



Transcriptional diversity and bioenergetic shift in human breast cancer metastasis revealed by single-cell RNA sequencing

Ryan T. Davis¹, Kerrigan Blake², Dennis Ma³, Mari B. Ishak Gabra⁴, Grace A. Hernandez¹, Anh T. Phung¹, Ying Yang⁴, Dustin Maurer², Austin E. Y. T. Lefebvre^{5,6}, Hamad Alshetaiwi^{3,7}, Zhengtao Xiao⁸, Juan Liu⁸, Jason W. Locasale⁸, Michelle A. Digman^{2,5,6}, Eric Mjolsness⁹, Mei Kong⁴, Zena Werb¹⁰ and Devon A. Lawson¹✉

Although metastasis remains the cause of most cancer-related mortality, mechanisms governing seeding in distal tissues are poorly understood. Here, we establish a robust method for the identification of global transcriptomic changes in rare metastatic cells during seeding using single-cell RNA sequencing and patient-derived-xenograft models of breast cancer. We find that both primary tumours and micrometastases display transcriptional heterogeneity but micrometastases harbour a distinct transcriptome program conserved across patient-derived-xenograft models that is highly predictive of poor survival of patients. Pathway analysis revealed mitochondrial oxidative phosphorylation as the top pathway upregulated in micrometastases, in contrast to higher levels of glycolytic enzymes in primary tumour cells, which we corroborated by flow cytometric and metabolomic analyses. Pharmacological inhibition of oxidative phosphorylation dramatically attenuated metastatic seeding in the lungs, which demonstrates the functional importance of oxidative phosphorylation in metastasis and highlights its potential as a therapeutic target to prevent metastatic spread in patients with breast cancer.

Despite major advances in the detection and treatment of early stage disease, metastasis remains the cause of nearly all mortality associated with breast cancer^{1,2}. Previous work suggests that metastasis is seeded by rare primary tumour cells with unique biological properties that enable them to surpass each step in the metastatic cascade^{3–5}. Although the properties promoting cell motility and migration have been well studied, the mechanisms governing the seeding and establishment of micrometastasis in distal tissues are poorly understood. This is in part because metastatic seeding cannot be studied in humans and because it is technically challenging to detect and analyse rare cells at this transient stage in animal models. Further insights into the mechanisms driving metastatic seeding are critical to inspire new strategies for the prevention of metastatic spread and reduction in mortality of patients with breast cancer.

We have developed a robust approach for the capture and analysis of individual cancer cells during the seeding of micrometastasis in human patient-derived-xenograft (PDX) models using single-cell RNA sequencing (scRNA-seq) technology. This enabled us to study the role of tumour heterogeneity in metastasis and identify cellular programs upregulated in micrometastatic cells that promote metastatic seeding. We found that micrometastases display a distinct transcriptome program and identified mitochondrial oxidative phosphorylation (OXPHOS) as a top pathway upregulated during metastatic seeding. This was confirmed at the molecular level,

where we demonstrated that micrometastatic cells display increased mitochondrial membrane potential and a distinct metabolic profile. We further found that the pharmacological inhibition of OXPHOS, specifically in cancer cells, attenuates metastatic seeding in the lungs using two different breast cancer experimental metastasis models. These data show that OXPHOS is functionally important for metastasis and suggest that different metabolic strategies may be advantageous for cells to complete discrete stages of the metastatic cascade.

Results

Single-cell RNA sequencing of matched primary tumours and micrometastases. To identify fundamental cellular programs important for seeding in metastatic tissues, we investigated transcriptome programs uniquely expressed by cancer cells during the seeding and establishment of micrometastatic lesions. We utilized PDX models of breast cancer because of their human disease authenticity, including intra-tumour heterogeneity, which is critical for determining the role of cellular heterogeneity in metastasis^{6–9}. We analysed three previously established PDX models of triple-negative breast cancer (TNBC): HCl001, HCl002 and HCl010 (Fig. 1a)⁹. As in many patients with breast cancer, metastatic progression is slow and sporadic in these models, where most animals display dispersed micrometastases in the lung and lymph nodes and very low metastatic burden at the endpoint (primary tumour diameter of 20–25 mm; Fig. 1b,c)^{9,10}. This enabled us to investigate the

¹Department of Physiology and Biophysics, University of California, Irvine, Irvine, CA, USA. ²Center for Complex Biological Systems, University of California, Irvine, Irvine, CA, USA. ³Department of Biological Chemistry, University of California, Irvine, Irvine, CA, USA. ⁴Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA, USA. ⁵Biomedical Engineering Department, University of California, Irvine, Irvine, CA, USA. ⁶Laboratory for Fluorescence Dynamics, University of California, Irvine, Irvine, CA, USA. ⁷Department of Pathology, University of Hail, Hail, Saudi Arabia. ⁸Department of Pharmacology and Cancer Biology, Duke University, Durham, NC, USA. ⁹Department of Computer Science, University of California, Irvine, Irvine, CA, USA. ¹⁰Department of Anatomy, University of California, San Francisco, CA, USA. ✉e-mail: dalawson@uci.edu

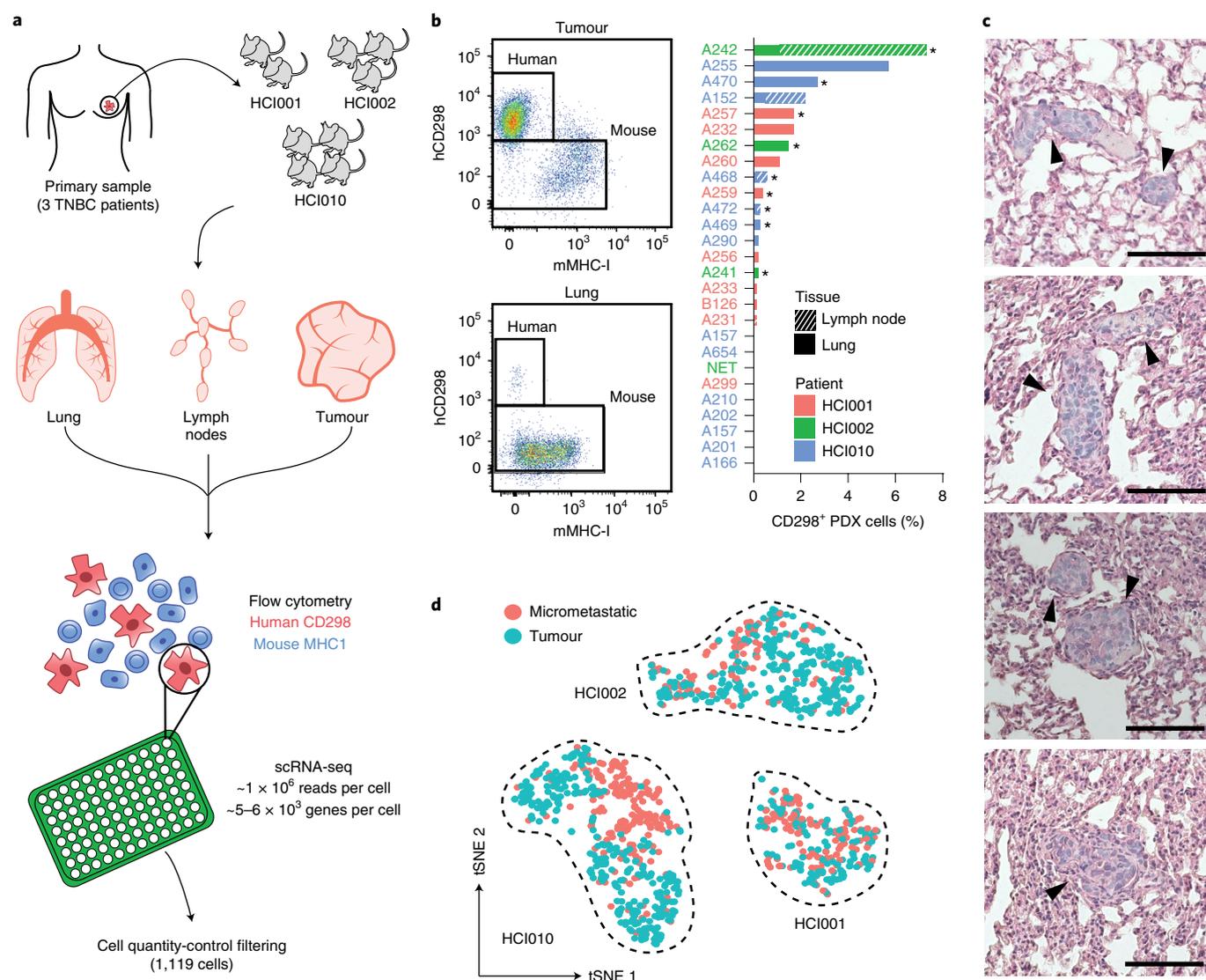


Fig. 1 | Single-cell RNA sequencing of micrometastatic and primary tumour cells. a, Overview of the experimental workflow. The primary tumour, lungs and lymph nodes of each PDX animal were digested to make single-cell suspensions. Single CD298⁺MHC-I⁻ human tumour cells were isolated by flow cytometry, deposited into individual wells of 96-well plates and single-cell cDNA libraries were prepared using Smartseq2 chemistry. Matched primary tumour and micrometastatic cells from nine mice and three PDX models (HCI001, HCI002 and HCI010) were analysed, and 1,119 cells passed quality-control filtering. **b**, Left, flow cytometry-based strategy for the isolation of human CD298⁺MHC-I⁻ cells from micrometastatic (bottom) and primary tumour (top) cells. Right, frequency of CD298⁺MHC-I⁻ cells in a cohort of PDX mice. Mice that were selected for sequencing are indicated with an asterisk. **c**, Haematoxylin and eosin stains of micrometastases in the lungs of HCI001 mice. Data are representative of $n=8$ biologically independent samples. Micrometastatic lesions are indicated by the arrows. Scale bars, 100 μm. **d**, T-distributed stochastic neighbour embedding (tSNE) plot showing clustering of 1,119 metastatic and primary tumour cells from the HCI001, HCI002 and HCI010 models.

transcriptional changes associated with early events in the seeding and establishment of micrometastasis.

We previously developed a robust protocol for the isolation of metastatic cells from PDX models using flow cytometry with human (CD298) and mouse (MHC-I) species-specific antibodies (Fig. 1a,b)¹⁰. Here, we used this approach to compare the cellular diversity in primary tumours and micrometastases by scRNA-seq. We sorted individual cancer cells from the lungs, lymph nodes and primary tumours of PDX mice into 96-well PCR plates (Fig. 1a,b). Matched metastatic and primary tumour cells were isolated from each animal. We optimized our scRNA-seq protocol specifically for flow cytometry-sorted cancer cells^{10,11}. Single-cell libraries were constructed using the Illumina platform and

sequenced at high depth (1×10^6 paired-end reads per cell). In contrast to high-throughput platforms (for example, droplet-based), our deep sequencing protocol enabled us to capture rare metastatic cells and detect more genes per cell¹². In total, we sequenced 1,707 tumour and metastatic cells from nine PDX mice and three tumour models (Fig. 1a,b). The cell libraries were filtered to remove low-quality cells using stringent parameters for low library complexity, mitochondrial gene overrepresentation and noisy gene expression to ensure that technical artefacts did not impact the downstream analysis (Extended Data Fig. 1a,c)^{13,14}. A total of 1,119 cells (66%) passed filtering and were carried forward for further analysis. Importantly, all patient models exhibited a similar library quality (Extended Data Fig. 1a–c).

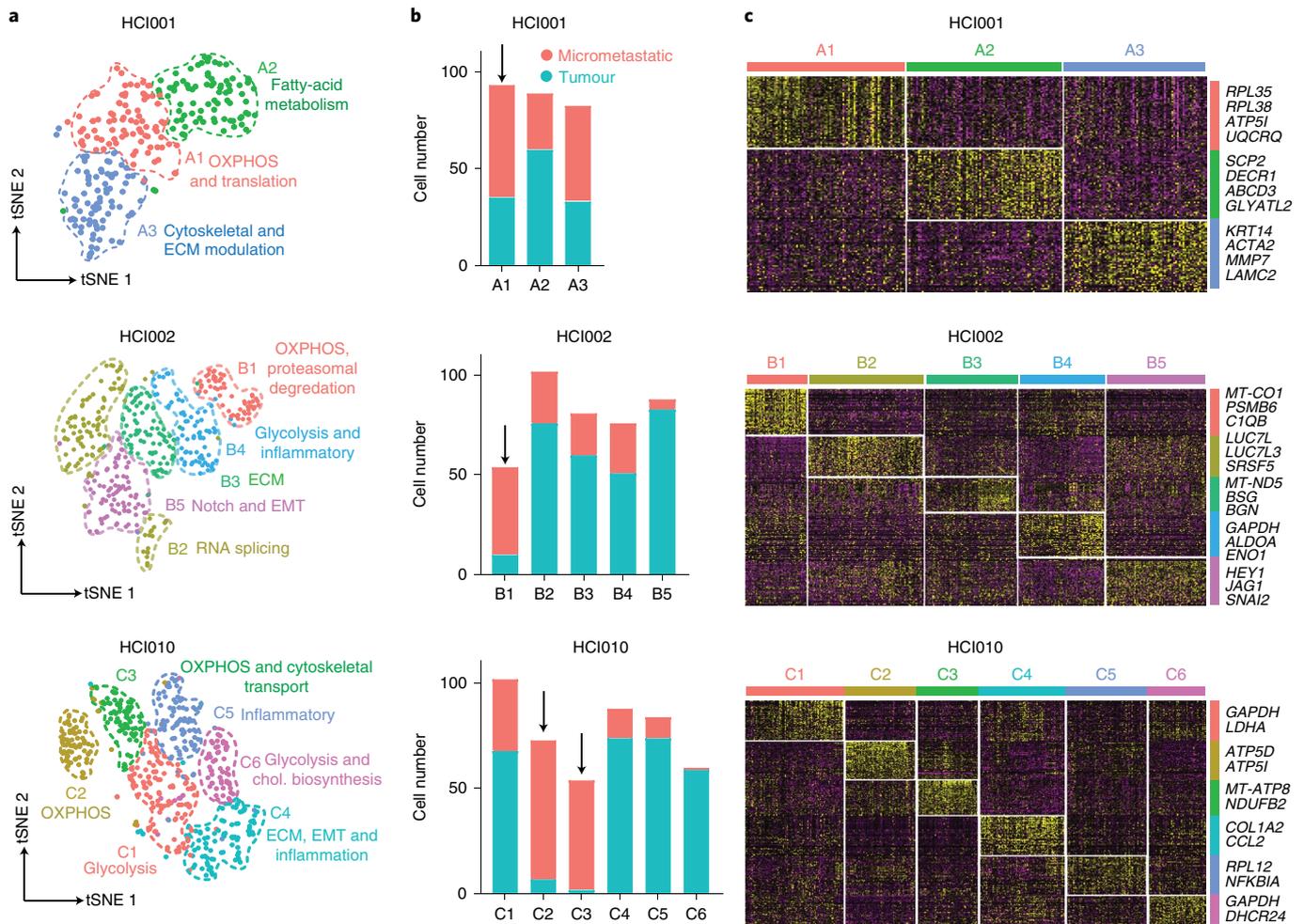


Fig. 2 | Transcriptional diversity in micrometastatic and primary tumour cells. a, Clustering of cells from the HCl001 (top; $n = 247$ cells), HCl002 (middle; $n = 401$ cells) and HCl010 (bottom; $n = 471$ cells) PDX models shown in tSNE plots. The cells are coloured according to their cluster identity. The biological features defining each population identified by GO term analysis of marker genes are indicated. Chol., cholesterol. **b**, Number of micrometastatic and primary tumour cells in each cluster. The arrows indicate clusters enriched for metastatic cells. **c**, Heat maps showing the top marker genes in each cluster. Select marker genes have been highlighted and the full gene lists are provided in Supplementary Table 1.

Transcriptional diversity in micrometastatic and primary tumour cells. To investigate intra- and inter-tumoural diversity, we performed K-nearest-neighbour graph-based clustering, followed by Louvain modularity optimization utilizing the Seurat analysis package^{13,14}. This revealed substantial inter-tumoural heterogeneity, which is consistent with previous bulk and scRNA-seq studies of malignant cells (Fig. 1d)^{9,15–19}. We subsequently analysed each model separately to explore intra-tumoural heterogeneity. Initial clustering in the HCl010 and HCl001 models was primarily driven by cell-cycle status (Extended Data Fig. 1d,e). We consequently regressed out the effects of cell cycle using a previously established method that infers the cell-cycle stage by calculating scores for each cell based on its expression of genes associated with the G1/S or G2/M phases (Extended Data Fig. 1d,e)^{15,16}. The mouse-to-mouse variation within each model was limited after regressing out the effects of cell cycle, suggesting minimal batch effects (Extended Data Fig. 2a). Clustering analysis identified three distinct cell populations in HCl001 (A1–A3), five in HCl002 (B1–B5) and six in HCl010 (C1–C6; Fig. 2a). Interestingly, metastatic and primary tumour cells were found in all clusters, suggesting that metastatic cells display substantial transcriptional diversity even at the early stages of micrometastasis (Fig. 2b and Extended Data Fig. 2b). However, clusters

A1, B1, C2 and C3 from each model contained higher proportions of metastatic cells relative to the other clusters, suggesting a skewing towards this cell state in micrometastasis (Fig. 2b).

We performed marker gene identification and Gene Ontology (GO) analyses using Seurat and Enrichr^{13,14,20–22} to identify cellular properties characteristic of each population. This revealed distinct markers and pathways associated with each cell cluster (Fig. 2c, Extended Data Fig. 2c and Supplementary Tables 1,2). For example, cluster A2 in HCl001 contained a population of cells that expressed high levels of genes associated with fatty-acid metabolism such as *BDH2*, *DECR1* and *CAV1* (Extended Data Fig. 2c,d). B1 in HCl002 expressed high levels of genes encoding proteasome function (Extended Data Fig. 2c,d). A3 in HCl001 contained cells with a robust extracellular-matrix (ECM) modulatory signature, which displayed increased expression of collagens (*COL12A1*, *COL15A1* and *COL16A1*), matrix metalloproteinases (*MMP11* and *MMP16*) and tissue inhibitors of metalloproteinases (*TIMP1*; Extended Data Fig. 2c,d). B3 in HCl002 and C4 in HCl010 showed analogous ECM modulatory gene signatures (Extended Data Fig. 2c,d). C4 also expressed genes associated with epithelial-to-mesenchymal (EMT) transition—such as increased *ZEB1*, *SNAI2* and *VIM* and decreased *EPCAM*, *CDH1* and *KRT19* expression

(Extended Data Fig. 2c,d). These data show that PDX models display substantial inter- and intra-tumoural heterogeneity, and distinct populations of cancer cells can be identified in both the primary tumour and micrometastases.

Micrometastatic cells display a distinct transcriptome program.

To determine whether micrometastatic cells possess unique transcriptome programs that may facilitate seeding in distal tissues, we performed supervised analysis to directly compare gene expression in primary tumour and micrometastatic cells (Fig. 3a)^{13,14,23}. This identified 330 differentially expressed genes ($P < 0.05$; $\log[\text{fold change (FC)}] > 0.25$), including 116 genes specifically upregulated in micrometastatic cells conserved in all three PDX models (Fig. 3b and Supplementary Table 3). Top micrometastasis-associated genes included several heat shock proteins (*HSPB1*, *HSPE1* and *HSPA8*), which are protein chaperones upregulated in response to environmental stress that play pleiotropic roles in protein folding, wound healing, antigen presentation, protection from apoptosis, and cellular proliferation and differentiation (Fig. 3c)^{24,25}. Several cyto-keratins (*KRT14*, *KRT16*, *KRT7* and *KRT17*) were also upregulated in micrometastatic cells from all three models as well as *ACTG2*, which encodes a smooth muscle protein involved in cell motility (Fig. 3c). Micrometastatic cells also expressed higher levels of several genes with lesser known roles in metastasis such as *CKB*, *PHLDA2*, *NME1*, *ASHA1*, *NOP16* and *S100A16* (Fig. 3c).

We next performed a series of survival analyses, logistic regression and in situ validation experiments to confirm increased expression of micrometastasis-associated genes in PDX mice and evaluate the disease relevance in patients with breast cancer. We first explored whether increased primary tumour expression of micrometastasis-associated genes is predictive of poor survival in patients with basal-like breast cancer using microarray data from the KM Plotter database (879 patients)²⁶. This was done to determine whether micrometastasis-associated genes are higher in more biologically aggressive tumours and whether they have value as predictive biomarkers for disease progression in patients. Remarkably, we found that 15 of our top-20 micrometastasis-associated genes were significantly predictive of relapse in basal-like patients (two genes were not predictive and the data for three genes were not available; $P < 0.05$, hazard ratio (HR) ≥ 1.4 ; Extended Data Fig. 3a). Combining all 15 genes further increased the power for predicting patient relapse in this cohort, showing a nearly threefold increased risk (HR = 2.73, $P = 2.9 \times 10^{-10}$; Fig. 3d).

We next utilized a stepwise logistic-regression model to identify the top biomarker candidates for micrometastasis in our dataset. We utilized the Akaike information criterion to determine the optimal number of genes to include in each subsampling ($n = 5$) and constructed ten models on equal subsamplings of micrometastatic and primary tumour cells in our dataset (Extended Data Fig. 3b,c). Using this approach, the top candidates will appear the most often in the ten data subsamplings. We found that *PHLDA2* was the top candidate for micrometastatic cells and was present in eight of ten models (Extended Data Fig. 3d).

PHLDA2 is a maternally imprinted gene that regulates placental growth and increases xenograft engraftment and cell invasion in vitro but has limited prior association with breast cancer metastasis^{27,28}. Our survival analysis showed that *PHLDA2* was significantly predictive of increased relapse ($P = 7.4 \times 10^{-3}$; HR = 1.44) in patients with basal-like breast cancer (Extended Data Fig. 3a). We compared the *PHLDA2* transcript levels in primary tumour and micrometastases using high-resolution single-molecule fluorescence in situ hybridization (RNAscope) to determine whether the *PHLDA2* transcripts are upregulated in micrometastatic lesions in situ. We found that the levels of *PHLDA2* transcripts were at least twofold higher in micrometastases relative to primary tumours, validating our approach for identifying genes upregulated in micrometastasis

(Fig. 3e,f). Interestingly, only rare primary tumour cells expressed *PHLDA2*, raising the question of whether it marks pre-metastatic cells (Fig. 3e). These data highlight our dataset as a resource for the identification of potential drivers of metastatic seeding and biomarkers to predict metastatic progression in patients with breast cancer.

Micrometastatic cells upregulate mitochondrial OXPHOS. We performed GO term analysis on our 330-gene signature to identify pathways and cellular properties upregulated in cancer cells during metastatic seeding (Supplementary Table 4). The top GO terms for micrometastatic cells included ‘epidermis development’ (*CALML5*, *KRT17*, *KRT16*, *KRT14* and *KLK5*) and ‘response to unfolded protein’ (*DNAJA1*, *HSPA8*, *HSPB1*, *HSPE1* and *HSPD1*; Fig. 4a and Supplementary Table 4). Interestingly, many top GO terms centred on mitochondrial biology and metabolism and included ‘respiratory electron transport chain’, ‘ATP metabolic process’, ‘mitochondrial transport’ and ‘oxidative phosphorylation’ (Fig. 4a). In contrast, ‘canonical glycolysis’ and ‘pyruvate metabolic process’ were top GO terms for primary tumour cells, suggesting clear metabolic differences between primary tumour and micrometastatic cells (Fig. 4a and Supplementary Table 4). The analysis of the GO term-associated genes showed that 27 genes related to OXPHOS, glycolysis and mitochondrial complexes were differentially expressed ($P < 0.05$; Fig. 4b). Micrometastatic cells expressed higher levels of numerous mitochondrial electron transport chain (ETC) genes, including *NDUFS6*, *NDUFAB1*, *NDUFB2*, *NDUF44*, *UQCRC3* and *COA6* as well as the ATP synthase subunits *ATP5I*, *ATP5G1* and *ATP5J2* (Fig. 4b). Micrometastatic cells also expressed higher levels of the mitochondrial transport genes *TOMM5*, *TOMM6* and *TIMM13* as well as the mitochondrial ribosome genes *MRPL14*, *MRPL55* and *MRPL51*, which translate mitochondrial genes encoding ETC proteins (Fig. 4b). Superoxide dismutase 1 (*SOD1*)—which encodes a mitochondrial enzyme that converts superoxide radicals to O_2 and H_2O_2 , and may thus protect micrometastatic cells from apoptosis due to oxidative stress from increased ETC activity—was also upregulated in micrometastatic cells (Fig. 4b). In contrast, primary tumour cells expressed higher levels of genes associated with glycolysis, including the glycolytic enzymes *ALDOA*, *ALDOB*, *ALDOC*, *PGM1* and *PGK1* (Fig. 4b). Our logistic regression analysis also identified *LDHA*, which promotes aerobic glycolysis by catalysing the conversion of pyruvate to lactate and diverting it from entry into the citric acid cycle (CAC), as the top gene most descriptive of primary tumour cells (Extended Data Fig. 3d).

We evaluated the expression of 1,402 genes associated with 37 metabolic pathways—such as the pentose phosphate pathway, the citric acid cycle and fatty-acid metabolism—to further investigate the metabolic differences between primary tumour cells and micrometastases (Fig. 4c, Extended Data Fig. 4 and Supplementary Table 5). Gene scoring for each pathway showed that glycolysis ($P = 2.20 \times 10^{-16}$) and OXPHOS ($P = 2.20 \times 10^{-16}$) were the most significantly differentially expressed of all 37 pathways (Fig. 4c). Importantly, we found that >70% of the genes were detected and passed quality filtering in 33 of the 37 pathways, indicating sufficient coverage across the metabolic pathways for comparative analysis (Extended Data Fig. 4).

We next performed several studies for molecular validation of the metabolic differences between primary tumour and micrometastasis cells. As conventional oxygen-consumption assays are not amenable to the small cell numbers extracted from micrometastases, we performed a combination of flow cytometry and high-resolution metabolomic analyses. We compared the mitochondrial activity of primary tumour and metastatic cells through flow cytometry using tetramethylrhodamine methyl ester (TMRM), which is a fluorescent dye that accumulates in the mitochondria as a result of a difference in membrane potential (Fig. 4d)²⁹. We found that

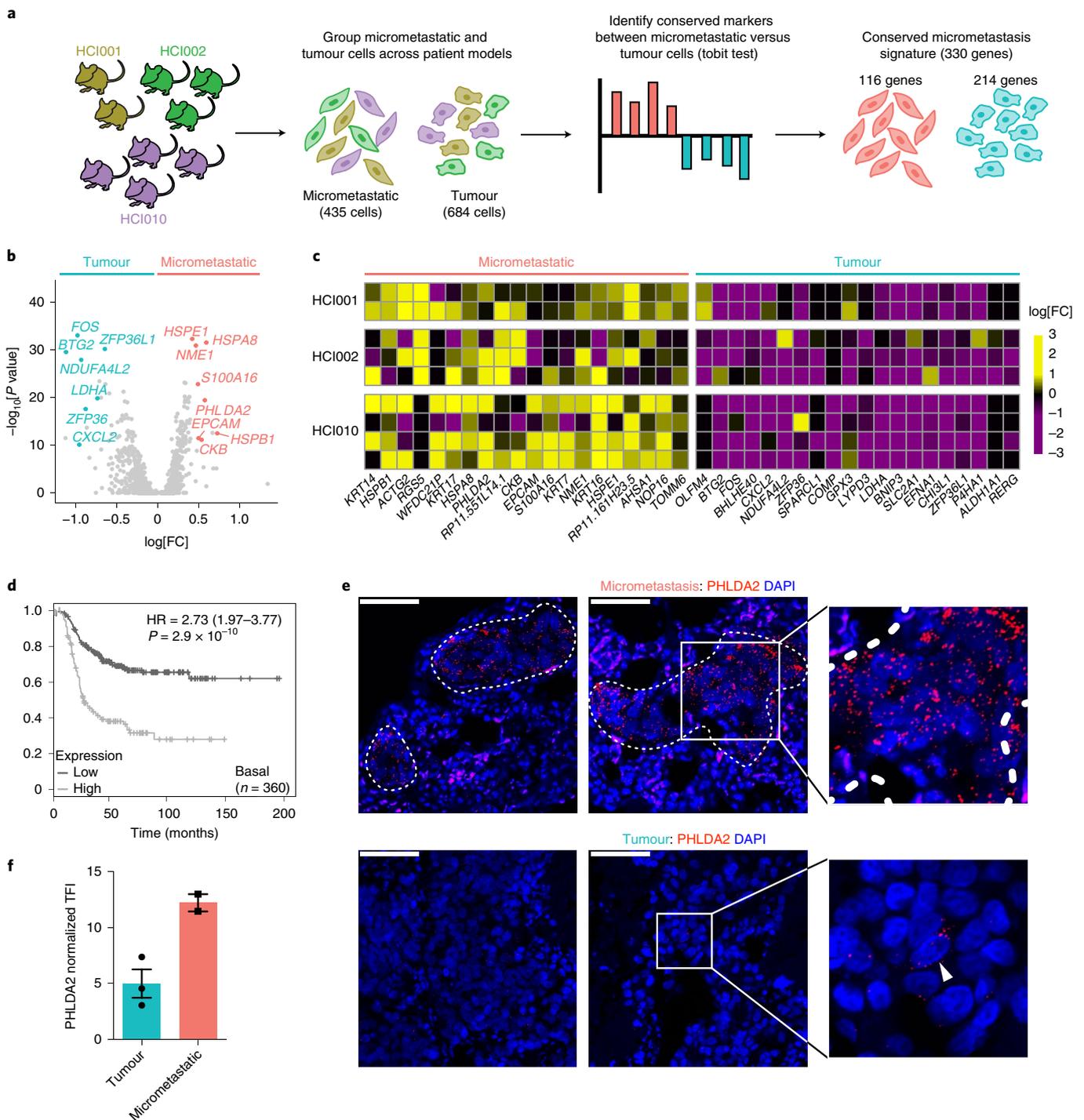


Fig. 3 | Micrometastatic cells display a distinct transcriptome program. **a**, Schematic of the analysis method to identify genes that were differentially expressed between micrometastatic and primary tumour cells, conserved in all three PDX models. All of the cells in the dataset were classified as micrometastatic or tumour cells and the differential gene expression between the groups was calculated using the tobit test in Seurat with the patient ID as a latent variable; 330 differentially expressed genes were identified ($P < 0.05$, $\text{min.pct} = 0.1$, $\log[\text{FC}]$ threshold = 0.25). **b**, Volcano plot showing all genes that were differentially expressed between micrometastases and tumours. The P values were determined using the tobit test in Seurat, which utilizes a likelihood-ratio test. **c**, Heat map of the top-20 marker genes for micrometastatic and primary tumour cells. The average FC of each gene (x axis) in micrometastatic relative to primary tumour cells was plotted for each mouse (y axis). Yellow indicates higher expression in micrometastatic cells and purple indicates higher expression in tumour cells. **d**, Kaplan-Meier survival curve showing decreased relapse-free survival in patients with basal-like breast cancer ($n = 360$) who expressed high levels of the top-15 micrometastasis-associated genes. The P values were determined using a log-rank test. **e**, Representative fluorescent in situ hybridization for PHLDA2 (RNAscope) on primary tumour ($n = 3$ biologically independent samples; bottom) and lung micrometastases ($n = 2$ biologically independent samples; top) from the PDX model HCI001. Insets, higher magnification of individual puncta. The white arrow indicates a tumour cell with high expression of PHLDA2. Scale bars, $50 \mu\text{m}$. **f**, Normalized total fluorescent intensity (TFI) of PHLDA2 in primary tumour and micrometastatic cells from the PDX model HCI001 ($n = 2$ lungs, >15 lesions; $n = 3$ tumours, 22 fields). Data are shown as the mean \pm s.e.m.

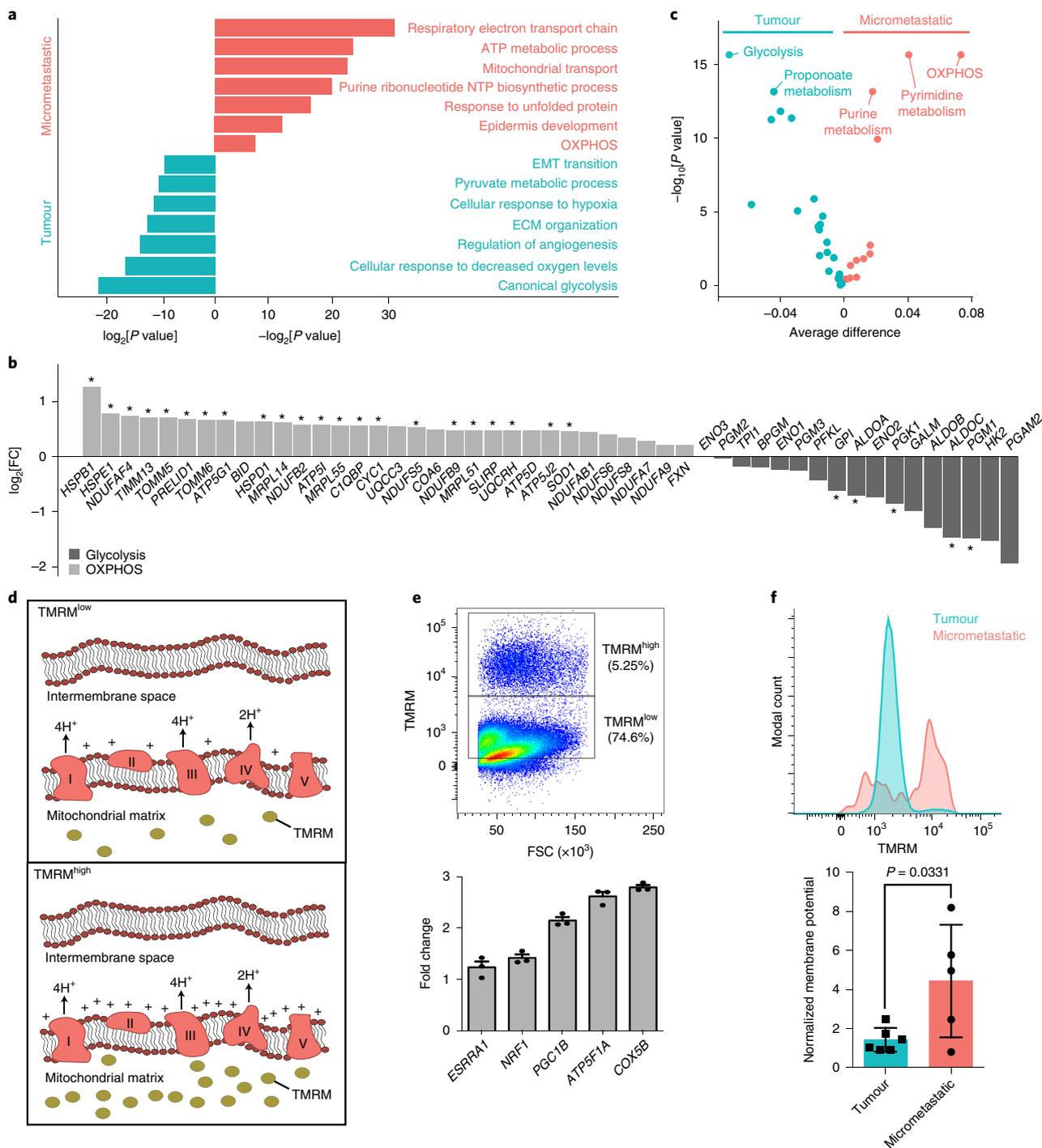


Fig. 4 | Micrometastatic cells display increased mitochondrial OXPHOS. **a**, Gene Ontology terms identified for primary tumour ($n=214$ genes) and micrometastatic cells ($n=116$ genes) based on the 330-micrometastasis-gene signature. The P values were determined using Fisher's exact test. **b**, Bar graph showing $\log_2[FC]$ values for differentially expressed OXPHOS- and glycolysis-associated genes in micrometastatic cells ($n=435$) relative to primary tumour cells ($n=684$). $*P < 0.05$, tobit test in Seurat (which utilizes a likelihood-ratio test) for genes in our 330-micrometastasis signature. The full micrometastatic gene signature and corresponding P values are provided in Supplementary Table 3. **c**, Gene scores of 37 unique metabolic pathways in micrometastatic ($n=435$) and primary tumour ($n=684$) cells. The average gene scores were calculated on individual tumour or micrometastatic cells and averaged across the three PDX models. Significance is shown on the y axis. The P values were determined using a two-sided Wilcoxon test. **d**, Schematic illustrating the mechanism of TMRM fluorescent staining in mitochondria. Increased ETC activity increases proton efflux from the mitochondrial matrix, resulting in a greater membrane potential and further accumulation of TMRM in the matrix. **e**, Top, flow cytometry analysis of TMRM^{high} and TMRM^{low} populations in primary tumour cells from an HCl010 mouse. Data are representative of $n=6$ biologically independent mice. Bottom, quantification of mitochondrial biogenesis and ETC genes in TMRM^{high} relative to TMRM^{low} tumour cells, determined by real-time quantitative reverse-transcription PCR; $n=3$ technical replicates. **f**, Top, histogram overlay of TMRM-fluorescence intensity in lung micrometastatic and primary tumour cells from an HCl010 mouse. Data are representative of $n=5$ biologically independent mice. Bottom, ratio of mean fluorescence intensity of TMRM to MitoTracker in CD298⁺MHC-I⁻ cells. The P value was determined using an unpaired two-sided Student's t -test. Data presented as the mean \pm s.d. of $n=6$ mice. One lung with no metastasis was excluded.

HCI010 primary tumours contain TMRM^{high} and TMRM^{low} populations, where TMRM^{high} cells are the minor population (Fig. 4e). Quantitative PCR analysis of sorted TMRM^{high} and TMRM^{low} cells confirmed that TMRM^{high} cells express higher levels of the genes involved in mitochondrial biogenesis and the ETC, validating the TMRM assay in PDX cells (Fig. 4e). We stained matched primary tumour and lung cell suspensions from six HCI010 PDX mice with TMRM. The cells were co-stained with a viability dye (Sytox), CD298, MHC-I and MitoTracker to identify viable human cells and assess the mitochondrial membrane potential relative to mitochondrial mass. Flow cytometry analysis showed that the mitochondrial membrane potential of micrometastases was threefold higher than primary tumour cells ($P=0.0331$; Fig. 4f). These data, combined with transcriptome analyses, strongly suggest that micrometastases utilize higher levels of OXPHOS for cellular metabolism.

Micrometastatic cells display a distinct metabolic profile. We next sought to determine whether micrometastatic cells are distinct from primary tumour cells at the global metabolic level. We used liquid chromatography coupled with high-resolution mass spectrometry (LC–HRMS) to perform steady-state metabolomic analysis of cells isolated from the lungs and primary tumours of six HCI010 PDX animals (Fig. 5a and Supplementary Table 6)³⁰. Although the less-abundant metabolites fell below the limit of detection due to limited cell numbers available from metastatic lesions, analysis of the 150 identified metabolites showed that metastatic cells display a distinct metabolic profile (Fig. 5b–d). This revealed modestly higher levels of fumarate and malate in metastatic cells ($P<0.05$), which are intermediates of the CAC, which feeds into OXPHOS (Fig. 5e and Supplementary Table 6). We also identified 2.9-fold higher levels of succinyl carnitine in metastatic cells, which accumulates as a waste intermediate of the CAC (Fig. 5e). Glutamine, glutamate, glucose and palmitoylcarnitine—an intermediate in fatty-acid metabolism—were also higher in metastatic cells (Fig. 5e). These metabolites are used by several pathways in the cell but mainly provide carbon sources for the CAC to generate NADH/FADH₂ for energy production through OXPHOS (Fig. 5f). We observed no difference in pyruvate or lactate between the primary tumour and metastatic cells (Fig. 5e). Interestingly, pathway analysis also showed increased amino acid metabolism in metastatic cells ($P<0.05$, FC > 1.5; Fig. 5g). Although amino acid accumulation may support diverse cellular functions—such as protein synthesis, signalling and proliferation—the catabolism of several of these amino acids—such as cysteine, lysine, arginine, phenylalanine and valine—can also support the CAC and OXPHOS³¹. These data show that metastatic cells display a distinct metabolic profile, which includes higher levels of several metabolites that may fuel OXPHOS.

Oxidative phosphorylation is critical for lung metastasis. We next investigated whether increased OXPHOS is functionally important for metastasis or simply represents a response to increased oxygen availability in the metastatic site. We used two experimental metastasis models of TNBC that metastasize to the lung with reproducible kinetics following intravenous delivery—that is, mouse 4T1 and human MDA-MB-231 cells. We used the complex V inhibitor oligomycin to inhibit OXPHOS in vitro and then measured its effects on metastatic outgrowth in the lung (Fig. 6a,b). We chose oligomycin because it is a highly selective inhibitor of OXPHOS with limited off-target effects³². Oligomycin induces cells to shift to glycolysis to produce ATP and has been used to study the consequences of this metabolic shift in diverse settings^{33–36}.

Given that oligomycin is cytotoxic if cells are unable to switch to increased glycolysis³³, we performed a series of analyses to test the health and metabolic state of MDA-MB-231 and 4T1 cells post treatment (Extended Data Fig. 5a,b). Flow cytometry analysis of propidium iodide and annexin V staining showed that oligomycin does

not reduce cell viability, increase apoptosis or alter cell morphology 72 h post treatment (Extended Data Fig. 5c,d). Furthermore, proliferation assays showed no decrease in the growth rate of treated cells (Extended Data Fig. 5e,g). Using the Seahorse XF assay, we found that oligomycin induces a sustained reduction in the oxygen-consumption rate (OCR) and a corresponding increase in the extracellular acidification rate (ECAR; Fig. 6c,d and Extended Data Fig. 5f,h), suggesting the cells shift from OXPHOS to glycolysis. We further utilized phasor fluorescence lifetime imaging (FLIM) of NADH to determine the durability of the metabolic shift. Fluorescence lifetime imaging relies on the fact that the fluorescence lifetime of NADH is longer when bound to enzymes involved in mitochondrial metabolism (approximately 3.4 ns) than when free floating in the cell during glycolysis (approximately 0.4 ns)^{37,38}. Analysis of the FLIM 6–24 h after treatment showed a significant decrease in bound NADH, indicating a durable shift from OXPHOS to glycolysis in both cell lines ($P<0.05$; Fig. 6e,f and Extended Data Fig. 5i,j). Collectively, these data demonstrate that oligomycin induces MDA-MB-231 and 4T1 cells to shift from OXPHOS to glycolytic metabolism without compromising cell viability or proliferation.

We injected NOD/SCID mice intravenously with oligomycin-treated or control MDA-MB-231 cells (8×10^5) and harvested lungs of the mice 21 d later (Fig. 6a). Flow cytometry analysis showed an almost threefold decrease in the frequency of metastatic cells in the lungs of the treatment group (Fig. 6g), showing that OXPHOS inhibition attenuates the metastatic capacity of MDA-MB-231 cells. We used a similar approach to test the effects of OXPHOS inhibition on 4T1 cell metastasis (Fig. 6b). We injected green fluorescent protein (GFP)-labelled 4T1 cells (1×10^5) into BALB/c mice and harvested the lungs of the mice 5 d later, as 4T1 cells metastasize rapidly in the lung. Remarkably, the animals injected with oligomycin-treated 4T1–GFP cells displayed a sevenfold decrease in metastatic cells in their lungs at this early time point (Fig. 6h), suggesting OXPHOS is important to facilitate early events in the metastatic cascade.

Finally, we tested the effects of oligomycin treatment on primary tumour growth to determine whether OXPHOS is specifically important for metastasis or generally important for cancer cell function. We treated 4T1 and MDA-MB-231 cells with oligomycin in vitro, performed bilateral orthotopic injections (1×10^6 cells), and harvested tumours 14 and 17 d later, respectively (Extended Data Fig. 5k,l). We found no significant difference in engraftment efficiency or tumour size in control versus oligomycin-treated tumours (Extended Data Fig. 5k,l), further establishing a role for OXPHOS in mediating the lung metastasis (Extended Data Fig. 6).

Discussion

Defining the molecular mechanisms that facilitate specific stages of the metastatic cascade is critical to understand how cells metastasize and inspire new strategies to prevent metastatic spread in cancer patients. We have developed a robust approach for the capture and analysis of rare cells during the seeding and establishment of micrometastasis in PDX models using scRNA-seq. We found that micrometastases display significant transcriptional heterogeneity but also display a distinct expression program that is predictive of poor survival in patients with breast cancer. We identify mitochondrial OXPHOS as the top pathway upregulated in micrometastases, in contrast to primary tumour cells that score higher for genes associated with aerobic glycolysis (Extended Data Fig. 6). We validated this by flow cytometry and HR-LCMS, where we found that micrometastases display a distinct metabolic profile and higher levels of several metabolites that feed into OXPHOS. Most importantly, we found that pharmacological inhibition of OXPHOS substantially attenuates lung metastasis, showing that OXPHOS is functionally critical for metastatic spread.

Our work sheds light on the controversial role of metabolism in metastasis. Historically, tumours were thought to contain dys-

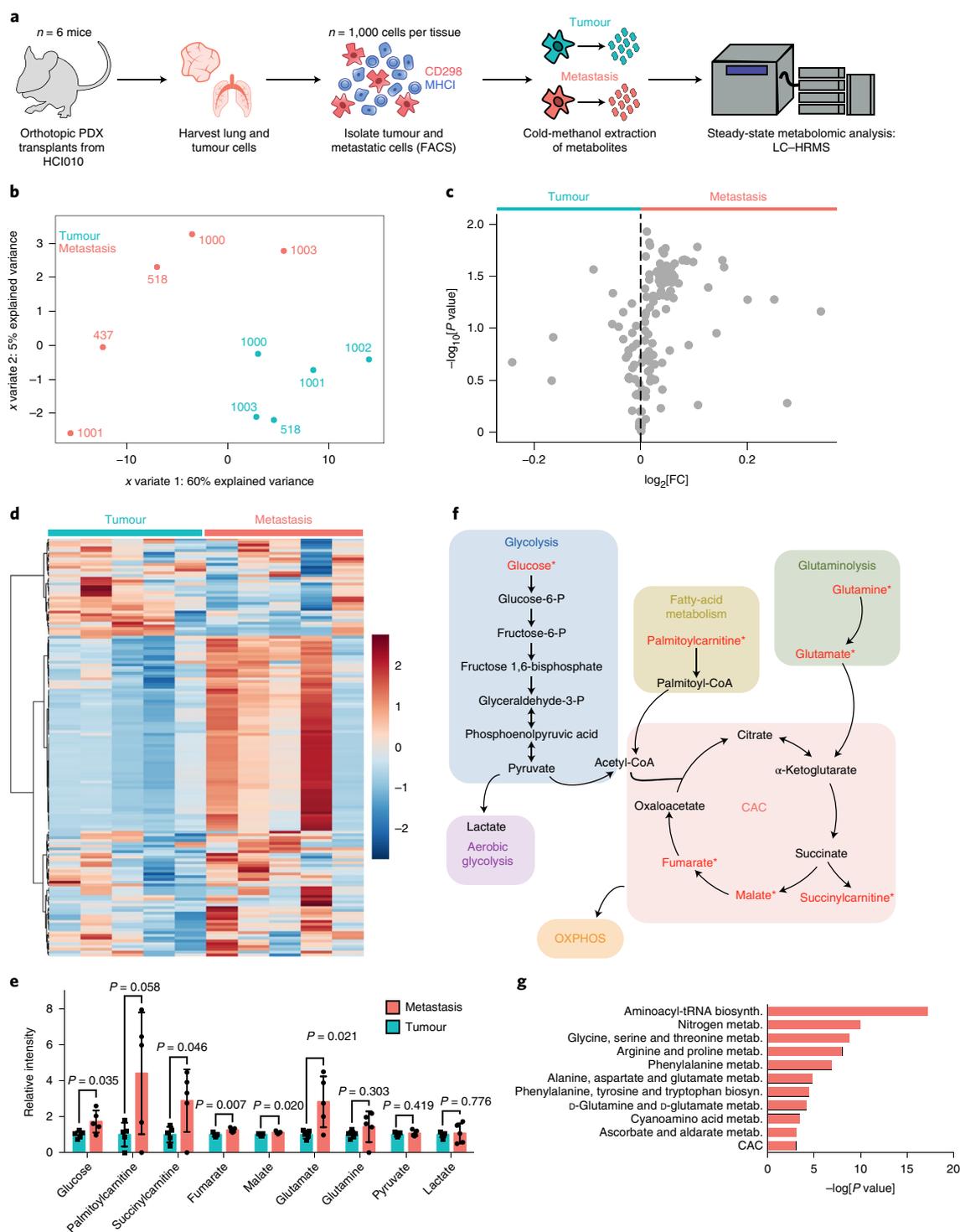


Fig. 5 | Metastatic cells display a distinct metabolic profile. **a**, Schematic overview of the workflow for steady-state global metabolomics profiling of matched primary tumour and lung metastatic cells from six HCl010 PDX animals using LC-HRMS. **b**, Partial least-squares discriminant analysis of the global metabolite data in metastatic and tumour cells. Individual points represent the bulk profiles of $n = 1,000$ cells isolated from either the lung (metastatic) or primary tumour ($n = 6$ mice). The individual points are labelled with their mouse ID. One tumour and one metastatic sample were excluded from the analysis due to low signal. **c**, Volcano plot showing metabolites that were differentially detected in metastatic or tumour samples. The P values were determined using an unpaired two-sided Student's t -test. **d**, Heat map of peak intensities measured by LC-MS showing unsupervised clustering of 150 metabolites identified in metastatic and primary tumour cell samples normalized by autoscaling on MetaboAnalyst and ranked by a Student's t -test. **e**, Levels of selected glycolysis, CAC and lipid metabolism intermediates in metastatic cells relative to primary tumour cells represented as the normalized intensity mean \pm s.d. of $n = 5$ micrometastatic and tumour samples. The values of individual samples were normalized to the tumour average. The P values were determined using an unpaired two-sided Student's t -test. **f**, Schematic showing the role of metabolites identified at higher levels in metastatic cells (indicated by an asterisk) in key metabolic pathways. **g**, Top metabolic pathways that were higher in metastatic cells identified by pathway analysis ($P < 0.05$; $FC > 1.5$; $n = 62$ metabolites). Metab., metabolism; biosynth., biosynthesis. The P values were determined using an unpaired two-sided Student's t -test in MetaboAnalyst.

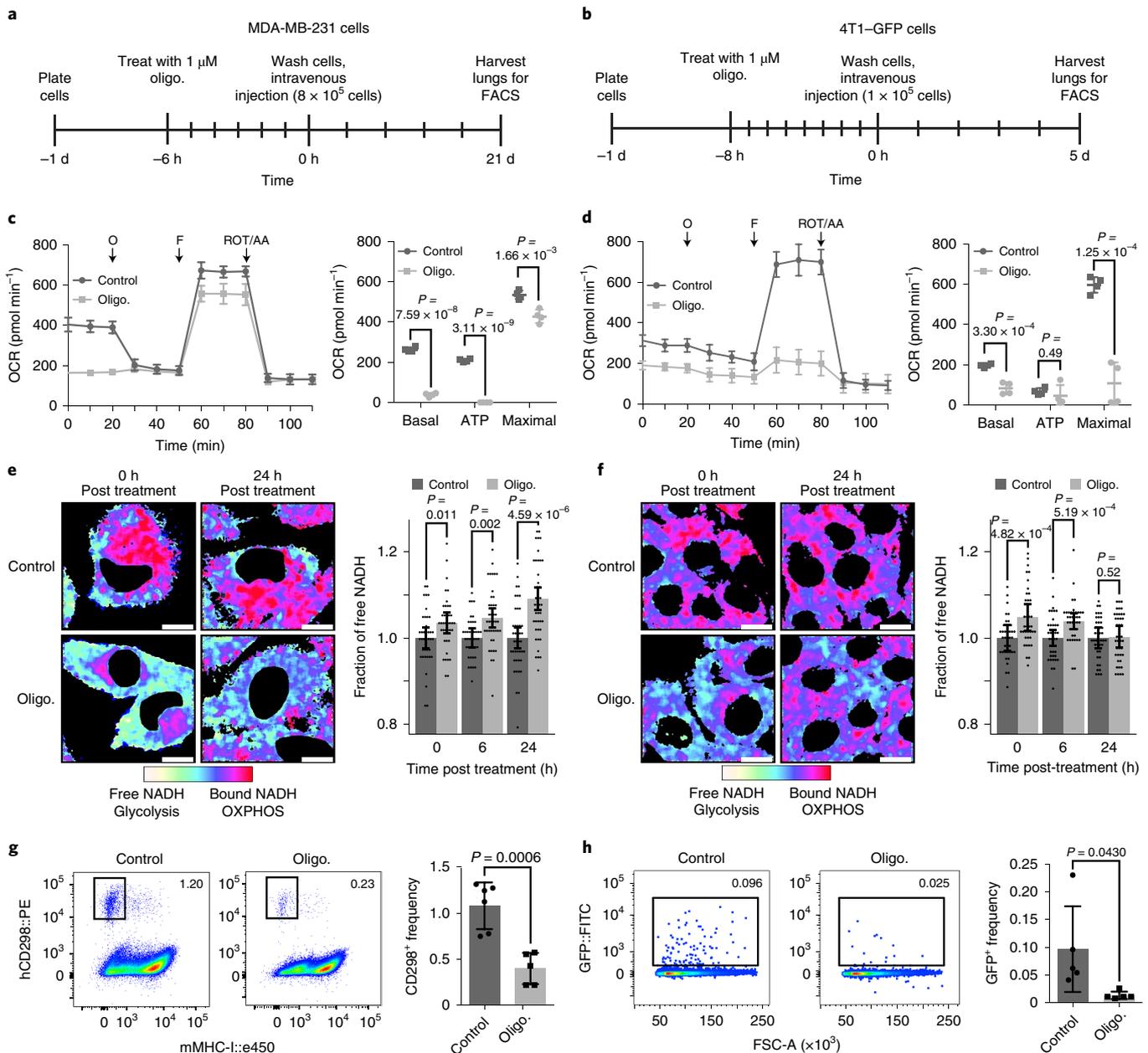


Fig. 6 | Oxidative phosphorylation is critical for lung metastasis. a, b, Schematic of the experimental approach used to determine the effects of complex V inhibition via oligomycin on metastasis of MDA-MB-231 (**a**) and 4T1-GFP cells (**b**). **c, d**, Left, OCR of MDA-MB-231 (**c**) and 4T1-GFP cells (**d**) versus control cells after treatment with oligomycin for 6–8 h in vitro. O, oligomycin; F, FCCP; ROT/AA, rotenone with antimycin A. Right, basal, ATP and maximal respiration rates for each condition. The P values were determined using an unpaired two-sided Student's t -test. Data presented as the mean \pm s.d. of $n = 4$ replicates. **e, f**, Left, representative FLIM images of the fluorescence lifetime of NADH in cultured MDA-MB-231 (**e**) or 4T1-GFP cells (**f**). The fields do not represent consecutive images of the same cell. Scale bars, 10 μm . Right, quantification of the free-NADH fraction in the FLIM images normalized to the control. The P values were determined using a two-sided Student's t -test with homoscedasticity determined using an F -test. Data presented as the mean \pm s.e.m.; $t = 0$ h, $n = 39$ fields; $t = 6$ h, $n = 41$ (MDA-MB-231) and 40 (4T1-GFP) fields; and $t = 24$ h, $n = 42$ fields. **g, h**, Flow cytometry analysis of the metastatic burden in the lungs of the animals injected intravenously with oligomycin-treated or control cells. Left, FACS plots show MDA-MB-231 cells (**g**), identified as CD298⁺MHC-1⁺, and 4T1-GFP cells (**h**), identified by GFP, in representative animals. Right, frequency of metastatic cells. Data presented as the mean \pm s.d. of the frequency of MDA-MB-231 ($n = 5$ oligomycin-treated, one mouse without detectable micrometastases was excluded; and $n = 6$ control mice) and 4T1-GFP ($n = 5$ oligomycin-treated and control mice) cells. The P value was determined using an unpaired two-sided Student's t -test. Oligo., oligomycin.

functional mitochondria and be principally sustained by Warburg metabolism^{39,40}. Recent work has challenged this dogma, showing the importance of glutamine^{41–43}, fatty-acid^{44,45}, proline⁴⁶ and pyruvate carboxylase-mediated metabolism⁴⁷ in metastatic progression^{48,49}. Recent work using the 4T1 mouse model also showed that

circulating cancer cells utilize increased OXPHOS and that suppression of a central regulator of mitochondrial biogenesis (PGC1 α) reduces cancer-cell invasion and metastasis⁵⁰. Another study showed increased expression of mitochondrial complex I and oxidative stress in lung micrometastases using in situ RNA sequencing⁵¹.

Our work utilized an unbiased transcriptomic approach to discover that OXPHOS promotes metastasis, specifically during the seeding step of the metastatic cascade, in three distinct human patient models of TNBC. Interestingly, many metabolic pathways implicated in metastasis—such as glutamine, fatty-acid and proline metabolism—also converge on or produce critical metabolites to drive OXPHOS, raising the question of whether OXPHOS represents an underlying mechanistic link between them.

Oxidative phosphorylation may promote metastatic seeding in several ways. Increased ATP generation through OXPHOS may provide energy for cytoskeleton remodelling for motility or to survive anoikis during cell detachment and migration⁵². Increased ROS production through OXPHOS may promote cell motility by activating oncogenic signalling, as mitochondrial ROS-inducing mutations are sufficient to trigger metastasis^{53–57}. Epidemiological data in humans also support a role for OXPHOS in cancer progression, showing that treatment with the diabetes drug metformin (complex I inhibitor) is protective against the relapse and metastasis of breast cancer^{58–61}. Several drugs targeting mitochondrial metabolism are already FDA approved or in clinical trials (for example, IACS-010759 (ref. ⁶²), plumbagin^{63,64} and atovaquone^{65,66}), providing an opportunity for re-indication in breast cancer patients to prevent metastasis. However, further studies are needed to define which specific steps of the metastatic cascade OXPHOS is critical for to determine when it might be applied for clinical benefit against metastasis.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41556-020-0477-0>.

Received: 28 October 2018; Accepted: 4 February 2020;

Published online: 6 March 2020

References

- Bianchini, G., Balko, J. M., Mayer, I. A., Sanders, M. E. & Gianni, L. Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. *Nat. Rev. Clin. Oncol.* **13**, 674–690 (2016).
- Weigelt, B., Peterse, J. L. & van't Veer, L. J. Breast cancer metastasis: markers and models. *Nat. Rev. Cancer* **5**, 591–602 (2005).
- Oskarsson, T., Battle, E. & Massagué, J. Metastatic stem cells: sources, niches, and vital pathways. *Cell Stem Cell* **14**, 306–321 (2014).
- Lawson, D. A., Kessenbrock, K., Davis, R. T., Pervolarakis, N. & Werb, Z. Tumour heterogeneity and metastasis at single-cell resolution. *Nat. Cell Biol.* **20**, 1349–1360 (2018).
- Hermann, P. C. et al. Distinct populations of cancer stem cells determine tumor growth and metastatic activity in human pancreatic cancer. *Cell Stem Cell* **1**, 313–323 (2007).
- Lodhia, K. A., Hadley, A. M., Haluska, P. & Scott, C. L. Prioritizing therapeutic targets using patient-derived xenograft models. *Biochim. Biophys. Acta* **1855**, 223–234 (2015).
- Hochhauser, D. & Caldas, C. Of mice and men: patient-derived xenografts in cancer medicine. *Ann. Oncol.* **28**, 2330–2331 (2017).
- Izumchenko, E. et al. Patient-derived xenografts effectively capture responses to oncology therapy in a heterogeneous cohort of patients with solid tumors. *Ann. Oncol.* **28**, 2595–2605 (2017).
- DeRose, Y. S. et al. Tumor grafts derived from women with breast cancer authentically reflect tumor pathology, growth, metastasis and disease outcomes. *Nat. Med.* **17**, 1514–1520 (2011).
- Lawson, D. A. et al. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526**, 131–135 (2015).
- Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
- Ziegenhain, C. et al. Comparative analysis of single-cell RNA sequencing methods. *Mol. Cell* **65**, 631–643.e4 (2017).
- Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
- Patel, A. P. et al. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
- Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
- Puram, S. V. et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck cancer. *Cell* **171**, 1611–1624 (2017).
- Venteicher, A. S. et al. Decoupling genetics, lineages, and microenvironment in IDH-mutant gliomas by single-cell RNA-seq. *Science* **355**, eaai8478 (2017).
- Bruna, A. et al. A biobank of breast cancer explants with preserved intra-tumor heterogeneity to screen anticancer compounds. *Cell* **167**, 260–274 (2016).
- Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
- The Gene Ontology Consortium. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **45**, D331–D338 (2017).
- Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
- Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
- Ikwegbue, P. C., Masamba, P., Oyinloye, B. E. & Kappo, A. P. Roles of heat shock proteins in apoptosis, oxidative stress, human inflammatory diseases, and cancer. *Pharmaceuticals* **11**, 1–18 (2018).
- Creagh, E. M., Sheehan, D. & Cotter, T. G. Heat shock proteins—modulators of apoptosis in tumour cells. *Leukemia* **14**, 1161–1173 (2000).
- Györfi, B. et al. An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* **123**, 725–731 (2010).
- Tunster, S. J., Creeth, H. D. J. & John, R. M. The imprinted *Phlda2* gene modulates a major endocrine compartment of the placenta to regulate placental demands for maternal resources. *Dev. Biol.* **409**, 251–260 (2016).
- Moon, H.-G. et al. Prognostic and functional importance of the engrainment-associated genes in the patient-derived xenograft models of triple-negative breast cancers. *Breast Cancer Res. Treat.* **154**, 13–22 (2015).
- Scaduto, R. C. & Grotyohann, L. W. Measurement of mitochondrial membrane potential using fluorescent rhodamine derivatives. *Biophys. J.* **76**, 469–477 (1999).
- Liu, X. et al. High-resolution metabolomics with Acyl-CoA profiling reveals widespread remodeling in response to diet. *Mol. Cell. Proteomics* **14**, 1489–1500 (2015).
- Mayers, J. R. et al. Elevation of circulating branched-chain amino acids is an early event in human pancreatic adenocarcinoma development. *Nat. Med.* **20**, 1193–1198 (2014).
- Fernyhough, P. & McGavock, J. in *Handbook of Clinical Neurology* Vol. 126 (eds Zochodne, D.W. and Malik, R.A.) 353–377 (Elsevier, 2014).
- Hao, W., Chang, C. P., Tsao, C. C. & Xu, J. Oligomycin-induced bioenergetic adaptation in cancer cells with heterogeneous bioenergetic organization. *J. Biol. Chem.* **285**, 12647–12654 (2010).
- Choi, S. W., Gerecsner, A. A. & Nicholls, D. G. Bioenergetic analysis of isolated cerebrocortical nerve terminals on a microgram scale: spare respiratory capacity and stochastic mitochondrial failure. *J. Neurochem.* **109**, 1179–1191 (2009).
- Dayan, F. et al. Activation of HIF-1 α in exponentially growing cells via hypoxic stimulation is independent of the Akt/mTOR pathway. *J. Cell. Physiol.* **218**, 167–174 (2009).
- Ward, M. W., Rego, A. C., Frenguelli, B. G. & Nicholls, D. G. Mitochondrial membrane potential and glutamate excitotoxicity in cultured cerebellar granule cells. *J. Neurosci.* **20**, 7208–7219 (2000).
- Digman, M. A., Caiolfa, V. R., Zamai, M. & Gratton, E. The phasor approach to fluorescence lifetime imaging analysis. *Biophys. J.* **94**, L14–L16 (2008).
- Sameni, S., Syed, A., Marsh, J. L. & Digman, M. A. The phasor-FLIM fingerprints reveal shifts from OXPHOS to enhanced glycolysis in Huntington Disease. *Sci. Rep.* **6**, 34755 (2016).
- Warburg, O. Origin of cancer cells. *Oncol.* **9**, 75–83 (1956).
- Koppenol, W. H., Bounds, P. L. & Dang, C. V. Otto Warburg's contributions to current concepts of cancer metabolism. *Nat. Rev. Cancer* **11**, 325–337 (2011).
- Yang, L. et al. Metabolic shifts toward glutamine regulate tumor growth, invasion and bioenergetics in ovarian cancer. *Mol. Syst. Biol.* **10**, 728 (2014).
- Dornier, E. et al. Glutaminolysis drives membrane trafficking to promote invasiveness of breast cancer cells. *Nat. Commun.* **8**, 2255 (2017).
- Rodrigues, M. F. et al. Enhanced OXPHOS, glutaminolysis and β -oxidation constitute the metastatic phenotype of melanoma cells. *Biochem. J.* **473**, 703–715 (2016).
- Camarda, R. et al. Inhibition of fatty acid oxidation as a therapy for MYC-overexpressing triple-negative breast cancer. *Nat. Med.* **22**, 427–432 (2016).

45. Antalis, C. J., Uchida, A., Buhman, K. K. & Siddiqui, R. A. Migration of MDA-MB-231 breast cancer cells depends on the availability of exogenous lipids and cholesterol esterification. *Clin. Exp. Metastasis* **28**, 733–741 (2011).
46. Elia, I. et al. Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. *Nat. Commun.* **8**, 15267 (2017).
47. Christen, S. et al. Breast cancer-derived lung metastases show increased pyruvate carboxylase-dependent anaplerosis. *Cell Rep.* **17**, 837–848 (2016).
48. Elia, I., Doglioni, G. & Fendt, S. M. Metabolic hallmarks of metastasis formation. *Trends Cell Biol.* **28**, 673–684 (2018).
49. Teoh, S. T. & Lunt, S. Y. Metabolism in cancer metastasis: bioenergetics, biosynthesis, and beyond. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **10**, e1406 (2018).
50. LeBleu, V. S. et al. PGC-1 α mediates mitochondrial biogenesis and oxidative phosphorylation in cancer cells to promote metastasis. *Nat. Cell Biol.* **16**, 992–1003 (2014).
51. Basnet, H. et al. Flura-seq identifies organ-specific metabolic adaptations during early metastatic colonization. *eLife* **8**, e43627 (2019).
52. Schafer, Z. T. et al. Antioxidant and oncogene rescue of metabolic defects caused by loss of matrix attachment. *Nature* **461**, 109–113 (2009).
53. Porporato, P. E. et al. A mitochondrial switch promotes tumor metastasis. *Cell Rep.* **8**, 754–766 (2014).
54. Ogura, M., Yamaki, J., Homma, M. K. & Homma, Y. Mitochondrial c-Src regulates cell survival through phosphorylation of respiratory chain components. *Biochem. J.* **447**, 281–289 (2012).
55. Zielonka, J. & Kalyanaraman, B. 'ROS-generating mitochondrial DNA mutations can regulate tumor cell metastasis'-a critical commentary. *Free Radic. Biol. Med.* **45**, 1217–1219 (2008).
56. Dai, X. et al. Breast cancer intrinsic subtype classification, clinical use and future trends. *Am. J. Cancer Res.* **5**, 2929–2943 (2015).
57. Giannoni, E. et al. Redox regulation of anoikis: reactive oxygen species as essential mediators of cell survival. *Cell Death Differ.* **15**, 867–878 (2008).
58. Kumar, S. et al. Metformin intake is associated with better survival in ovarian cancer: a case-control study. *Cancer* **119**, 555–562 (2013).
59. Bodmer, M., Becker, C., Meier, C., Jick, S. S. & Meier, C. R. Use of metformin and the risk of ovarian cancer: a case-control analysis. *Gynecol. Oncol.* **123**, 200–204 (2011).
60. Romero, I. L. et al. Relationship of type II diabetes and metformin use to ovarian cancer progression, survival, and chemosensitivity. *Obstet. Gynecol.* **119**, 61–67 (2012).
61. Col, N. F., Ochs, L., Springmann, V., Aragaki, A. K. & Chlebowski, R. T. Metformin and breast cancer risk: a meta-analysis and critical literature review. *Breast Cancer Res. Treat.* **135**, 639–646 (2012).
62. Yap, T. A. et al. Phase I trial of IACS-010759 (IACS), a potent, selective inhibitor of complex I of the mitochondrial electron transport chain, in patients (pts) with advanced solid tumors. *J. Clin. Oncol.* **37**, 3014–3014 (2019).
63. Hafeez, B. Bin et al. Plumbagin inhibits prostate cancer development in TRAMP mice via targeting PKC ϵ , Stat3 and neuroendocrine markers. *Carcinogenesis* **33**, 2586–2592 (2012).
64. Hafeez, B. Bin et al. Plumbagin inhibits prostate carcinogenesis in intact and castrated PTEN knockout mice via targeting PKC ϵ , Stat3, and epithelial-to-mesenchymal transition markers. *Cancer Prev. Res.* **8**, 375–386 (2015).
65. Fiorillo, M. et al. Repurposing atovaquone: targeting mitochondrial complex III and OXPHOS to eradicate cancer stem cells. *Oncotarget* **7**, 34084–34099 (2016).
66. Lv, Z., Yan, X., Lu, L., Su, C. & He, Y. Atovaquone enhances doxorubicin's efficacy via inhibiting mitochondrial respiration and STAT3 in aggressive thyroid cancer. *J. Bioenerg. Biomembr.* **50**, 263–270 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

PDX models. The samples from patients were provided by A. L. Welm at the Department of Oncological Sciences at the Huntsman Cancer Institute (HCI). All of the tissue samples were collected with informed consent from individuals being treated at the Huntsman Cancer Hospital and the University of Utah under a protocol approved by the Institutional Review Board of the University of Utah¹². HCI001 was acquired from a primary tumour biopsy of a female patient diagnosed with Stage IV ER⁺PR⁺Her2⁻ basal-like invasive ductal carcinoma with no previous systemic treatment. HCI002 was acquired from a primary tumour biopsy of a female patient diagnosed with Stage IIIA ER⁺PR⁺Her2⁻ basal-like medullary-type invasive ductal carcinoma with no previous systemic treatment. HCI010 was acquired from a pleural effusion of a female patient diagnosed with Stage IIIC ER⁺PR⁺Her2⁻ basal-like (PAM50) invasive ductal carcinoma treated with several rounds of chemotherapies¹². Additional clinical details of each patient tumour can be found in Supplementary Table 1 of ref. ¹². The samples were collected and identified by the Huntsman Cancer Institute Tissue Resource and Application Core facility before being obtained for implantation. The study is compliant with all of the relevant ethical regulations regarding research involving human participants.

Animal experiments. The Institutional Animal Care and Use Committee of the University of California, Irvine reviewed and approved all of the animal experiments. Orthotopic transplants of serially passaged human tumour samples were performed on immunocompromised three- to four-week-old NOD/SCID or NSG mice after clearing the mammary fat pads following established protocols¹⁰. Tumour growth was monitored by weekly caliper measurements and volumes were calculated as: length × width² × 0.51. The animals were euthanized and tissues were harvested when the tumours reached a length or width of 2.0–2.5 cm. For the experimental metastasis studies, cultured MDA-MB-231 or 4T1-GFP cells were suspended in 100 μl sterile PBS and injected into the tail vein of eight- to ten-week-old female NOD/SCID (MDA-MB-231) or BALB/c (4T1-GFP) mice. The mice were euthanized by asphyxiation with CO₂ followed by cervical dislocation and perfusion with 10 mM EDTA in D-PBS 21 d (MDA-MB-231) or 5 d (4T1-GFP) post injection. For the cell-line orthotopic tumour injections, the mice were placed under isoflurane anaesthesia (1.5–2.0%). The area was shaved and cleaned with an alcohol swab before injection. MDA-MB-231 or 4T1-GFP cells (1 × 10⁶ cells in 100 μl PBS) were injected bilaterally into the fourth mammary fat pad of 10- to 12-week-old NSG mice. The study is compliant with all of the relevant ethical regulations regarding animal research.

Tissue harvest and dissociation. Animals at the endpoint were euthanized by asphyxiation with CO₂ followed by cervical dislocation and perfusion with 10 mM EDTA in D-PBS. Evan's Blue (Sigma-Aldrich, cat. no. E2129-10G) was injected into the footpads and ears of the anaesthetized mice before perfusion to aid visualization of the lymph nodes. The solid tissues from the mice—which included the primary tumour, lungs and lymph nodes—were processed for flow cytometry by mechanical chopping with blades, followed by collagenase IV (Sigma-Aldrich cat. no. C5138-1G) digestion in medium (DMEM-F12 medium with 5% FBS, 5 μg ml⁻¹ insulin and 1% penicillin/streptomycin solution) for 45 min at 37 °C. The cell suspensions were washed with 2 μg ml⁻¹ DNase I (Worthington Biochemical, cat. no. LS002139) for 5 min and further dissociated with 0.05% trypsin for 10 min. Following a wash with Hanks balanced salt solution with 2% FBS, the cells were passed through a 70-μm filter. Lung and primary tumour cells were treated with 1 × RBC lysis buffer, followed by resuspension in DMEM-F12 with 10% FBS for FACS.

Cell culture. MDA-MB-231 and 4T1-GFP cells were cultured in either 10 cm or six-well plates (Genesee Scientific, cat. no. 25-105) containing DMEM-F12 (Fisher Scientific, cat. no. MT10090CV), 5% heat-inactivated fetal bovine serum (Sigma-Aldrich, cat. no. 12306C) and 1% penicillin–streptomycin 100× solution (Hyclone, cat. no. SV30010). The cells were passaged with 0.05% trypsin (Corning, cat. no. 25-052-Cl) after reaching 70% confluency. For the control-treated cells, the medium was replaced with fresh medium. For the oligomycin-treated cells, medium containing 1 μM oligomycin was added for either 6 h (MDA-MB-231) or 8 h (4T1-GFP), the cells were then washed and fresh medium was added. For the FLIM analysis, 4T1-GFP mouse breast cancer cells and MDA-MB-231 human breast cancer cells were cultured in DMEM (Thermo Fisher Scientific, cat. no. 11965092) supplemented with 10% heat-inactivated fetal bovine serum (Thermo Fisher Scientific, cat. no. 10082139) and 1% penicillin–streptomycin 100× solution (Hyclone, cat. no. SV30010). The cells were plated at a confluency of 2 × 10⁴ cells cm⁻² in eight-well glass-bottomed fibronectin-coated imaging dishes (Thermo Fisher Scientific, cat. no. 155411) approximately 24 h before drug treatment and left to adhere. The medium was replaced the following day with medium containing 1 μM oligomycin (MP Biomedicals, cat. no. 0215178610), whereas the control cells had their medium replaced. The cells were washed with PBS and their medium was replaced after 8 h (4T1-GFP) or 6 h (MDA-MB-231) of treatment. The cells were then imaged at 0, 6 and 24 h post washout for no more than 1 h per experimental condition. For the proliferation studies, 2 × 10⁵ MDA-MB-231 cells were seeded on six-well plates (Genesee Scientific, cat. no. 25-105) containing DMEM-F12 (Fisher Scientific, cat. no. MT10090CV),

10% heat-inactivated fetal bovine serum (Sigma-Aldrich, cat. no. 12306C), 1% penicillin–streptomycin 100× solution (Hyclone, cat. no. SV30010) and 5 mM HEPES (Gibco, cat. no. 15630-080) approximately 24 h before drug treatment and left to adhere; 4T1 (2 × 10⁵) cells were seeded on six-well plates (Genesee Scientific, cat. no. 25-105) containing DMEM-F12 (Fisher Scientific, cat. no. MT10090CV), 5% heat-inactivated fetal bovine serum (Sigma-Aldrich, cat. no. 12306C) and 1% penicillin–streptomycin 100× solution (Hyclone, cat. no. SV30010) approximately 24 h before drug treatment and left to adhere. The medium was replaced the following day with medium containing 1 μM oligomycin (MP Biomedicals, cat. no. 0215178610), whereas the control cells had their medium replaced. The cells were washed with PBS and their medium was replaced after 8 h (4T1-GFP) or 6 h (MDA-MB-231) of treatment. The cells were collected with 0.05% trypsin (Corning, cat. no. 25-052-Cl) at 24, 48 and 72 h post treatment and counted using a Countess II automated cell counter (Thermo Fisher Scientific, cat. no. AMQAX1000). The growth rate between 24 and 48 h, and between 48 and 72 h post treatment was calculated using the following equation:

$$\text{Growth rate} = \frac{\ln\left(\frac{N(t)}{N(0)}\right)}{t}$$

Where $N(t)$ is the number of cells at either 48 or 72 h, $N(0)$ is the number of cells at either 24 or 48 h and t is the time between measurements.

Flow cytometry. We used the human-specific antibody CD298 (diluted 1:100; PE; BioLegend, cat. no. 341704) and the mouse-specific antibody MHC-I (diluted 1:150; APC; Thermo Fisher Scientific, cat. no. 17-5957-80). Flow cytometry was performed using a BD FACSAria Fusion cell sorter. Cell viability was determined by negative staining with SYTOX blue (diluted 1:1,000; Thermo Fisher Scientific, cat. no. S34857). The forward-scatter area by forward-scatter width (FSC-H × FSC-A) and side-scatter area by side-scatter width (SSC-H × SSC-A) was used to discriminate single cells from doublet and multiplet cells. Mouse cells were excluded by gating out CD298⁻MHC-I⁺ cells. Human primary tumour cells and metastatic cells were selected by gating on Sytox⁻CD298⁺MHC-I⁻ cells. For the mitochondrial membrane potential readouts, the cells were stained with TMRM (diluted 1:500; Thermo Fisher Scientific, cat. no. T668), MitoTracker-Green (diluted 1:100 from a 10 μM stock; Thermo Fisher Scientific, cat. no. M7514), human-specific antibody CD298 (diluted 1:100; APC; BioLegend, cat. no. 341706) and the mouse-specific antibody MHC-I (diluted 1:100; PE/Cy7; BioLegend, cat. no. 114717). Cell viability was determined by negative staining with SYTOX blue (diluted 1:1,000; Thermo Fisher Scientific, cat. no. S34857). Compensation controls and fluorescence minus one controls were used to determine the TMRM^{high} and TMRM^{low} populations in the PDX tumour cells. For the analysis of the MDA-MB-231 metastatic burden, FSC-W × FSC-A and SSC-W × SSC-A was used to discriminate single cells from doublet and multiplet cells. The MDA-MB-231 cells were defined by gating on Sytox⁻CD298⁺MHC-I⁻ cells and the level of metastatic burden was determined using the frequency of that population. One mouse from the oligomycin-treated group did not contain any detectable MDA-MB-231 cells by flow cytometry and was thus excluded from the analysis. For the analysis of the 4T1-GFP metastatic burden, FSC-H × FSC-A and SSC-H × SSC-A was used to discriminate single cells from doublet and multiplet cells. The 4T1-GFP cells were defined by gating on Sytox⁻GFP⁺ cells and the level of metastatic burden was determined using the frequency of that population. For the analysis of cell viability following oligomycin treatment, the cells were stained with annexin V-FITC (diluted 1:100; GeneTex, cat. no. GTX14082) and propidium iodide (diluted 1:100; Thermo Fisher Scientific, cat. no. P3566). FSC-W × FSC-A and SSC-W × SSC-A was used to discriminate single cells from doublet and multiplet cells. Viability was determined by the frequency of annexin V⁻propidium iodide⁻ cells in the population.

Generation of scRNA-seq data. Single cells were sorted directly into each well of a skirted 96-well PCR plate (Fisher Scientific, Eppendorf, cat. no. E951020443) containing lysis buffer (0.2% Triton X-100 (Sigma-Aldrich, cat. no. T9284), 2 U μl⁻¹ RNaseOUT (Thermo Fisher Scientific, cat. no. 10777019), 10 μM oligo-dT30VN and 10 μM dNTPs (Thermo Fisher Scientific, cat. no. 18427088)) as described previously¹¹. The plates were snap frozen on dry ice and stored at -80 °C until further processing. Total RNA was converted into complementary DNA using the SmartSeq2 protocol and prepared for Illumina sequencing using the Nextera XT DNA library preparation kit (Illumina, cat. no. FC-131-1096). The cells were sequenced at a depth of 1 × 10⁶ reads per cell on a HiSeq 2500 system.

Processing of scRNA-seq data. Files from the HiSeq 2500 were demultiplexed and converted to FASTQ files. Paired-end 100 bp reads were aligned to the Gencode 21 human transcriptome using Bowtie 2 and quantified using RSEM with the following parameters: rsem-calculate-expression -p \$CORES --bowtie2 --paired-end READ1 READ2 gencodehg21. The expression values were log-transformed into log[transcripts per kilobase million + 1] matrices and loaded into the Seurat analysis package with the following parameters: p10 <- CreateSeuratObject(raw.data = p10.mat, min.cells = 8, min.genes = 1,000, project = 'HCI010'). We removed any cells identified as visual outliers by library complexity (<2,500 genes per cell) or overrepresentation of mitochondrial gene expression (>50%) as a further quality

control. In addition, we removed any genes that were not represented in a robust population of cells (<8 cells per gene) from the downstream analysis. This resulted in a final analysis of 1,119 single-cell profiles. Using the `RegressOut` feature in Seurat, we calculated the *z*-score residuals using `nGene` and `percent.mito` as co-variables, which was used to perform principal-component analysis and tSNE. A G1/S and G2/M score was calculated using the gene score method described below and regressed out as well for HCl001 and HCl010.

Dimensionality reduction, cell cluster identification and differential gene expression analysis. Dimensionality reduction and differential gene expression was performed using the Seurat analysis package version 2.1.0 (refs. ^{13,14}). For the combined and individual patient analysis, highly variable genes in our dataset were identified using the `MeanVarPlot` function with the following parameters: `FindVariableGenes(object = comb, mean.function = ExpMean, dispersion.function = LogVMR, x.low.cutoff = 0.0125, x.high.cutoff = 3, y.cutoff = 0.5)`. These variable genes were then used for principal-component analysis. The principal components generated were then used to perform tSNE of the data. For the individual patient analysis, using the `FindClusters` function in Seurat and a granularity parameter of 1.0, we identified distinct subpopulations and defined marker genes for each of them with the `FindAllMarkers` function in Seurat with the default settings for the `FindAllMarkers` function and the 'bimod' statistical test. For the generation of the 330-micrometastatic-gene signature, metastatic cells from all PDX models were grouped together separate from tumour cells and we calculated a differential expression test in Seurat using the 'tobit' test with the following parameters: `comb <- FindAllMarkers(object = comb, only.pos = TRUE, min.pct = 0.1, logfc.threshold = 0.25, test.use = 'tobit', latent.vars = 'orig.ident')`. The 'orig.ident' command in the 'latent.vars' variable represents the patient ID (that is, HCl001, and so on). By including this variable, the tobit model identifies conserved marker genes^{13,14,23}. The 'min.pct' variable in Seurat's differential expression (DE) tobit test is defined as the minimum percent of cells per group that must express a gene ($\log(\text{TPM} + 1) > 0$) to be considered in the output of the test. Gene Ontology analysis was performed using the Enrichr web resource, where the input gene set for each population was the markers identified by `FindAllMarkers`²². For the analysis of specific OXPHOS and glycolysis genes, we calculated the average $\log_2[\text{FC}]$ of the raw matrix values using the `FetchData(use.raw=T)` function in Seurat.

Development of logistic regression model for identifying candidate biomarkers.

To calculate the logistic regression model used to identify candidate biomarker targets, the gene expressions were normalized across all cells such that each gene had a mean expression of zero and a s.d. of one. For model fitting, the data were sampled equally ten times from each mouse and cell-identity (tumour versus micrometastasis) category to avoid systematic bias. For each sampling, a stepwise regression with forward selection was performed, where at each step the model that minimized the Akaike information criterion was chosen to be used as a base model for the next step. Our logistic regression model used a conservative cutoff of five genes based on the elbow in the Akaike information criterion plot to minimize the size of our gene set while maintaining descriptive power.

Gene scoring. To quantitatively compare gene groups (for example, glycolysis and OXPHOS) in our cell populations, we utilized individual gene scores as described previously^{15,16}. A curated list of genes was obtained from the Kyoto Encyclopedia of Genes and Genomes, a database used for GO term enrichment. We utilized the `AddModuleScore()` function from Seurat to calculate the gene scores for each population. Full gene lists for all pathways used in the gene-scoring analysis are provided in Supplementary Table 5.

Relapse-free-survival analysis. For the relapse-free-survival analysis, we generated Kaplan–Meier survival curves on primary tumour microarray data of patients with basal-like breast cancer from the KM Plotter database for the top-20 micrometastasis-associated genes in our 330-gene signature^{26,67}. We found that two genes were not predictive and the data for three genes were not available, and these genes were thus excluded. To generate the combined survival analysis, we calculated a weighted average of the 15 predictive genes analysed using the 'Use Multiple Genes' function in KM Plotter. All Kaplan–Meier plots are displayed using the 'auto select best cutoff' parameter.

Histology. Tumour and lung tissues from the PDX mouse models were fixed overnight in 4% paraformaldehyde and then dehydrated and processed for paraffin embedding in a Leica tissue processor using standard protocols. The paraffin blocks were cut into 5- μm -thick sections using a Leica microtome, rehydrated and then stained with haematoxylin and eosin. Bright-field imaging was performed using a BZ-X700 Keyence microscope.

Fluorescent in situ hybridization. Fluorescent in situ hybridization was performed on formalin-fixed paraffin-embedded sections using the RNAscope multiplex fluorescent reagent kit v2 (ACD, cat. no. 323110) according to the manufacturer's instructions. Briefly, the formalin-fixed paraffin-embedded sections were rehydrated in HistoClear and 100% ethanol before antigen retrieval using the RNAscope antigen retrieval solution and mild boiling at 100 °C for 15 min.

PHLDA2 probe (ACD, cat. no. 551441) amplification was performed according to the manufacturer's instructions with the TSA plus cyanine 3 (diluted 1:1,000; PerkinElmer, cat. no. PN NEL744001KT) fluorophore, stained with DAPI and mounted with Prolong Gold. The slides were visualized using a Zeiss LS700 confocal microscope. Image analysis was performed in ImageJ. The normalized total fluorescence intensity of the PHLDA2 probe was calculated on regions of interest across at least five different fields of view on two mouse lungs and three tumours from HCl001 according to the following equation:

$$\text{Normalized TFI} = \frac{\text{Fluorescence integrated density}}{\text{Area of region of interest}}$$

The surrounding mouse stroma was excluded from the analysis for the quantification of lung micrometastatic regions of interest and the necrotic regions or mouse stroma were excluded from the analysis for the tumour regions of interest.

Quantitative real-time PCR. Total RNA was extracted using a Quick-RNA MicroPrep kit (Zymo, cat. no. R1050). The messenger RNA was reverse transcribed into cDNA using the iScript cDNA synthesis kit (BioRad, cat. no. 1708891) according to the manufacturer's protocol. The quantitative real-time PCR reactions were performed using PowerUp SYBRGreen master mix (Applied Biosystems, cat. no. A25742). The primer sequences used for each gene are listed in Supplementary Table 7.

OCR and ECAR measurements. MDA-MB-231 or 4T1-GFP (8×10^4) cells were seeded into Seahorse XF24 microplates (Agilent, cat. no. 100850), and the XF24 cartridge (Agilent, cat. no. 100850) was calibrated in the Seahorse prep station (Agilent) overnight. The medium was replaced with 0.5 ml XF base medium (Agilent, cat. no. 102353) supplemented with 10 mM glucose, 2 mM glutamine and 1 mM pyruvate before the assay. The cells were incubated at 37 °C for 1 h in the Seahorse prep station. Next, 56 μl oligomycin (2 μM ; Sigma, cat. no. 75351), 62 μl FCCP (2 μM ; Sigma, cat. no. C2920) and 69 μl rotenone with antimycin A (1 μM ; Sigma, cat. no. 557368, A8674) were added into the cartridge wells. The OCR and ECAR levels were determined using Seahorse bioscience XF24 extracellular flux analyzer (Agilent) and each cycle of measurement involved mixing (3 min), waiting (2 min) and measuring (3 min) cycles. To quantify the basal respiration, we calculated the difference in the average OCR before the addition of oligomycin and the average OCR after the addition of rotenone with antimycin A. For ATP-dependent respiration, we calculated the difference in average OCR before and after the addition of oligomycin. The maximal respiration rates were calculated as the difference in average OCR after the addition of FCCP and the average OCR after the addition of oligomycin. Any negative values were set to zero for visualization and when calculating differences between groups.

FLIM. All FLIM imaging experiments were performed as previously described⁶⁸. The FLIM images were acquired on an inverted laser scanning confocal microscope with a $\times 40/1.2$ numerical aperture water-immersion objective with cells kept under biological conditions throughout. The cells (4T1-GFP and MDA-MB-231) were excited at approximately 2 mW with a two-photon Ti:Sapphire laser (Spectra-Physics, MaiTai) at 740 nm, which was passed through a 690 nm dichroic filter. The fluorescence emission was separated through a bandpass filter (442/46 nm) to capture the cell auto-fluorescence and exclude any GFP emission, and was then detected using a photomultiplier tube (Hamamatsu, cat. no. H7422P-40). The fluorescence lifetime decays were captured in the frequency domain via an A320 FastFLIM box (ISS) and subsequently mapped onto the phasor plot using the SimFCS software—developed at the Laboratory for Fluorescence Dynamics at the University of California, Irvine—for quantitative NADH lifetime analysis. Coumarin-6 in ethanol, with a known single exponential fluorescence lifetime of 2.5 ns, was imaged before each experiment and used as a calibration sample for the instrument response time.

FLIM data analysis. Each pixel of the FLIM image was Fourier transformed and plotted on the phasor plot in SimFCS, as previously described⁶⁸. Each FLIM image was individually masked to exclude any nuclei from the analysis. The cluster of phasors of each masked frame were then averaged to obtain its G and S coordinate, which was then used to determine the fraction of free NADH of each cluster. This free NADH fraction was calculated on the fact that 100% free NADH has a decay of approximately 0.4 ns and 100% bound NADH has a decay of approximately 3.2 ns. Any pixel with a fluorescence decay containing a mixture of both free and bound NADH will lie within the linear combination of 100% free and 100% bound NADH when on the phasor plot in the frequency domain. Consequently, the fraction of free NADH of the cluster can be calculated by the mathematical distance between the positions of the 100% free NADH and 100% bound NADH on the universal circle of the cluster. For each independent experiment, the fraction of free NADH of the treated sample was subtracted from the average of the fraction of free NADH of the control sample to obtain the fraction of free NADH difference. Significance was determined using a two-tailed Student's *t*-test with homoscedasticity determined using an *F*-test. The error bars represent the s.e.m.

Generation of steady-state metabolomics data using LC–HRMS. Approximately 1,000 cells were sorted using flow cytometry from primary tumours and lung metastatic cells from six HClO10 PDX transplants in NSG mice. The cells were centrifuged to remove the medium and extracted in pre-cooled 80% high performance liquid chromatography-grade methanol in high performance liquid chromatography-grade water. The samples were placed at -80°C for 15 min. The samples were then placed on ice and mixed several times. All samples were then centrifuged at 20,000 r.c.f. for 10 min at 4°C . The supernatant was transferred to a new tube and dried using a speed vacuum. The samples were prepared and analysed by LC–HRMS, as previously described³⁰. The integrated peak intensities were used. Hierarchical clustering and heat maps were generated using the MetaboAnalyst software (<https://www.metaboanalyst.ca/>) with the samples normalized using the Autoscaling feature, the false detection rate was set at 5% and metabolites were ranked by *t*-tests. Pathway enrichment analysis was conducted using MetaboAnalyst; briefly, metabolite identifications from the human metabolome database from the metabolites that were significantly enriched ($\text{FC} > 1.5$ and $P > 0.05$) were inputted. The pathway library used was *Homo sapiens* and a hypergeometric test was used for overrepresentation analysis. Partial least-squares discriminant analysis was implemented using the *plsda* function in the R package 'mixOmics'⁶⁹.

Statistics and reproducibility. Quantitative data are presented as either the mean \pm s.e.m. or mean \pm s.d. from at least three independent experiments unless otherwise specified. When appropriate, statistically significant differences between groups were analysed using a two-way unpaired Student's *t*-test. Differences were considered significant at $P < 0.05$. All samples—if preserved and properly processed—were included in the analyses and no samples or animals were excluded, unless otherwise specified. No statistical method was used to predetermine the sample sizes and the female mice were randomly assigned PDX tumours for transplant. The investigators were not blinded to allocation during experiments and outcome assessment.

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

Data availability

The authors declare that all data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding author on reasonable request. All RNA-seq data files along with their associated metadata have been deposited in the GEO database under the accession code GSE123837. Previously published microarray data that were reanalysed here with KM Plotter^{26,67} are available under the following accession codes: E-MTAB-365, E-TABM-43, GSE11121, GSE12093, GSE12276, GSE1456, GSE16391, GSE16446, GSE16716, GSE17705, GSE17907, GSE18728, GSE19615, GSE20194, GSE20271, GSE2034, GSE20685, GSE20711, GSE21653, GSE2603, GSE26971, GSE2990, GSE31448, GSE31519, GSE32646, GSE3494, GSE37946, GSE41998, GSE42568, GSE45255, GSE4611, GSE5327, GSE6532, GSE7390 and GSE9195.

Code availability

Custom scripts are available at https://github.com/lawsonlab/Single_Cell_Metastasis.

References

- Nagy, Á., Lániczky, A., Menyhart, O. & Györfy, B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. *Sci. Rep.* **8**, 9227 (2018).
- Mah, E. J., Lefebvre, A. E. Y. T., McGahey, G. E., Yee, A. F. & Digman, M. A. Collagen density modulates triple-negative breast cancer cell metabolism through adhesion-mediated contractility. *Sci. Rep.* **8**, 17094 (2018).
- Rohart, F., Gautier, B., Singh, A. & Lê Cao, K. A. mixOmics: an R package for 'omics feature selection and multiple data integration. *PLoS Comput. Biol.* **13**, e1005752 (2017).

Acknowledgements

We thank G. Gutierrez and M. Masoud for technical assistance and animal handling. We thank N. Pervolarakis for helpful advice on data analysis. We thank K. Kessenbrock for thoughtful feedback on experimental design and careful review of the manuscript. We thank A. L. Welm for providing the PDX samples. Image acquisition was made possible in part through access to the Optical Biology Core Facility of the Developmental Biology Center, a shared resource supported by the Cancer Center Support Grant (grant no. CA-62203), with assistance from A. Syed and a Center for Complex Biological Systems Support Grant (grant no. GM-076516) at the University of California, Irvine. This study was supported by funds from the National Cancer Institute (grant nos R01 CA057621 and U01 CA199315 to Z.W., and grant no. K22 CA190511 to D.A.L.), National Institutes of Health (grant no. R01HD073179 to E.M., P41-GM103540 to M.A.D. and A.E.Y.T.L., and T32CA009054 to M.B.G. and R.T.D., through matched university funds through matched university funds), National Science Foundation (grant no. 1847005 to M.A.D. and NSF GRFP DGE-1839285 to A.E.Y.T.L.), Team Michelle and Friends non-profit organization, Suzette Kirby Breast Cancer Research Fund, V Foundation (grant no. V2019-019) as well as an Opportunity Award funded by the UCI Center for Complex Biological Systems (CCBS; NIGMS, grant no. P50-GM076516 to R.T.D., K.B., D.Maurer, E.M. and D.A.L.). H.A. was supported by the University of Hail, Hail, Saudi Arabia for the PhD Fellowship. D.Ma was supported by a Canadian Institutes of Health Research Postdoctoral Fellowship.

Author contributions

D.A.L., E.M., M.K., M.A.D., J.W.L. and Z.W. designed and supervised the research. R.T.D., Y.Y., K.B., M.B.G., D.Ma, A.E.Y.T.L., A.T.P., H.A., G.A.H., J.L. and D.A.L. performed the research. R.T.D., K.B., Z.X. and D.Maurer performed the bioinformatic analyses. R.T.D. and D.A.L. wrote the manuscript, and all authors discussed the results and provided comments and feedback.

Competing interests

The authors declare no competing interests.

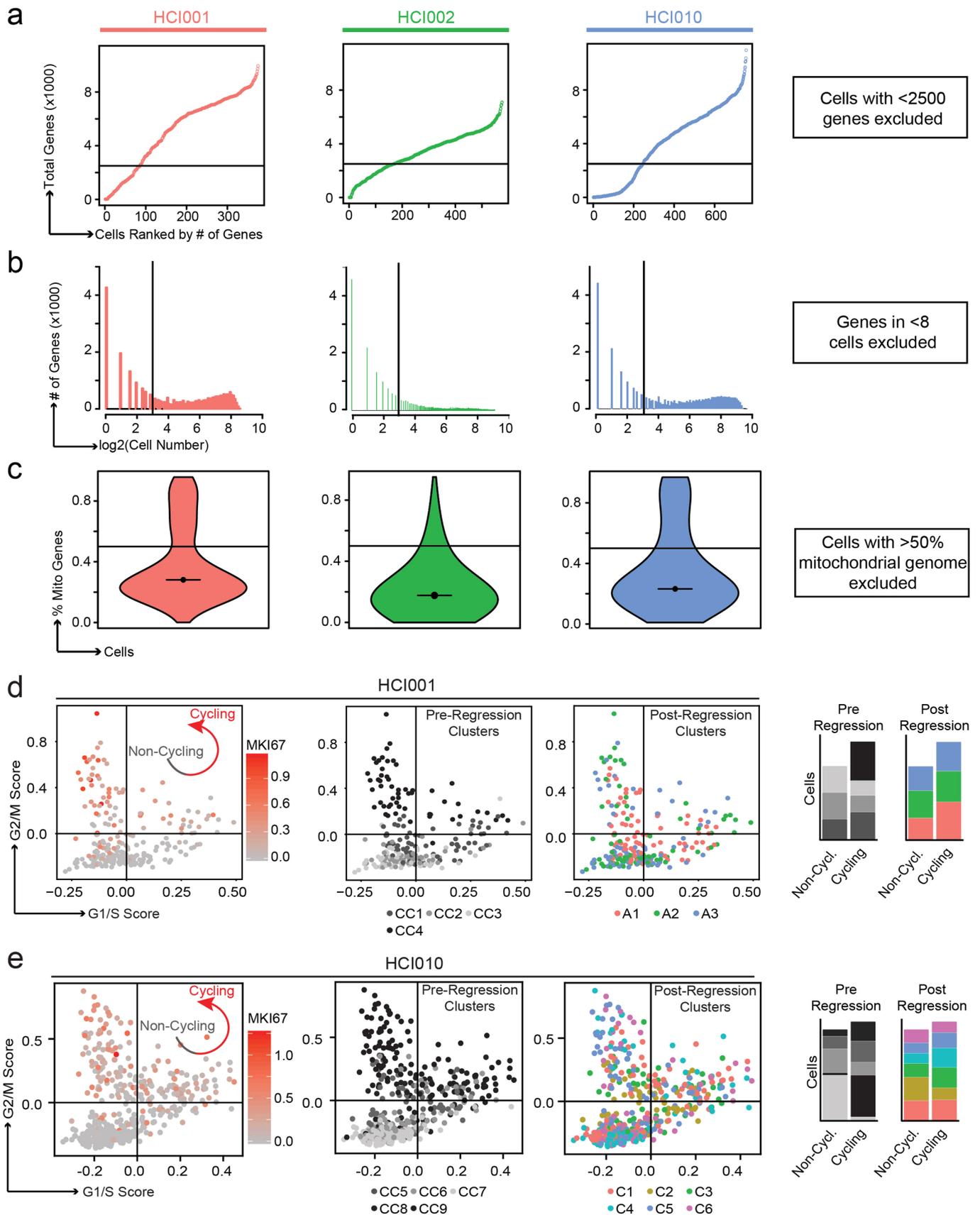
Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41556-020-0477-0>.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41556-020-0477-0>.

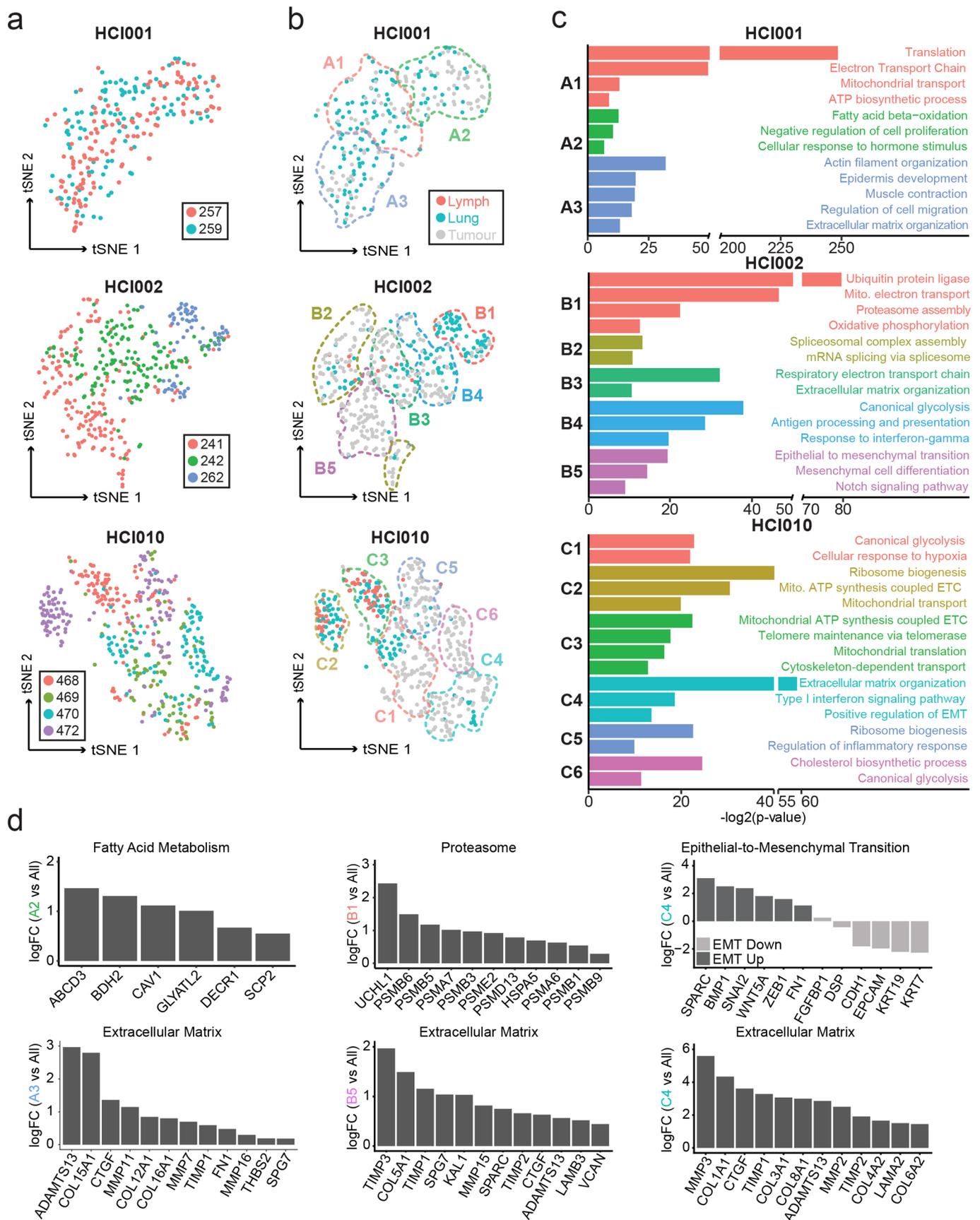
Correspondence and requests for materials should be addressed to D.A.L.

Reprints and permissions information is available at www.nature.com/reprints.



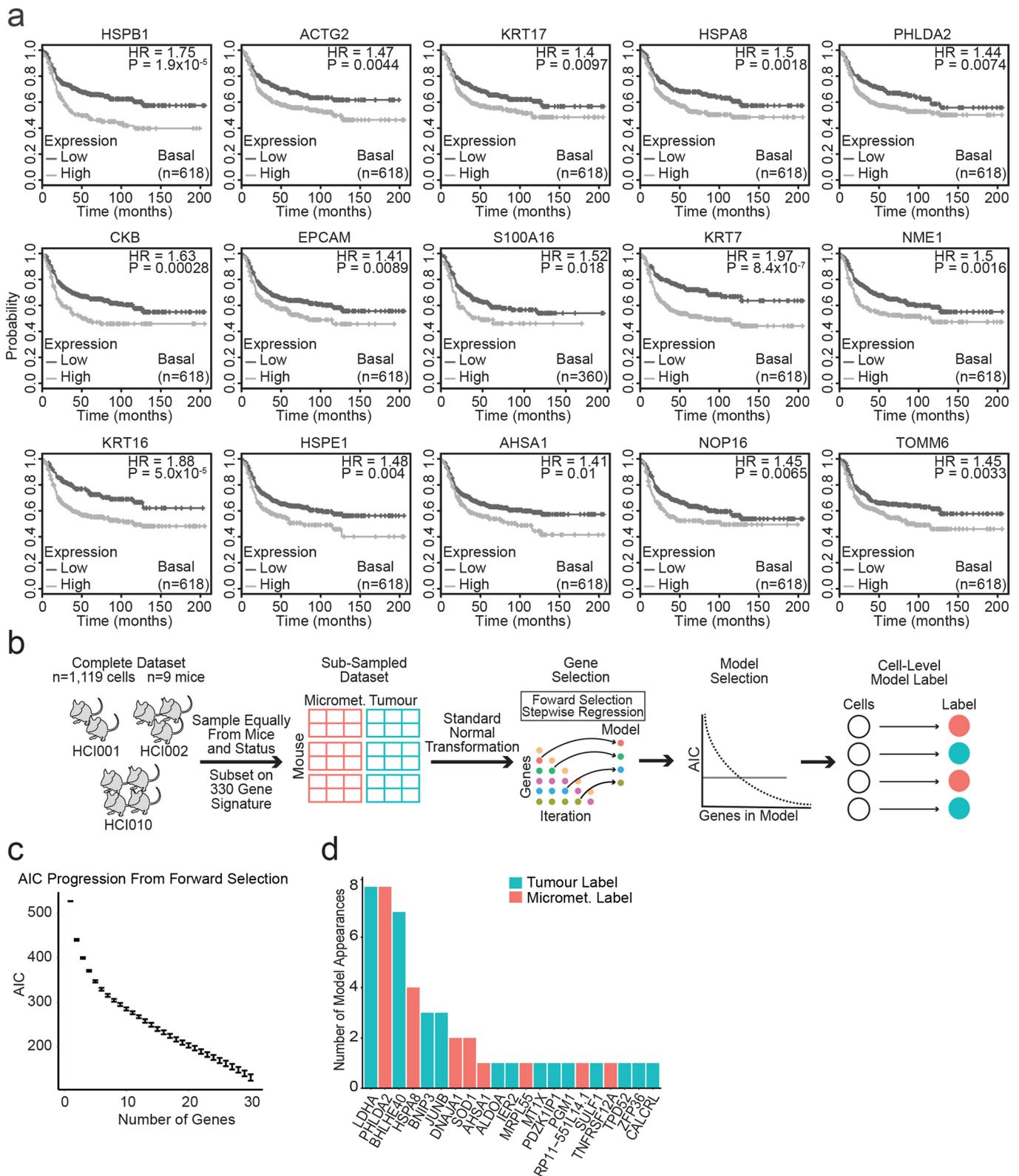
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Quality control and exclusion criteria for single-cell RNA sequencing. **a**, Identification and removal of poor-quality cell libraries. Plots show the number of genes detected in each cell from each PDX model. Cells (x axis) are ordered from fewest to most genes detected. Cells with fewer than 2500 genes detected (horizontal line) were excluded. **b**, Identification and removal of noisy/background genes. Plots show the distribution of genes detected per cell. Vertical line indicates that genes detected in fewer than 8 cells were excluded. **c**, Violin plots indicate the expression of mitochondrial genes as a percentage of total gene expression in each cell separated by HCl001 ($n=375$ cells), HCl002 ($n=576$ cells), and HCl010 ($n=756$ cells). Cells were excluded if $>50\%$ of genes detected were mitochondrial (horizontal line). Black line with dot indicates median. **(d, e)** Cells from HCl001 (**d**) and HCl010 (**e**) are plotted based on the relative expression of gene sets associated with G1/S (x-axis) and G2/M (y-axis) stages of the cell cycle. Left plots: Relative expression of the proliferation-associated gene *MKI67* is shown in HCl001 and HCl010. Middle plots: Cell clustering before cell cycle regression. Cluster identities are shown in grey. Right plots: Cell clustering after cell cycle regression. Colours indicate new cluster identity of each cell and correspond with clusters shown in Fig. 2a. Bar graphs show quantification of cell clusters by cycling status before and after regression.

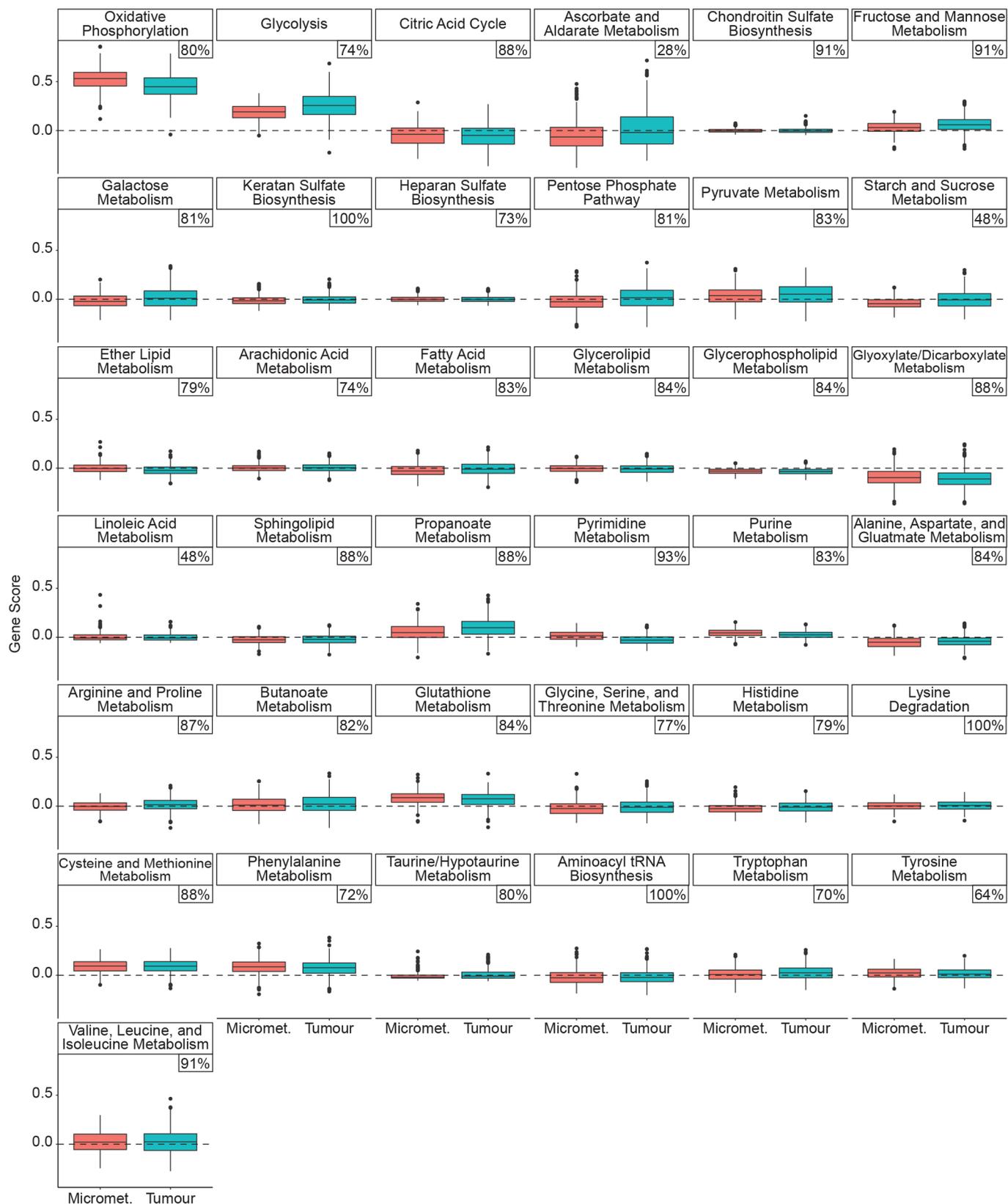


Extended Data Fig. 2 | See next page for caption.

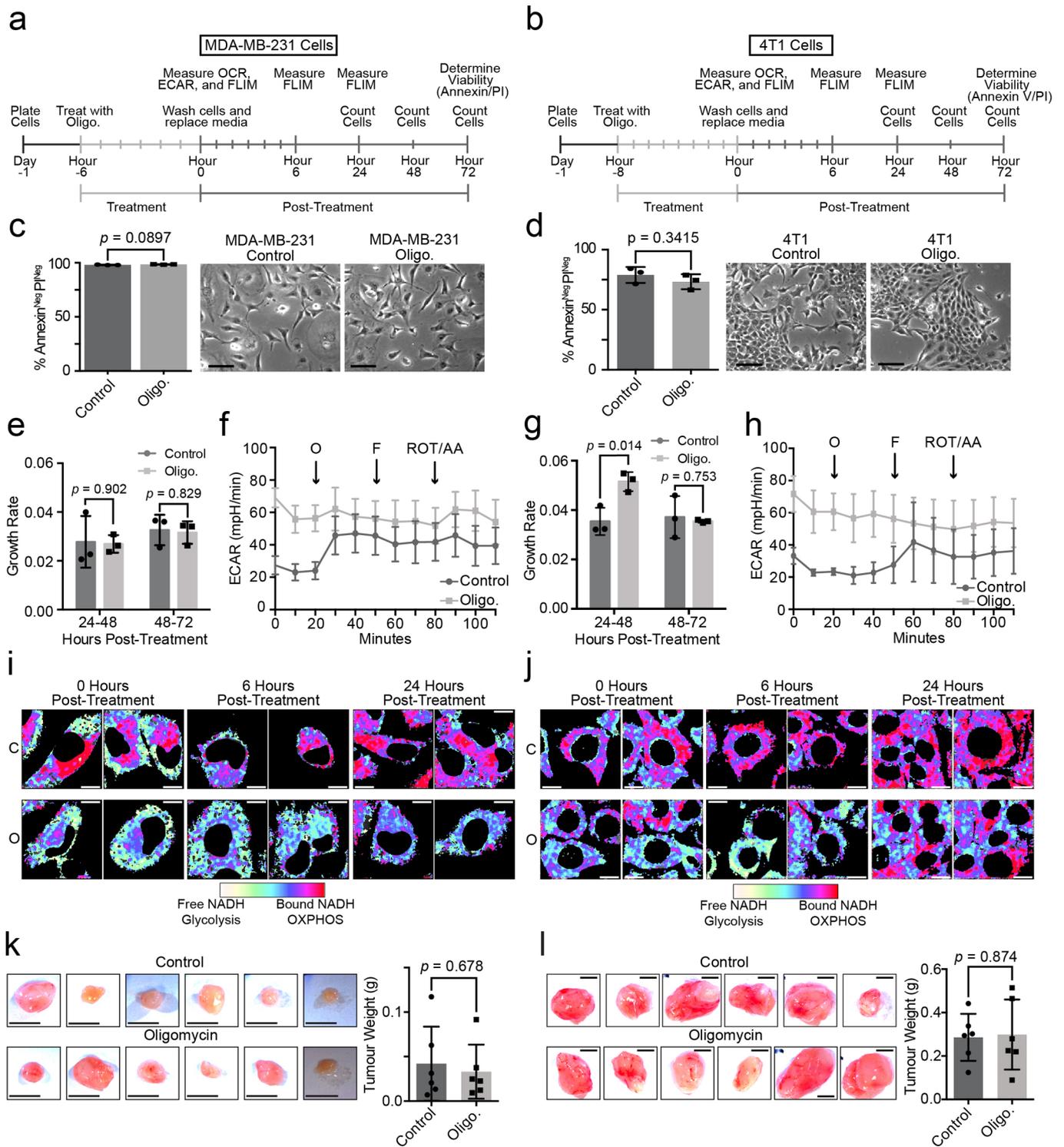
Extended Data Fig. 2 | Marker and GO term analysis of cell clusters from each PDX model. **a**, tSNE plots display clustering of cells coloured by mouse of origin from PDX models HCl001 ($n=247$ cells), HCl002 ($n=401$ cells) and HCl010 ($n=471$ cells). **b**, tSNE plots display clustering of cells coloured by tissue of origin from PDX models HCl001 ($n=247$ cells), HCl002 ($n=401$ cells) and HCl010 ($n=471$ cells). **c**, Bar plots show selected top GO terms determined by the marker genes identified for each cell cluster. P values are determined by the Fisher exact test. Full marker gene lists utilized are listed in Supplementary Table 1. Specifically, for HCl001, $n=162$ A1 genes, $n=107$ A2 genes, and $n=199$ A3 genes. For HCl002, $n=490$ B1 genes, $n=173$ B2 genes, $n=34$ B3 genes, $n=181$ B4 genes, and $n=194$ B5 genes. For HCl010, $n=96$ C1 genes, $n=247$ C2 genes, $n=198$ C3 genes, $n=357$ C4 genes, $n=54$ C5 genes, $n=110$ C6 genes. **d**, Bar graphs show the log fold change (logFC) for selected genes from GO term pathways. Values indicate the logFC of the average gene expression for the indicated cell cluster relative to all other clusters within that PDX model.



Extended Data Fig. 3 | Prognostic value of micrometastasis-associated genes in basal-like breast cancer patients. **a**, Kaplan-Meier curves show relapse free survival (RFS) in basal-like breast cancer patients from the KM plotter database (879 patients), based on their primary tumour expression of specified micrometastasis-associated genes. *P* values were determined via a log-rank test. **b**, Schematic for the construction of a stepwise logistic regression model to identify top biomarker candidates descriptive of primary tumour or micrometastatic cells. The data was subsampled to analyse equal numbers of micrometastatic and tumour cells from each mouse. The model was run on 10 subsamplings of the data, with the number of genes in each model determined by AIC. **c**, Plot demonstrating the AIC versus the number of genes included in the model. AIC is used to balance parameter additions (that is gene additions) with the descriptive power of a model. Data is presented as the 10% and 90% quantiles of the 10 data subsamplings. **d**, Bar plot showing the number of model appearances for each gene out of 10 data subsamplings.

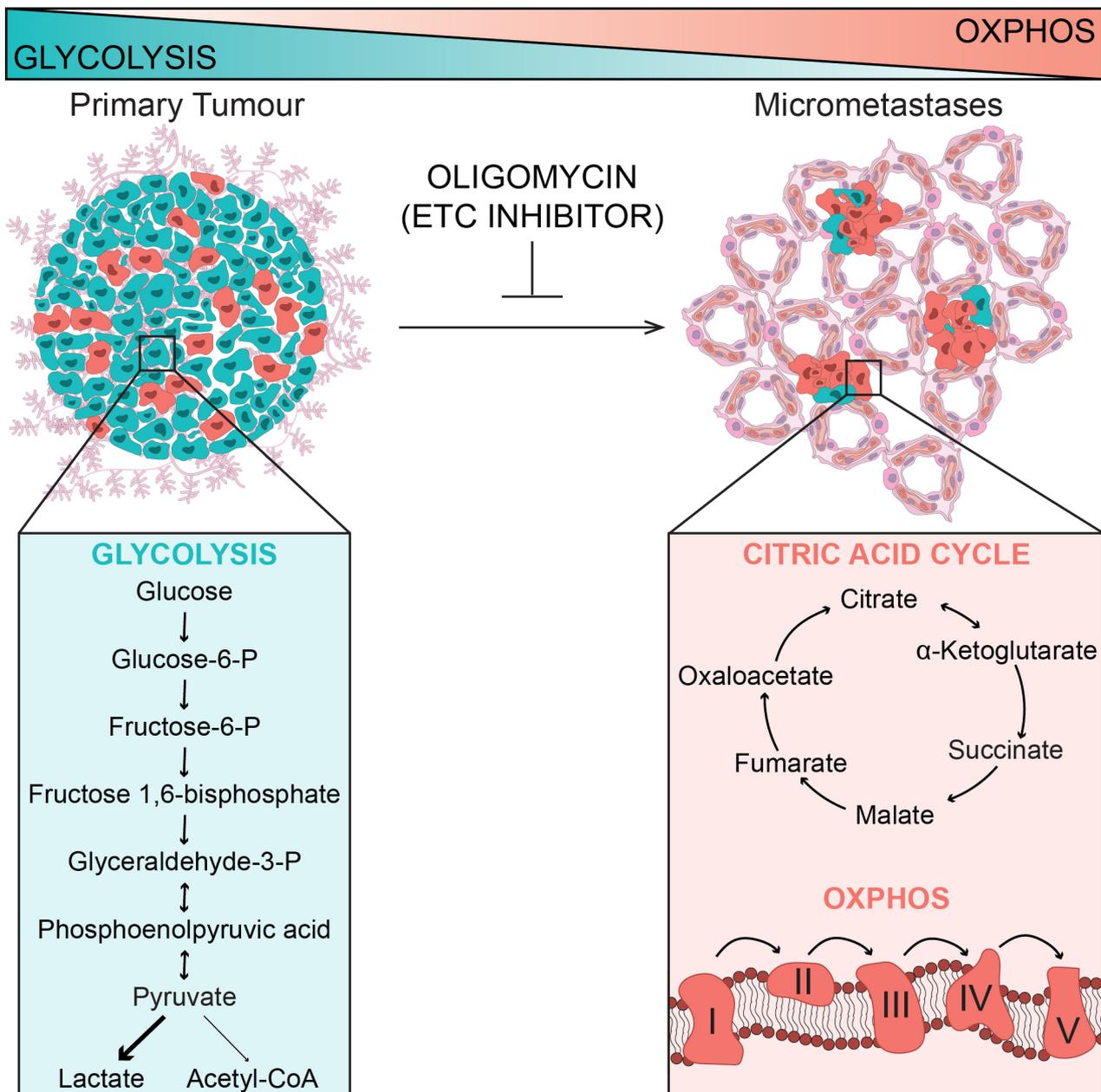


Extended Data Fig. 4 | Gene scoring identifies OXPPOS and glycolysis as top metabolic pathways differentially expressed between micrometastases and primary tumour cells. Gene scores for each metabolic pathway in micrometastatic (red, $n = 435$ cells) or primary tumour cells (blue, $n = 684$ cells). Each cell in the dataset was scored by calculating the difference between the average gene expression for all the genes in each metabolic pathway versus the average gene expression of a randomly selected background gene set. Dotted line represents a zero score, which indicates the metabolic pathway is not differentially expressed relative to the background gene set. The boxed value (top right of each plot) indicates the percent of genes in the pathway that was detected in the dataset. The lower and upper hinges correspond to the first and third quartiles, and the midline represents the median. The upper and lower whiskers extend from the hinge up to $1.5 \times$ IQR (inter-quartile range). Outlier points are indicated if they extend beyond this range.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Oligomycin treatment inhibits OXPHOS in MDA-MB-231 and 4T1-GFP cells. **a,b**, Schematic of the experimental setup to determine the effects of oligomycin treatment on MDA-MB-231 (**a**) or 4T1-GFP (**b**) cells. Oligo. = oligomycin. (**c-d**) Bar graph (left) of the viability of MDA-MB-231 (**c**) or 4T1-GFP (**d**) cells 72-hours post-treatment determined via flow cytometry. Data is presented as the mean \pm s.d. for $n=3$ replicates. P values determined by unpaired, two-sided Student's t -test. Brightfield images (right) of in vitro MDA-MB-231 (**c**) or 4T1-GFP (**d**) cells 72 hours after cessation of oligomycin treatment. Scale bar = 100 μm . **e**, Growth rate of MDA-MB-231 cells for the indicated time period after cessation of oligomycin treatment. Data is presented as mean \pm s.d. of $n=3$ replicates. P values determined by unpaired, two-sided Student's t -test. **f**, Extracellular acidification rate (ECAR) of MDA-MB-231 treated with oligomycin compared to control cells. ECAR was measured at the conclusion of treatment with oligomycin as described in (**a**). Arrows indicate when drugs were added. O=oligomycin, F=FCCP, ROT/AA=Rotenone/ Antimycin A. Data is presented as mean \pm s.d. of $n=4$ replicates. **g**, Same as (**e**) for 4T1-GFP cells. **h**, Same as (**f**) for 4T1-GFP cells. Data is presented as mean \pm s.d. of $n=3$ replicates. P values determined by unpaired, two-sided Student's t -test. (**i, j**) Additional FLIM images of the fluorescence lifetime of NADH in cultured MDA-MB-231 (**i**) or 4T1-GFP cells (**j**) as shown in Fig. 5e, f. Fields do not represent consecutive images of the same cell. O=oligomycin, C=control. Scale bar=10 μm . (**k-l**) Brightfield images of tumours from orthotopically injected control or oligomycin treated MDA-MB-231 (**k**) or 4T1-GFP (**l**) cells. Bar graphs indicate tumour weights (right). Data presented as mean \pm s.d. of MDA-MB-231 ($n=6$ oligomycin-treated, $n=6$ control) and 4T1-GFP ($n=6$ oligomycin-treated, $n=6$ control) tumours. P -values were determined by unpaired, two-sided Student's t -test. Scale bar = 0.5 cm.



Extended Data Fig. 6 | Model for metabolic shift associated with metastatic seeding in TNBC. Metastatic cells in the lung and lymph nodes display increased OXPHOS, in contrast to primary tumour cells that express higher levels of genes associated with aerobic glycolysis. Pharmacological inhibition of OXPHOS with the complex V inhibitor oligomycin substantially attenuates lung metastasis in experimental models of TNBC, showing that OXPHOS is functionally important for metastasis.

Reporting Summary

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

Statistics

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

n/a Confirmed

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

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Paired-end 100bp reads were aligned to the Gencode 21 human transcriptome using Bowtie2 2.2.3 and quantified using RSEM 1.2.12. FACS data was collected using FACS diva software v6.1.1

Data analysis

Analysis was performed in R (version 3.4.1) using publicly available packages. Dimensionality reduction and differential gene expression was performed using the Seurat analysis package (version 2.1.0). Gene ontology analysis was performed using the Enrichr web resource. For the relapse-free survival analysis, we generated Kaplan-Meier (KM) survival curves on basal-like breast cancer patient primary tumour microarray data from the KM plotter database. Image analysis was performed in ImageJ. Each pixel of the FLIM image was Fourier transformed and plotted on the phasor plot in SimFCS. Hierarchical clustering and heat map were generated using MetaboAnalyst software. Partial least squares-discriminant analysis (PLS-DA) was implemented using the plsda function in R package "mixOmics". Flow cytometry data was analyzed using FlowJo software (version 10).

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

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

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

The authors declare that all data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding author upon reasonable request. All RNAseq data files along with their associated meta data have been deposited in the GEO database under accession code GSE123837.

Field-specific reporting

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

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

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

| | |
|-----------------|---|
| Sample size | Sample sizes were not pre-determined; however, we ensured that each patient was represented by multiple xenograft passages and that the major differences in cell populations were not due to mouse-to-mouse variation. |
| Data exclusions | No data were excluded from analyses presented in the manuscript unless otherwise specified. |
| Replication | All attempts at replication were successful. |
| Randomization | Female mice were randomly assigned PDX tumors for transplant. |
| Blinding | Blinding of study parameters was not performed in order to monitor tumor growth effectively. |

Reporting for specific materials, systems and methods

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

Materials & experimental systems

| n/a | Involvement in the study |
|-------------------------------------|---|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Antibodies |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Eukaryotic cell lines |
| <input type="checkbox"/> | <input type="checkbox"/> Palaeontology |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Animals and other organisms |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Human research participants |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |

Methods

| n/a | Involvement in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Antibodies

| | |
|-----------------|---|
| Antibodies used | Human-specific antibody CD298, diluted 1:100 (PE, BioLegend, Cat. No. 341704, Clone No. LNH-94, Lot No. B269993), the mouse-specific antibody MHC-I, diluted 1:150 (APC, ThermoFisher, Scientific Cat. No. 17-5957-80, Clone No. SF1-1.1.1, Lot No. 2044729) |
| Validation | <p>Human-specific antibody CD298: The antibody was purified by affinity chromatography, and conjugated with PE under optimal conditions. The solution is free of unconjugated PE and unconjugated antibody. This LNH-94 antibody has been reported for use in flow cytometric analysis in human peripheral blood lymphocytes. References: Chiampanichayakul S, et al. Tissue Antigens 2006; Malik N, et al. J. Biol. Chem. 1996</p> <p>Mouse-specific antibody MHC-1: This SF1-1.1.1 antibody has been reported for use in flow cytometric analysis. This SF1-1.1.1 antibody has been tested by flow cytometric analysis on Balb/c splenocytes. This antibody can be used at less than or equal to 0.125 µg per test. A test is defined as the amount (µg) of antibody that will stain a cell sample in a final volume of 100 µL. Cell number should be determined empirically but can range from 10⁵ to 10⁸ cells/test. It is recommended that the antibody be carefully titrated for optimal performance in the assay of interest. References: Kindy MS, et al. J. Transl. Med. 2016; Larsen J, et al. Eur. J. Immunol. 2014; Luo Y, et al. Stem Cells 2012; Horst D, et al. Cancer Res. 2012</p> <p>We have independently performed extensive validation analysis for CD298 and MHCI (Lawson DA et al., Nature, 2015).</p> |

Eukaryotic cell lines

Policy information about [cell lines](#)

| | |
|--|--|
| Cell line source(s) | MDA-MB-231 and 4T1 cell lines were obtained from the American Type Culture Collection (ATCC) |
| Authentication | MDA-MB-231 cells were validated by short-tandem repeat profiling by ATCC |
| Mycoplasma contamination | The cell lines were not confirmed to be mycoplasma-free |
| Commonly misidentified lines (See ICLAC register) | No commonly misidentified cell lines were used. |

Palaeontology

| | |
|---------------------|--|
| Specimen provenance | <i>Provide provenance information for specimens and describe permits that were obtained for the work (including the name of the issuing authority, the date of issue, and any identifying information).</i> |
| Specimen deposition | <i>Indicate where the specimens have been deposited to permit free access by other researchers.</i> |
| Dating methods | <i>If new dates are provided, describe how they were obtained (e.g. collection, storage, sample pretreatment and measurement), where they were obtained (i.e. lab name), the calibration program and the protocol for quality assurance OR state that no new dates are provided.</i> |

Tick this box to confirm that the raw and calibrated dates are available in the paper or in Supplementary Information.

Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

| | |
|-------------------------|---|
| Laboratory animals | Orthotopic transplants of serially passaged human tumour samples were performed on immunocompromised 3-4-week-old female NOD/SCID or NSG mice after clearing the mammary fat pads following established protocols. For experimental metastasis studies, cultured MDA-MB-231 or 4T1-GFP were suspended in 100ul of sterile PBS and injected into the tail vein of 8-10 week old female Nod/SCID (MDA-MB-231) or BALB/c (4T1-GFP) mice. For cell line orthotopic tumor injections, 1x10 ⁶ 4T1-GFP or MDA-MB-231 cells in 100 µl phosphate buffered saline (PBS) were injected bilaterally into the fourth mammary fat pad of 10-12 week old female NSG mice. |
| Wild animals | This study did not include the use of wild animals |
| Field-collected samples | This study did not include field-collected samples. |
| Ethics oversight | The University of California, Irvine Institutional Animal Care and Use Committee (IACUC) reviewed and approved all animal experiments under AUP-19-051. |

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

Human research participants

Policy information about [studies involving human research participants](#)

| | |
|----------------------------|--|
| Population characteristics | Samples from patients were generously provided by A.L. Welm in the Department of Oncological Sciences at the Huntsman Cancer Institute (HCI). Briefly, HCI001 was acquired from a primary tumour biopsy of a Stage IV female patient diagnosed with ER-PR-Her2-, basal-like invasive ductal carcinoma (IDC) with no previous systemic treatment. HCI002 was acquired from a primary tumour biopsy of a female patient diagnosed with ER-PR-Her2-, basal-like Stage IIIA medullary type IDC with no previous systemic treatment. HCI010 was acquired from a pleural effusion of a Stage IIIC female patient diagnosed with ER-PR-Her2-, basal-like (PAM50) IDC treated with several rounds of chemotherapies. |
| Recruitment | All tissue samples were collected with informed consent from individuals being treated at the Huntsman Cancer Hospital and the University of Utah. |
| Ethics oversight | Samples were collected and de-identified by the Huntsman Cancer Institute Tissue Resource and Application Core facility before being obtained for implantation under a protocol approved by the University of Utah Institutional Review Board. |

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

Plots

Confirm that:

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

Methodology

Sample preparation

Animals were euthanized, and tissues were harvested when tumours reached 2.0-2.5cm by length or width. Animals at endpoint were euthanized by asphyxiation with CO₂ followed by cervical dislocation and perfusion with 10mM EDTA in D-PBS. Prior to perfusion, Evan's Blue (Sigma-Aldrich, Cat. No. E2129-10G) was injected into the footpads and ears of anesthetized mice to aid in lymph node visualization. Solid tissues from the mice, which includes the primary tumour, lung, and lymph nodes were processed for flow cytometry by mechanically chopping with blades, followed by Collagenase IV digest (Sigma-Aldrich Cat. No. C5138-1G) in media (DMEM/F12 with 5% FBS, 5µg/mL insulin, and 1% Penstrep/Ampho B) for 45 min at 37°C. Cell suspensions were washed with 2 µg/mL DNaseI for 5 min and further dissociated with 0.05% Trypsin for 10 min. Following a wash with HBBS/2% FBS, cells were passed through a 70µm filter. Lung and primary tumour cells were treated with 1X RBC lysis buffer, followed by resuspension in DMEM/F12 with 10% FBS for FACS. We used the human-specific antibody CD298, diluted 1:100 (PE, BioLegend, Cat. No. 341704) and the mouse-specific antibody MHC-I diluted, diluted 1:200 (APC, ThermoFisher, Scientific Cat. No. 17-5957-80). Flow cytometry was performed using the BD FACSAria Fusion cell sorter. Cell viability was determined by negative staining for SYTOX Blue, diluted 1:1000 (ThermoFisher Scientific Cat. No. S34857). Forward scatter area by forward scatter width (FSC H x FSC A) and side scatter area by side scatter width (SSC H x SSC A) was used to discriminate single cells from doublet and multiplet cells. Mouse cells were excluded by gating out CD298-MHCI+. Human primary tumour cells and metastatic cells were selected by gating on Sytox-CD298+MHC-I. For mitochondrial membrane potential readouts, cells were stained with TMRM (ThermoFisher Scientific Cat. No. T668) and MitoTracker-Green (ThermoFisher Scientific Cat. No. M7514). Compensation controls and FMOs were used to determine TMRM(High) and TMRM(Low) populations in the PDX tumour cells. For analysis of MDA-MB-231 metastatic burden, FSC W x FSC A and SSC W x SSC A was used to discriminate single cells from doublet and multiplet cells. MDA-MB-231 cells were defined by gating on Sytox-CD298+MHC-I, and the level of metastatic burden was determined by the frequency of that population. For analysis of 4T1-GFP metastatic burden, FSC H x FSC A and SSC H x SSC A was used to discriminate single cells from doublet and multiplet cells. 4T1-GFP cells were defined by gating on Sytox-GFP+ cells, and the level of metastatic burden was determined by the frequency of that population.

Instrument

BD FACSAria Fusion cell sorter

Software

FACS data was collected using FACS diva software v6.1.1, and analyzed using FlowJo software v10.

Cell population abundance

Post-sort analyses were performed in validation studies to determine the purity of sorted samples, as well as the accuracy and yield of single-cell sort experiments into 96 well plates. Using stringent single-cell sorting parameters, we routinely achieve >95% purity of cells sorted using the CD298/MHCI paradigm. In post-sort analyses of single-cell RNA sequencing data, we remove cells which do not meet standards for quality, viability, or species identity.

Gating strategy

Cell viability was determined by negative staining for SYTOX Blue (ThermoFisher Scientific Cat. No. S34857). Forward scatter area by forward scatter width (FSC W x FSC A) and side scatter area by side scatter width (SSC W x SSC A) was used to discriminate single cells from doublet and multiplet cells. Mouse cells were excluded by gating out CD298-MHCI+. Human primary tumour cells and metastatic cells were selected by gating on Sytox-CD298+MHC-I. or analysis of MDA-MB-231 metastatic burden, FSC W x FSC A and SSC W x SSC A was used to discriminate single cells from doublet and multiplet cells. MDA-MB-231 cells were defined by gating on Sytox-CD298+MHC-I, and the level of metastatic burden was determined by the frequency of that population. For analysis of 4T1-GFP metastatic burden, FSC H x FSC A and SSC H x SSC A was used to discriminate single cells from doublet and multiplet cells. 4T1-GFP cells were defined by gating on Sytox-GFP+ cells, and the level of metastatic burden was determined by the frequency of that population. For analysis of cell viability following oligomycin treatment, cells were stained with Annexin VFITC, diluted 1:100 (GeneTex Cat. No. GTX14082) and propidium iodide (PI), diluted 1:100 (ThermoFisher Scientific Cat. No. P3566). FSC W x FSC A and SSC W x SSC A was used to discriminate single cells from doublet and multiplet cells. Viability was determined by the frequency of Annexin VNegPINeg cells in the population.

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