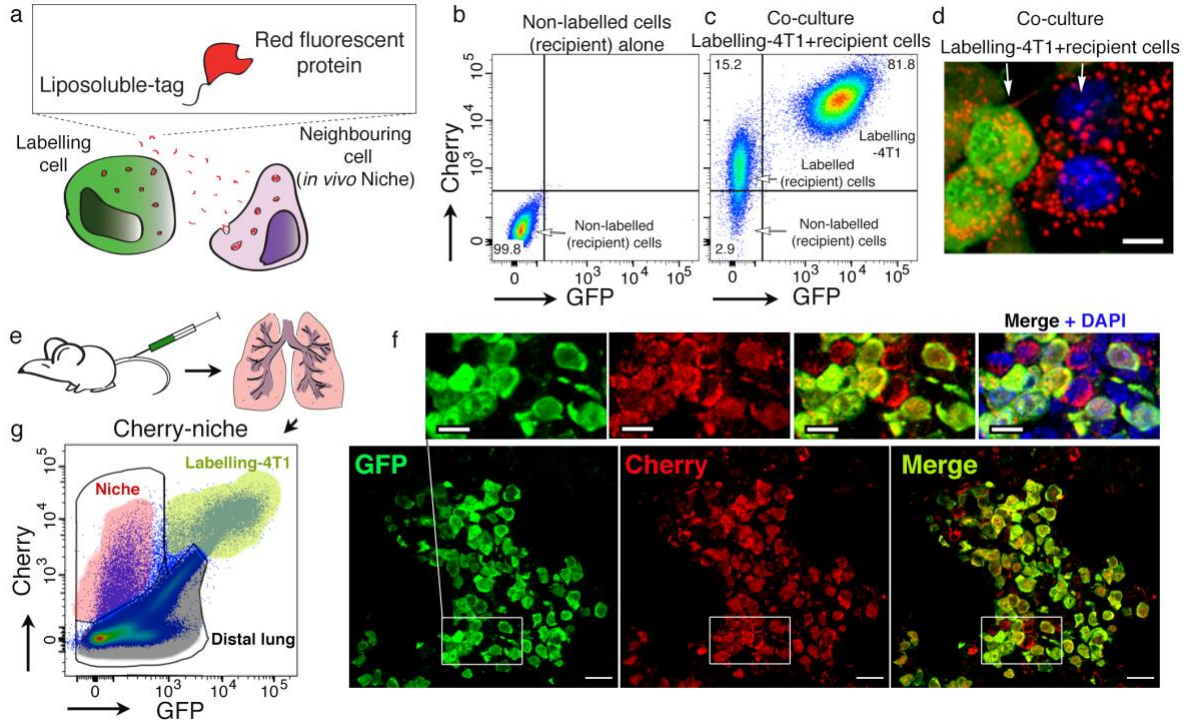


Figure 1 - Cherry-niche labelling strategy.

a, Labelling design. **b-d,** FACS plots of (b) non-labelled (recipient) 4T1 cells alone or (c) co-cultured with Labelling-4T1 and (d) fluorescence image from co-culture (Scale bar 10μm). **e-g,** *In vivo* labelling: (e) schematic experimental design; (f) representative immuno-fluorescence (IF) images of Labelling-4T1 metastasis: cancer cells GFP (green) and Cherry (red), niche cells (Cherry only). DAPI (blue). Scale bars: main 20μm, inset 10μm; (g) representative FACS plot of a dissociated metastatic lung.

Figure 2

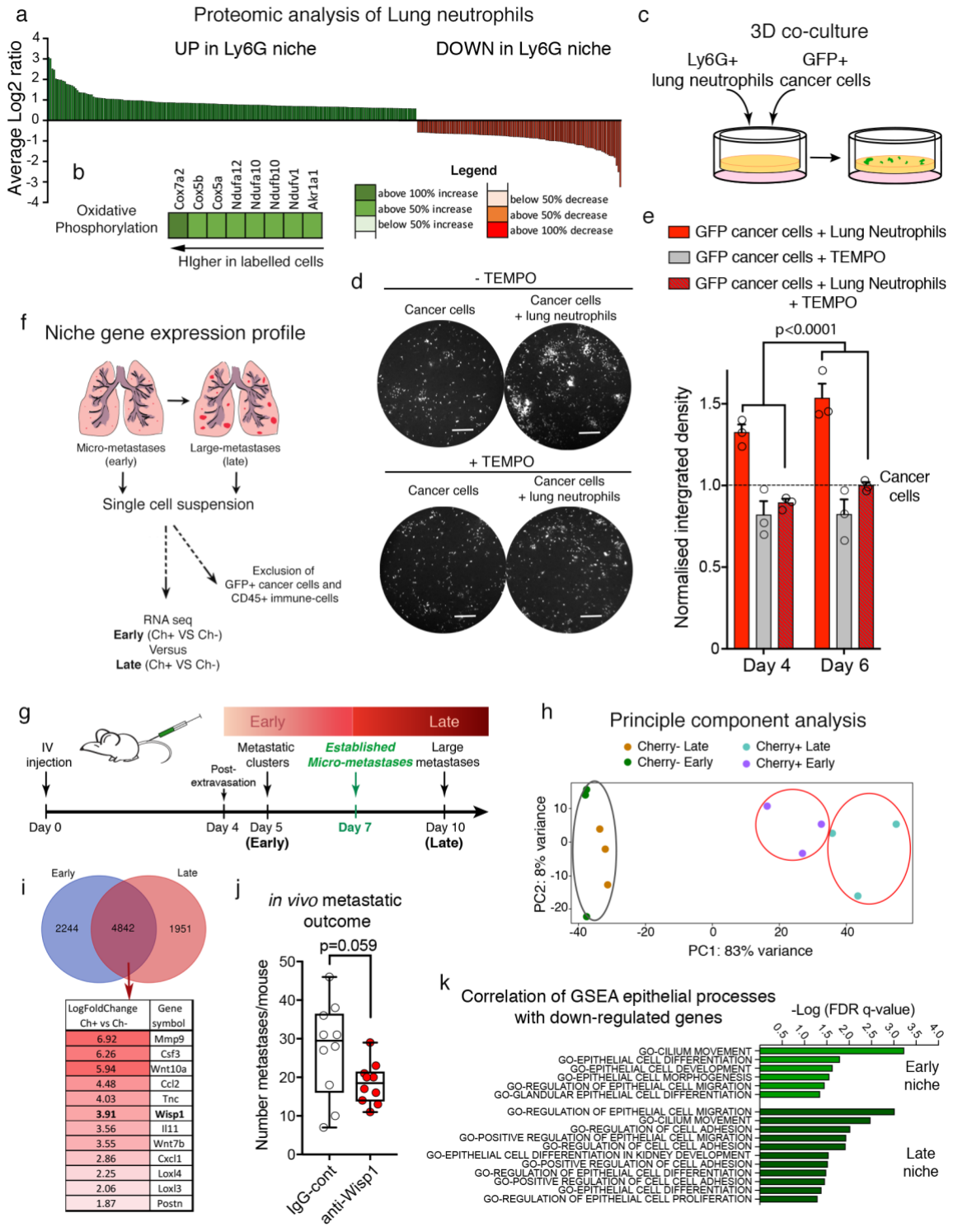


Figure 2 - Cherry-niche allows detection of niche cells and identification of niche factors.

a,b, Proteomic analysis of Ly6G$^+$ FACS-sorted cells: (a) all differentially detected proteins and (b) heatmap of oxidative phosphorylation associated proteins. **c-e,** 3D co-culture with MMTV-PyMT-GFP$^+$ cancer cells, Ly6G$^+$ MACS-sorted cells with or without the ROS inhibitor TEMPO:

(c) co-culture scheme; (d) representative pictures at day 6 (scale bar 400μm); (e) GFP signal quantification. Data normalised to cancer cell growth. Data show three independent experiments (dots), each performed with 3 to 10 technical replicates. Data are represented as mean ±SEM. **f,** Schematic experimental design for RNA-seq. **g,** Schematic of metastatic progression using 4T1 cells. **h,** Principle Component Analysis (PCA) diagram of $CD45^{-}Ter119^{-}$ cell signatures. **i,** Venn diagram of differentially expressed genes in Cherry-niche from RNA-seq and selected factors common at early and late stages. **j,** Anti-Wisp1 blocking antibody treatment *in vivo* (n=10 from two independent experiments; each dot represents one animal on a Tukey plot). A third experiment with higher overall metastatic frequency is quantified in Extended Data Figure 5d, e. **k,** GSEA correlation from RNA-seq data. Statistical analysis by Two-way ANOVA (e) and unpaired two-tailed t-test with Welch's correction (j).

Figure 3

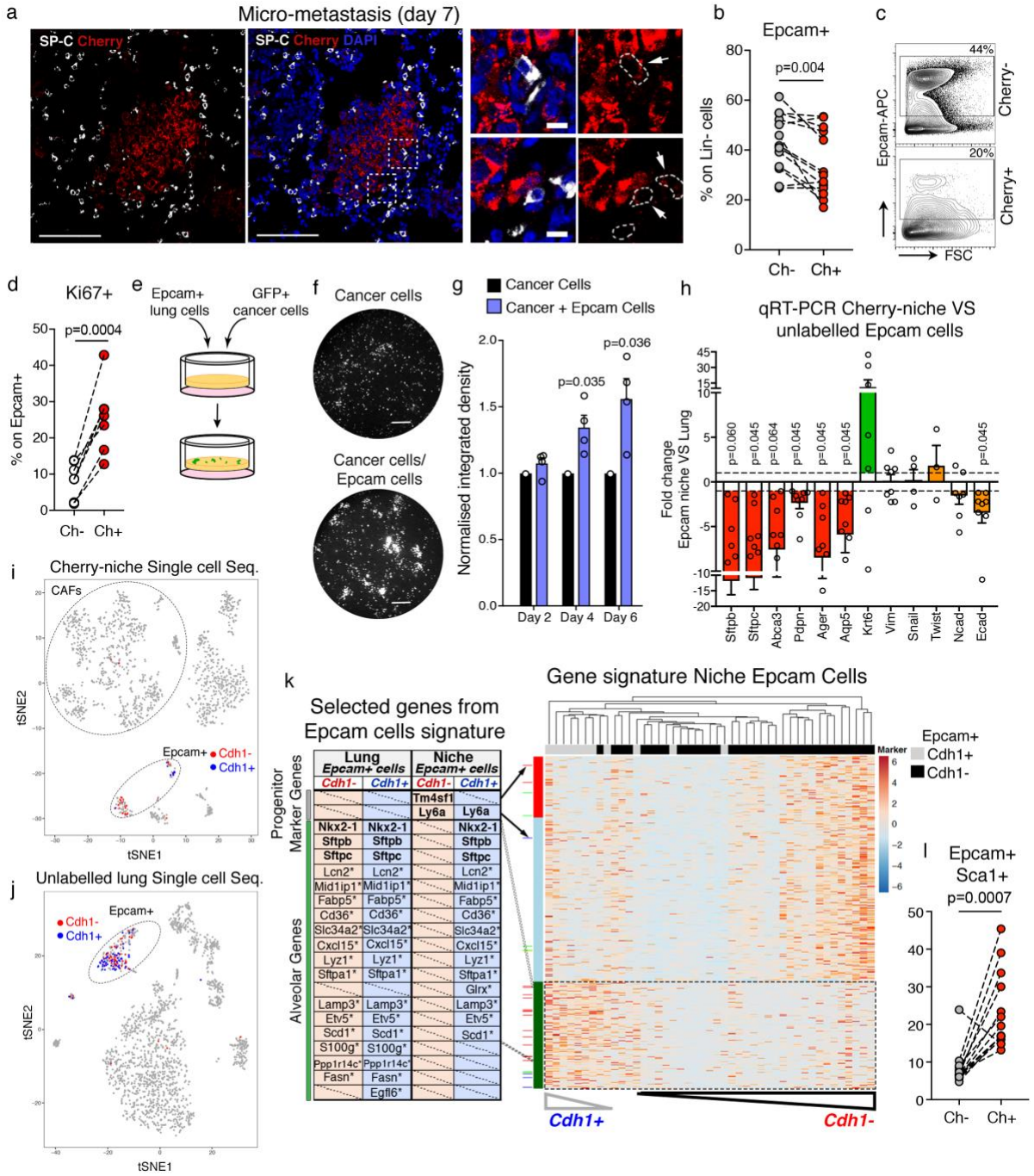


Figure 3 - Cherry-niche reveals lung epithelial cells with progenitor phenotype in the metastatic niche.

a, Representative IF image of lung tissue showing: mCherry-labelled micro-metastasis (red), Surfactant protein C (SP-C) (white) and DAPI (blue). Scale bars: main 100μm, inset 10μm (white arrows show mCherry labelled SP-C$^+$ cells). **b,** Epcam$^+$ cell frequency on Lin- (CD45$^-$ CD31$^-$Ter119$^-$) cells in distal lung (Ch$^-$) and Cherry-niche (Ch$^+$) by FACS (n=13). **c,** Representative FACS plots from (b). **d,** Scatter plot of Epcam$^+$ cell proliferation tested by Ki67 staining on FACS-sorted cells (n=6 from independent sorts). **e-g,** MMTV-PyMT-GFP$^+$ cancer

cell growth in 3D co-culture with MACS-sorted Epcam$^+$ cells: (e) co-culture scheme, (f) representative pictures at day 6 (scale bar 400μm), (g) GFP signal quantification; 3-4 technical replicates per experiment; four independent sorts (dots); data normalised to cancer cell growth. **h,** qRT-PCR analysis of Epcam$^+$ FACS-sorted cells (three samples from independent sorts, all technical replicates shown). Data normalised to Ch$^-$ Lung Epcam cells. **i, j**, tSNE plots of CD45$^-$ cells from (i) Cherry-niche or (j) distal lung after scRNA-seq analysis. **k,** Heatmap showing differences in gene expression of niche Epcam$^+$ cells (Cdh1$^+$ or Cdh1$^-$); genes hierarchically clustered in rows and cells in columns; inset table shows established lineage markers (bold) and putative alveolar markers[17] (*). **l,** Epcam$^+$Sca1$^+$ cell frequency on Lin- (CD45$^-$CD31$^-$Ter119$^-$) cells by FACS (n=13). Statistical analysis was performed by Wilcoxon matched-pairs signed rank test (l), paired two-tailed t-test (b, d), one sample t-test (g) with a Benjamini-Hochberg correction (h). Data represented as mean ±SEM.

Figure 4

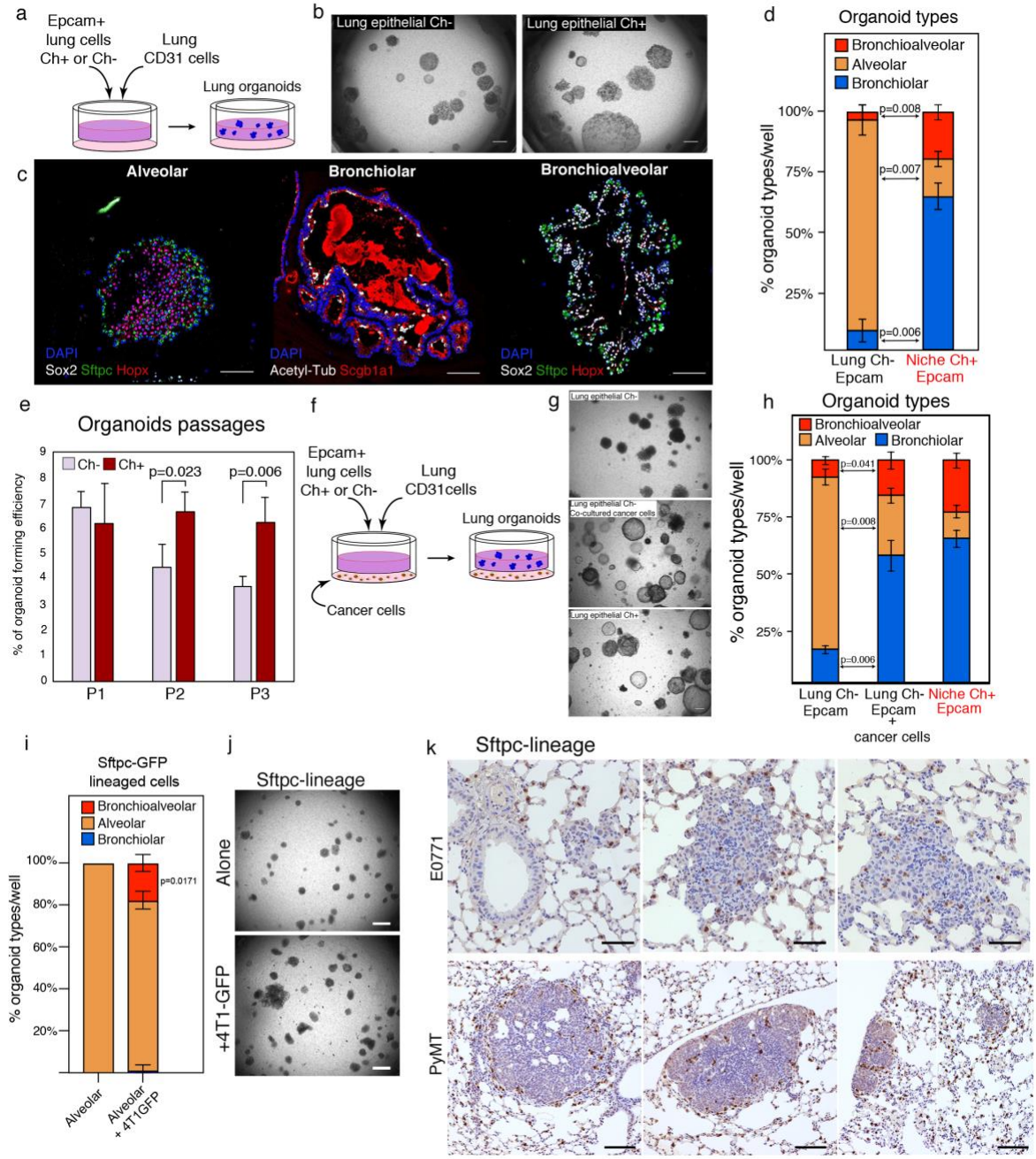


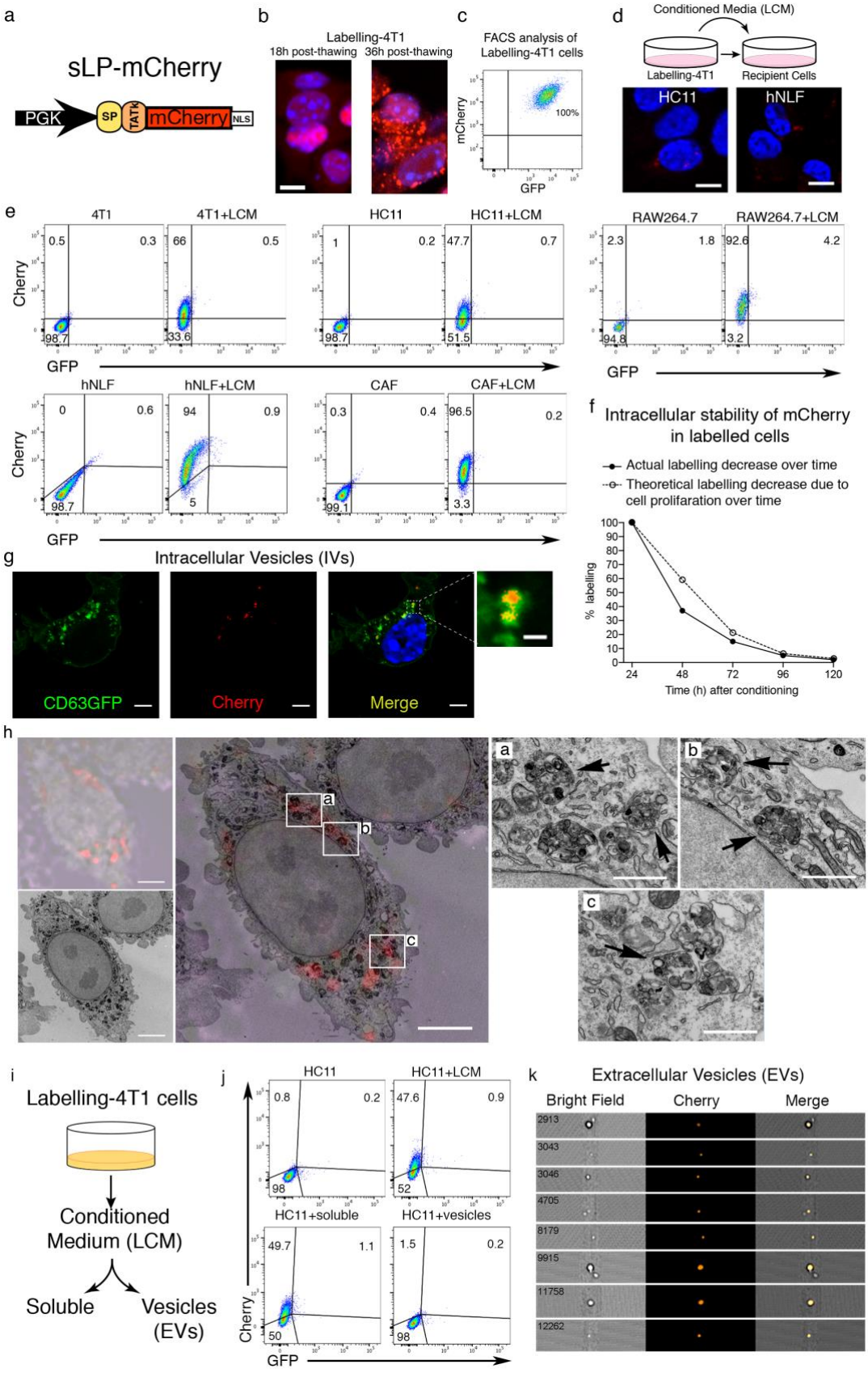
Figure 4 - Cancer associated parenchymal cells (CAPs) show multilineage differentiation potential.

a-e, Lung organoids: (a) co-culture scheme; (b) representative bright-field images (scale bar 100μm); (c) representative IF of organoid sections stained with the indicated markers (scale bar 50μm); (d) quantification; (e) organoid formation efficiency over passages. **f-h,** Lung organoids with or without Labelling-4T1: (f) co-culture scheme, (h) quantification and (g) representative bright-field pictures (scale bar 100μm). **i, j,** Lung organoids with Sftpc-CreERT2 lineage cells with or without 4T1-GFP: (i) quantification and (j) representative bright-field pictures, scale bar 150μm. **k**, Representative staining of lineage cells in metastatic lungs from

Sftpc-CreERT2 mice injected with cancer cells, either E0771 (scale bar 50μm) or MMTV-PyMT (scale bar 100μm). Data presented are the mean of two independent sorts with triplicate wells ±SD (d, e, h) and sorted AT2-lineaged cells co-cultured in triplicate wells ±SD (i). Statistical analysis was performed by two-tailed t-test (d, e, h) and one sample t-test (i).

Extended Data Figures and Legends

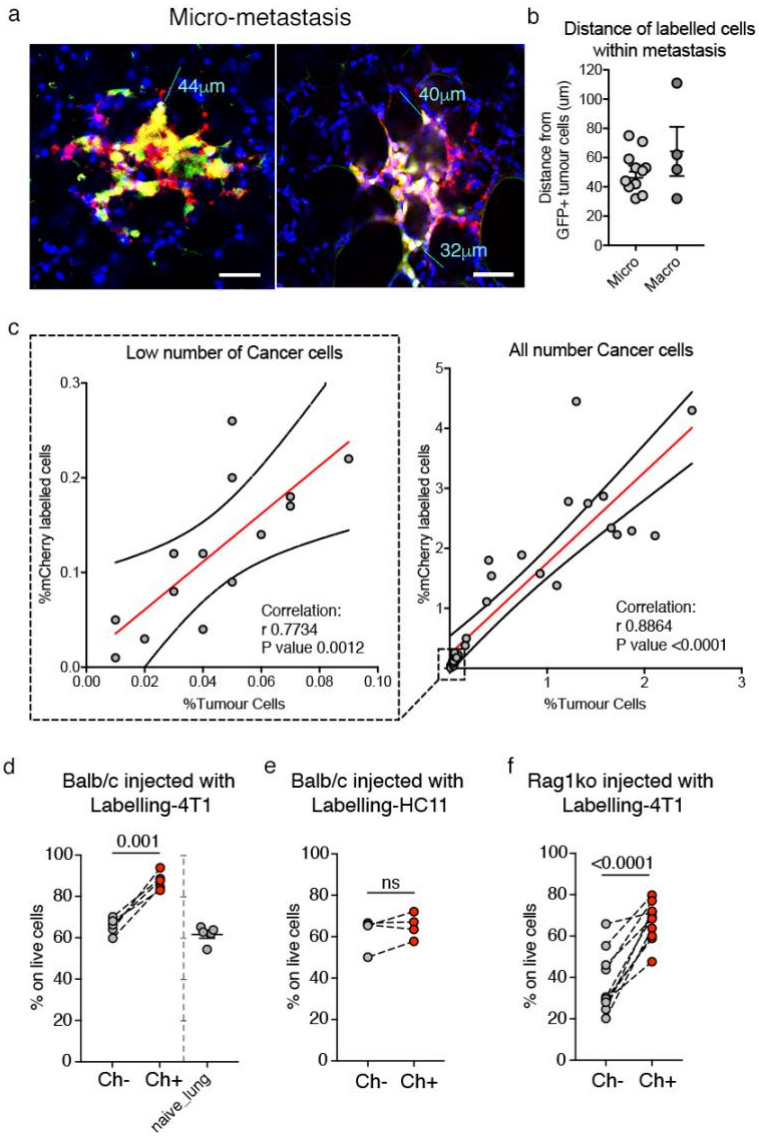
Extended Data Figure 1



Extended data Figure 1 - Cherry-niche *in vitro*

a, sLP-mCherry design. **b,** Fluorescence images of Labelling-4T1 cells post-thawing. Scale bar10μm. **c,** Representative FACS plot of Labelling-4T1 cells. **d,** *In vitro* cultures of the indicated cell types with Labelling-4T1 cell conditioned media (LCM): culture scheme and representative fluorescence images of HC11 (murine mammary epithelial cells) and hNLF (human normal lung fibroblasts) with LCM (scale bar 10μm). **e,** FACS plots of 4T1, HC11, RAW264.7 (murine macrophages), hNLF, murine breast Carcinoma Associated Fibroblasts (CAF) cultured with LCM. **f,** FACS analysis of 293T cells cultured with LCM, at different time-points after LCM removal (black dots); white dots show the theoretical decrease considering the cell proliferation rate only (the amount of 293T cells mCherry labelled after 24h incubation with LCM was set to 100%). **g,** Representative fluorescence image of 4T1-CD63GFP cells cultured with LCM. Scale bars: main 5μm, inset 1μm. **h,** Representative Correlative Light and Electron Microscopy (CLEM) of Labelling-4T1 cells re-up-taking sLP-mCherry: upper-left panel shows bright-field image overlaid mCherry IF (~700nm optical section); lower-left panel shows EM of same cell (~70nm section thickness); large central panel shows best approximation of IF/bright-field/EM overlay (scale bar 5μm); right panels show EM insets from indicated areas in the large central panel (black arrows point at vesicular structures containing the mCherry, scale bar 1μm). **i, j,** Analysis of *in vitro* labelling potential of soluble fraction and extracellular vesicles (EVs) isolated from LCM by FACS: (i) schematic representation of LCM fractionation; (j) HC11 cells cultured with either LCM, soluble fraction after EVs depletion (soluble) or purified EVs. **k,** ImageStream analysis of $Cherry^+$ EVs in LCM (16% of total EVs are $Cherry^+$).

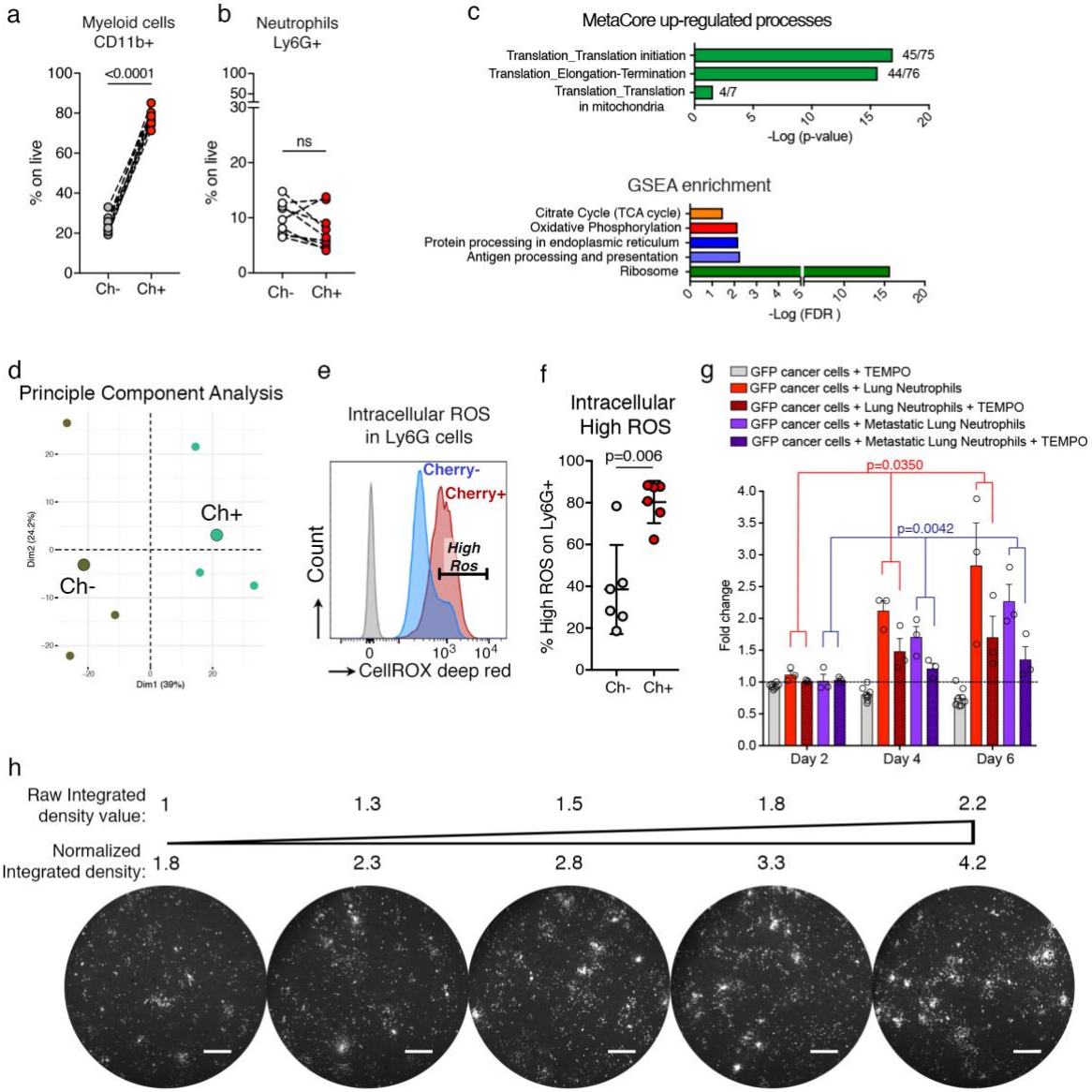
Extended Data Figure 2



Extended data Figure 2 - Cherry-niche *in vivo.*

a, b, Distance of labelled cells within metastases: (a) representative fluorescence images (lines measure the maximum distance of labelled cells (mCherry) from Labelling-4T1 cells (mCherry/GFP); scale bar 50μm); (b) quantification of labelling distance in micro-metastases and macro-metastases (n=11 micro-metastases and n=4 macro-metastases). **c,** Correlation between the percentage of mCherry labelled niche cells and the percentage of cancer cells in metastatic lungs analysed by FACS: (left) only low number of cancer cells (right) all cancer cell frequencies. Statistical analysis by linear regression. **d-f,** $CD45^+$ cell frequency on live cells in distal lung, Cherry-niche and naïve lungs (collected from mice which were not injected) by FACS: (d) Balb/c mice injected with Labelling-4T1 cells (n=5 per group); (e) Balb/c mice injected with Labelling-HC11 cells (n=4); (f) Rag1ko mice injected with Labelling-4T1 cells (n=10). Statistical analysis by paired two-tailed t-test. Data are represented as mean ±SEM.

Extended Data Figure 3

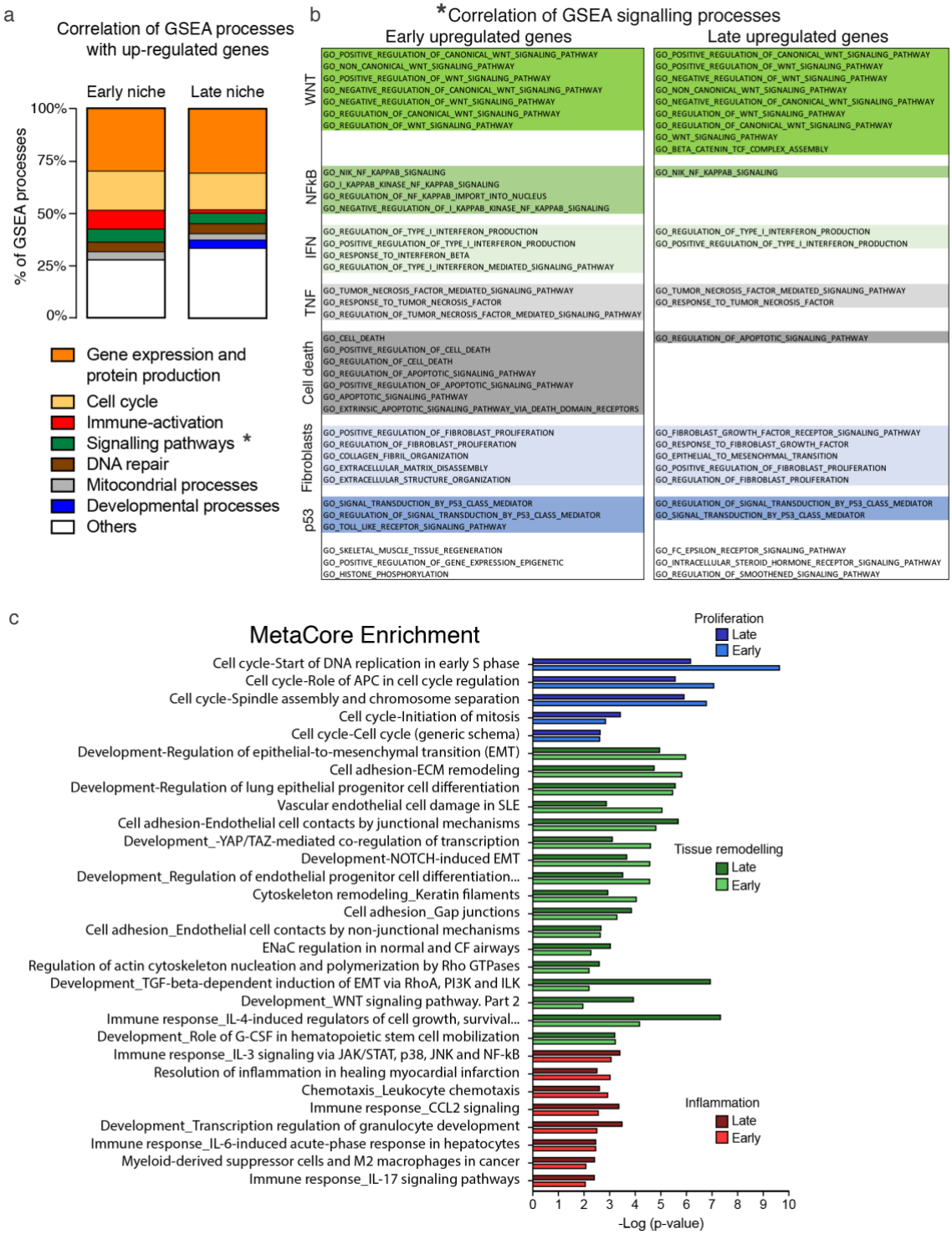


Extended data Figure 3 - Cherry-niche neutrophils increase ROS production.

a, b, (a) CD11b^{+} and (b) Ly6G^{+}cell frequencies on live cells in distal lung and Cherry-niche by FACS (n=9 per group). **c,** Enriched processes by MetaCore analysis and GSEA on Cherry-niche neutrophils dominant proteins using WebGestalt (http://www.webgestalt.org/option.php). **d,** PCA of proteins found in unlabelled or Cherry-niche neutrophils: small circles represent proteomic data from each independent sorts and large circles represent the average of the triplicates. **e,** Representative FACS plot and **f,** scatter plot of intrinsic ROS in Ly6G cells (n=6). **g,** GFP signal quantification of 3D co-culture with MMTV-PyMT-GFP^{+} cancer cells, Ly6G^{+} MACS-sorted cells from either naïve or metastatic lungs with or without the ROS inhibitor TEMPO. Data is normalised to cancer cell

growth. Each dot represents an independent experiment performed in triplicate. **h,** Representative cancer cell growth on the scaffold: integrated density of the GFP signal measure on the scaffold using ImageJ and the corresponding fluorescent image of GFP$^+$ cancer cell growth (scale bar 400μm). Statistical analysis by paired two-tailed t-test (a, b, f) and Two-way ANOVA (g). Data represented as mean ±SEM.

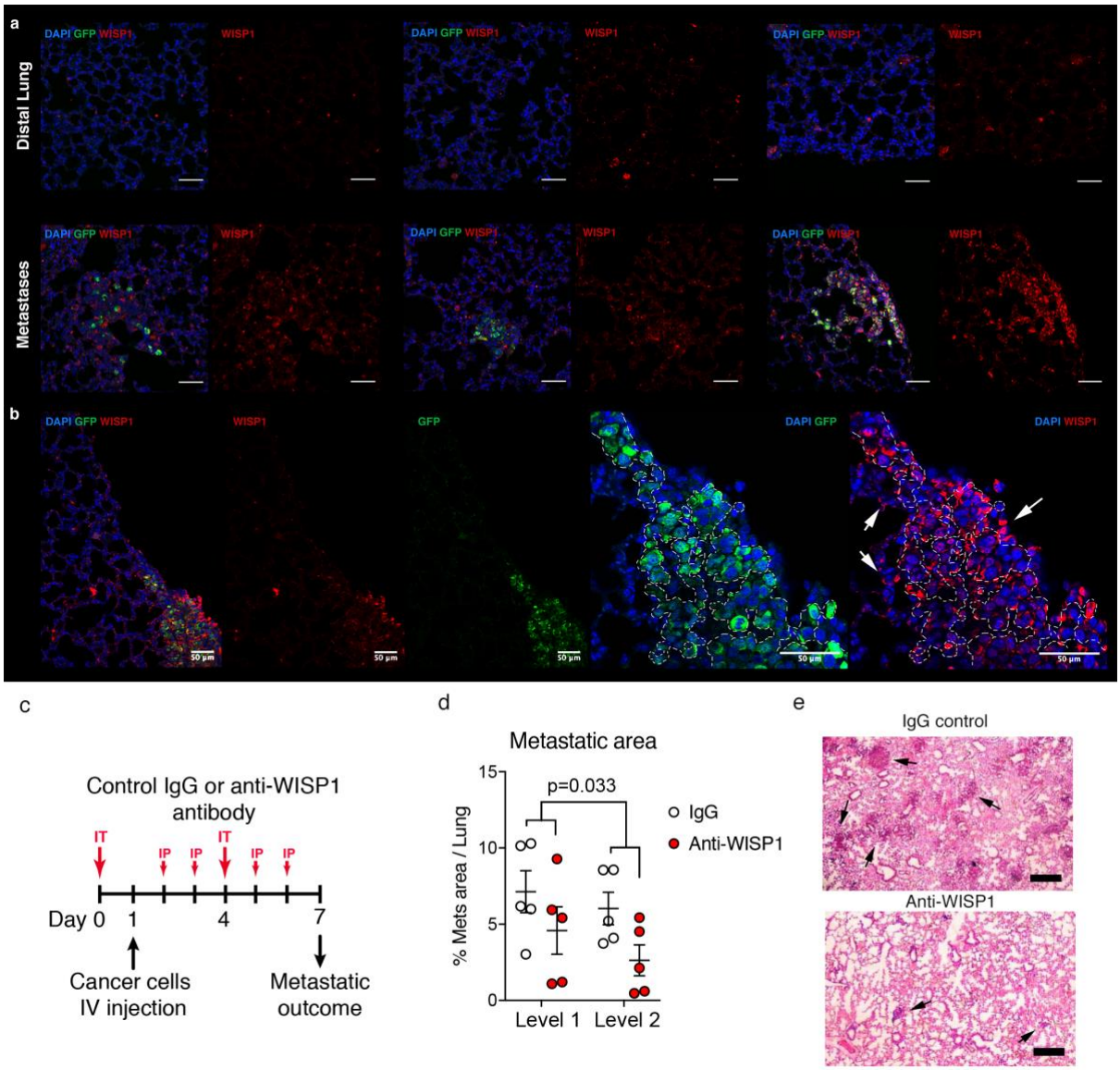
Extended Data Figure 4



Extended data Figure 4 - RNA sequencing of non-immune Cherry-niche cells.

a, b, GSEA on Cherry-niche upregulated genes: (a) percentage of correlating processes related to the indicated activity and (b) specific signalling pathways (indicated by the * in (a) either at early or late time point). **c,** MetaCore analysis on genes differentially expressed in Cherry-niche versus unlabelled lung cells by RNA-seq.

Extended data Figure 5

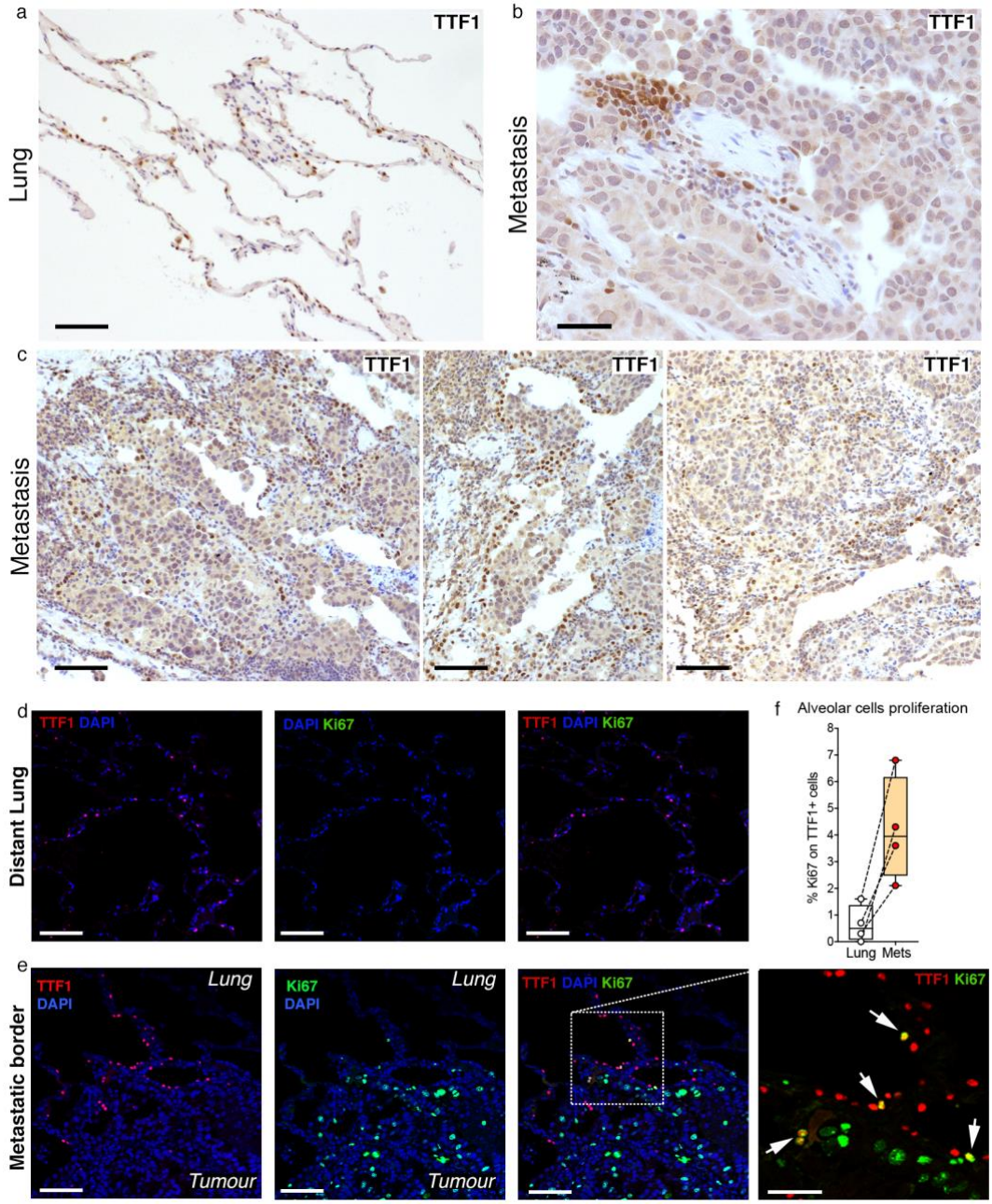


Extended data Figure 5 - Wisp1 supports metastatic growth.

a, b, Representative IF images of lung metastatic tissues stained with GFP (green) to detect Labelling-4T1 cells, WISP1 (red) and DAPI (blue) showing distal lung and metastatic areas, scale bar 50 μm; (b) a representative image showing the enrichment of $Wisp1^{+}$ cells within lung metastasis including niche cells (white arrows); scale bar 50μm. **c-e,** Anti-WISP1 blocking antibody treatment *in vivo:* (c) experimental design (IT, intratracheal injection; IP,

intraperitoneal injection); (d) metastatic outcome measured as the percentage of lung area covered by metastases (quantification was performed on two lung levels 100μm apart); (e) representative H&E pictures (black arrows show metastatic foci), scale bar 500μm.

Extended Data Figure 6

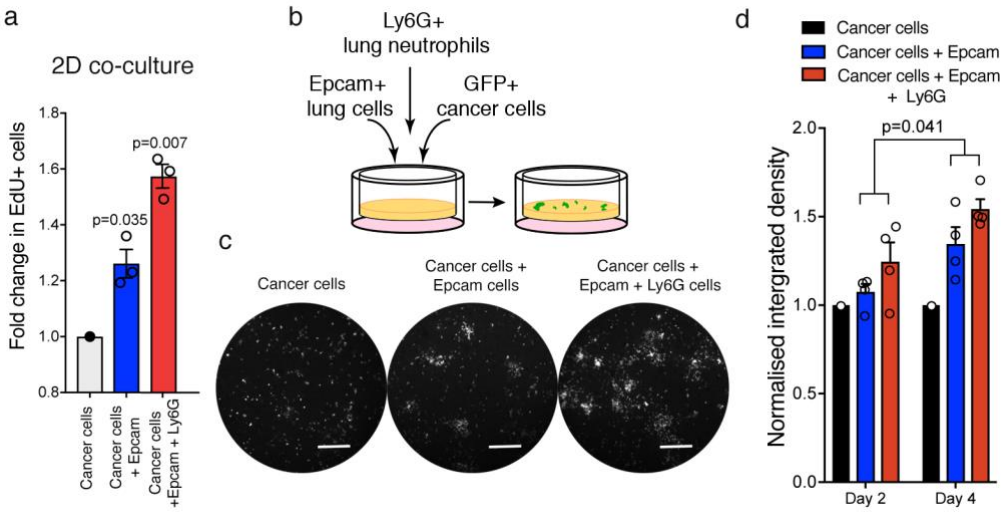


Extended data Figure 6 – Lung pneumocytes react to cancer cells in human breast pulmonary metastases.

a-c, Histology on human breast tumour lung metastases sections: (a) representative image of distal lung (scale bar 100μm); (b) image from the tumour-lung interface showing TTF1 cluster (scale bar 50μm); (c) representative histology images from metastatic border (scale bar 100μm). **d-f,** Alveolar cell proliferation in human breast tumour lung metastases analysed by

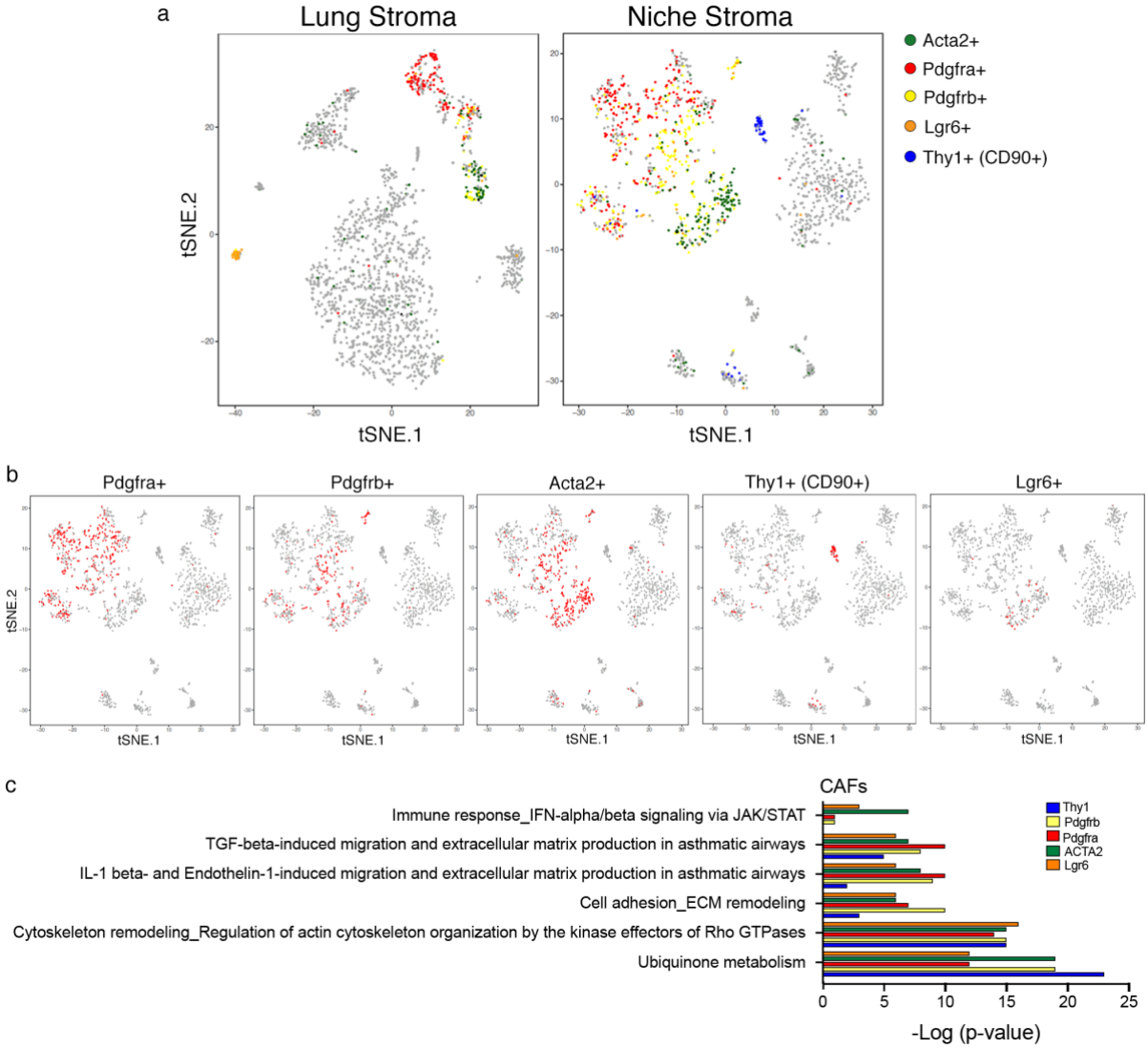
IF: representative pictures from (d) distal lung and (e) metastatic border showing TTF1 (red), Ki67 (green) and DAPI (blue), scale bars: main 100μm, inset 50μm; (f) quantification.

Extended Data Figure 7



Extended data Figure 7 - Epithelial cells support cancer cell growth *ex vivo*.
a, MMTV-PyMT-GFP$^+$ cancer cell proliferation in 2D co-culture with MACS-sorted Epcam$^+$ and Ly6G$^+$ cells stained with EdU and analysed by FACS (n=3 independent experiments). Data normalised to cancer cell proliferation. **b-d,** 3D co-culture of MMTV-PyMT-GFP$^+$ cancer cell with MACS-sorted Epcam$^+$ and Ly6G$^+$ cells: (b) co-culture scheme; (c) representative pictures at day 4, scale bar 400μm; (d) GFP signal quantification. Data normalised to cancer cell growth (Dots represent independent sorting experiments performed with 3-4 technical replicates). Statistical analysis by one sample t-test (a) and Two-way ANOVA (d). Data are represented as mean ±SEM.

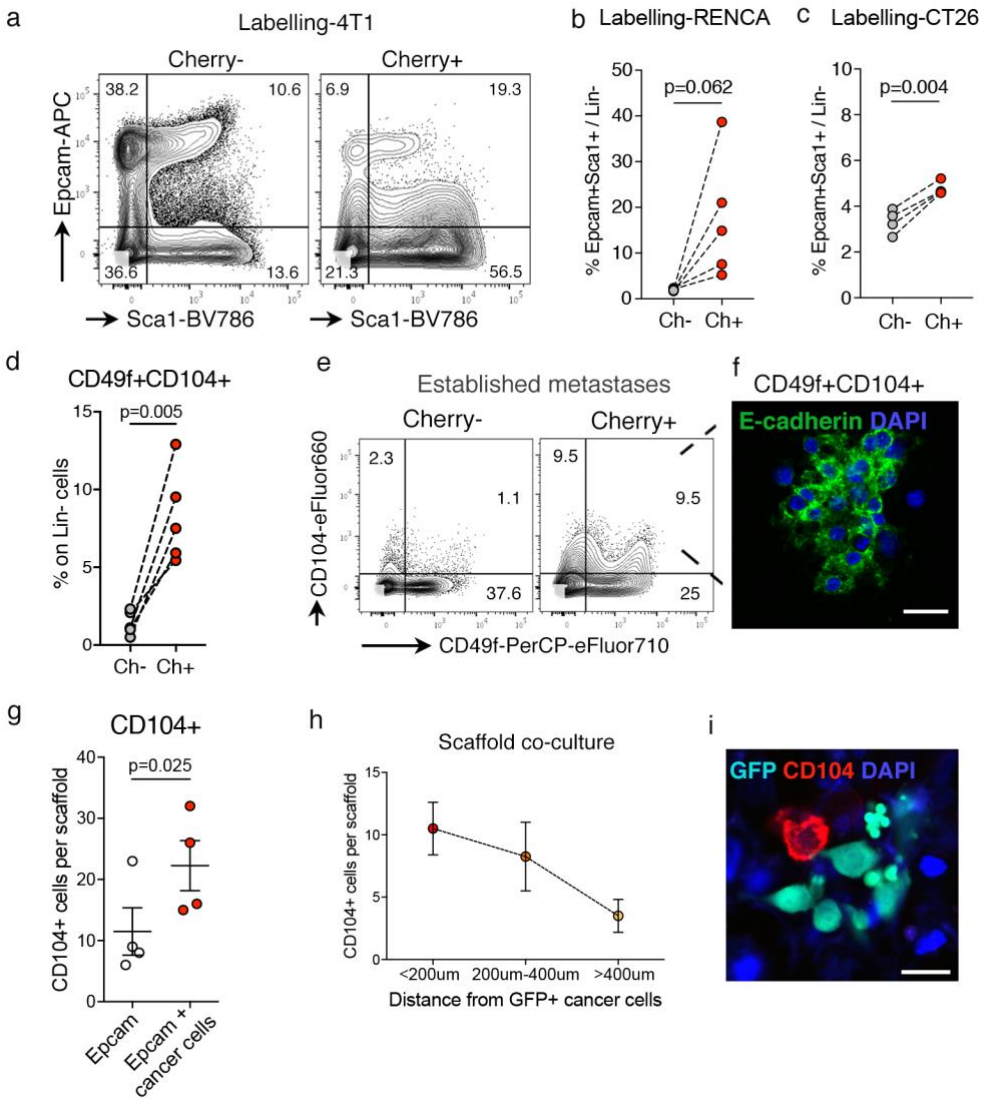
Extended Data Figure 8



Extended data Figure 8 - scRNA-seq analysis reveals different sub-pools of stromal cells in the niche.

a, tSNE plots of CD45^{-} cells isolated from distal lung or Cherry-niche after scRNA-seq analysis: the CAFs are coloured based on the expression levels of the indicated genes. **b,** tSNE plots split up from the niche plot in (a), where each plot shows in red the cells expressing the indicated stromal marker. **c,** MetaCore pathway enrichment analysis using the list of genes detected in at least 50% of the indicated marker defined cells.

Extended Data Figure 9

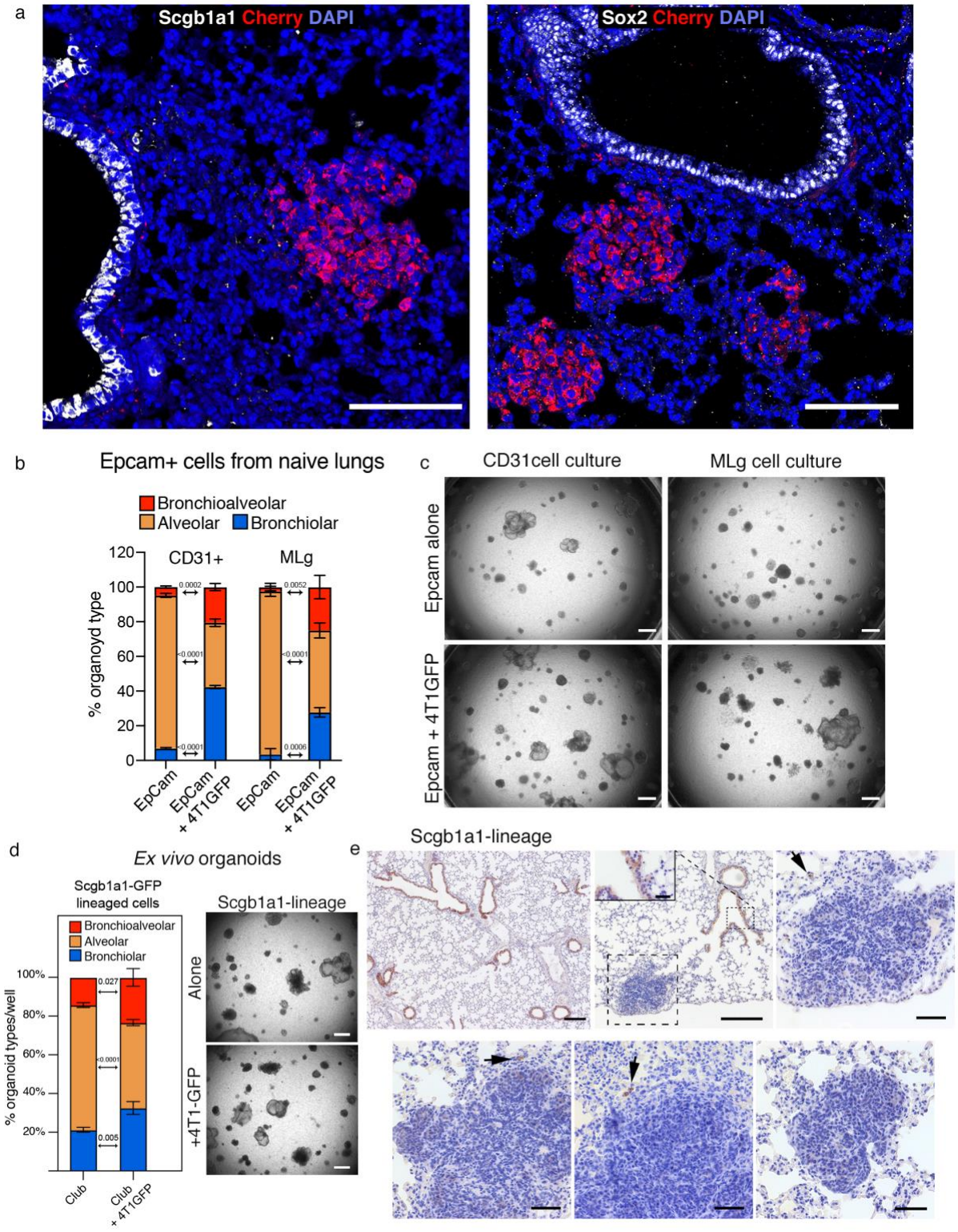


Extended data Figure 9 - Cherry-niche epithelial cells are enriched for stem cell markers.

a, Representative FACS plots showing Lin- ($CD45^-CD31^-Ter119^-$) cells in distal lung and Cherry-niche from Labelling-4T1 injected mice (quantification in Fig.3l). **b, c,** Scatter plots showing FACS quantification of $Epcam^+Sca1^+$ cell frequency on Lin- ($CD45^-CD31^-Ter119^-$) cells in distal lung and Cherry-niche with (b) Labelling-RENCA (n=5) and (c) Labelling-CT26 (n=4). **d-f,** (d) Scatter plot of $CD49f^+CD104^+$ cell frequency on Lin- ($CD45^-CD31^-Ter119^-$) cells in distal lung and Cherry-niche by FACS (n=5); (e) representative FACS plots; (f) representative IF image of FACS-sorted Cherry-niche $CD49f^+CD104^+$ cells using E-cadherin (green) and DAPI (blue); scale bar 20μm. **g-i,** 3D co-culture of MMTV-PyMT-GFP^+ cancer cell with MACS-sorted $Epcam^+$ cells: (g) quantification of integrin β4 (CD104) expression on $Epcam^+$ cells; (h) number of $CD104^+$ cells proximal to cancer cells (n=4 from three independent sorts); (i) representative IF image from the co-culture stained with CD104 (red);

GFP⁺ cancer cells (green) and DAPI (blue); scale bar 20μm. Statistical analysis by paired two-tailed t-test (b-d, g). Data represented as mean ±SEM.

Extended Data Figure 10



Extended data Figure 10 - Cancer cells change lung epithelial cell lineage commitment *ex-vivo*.

a, Representative IF images of lung metastatic sections co-stained with airway markers, either (a) Scgb1a1 (white) or (b) Sox2 (white), Cherry (red) and DAPI (blue); scale bar 100μm. **b, c,** Lung organoids with Epcam$^+$ FACS-sorted cells in co-culture with either lung stromal CD31$^+$ cells or MLg fibroblasts alone or in presence 4T1-GFP cells from metastatic lungs in the lower chamber: (b) quantification and (c) representative bright-field images of organoids, scale bar 150μm. **d,** Lung organoids with Scgb1a1-CreERT2 lineage cells with or without 4T1-GFP: quantification and representative bright-field pictures, scale bar 150μm. **e**, Representative staining of lineage cells in metastatic lungs from Scgb1a1-CreERT2 mice injected with MMTV-PyMT cancer cells. Scale bars: main 50μm, apart from the first 2 panels where is 200μm (inset 25μm). Sorted Epcam (b) or club-lineage cells (d) were co-cultured in triplicate wells ±SD. Statistical analysis was performed by two-tailed t-test.

Methods

Statistical Analysis

Statistical analyses were performed using Prism software (version 7.0c, GraphPad Software, USA) with the exception of the qRTPCR data, for which R was used. P values were obtained from two-tailed Student t-tests with paired or unpaired adjustment. When needed, unpaired t-test were adjusted using Welch's correction for unequal variance. In one instance (Fig. 3l) data in one of the groups did not pass D'Agostino & Pearson normality test, therefore a Wilcoxon matched-pairs signed rank test was performed. For qRTPCR data, single-sample tests with a Benjamini-Hochberg correction to account for multiple testing were performed. Single-sample tests were also used for comparisons of co-cultured cancer cell growth on scaffolds to the normalized value of cancer cells alone. For comparisons between two scaffold conditions of growth over time or to perform multiple analysis between experimental groups, Two-way ANOVA was used.

Data availability

The single cell RNA sequencing datasets are being deposited in the Gene Expression Omnibus (GEO, NCBI) repository. RNA sequencing datasets are deposited with GEO (GSE117930) and the proteomic datasets in PRoteomics IDEntifications (PRIDE) repository (PXD010597). All data are available upon request during the review process.

Mouse strains

All mice used are available from Jackson Laboratory. MMTV-PyMT mice[25] on FVB and C57BL/6 background, actin-GFP[26] mice and Rag1KO are on FVB background (gift from J. Huelsken laboratory (EPFL, Lausanne, Switzerland)). *Sftpc-CreERT2*[27], *Rosa26R-YFP*[28] (*Sftpc-CreERT2;R26R-YFP)* are on a C57BL/6 background. Balb/cj mice and the above-mentioned lines were bred and maintained under specific-pathogen-free conditions by The Francis Crick Biological Research Facility and female mice were used between 6 to 10 weeks of age. Breeding and all animal procedures were performed at the Francis Crick in accordance with UK Home Office regulations under project license P83B37B3C.

For *ex-vivo* organoids lineage tracing experiments, *Scgb1a1-CreERT2*[29], *Sftpc-CreERT2*[27], and *Rosa26R-fGFP*[29] (*Sftpc-CreERT2;R26R-fGFP* and *Scgb1a1-CreERT2;R26R-fGFP)* on a C57BL/6 background were bred and maintained under specific-pathogen-free conditions at the Gurdon Institute of University of Cambridge in accordance with UK Home Office project licence PC7F8AE82.

Tamoxifen administration

Tamoxifen (Merck Sigma-Aldrich, Germany) was dissolved in Mazola corn oil (Merck Sigma-Aldrich, Germany) in a 20mg/ml stock solution. Two doses of tamoxifen (0.2mg/g body weight) were given via oral gavage every other day and lung tissues were collected two days after tamoxifen administration to isolate cells for lung organoids. For *in vivo* lineage tracing three doses of tamoxifen (0.2mg/g body weight) were given via oral gavage over consecutive days and mice were injected two weeks later.

Cells

MLg cells (murine normal lung fibroblasts) were purchased from ATCC (USA). CAF (cancer associated fibroblasts) isolated from MMTV-PyMT tumours and human normal fibroblast (hNLF) were a gift from E.Sahai. All other cell lines were provided by the Cell Services Unit of The Francis Crick Institute. MMTV-PyMT cells were cultured on collagen solution coated dishes in MEM medium (DMEM/F12 (ThermoFisher Scientific, USA) with 2% fetal bovine serum (FBS; Labtech, UK), 100 U ml^{-1} penicillin-streptomycin (ThermoFisher Scientific, USA), 20 ng ml^{-1} EGF (ThermoFisher Scientific, USA) and 10 µg ml^{-1} insulin (Merck Sigma-Aldrich, Germany)). Collagen solution is made by 30 µg/mL PureCol collagen (Advanced Biomatrix, USA), 0.1% bovine serum albumin (BSA), 20 mM HEPES in HBSS (ThermoFisher Scientific, USA). HC11 cells were cultured in RPMI (ThermoFisher Scientific, USA) supplemented with 10% FBS, 100 U ml^{-1} penicillin-streptomycin, 10 ng ml^{-1} EGF (ThermoFisher Scientific, USA) and 5 µg ml^{-1} insulin. All other cell lines were cultured in DMEM (ThermoFisher Scientific, USA) supplemented with 10% FBS and 100 U ml^{-1} penicillin-streptomycin. All cells were cultured at 37ºC and 5% CO_2.

Human Samples

Human pulmonary breast cancer metastases from independent patients were obtained from the Grampian Biorepository, Aberdeen Royal Infirmary (REC approval: 16/NS/0055). Four samples were stained by Immunohistochemistry and Immunofluorescence and epithelial cells proliferation was quantified.

Labelling system

A soluble peptide (SP)[2] and a modified TAT peptide[3] were cloned upstream of the mCherry cDNA, under the control of a mouse PGK promoter (sLP-Cherry). The sLP-Cherry sequence was cloned into a pRRL lentiviral backbone. 4T1, Renca, CT26 and HC11 cells were stably infected with sLP-Cherry and pLentiGFP lentiviral particles and subsequently sorted to isolate $Cherry^{+}GFP^{+}$ cells.

Induction of experimental metastases

For experimental metastases, 4T1 (1,000,000 cells), Renca (500,000 cells), CT26 (200,000 cells) were re-suspended in 100 µl PBS and tail-vein injected in Balb/cJ mice.

***In vivo* lineage tracing experiments**

Sftpc-CreERT2 and *Scgb1a1-CreERT2* mice on C57BL/6 background were tail-vein injected either with 175,000 MMTV-PyMT C57BL/6 cells and lungs collected 4 weeks later or with 700,000 E0771 cells and lungs collected 12 days later.

Tissue digestion for cell isolation or analysis

Lung tissues were dissociated as previously described[14]. Briefly, lungs were removed at day 7 after tumour cell injection (unless otherwise specified), minced manually and then digested for 30 min in a shaker at 37ºC with a mixture of DNase I (Merck Sigma-Aldrich, Germany) and Liberase TM and TH (Roche Diagnostics, Switzerland) in HBSS solution. Samples were then washed, passed through a 100 µm filter and incubated in Red Blood Cell Lysis buffer (Miltenyi Biotec, Germany) for 3-5 min at room temperature. After a wash with MACS buffer (0.5% BSA and 250 mM EDTA in PBS), samples were passed through a 40 µm filter and a 20 µm strainer-capped flow cytometry tube to generate a single cell suspension to use for flow cytometric analysis or further purification.

FACS analysis and cell sorting

Prepared single-cell suspensions of mouse lung tissues and *in vitro* cell lines were incubated with mouse FcR Blocking Reagent (Miltenyi Biotec, Germany) for 10 min at 4ºC followed by an incubation with a mix of pre-labelled antibodies (antibody information is provided in the table below) for 30 min at 4ºC. After two washes with MACS buffer, dead cells were stained with 4′,6-diamidino-2-phenylindole (DAPI). Flow cytometry analyses were carried out on a BD LSR-Fortessa (BD Biosciences, USA) and FlowJo 10.4.2 (FlowJO, LCC 2006-2018, USA) was used for further analysis. All cell-sorting experiments were carried out on a BD Influx cell sorter (BD Biosciences, USA).

Tissue digestion and FACS analysis in *ex-vivo* lineage tracing experiments

Lung tissues were dissociated with a collagenase/dispase solution as previously described[30]. Briefly, after lungs were cleared by perfusion with cold PBS through the right ventricle, 2 mL of dispase (50 U/ml, BD Biosciences, USA) was instilled into the lungs through the trachea until the lungs inflated, followed by instillation of 1% low melting agarose (Bio-Rad Laboratories, USA) through the trachea to prevent leakage of dispase. Each lobe was

dissected and minced into small pieces in a conical tube containing 3 ml of PBS, 60 µL of collagenase/dispase (Roche, Switzerland), and 7.5 µL of 1% DNase I (Merck Sigma-Aldrich, Germany) followed by rotating incubation for 45 min at 37°C. The cells were then filtered sequentially through 100- and 40-µm strainers and centrifuged at 1000rpm for 5 min at 4°C. The cell pellet was resuspended in 1 ml of ACK lysis buffer (0.15 M NH4Cl, 10mM KHCO3, 0.1 mM EDTA) and lysed for 90 s at room temperature. 6 ml basic F12 media (ThermoFisher Scientific, USA) was added and 500 µl of FBS (Fisher Scientific, USA) was slowly added in the bottom of tube. Cells were centrifuged at 1500 rpm for 5 min at 4°C. The cell pellet was resuspended in PF10 buffer (PBS with 10% FBS) for further staining. The antibodies used were as follows: CD45 (30-F11)-APC (BD Biosciences, USA), CD31 (MEC13.3)-APC (BD Biosciences, USA), and EpCAM (G8.8)-PE-Cy7 (BioLegend, USA). MOFLO system (Beckman Coulter, USA) was used for the sorting at Wellcome-MRC Stem Cell Institute Flow Cytometry Facility.

Lung organoid assay

Lung organoid co-culture assays were previously reported[3]. Briefly, freshly sorted epithelial cells ($Epcam^+CD45^-CD31^-Ter119^-GFP^-$) from either the metastatic niche or the distal lung were resuspended in 3D basic media (DMEM/F12, supplemented with 10% FBS, penicillin/streptomycin, 1 mM HEPES, and insulin/transferrin/selenium (ITS) (Merck Sigma-Aldrich, Germany), and mixed with MACS-sorted $CD31^+$ lung stromal cells or MLg cells followed by resuspension in growth factor-reduced (GFR) Matrigel (BD Biosciences, USA) at a ratio of 1:1. 100 µl of mixture was then placed in a 24-well transwell insert with a 0.4 µm pore (Corning, USA). 1-2.5 $x10^3$ distal lung or niche epithelial cells and 25,000 $CD31^+$ or MLg cells were seeded in each insert. 500 µl of 3D basic media was placed in the lower chamber and media was changed every other day. In addition, freshly sorted lineage-labelled $Scgb1a1^+$ club cells or $Sftpc^+$ AT2 cells were resuspended in 3D basic media followed by mixing with GFR matrigel retaining $CD31^+$ stromal cells as described above. For co-culture of lung epithelial cells with tumour cells, a mixture of 1-2.5 $x10^3$ distal lung epithelial cells and 25,000 $CD31^+$ cells in Matrigel was placed in the transwell insert, and 2,000 tumour cells FACS-sorted from metastatic lungs were seeded in the lower chamber. Plates were scored for colony number after 14 days. Colony-forming efficiency was calculated as number of colonies formed/number of cells plated per well as a percentage. Quantification of distinct types of differentiated colonies was performed by scoring the colonies expressing Sox2 or SP-C/Hopx by IF staining from at least five step sections (20 µm apart) per individual well. Bright-field images were acquired after 14 days using an EVOS microscope (ThermoFisher Scientific, USA).

3D Cell culture

Primary MMTV-PyMT actin-GFP cells were seeded at a density of 5,000 cells/well in a collagen solution coated Alvetex Scaffold 96-well plate (ReproCELL, Europe). The following day, $Ly6G^+$ lung cells and/or $Epcam^+$ lung epithelial cells were MACS sorted and seeded on top of the cancer cells at a density of 50,000 cells per well. In selected experiments, wells were supplemented with 4-Hydroxy-TEMPO (Merck Sigma-Aldrich, 200 µM) or mouse anti-Wisp1 (250 ng/mL, MAB1680, R&D, USA). The growth of GFP^+ cells was monitored daily for 6 days using the SteREO Lumar.V12 stereomicroscope (Zeiss, Germany), and images were quantified using ImageJ (NIH, USA). For quantification, the Li's Minimum Cross Entropy thresholding algorithm was performed on the stacked images.

For the CD104 staining experiment, $Epcam^+$ lung cells were harvested from mouse lung tissue via MACS sorting and seeded at a density of 1,500,000 cells per well on collagen solution coated Alvetex Scaffold 12-well inserts. After 48 h, MMTV-PyMT actin-GFP cells were seeded on top of the $Epcam^+$ cells at a density of 2,000 cells per scaffold insert.

Immunofluorescence and immunohistochemistry

Mouse lungs were fixed in 4% PFA in PBS for 24 h and embedded in paraffin blocks. 4 µm thick tissue sections were cut, deparaffinised and rehydrated using standard methods. After heat-mediated antigen retrieval in citrate buffer (unless stated otherwise), sections were blocked with a solution of 1% BSA, 10% Donkey serum in PBS.

mCherry and GFP staining. An overnight incubation at 4ºC with goat anti-GFP and rabbit anti-mCherry antibodies was followed by 1 h incubation at room temperature with anti-goat AlexaFluor 488 and anti-rabbit AlexaFluor 555 (both secondary antibodies were purchased from ThermoFisher Scientific (USA) and used at 1:400). Next, the slides were incubated with Sudan Black B for 20 min and mounted with Vectashield Mounting Medium with DAPI (Vector Laboratories, USA).

Lineage staining. An overnight incubation at 4ºC with goat anti-GFP antibody was followed by 45 min incubation at room temperature with secondary biotinylated-conjugated antibodies. Next, the VECTASTAIN Elite ABC kit (Vector Laboratories, USA) was used according to the manufacturer's instructions. The visualization of cell nuclei was performed with hematoxylin and analysis employed the Nikon Eclipse 90i light microscope and NIS-elements software.

WISP1 staining. An overnight incubation at 4ºC with goat anti-GFP and rabbit anti-mCherry antibodies was followed by 30 min incubation at room temperature with anti-goat AlexaFluor 488 and anti-rabbit AlexaFluor 555 (both secondary antibodies were purchased from ThermoFisher Scientific (USA) and used at 1:500). Next, the slides were incubated with Sudan Black B for 20 minutes and mounted with Vectashield Mounting Medium with DAPI (Vector Laboratories, USA).

Ki67 staining. Epcam^{+}CD45^{-}CD31^{-}Ter119^{-}GFP^{-} cells were sorted from lung suspensions, plated on poly-lysine glass coverslips for 15 min at room temperature and fixed in 4% PFA in PBS for 10 min. After fixation, cells were permeabilized with 0.1% Triton-X-100 in PBS for 5 min and incubated with a blocking solution (1% BSA, 10% goat serum, 0.3 M glycine, 0.1% Tween in PBS) for 1 h at room temperature. Next, cells were incubated overnight with an anti-mouse Ki67 antibody diluted in blocking solution followed by a 1 h incubation with a goat anti-rabbit AlexaFluor 488 (1:500, ThermoFisher Scientific (USA)). Finally, cells were mounted with Vectashield Mounting Medium with DAPI for imaging.

E-cadherin staining. CD49f^{+}CD104^{+}CD45^{-}CD31^{-}Ter119^{-}GFP^{-} cells were sorted from lung suspensions, cytospun on glass slides and fixed in 4% PFA in PBS for 10 min. Next, cells were permeabilized with 0.5% TritonX-100 for 30 min and incubated in blocking solution (4% BSA, 0.05% Tween20 in PBS) for 45 min at room temperature. Then, cells were incubated with a rat anti-E-cadherin antibody in blocking solution overnight at 4ºC followed by an incubation with a goat anti-rat AlexaFluor 647 (1:500, ThermoFisher Scientific (USA)). Finally, cells were mounted with Vectashield Mounting Medium with DAPI for imaging.

CD104 staining. Epcam^{+} cells were MACS sorted and plated on Alvetex scaffold inserts as described above. 7 days after plating the whole scaffold was collected, washed with PBS and incubated in blocking solution (10% goat serum in PBS) for 1 h at room temperature. Next, the samples were incubated with a conjugated anti-CD104-eFluor660 antibody (1:100 in PBS with 1:10 FcR blocking (Miltenyi Biotec, Germany)) for 1 h at room temperature. Then, the samples were fixed with 4% PFA in PBS for 10 min and mounted with Vectashield Mounting Medium with DAPI. Pictures were captured with the Axio Scan.Z1 slide scanner (Zeiss, Germany).

Lung organoid staining. Cultured colonies were fixed with 4% PFA in PBS for 2-4 h at room temperature followed by immobilization with Histogel (ThermoFisher Scientific, USA) for paraffin embedding. At least five step sections (20 µm apart) per individual well were stained. Fluorescence images were acquired using a confocal microscope Leica TCS SP5 (Leica Microsystems, Germany). All the images were further processed with Fiji software.

TTF1 and Ki67 co-staining. Target retrieval solution pH9 (Agilent DAKO, USA) was used as antigen retrieval. For histology, 1h incubation at room temperature with mouse anti-TTF1 was followed by 45 min incubation at room temperature with secondary biotinylated-conjugated antibodies. Next, the VECTASTAIN Elite ABC kit (Vector Laboratories, USA) was used according to the manufacturer's instructions. The visualization of cell nuclei was performed with hematoxylin and analysis employed the Nikon Eclipse 90i light microscope and NIS-elements software. For immune-fluorescence, 1h incubation at room temperature with mouse anti-TTF1 and rabbit anti-Ki67 was followed by 45 min incubation at room temperature with anti-mouse AlexaFluor 555 and anti-rabbit AlexaFluor 488 (both secondary antibodies were

purchased from ThermoFisher Scientific (USA) and used at 1:250). Next, the slides were incubated with Sudan Black B for 20 min and mounted with Vectashield Mounting Medium with DAPI (Vector Laboratories, USA.

All pictures were captured with either a Zeiss Upright710 confocal microscope or a Zeiss Upright780 confocal microscope unless differently stated.

Quantitative real time PCR

RNA preparation was performed using the MagMax-96 Total RNA Isolation Kit. cDNA synthesis was performed using a SuperScript III First-Strand Synthesis System (ThermoFisher Scientific, USA), according to the manufacturer's protocol. Quantitative real-time PCR samples were prepared with 50-100 ng total cDNA for each PCR reaction. The PCR, data collection and data analysis were performed on a 7500 FAST Real-Time PCR System (ThermoFisher Scientific, USA). GAPDH was used as internal expression reference.

Anti-Wisp1 treatment *in vivo*

BALB/cJ female mice (6 - 8 weeks old) were administered with anti-Wisp1 (5µg AF1680 and 5µg MAB1680, R&D, USA) or a control-IgG antibody via an intra-tracheal injection (50µl/mouse). The following day, mice were intravenously injected with 250,000 4T1 cells. Anti-Wisp1/control-IgG treatment was repeated daily, via a second intra-tracheal injection on day 4, and intra-peritoneal injections on days 2,3,5 and 6. Mice were harvested 7 days after the first treatment and lungs were embedded, cut and H&E stained. The lung metastatic burden was assessed by counting number of metastases on four levels (100µm interval) from two lung lobes (n=10 per group).

EdU *in vitro* proliferation assay

MMTV-PyMT actin-GFP cells were seeded at a density of 10,000 cells per well into collagen solution coated 6-well plates. The following day, $Ly6G^+$ lung cells and/or $Epcam^+$ lung cells were isolated via MACS sorting and added to the wells at a density of 100,000 cells/well. After 60h, wells were supplemented with 20 µM EdU (5-ethynyl-2´-deoxyuridine). Cells were harvested 6h later, and EdU incorporation was assessed using the Click-iT Plus EdU Flow Cytometry Assay Kit (ThermoFisher Scientific, USA), according to the manufacturer's instructions. Sample data were acquired on a BD LSR-Fortessa flow cytometer and analysed using FlowJo 10 software.

Conditioned media preparation and vesicles isolation

4T1-sLP-mCherry-GFP cells (Labelling-4T1) were plated on 10cm petri dishes. When cells were 80-90% confluent, 7 mL of DMEM with 10% FCS was added to be conditioned for 24 h. The conditioned media preparation and vesicles isolation were performed as previously described[31]. Briefly, the media was collected and spun at 300g for 10 min. Next, the supernatant was collected and spun at 2,000g for 10 min. The supernatant after this second centrifugation was collected and used as conditioned media. For vesicles isolation, the conditioned media was subsequently ultra-centrifuged at 10,000g for 30 min and at 100,000g for 70 min. The vesicle pellet at this stage was washed with PBS, spun at 100,000g for 70 min and resuspended again in PBS for *in vitro* uptake experiments.

ImageStream analysis

Image stream analyses were carried out on an ImageStream Mark X II Imaging Flow Cytometer (Amnis Merck, USA). The acquired data were analysed using IDEA software (Amnis Merck, USA).

Electron Microscopy (EM)

Experiments were performed on glass bottom dishes with a numbered grid (MatTek, USA) to enable subsequent location of the same cell imaged by confocal microscopy. After confocal imaging, cells were fixed in 8% formaldehyde in 0.1 M phosphate buffer (pH 7.4) added in equal quantities to cell media for 15 min and then further fixed in 2.5% glutaraldehyde and 4% formaldehyde in 0.1 M phosphate buffer (pH 7.4) for 1 h and then processed using the National Center for Microscopy and Imaging Research (NCMIR) protocol (Deerinck, T.J., et al., NCMIR methods for 3D EM: a new protocol for preparation of biological specimens for serial block face scanning electron microscopy. National Center for Microscopy and Imaging Research (2010) (available from https://ncmir.ucsd.edu/sbem-protocol).

For transmission electron microscopy **(**TEM), 70 nm serial sections were cut using a UC6 ultramicrotome (Leica Microsystems, Germany) and collected on formvar-coated slot grids. No post-staining was required due to the density of metal deposited using the NCMIR protocol. Images were acquired using a 120 kV Tecnai G2 Spirit TEM (FEI Company ThermoFisher Scientific, USA) and an Orius CCD camera (Gatan, USA).

RNA sequencing sample preparation

Bulk RNA sequencing: $CD45^{-}Ter119^{-}$ (CD45-ve) cells were sorted from single cell suspensions of metastatic lungs stained with anti-mouse CD45 and Ter119 antibodies and DAPI. RNA isolation was performed using the MagMax-96 Total RNA Isolation Kit (ThermoFisher Scientific, USA) that allows high quality RNA extraction from samples with low

cell numbers (<10,000 cells). RNA quality for each sample was assessed using the Agilent RNA 6000 Pico Kit (Agilent Technologies, USA). RNA was amplified and analysed at the Barts and London Genome Centre.

Single cell RNA sequencing: $CD45^{-}Ter119^{-}$ cells were sorted from single cell suspensions of metastatic lungs stained with anti-mouse CD45 and Ter119 antibodies and DAPI. Library generation for 10× Genomics were performed following the Chromium Single Cell 3' Reagents Kits (10X Genomics, USA) and sequenced on an Hiseq4000 (Illumina, USA), to achieve an average of 50,000 reads per cell.

Determination of intracellular ROS levels

Single cell suspensions from mouse lungs were incubated with mouse FcR Blocking Reagent for 5 min on ice and subsequently incubated with CellROX® Deep Red Reagent (ThermoFisher Scientific, USA) for 30 min at 37ºC following manufacturer's recommendations. Next, cells were washed twice with MACS buffer, stained with DAPI and analysed by flow cytometry.

Quantitative proteomic analysis of Ly6G cells

Neutrophils were FACS-sorted from single cell suspensions of metastatic lungs stained with a conjugated anti-mouse Ly6G-APC antibody (4 samples from independent sorts). Ly6G cells from the metastatic niche (Ch^{+}) and the distal lung (Ch^{-}) were digested into peptides using a previously described protocol (https://doi.org/10.1101/220343) and analysed by Data Independent Acquisition (DIA) mass spectrometry[32] on a Orbitrap Fusion Lumos instrument (ThermoFisher Scientific, USA). A hybrid spectral library was generated using the search engine Pulsar in Spectronaut Professional+ (version 11.0.15038, Biognosys AG, Switzerland) by combing Data Dependent Acquisition (DDA) runs obtained from a pooled sample of Ly6G cells, and the DIA data. Data analysis and differential protein expression was performed using Spectronaut Professional+. A detailed description of sample processing, data acquisition and processing are provided upon request.

Bioinformatic analysis

Bulk RNA sequencing: the sequencing was performed on biological triplicates for each condition generating approximately 35 million 76bp paired end reads. The RSEM package (version 1.2.29) [33] and Bowtie2 were used to align reads to the mouse mm10 transcriptome, taken from known Gene reference table available at UCSC (https://genome.ucsc.edu/). For RSEM, all parameters were run as default except "--forward-prob" which was set to "0.5". Differential expression analysis was carried out with DESeq2 package[34] (version 1.12.4) within R version 3.3.1 (https://www.r-project.org/). Genes were considered to be differentially

expressed if the adjusted p value was less than 0.05. Differentially expressed genes were taken forward and their pathway and process enrichments were analysed using Metacore (https://portal.genego.com).

Gene Set Enrichment Analysis, GSEA, (version 2.2.3) [35,36] was carried out using ranked gene lists using the Wald statistic and the gene sets of C2 canonical pathways and C5 biological processes. All parameters were kept as default except for enrichment statistic (classic) and max size which was changed to 5000 respectively. Gene signatures with FDR q-value equal or less than 0.05 were considered statistically significant.

Heatmaps of differentially expressed genes were generated using the gplots (Gregory et al., gplots: Various R Programming Tools for Plotting Data. R package version 3.0.1. (2016). https://CRAN.R-project.org/package=gplots) CRAN package (version 3.0.1). Genes were clustered using an Eisen distance matrix and average linkage clustering.

Single cell RNA sequencing: raw reads were initially processed by the Cell Ranger v2.1.1 pipeline, using STAR (v2.5.1b) to align to the mm10 transcriptome, deconvolve reads to their cell of origin using the UMI tags and report cell-specific gene expression count estimates. All subsequent analyses were performed in R-3.4.1 using the cellrangerRkit, monocle and pheatmap packages. Genes were considered to be "expressed" if the estimated ($\log_{10}$) count was at least 0.1. Primary filtering was then performed by removing from consideration: genes expressed in fewer than 20 cells; cells expressing fewer than 50 genes; cells for which the total yield (i.e. sum of expression across all genes) was more than 2 standard deviations from the mean across all cells in that sample; cells for which mitochondrial genes made up greater than 10% of all expressed genes. PCA decomposition was performed and, after consideration of the eigenvalue "elbow-plots", the first 25 components were used to construct t-SNE plots for both samples. Niche cells expressing Epcam were subdivided into those also expressing Cdh1 and those not expressing Cdh1. Other genes expressed in at least 50% of cells in a given group were said to be co-expressed and the set of genes co-expressed in one or more groups was presented as a heatmap, with the columns (cells) clustered using the standard Euclidean hierarchical method.

933 **Antibodies**

ANTIBODY	COMPANY	CATALOGUE No	CLONAL (CLONE)	DILUTION (TECHNIQUE)
Acetylated-tubulin	Sigma-Aldrich	T7451	Mouse monoclonal (6-11B-1)	1:1000 (IF)
CC10 (SCGB1A1)	Santa Cruz	sc-25555	Rabbit polyclonal (FL-96)	1:200 (IF)
CD11b-APC	Biolegend	10121	Rat monoclonal (M1/70)	1:100 (FC)
CD11b-APCCy7	Biolegend	101226	Rat monoclonal (M1/70)	1:100 (FC)
CD45-BV421	Biolegend	103133	Rat monoclonal (30-F11)	1:200 (FC)
CD45-APC	eBioscience	17-0451-83	Rat monoclonal (30-F11)	1:200 (FC)
CD45-APC-eFluor780	eBioscience	47-0451-82	Rat monoclonal (30-F11)	1:200 (FC)
CD49f-PerCP-eFluor710	eBioscience	46-0495-82	Rat monoclonal (ebioGOH3)	1:200 (FC)
CD104-eFluor660	eBioscience	50-1049-82	Rat monoclonal (439-9b)	1:100 (FC; IF)
CD326(EPCAM)-APC	eBioscience	17-5791-81	Rat monoclonal (G8.8)	1:200 (FC)
CD326(EPCAM)-APC750Fire	Biolegend	118230	Rat monoclonal (G8.8)	1:200 (FC)
E-CADHERIN	Abcam	Ab11512	Rat monoclonal (DECMA-1)	1:200 (IF)
GFP	Abcam	ab6673	Goat polyclonal	1:300 (IF)
HOPX	Santa Cruz	sc-30216	Rabbit polyclonal (FL-73)	1:250 (IF)
Ki67	Abcam	Ab16667	Rabbit monoclonal (SP6)	1:300 (IF)
Ly6A/E(SCA-1)-APC	Biolegend	108111	Rat monoclonal (D7)	1:200 (FC)
Ly6A/E(SCA-1)-APC750Fire	Biolegend	127652	Rat monoclonal (D7)	1:200 (FC)
Ly6A/E(SCA-1)-BV786	BD Bioscience	563991	Rat monoclonal (D7)	1:200 (FC)
Ly6G-APC	BD Bioscience	560599	Rat monoclonal (1A8)	1:150 (FC)
Ly6G-APC750Fire	Biolegend	127652	Rat monoclonal (1A8)	1:150 (FC)
Ly6G-V450	BD Bioscience	560603	Rat monoclonal (1A8)	1:150 (FC)
mCHERRY	Abcam	ab183628	Rabbit polyclonal	1:750 (IF)
SOX2	eBioscience	14-9811-80	Rat monoclonal (Btjce)	1:500 (IF)
SP-C	Santa Cruz	sc-7706	Goat polyclonal (M-20)	1:200 (IF tissue)
TER-119	Biolegend	116233	Rat monoclonal (TER-119)	1:200 (FC)
TTF1	DAKO	M3575	Mouse monoclonal (8G7G3/1)	1:50 *IF)
WISP1	Abcam	Ab178547	Rabbit polyclonal	1:100 (IF)

934

Methods references

25. 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).
26. Okabe, M., Ikawa, M., Kominami, K., Nakanishi, T. & Nishimune, Y. `Green mice' as a source of ubiquitous green cells. *FEBS Letters* **407,** 313–319 (1997).
27. Rock, J. R. *et al.* Multiple stromal populations contribute to pulmonary fibrosis without evidence for epithelial to mesenchymal transition. *Proc Natl Acad Sci USA* **108,** E1475–83 (2011).
28. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. *BMC Dev. Biol.* **1,** 4 (2001).
29. Rawlins, E. L. *et al.* The role of Scgb1a1+ Clara cells in the long-term maintenance and repair of lung airway, but not alveolar, epithelium. *Cell Stem Cell* **4,** 525–534 (2009).
30. Lee, J.-H. *et al.* Anatomically and Functionally Distinct Lung Mesenchymal Populations Marked by Lgr5 and Lgr6. *Cell* **170,** 1149–1156.e12 (2017).
31. Théry, C., Amigorena, S., Raposo, G. & Clayton, A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. *Curr Protoc Cell Biol* **Chapter 3,** Unit 3.22–3.22.29 (2006).
32. Bruderer, R. *et al.* Optimization of Experimental Parameters in Data-Independent Mass Spectrometry Significantly Increases Depth and Reproducibility of Results. *Mol. Cell Proteomics* **16,** 2296–2309 (2017).
33. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
34. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
35. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 15545–15550 (2005).
36. Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34,** 267–273 (2003).