Lymphatic system is the mainstream for breast cancer dissemination and metastasis revealed by single-cell lineage tracing

Kai Miao

University of macau

Aiping Zhang

University of macau

Fangyuan Shao

University of Macau

Lijian Wang

1, 2 Cancer Center, Centre for Precision Medicine Research and Training, Faculty of Health Sciences, University of Macau, Macau SAR

Xin Zhang

University of Macau

Heng Sun

Cancer Centre, Faculty of Health Sciences, University of Macau

Jun Xu

University of macau

Jingyao Zhang

China Agricultural University

Yuzhao Feng

University of Macau

Sen Guo

University of Macau <https://orcid.org/0000-0002-2681-0576>

Zhihui Weng

University of Macau

Dong Wang

Southern Medical University <https://orcid.org/0000-0002-6860-6864>

Shuai Gao

China Agricultural University

Xiaoyang Zhao

Southern Medical University <https://orcid.org/0000-0003-2544-8293>

Xiaoling Xu

University of Macau <https://orcid.org/0000-0002-0852-0225>

Chu-Xia Deng (✉ cx deng@um.edu.mo)

Faculty of Health Sciences, University of Macau

Article

Keywords: lymphatic metastasis, lineage tracing, continuous dissemination model, IL-2

Posted Date: December 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1193629/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Cancer metastasis is the primary cause of cancer-related death, yet the forces that drive cancer cells through various steps and different routes to distinct target organs/tissues remain elusive. In this study, we applied a CellTag system-based single-cell lineage tracing approach to show the metastasis rate and route of breast cancer cells and their interactions with the tumour microenvironment (TME) during metastasis. The results indicate that only a small fraction of cells can intravasate from the primary site into the blood circulation, whereas more cells disseminate through the lymphatic system to different organs. Tumour cells derived from the same progenitor cell exhibit different gene expression patterns in different soils, and the cancer cell-TME communication paradigm varies significantly between primary and metastatic tumours. Furthermore, metastable cells require a prewired IL-2 expression ability to migrate in vivo. In summary, leveraging a single-cell lineage tracing system, we demonstrate that the crosstalk between tumour cells and the TME is the driving force controlling the preferential metastatic fate of cancer cells through the lymphatic system and that this metastasis can be suppressed by knockdown of IL-2.

Introduction

Cancer metastasis is a process through which tumour cells from the primary tumour invade through the basement membrane into blood or lymphatic vessels (intravasation), travel through the blood or lymphatic system, and colonize distant organs/tissues for further development. Metastasis is responsible for up to 70-90% of cancer-related mortality ^{1,2}. Shedding of tumour cells into the vasculature and lymphatic vessels occurs simultaneously in the vast majority of cancers ^{1,3,4}. Clonal origin analysis has demonstrated distinct lineage relationships between the lymphatic system and blood circulation ^{5,6}, indicating that tumour cell dissemination might play a distinct role in the metastatic process.

The metastasis of different tumour types demonstrates a biased distribution among distant organs. This phenomenon is called “organotropism” or “organ-specific metastasis” and is believed to be caused primarily by coordination between metastatic cancer cells (“seeds”) and their organ/tissue microenvironment (“soil”), i.e., the “seed and soil” hypothesis ^{7,8}. Tissue-specific gene signatures and signalling pathways in metastatic breast cancers have been identified by bulk RNA sequencing (RNA-seq) by comparing the primary tumour and distant lesions in different organs ⁹. However, whether these distinct signatures represent different cell populations within the primary tumour or the same type of cell educated to exhibit different gene expression profiles remains unknown.

A recently developed lineage tracing system has shed light on this question, as it allows the derivation of metastasis maps ¹⁰ and the phylogenetic development of colonized metastatic tumours in cancer xenografts ¹¹. Unfortunately, in immunodeficient mouse models, the effect of the microenvironment on tumour migration is neglected. Further application of the evolving barcode system has demonstrated that the bone microenvironment primes tumour cells for later progressive processes ¹². However, some key

questions remain unanswered. For example, do tumour cells demonstrate different affinities for lymphatic and haematogenous dissemination? Are there any intrinsic factors that dictate the route selection of cancer cells in the initial steps of metastasis? How does the crosstalk between cancer cells and the tumour microenvironment (TME) determine tissue-specific metastasis? Therefore, prospective studies providing a systematic demonstration of the tumour cell metastasis route are urgently needed.

In this study, we used the CellTag system to label individual tumour cells to achieve lineage tracing in order to show cell fate during metastasis, including the migration ability and route of each cell. Combining this analysis with single-cell RNA sequencing (scRNA-seq), we demonstrated that the lymphatic system is the main route of breast cancer dissemination and metastasis with the aid of prewired IL-2 expression, which serves as an intrinsic driver for tumour cell crosstalk with the TME. Our data also provide evidence that cancer cells derived from the same progenitor cell show different gene expression patterns and a different tumour cell-TME communication paradigm in different soils.

Results

Intravasation is the bottleneck for tumour metastasis through the blood circulation

We applied the CellTag system¹³ to the mammary carcinoma cell lines 4T1 and EMT6 to trace tumour cell metastasis in mouse cancer models. Because both cell lines can spontaneously produce highly metastatic tumours in immunocompetent mice, they are ideal models to show the tumour cell dissemination route and fate for studying the intrinsic drivers and their communication patterns with stromal cells in the tumour immune microenvironment during metastasis.

First, we transfected cells with a combinatory cell indexing barcode tag fused to a GFP marker and then sorted the cells by fluorescence-activated cell sorting (FACS) to select labelled tumour cells. To ensure that each cell contained a unique barcode, we used a low multiplicity of infection (MOI) of 0.01 for infection and sorted 10,000 GFP-positive cells, a much lower number than the number of different barcodes (19,973). Then, the GFP-positive cells were expanded in culture for 7 days prior to DNA sequencing (DNA-seq), which detected 7,665 and 7,809 unique barcodes in 4T1 and EMT6 cells, respectively. We then inoculated 2 million cells into the 4th mammary fat pad of BALB/c mice to model tumour metastasis (Fig. 1a; n=10 mice per cell line). Mice were monitored twice a week, and blood samples were collected at 10-day intervals. A complete autopsy was performed to assess metastatic tumours when the mice were moribund (30 days for 4T1-derived tumours and 47 days for EMT6-derived tumours) (Supplementary Fig. 1a).

All recipient mice exhibited severe metastasis in distant organs, especially the lung and liver (Fig. 1b). DNA-seq detected 4,741 unique barcodes for 4T1 cells and 5,227 unique barcodes for EMT6 cells, on average, from each primary tumour, indicating that approximately 62% (4,741/7,665) of the 4T1 cells and 67% (5,227/7,809) of the EMT6 cells could form tumours in BALB/c mice (Supplementary Table 1, Supplementary Table 2). Notably, the diversity of the barcodes identified in the blood and distant organs

was decreased dramatically to approximately 0.67-3.10% of that in the primary tumour (Fig. 1c). A similar pattern was observed in all recipient mice for both 4T1 and EMT6 cells (Supplementary Fig. 1b, c; Supplementary Table 1, Supplementary Table 2). Further analysis indicated that only approximately one hundred distinct 4T1 cells entered the blood circulation from the primary tumour continuously at different time points and that eventually, approximately comparable number of cell types colonized the lung or liver (Fig. 1d, e; Supplementary Table 1). Similar analysis was conducted in the EMT6 model, and the decrease in heterogeneity after tumour cell intravasation was also detected in this model (Supplementary Fig. 1d, Supplementary Table 2). These data suggest that the step of cancer cell intravasation from the primary tumour is a major bottleneck for tumour metastasis. To verify this hypothesis, we directly injected the same amount of tumour cells into a blood vessel through the tail vein. By DNA-seq, we found that the average number of barcodes in lung metastatic tumours was 1,839, approximately 15 times higher than that (121 CellTags) detected in the lungs of mice implanted with the same number of cancer cells in the mammary fat pad (Fig. 1f, Supplementary Fig. 1e, Supplementary Table 3).

To further distinguish the migratory fate of individual tumour cells, we compared the barcode composition among the blood, liver, and lung. In each individual mouse, more than half (59%) of the intravascular cancer cells formed lung and/or liver metastatic tumours (Supplementary Fig. 2a, b; Supplementary Table 4). Moreover, approximately 41% of the cells were detected only in the blood but not in the metastatic tumours, suggesting that they might lack extravasation ability for further colonization or fail to evade immunosurveillance while circulating (Supplementary Fig. 2a-c, Supplementary Table 4).

Notably, substantial numbers of lung and liver metastatic cells (25.6% in the 4T1 model, 22.9% in the EMT6 model) could not be detected in the blood, and this phenomenon was highly consistent among the 10 replicate mice for both cell lines (Fig. 1g; Supplementary Fig. 2c, d; Supplementary Table 4), suggesting that these cells might disseminate through other routes.

The lymphatic system is the main route of tumour dissemination

To identify the tumour cell migration route, we specifically checked the CellTag information of the lymphatic-disseminated tumour cells in the inguinal, axillary, and brachial lymph nodes (Fig. 2a, Supplementary Table 5). The results showed that the most cells (2,546 unique barcodes) were drained from the inguinal lymph nodes, followed by the axillary lymph nodes and brachial lymph nodes, which contained 696 and 598 unique barcodes, respectively (Fig. 2b). To further show the cell migration route, we first defined lung and liver metastatic cells based on the barcode information in the lung and liver on day 30. Then, the barcode composition was compared with that of lymph node-disseminated tumour cells collected at different time points. As the Venn diagrams show, all the tumour cells in the blood circulation that could not colonize distant organs (Blood-trap cells) were found in the lymphatic system, and more metastatic tumour cells were detected in lymph nodes than in the blood (Fig. 2c). Moreover, comparison of the lymphatic system-disseminated tumour cells at different time points indicated that the

lymph nodes contained the most tumour cells (2,583) on day 10 and that the number was decreased on day 20 (474) and day 30 (539) (Fig. 2d). This result may indicate that tumour cells have a stronger ability to intravasate into the lymphatic system than into the blood circulation, while during lymphatic circulation, the majority of tumour cells gradually die during the journey before colonizing remote tissues/organs to establish metastatic tumours.

Further global cell fate comparison defined based on 7 replicates also indicated that lymphatic-disseminated tumour cells exhibited the highest diversity, covering almost all blood circulating tumour cells (CTCs), and formed metastatic tumours in distinct organs (Fig. 2e). These data indicate that the lymphatic system is the route of tumour metastasis.

To facilitate subsequent analysis, we classified the tumour cells based on the barcode information and appearance frequency (in >3 samples or >5% of samples) into different types of metastatic cells (Fig. 2f), including cells that underwent either lymphatic or haematogenous dissemination to other organs (LyN-met, and Blood-met, respectively), cells that underwent either lymphatic or haematogenous dissemination but could not colonize distant organs (LyN-trap and Blood-trap, respectively), and some tagged cells detected only in distant organs (Other-met). To facilitate further comparative analysis, we also defined a group of cells (Harmony) that never appeared in either the blood or lymphatic circulation or in distant organs. In addition, we used lymphatic CTCs, blood CTCs, liver metastatic tumours, and lung metastatic tumours for further analysis (Fig. 2f).

Thus, by using the CellTag system, we successfully discovered the tumour cell metastatic route and consequently defined different types of metastatic tumour cells. Therefore, based on this lineage information, we can determine the fates of cells in primary tumours, i.e., which cells could migrate into the blood and/or lymphatic system, which cells could colonize the lung or liver via the blood circulation and which cells could disseminate into the lung or liver through the lymphatic system.

These results indicate that tumour cells may already be wired for a particular migration route before disseminating from the primary site, a finding that demonstrated heritability and reproducibility among the mouse models.

Communication interactions between tumour cells and stromal cells correlate with their metastasis route

To identify the intrinsic features that govern the metastatic fate of tumour cells, we conducted scRNA-seq on the primary tumours, and 6,066 cells passed quality control. Dimensionality reduction via uniform manifold approximation and projection (UMAP) clustering yielded 10 distinct clusters (Fig. 3a), including tumour cells, fibroblasts, macrophages, T cells, and other types of stromal cells, based on distinct marker gene expression patterns (Supplementary Fig. 3). Then, we focused specifically on tumour cells for dimensionality reduction to reveal the differences between different types of metastatic cells and nonmetastatic cells. As shown in the UMAP plot, the tumour cells were mixed together and did not

exhibit clustering patterns correlated with cell type, indicating that the gene expression patterns of different types of cells are highly similar (Fig. 3b).

To investigate the driving force responsible for the metastatic fate of different types of tumour cells, we conducted cell–cell interaction analysis by using the CellCall program¹⁴. The results indicated that tumour cells communicated closely with a variety of stromal cells, especially macrophages, fibroblasts, endothelial cells, and dendritic cells (Fig. 3c). Different types of cells interacted differently with immune cells (Supplementary Fig. 4a, b); for example, fibroblasts communicated comprehensively with tumour cells, indicating the important role of fibroblasts in the TME. Endothelial cells, macrophages, and monocytes detected signals from tumour cells (Fig. 3d, Supplementary Fig. 4b, c). Analysis of signal transduction from tumour cells to stromal cells showed that different types of cells secreted different cytokines or chemokines into the microenvironment, which might help to direct their migration route (Fig. 3d, e). Specifically, most of the tumour cells, except the Blood-trap tumour cells, demonstrated a high expression level of Tgfb2 to interact with endothelial cells and macrophages. In contrast, the Blood-trap tumour cells communicated strongly with macrophages and monocytes. The LyN-trap and Other-met tumour cells showed very similar interaction patterns, suggesting that Other-met tumour cells also disseminated through the lymphatic system but could not be detected, possibly because they were moderately difficult to detect. Notably, all cells identified in the blood were also detected in lymph nodes (Fig. 2c, f); therefore, the Blood-trap and Blood-met groups might not be representative of intravasation even though they demonstrated distinct cell–cell interaction patterns (Fig. 3d, Supplementary Fig. 4b, c).

Through comparative analysis of the cell–cell interactions between the abovementioned groups as well as the nonmetastasized cells, we identified several specific features that correlated with lymphatic and haematogenous dissemination and with lung or liver metastasis. Notably, the observation that the Harmony cells lacked Il2-Il2rg–mediated communication with stromal cells reliably demonstrated the key role of Il2-Il2rg signalling in tumour metastasis (Fig. 3d, e). This finding is consistent with clinical data indicating that metastatic tumours express high levels of IL2 (Supplementary Fig. 5a) and predispose patients to worse outcomes (Supplementary Fig. 5b).

Il2 facilitates tumour cell lymphatic dissemination

To further validate the driving role of Il2 during metastasis, we knocked down Il2 in 4T1 cells and then inoculated these cells into the fat pads of BALB/c mice (Fig. 4a, b; Supplementary Table 6). Then, we analysed the CellTag information in the mouse model to study tumour cell metastasis after Il2 knockdown.

Consistent with our earlier observation, tumour cells demonstrated high diversity in lymph nodes and covered the majority of the lung or liver metastasis tumour cells. However, after Il2 knockdown (Fig. 4b), the diversity of lymphatic-disseminated tumour cells was dramatically decreased (Fig. 4c), while in the

shRNA3-mediated knockdown group, the tumour cells in distant organs showed less diversity (Fig. 4d, e). Comparative analysis indicated that the CellTag composition in metastatic tumours also exhibited reduced overlap with that in lymph nodes (Fig. 4f). These data suggest that IL2 is critical for tumour cell lymphatic migration and could constitute a therapeutic target in cancers.

Tumour cells demonstrate distinct properties in congenial soil

It is accepted that metastases develop only when the seed and soil are compatible. Samples from the same organ/tissue demonstrated comparable CellTag diversity even in different mice (Supplementary Fig. 6a, b), indicating that tumour cells exhibited greater similarity in more similar sample types and that highly conserved features existed in the tumour cells to modulate the migration fate, which could be identified by the dissimilar CellTags.

To further demonstrate the features of cancer cells during metastasis, we also conducted scRNA-seq on lung and liver metastatic tumours to provide greater insights into the features of organ-specific tumour metastasis. A total of 23,843 cells in the 4T1 model passed quality control, of which 6,066, 8,582, and 9,195 cells were derived from the primary tumours, lung metastatic tumours, and liver metastatic tumours, respectively. Dimensionality reduction via UMAP clustering analysis yielded 10 distinct clusters (Fig. 5a), including tumour cells, fibroblasts, macrophages, T cells, and other types of stromal cells, based on distinct specific marker gene expression patterns (Supplementary Fig. 6c, d). Comparison of the TME composition demonstrated similarity across different stages during metastasis, except for the higher proportion of granulocytes in metastatic tumours (Supplementary Fig. 7a).

To assess the dynamic changes in tumour cells during metastasis, we conducted gene set enrichment analysis (GSEA) specifically on cancer cells to show the differences among primary tumours and metastatic tumours. First, we included all tumour cells, similar to the traditional bulk approach, for comparative analysis (Supplementary Fig. 7b). After tumour cells migrated into the lung, the cytokine response, including the TGF β signalling pathway, was upregulated, whereas the Toll-like receptor-related pathway and metabolism-regulated growth-related pathway were downregulated (Supplementary Fig. 7c). Tumour cells demonstrated similar gene expression patterns, except for the upregulation of Myc related pathways, after they colonized the liver (Supplementary Fig. 7d).

As demonstrated in our previous results, only a small portion of tumour cells could eventually colonize distant organs. To demonstrate the faithful properties of seeds in different soils, we focused specifically on cells that carried the same tag in primary tumours, lung metastatic tumours, and liver metastatic tumours (Fig. 5b). Based on the scRNA-seq data, eight types of cells fit the criteria and were therefore used for further analysis (Fig. 5c). As predicted, we identified a group of genes that were specifically differentially regulated after cell colonization of the lung and/or liver (Fig. 5d). Compared with the traditional bulk approach, the CellTag-based analysis identified different marker genes during tumour cell migration (Fig. 5e, Supplementary Fig. 7e, f). GSEA indicated that in addition to the terms identified by the

bulk sequencing approach, the CellTag-based analysis approach also allowed us to identify more specific functional terms related to tumour metastasis, i.e., reactive oxygen species detoxification, hypoxia, DNA methylation, and mitochondria terms (Fig. 5f). These data indicate that identification of the cell migration fate provides detailed information on tumour metastasis.

The seed-soil communication paradigm differs among primary tumours and metastatic tumours

To further investigate cell–cell interactions and communication in the TME, we used CellCall to estimate the putative intercellular networks based on not only ligand and receptor expression but also receptor downstream transcription factors (TFs). These analyses revealed comprehensive crosstalk among tumour cells and stromal cells (Supplementary Fig. 8a-c).

Furthermore, we focused specifically on the communication between tumour cells and other stromal cells (Fig. 3c). Tumour cells showed varying levels of responses to signals from immune cells and other types of stromal cells. Compared with the traditional bulk approach (Supplementary Fig. 8d), cell identification by the CellTag approach provided much information on cell–cell crosstalk (Supplementary Fig. 8e). Specifically, tumour cells at the primary site demonstrated much stronger interactions with endothelial cells and fibroblasts, for example, Vegfc-Kdr/Flt4/Flt1 and Pdgfc-Pdgfrb/Pdgfra/Kdr/Flt4/Flt1 interactions (Fig. 6c). Functional analysis indicated that some of these genes are mainly growth factors that consistently exhibit high expression levels in primary tumours. In addition, groups of ligand–receptor-transcription factor interaction pairs were commonly upregulated in all tumour metastasis steps, including Tslp-Il17r/Crlf2, Il11-Il16st/Il16ra, Tgfb1/3-Tfgbe1/2, Csf3-Csf3r, and Csf1-Csf1r. These interactions were functionally enriched in cytokine-, chemokine- and interferon-gamma response-related genes, as well as the Tgfb signalling pathway.

In addition, tumour cells exhibited distinct responses to stromal signals in different soils through differentially expressed receptors, which were also reflected by the levels of their downstream transcription factors (Fig. 6d, e). For example, tumour cells highly expressed Bmpr2 after migrating into the lungs, with subsequent activation of its downstream targets, including Id1, Id2, Id4, and Smad4.

These results are consistent with those of GSEA, indicating that tumour cells have different gene expression patterns in different soils and that the cancer cell-TME communication paradigm varies significantly between primary tumours and metastatic tumours.

Discussion

Tumour cells metastasize mainly through the blood circulation and the lymphatic system. Numerous studies have been carried out to distinguish the role of both systems in the various stages of cancer metastasis^{1,5,6,9,15}, yet which one is the “preferred” route for distant metastasis remains unknown. Phylogenetic analysis of colorectal and breast cancers has shown that in the majority of cases,

lymphatic and distant metastases arose from independent subclones in the primary tumour;; moreover, the two sites are subject to different levels of selection, indicating that the origin cells of lymphatic and distant metastases are fundamentally different ^{5,6,15}.

Cancer cells enter the bloodstream through intravasation to become CTCs ¹, which have been proven to be highly correlated with the survival outcomes of breast cancer patients ¹⁶. CTCs tend to recruit platelets and different types of leukocytes to form a cluster for travel through the circulation and further evade immunosurveillance ^{9,17,18}. Due to the convenience the blood tests, CTCs are considered a component of liquid biopsy for early tumour metastasis prediction, detection, and therapeutic indication ¹⁹. On the other hand, the lymphatic system is also reported to be critical for metastasis. Lymphatic system-experienced melanoma cells were more resistant to ferroptosis and formed more metastases ²⁰. In primary tumour resection models, cancer cells can further disseminate from lymph nodes into blood vessels and form lung metastatic tumours ²¹. Although metastatic spread is feasible via both routes, the relative importance of each process for a given cancer type and target organ is unknown owing to the lack of appropriate quantitative tools. Leveraging the CellTag system, we demonstrated that the majority of lung and liver metastatic cells can be detected in the lymphatic system instead of as blood-derived CTCs, indicating that lymphatic vessels could be the preferred route for tumour cell metastasis, at least in breast cancer.

Tumour cells demonstrate high intra- and intertumour heterogeneity ²² and exhibit multiclonal evolution during tumorigenesis, drug treatment, and metastasis ^{22–24}. Via the cell labelling system, we were able to perform lineage tracing to define individual cell fates. In our metastasis models, different tagged cells, even in established cell lines, consistently demonstrated migration route bias and organotropic metastasis, indicating that intrinsic mechanisms were prewired in each tumour cell before it disseminated from the primary site, controlling its heritable and reproducible migration fate in individual mice. Owing to its tremendous complexity and diversity, the immune system performs multifaceted functions in the TME, with different immune cells playing distinct or sometimes even opposing roles. Emerging evidence indicates that intrinsic tumour genetic aberrations shape the tumour milieu. Intratumour heterogeneity predisposes tumours to an elusive tumour-TME crosstalk pattern, making exploitation of the intrinsic properties of cancer cells impossible without high-throughput approaches for cell fate classification. Further scRNA seq and cell–cell interaction analyses allowed us to identify the unique features of metastatic tumour cells in primary tumours, for example, Il2-Il2rg and Flt3l-Flt3, which mediate interactions between dendritic cells and endothelial cells.

Tumour cells colonizing distinct organs/tissues to form metastatic tumours usually exhibit tremendously different gene expression patterns ^{7,9}. Because only a small portion of shed tumour cells can form metastatic tumours, RNA-seq (at either the bulk or single-cell level) cannot determine whether organotropism is generated because different seeds preferentially migrate to different organs or because the same seeds are educated for differential gene expression in different soils. Therefore, this lineage tracing approach coupled with scRNA-seq analysis provided a powerful means to distinguish between

these alternatives. In this regard, our data precisely elucidated that tumour cells derived from the same progenitor cell exhibit different gene expression patterns in different soils and that the cancer cell-TME communication paradigm varies significantly between primary tumours and metastatic tumours.

In summary, leveraging the CellTag system, we demonstrated the route of breast cancer metastasis and identified the potent intrinsic drivers of lymphatic dissemination (Supplementary Fig. 9). To strengthen the clinical utility of tumour metastasis prediction and therapy, prospective studies should exploit effective cellular markers for cellular and molecular properties and identify targets for pharmacological treatment to block or eliminate cancer cell dissemination.

Methods

Cell indexing with the CellTag system

4T1 cells (American Type Culture Collection (ATCC), CRL-2539) and EMT6 cells (ATCC, CRL-2755) were cultured in RPMI 1640 medium supplemented with 10% FBS, glutamine (Thermo), and penicillin/streptomycin (pen/strep). All cells were routinely tested to exclude mycoplasma contamination.

Lentivirus was produced by transfecting HEK293FT cells with the CellTag library (pSMAL-CellTag-V1, Addgene, 115643) along with the packaging plasmids pCMV-dR8.2 dvpr (Addgene, 8455) and pCMV-VSV-G (Addgene, 8454). Viral particles were collected 60 h after transfection and filtered through a 0.45- μ m filter, and they were then applied to cells immediately and incubated overnight. Forty-eight hours after infection, GFP-positive cells were sorted by FACS using a BD FACSAria III (BD Biosciences). To ensure that each cell contained a unique barcode, we used a low MOI of 0.01 for infection and sorted 10,000 GFP-positive cells, a much lower number than the number of different barcodes (19,973). To maintain diversity, cells were expanded after sorting with a minimal time in culture (<7 days). For lineage tracing experiments, cells were expanded only after thawing for 2-4 days prior to orthotopic injection.

Orthotopic metastasis model

All mouse experiments were performed under the ethical guidelines of the University of Macau (animal protocol number: UMAEC-015-2019). All mice were maintained in a specific pathogen-free (SPF) facility at 23-25 °C on a 12-h light/dark cycle. Animals were monitored twice a week.

Eight-week-old female BALB/c and athymic nude mice were used for orthotopic injection. CellTag-labelled cells (4T1 or EMT6) were surgically implanted into the 4th mammary fat pads of anaesthetized (Avertin) mice. Skin flaps were sutured by using clips. For tail vein injection, 2 million tumour cells were suspended in 100 μ l of saline solution. Animals were first kept under red light to warm the tail veins and were then transferred to a restraint device prior to tail vein injection.

Sample collection and preparation

Mice were euthanized on day 30 (4T1 tumour model) or day 47 (EMT6 tumour model) post tumour cell implantation for collection of primary tumours, lungs, livers, lymph nodes, and blood. In addition, blood samples were collected at 10-day intervals; i.e., day 10, day 20, and day 30.

For blood samples, EDTA Vacutainer tubes were used to prevent coagulation, and the cells were then lysed with red blood cell (RBC) lysis buffer (Thermo Fisher, 00-4333-57). The remaining cells were pelleted by centrifugation at $500 \times g$ for 5 min for further use.

Primary tumours, lung metastatic tumours, and liver metastatic tumours were excised from the surrounding tissue, removing as much normal surrounding tissue as possible, and were then dissociated for preparation of single-cell suspensions as previously described²⁵. In brief, tissue was minced and digested in DMEM/F-12 containing 300 units/ml collagenase III (Worthington, S4M7602S), 100 units/ml hyaluronidase (Sigma, H3506), 5% FBS (Thermo Fisher), 5 µg/ml insulin (Sigma, I1882), 10 ng/ml EGF (Thermo Fisher, PHG0311L), and 500 ng/ml hydrocortisone (H0888; Sigma) for approximately 1 h at 37 °C in 5% CO₂. The resultant suspension was sequentially resuspended in DMEM/F-12 supplemented with 5 mg/ml dispase II (Roche Diagnostics, 825-001) and 0.1 mg/ml DNase I (Worthington, 58C10349) for 5 min at 37 °C and treated with RBC lysis buffer (Thermo Fisher, 00-4333-57) for 3 min to remove RBCs before filtration through a 40 µm cell strainer (Jet Bio-Filtration, CSS013040) to obtain single-cell suspensions.

Bulk DNA barcode sequencing

Primary tumours, lungs, livers, lymph nodes, and blood containing tumour cells were used for DNA extraction. Two micrograms of total DNA was used for CellTag barcode amplification. Barcodes were amplified from genomic DNA via a nested PCR approach with predesigned dual index primers (Supplementary Table 7) containing sequencing adaptors, sample indexes, and flow cell adaptors. AMPure beads (Beckman Coulter, A63880) were used to purify the PCR products.

scRNA-seq

scRNA-seq libraries were prepared using 10x Chromium Next GEM Single Cell 3' Reagent Kits (v3.1) according to the user guide instructions. Single-cell transcriptome libraries were sequenced on the Illumina HiSeq X-Ten system.

scRNA-seq data preprocessing

We first generated the expanded reference genome by adding the GFP sequence with random CellTag sequences to the mm10 mouse genome FASTA file. In the GTF files, we also added the 'GFP' gene as the expanded annotation file.

Single-cell transcriptome data were mapped to the expanded mm10 mouse reference genome, and unique molecular identifiers (UMIs) were counted by using the 10X Genomics Cell Ranger (v4.0.0) toolkit. All samples were then processed for quality control by the following criteria: 1, cells with high mitochondrial gene expression levels were filtered out by evaluating the median-centred median absolute deviation (MAD) and variance for each sample; 2, cells in which the number of expressed genes was less than 200 or more than 10,000 were removed; and 3, the potential doublets were identified and removed by DoubletFinder (v2.0.3) with default parameters.

Cluster and cell type identification and visualization

The preprocessed data were then imported into Seurat (v4) to perform UMI count matrix normalization and find variable features with the `NormalizeData` and `FindVariableFeatures` functions, respectively. Here, we analysed only genes expressed in more than 10 cells. Then, data for all samples (primary tumour, lung metastatic tumour, and liver metastatic tumour) were integrated with the `IntegrateData` function. To prevent the highly expressed genes from dominating the downstream analysis, we applied the `ScaleData` function to weight each gene equally. Then, we performed linear dimensionality reduction and graph-based clustering to group cells. To annotate and identify different cell types, we mainly applied two complementary strategies. First, we applied SingleR (v1.4.0) to annotate cell types for each cluster. Second, we used `FindMarkers` to identify marker genes for each cluster and then performed a database search to confirm whether the top-ranked genes were typical cell marker genes. Finally, we defined cell types and plotted the expression levels of well-known cell type marker genes.

Tumour cells were also confirmed by the expression of GFP with the criterion of a GFP depth greater than 5.

CellTag extraction and quantification for single cells

CellTag extraction and quantification were conducted by calling cell tag motifs from the BAM files containing the mapped reads. First, we used the binary viewing tool SAMtools (1.9) to view those BAM files. We then used the 'grep' command to search CellTag-containing reads and output the search results. CellTag information for each cell was then extracted based on the flanking sequence. Finally, the CellTag depth for each cell was calculated, and the CellTag-cell barcode matrix was generated.

Differential expression analysis

To identify the differentially expressed genes (DEGs) between primary tumour and metastatic tumour cells, we applied the negative binomial distribution method DESeq2 (v1.30.0), which estimates the variance and mean dependence in count data for differential expression analysis. Here, we performed comparisons with bulk cells and with the same tagged cells. For analysis of bulk cells, we compared the set of all tagged primary cells with the set of all tagged metastatic cells. For analysis of the same tagged cells, we deemed that the cells were the same cells in different microenvironments (before and after metastasis). For both approaches, we aimed to determine whether our tagged cell technique provides more detailed information. Here, the valid differentially expressed genes were identified as those with $\text{Padj} < 0.05$ and $\log\text{FC} > 1$.

Cell–cell communication analysis

Cell–cell communication interactions were identified based on scRNA-seq data by using CellCall (0.0.0.9000), an R package that leverages not only ligand–receptor expression levels but also the levels of the downstream transcription factors of these target genes. To address whether the cell communication interactions are identical in different TMEs (primary tumour, lung metastatic tumour, and liver metastatic tumour), we performed separate cell–cell communication analyses for the same tagged cells in different tumours. For analysis of metastatic features, we split the tumour cells in the primary tumour into different subsets based on our subgroup definition.

Data availability

The scRNA-seq data have been deposited in the Sequence Read Archive (SRA) database under accession code PRJNA751409. All other data supporting the findings of this study are included in the supplementary files. The reporting summary for this article is available as a Supplementary Information file.

Statistics and reproducibility

Microsoft Excel, GraphPad Prism (version 6) and R (version 4.0.4) software were used for statistical calculations. Specific statistical tests, sample numbers and other information are indicated in the main text or figure legends. A two-tailed *t* test analysis approach was used to determine the significance of differences between the different sets of data unless otherwise specified. Differences were considered

statistically significant when $p < 0.05$. The cutoff value for differential gene expression and pathway enrichment was $FDR < 0.05$.

All experiments were conducted at least three times independently, and similar results were used for further analysis to guarantee reproducibility.

Declarations

Acknowledgements

We thank the Animal Research Core for providing the animal housing and the Information and Communication Technology Office (ICTO) for providing the high-performance computer (HPC) for data processing. This work was supported by multiyear research grant (MYRG) 2018-00186-FHS to K. M. by the University of Macau, Macau SAR, China; Macao Science and Technology Development Fund (FDCT) grants 0011/2019/AKP, 0034/2019/AGJ, and 0048/2019/A1 to C. D., 0101/2018/A3 to X. X., and 0073/2021/A2 and 111/2017/A to K. M.; and National Natural Science Foundation of China (NSFC) grant 82030094 to C. D.

Author contributions

K. M. and C. D. designed this research; K. M., L. W., F. S., X. Z., S. Guo, and Y. F. performed experiments; A. Z. performed the bioinformatics analyses; H. S. and J. X. performed cell sorting; J. Z. assisted with the data analysis; S. Gao, X. Y. Z., H. S., and Z. W. supervised the data analysis; C. D. and X. X. supervised the experiments; and K. M. and C. D. wrote the manuscript.

Competing interests:

The authors declare no competing interests.

References

1. Chaffer, C.L. & Weinberg, R.A. A Perspective on Cancer Cell Metastasis. *Science* **331**, 1559–1564 (2011).
2. Dillekas, H., Rogers, M.S. & Straume, O. Are 90% of deaths from cancer caused by metastases? *Cancer Med* **8**, 5574–5576 (2019).
3. Nguyen, D.X., Bos, P.D. & Massague, J. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer* **9**, 274–84 (2009).
4. Lambert, A.W., Pattabiraman, D.R. & Weinberg, R.A. Emerging Biological Principles of Metastasis. *Cell* **168**, 670–691 (2017).

5. Naxerova, K. *et al.* Origins of lymphatic and distant metastases in human colorectal cancer. *Science* **357**, 55–60 (2017).
6. Reiter, J.G. *et al.* Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases. *Nature Genetics* **52**, 692+ (2020).
7. Gao, Y. *et al.* Metastasis Organotropism: Redefining the Congenial Soil. *Dev Cell* **49**, 375–391 (2019).
8. Peinado, H. *et al.* Pre-metastatic niches: organ-specific homes for metastases. *Nat Rev Cancer* **17**, 302–317 (2017).
9. Chen, W.J., Hoffmann, A.D., Liu, H.P. & Liu, X. Organotropism: new insights into molecular mechanisms of breast cancer metastasis. *Npj Precision Oncology* **2**(2018).
10. Jin, X. *et al.* A metastasis map of human cancer cell lines. *Nature* **588**, 331–336 (2020).
11. Quinn, J.J. *et al.* Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts. *Science* **371**(2021).
12. Zhang, W.J. *et al.* The bone microenvironment invigorates metastatic seeds for further dissemination. *Cell* **184**, 2471+ (2021).
13. Biddy, B.A. *et al.* Single-cell mapping of lineage and identity in direct reprogramming. *Nature* **564**, 219+ (2018).
14. Zhang, Y. *et al.* CellCall: integrating paired ligand-receptor and transcription factor activities for cell-cell communication. *Nucleic Acids Res* (2021).
15. Venet, D. *et al.* Phylogenetic reconstruction of breast cancer reveals two routes of metastatic dissemination associated with distinct clinical outcome. *EBioMedicine* **56**, 102793 (2020).
16. Rack, B. *et al.* Circulating tumor cells predict survival in early average-to-high risk breast cancer patients. *J Natl Cancer Inst* **106**(2014).
17. Szczerba, B.M. *et al.* Neutrophils escort circulating tumour cells to enable cell cycle progression. *Nature* **566**, 553–557 (2019).
18. Liu, X. *et al.* Homophilic CD44 Interactions Mediate Tumor Cell Aggregation and Polyclonal Metastasis in Patient-Derived Breast Cancer Models. *Cancer Discov* **9**, 96–113 (2019).
19. Sinkala, E. *et al.* Profiling protein expression in circulating tumour cells using microfluidic western blotting. *Nature Communications* **8**(2017).
20. Ubellacker, J.M. *et al.* Lymph protects metastasizing melanoma cells from ferroptosis. *Nature* **585**, 113–118 (2020).
21. Pereira, E.R. *et al.* Lymph node metastases can invade local blood vessels, exit the node, and colonize distant organs in mice. *Science* **359**, 1403–1407 (2018).
22. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* **168**, 613–628 (2017).
23. Dagogo-Jack, I. & Shaw, A.T. Tumour heterogeneity and resistance to cancer therapies. *Nat Rev Clin Oncol* **15**, 81–94 (2018).

24. Lawson, D.A., Kessenbrock, K., Davis, R.T., Pervolarakis, N. & Werb, Z. Tumour heterogeneity and metastasis at single-cell resolution. *Nature Cell Biology* **20**, 1349–1360 (2018).

25. Sun, H. *et al.* Single-cell RNA-Seq reveals cell heterogeneity and hierarchy within mouse mammary epithelia. *Journal of Biological Chemistry* **293**, 8315–8329 (2018).

Figures

Figure 1

Figure 1

Lineage tracing of tumour cell metastatic fate. **a**, Schematic illustration of the strategy for tracing tumour cell migration fate using single-cell barcoding integrated with scRNA-seq (n=10). **b**, Representative pictures of lung and liver metastatic tumours. **c**, Connected bar chart showing the 4T1 tumour cell dissemination pattern in mouse #540. The different colours represent different tagged cells, and the grey colour in the primary tumours indicates cells detected exclusively in the primary tumours. The size of the different coloured bars indicates the proportion of called clones labelled with different CellTag barcodes. **d**, Venn diagram showing the CellTag compositions among blood CTCs on different days in mouse #540. **e**, Summary of CellTag barcode types counted in different organs/tissues in the 4T1 metastasis model (n=10 mice). **f**, CellTag barcode types identified in lung and liver metastatic tumours were analysed in the tail vein injection model and compared with those in the mammary fat pad inoculation model. **g**, Oncoplot showing the CellTag composition in blood at different time points and in distant organs in the 4T1 cell metastasis model. CellTags were defined based on their presence in at least 5% of tumours or more than 3 samples.

Figure 2

Figure 2

Tumour cells disseminate preferentially through the lymphatic system. **a**, Workflow of the sample collection strategy to demonstrate the tumour cell lymphatic and haematogenous migration routes. **b**, CellTag comparison between sentinel lymph nodes (inguinal lymph nodes), brachial lymph nodes, and axillary lymph nodes on day 10 (n=7 mice). **c**, The CellTag barcode types were compared in different organs/tissues in the 4T1 metastasis model on day 30 (n=7 mice). **d**, Comparison of CellTag barcodes in lymph nodes at different time points (n=7 mice). **e**, Heatmap of different tagged cells showing their metastatic fates. Based on their migration route, cells could be classified into different groups, which are

indicated by the cell types. **f**, Oncoplot presentation of the CellTag composition in the blood, lymph nodes, and distant organs in the 4T1 cell metastasis model. CellTags were defined based on their presence in at least 5% of tumours or more than 3 samples.

Figure 3

Figure 3

scRNA-seq identifies the intrinsic features correlated with distinct metastatic fates. **a**, Dimensionality reduction by UMAP showed the cell type classification in primary tumours. **b**, Tumour cell-specific UMAP analysis indicated that the different types of cells, which exhibited distinct metastatic fates, had similar gene expression patterns. **c**, Global pattern of interactions between tumour cells and other types of

stromal cells in primary tumours. The different colours represent the types of cells, and the size of the coloured bar indicates the abundance of the interaction. **d**, Ligand–receptor pairs in tumour–stromal interactions; only the differential pairs among a variety of tumour cells were used in the plot. The colour and size of the circles indicate the significance and the enrichment score, respectively. **e**, Expression levels of potential candidates in different types of tumour cells.

Figure 4

Figure 4

Functional validation of IL2 in regulating tumour cell lymphatic dissemination. **a**, Workflow for validation of the experimental design. CellTag-labelled cells with shRNA-mediated IL2 knockdown or treatment with the control shRNA were inoculated into the mammary fat pad, and the metastasis pattern was monitored (n=10 mice per shRNA type). **b**, Western blot analysis of IL2 showing the knockdown efficiency. **c-e**, CellTag barcode types were counted in lymph nodes (c), lungs (d), and livers (e) in the 4T1 metastasis model on day 30 (n=7 mice). **f**, Analysis of the CellTag composition in blood, lymph nodes, lungs, and livers between the IL2 knockdown group and the control group. %: percentage of CellTags overlapping with lung or liver metastatic tumour cells.

Figure 5

Tumour cell gene expression profile during metastasis. **a**, UMAP analysis of the lung and liver metastatic tumours as well as the primary tumours with classification into 10 distinct cell types. **b**, Schematic showing the strategy for comparative analysis of gene expression. Cells with the same tag in both primary tumours and metastatic tumours were used for the following analysis. **c**, Venn diagram showing the CellTags detected by the scRNA-seq of primary, lung metastatic, and liver metastatic tumours. **d**, Heatmap showing that the same cells labelled with identical tags were educated for differential regulation of their gene expression pattern in distinct organs during metastasis. **e**, Differentially expressed genes identified by the traditional bulk RNA analysis approach and the identical tag approach were compared. **f**, GSEA of the differentially expressed genes in identical tagged cells during metastasis from the primary tumour to the lung and from the primary tumour to the liver.

Figure 6

The tumour-stromal cell communication paradigm differs between primary tumours and metastatic tumours. **a-b**, Global pattern of interactions between tumour cells and other types of stromal cells in lung (a) and liver metastatic tumours (b). The different colours represent different types of cells, and the size of the coloured bar indicates the abundance of the interaction. **c**, The ligand–receptor pairs in the tumour–stromal interactions; only the differential pairs among a variety of tumour cells were used in the plot. The colour and size of the circles indicate the significance and the enrichment score, respectively. The expression levels of ligands in primary, lung metastatic, and liver metastatic tumours are shown in violin plots in the figure. **d**, Ligand–receptor pairs in the tumour-stromal interactions; only the differential pairs among a variety of tumour cells were used in the plot. The colour and size of the circles indicate the significance and the enrichment score, respectively. **e**, Expression levels of receptor genes as well as the downstream target genes in different tumours.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigS1.pdf](#)
- [FigS2.pdf](#)
- [FigS3.pdf](#)
- [FigS4.pdf](#)
- [FigS5.pdf](#)
- [FigS6.pdf](#)
- [FigS7.pdf](#)
- [FigS8.pdf](#)
- [FigS9.pdf](#)
- [SupplementaryTable1Celltaginformationin4T1tumorbearingmice.xlsx](#)
- [SupplementaryTable2CelltaginformationinEMT6tumorbearingmice.xlsx](#)
- [SupplementaryTable3Celltaginformationintailveininjectionmicemodel.xlsx](#)
- [SupplementaryTable4Candidatecelltagfortumorcelldissemination.xlsx](#)
- [SupplementaryTable5LymphnodeCelltaginformationin4T1tumorbearingmice.xlsx](#)
- [SupplementaryTable6CelltaginformationinIl2knockdown4T1tumorbearingmice.xlsx](#)
- [SupplementaryTable7primersforBulkDNAbarcodesequencing.xlsx](#)
- [nreditorialpolicychecklist.pdf](#)
- [nrreportingsummary.pdf](#)