ARTICLE

Epithelial–to–mesenchymal transition is not required for lung metastasis but contributes to chemoresistance

Kari R. Fischer[1,2,3,4], Anna Durrans[1,2,3], Sharrell Lee[1,2,3], Jianting Sheng[5], Fuhai Li[5], Stephen T. C. Wong[5,6], Hyejin Choi[1,2,3,4], Tina El Rayes[1,2,3,4], Seongho Ryu[1,3,7], Juliane Troeger[8,9], Robert F. Schwabe[8,9], Linda T. Vahdat[10], Nasser K. Altorki[1,3], Vivek Mittal[1,2,3] & Dingcheng Gao[1,2,3]

The role of epithelial–to–mesenchymal transition (EMT) in metastasis is a longstanding source of debate, largely owing to an inability to monitor transient and reversible EMT phenotypes *in vivo*. Here we establish an EMT lineage–tracing system to monitor this process in mice, using a mesenchymal–specific Cre–mediated fluorescent marker switch system in spontaneous breast–to–lung metastasis models. We show that within a predominantly epithelial primary tumour, a small proportion of tumour cells undergo EMT. Notably, lung metastases mainly consist of non–EMT tumour cells that maintain their epithelial phenotype. Inhibiting EMT by overexpressing the microRNA miR–200 does not affect lung metastasis development. However, EMT cells significantly contribute to recurrent lung metastasis formation after chemotherapy. These cells survived cyclophosphamide treatment owing to reduced proliferation, apoptotic tolerance and increased expression of chemoresistance–related genes. Overexpression of miR–200 abrogated this resistance. This study suggests the potential of an EMT–targeting strategy, in conjunction with conventional chemotherapies, for breast cancer treatment.
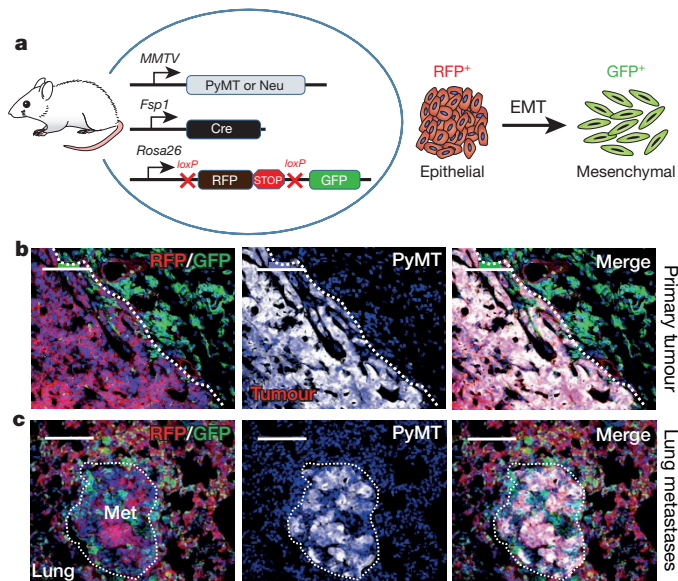
Despite significant advances in diagnosing and treating cancer, metastasis persists as a barrier to successful therapy and the main cause of cancer-related death[1]. The EMT, wherein epithelial cells depolarize, lose their cell–cell contacts, and gain an elongated, fibroblast-like morphology, is a potential mechanism by which tumour cells gain metastatic features. Functional implications of EMT include enhanced mobility, invasion and resistance to apoptotic stimuli[2,3]. Moreover, through EMT tumour cells acquire cancer stem cell, secondary tumour-initiating and chemoresistance properties[4–6]. However, the importance of EMT *in vivo* is fiercely debated owing to major challenges. Mesenchymal tumour cells cannot easily be distinguished from neighbouring stromal cells, and metastatic lesions mostly exhibit epithelial phenotypes[7]. The latter may be due to the hypothesized reverse process, mesenchymal to epithelial transition (MET), of the disseminated tumour cells. Studies have confirmed that mesenchymal cells are more capable of escaping the primary tumour, and of reaching distant sites, but it remains unproven that those same cells complete the full metastatic cascade in the form of a secondary nodule. Without evidence for the dissemination, colonization and metastatic outgrowth of mesenchymal tumour cells, the role of EMT will remain contested. In this study, we employed multiple transgenic mouse models, establishing a cell lineage tracing approach together with characterization of epithelial and mesenchymal markers, to address the requirement of EMT in metastasis. The newly established transgenic model also provided us a unique opportunity to study the contribution of EMT to chemoresistance.

EMT lineage tracing during metastasis

To track EMT during metastasis *in vivo*, we generated a mesenchymal-specific, Cre-mediated fluorescent marker switch strategy and established a triple-transgenic mouse model (*MMTV-PyMT/Rosa26-RFP-GFP/Fsp1-cre*, tri-PyMT, Fig. 1a). In these mice, spontaneous multifocal breast adenocarcinomas with distinct epithelial characteristics resembling the human luminal subtype develop in the mammary glands, and give rise to lung metastases with high penetrance[8,9]. The *Fsp1* (fibroblast specific protein 1) promoter drives expression of Cre recombinase in cells of mesenchymal lineage[10]. A Cre-switchable fluorescent marker (*lox*-RFP-STOP-*lox*-GFP) is ubiquitously expressed under the control of the β-actin promoter in the *Rosa26* locus[11]. *Fsp1* is the critical gatekeeping gene of EMT initiation[12], and its early activation in this process[13] allows for lineage tracing of tumour cells that have undergone EMT *in vivo*. Importantly, the colour switch system is irreversible—even if the mesenchymal tumour cells undergo MET in the metastatic organs[14], they would remain GFP+.

Primary breast tumours developed in the tri-PyMT mice at 8 weeks of age. Immunofluorescence revealed that the majority of tumour cells, identified by PyMT oncogene expression, were RFP positive (Fig. 1b). These cells expressed E-cadherin and lacked vimentin (Extended Data Fig. 1a), indicating their epithelial phenotype. The GFP+ cells detected in the tumour bed were largely haematopoietic cells as they are PyMT negative and express CD45, a pan-haematopoietic marker (Extended Data Fig. 1a), which is consistent with previous reports[15]. Altogether,

Figure 1 | Establishing an EMT lineage tracing system in triple-transgenic mice. a, Schematic of triple-transgenic mice carrying polyoma middle-T (PyMT) or Neu oncogenes driven by the *MMTV* promoter, Cre recombinase under the control of the *Fsp1* promoter, and floxed RFP-STOP followed by GFP under control of the β-actin promoter in the *Rosa26* locus. RFP⁺ epithelial tumour cells undergoing EMT permanently convert into GFP⁺ cells following activation of *Fsp1*-Cre. **b, c**, Immunofluorescent microscopy images of tri-PyMT primary tumours (**b**) and lung metastases (met; **c**) (>10 sections from 3 mice), depicting RFP⁺ and GFP⁺ cells within the tumour bed, and staining (white, pseudo-coloured) for PyMT. Scale bars, 100 μm.

this data suggests that tumour cells maintain their original RFP expression and epithelial phenotype in the primary tumour.

Lung metastasis developed spontaneously in tri-PyMT lungs at 12 weeks of age. Surprisingly, the PyMT-positive metastatic lesions were RFP⁺ (Fig. 1c), and epithelial (E-cadherin⁺/vimentin⁻) (Extended Data Fig. 1b), whereas only non-tumour cells expressed GFP. These results indicate that tumour cells did not activate the mesenchymal-specific *Fsp1* promoter, and retained their epithelial phenotype during metastasis. Thus, tumour cells may not undergo EMT to form metastatic lesions.

Lineage tracing in additional models

To exclude the possibility that the absence of EMT in metastasis may be unique to PyMT-driven breast tumours, we established EMT lineage tracing in the Neu oncogene-driven[16] spontaneous breast cancer model (*MMTV-neu/Rosa26-RFP-GFP/Fsp1-Cre*, tri-Neu mouse). The Neu (ErbB-2) proto-oncogene is associated with 20–30% of human breast cancers, and *MMTV*–Neu transgenic mice spontaneously develop focal adenocarcinomas resembling human luminal phenotypes after an extended latency at 6–8 months of age. Lung metastases are frequently (72%) observed in these transgenic mice at 9–12 months of age. Mirroring the tri-PyMT model, the Neu⁺ tumour cells in both primary and metastatic lesions in tri-Neu mice were also RFP⁺ and epithelial (E-cad⁺/Vim⁻) (Extended Data Fig. 2). Therefore, the absence of EMT during metastasis formation is an oncogene-independent phenomenon, manifesting in both PyMT and Neu-driven tumours.

To overcome the limitation of using solely *Fsp1*–Cre to indicate EMT, we acquired the vimentin–CreER transgenic mouse, which successfully traced mesenchymal lineage cells during liver fibrosis[17], and generated an additional EMT lineage tracing model (tri-PyMT/Vim mice, *MMTV-PyMT/Rosa26-RFP-GFP/Vimentin-creER*). After continuous induction of Cre activity by Tamoxifen injection (2 mg, intraperitoneal, three times per week starting when the primary tumours appear at 8 weeks of age) the majority of tumour cells in both the primary and metastatic lesions in tri-PyMT/Vim mice were RFP⁺ (Extended Data Fig. 3)—suggesting an absence of vimentin promoter activation
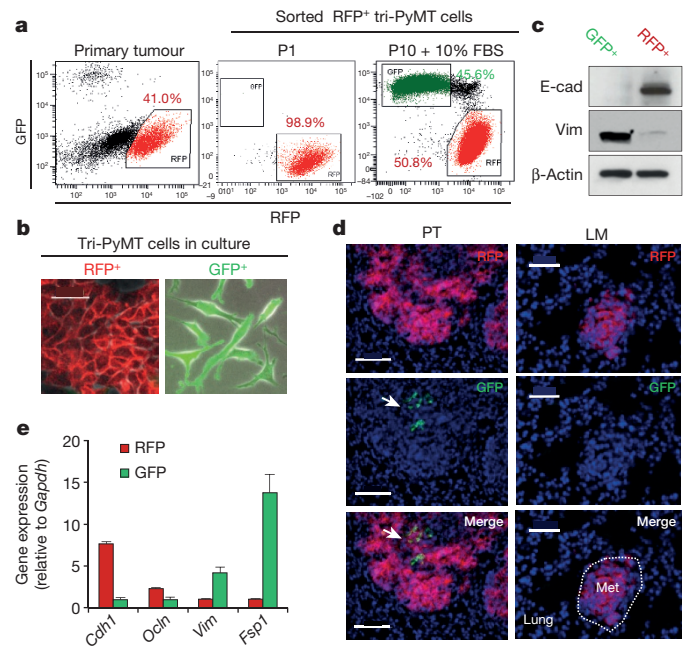
Figure 2 | The EMT lineage tracing system reports EMT in tumour cells with high fidelity. a, Scatter plots from flow cytometry analysis of tri-PyMT primary tumour cells, depicting GFP⁺ and RFP⁺ populations in the primary tumour immediately after sorting of RFP⁺ cells (P1), and after ten passages in culture with 10% FBS (P10 + 10% FBS). Numbers indicate the percentage of RFP⁺ and GFP⁺ cells in the total population. **b**, Phase contrast/fluorescent overlay image of tri-PyMT cells in culture. Scale bar, 50 μm. **c**, Western blot of sorted RFP⁺ and GFP⁺ tri-PyMT cells for E-cadherin, vimentin and β-actin as a loading control. Representative of two individual experiments. For original gel images, see Supplementary Fig. 1. **d**, Representative imaging of GFP⁺ and RFP⁺ tumour cells in primary tumours (PT) and lung metastases (LM) in the orthotopic model (*n* = 8 mice). Arrow indicates scattered GFP⁺ EMT tumour cells in the primary tumour. Scale bars, 100 μm (PT) and 50 μm (LM). **e**, qRT–PCR analysis of relative expression of EMT markers in RFP⁺ and GFP⁺ cells sorted from orthotopic tri-PyMT primary tumours. *Gapdh* served as the internal control. E-cadherin is encoded by the *Cdh1* gene. Occludin is encoded by the *Ocln* gene. Data are reported as mean ± s.e.m., *n* = 4 primary tumours.

during lung metastasis formation. EMT marker staining also revealed the epithelial phenotype (E-cad⁺/Vim⁻) of the tumour cells in both primary and metastatic lesions (Extended Data Fig. 3).

Together, results from two oncogene-driven metastatic tumour models (*MMTV*–PyMT and *MMTV*–Neu) and two independent mesenchymal-specific reporters (*Vim*–Cre and *Fsp1*–Cre) suggest that EMT does not significantly contribute to the development of lung metastases.

Validating EMT lineage tracing

To evaluate the specificity and sensitivity of the EMT lineage tracing system, we established a cell line from the tri-PyMT breast tumours. In culture, RFP⁺ tri-PyMT cells switched their fluorescent marker expression to GFP, as indicated by the presence of a RFP⁺/GFP⁺ double-positive transitioning population (Fig. 2a). The cells were cultured in 10% FBS, and serum is known to be enriched for many EMT promoting factors including TGFs[18]. Moreover, addition of TGF-β1 in low-serum conditions (2% FBS), yielded an increase in GFP⁺ cells (Extended Data Fig. 4a). In concert with the fluorescent marker switch, tri-PyMT cells changed their morphology from cobblestone-like clusters of epithelial cells to dispersed spindle-shaped mesenchymal cells (Fig. 2b). Reflecting the morphologic differences, the GFP⁺ cells were more motile than RFP⁺ cells (Extended Data Fig. 4b).

The fidelity of the EMT lineage tracing system was confirmed by analysis of EMT marker expression in sorted RFP⁺ and GFP⁺ tri-PyMT cells. RFP⁺ cells expressed elevated levels of epithelial markers including E-cadherin and Occludin, while GFP⁺ cells expressed

several mesenchymal markers including vimentin, FSP1, Twist, Zeb1 and Zeb2 as determined by quantitative reverse transcription PCR (qRT–PCR) (Extended Data Fig. 4c). Both RFP$^+$ and GFP$^+$ tri-PyMT cells expressed the PyMT oncogene. Consistently, western blot analysis confirmed the differential expression of E-cadherin and vimentin in RFP$^+$ and GFP$^+$ cells (Fig. 2c). Flow cytometry for E-cadherin revealed that the majority of E-cadherin$^-$ cells were GFP$^+$ (97.4%) (Extended Data Fig. 4d). Of note, the E-cadherin$^+$ cells were either RFP$^+$ (93.6%) or RFP$^+$/GFP$^+$ (6.0%), demonstrating that tumour cells switch their fluorescent marker expression before the loss of epithelial markers, and validating the early reporting of EMT in our system. These results confirm that the *Fsp1*–Cre-mediated fluorescent marker switch in tumour cells reports EMT with high fidelity and efficiency.

Rare EMT events in tumour progression

In the triple-transgenic models, ubiquitous expression of GFP in the tumour microenvironment precluded detection of potentially rare GFP$^+$ tumour cells. To confine the fluorescence to tumour cells, we established an orthotopic model by implanting purified RFP$^+$ tri-PyMT cells in wild-type mice (Extended Data Fig. 5a, b). Consistent with observations in the triple-transgenic mice, primary tumours contained RFP$^+$ epithelial cells (Fig. 2d). However, GFP$^+$ cells were detected, indicating tumour EMT (Fig. 2d and Extended Data Fig. 5c, upper panel). These cells lacked E-cadherin (Extended Data Fig. 5c, upper panel) and made up $1.98 \pm 1.40\%$ ($n = 6$) of the total tumour cells (Extended Data Fig. 5d). qRT–PCR analysis of EMT markers comparing sorted RFP$^+$ and GFP$^+$ cells from the same primary tumour confirmed the mesenchymal phenotype of the GFP$^+$ cells (Fig. 2e). Importantly, these GFP$^+$ EMT tumour cells did not contribute to lung metastasis. Early disseminated tumour cells detected in the lungs were epithelial and RFP$^+$ (Extended Data Fig. 5c, middle panel), and 28 lung nodules detected in 8 mice maintained the epithelial phenotype (Fig. 2d and Extended Data Fig. 5c, lower panel).

We also established an orthotopic tri-PyMT/Vim model, wherein Tamoxifen was administered directly after orthotopic injection to ensure immediate tracing of EMT events. Consistently, the majority of tumour cells in both primary and metastatic tumours were RFP$^+$ and epithelial (Extended Data Fig. 6). Again, GFP$^+$ EMT events ($4.46 \pm 1.0\%$ of total tumour cells, $n = 3$) were detected in the primary tumours.

To further dissect the metastatic cascade, we quantified the relative numbers of RFP$^+$ and GFP$^+$ cells in the primary tumour, blood and metastases of the tri-PyMT orthotopic model by flow cytometry. An RFP to GFP ratio of ~100:1 in the primary tumour and ~15:1 in the blood was observed (Extended Data Fig. 7a, b). However, gain by the enrichment of GFP$^+$ cells in circulation did not translate to an advantage in metastatic outgrowth, as the RFP:GFP ratio in the lung was ~150:1. Altogether, these findings are consistent with our observations in the triple-transgenic models, suggesting that the majority of breast tumour cells persist in an epithelial state during primary tumour growth and lung metastasis formation.

EMT inhibition and metastasis formation

In spite of the extensive characterization of the EMT reporter system, there was still the distant possibility of our reporter failing to manifest all EMT events *in vivo*. Therefore, we sought to inhibit EMT and determine its impact on metastasis. We ectopically expressed miR-200, a well-known inhibitor of EMT that directly targets *Zeb1* and *Zeb2*—the transcriptional repressors of E-cadherin[19,20]. We posited that stably expressing miR-200 in tri-PyMT cells would block EMT and trap tumour cells in a permanent epithelial state. Compared with control cells, miR-200 overexpressing cells (Extended Data Fig. 7c) showed elevated expression of epithelial cell markers and reduced expression of mesenchymal markers (Extended Data Fig. 7d). As expected, overexpression of miR-200 inhibited the RFP to GFP conversion (>90% remaining RFP$^+$, Fig. 3a). These results substantiate effective miR-200 suppression of EMT in the tri-PyMT cells.
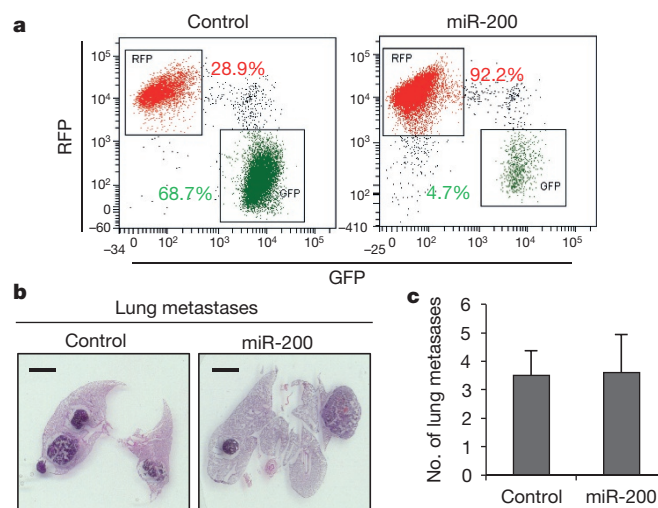
Figure 3 | mir-200 inhibition of EMT in tri-PyMT cells did not impact lung metastasis. a, Flow cytometry analysis of tri-PyMT control and mir-200-expressing cells, indicating the percentage of RFP$^+$ and GFP$^+$ cells. **b**, Representative histologic lung images in tri-PyMT control and mir-200-expressing orthotopic mice ($n = 5$). Scale bar, 1.5 mm. **c**, Quantification of lung metastasis formation (number of individual nodules) in tri-PyMT control and mir-200-expressing tumour-bearing mice ($n = 5$). Data reported as the mean $\pm$ s.e.m.
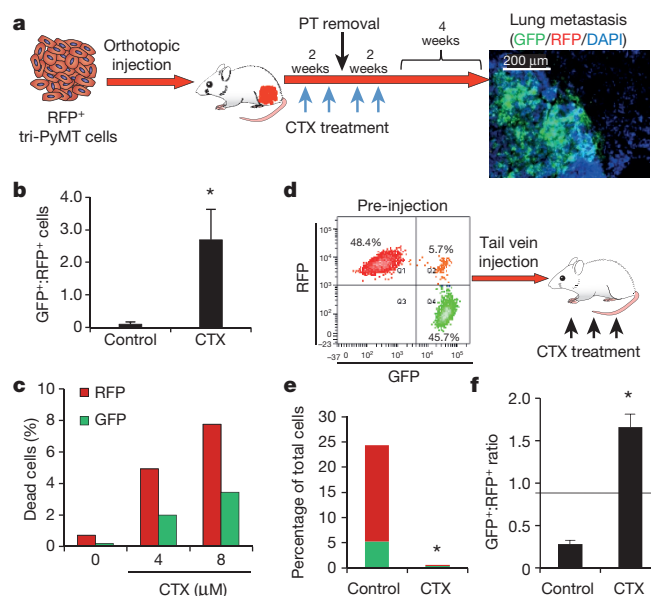
To explore the impact of inhibiting EMT on metastasis formation *in vivo*, we orthotopically injected miR-200 overexpressing tri-PyMT cells. We identified 18 metastases in 5 mice, a similar ratio to that observed in mice bearing control tri-PyMT cells (28 metastases in 8 mice) (Fig. 3b, c). These results demonstrate that inhibition of EMT by miR-200 overexpression does not impair the ability of tumour cells to form distant lung metastases.

EMT is involved in chemoresistance

Emerging evidence suggests a molecular and phenotypic association between EMT and chemoresistance in several cancers[21–23]. Compellingly, residual breast cancers following chemotherapy display a mesenchymal phenotype and tumour-initiating features[23]. To determine if the acquisition of chemoresistance induces specific molecular changes consistent with EMT, we evaluated the orthotopic tri-PyMT model under chemotherapy. Animals with established primary tumours were treated with cyclophosphamide (CTX), a commonly used drug in breast cancer treatment[24] ($100 \, \text{mg} \, \text{kg}^{-1}$, once per week, for two weeks prior, and two weeks after, surgery; Fig. 4a). The tumours responded to chemotherapy, manifesting a 60% reduction in growth and markedly enhanced apoptotic activity (Extended Data Fig. 8a–c). Of note, the RFP$^+$ cells were highly proliferative and apoptotic in comparison with GFP$^+$ cells in CTX-treated mice (Extended Data Fig. 8d–g), suggesting that GFP$^+$ cells have reduced susceptibility to chemotherapy. However, in the primary tumour, the GFP$^+$ cell percentage remained static under CTX treatment (Extended Data Fig. 8h).

Remarkably, in the early metastatic lungs (four weeks after tumour inoculation), flow cytometry analysis revealed a 2.7:1 ratio of GFP$^+$ to RFP$^+$ cells in CTX-treated mice (Fig. 4b). Subsequently at four weeks after cessation of treatment, a notable contribution of GFP$^+$ tumour cells was detected in 5 out of 17 metastatic lesions (Fig. 4a). This is in contrast to untreated mice, where all metastatic lesions were derived from RFP$^+$ cells (Fig. 2d), suggesting that the EMT process may be involved in metastatic outgrowth in the context of chemotherapy.

To evaluate the effects of CTX on the EMT and non-EMT cell populations, sorted GFP$^+$ and RFP$^+$ cells were incubated with CTX *in vitro*—the GFP$^+$ cells were markedly more resistant to both short- and long-term treatment (Fig. 4c and Extended Data Fig. 9a, b). The selective advantage of mesenchymal tumour cells in the context of chemotherapy

Figure 4 | EMT tumour cells are resistant to chemotherapy. a, Schema of CTX treatment in tri-PyMT orthotopic model. Mice bearing an RFP$^+$ primary tumour were treated with CTX (100 mg kg^{-1}, once per week, for 4 weeks, as indicated by blue arrows). After 2 weeks of treatment, primary tumour (PT) was removed (black arrow). Lung metastasis growth was permitted for 4 weeks post CTX treatment. Fluorescent imaging of lungs revealed the contribution of GFP$^+$ tumour cells to lung metastases ($n = 9$ mice). **b**, Ratio of GFP$^+$ to RFP$^+$ cells in early metastatic lungs (4 weeks post orthotopic injection) of untreated control and CTX-treated mice as quantified by flow cytometry ($n = 4$, *$P < 0.05$). Data reported as the mean ± s.e.m. **c**, Apoptosis (as measured by Annexin binding) of RFP$^+$ and GFP$^+$ tri-PyMT cells treated with CTX ($n = 2$ biological replicates). **d**, Flow cytometry scatter plot showing the proportions of RFP$^+$ and GFP$^+$ tri-PyMT cells before intravenous injection. Mice were treated with CTX (100 mg kg^{-1} per week for 3 weeks, $n = 5$ mice per group). **e**, Quantification of flow cytometry data showing the percentage of RFP$^+$ and GFP$^+$ tumour cells (red and green bars, respectively) of total cells in the lung of control and CTX-treated mice ($n = 5$ mice per group, *$P < 0.05$). **f**, Quantification of flow cytometry data showing the ratio of GFP$^+$ to RFP$^+$ cells in lungs of control and CTX-treated mice. Black line represents the starting ratio of GFP$^+$ to RFP$^+$ cells before injection as derived from the data in Fig. 4d (*$P < 0.05$). Data reported as the mean ± s.e.m.

was then corroborated by a competitive survival assay *in vivo* (Fig. 4d). Mice were injected intravenously with an equivalent number of RFP$^+$ and GFP$^+$ cells, and immediately received CTX (100 mg kg^{-1}, once per week). After three weeks, lungs were harvested and the ratio of RFP$^+$ and GFP$^+$ cells was assessed by flow cytometry. CTX significantly inhibited outgrowth of lung metastasis from both RFP$^+$ and GFP$^+$ cells (Fig. 4e). The untreated lungs were morbidly overwhelmed with tumours, with nearly 80% of the tumour cells detected as RFP$^+$. Conversely, in CTX-treated mice, more than 60% of the surviving tumour cells were GFP$^+$, producing a significantly higher ratio of GFP:RFP cells in these mice (Fig. 4f). These results indicate that GFP$^+$ EMT cells are more resistant to chemotherapy both *in vitro* and *in vivo*.

Immunostaining revealed that in the untreated mice, both RFP$^+$ and GFP$^+$ cells formed epithelial metastatic lesions (E-cad$^+$/Vim$^-$) (Extended Data Fig. 9c). Given the initial mesenchymal phenotypes of GFP$^+$ cells before injection, this suggests that the GFP$^+$ tumour cells have undergone MET in the metastatic organ. On the other hand, in CTX-treated mice the majority of surviving tumour cells were scattered mesenchymal GFP$^+$ cells (E-cad$^-$/Vim$^+$) (Extended Data Fig. 9d). Together, these observations suggest that EMT tumour cells that sustain a mesenchymal phenotype are resistant to chemotherapy.

To begin to investigate the molecular underpinnings of mesenchymal tumour cell resistance, we analysed the transcriptomic changes of
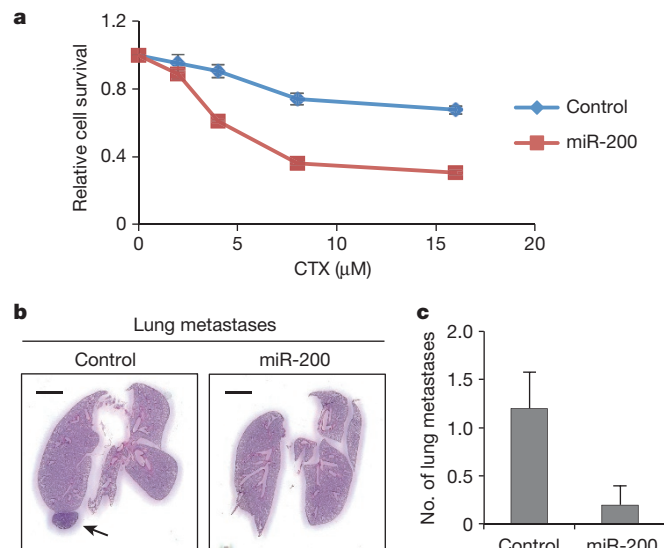
Figure 5 | miR-200 overexpression abrogates CTX resistance.
a, Sensitivity of Control and miR-200-expressing tri-PyMT tumour cells to CTX treatment as measured by CellTiter-Glo. $n = 4$ biological replicates per condition **b**, Representative histologic lung images in tri-PyMT control and mir-200-expressing tumour-bearing mice treated with CTX ($n = 5$). Scale bar, 1.5 mm. **c**, Quantification of lung metastasis formation (number of individual nodules) in CTX-treated tri-PyMT control and mir-200-expressing tumour-bearing mice ($n = 5$). Data reported as the mean ± s.e.m.

EMT tumour cells. We sorted RFP$^+$ and GFP$^+$ cells and performed RNA-sequencing analysis (Supplementary Information Table 1). In addition to the expected changes in EMT marker expression (Extended Data Fig. 10a), the expression of many cell-proliferation-related genes was reduced in GFP$^+$ cells (Extended Data Fig. 10b), mirroring their phenotype of reduced proliferation *in vivo*. The GFP$^+$ cells also showed increased expression of proven chemoresistance-related factors including IL6, Periostin, Enpp2 and Pdgfr[25–28]. Additionally, the CTX-treated GFP$^+$ cells elevated their expression of many drug-metabolizing enzymes including drug transporters (Abcb1a, Abcb1b and Abcc1), aldehyde dehydrogenases (ALDHs), cytochrome P450s, and glutathione-metabolism-related enzymes (Extended Data Fig. 10c). The main toxicity of CTX is due to its metabolite phosphoramide mustard, which is only formed in cells with low levels of ALDHs. ALDH converts the CTX-metabolite aldophosphamide into the non-toxic carboxyphosphamide[29]. In accordance with the transcriptomic data, GFP$^+$ cells had significantly higher ALDH activity compared with RFP$^+$ cells (Extended Data Fig. 10d). These properties of reduced proliferation, increased apoptotic resistance, and upregulation of chemoresistance and drug metabolizing genes in GFP$^+$ EMT tumour cells may contribute to their insensitivity to CTX. Notably, GFP$^+$ cells were also refractory to other commonly used chemotherapies including doxorubicin, paclitaxel, and fluorouracil treatment (Extended Data Fig. 10e).

To demonstrate that the EMT is required for the generation of CTX resistance, we first tested *in vitro* the effect of treatment on control and miR-200 overexpressing tri-PyMT cells. With increasing concentrations of CTX, the miR-200 cells were significantly more susceptible to therapy (Fig. 5a). We then expanded upon this finding *in vivo*, establishing orthotopic control and miR-200 primary tumours, and applying the pre- and post-surgery CTX regimen. We found that by blocking EMT in tumour cells, we effectively ablated metastatic growth (Fig. 5b, c). Thus, EMT contributes to the development of chemo-resistant metastasis.

Discussion

Using two independent EMT lineage tracing strategies in two disparate oncogene-driven autochthonous models of breast cancer, we demonstrated that lung metastases are derived from non-EMT tumour

cells, contradicting the original EMT/MET hypothesis[2,30]. In a tracing system similar to our own, EMT was identified in primary tumours, but the mesenchymal lineage status of the metastatic nodules was not pursued[31]. Ultimately in our models we found that tumour cells disseminate and form metastases while persisting in their epithelial phenotype, in accordance with a recent study[32]. To underline that EMT is not required for metastasis, overexpression of miR-200—a microRNA that is incongruously associated with both reduced invasion[19,20] and increased metastasis[33]—resulted in combined suppression of the EMT-promoting transcription factors Snail1/2, Twist, Zeb1 and Zeb2, but had no effect on metastasis. Given that both epithelial and mesenchymal tumour cells have the potential to disseminate, it is plausible that the larger fraction of highly proliferative epithelial cells outcompete the minor EMT tumour cell population in generating macrometastatic lesions.

Until now, the majority of data connecting EMT with chemoresistance was largely derived from *in vitro* studies, or clinical prognostic data. Here we demonstrate that highly proliferative non-EMT cells are sensitive to chemotherapy, and observe the emergence of recurrent EMT-derived metastases after treatment. There is a great emphasis towards developing EMT-targeting therapies[34,35], and our studies suggest that while EMT blockade may not affect metastasis formation, specifically targeting EMT tumour cells will be synergistic with conventional chemotherapy. Thus, our EMT lineage tracing system provides a unique preclinical platform to develop combination therapies that will eliminate both populations, and combat chemoresistance.

1. Bastid, J. EMT in carcinoma progression and dissemination: facts, unanswered questions, and clinical considerations. *Cancer Metastasis Rev.* **31,** 277–283 (2012).
2. Kalluri, R. & Weinberg, R. A. The basics of epithelial–mesenchymal transition. *J. Clin. Invest.* **119,** 1420–1428 (2009).
3. Scheel, C. & Weinberg, R. A. Phenotypic plasticity and epithelial–mesenchymal transitions in cancer and normal stem cells? *Int. J. Cancer* **129,** 2310–2314 (2011).
4. Mani, S. A. *et al.* The epithelial–mesenchymal transition generates cells with properties of stem cells. *Cell* **133,** 704–715 (2008).
5. Gal, A. *et al.* Sustained TGF beta exposure suppresses Smad and non-Smad signalling in mammary epithelial cells, leading to EMT and inhibition of growth arrest and apoptosis. *Oncogene* **27,** 1218–1230 (2008).
6. Hennessy, B. T. *et al.* Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Res.* **69,** 4116–4124 (2009).
7. Gao, D. *et al.* Myeloid progenitor cells in the premetastatic lung promote metastases by inducing mesenchymal to epithelial transition. *Cancer Res.* **72,** 1384–1394 (2012).
8. Lin, E. Y. *et al.* Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *Am. J. Pathol.* **163,** 2113–2126 (2003).
9. 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).
10. Bhowmick, N. A. *et al.* TGF-β signaling in fibroblasts modulates the oncogenic potential of adjacent epithelia. *Science* **303,** 848–851 (2004).
11. Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, L. & Luo, L. A global double-fluorescent Cre reporter mouse. *Genesis* **45,** 593–605 (2007).
12. Xue, C., Plieth, D., Venkov, C., Xu, C. & Neilson, E. G. The gatekeeper effect of epithelial–mesenchymal transition regulates the frequency of breast cancer metastasis. *Cancer Res.* **63,** 3386–3394 (2003).
13. Okada, H., Danoff, T. M., Kalluri, R. & Neilson, E. G. Early role of Fsp1 in epithelial–mesenchymal transformation. *Am. J. Physiol.* **273,** F563–F574 (1997).
14. Gunasinghe, N. P., Wells, A., Thompson, E. W. & Hugo, H. J. Mesenchymal–epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer. *Cancer Metastasis Rev.* **31,** 469–478 (2012).
15. Cabezón, T. *et al.* Expression of S100A4 by a variety of cell types present in the tumor microenvironment of human breast cancer. *Int. J. Cancer* **121,** 1433–1444 (2007).
16. Guy, C. T. *et al.* Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proc. Natl Acad. Sci. USA* **89,** 10578–10582 (1992).
17. Troeger, J. S. *et al.* Deactivation of hepatic stellate cells during liver fibrosis resolution in mice. *Gastroenterology* **143,** 1073–1083 (2012).
18. Dumont, N. *et al.* Sustained induction of epithelial to mesenchymal transition activates DNA methylation of genes silenced in basal-like breast cancers. *Proc. Natl Acad. Sci. USA* **105,** 14867–14872 (2008).
19. Park, S. M., Gaur, A. B., Lengyel, E. & Peter, M. E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev.* **22,** 894–907 (2008).
20. Gregory, P. A. *et al.* The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nature Cell Biol.* **10,** 593–601 (2008).
21. Singh, A. & Settleman, J. EMT, cancer stem cells and drug resistance: an emerging axis of evil in the war on cancer. *Oncogene* **29,** 4741–4751 (2010).
22. Zhang, Y., Toy, K. A. & Kleer, C. G. Metaplastic breast carcinomas are enriched in markers of tumor-initiating cells and epithelial to mesenchymal transition. *Mod. Pathol.* **25,** 178–184 (2012).
23. Creighton, C. J. *et al.* Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. *Proc. Natl Acad. Sci. USA* **106,** 13820–13825 (2009).
24. von Minckwitz, G. Docetaxel/anthracycline combinations for breast cancer treatment. *Expert Opin. Pharmacother.* **8,** 485–495 (2007).
25. Lau, C. K. *et al.* An Akt/hypoxia-inducible factor-1α/platelet-derived growth factor-BB autocrine loop mediates hypoxia-induced chemoresistance in liver cancer cells and tumorigenic hepatic progenitor cells. *Clin. Cancer Res.* **15,** 3462–3471 (2009).
26. Xiao, Z. M., Wang, X. Y. & Wang, A. M. Periostin induces chemoresistance in colon cancer cells through activation of the PI3K/Akt/survivin pathway. *Biotechnol. Appl. Biochem.* **62,** 401–406 (2015).
27. Yamada, D. *et al.* Role of crosstalk between interleukin-6 and transforming growth factor-beta 1 in epithelial-mesenchymal transition and chemoresistance in biliary tract cancer. *Eur. J. Cancer* **49,** 1725–1740 (2013).
28. Yao, Z. *et al.* TGF-β IL-6 axis mediates selective and adaptive mechanisms of resistance to molecular targeted therapy in lung cancer. *Proc. Natl Acad. Sci. USA* **107,** 15535–15540 (2010).
29. Russo, J. E. & Hilton, J. Characterization of cytosolic aldehyde dehydrogenase from cyclophosphamide resistant L1210 cells. *Cancer Res.* **48,** 2963–2968 (1988).
30. Tam, W. L. & Weinberg, R. A. The epigenetics of epithelial–mesenchymal plasticity in cancer. *Nature Med.* **19,** 1438–1449 (2013).
31. Trimboli, A. J. *et al.* Direct evidence for epithelial–mesenchymal transitions in breast cancer. *Cancer Res.* **68,** 937–945 (2008).
32. Yu, M. *et al.* Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339,** 580–584 (2013).
33. Korpal, M. *et al.* Direct targeting of Sec23a by miR-200s influences cancer cell secretome and promotes metastatic colonization. *Nature Med.* **17,** 1101–1108 (2011).
34. Gupta, P. B. *et al.* Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138,** 645–659 (2009).
35. Diessner, J. *et al.* Targeting of preexisting and induced breast cancer stem cells with trastuzumab and trastuzumab emtansine (T-DM1). *Cell Death Dis.* **5,** e1149 (2014).

Author Contributions D.G., K.R.F. and V.M. designed the experiments. K.R.F. and D.G. performed the experiments. A.D., S.L., H.C., T.E.R. and S.R. provided technical support with experiments and animal work. L.T.V. and N.K.A. made critical comments to improve the study design. J.S., F.L. and S.T.C.W. performed RNA-sequencing analysis. J.T. and R.F.S. generated the Vim-Cre transgenic mice. K.R.F., D.G. and V.M. wrote and edited the manuscript with input from the other authors. All authors discussed the results and conclusions drawn from them.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to V.M. (vim2010@med.cornell.edu) or D.G. (dig2009@med.cornell.edu).

METHODS

Animals. Wild-type C57BL/6 and FVB/n mice, and transgenic mice with ACTB–tdTomato–eGFP (stock no. 007676), *Fsp1*–Cre (stock no. 012641), MMTV–PyMT (stock no. 002374), and MMTV–Neu (stock no. 002376) were obtained from The Jackson Laboratory. The vimentin–CreER mouse was a kind gift from the laboratory of R. F. Schwabe at Columbia University. CB-17 SCID mice were obtained from Charles River Laboratories. All mouse strains obtained were bred in the animal facility at Weill Cornell Medical College. All animal work was conducted in accordance with a protocol approved by the Institutional Animal Care and Use Committee at Weill Cornell Medical College.

The ACTB–tdTomato–EGFP and *Fsp1*–Cre mice were bred together to obtain double transgenic mice and then bred with MMTV–PyMT or MMTV–Neu mice to obtain the tri-PyMT and tri-Neu triple-transgenic mice, respectively. Double transgenic male mice carrying ACTB–tdTomato–eGFP and MMTV–PyMT were crossed with the vimentin–CreER mice to obtain the tri-PyMT/Vim triple-transgenic mice. Genotyping for each transgenic line was performed following the standardized protocols as described in the website of The Jackson Laboratory. Genotyping for vimentin–CreER was done using forward primer 5′-CCCCTTCCTCACTTCTTTCC and reverse primer 5′-ATGTTTAGCTGGCCCAAATG.

Tamoxifen injection. To induce vimentin–CreER activity in the tri-PyMT/Vim mice, Tamoxifen (Sigma-Aldrich, 2 mg per mouse, dissolved in corn oil) was administered through intraperitoneal injections, three times per week starting when the primary tumours appear (at 8 weeks of age) and continuing for 6 weeks until metastasis developed in the lung.

Establishing tri-PyMT cell line. The primary tumour of the tri-PyMT mouse (12-week-old female) was surgically removed under sterile conditions. Tumour tissue was sliced into ~1 mm^3 blocks and implanted into the fat pad (no. 4 on the right side) of CB-17 SCID mice. The secondary tumour was used to establish the tri-PyMT cell line, eliminating the contamination of fluorescent positive stromal cells in the tumour tissue from tri-PyMT transgenic mice.

Tumour tissue was minced and digested with an enzyme cocktail (Collagenase A, elastase, and DNase I, Roche Applied Science) in HBSS buffer at 37 °C for 30 min. The cell suspension was strained through a 40-μm cell strainer (BD Biosciences). Cells were washed with PBS three times and uploaded in the Aria III cell sorter (BD Biosciences). The sorted RFP$^+$ cells were cultured in DMEM supplemented with 10% fetal bovine serum. The PyMT oncogene expression in the established cell line was confirmed by RT–PCR (Extended Data Fig. 4c). The tumorigenic ability of these cells was confirmed throughout the study.

To determine EMT induction by TGF-β, cells were cultured for one week in DMEM with 2% FBS and 2 ng ml^{-1} TGF-β1 (R&D Systems). The GFP$^+$ cell ratio was quantified by flow cytometry.

To generate the miR-200 overexpressing cell line, a pLenti 4.1 Ex miR-200b-200a-429 construct[20], was obtained from Addgene. To eliminate the contamination of fluorescent marker expression in targeted cells, the GFP gene in this construct was removed by BstBI/XbaI digestion followed by blunted self-ligation. Lentivirus was packaged by co-transfection of the pLenti-miR-200 construct and packaging plasmids into HEK293T cells. tri-PyMT cells (passage 2) were infected with the lentivirus. Infected cells (tri-PyMT miR-200) were selected by culturing with puromycin (2 μg ml^{-1}) for 14 days. A control tri-PyMT cell line was generated by infecting cells with lentivirus carrying the puromycin resistance gene, following the same procedure in parallel.

Orthotopic breast tumour model. To establish an orthotopic breast tumour model, we first purified RFP$^+$ cells from passages 10–15 of tri-PyMT cell culture by FACS. The purified RFP$^+$ tri-PyMT cells ($1 × 10^6$ cells with purity >99%, Extended Data Fig. 5a) were injected into the mammary fat pad of 8-week-old female CB-17 SCID mice. The growth of the primary tumour was monitored by external calliper measurement once a week. In approximately 4 weeks, the primary tumour was surgically removed and the incision was closed with wound clips. The tumour size did not exceed 5% of total body weight as permitted in the IACUC protocol. Animals were euthanized 4 weeks after primary tumour removal to analyse the development of pulmonary metastasis. For animals subjected to chemotherapy, Cyclophosphamide (CTX, Sigma-Aldrich, 100 mg kg^{-1}) was administered once per week, for 2 weeks prior and 2 weeks after surgery.

Tissue processing, immunofluorescence and microscopy. The harvested primary tumours and PBS-perfused lungs bearing metastases were fixed in 4% paraformaldehyde overnight, followed by 30% sucrose for 2 days, and then embedded in Tissue-tek O.C.T. embedding compound (Electron Microscopy Sciences). Serial sections (10 μm, at least 10 sections) were prepared for histological analysis by haematoxylin and eosin staining, and immunofluorescent staining following standardized protocols.

Primary antibodies used in this study include CD45 (30-F11, BioLegend), E-cadherin (DECMA-1, BioLegend), vimentin (sc-7557, Santa Cruz), PyMT (ab15085, Abcam), Neu (sc-284, Santa Cruz), Ki67 (ab15580, Abcam), and active caspase-3 (C92-605, BD Pharmingen). Primary antibodies were directly conjugated to Alexa Fluor 647 using an antibody labelling kit (Invitrogen) performed as per manufacturer's instructions and purified over BioSpin P30 columns (Bio-Rad). GFP$^+$ and RFP$^+$ cells were detected by inherent fluorescence.

Fluorescent images were obtained using a computerized Zeiss fluorescent microscope (Axiovert 200M), fitted with an apotome and an HRM camera. Images were analysed using Axiovision 4.6 software (Carl Zeiss).

Flow cytometry and cell sorting. For the metastatic lungs and primary tumours, cell suspensions were prepared by digesting tissues with an enzyme cocktail (collagenase A, elastase, and DNase I, Roche Applied Science) in HBSS buffer at 37 °C for 30 min. For cultured cells, cells were collected through trypsinization. A single-cell suspension was prepared by filtering through a 30-μm cell strainer (BD Biosciences). Then cells were stained following a standard immunostaining protocol. In brief, cells were pre-blocked with 2% FBS plus Fc block (CD16/CD32, 1:30, BD Biosciences) and then incubated with the primary antibody against E-cadherin (DECMA-1, BioLegend). SYTOX Blue (Invitrogen) was added to the staining tube in the last 5 min to facilitate the elimination of dead cells. GFP$^+$ and RFP$^+$ cells were detected by their intrinsic signals. The stained samples were analysed using the LSRII flow cytometer coupled with FACS Diva software (BD Biosciences). Flow cytometry analysis was performed using a variety of controls including isotype antibodies, unstained and single-colour stained samples for determining appropriate gates, voltages and compensations required in multivariate flow cytometry.

For sorting live cells back for further culturing or injection into animals, we used the Aria II cell sorter coupled with FACS Diva software (BD Biosciences). The preparation of cells for sorting was performed under sterile conditions. The purity of subpopulations after sorting was confirmed by analysing post-sort samples in the sorter again.

Quantitative RT–PCR analysis. Total RNA was extracted by using the RNeasy Kit (Qiagen), and miRNA via the mirVana miRNA isolation kit (Life Technologies), and converted to cDNA using qScript cDNA SuperMix (Quanta Biosciences) and RT–PCR. qPCR was performed with the appropriate primers (sequences shown in the table) and iQ^{TM} SYBR Green master mix (Bio-Rad). PCR protocol: initial denaturing at 95 °C for 3 min, 40 cycles of 95 °C for 20 s, 60 °C for 30 s, and 72 °C for 30 s, followed by final extension at 72 °C for 5 min and melt curve analysis was applied on a Bio-Rad CFX96 Real Time System (Bio-Rad) coupled with Bio-Rad-CFX Manager software. Primers used are as follows: *GAPDH*, forward, 5′-GGTCCTCAGTGTAGCCCAAG-3′; reverse 5′-AATGTGTC CGTCGTGGATCT-3′; *Cdh1* (E-cadherin), forward, 5′-ACACCGATGGTGAGGG TACACAGG-3′; reverse, 5′-GCCGCCACACACAGCATAGTCTC-3′; *Ocln*, forward, 5′-TGCTAAGGCAGTTTTGGCTAAGTCT-3′, reverse, 5′-AAAA ACAGTGGTGGGGAACGTG-3′; *Vim*, forward, 5′-TGACCTCTCTGAGG CTGCCAACC-3′; reverse, 5′-TTCCATCTCACGCATCTGGCGCTC-3′; *Cdh2* (N-cadherin), forward, 5′-AAAGAGCGCCAAGCCAAGCAGC-3′; reverse, 5′-TGCGGATCGGACTGGGTACTGTG-3′; *FSP-1*, forward, 5′-CCTG TCCTGCATTGCCATGAT-3′, reverse, 5′-CCCACTGGCAAACTACACCC-3′; *Snai1*, forward, 5′-ACTGGTGAGAAGCCATTCTCCT-3′; reverse, 5′-CTGGC ACTGGTATCTCTTCACA-3′; *Snai2*, forward, 5′-TTGCAGACAGATCA AACCTGAG-3′; reverse, 5′-TGTTTATGCAGAAGCGACATTC-3′; *Twist1*, forward, 5′-AGCTACGCCTTCTCCGTCTG-3′; reverse, 5′-CTCCTTCT CTGGAAACAATGACA-3′; *Zeb-1*, forward, 5′-GATTCCCCAAGTGGC ATATACA-3′; reverse, 5′-TGGAGACTCCTTCTGAGCTAGTG-3′; *Zeb-2*, forward, 5′-TGGATCAGATGAGCTTCCTACC-3′; reverse, 5′-AGCAA GTCTCCCTGAAATCCTT-3′; *PyMT*, forward, 5′-ACTGCTACTGCA CCCAGACA-3′; reverse, 5′-CTGGAAGCCGGTTCCTCCTA-3′; *GFP*, forward, 5′-CCACATGAAGCAGCACGACT-3′; reverse, 5′-GGGTCTTG TAGTTGCCGTCG-3′; *RFP*, forward, 5′-AGCGCGTGATGAACTTCGAG-3′; reverse, 5′-CCGCGCATCTTCACCTTGTA-3′.

RNA-sequencing analysis. Total RNA was extracted from sorted RFP$^+$ and GFP$^+$ tri-PyMT cells with the RNeasy Kit (Qiagen). RNA-seq libraries was constructed and sequenced following standard protocols (Illumina). Single-end RNA-seq reads were mapped to UCSC mouse genome (GRCm38/mm10) using Tophat2. FPKM values for each gene were estimated by Cufflinks and statistical analysis was done using Cuffdiff2. Heat maps for differentially expressed genes with adjusted *P* values <0.05 were drawn using gplots R package.

Western blot analysis. Cells were homogenized in 1× RIPA lysis buffer (Millipore) with protease inhibitors (Roche Applied Science). Samples were boiled in 1× Laemmli buffer and 10% β-mercaptoethanol, and loaded onto 12% gradient Tris-glycine gels (Bio-Rad). Western blotting was performed using antibodies specific for E-cadherin (clone DECMA-1), vimentin (clone RV202, BD Pharmingen), and β-actin (clone AC-15, Sigma-Aldrich).

Cell apoptosis and viability assays. To determine apoptosis of RFP$^+$ and GFP$^+$ cells, tri-PyMT cells (Passage 10) were seeded on adherent six-well plates

(1 × 10⁶ cells), and treated with 4-hydroperoxy cyclophosphamide (Santa Cruz) for 48 h. After treatment, cells were trypsinized and stained with APC-conjugated Annexin V (BD Biosciences) and SYTOX Blue (Invitrogen) for apoptotic-cell labelling. The stained cells were analysed in the LSRII flow cytometer to quantify the percentage of apoptotic, dead, and live RFP⁺ and GFP⁺ cells by FACS Diva software. To determine the viability of tri-PyMT control and miR-200-expressing cells treated with CTX, cells were plated in 96-well adherent black-walled plates (1 × 10⁴ cells), and treated with 4-hydroperoxy cyclophosphamide for 48 h. After treatment, cell viability was measured with the CellTiter-Glo Luminescent Cell Viability Assay (Promega).

Cell migration assay. 1 × 10⁵ tri-PyMT cells were seeded in a six-well plate. Real-time images of cells (including phase, GFP and RFP channels) were taken under a computerized Zeiss microscope (Axiovert observation) every 10 min for 10 h. Movement of individual cells (>10 RFP⁺ and >10 GFP⁺ cells in each field, >2 fields were analysed) were tracked with ImageJ software, and the distance that was travelled during that time was measured as indicated.

ALDH activity assay. RFP⁺ and GFP⁺ tri-PyMT cells (1 × 10⁶ cells each) were freshly sorted from culture by FACS and then homogenized in cold ALDH Assay buffer provided in the ALDH Activity Colorimetric Assay Kit (Biovision Inc.) Following the protocol, ALDH substrate and acetaldehyde were added. ALDH activities in samples were measured by OD at 450 nm in kinetic mode (every 3 min for 60 min).

Statistical analysis. To determine the sample size of animal experiments, we used power analysis assuming $\frac{\text{difference in means}}{\text{standard deviation}} > 2.5$. Therefore, all animal experiments were conducted with ≥5 mice per group to ensure adequate power between groups by two-sample t-test comparison. Animals were randomized within each experimental group. No blinding was applied in performing experiments. Results are expressed as mean ± s.e.m. Data distribution in groups and significance between different treatment groups was analysed by using the Mann–Whitney U-test in GraphPad Prism software. P values <0.05 were considered significant. Error bars depict s.e.m., except where indicated otherwise.
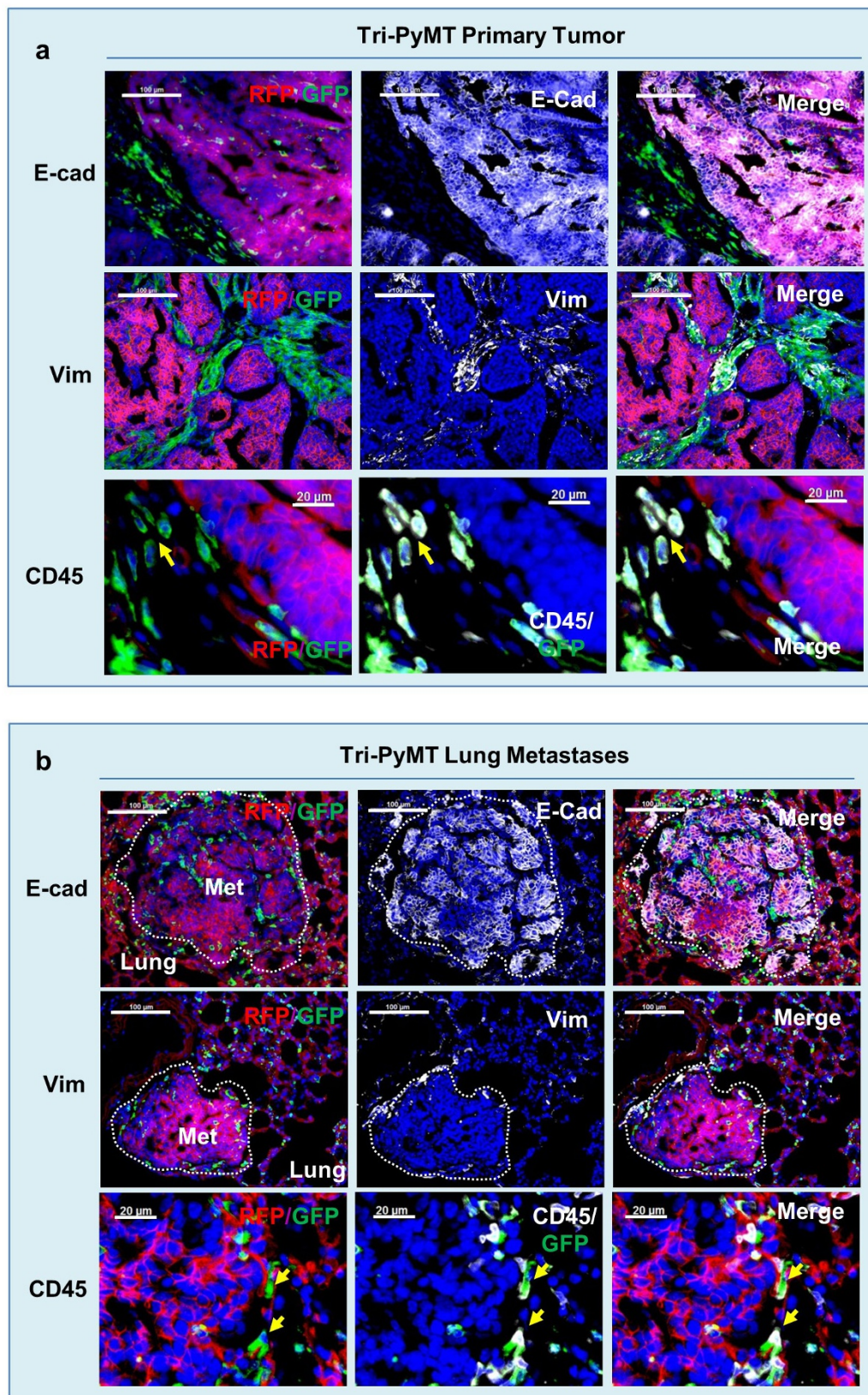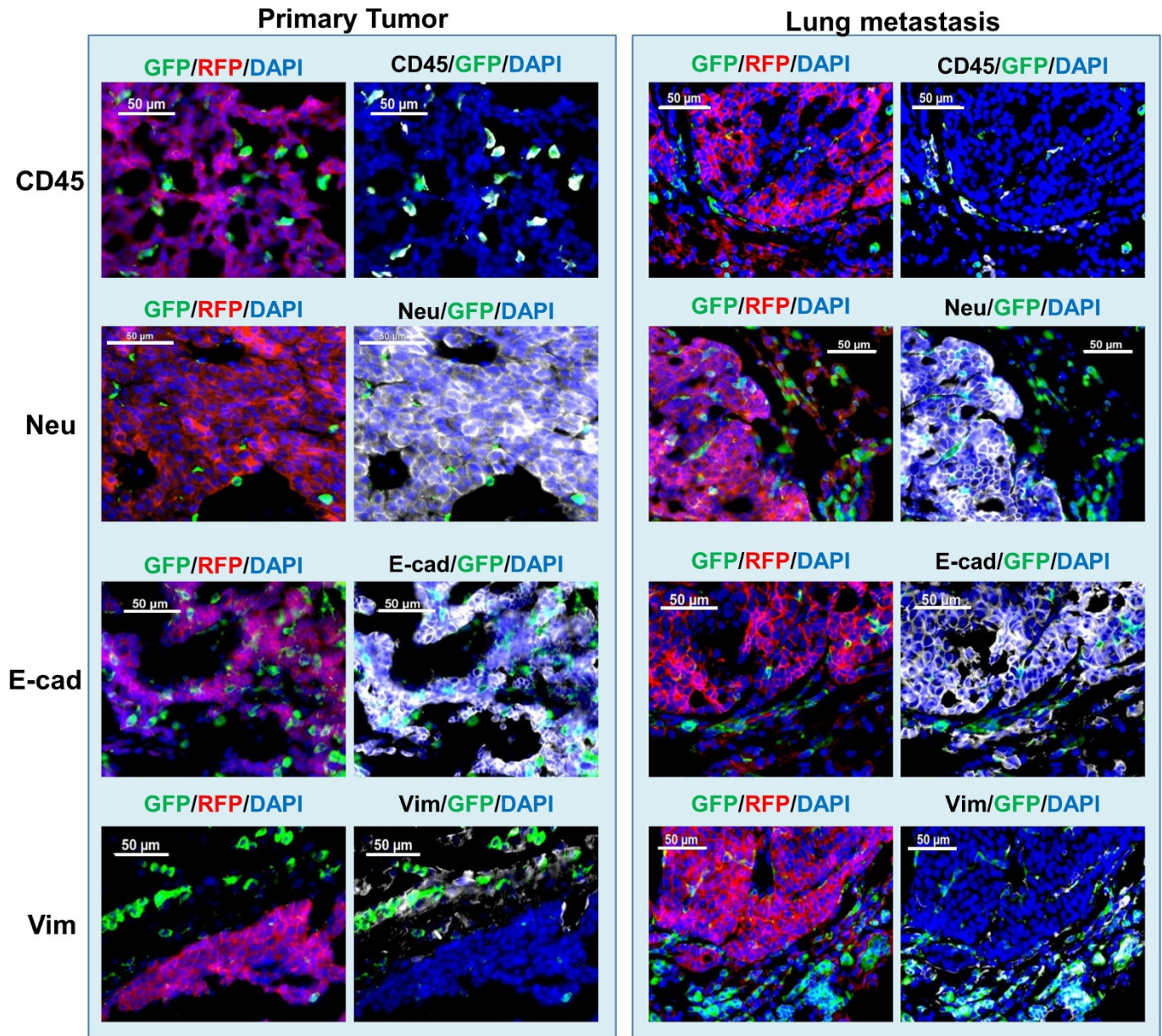
Extended Data Figure 1 | Characterization of the primary tumour and lung metastasis of tri-PyMT mice. a, b, Sections of primary tumours (**a**) and lungs (**b**) from tri-PyMT mice were immunostained for E-cadherin (E-cad, top), vimentin (Vim, middle) and CD45 (bottom) in white pseudo-colour. Representative images are shown ($n > 5$ mice). Note the co-localization of PyMT with RFP, and CD45 with GFP (as indicated by arrows), in both primary tumours and lung metastases.
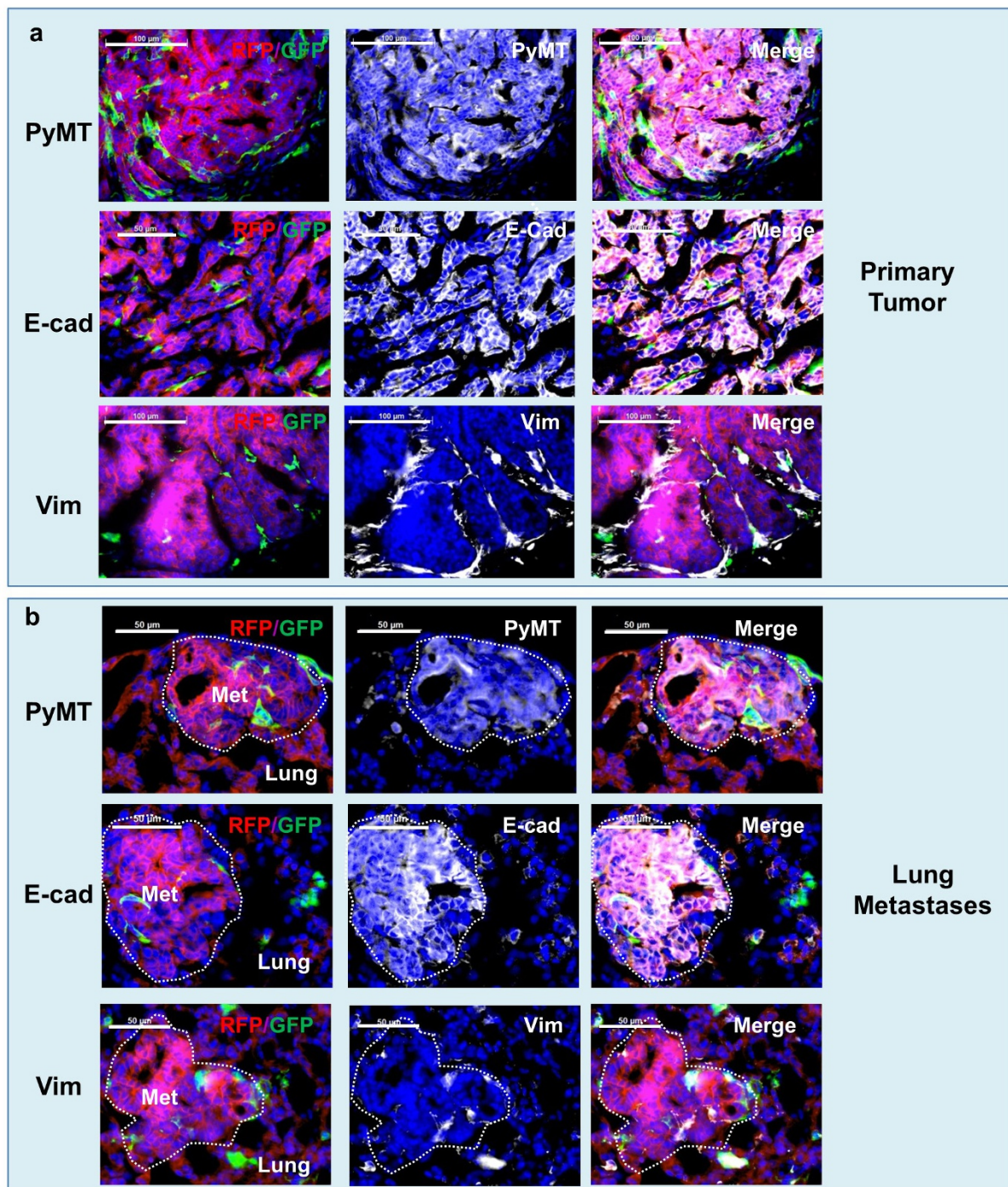
Tri-Neu mouse (MMTV-Neu/FSP1-Cre/Rosa-RGFP)

Primary Tumor

Lung metastasis

Extended Data Figure 2 | Characterization of the primary tumour and lung metastasis of tri-Neu mice. Sections of primary tumours (left panel) and lungs (right panel) from tri-Neu mice were immunostained for CD45, Neu, E-cadherin, and vimentin (in white pseudo-colour). Representative images are shown ($n > 5$ mice). Note that both primary tumours and lung metastases are largely composed of epithelial RFP$^+$ tumour cells.

Tri-PyMT/Vim mouse (MMTV-PyMT/Vim-creER/Rosa-RGFP)

Extended Data Figure 3 | Characterization of the primary tumour and lung metastasis of tri-PyMT/Vim mice. Tri-PyMT/Vim mice were obtained by crossing MMTV–PyMT, vimentin–Cre and Rosa26–RFP–GFP transgenic mice. **a, b,** Sections of primary tumours (**a**) and lungs (**b**) were immunostained for PyMT, E-cadherin and vimentin (in white pseudo-colour). Representative images are shown ($n > 5$ mice). Note that both primary tumours and lung metastases are largely composed of epithelial RFP$^+$ tumour cells.
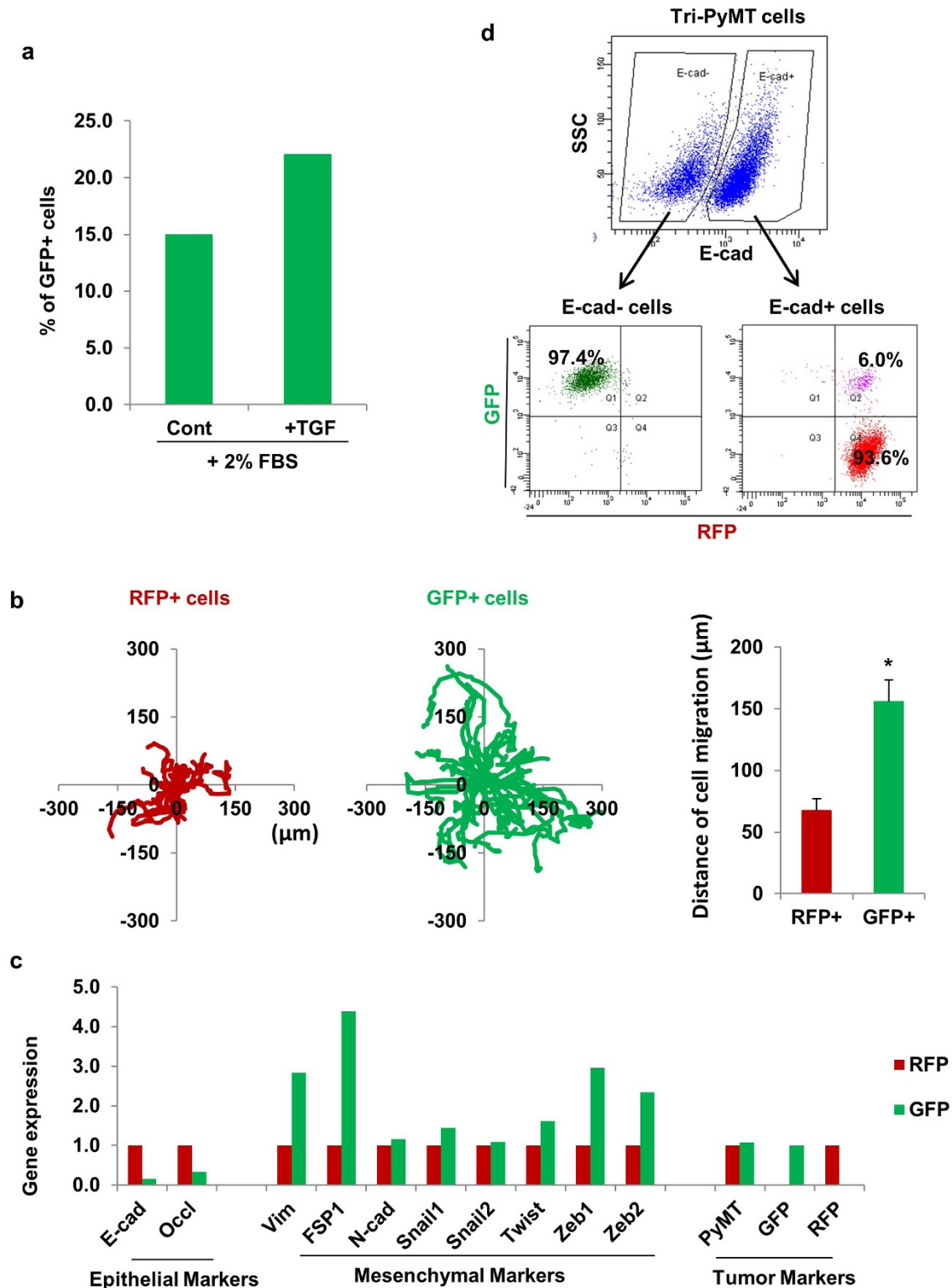
Extended Data Figure 4 | Characterization of tri-PyMT cells. a, EMT of tri-PyMT cells with TGF. RFP[+] tri-PyMT cells were sorted by flow cytometry and cultured in medium containing 2% FBS with or without TGF-β1 (2 ng ml[−1]) for 3 days. Plot shows quantification of the percentage of GFP[+] cells analysed by flow cytometry ($n = 2$ biological replicates). **b**, Cell migration assay of tri-PyMT cells. The tracing plots show the movement of individual RFP[+] and GFP[+] cells in 10 h of live imaging. Quantification plot (right panel) showed the average distance that RFP[+] and GFP[+] cells have moved during the time frame ($n > 20$, *$P < 0.01$). **c**, Relative expression of epithelial, mesenchymal and tumour markers in

sorted RFP[+] and GFP[+] tri-PyMT cells as determined by qRT–PCR with *Gapdh* as the internal control. $n = 2$ individual experiments. **d**, EMT of tri-PyMT cells is reported by fluorescent marker switch. Flow cytometry plot shows E-cadherin[−] (E-cad[−]) and E-cadherin[+] (E-cad[+]) subpopulations of tri-PyMT cells (upper panel). Of the E-cad[−] and E-cad[+] subsets, the populations were further dissected according to innate fluorescence (lower panel). Numbers indicate the percentage of GFP[+], RFP[+], or transitioning (Q2) cells in the parental E-cad[−] or E-cad[+] subsets, respectively.
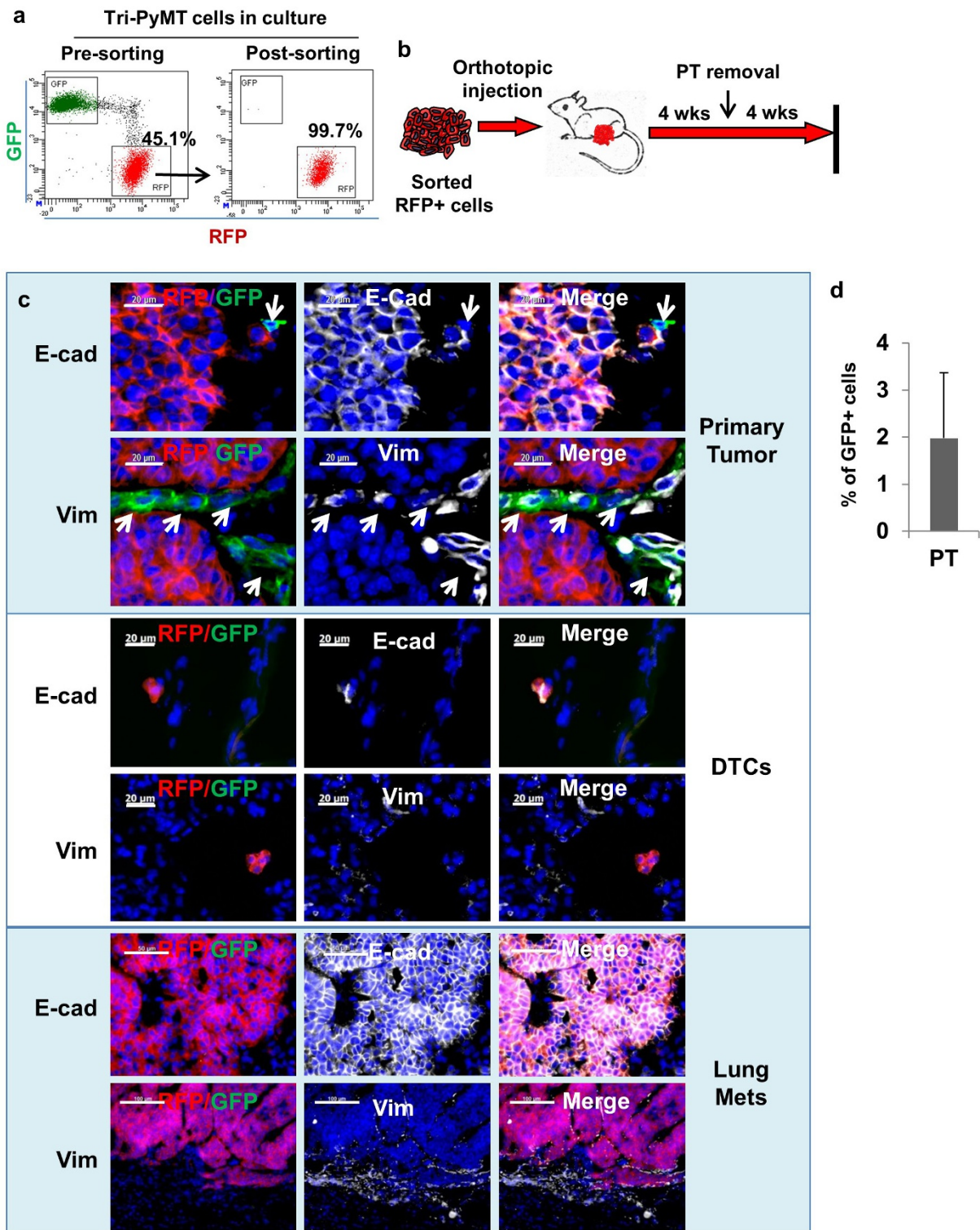
Extended Data Figure 5 | Establishing an orthotopic model with sorted RFP+ tri-PyMT cells. a, Flow cytometry plots show tri-PyMT cells before and after sorting for RFP+ cells. Numbers indicate the percentage and purity of RFP+ cells used for establishing orthotopic breast tumours in mice. **b,** Schematic of the orthotopic breast tumour model with sorted RFP+ tri-PyMT cells. Cells are injected into the mammary gland of wild-type mice to generate primary breast tumours, resection of primary tumour at 4 weeks and lung metastases evaluation in another 4 weeks. **c,** Characterization of tumour cells in the primary tumour, disseminated tumour cells (DTCs) and tumour cells in the lung metastasis of the tri-PyMT orthotopic model. Sections of primary tumours and lungs from tri-PyMT orthotopic mice were immunostained for E-cadherin and vimentin (in white pseudo-colour). Essentially all RFP+ tumour cells are detected as E-cad+/Vim−, while the scattered GFP+ tumour cells in the primary tumour are E-cad−/Vim+ (as indicated by arrows in the top panel). Representative images are shown ($n = 8$). **d,** Plot shows the percentage of GFP+ cells out of total tumour cells (GFP+ plus RFP+, $n = 6$).
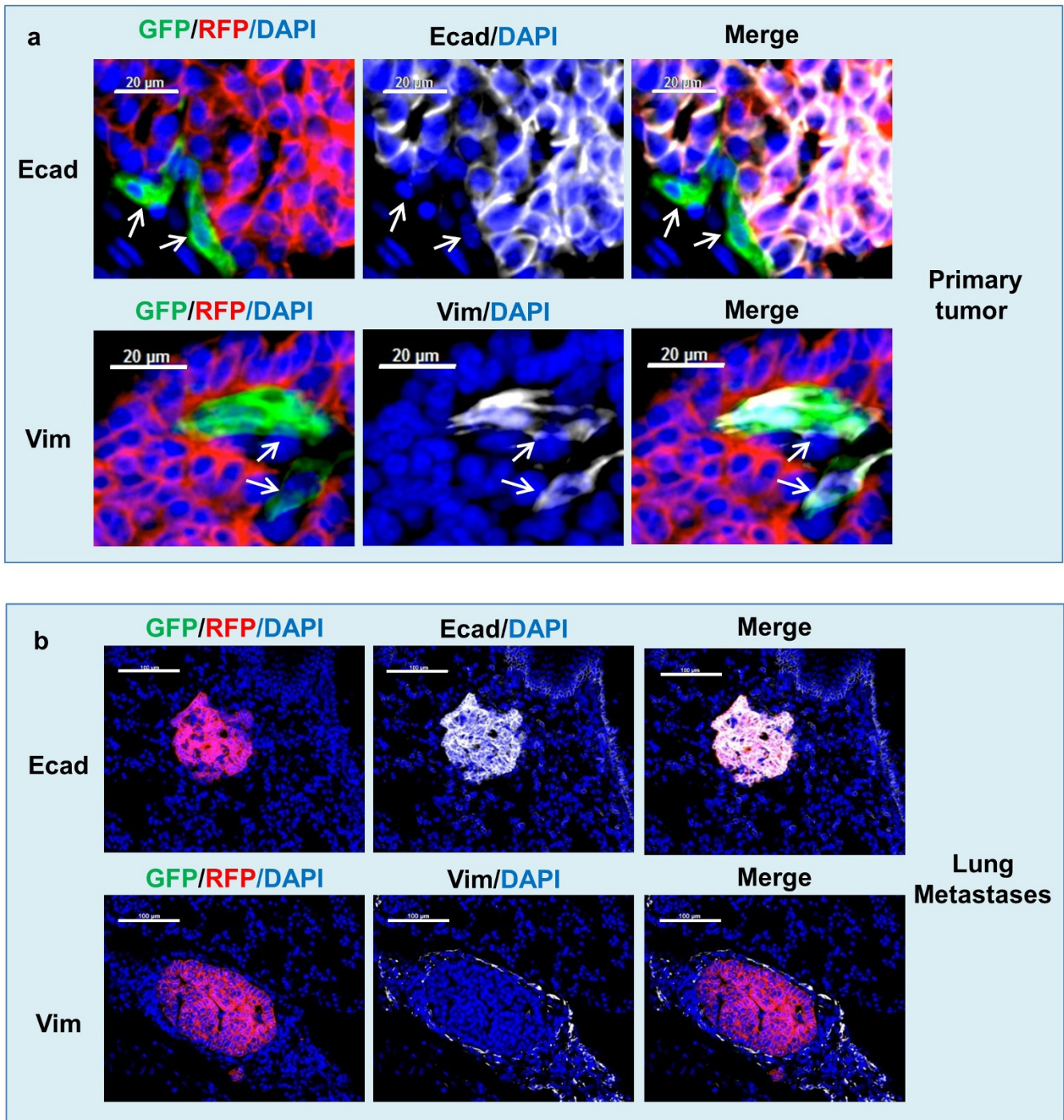
Extended Data Figure 6 | Characterization of EMT status of orthotopic tri-Vim–PyMT primary tumours. a, b, Sections of tri-Vim–PyMT orthotopic primary tumours (**a**) and metastatic lung (**b**) were immunostained for E-cadherin and vimentin (in white pseudo-colour). As expected, RFP[+] tumour cells are entirely E-cadherin-positive and vimentin-negative, GFP[+] tumour cells are vimentin-positive and E-cadherin-negative, and lung metastases are epithelial and RFP[+].
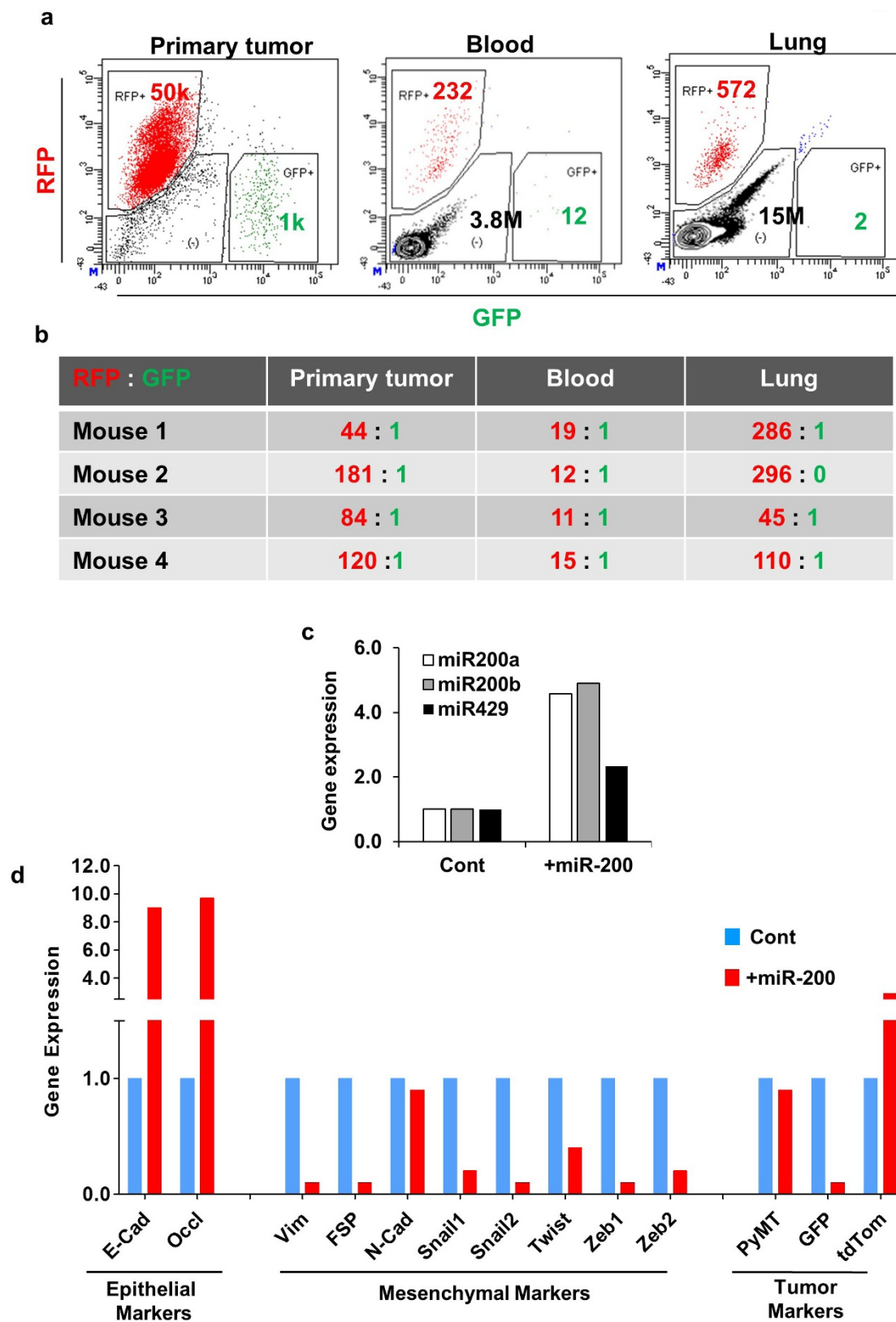
**Extended Data Figure 7 | Dissemination of tri-PyMT cells *in vivo*.**
a, Disseminated tumour cells are RFP+ and epithelial. RFP+ tri-PyMT cells were injected into the fat pad of mice. The fluorescence of the primary tumour, circulating tumour cells in the blood and disseminated tumour cells in the lung were analysed by flow cytometry. The flow cytometry plots depicted are the enumeration of RFP+ and GFP+ cells. **b**, The ratios of detected RFP+ versus GFP+ cells are shown in the chart ($n = 4$ mice). **c**, Relative expression of miR-200-family microRNAs in tri-PyMT control and miR-200-expressing cells. $n = 2$ individual experiments. **d**, Relative expression of EMT markers and tumour markers in tri-PyMT control and mir-200-expressing cells as determined by qRT–PCR with *Gapdh* as the internal control, $n = 2$ individual experiments.
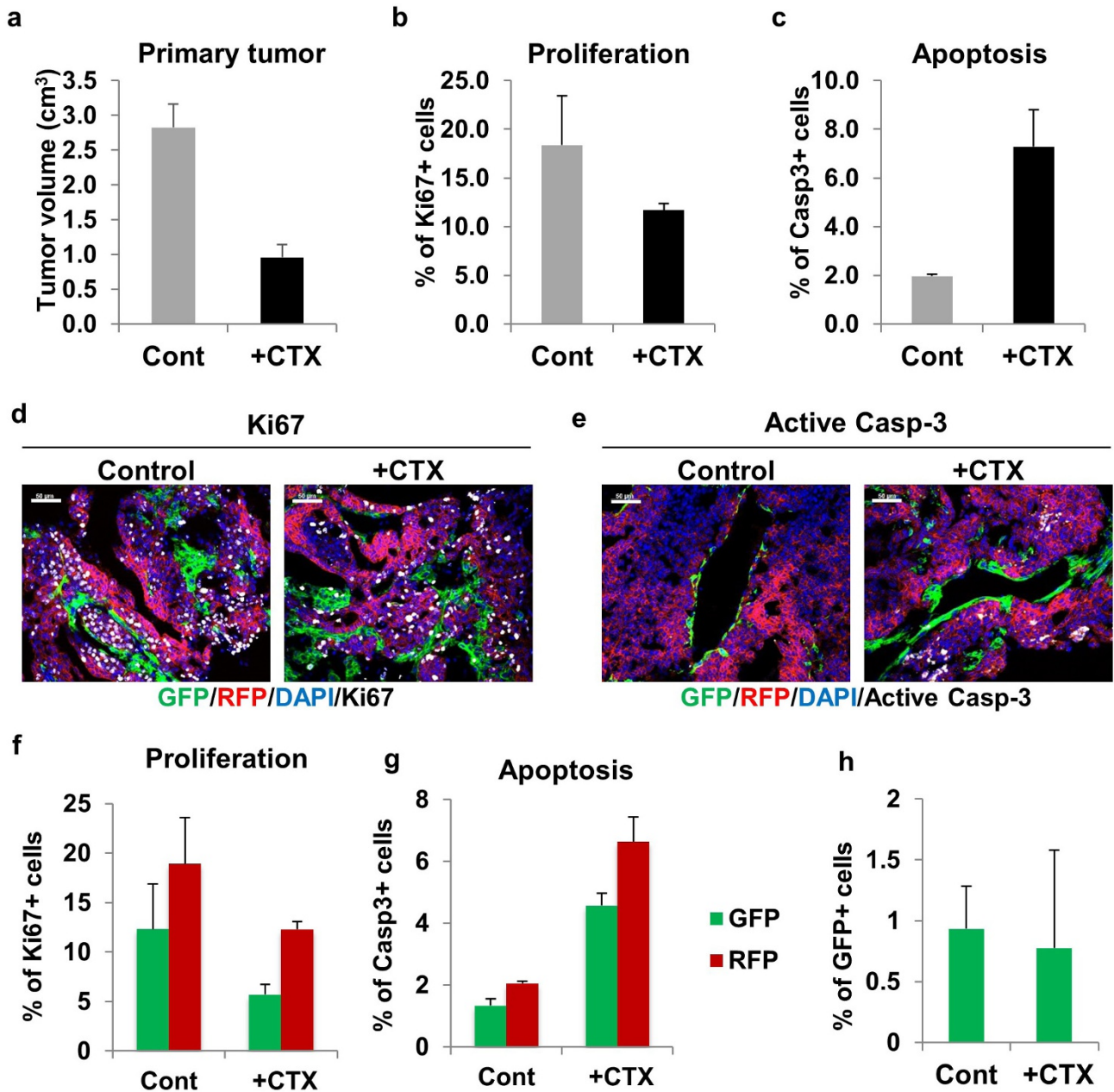
Extended Data Figure 8 | Effects of CTX therapy on primary tumours.
a, Quantification of primary tumour growth after 2 weeks of CTX therapy. For tumour growth data see accompanying Source Data. **b**, Proliferation status of primary tumour cells as detected by Ki67 staining in control mice and after 2 weeks of CTX therapy. **c**, Level of apoptosis in primary tumours as detected by active caspase-3 staining in control mice and after 2 weeks of CTX therapy. **d**, **e**, Representative images of Ki67 (**d**) and active caspase-3 staining (**e**) (white pseudo-colour) of primary tumours in control mice and CTX-treated mice. Scale bars, 50 μm. **f**, Proliferation status of RFP[+] and GFP[+] primary tumour cells as detected by Ki67 staining in control and CTX-treated mice. **g**, Level of apoptosis in RFP[+] and GFP[+] primary tumours as detected by active caspase-3 staining in control and CTX-treated mice. **h**, Percentage of GFP[+] tumour cells in control and CTX-treated primary tumours. $n = 3$ mice for all figures described above. Quantification performed using ImageJ software.
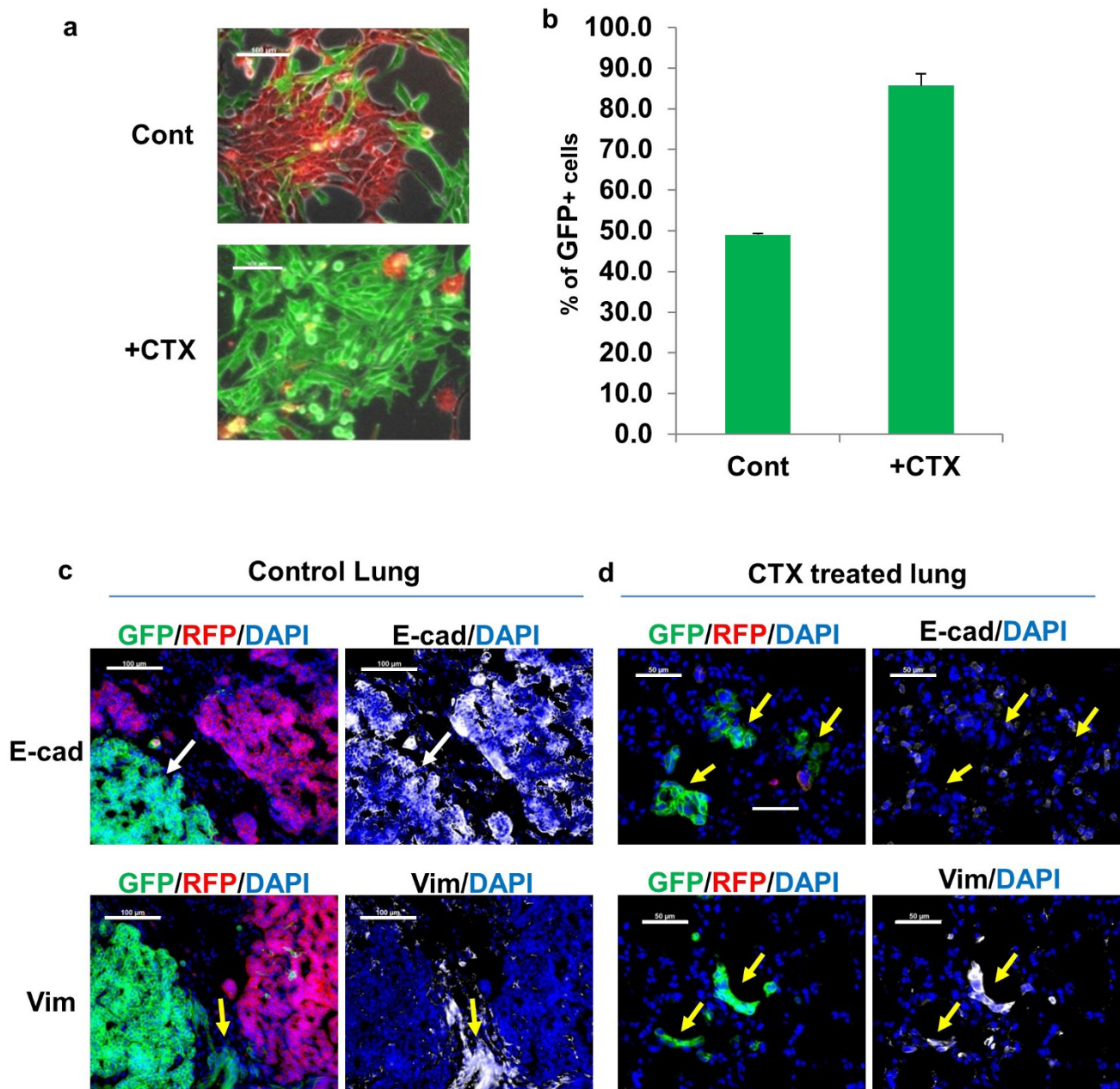
Extended Data Figure 9 | EMT tumour cells are resistant to CTX treatment both *in vitro* and *in vivo*. a, b, Long-term CTX treatment *in vitro* results in a GFP$^+$ population. Tri-PyMT cells were subjected to 2 weeks cyclophosphamide (+CTX) treatment (4 µM). Fluorescent imaging (**a**) and flow cytometry (quantified, **b**, $n = 3$) exhibit the percentage of GFP$^+$ cells in the CTX-treated culture compared to untreated control cells **c, d,** EMT status of lung nodules in competitive survival assay. Representative fluorescent images of tri-PyMT lung metastases in untreated control lungs (**c**) and CTX-treated lungs (**d**), depicting RFP$^+$ and GFP$^+$ tumour cells. Immunostaining showing E-cadherin (E-cad) or vimentin (Vim) in white pseudo-colour. White arrow indicates GFP$^+$ tumour cells with epithelial phenotypes (E-cad$^+$/Vim$^-$), while the yellow arrow indicates GFP$^+$ cells with mesenchymal phenotypes (E-cad$^-$/Vim$^+$). Nuclei were counter-stained with DAPI. $n = 5$.
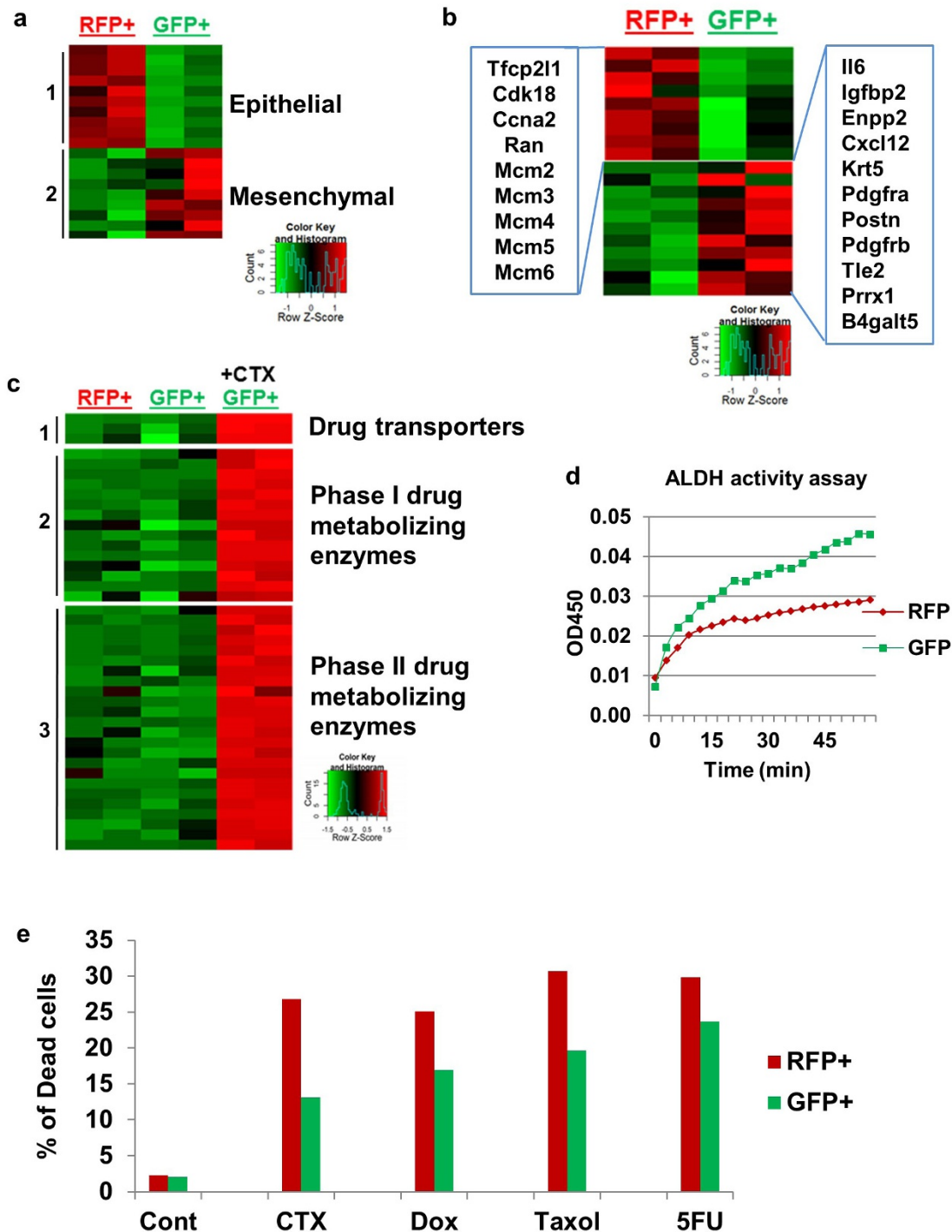
Extended Data Figure 10 | Gene expression profile analysis of RFP+ and GFP+ tri-PyMT cells. RFP+ and GFP+ tri-PyMT cells were sorted by flow cytometry and subjected to transcriptomic analysis by RNA-sequencing. **a**, Heat map of differentially expressed genes (adjusted $P < 0.05$) from RNA-seq of sorted RFP+ and GFP+ tri-PyMT cells, biologically duplicated. Genes that are established epithelial markers (Group 1) include *Cdh1* (which encodes E-cad), *Dsp, Epcam, Fgfbp1, Krt18, Krt19, Ocln, Tjp3, Krt14* and *Tjp2*; the mesenchymal markers (Group 2) include *Cdh2* (which encodes N-cad), *Col23a1, Col3a1, Col5a1, Col6a2, Fsp1, Mmp3, Wnt5a* and *Zeb1*. **b**, Cell cycle (left panel) and chemoresistance-related (right panel) genes alternatively regulated in RFP+ and GFP+ cells. **c**, GFP+ tri-PyMT cells were also sorted from CTX-treated (4 μM) samples. Interestingly, a branch of genes related to drug metabolism were significantly elevated in CTX-treated GFP+ cells. Group 1 genes are drug transporters including *Abcb1a, Abcb1b* and *Abcc1*. Group 2 genes are phase I drug-metabolizing enzymes including *Adh7, Aldh1a1, Aldh1a3, Aldh1l1, Aldh1l2, Aldh2,*

Aldh3a1, Aldh3a2, Aldh3b2, Aldh4a1, Cyp1a1, Cyp2f2, Cyp2j6, Ptgs1 and *Ptgs2*. Group 3 genes are phase II drug metabolizing enzymes including *Aox1, Blvrb, Ces2e, Ces2f, Ces2g, Chst1, Ephx1, Fmo1, Gpx2, Gsta3, Gsta4, Gstm2, Gsto1, Gstp1, Gstt3, Maoa, Mgst1, Mgst2, Nat6, Nat9, Nqo1, Pon3, Ugt1a6a* and *Ugt1a7c*. **d**, Aldehyde dehydrogenase (ALDH) activity assay. Cell lysates were prepared from flow cytometry-sorted RFP+ and GFP+ tri-PyMT cells. ALDH activity in samples was measured by OD at 450 nm in a kinetic mode (every 3 min for 60 min). Representative result from two independent experiments depicted. **e**, EMT tumour cells (GFP+ cells) showed resistance to multiple commonly used chemotherapies. Tri-PyMT cells were subjected to treatment with CTX (8 μM), doxorubicin (Dox, 2 μM), paclitaxel (Taxol, 10 μM) and fluorouracil (5FU; 1.6 μM) for 3 days. Flow cytometry analysis of apoptotic cells was performed after Annexin staining. The percentage of dead cells (Annexin+) in RFP+ and GFP+ cells, respectively, was quantified. $n = 2$ biological replicates.