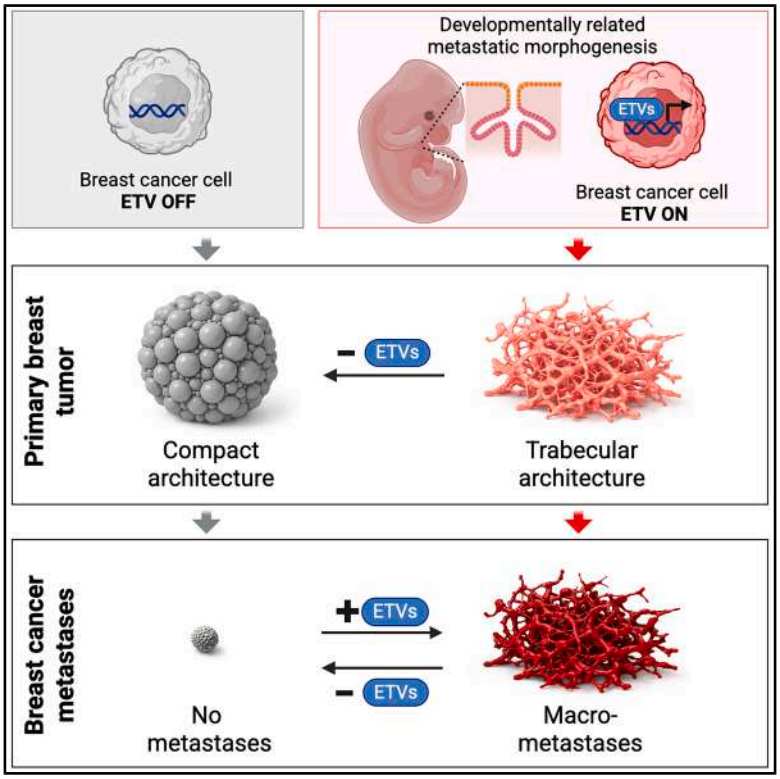
Cell

A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer

Graphical abstract



Highlights

- Imaging shows macrometastases built as a 3D lattice of epithelial cords
- Metastases adopt a developmental 3D branching morphogenesis blueprint (MTM)
- MTM^{HIGH} tumors pre-exist in primaries that metastasize, and MTM^{LOW} stays localized
- ETV1/4/5 act as master TFs for MTM required for metastasis, not for the primary tumor

Authors

Robin Caire, Roberta Bordo, Francesca Zanconato, ..., Massimiliano Pagani, Michelangelo Cordenonsi, Stefano Piccolo

Correspondence

massimiliano.pagani@ifom.eu (M.P.), michelangelo.cordenonsi@unipd.it (M.C.), stefano.piccolo@unipd.it (S.P.)

In brief

Breast cancer macrometastases expand by activating a metastatic trabecular morphogenesis (MTM) program that redeploys developmental branching to build a fractal 3D lattice of epithelial cords, a state already and selectively present in primary tumors destined to metastasize.

Caire et al., 2026, Cell *189*, 1–18
June 11, 2026
https://doi.org/10.1016/j.cell.2026.03.009

CellPress

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

Cell

CellPress
OPEN ACCESS

Article

A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer

Robin Caire,[1] Roberta Bordo,[2,10] Francesca Zanconato,[1] Tito Panciera,[1] Estelle Audoux,[1] Paolo Contessotto,[1] Michaela Fakiola,[2,10] Ramona Bason,[2,10] Oriana Romano,[1] Ambela Suli,[1] Giusy Battilana,[1] Matteo Marchionni,[1] Mattia Forcato,[1] Sara Donzelli,[3] Maria Vittoria Dieci,[4] Gaia Griguolo,[4] Mariantonia Carosi,[3] Matteo Fassan,[5] Vincenza Guzzardo,[5] Angelo Paolo Dei Tos,[5] Silvia Marsoni,[2] Pei-Hsun Wu,[7] Denis Wirtz,[7] Shanshan He,[6] Cecilia Casali,[8] Francesco Volpin,[9] Giovanni Blandino,[3] Claudio Tripodo,[10] Silvio Bicciato,[1] Valentina Guarneri,[4] Massimiliano Pagani,[2,10,11,12,*] Michelangelo Cordenonsi,[1,11,12,*] and Stefano Piccolo[1,10,11,12,13,*]

[1]Department of Molecular Medicine DMM, University of Padua, Padua, Italy
[2]University of Milan, Milan, Italy
[3]IRCCS Regina Elena National Cancer Institute, Rome, Italy
[4]Istituto Oncologico Veneto IOV IRCCS Padua - Department of Surgery, Oncology and Gastroenterology DISCOG, University of Padua, Padua, Italy
[5]Department of Medicine DIMED, University of Padua, Padua, Italy
[6]Bruker Spatial Biology, Inc., Seattle, WA, USA
[7]Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA
[8]Fondazione IRCCS Istituto Neurologico "Carlo Besta," Neurosurgery Department, Milan, Italy
[9]Neurosurgery Unit, Azienda Ospedale, Università di Padova, 35128 Padua, Italy
[10]IFOM ETS - the AIRC Institute of Molecular Oncology, Milan, Italy
[11]Senior authors
[12]These authors contributed equally
[13]Lead contact
*Correspondence: massimiliano.pagani@ifom.eu (M.P.), michelangelo.cordenonsi@unipd.it (M.C.), stefano.piccolo@unipd.it (S.P.)
https://doi.org/10.1016/j.cell.2026.03.009

SUMMARY

The tissue-level processes underpinning metastatic outgrowth remain unclear. We combined single-cell RNA sequencing, spatial transcriptomics, and AI-supported 3D imaging in human breast cancer with functional investigations in mice to uncover a 3D morphogenetic process essential for macrometastatic expansion. Macrometastases pervasively activate a metastatic trabecular morphogenesis (MTM) gene-expression program that redeploys developmental branching morphogenesis to build macrometastases as a 3D trabecular lattice of epithelial cords. MTM^{HIGH} cells pre-exist in primary tumors destined to metastasize, whereas MTM^{LOW} primaries are non-metastatic and display a compact, expansile growth architecture. Chromatin immunoprecipitation sequencing (ChIP-seq) on metastatic organoids identifies ETV1/4/5 as master regulators of MTM and branching cancer morphogenesis, required for metastatic outgrowth but dispensable for primary tumor take, bulk growth, and initial metastatic dissemination. Spatial and functional analyses reveal stromal fibroblast growth factor (FGF)→fibroblast growth factor receptor (FGFR) signaling as an actionable MTM dependency. Thus, we link metastatic outgrowth to a 3D developmental morphogenetic process, exposing therapeutic vulnerabilities specific to the lethal macrometastatic stage.

INTRODUCTION

The outgrowth of distant metastases signals, with few exceptions, the reach of an incurable stage of cancer progression.[1] Resolving the nature of the aggressive cell state(s) that populate macrometastases—their molecular programs, dependencies, and emergence in primary tumors—remains a fundamental quest in cancer biology. Prior work showed that metastases, unlike primary tumor cells, are largely driven by non-genetic adaptations.[2] Although the specific nature of this epigenetic plasticity remains unaddressed, this adaptability enables metastatic cells to grow organ-like structures at distant sites, in spite of the various constraints imposed by foreign tissues.[1] But is there a "metastatic *Bauplan*"—an underlying 3D morphogenetic program and architectural blueprint—guiding metastatic outgrowth? Deciphering these programs could uncover regulatory mechanisms operating at the tissue level, revealing currently unsuspected vulnerabilities.

Here, we started to shed light on these outstanding questions by focusing on the single-cell characterization of metastatic breast cancer (BC), a worldwide leading cause of death.[3] Little

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

is known about the nature of the malignant cells populating BC metastatic lesions. Most single-cell analyses have focused on primary BC tumors, with only recent advances targeting metastases.[4-7] Although these studies are very valuable resources, they left unaddressed the fundamental changes in cell state that define metastasis and distinguish it from primary tumors. In this study, we compared human BC metastases with primary tumors at the single-cell level and discovered that overt human BC metastases are formed by cells activating a peculiar gene-expression program, combining adult luminal mammary gland identity with reactivation of a morphogenetic program typical of developing tubular organs. By 3D imaging, we show that human and murine metastases indeed outgrow as a trabecular lattice of thick epithelial cords, optimally exploiting their available space by adopting a defined architectural *Bauplan*. Epigenetic investigations on primary metastatic BC organoids led to the discovery of ETV1/4/5 as the master transcription factors (TFs) of the metastatic morphogenetic program, enabling functional investigations revealing that this program is essential for metastatic outgrowth but dispensable for primary tumor growth and micrometastatic dissemination.

RESULTS

Linking human macrometastases to tubular morphogenesis

We prospectively collected fresh surgical specimens of BC primary tumors and metastases along with healthy mammary glands (normal MGs) (Figures 1A and S1A; Table S1A). Dissociated tissues were profiled by single-cell RNA sequencing (scRNA-seq) with the 10× Genomics droplet-based method. We included scRNA-seq data of two liver metastases and eleven primaries from public databases[4,8] (Figure 1A). This collection included the major human BC subtypes, namely estrogen-receptor positive (ER+), HER2-amplified (HER2+), and triple-negative (TNBC) tumors. In total, we retained high-quality transcriptional profiles relative to 48,085 cells from metastases, 91,543 cells from primary BCs, and 32,724 cells from normal MGs.

Cells were subjected to clustering to identify the major cell types (Figure 1B, left) and then annotated using canonical lineage markers (Figure 1B, right). As expected, the normal MG contained three well-defined basal, luminal type-1 (L1), and luminal type-2 (L2) epithelial cell populations (Figure S1A). In all tumor samples, we identified clusters of epithelial, endothelial, stromal, and immune cells (Figures 1B, 1C, S1B, and S1C). We unambiguously identified tumor cells by estimating single-cell copy-number variation (CNV) profiles[3] (Figure S1D). Primary and metastatic BCs could also be stratified according to the expected PAM50 molecular subtypes,[9] that is, luminal-like, Her2, and basal-like (Figures S1E and S1F).

All main BC subtypes are known to arise from luminal cells of the adult mammary gland.[3] Consistently, using a luminal-specific gene signature (Table S1B), primary and metastatic BC cells express typical adult luminal markers, such as EPCAM, KRT18, KRT19, CD24, and others (Figures S2A and S2B). Only basal-like lesions co-expressed luminal markers with some basal cell markers (Figure S2C). Thus, metastatic human BC lesions remain well-rooted to their original mammary luminal cell state with no overt diversion to other epithelial lineages.

Next, we sought to identify features consistently enriched in metastatic cancer cells relative to primary tumors. We leveraged our scRNA-seq datasets and performed differential gene expression (DGE) analyses between metastatic and primary tumor samples, using a pseudobulk approach,[10] such as normalizing for tumor cell numerosity and minimizing false discovery rates (FDRs) (Figure 1D). The resulting ranked gene list was interrogated by gene-set enrichment analysis (GSEA) for Gene Ontology (GO) biological processes (Figure 1D; Table S1C). Unsupervised hierarchical clustering of enriched GO terms (FDR < 0.001) suggested two main biological underpinnings: (1) cell cycle, in agreement with bulk RNA sequencing (RNA-seq) reports that metastatic BC exhibits elevated proliferation scores compared to primaries,[2] and (2) an unexpected cluster centered on tubular morphogenesis, comprising redundant GO terms for the development of tubular organs, epithelial morphogenesis, and vascular development (Figure S2D; Table S1C).

Prompted by this result, we asked whether metastatic BC cells, while retaining adult mammary luminal identity, also reactivate developmental programs of the mammary gland—an organ that grows via branching tubulogenesis. We therefore derived a fetal mammary morphogenesis (FMM) signature by re-analyzing published mouse scRNA-seq data purged of any gene related to cell cycle or shared with human adult mammary epithelium (see STAR Methods). By applying GSEA to our metastases vs. primary tumors DGE list, the FMM signature was significantly enriched in metastases (Figure 1E; Table S1D). Similar results were obtained with two independent fetal mammary signatures (Table S1E). By contrast, there was no enrichment of signatures denoting cell states of the murine pubertal (postnatal) or of the human adult mammary gland or denoting undifferentiated cell states of other organs (Tables S1D–S1F).

A fetal morphogenesis program is activated in macrometastases

To validate the notion that metastases show an enrichment for the FMM signature, we calculated the average expression of the FMM gene set in each tumor of our scRNA-seq cohort. All metastatic lesions displayed elevated FMM expression (Figure 1F). We next investigated to what extent the FMM signature is activated within individual cells composing primary tumors and metastases. Remarkably, all human metastases were composed almost entirely of cells with an activated FMM program compared with normal mammary gland cells, and this occurred independently of tumor subtype, as classified by the PAM50 classifier (Figure S2E). The situation across primary tumors appeared more heterogeneous, with most primary lesions being composed by tumor cells essentially void of FMM expression (i.e., expressing it at levels comparable to the normal MG) (Figure S2E).

Since our scRNA-seq cohort did not include primary-metastasis pairs from the same patient (i.e., matched), we analyzed the AURORA-US consortium dataset of matched primaries and metastases,[2] obtained from RNA extracted from archival samples and then profiled by bulk RNA-seq. Notably, the FMM signature

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

CellPress
OPEN ACCESS

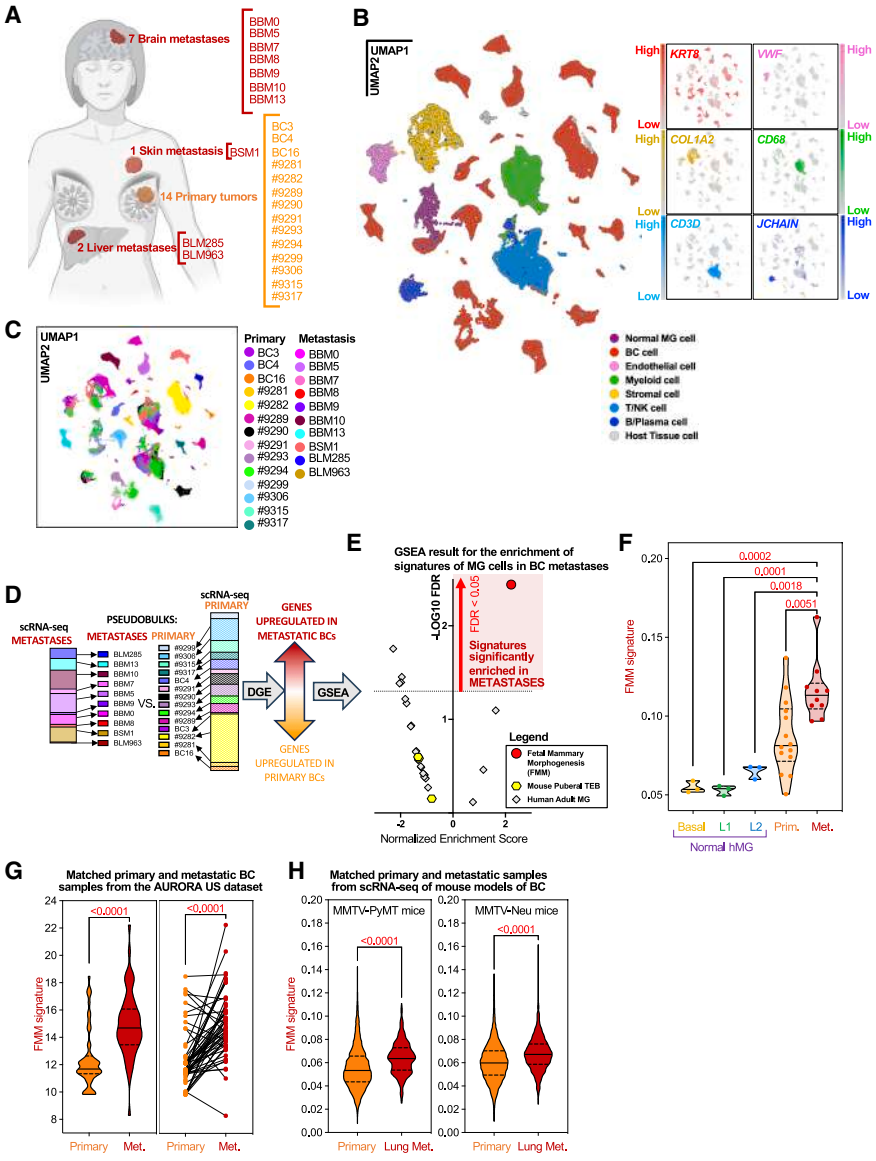


Figure 1. scRNA-seq analyses of metastatic BC cells reveal activation of a tubular morphogenesis program

(A) BC samples used in this study. BBMs, breast brain metastases; BSMs, breast skin metastases; BLMs, breast liver metastases[8]; BC3, BC4, and BC16, primary BCs. Samples from Pal et al.[4] are indicated with part of their GEO number. Created with Biorender.com.

(B) Uniform manifold approximation and projection (UMAP) of cells from BC tissues colored according to their annotation (left panel) or according to the expression of selected cell markers (right panels): *KRT8*, tumoral and normal epithelial cells; *VWF*, endothelial cells; *COL1A2*, stromal cells; *CD68*, myeloid cells; *CD3D*, T cells; *JCHAIN*, B cells and plasma cells.

(C) UMAP as in (B), with cells colored according to patients.

(D) Schematic representation of the procedure to identify signatures differentially expressed between metastatic and primary BC cells.

(E) Volcano plot of GSEA results for enrichment of signatures for normal epithelial cell types of the human adult MG and of the developing mouse MG (fetal MG or puberal terminal end buds [TEBs]) in metastatic vs. primary BC cells. Threshold for significant enrichment: FDR = 0.05 (dotted line). See Table S1D for full results.

(F) Average expression of the FMM signature from scRNA-seq data of epithelial cell populations of normal human MGs (hMG; $n = 3$ for each population), representing background baseline levels, and in the cancer cells of individual primary ($n = 14$) and metastatic BCs ($n = 10$).

(G) Expression of the FMM signature from bulk RNA-seq of 54 matched primary and distant (non-lymph node) metastatic BC samples of the Aurora US dataset.

(H) Expression of the FMM signature in cancer cells from primary tumors and lung metastases of mouse models of BC, as calculated from scRNA-seq data (see Figure S2F for the full cell composition of these samples).

p values: two-tailed one-way ANOVA with Dunnett's T3 multiple comparisons in (F), two-tailed paired *t* test in (G), and two-tailed *t* test with Welch's correction in (H).

See also Figures S1 and S2 and Table S1.

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

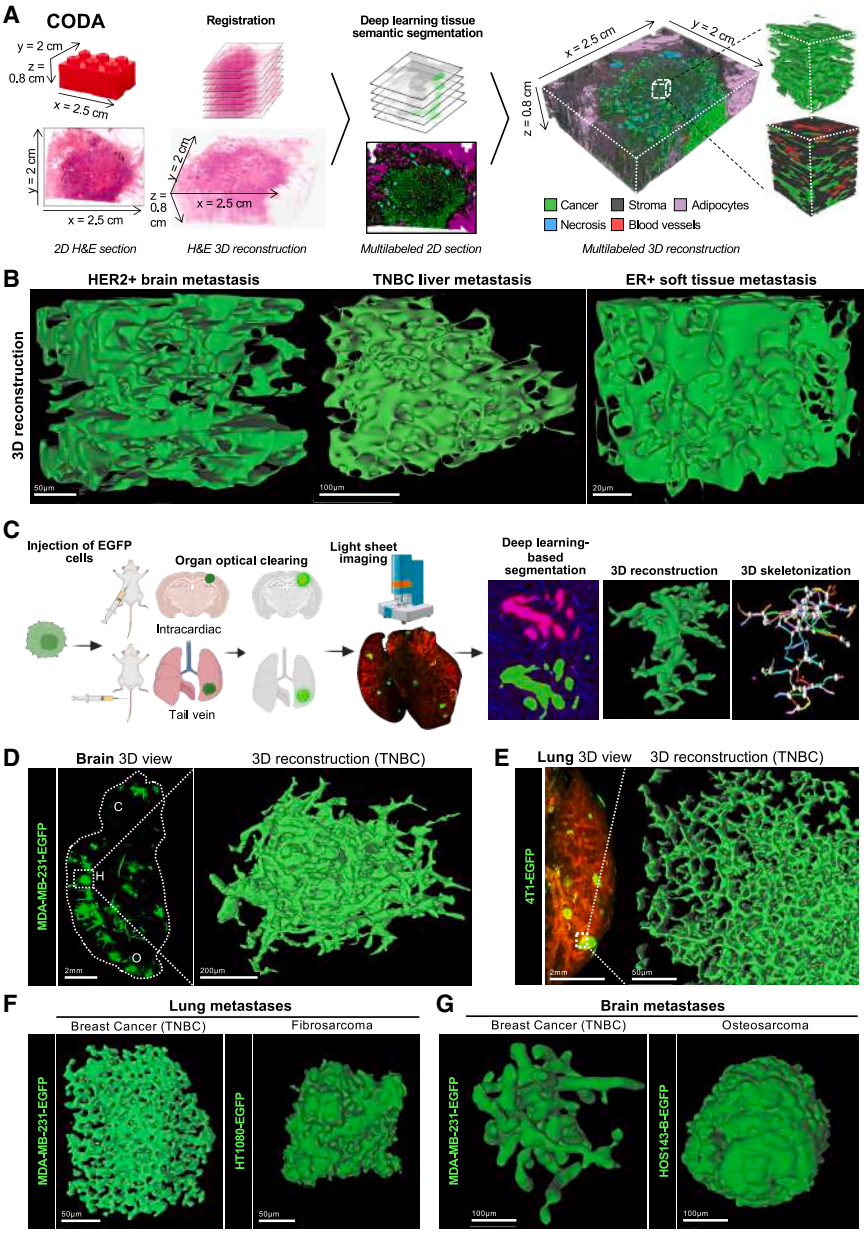


Figure 2. The 3D architecture of BC metastases

(A) Schematics of the 3D reconstruction procedure using CODA. Created with Biorender.com.

(B) 3D reconstructions (green: cancer cells) of an HER2$^+$ brain metastasis (left panel), a TNBC liver metastasis (middle panel), and an ER$^+$/HER2$^-$ soft-tissue metastasis (right panel).

(C) Experimental procedure for obtaining 3D reconstructions of experimental metastases in mice (see STAR Methods for details). Created with Biorender.com.

(D) 3D reconstruction of brain metastases obtained by injecting EGFP+MDA-MB-231 brain-trained TNBC cells in non-obese diabetic (NOD)/severe combined immunodeficiency (SCID) mice. Left panel: 3D view of a brain hemisphere. Right panel: representative 3D reconstruction of a single metastatic nodule.

(E) 3D reconstruction of lung metastases obtained after tail vein injection of EGFP+ 4T1 mouse TNBC cells in BALB/c mice. Left panel: 3D view of a lung lobe (green: tumor cells, red: autofluorescence). Right panel: representative 3D reconstruction of a single metastatic nodule.

(F and G) 3D reconstructions of similar-sized metastases formed by BC cells (MDA-MB-231) or non-epithelial cancer cells (HT1080: fibrosarcoma, HOS143-B: osteosarcoma) from lung (F) or brain metastases (G), obtained after intracardiac injection in NOD-SCID mice.

See also Figure S3 and Videos S1–S3.

was significantly higher in metastases than in their matched primaries (Figure 1G). To independently validate these findings, we performed scRNA-seq on matched primary tumors and spontaneous lung metastases from two murine models, MMTV-PyMT and MMTV-HER2/Neu (modeling TNBC and HER2+ disease, respectively) (Figure S2F). In both models, metastases displayed higher FMM expression than their corresponding primaries (Figure 1H). These findings suggest a conserved developmental constraint associated with metastatization.

The architectural blueprint of metastatic outgrowth

We next asked whether metastases are organized as tubular structures in 3D. For this, we used the CODA deep-learning workflow[11] to reconstruct 3D shape and microanatomy from 2D H&E serial sections (Figure 2A). We analyzed a human HER2+ brain metastasis, a TNBC liver metastasis, and an ER+ soft-tissue metastasis (Figure 2B). Across sites and lesions, metastatic BC cells formed a 3D lattice of multicellular, lumen-absent epithelial cords arranged in a fractal, trabecular architecture, expanding by branching morphogenesis with extensive anastomoses (Figure 2B; see Video S1 for a fly-through of the structure). This pattern was distinct from the tumor vasculature, and the epithelial cord network was embedded within an extracellular-matrix-rich stroma that was continuous throughout the tissue (Video S1).

We next extended these findings to experimental metastases generated by injecting human and mouse BC cell lines via the tail vein or intracardially. As schematized in Figure 2C, optical clearing and light-sheet microscopy of metastatic lesions formed by TNBC models (human MDA-MB-231 and mouse 4T1) and an HER2+ model (human HCC1954) again revealed a 3D lattice of branching and anastomosed epithelial cords in the brain (Figures 2D, 2G, left, and S3A; Video S2), lung (Figures 2E, 2F, left, and S3B; Video S3), and soft tissue (Figure S3C). Subcutaneous outgrowths of metastatic organoids derived from a human BC brain metastasis likewise exhibited extensive trabecular morphology (Figure S3D).

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

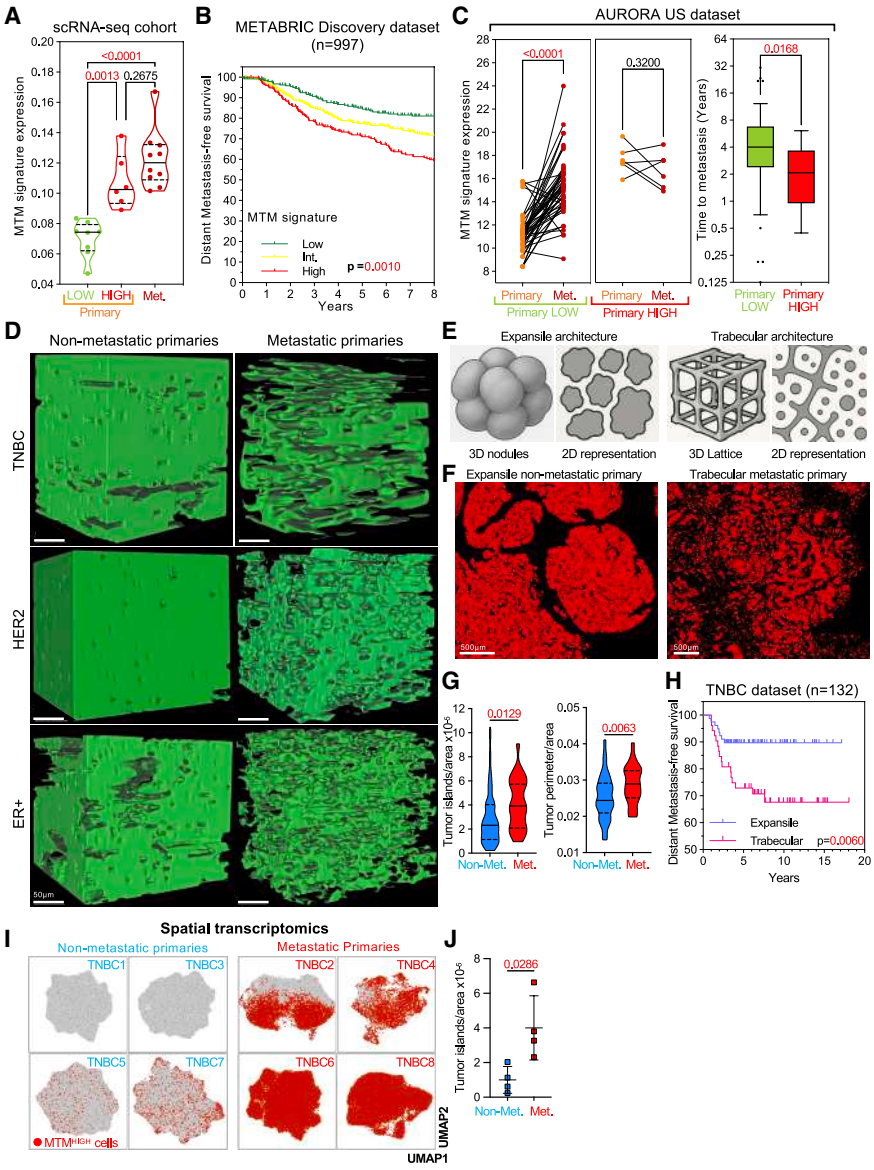


Figure 3. The MTM program is associated with the 3D architecture and metastatic proclivity of primary BCs

(A) Average expression of the MTM signature in primary BCs of our scRNA-seq cohort, classified as "MTM low" ($n = 8$) or "MTM high" ($n = 6$) (see STAR Methods); MTM expression in metastases ($n = 10$) serves as a reference.

(B) Kaplan-Meier analysis for metastasis-free survival in $n = 997$ BC patients from the METABRIC Discovery dataset, stratified according to high, intermediate (Int), or low expression of the MTM signature.

(C) Correlation between the expression of the MTM signature in primary BCs (left and middle panels) and time to metastasis in the Aurora US dataset (right panel). Primary tumors were stratified for having low ($n = 48$) or high ($n = 6$) expression of the MTM signature and then compared for the time (years) they required to form clinically detectable metastases.

(D) 3D reconstruction (via CODA as in Figure 2A) of cancer cells (green) of primary TNBCs (top panels), HER2-positive BCs (middle panels), or ER-positive BCs (bottom panels) from patients who either developed metastases (right panels) or experienced no distant recurrence (left panels).

(E) Schematics on how tumors with different 3D architecture (left, expansile; right, trabecular) are predicted to appear in 2D sections. See also Figure S4F for examples of experimental metastases displaying the two types of architectures.

(F) Spatial visualization of the shape of tumor tissue (red) from representative portions of 2D sections of the same TNBC primaries shown in (D).

(G) Quantification of the topo-morphometric parameters in the primary tumors of the TNBC cohort, from patients that either did not experience distant recurrence (non-met, $n = 108$) or developed metastases (met., $n = 24$).

(H) Kaplan-Meier analysis for metastasis-free survival in $n = 132$ patients from the TNBC cohort, stratified according to their expansile or trabecular architecture, inferred from the quantitation of tumor islands-area ratio, as in (G).

(I) UMAPs of tumor cells depicting the presence of MTM^{HIGH} cells in primary TNBCs representative of the cohort used for the morphological study in (G-H), as determined by CosMx-based ST (see STAR Methods).

(legend continued on next page)

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

To determine whether this architecture is tumor-intrinsic, rather than imposed by host tissue, we compared epithelial and non-epithelial metastases within the same organs. Following intracardiac injection of fibrosarcoma (HT1080) and osteosarcoma (HOS143B) cells, lung and brain metastases reconstructed by light-sheet microscopy were uniformly compact and nodular, displaying smooth, rounded borders typical of expansile growth and lacking the trabecular lattice of multicellular cords seen in size-matched BC metastases (compare the right vs. left panels of Figures 2F and 2G; Videos S2 and S3), as confirmed by quantifications of sphericity and branching index (Figures S3F and S3G). Primary brain tumors represent another control, growing as compact nodules with slender, radially oriented infiltrative projections (Figures S3E and S3G; Video S2). Finally, aspects of branching morphogenesis can be partially recapitulated *in vitro*, as shown by extensive branched outgrowth of 4T1 cells in thick collagen gels (Figure S3H). Collectively, these imaging analyses support metastatic morphogenesis as, at least in part, a self-organizing property of malignant epithelial cells. Thus, trabeculation is not imposed by the organ microenvironment but reflects a tumor-intrinsic capacity of epithelial metastatic BC cells.

To better define the developmental program supporting this malignant architecture, we intersected the fetal FMM gene list with genes enriched in metastasis from our scRNA-seq pseudo-bulk analyses and the AURORA-US dataset. This filtering removed developmental genes not enriched in metastasis, yielding a distilled signature that we term metastatic trabecular morphogenesis (MTM) (Table S1G). MTM contains signaling modules and effectors known to govern branching organs, including (1) fibroblast growth factor (FGF) signaling mediators and targets central to branching morphogenesis[12] (e.g., HS6ST1 and SPRY4); (2) non-canonical Wnt/planar cell polarity components (VANGL2, FZD2, FZD9, DVL2, and ROR2), implicated in epithelial tube formation in embryonic mammary glands, lungs, and kidneys[13]; (3) CXCR4, required for branching in several glands, angiogenesis, and BC metastasis[14]; and (4) Notch pathway regulators (EGFL7, DLL4, and HEY1), essential for vascular tree development.[15] The signature also includes TFs involved in mammary development and epithelial organization, such as GRHL3, PYGO2, and SOX11/12.[16-18]

Next, we monitored the expression of the MTM signature in our scRNA-seq cohort. Using as a reference the epithelial cell population of the normal mammary gland, metastases were almost entirely composed by cancer cells expressing the MTM signature, well above the background levels of normal mammary gland cells (MTM^{HIGH} cells) (Figure S4A; Table S1J). Instead, signatures measuring other properties associated with malignancy or early steps of metastatization[19-22]—such as cell-cycle, epithelial-to-mesenchymal transition (EMT), or basal-trait signatures—were not significantly enriched, being expressed at comparable or lower levels in respect to normal mammary gland luminal cells (Figure S4B). This is in line with the view that the MTM captures a morphogenetic program substantially distinct from prior correlates of aggressiveness.

MTM program in primary tumors has prognostic value

Having established that metastases are uniformly MTM^{HIGH}, we next asked to what extent this morphogenetic program is already activated in primary tumor cells. In our scRNA-seq cohort, primary tumors appeared heterogeneous, ranging from lesions composed entirely by cells essentially void of MTM expression (i.e., MTM^{LOW}) to lesions expressing it at levels comparable to metastases (MTM^{HIGH}) (Figures 3A and S4C). This variability suggests that the MTM program is not exclusive to overt metastases but can also emerge within a subset of primaries. To verify this in a larger clinical dataset, we analyzed the METABRIC Discovery cohort of treatment-naive primary BCs.[23] As in our scRNA-seq data, each intrinsic subtype (luminal A/B, HER2-enriched, basal-like, and normal-like) contained both MTM^{HIGH} and MTM^{LOW} tumors, indicating that MTM heterogeneity is not confined to a single subtype (Figure S4D).

We then asked whether MTM heterogeneity correlates with metastatic risk. In METABRIC, Kaplan-Meier analyses showed that MTM^{HIGH} tumors had a significantly higher probability of developing metastases than MTM^{LOW} tumors (Figure 3B). Importantly, in multivariable Cox proportional hazards models, the MTM signature remained an independent predictor of metastatic proclivity over established clinical covariates (subtype, lymph node status, histological grade, age, and systemic therapies) (Figure S4E). Consistent with this, in the AURORA-US cohort of matched cases, MTM expression in the primary tumor inversely correlated with time to metastatic emergence, such that primaries with the highest MTM levels metastasized earlier (Figure 3C).

Cancer 3D architectures are per se associated with outcome

Building on the two above observations—first, that overt metastases organize as trabecular lattices, and second, that the MTM program is enriched in metastases but also marks a subset of primaries with higher metastatic risk—we asked whether primary tumor architecture itself correlates with clinical outcome and with MTM status. For this, we first reconstructed six primary BCs with long-term prospective follow-up, selecting metastatic and non-metastatic cases across ER+, HER2+, and TNBC subtypes, and outcome windows were $\geq$20 years metastasis-free for ER+, up to 14 years for HER2+ (beyond which recurrence risk is negligible), and 14 years for TNBC (where relapses typically occur within three years). ER+ and TNBC primary tumors were matched with the 3D reconstructed metastatic lesions shown above (Figure 2B). All prospectively metastatic primary tumors displayed a trabecular lattice comparable to distant

(J) Quantitation of the topo-morphometric parameter tumor islands/area of TNBC samples shown in (I).
Graphs are box-and-whiskers plots (box, 25th–75th percentile; whiskers, 10th–90th percentile) in the right panel of (C) and scatter plots, mean and Standard Deviation (SD), in (J).
p values: two-tailed one-way ANOVA with Dunnett's T3 multiple comparisons in (A), log-rank (Mantel-Cox) test in (B) and (H), two-tailed paired *t* test in the left and middle panels of (C), two-tailed *t* test with Welch's correction in the right panel of (C) and in (G), and two-tailed Mann-Whitney test in (J).
See also Figure S4, Table S1, and Video S4.

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

CellPress
OPEN ACCESS

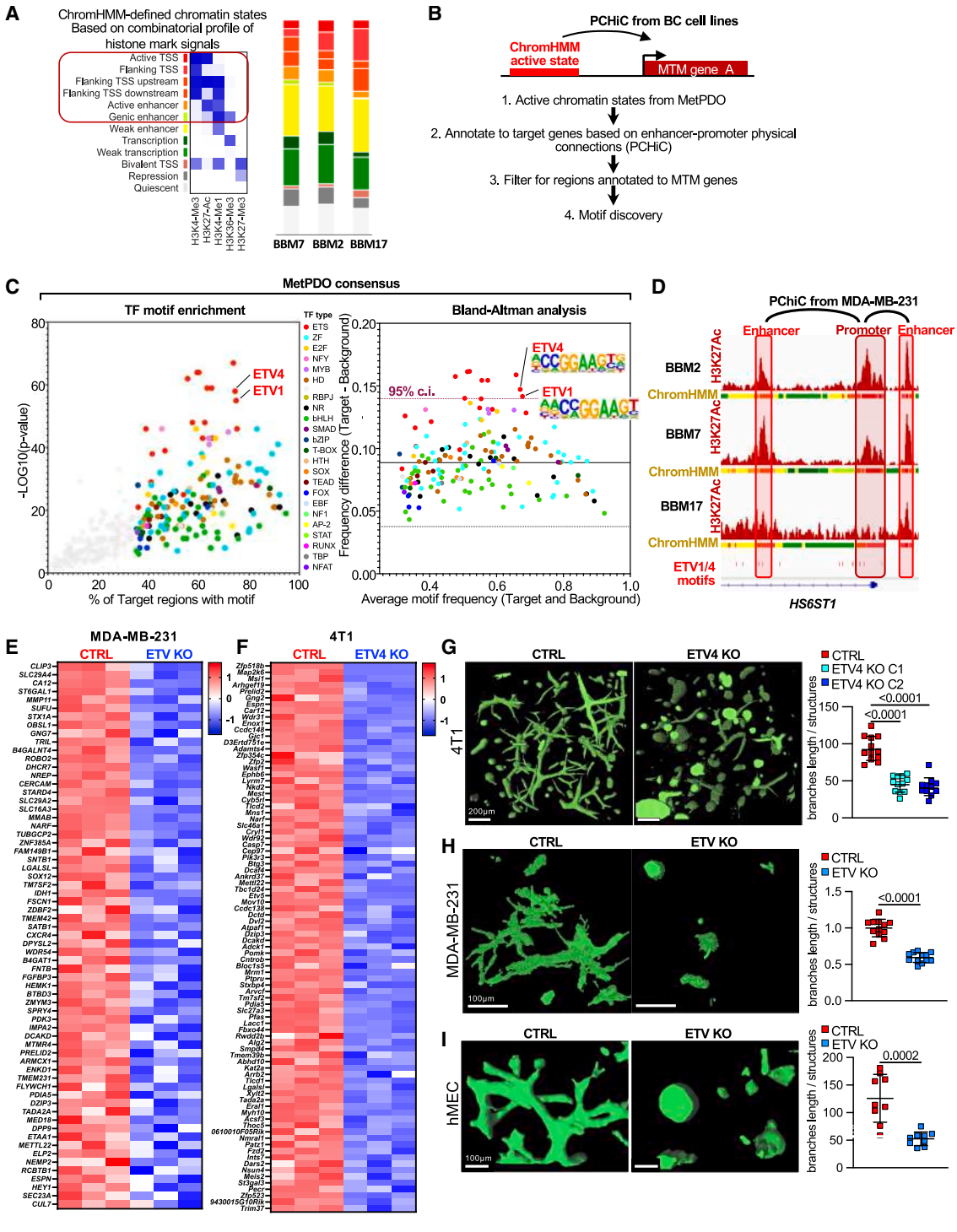


Figure 4. ETV1/4/5 are master genes of the MTM program

(A) Left panel: heatmaps depicting the average intensity of ChIP-seq signals for the indicated chromatin markers in the different ChromHMM-defined epigenetic states obtained from metPDOs (see Table S1A for clinical details). Active chromatin regions (red box) were used for inference of the master TFs of the MTM program. Right panel: bar plots showing the distribution of ChromHMM states in each metPDO.

(B) Experimental flow to identify TF motifs enriched in the regulatory regions of the genes of the MTM signature.

(C) Enrichment analyses of TF motifs in the regulatory regions associated with MTM signature genes in the metPDO consensus. Left panel: scatterplot showing the percentage of regulatory regions of the MTM genes containing individual TF motifs vs. the associated *p* value. TF motifs with an FDR q-value < 0.0001 ($n = 167$) are colored according to their TF family and were probed by Bland-Altman analysis (right panel) to identify those that are overrepresented in the regulatory regions

(legend continued on next page)

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

metastases, whereas non-metastatic primaries maintained a compact, solid organization indicative of expansile growth (Figure 3D; Video S4). Thus, 3D architectures of primary tumors correlate with clinical outcome and mirror the architectural blueprint of overt metastases.

To generalize this link in a larger series, we focused on TNBC because this is a paradigm of BC for which there is a clear-cut dichotomy in metastatic potential: primary tumors either display an early recurrence within a few years (typically 2–3 years) or are considered cured.[24] We assembled a real-world cohort of 132 upfront-resected TNBC primaries with longitudinal follow-up at our institution that did not receive any neoadjuvant therapy (Table S1H), as such allowing interrogation of baseline tumor biology (24 metastatic, 108 non-metastatic, proportions consistent with the literature[24]).

To scale up morphological analysis beyond individual 3D reconstructions to the cohort-scale level, we developed 2D topological descriptors as surrogates that capture the topological hallmarks of trabecular vs. expansile growth, enabling scalable pathology-ready quantification on histological sections. To start, as schematized in Figure 3E, we posited that 2D cross-sections through a trabecular 3D lattice would yield numerous small, near-circular or slightly elongated profiles, whereas an expansile mass would present as fewer, larger areas. We first validated this concept on sections of the very same TNBC cases for which we had full 3D reconstructions (metastatic and non-metastatic primaries). For this, we developed a supervised machine-learning pipeline to segment tumor epithelium on histological slides and generated tumor-cell masks for morphometric analysis. In the 3D-validated TNBC samples, the metastatic primary tumor showed a markedly higher density of components than the non-metastatic TNBC with expansile morphology, capturing in 2D the hallmarks of the underlying 3D architecture (Figure 3F). The opposing 2D morphologies of branched vs. compact lesions were also revalidated in sections of the experimental metastases from BC and sarcoma cell lines (i.e., shown above in Figure 2F; Figure S4F). We supported these morphological visual assessments with unbiased quantifications of topological descriptors of spatial organization (see STAR Methods), such as density of tumor islands (components per mm^2) and perimeter-to-area ratio (Figure S4G). Thus, the 3D organization of BC can be inferred from quantitative morphometrics on 2D sections. Remarkably, application of these 2D topo-morphometric descriptors to the whole cohort of 132 cases revealed that the same features were strongly enriched in tumors from patients who later developed metastases, but not in non-metastatic cases (Figure 3G).

Finally, we tested the prognostic value of tumor architecture per se. As shown in Figure 3H, Kaplan-Meier analyses revealed that primary tumors with a trabecular architecture—displaying a higher density of tumor islands—had a significantly higher probability of developing metastases than those with an expansile structure. Thus, tumor architecture, inferred from quantitative morphometrics on routine histopathological sections, has prognostic value, recapitulating the risk stratification that could be obtained molecularly with the MTM signature in METABRIC.

Linking the MTM^{HIGH} cell state to 3D organization in primary tumors

Do MTM^{HIGH} tumor cells physically build the trabecular architecture in primary tumors that later metastasize? To answer this, we analyzed archival TNBC primaries with known outcomes that had been classified—by 3D reconstruction and/or morphometrics—as trabecular or expansile, respectively. Using high-resolution single-cell spatial transcriptomics (ST) (CosMx[25]) on corresponding formalin-fixed/paraffin-embedded (FFPE) sections, we profiled a total of 999,262 good-quality cells. Next, from these data, we quantified MTM^{HIGH} and MTM^{LOW} tumor cells within each lesion. A striking dichotomy emerged: metastatic primaries with trabecular organization were composed almost entirely of MTM^{HIGH} cells, whereas compact, expansile non-metastatic primaries consisted almost exclusively of MTM^{LOW} cells (Figure 3I; Table S1K; see also Figure 3J for corresponding topo-morphometric quantifications of the same tumors).

Collectively, the data presented so far integrate the molecular and architectural layers of the study: MTM^{HIGH} cells—endowed with the developmental capacity for branching morphogenesis—build the same 3D structures in overt metastases and in primary tumors that will later metastasize, whereas MTM^{LOW} tumors adopt a solid, non-branching organization compatible with local growth but insufficient for metastasis formation.

ETV1/4/5 as master transcriptional regulators of the metastatic cell state

Is the trabecular MTM^{HIGH} program so far associated with overt metastases also functionally relevant for metastatic outgrowth? Gene networks in developmental processes are often controlled by upstream master TFs. To identify the master TFs of the MTM signature, we took advantage of patient-derived organoids for

associated with MTM signature genes (target regions) compared with background regions. TF motifs surpassing the 95th percentile of the confidence interval (grape dotted line) are those specifically enriched in the regulatory regions of the MTM signature genes.

(D) Representative tracks of H3K27ac and ChromHMM profiles illustrating one gene of the MTM signature whose regulatory regions contain motifs predicted to bind ETV1/4 (red vertical lines). Interactions between enhancers and promoter regions based on PCHiC data from human MDA-MB-231 BC cells are shown above the graphs.

(E and F) Heatmaps from RNA-seq of 3D collagen cultures (n = 3 biological replicas each) showing standardized expression of MTM signature genes displaying the most significant downregulation upon ETV1/4/5 KO in MDA-MB-231 (E) or in 4T1 (F) BC cells, compared with their matched control cells (CTRL).

(G–I) Representative 3D reconstructions (left and middle panels) and quantifications of branch lengths (right panels) of CTRL or ETV4 KO (two KO clones, C1 and C2) 4T1 cells (G), CTRL or ETV-KO MDA-MB-231 cells (H), and normal human mammary epithelial cells (hMECs) (I) cultured in collagen I gels for 4 days. Results are from 3 independent experiments, each with at least 3 biological replicates.

Graphs in (G)–(I) are scatter plots, mean and SD. p values: one-way ANOVA with Sidak's multiple comparison post hoc test in (G) and unpaired two-tailed t test in (H) and (I).

See also Figure S5.

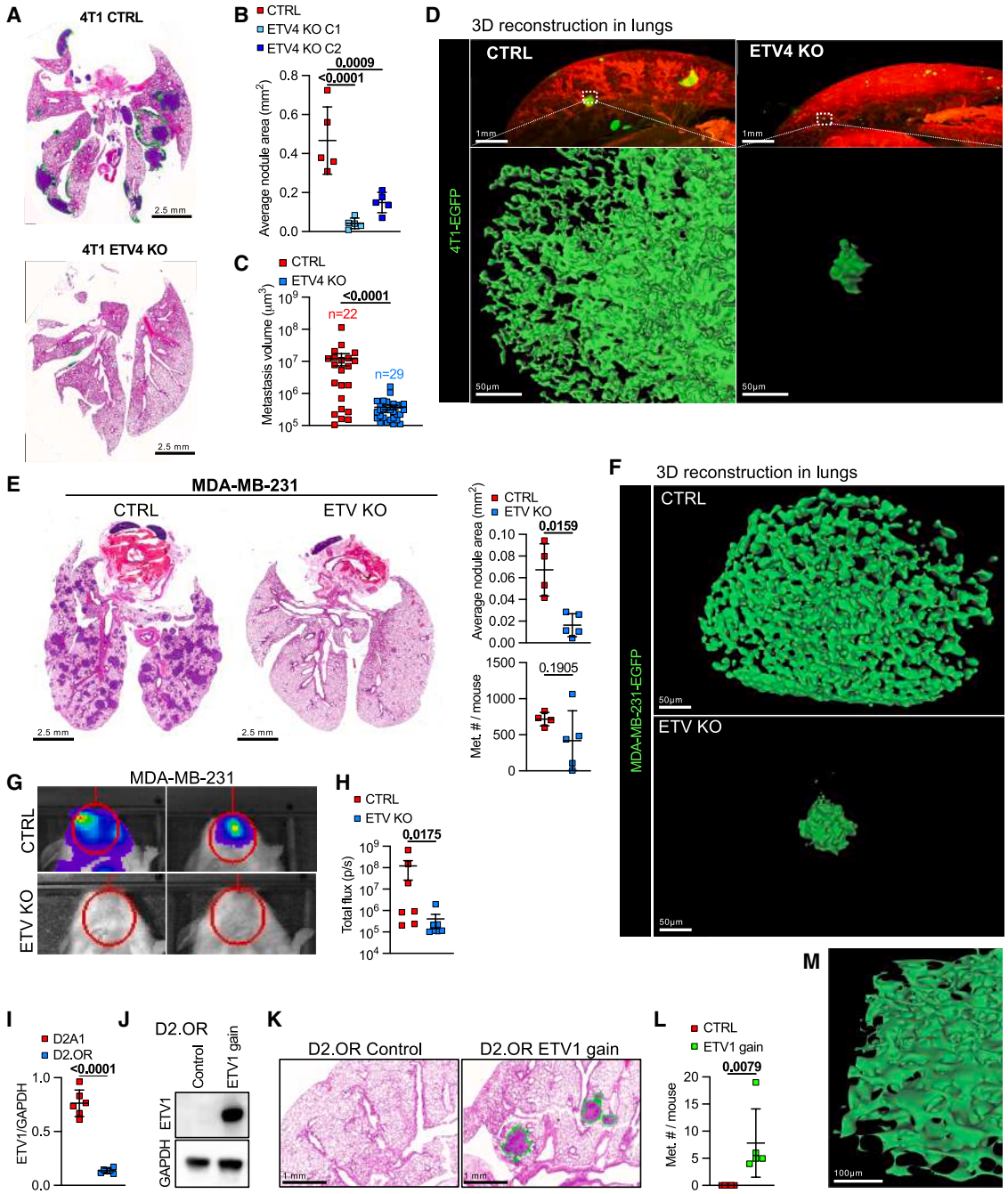


Figure 5. ETV1/4/5 are required for metastatic 3D morphogenesis and outgrowth

(A and B) BALB/c mice were injected in the tail vein with 5×10^4 CTRL ($n = 6$) or ETV4 KO (two KO clones C1 and C2, $n = 5$ each) 4T1 cells and killed after 3 weeks. (A) Representative H&E-stained paraffin sections of lungs. (B) Quantification of metastasis area.

(C and D) EGFP-expressing CTRL or ETV4 KO 4T1 cells were injected into the tail vein of BALB/c mice ($n = 5$ each) and collected after 12 days. (C) Quantification of the volume of individual nodules ($n = 22$ in CTRL, $n = 29$ in ETV4 KO) in representative mice (log10 scale). (D) Representative 3D views and 3D reconstructions of lung metastasis. See also Figures S6B and S6C for quantification of morphological parameters.

(E–H) Effects of ETV1/4/5 KO on the outgrowth of metastases from MDA-MB-231 cells injected in NOD/SCID mice.

(E) Left panels: representative H&E-stained sections of lungs from mice injected in the tail vein with 3×10^5 CTRL ($n = 4$ mice) or ETV-KO ($n = 5$ mice) cells. Right panels: quantification of average area (top) and number of metastatic nodules (bottom).

(F) Representative 3D reconstructions of lung metastases obtained after intracardiac injection of EGFP-expressing CTRL or ETV-KO cells. See also Video S5 for comparison between similar-sized metastases.

(G) Representative bioluminescence images of the heads of mice 3 weeks after intracardiac injection with 3×10^5 CTRL ($n = 7$ mice) or ETV-KO ($n = 7$ mice) luciferase-expressing brain-trained MDA-MB-231 cells.

(legend continued on next page)

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

three metastases established during this study (metPDOs). We used these models, as these more faithfully recapitulate the transcriptional landscape of *in vivo* tumors, limiting idiosyncrasies associated with established cell lines. We first carried out chromatin immunoprecipitation sequencing (ChIP-seq) from metPDOs, analyzing a panel of histone marks denoting promoters (H3K4-Me3), enhancers (H3K4-Me1), active chromatin states (H3K27-Ac), transcriptionally active chromatin (H3K36-Me3), and heterochromatin (H3K27-Me3). We then leveraged an established machine-learning approach (ChromHMM[26]) to perform *de novo* chromatin state characterization from ChIP-seq data (Figure 4A). To identify the genes regulated by these elements, we searched within the genome-wide maps of enhancer-promoter physical connections, previously established by 3D chromatin conformation studies in BC[27] (promoter-capture high-throughput chromatin conformation capture [PCHiC]), focusing our attention on *cis*-regulatory elements associated with the genes of the MTM signature (Figure 4B). *Cis*-regulatory elements connected to MTM genes were then searched for TF binding motifs using HOMER.

All analyses were performed in each organoid line, and epigenetic results were then merged to identify *cis*-acting regulatory elements shared across the three metastases (Figure S5A). We prioritized TF families whose motifs were significantly enriched in more than 50% of MTM-linked regulatory elements compared with elements associated with all other expressed genes (Figures 4C and S5B). Although motifs from 18 TF families appeared across samples, only one family stood out as consistently and strongly enriched in MTM regulatory regions: ETS factors. Notably, motifs recognized by the PEA3 subfamily of ETS factors, that is, the ETV1, ETV4, and ETV5 paralog genes, were the most overrepresented, occurring in approximately 75% of MTM-linked elements (examples in Figures 4D and S5C). The data put forward ETV1/4/5 as candidate master regulators of the MTM program. We further noted that ETV1/4/5 have been implicated in developmental lung branching morphogenesis,[28] an intriguing convergence that echoes the developmental program of BC metastases.

Next, we asked whether ETV TFs are causally involved in controlling the MTM state and the associated metastatic architecture in BC. We used CRISPR-Cas9 to inactivate ETVs according to their endogenous expression: 4T1 cells express Etv4 only, whereas MDA-MB-231 express ETV1/4/5 at comparable levels (Figure S5D). We generated an Etv4 knockout (KO) in 4T1 and a triple ETV1/4/5 KO in MDA-MB-231 (hereafter, ETV KO) (Figures S5E and S5F). RNA-seq showed a significant reduction of the MTM signature upon ETV inactivation (Figures 4E, 4F, and S5G). Functionally, Etv4 KO in 4T1 and triple ETV KO in MDA-MB-231 profoundly impaired branching in collagen gels, as cells formed spherical aggregates with an expansile growth pattern (Figures 4G and 4H). ETV loss did not alter 2D proliferation, wound-healing migration/polarization, or clonogenicity in mammosphere assays (Figures S5H–S5K), indicating that these factors are dispensable for generic growth or motility. More relevantly, these data establish that the ETV1/4/5 family is a master regulator of the MTM program and epithelial branching morphogenesis.

To validate the notion that ETVs govern branching as a redeployed developmental module rather than a cancer-specific invention, we examined normal human MG (hMG) organoids. For this, we took advantage of a well-established *ex vivo* phenomenon in which adult mammary epithelial cells transiently regain embryonic-like multipotency and branching capacity when removed from their native context and cultured in 3D.[29,30] ETV1/4/5 KO in primary hMG cells markedly reduced branching (Figures 4I and S5L), growing instead as compact spheroids. These data indicate that ETV factors are required for the deployment of an intrinsic branching morphogenesis program in normal human mammary epithelium.

Metastatic architecture drives outgrowth

We then assessed the *in vivo* relevance of ETVs for metastatic outgrowth. Following tail-vein injection of control vs. ETV4-KO 4T1 cells into BALB/c mice, ETV4 loss caused a marked reduction in macrometastatic burden in lung and mediastinal soft tissues (Figures 5A, 5B, and S6A). The number of micrometastatic foci per lung in the KO was comparable to the number of macrometastases in controls, but KO lesions failed to outgrow (see *n* vs. volume in Figure 5C), and 3D reconstructions of light-sheet images and topology quantification showed that ETV4KO micrometastases were indeed compact aggregates lacking extended tubular meshes (Figures 5D, S6B, and S6C). Importantly, this architectural difference persisted in size-matched lesions (Video S5). Similar results were obtained in the HER2+ TUBO model, where KO of Etv5 (the only expressed paralog) (Figure S5D) reduced macrometastatic outgrowth (Figures S6D–S6F).

Extending to human cells, we compared control and triple ETV-KO MDA-MB-231-lux cells delivered by tail vein (lung) or intracardiac injection (brain). ETV-KO dramatically reduced macrometastatic outgrowth in both organs (Figures 5E–5H). In lungs, residual KO lesions remained small with reduced branching, despite their number being comparable to control lesions

(H) Quantification of luminescence (log10 scale).
(I) RT-qPCR quantification of *Etv1* expression normalized to *Gapdh* in D2A1 and D2.OR mouse BC cell lines (3 independent experiments with 2 biological replicates per group).
(J–M) ETV1 gain of function confers lung metastatic abilities to D2.OR cells.
(J) Immunoblot for ETV1 expression in control and ETV1-gain D2.OR cells. GAPDH is a loading control.
(K) Representative H&E-stained paraffin sections of lungs of NOD-SCID mice injected in the tail vein with 3×10^5 D2.OR control ($n = 5$ mice) or ETV1 gain ($n = 5$ mice) cells and collected after 8 weeks.
(L) Quantification of the number of metastatic nodules per mouse.
(M) Representative 3D reconstructions of lung metastases from ETV1 gain D2.OR cells based on serial sections stained for CK8 (see STAR Methods).
Results in (B), (E), and (H) are representative of 2 independent experiments. Graphs are scatter plots, mean and SD. *p* values: Kruskal-Wallis test with Dunn's multiple comparison correction post hoc test for (B), two-tailed Mann-Whitney test for (C), (E), (H), and (L), and unpaired two-tailed *t* test for (I).
See also Figure S6.

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

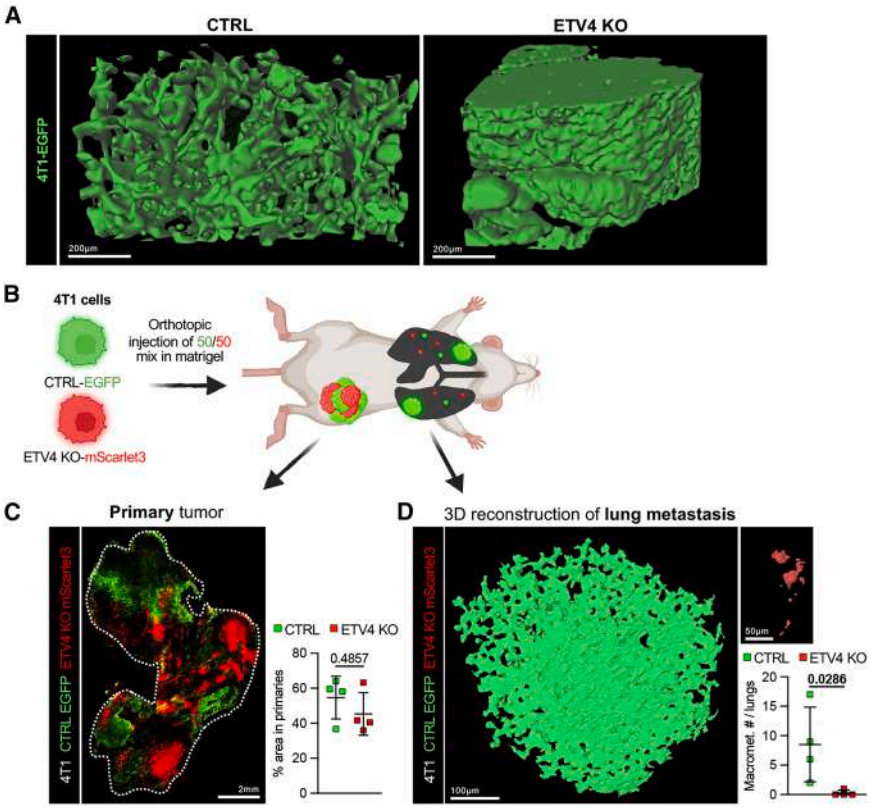


Figure 6. The MTM program is required for trabecular morphogenesis in primary tumors and to sustain their ability to form macrometastases

(A) 3D reconstructions of representative primary tumors obtained after orthotopic injection of 2×10^4 EGFP-expressing CTRL or ETV4 KO 4T1 cells in NOD/SCID mice.

(B–D) NOD/SCID mice ($n = 4$) were injected in the orthotopic site with an equal amount (10^4 each) of CTRL/MTM^{HIGH} (EGFP-expressing) and ETV4 KO/MTM^{LOW} (mScarlet3-expressing) 4T1 cells and killed after 4 weeks, as depicted in (B). (C) Fluorescent imaging of a frozen section from a representative primary tumor (left panel) and quantifications of the relative abundance of each cell type in all primary tumors (right panel). (D) 3D reconstructions from light-sheet imaging of representative metastases formed by CTRL (green, left panel) or ETV4 KO (red, top right panel) cells, and quantification of macrometastases formed by each cell type in the lungs (bottom right panel).

Graphs are scatter plots, mean and SD.

p values: two-tailed Mann-Whitney test.

See also Figure S6.

(Figures 5E, 5F, and S6G; Video S5). In the brain, only small human cytokeratin-positive clusters were detectable (Figure S6H). Size-matched incipient lesions from control and KO groups had similar Ki67 fractions (Figures S6I and S6J), consistent with *in vitro* data indicating that ETVs do not directly control proliferation in these contexts. Collectively, results in murine and human cellular models indicate that ETVs are dispensable for survival in circulation and extravasation but essential for establishing the metastatic *Bauplan* driving outgrowth.

Finally, we asked whether ETV activation may be sufficient to empower metastatic outgrowth in otherwise poorly metastatic cells. The isogenic D2A1/D2.OR pair provides a metastasis model in which D2A1 forms overt lung metastases after tail-vein injection, whereas D2.OR remains dormant.[31] D2A1, but not D2.OR, expresses a high level of Etv1 (Figure 5I), prompting us to test the pro-metastatic effect of ETV1 gain of function in D2.OR. Remarkably, overexpression of a constitutively active version of ETV1 was sufficient to induce metastatic outgrowths (Figures 5J–5L) that, as revealed by 3D reconstruction, also displayed the typical trabecular branching pattern of metastases so far analyzed in this study (Figure 5M).

ETVs control primary tumor morphogenesis and metastatic potential

What is the role of ETVs in primary tumors? Primary tumors arising from orthotopic injection of MTM^{HIGH}/control 4T1 cells displayed the same trabecular morphogenesis observed in their metastases (Figure 6A, left), consistent with the highly aggressive nature of this model. By contrast, MTM^{LOW}/ETV-deficient tumors lost branching capacity and adopted an expansile architecture (Figure 6A, right), reminiscent of MTM^{LOW}/non-metastatic human primaries. This architectural switch did not affect primary tumorigenesis or bulk growth (as tumor size, Ki67 index, and limiting-dilution tumor take were unchanged) (Figures S6K–S6O).

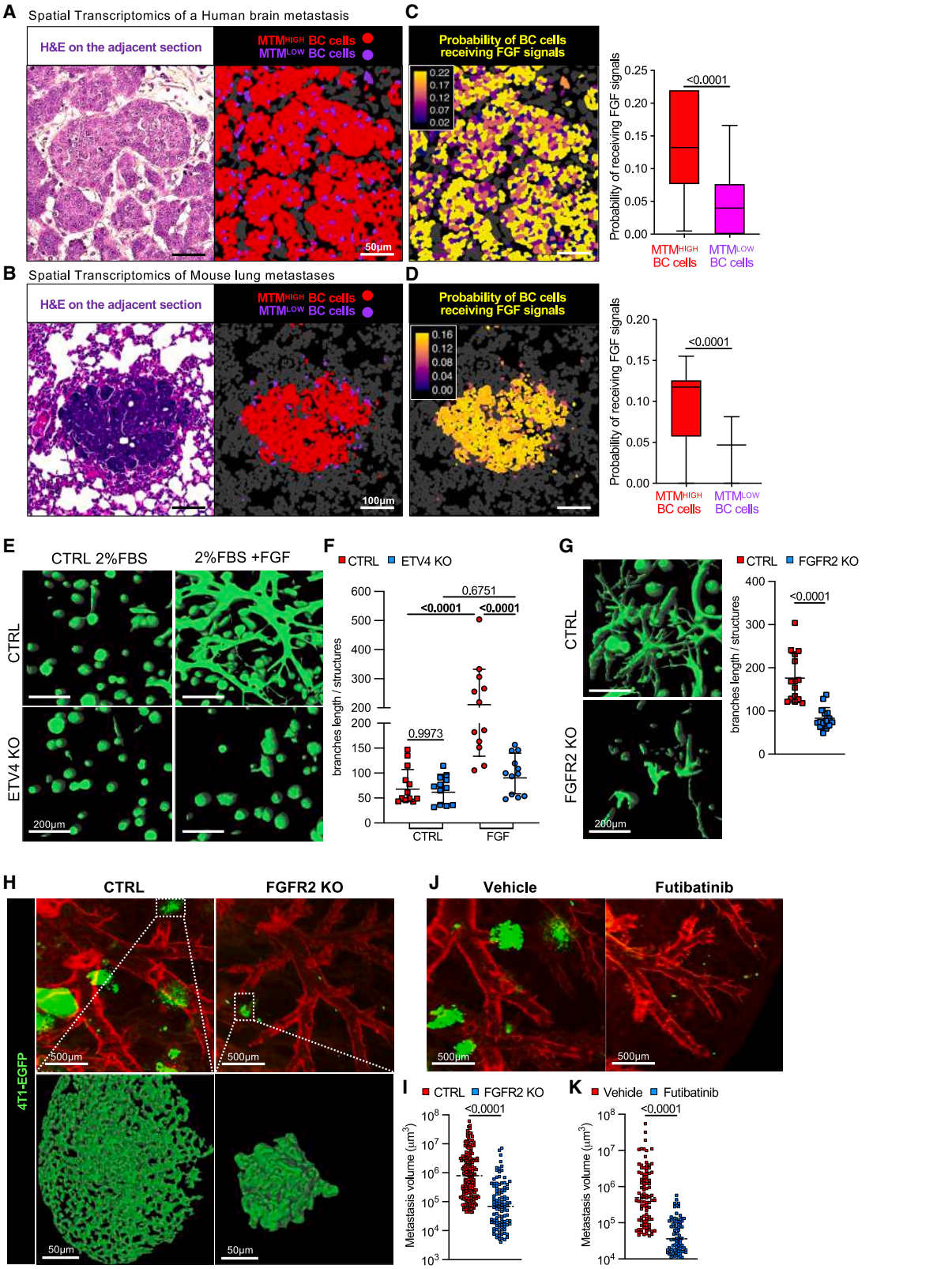
We next sought to determine whether MTM^{HIGH}/ETV-positive cells are specifically able to seed and/or generate overt distant metastases. For this, we orthotopically co-injected equal ratios of MTM^{HIGH}/ETV-intact (EGFP) and MTM^{LOW}/ETV-deficient (mScarlet) 4T1 cells (Figure 6B). Primary tumors retained both lineages at comparable proportions without growth bias (Figure 6C). Whole-organ light-sheet imaging revealed that both lineages seeded single cells and micrometastases in the lung; however, macrometastases were composed exclusively of MTM^{HIGH}/ETV-positive cells (Figure 6D).

In sum, combined with the results above, the data indicate that within primary tumors, the MTM^{HIGH} trabecular program driven by ETV1/4/5 prospectively earmarks lesions that will metastasize. Trabecular growth is not required for primary tumor formation per se, as non-metastatic primaries can also arise via alternative, expansile architectures, yet trabecular growth is essential for distant metastatic outgrowth.

The spatial landscape of metastases reveals actionable vulnerabilities

The MTM signature includes multiple transducers of canonical developmental pathways. We thus sought to investigate how the MTM program interfaces with extrinsic instructions orchestrated by external morphogens. For this, we integrated our

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009



(legend on next page)

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

scRNA-seq analyses of metastases with high-resolution ST using the CosMx 1,000-plex RNA imaging system. This allowed MTM scoring in individual cells within the tumor landscape, identification of stromal signaling sources, and inference of ligand-receptor communication across space.[25] We selected FFPE sections from one human brain metastasis and from spontaneous lung metastases arising in MMTV-PyMT mice. For each lesion, we imaged multiple areas, yielding 145,871 and 37,896 high-quality cells for the human and mouse samples, respectively. Cell types annotated from our scRNA-seq data were mapped onto these spatial transcriptomic datasets using Insitutype, which integrates transcript levels with spatial context for automated assignments.[32] Annotations were then validated and refined by marker-gene expression (Figure S7A).

In both metastatic specimens, tumor tissue consisted mainly of BC cells with high MTM activity (Figures 7A and 7B). In the ostensibly normal lung parenchyma of PyMT mice, spatial profiling also revealed isolated single cells or small clusters that were either MTM^{HIGH} or MTM^{LOW}, whereas growing lesions are almost entirely composed by MTM^{HIGH} BC cells, consistent with the above experimental observation that MTM^{LOW} cells can disseminate but fail to progress to macrometastases (Figure S7B).

To understand how intercellular communication may control trabecular morphogenesis, we focused on FGF signaling not only because regulators of this pathway are part of the MTM signature but also because it plays an evolutionarily conserved developmental role in branching morphogenesis of multiple organs.[12] To infer cell-cell signaling from ST data, we applied CellChat,[33] which uses gene expression to estimate ligand-receptor signaling probabilities. Remarkably, both human and mouse metastases showed strong predictions for FGF signals reaching tumor cells from the tumor microenvironment (TME), with MTM^{HIGH} BC cells being the main target of fibroblast growth factor receptor (FGFR) signaling (Figures 7C and 7D). In human brain metastasis, astrocytes were the dominant predicted FGF source (Figure S7D), whereas fibroblasts were the FGF source in mouse lung metastases (Figure S7E).

We next tested whether reception of FGF signaling is required for branching and outgrowth. *In vitro*, under low growth-factor conditions (2% FBS), exogenous FGF increased branch length in 4T1 cultures, without elevating MTM signature expression by RNA-seq (Figures 7E, top, 7F, and S7F). By contrast, ETV4-KO (MTM^{LOW}) 4T1 cells failed to branch in response to FGF, and MTM expression was downregulated irrespective of the presence of FGF (Figure 7E, bottom, 7F, and S7F). This indicates that FGF functions to induce branching morphogenesis downstream of the MTM program, but not as an inducer of the MTM^{HIGH} state. Consistently, CRISPR KO of FGFR2, the only expressed FGF receptor in 4T1 cells, abolished 3D branching without altering MTM transcriptional levels (Figures 7G and S7G–S7I), reinforcing the notion that FGF signaling is necessary to execute branching morphogenesis but not for the establishment of the MTM program.

In vivo, tail-vein injection of control vs. FGFR2-KO 4T1 cells demonstrated that FGF reception by tumor cells is essential for metastatic outgrowth and trabecular morphogenesis: control cells formed extensive branched networks, whereas FGFR2-KO cells produced only rounded, compact micrometastases (Figures 7H, 7I, and S7J). Primary tumor growth was indistinguishable between genotypes (Figure S7K), indicating that the FGF requirement is specific to metastatic expansion rather than tumorigenesis per se.

Finally, to test therapeutic leverage, we treated mice bearing disseminated metastatic cells with the clinical FGFR inhibitor futibatinib.[34] As shown in Figures 7J and 7K, macrometastatic burden was markedly reduced, while seeding of micrometastases persisted, in line with the genetic results with FGFR-KO. All in all, this establishes a proof of principle that an attribute of the MTM^{HIGH} program is activation of FGF-mediated branching, and this constitutes an actionable vulnerability for metastasis prevention.

Figure 7. FGF signaling promotes metastatic branching morphogenesis downstream of ETV1/4/5

(A and B) ST analysis of a brain metastasis from an HER2+BC patient (A) and a metastatic nodule spontaneously arising in the lungs of a PyMT mouse (B). Left panels are pictures from a field of view (FOV) stained by H&E. The right images are *in silico* reconstructions of the same FOVs, showing the spatial distribution of BC cells, classified according to MTM expression. TME cells are in dark gray. See also Figure S7C, depicting vascular cells in the same FOVs.

(C and D) Left panels: spatial distribution of the probability of receiving FGF signaling ($\times 10^{-3}$) by BC cells of the human brain (C) and mouse lung metastases (D) in the same FOVs shown in (A and B). Right panels: quantifications of the probability of receiving FGF signaling by BC cells classified according to their MTM^{HIGH} or MTM^{LOW} status.

(E and F) Representative 3D reconstructions (E) and branching quantification (n = 12 per group) (F) of 3D collagen cultures of CTRL and ETV4 KO 4T1 cells with low (2%) serum ±FGF.

(G) Representative 3D reconstructions (left panels) and quantifications of branch lengths (right panel, n = 15 per group) 3D collagen cultures of CTRL or FGFR2 KO 4T1 cells with 10% serum.

(H and I) BALB/c mice were injected in the tail vein with 5×10^4 CTRL or FGFR2 KO EGFP-expressing 4T1 cells, and lungs were collected after 12 days.

(H) Representative 3D views and 3D reconstructions of lung metastases.

(I) Quantification of the volume of individual nodules (CTRL n = 198; FGFR2 KO *n* = 109) in representative mice.

(J and K) BALB/c mice were injected in the tail vein with 5×10^4 4T1 cells, treated with the FGFRi futibatinib or vehicle, and killed after 12 days.

(J) Representative 3D views of lung metastases.

(K) Quantification of the volume of individual nodules (vehicle *n* = 97; futibatinib *n* = 83) in representative mice.

Results in (F) and (G) are from 3 independent experiments. Results in (I) and (K) are presented in a log10 scale and are representative of two independent experiments with n = 5 mice per group. Graphs in (C) and (D) are box-and-whiskers plots (box, 25th–75th percentile; whiskers, 5th–95th percentile). Graphs in (F), (G), (I), and (K) are scatter plots, mean and SD. *p* values: two-tailed *t* test with Welch's correction in (C) and (D), one-way ANOVA with Sidak's multiple comparison post hoc test in (F), unpaired two-tailed *t* test in (G), and two-tailed Mann-Whitney test in (I) and (K).

See also Figure S7.

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

DISCUSSION

Here, we report on the cellular constituents of human BC metastases, revealing that BC metastatic cells display chimeric features that superimpose adult and embryonic gene-expression programs. Metastatic cells have been proposed to transiently occupy a phenotypically "plastic" state,[1] though their concrete definition and selective advantage remain unclear. We posit that such plasticity in fact promotes the adoption of the MTM^{HIGH} state described here, thereby coupling metastatic outgrowth to a powerful modality of building epithelial tissues.

This work advances on the poorly explored view of malignancies as 3D entities that assemble into pathological organs. In this perspective, cancer's 3D shape merges with cancer biology: molecular circuitry is intrinsic to and projects into malignancy's architecture, setting the attributes and liabilities of metastatic outgrowth. Through 3D reconstructions and light-sheet imaging, metastases emerge not as chaotic masses but as spatially highly organized tissues: multicellular epithelial cords assemble into an anastomosing trabecular lattice that expands by branching morphogenesis (Video S6).

This ordered, almost fractal modality of BC metastatic outgrowth empowered by this cell state might represent an optimal strategy for maximizing epithelial surface area, thereby enhancing cell survival and proliferation by maximizing exposure to the limited availability of survival and growth signals from the host organ microenvironment. Indeed, branching morphogenesis of epithelial tubes has been modeled as an ideal modality to tile space in an optimal space-filling manner.[35,36]

Critically, we extend these principles to human primary tumors. Here, we found that metastatic primaries are indeed enriched in MTM^{HIGH} cells, and their Bauplan recapitulates the trabecular lattice observed in overt metastases. By contrast, MTM^{LOW} primaries adopt a different morphogenetic strategy, as these maintain a compact organization typical of expansile growth. Unable to turn on trabecular morphogenesis, MTM^{LOW} primaries are indeed non-metastatic. In turn, this is further validated by the fact that the MTM signature predicts primary tumor metastasis and metastatic latency in large cohorts, independently of subtype.

To scale architectural readouts beyond case studies, we established topological, morphometric 2D surrogates of the primary tumor's 3D organization: applying these on histological sections faithfully encodes the dichotomic architecture of metastatic (trabecular) vs. non-metastatic (expansile) primary BCs and stratifies patients according to metastasis-free survival. Thus, not only may the MTM signature serve as a prognostic biomarker identifying "bound-to-relapse" primary tumors, but also their architectural fingerprint is by itself measurable at diagnosis and prognostically informative.

By identifying the ETV1/4/5 subfamily of developmental ETS factors as the master TF of the MTM state, we reveal a coherent rationale for the redeployment of a branching program in metastasis. Importantly, the ETVs enabled us to shift from correlation to causality. Genetic perturbations demonstrated that ETV1/4/5 are necessary and sufficient for the trabecular growth modality and, as such, for metastatic outgrowth. Interestingly, ETVs appear dispensable for primary tumor take, baseline proliferation, clonogenicity, metastatic seeding, or general migration. Thus, ETV1/4/5, through their downstream MTM transcriptional program, are a primary architectural determinant of cancer shape and metastatic proclivity. Collectively, our results suggest the following cascade: ETV1/4/5 activity → MTM^{HIGH} state → trabecular morphogenesis → metastatic outgrowth.

This study holds manifold implications. Resolving the components of the MTM morphogenetic program may unveil potential actionable liabilities specific to the lethal macrometastatic stage. In particular, modules of the MTM state that govern responsiveness to microenvironmental cues may represent prime therapeutic entry points. Our results on the requirement of FGF signaling provide proof of principle that an in-depth functional dissection of the signals exchanged by metastatic cells with their microenvironment could be important to dissect additional therapeutically actionable options against metastasis.

Limitations of the study

This study interrogates the coupling between 3D form and function in malignant BC, yet we do not claim universality beyond BC. Other malignancies are likely to deploy distinct morphogenetic modules (e.g., expansile growth—as already shown here for sarcoma metastases—alveolar budding, rosette/cribriform patterning, and epithelial folding/buckling), each shaped by intrinsic lineage constraints, tissue mechanics, and local microenvironmental cues. Systematically charting these alternative *Baupläne* and testing whether any is rate-limiting for metastatic outgrowth in each lineage will require tumor-specific discovery pipelines and annotated cohorts analogous to those assembled here. A deeper understanding of morphogenetic logics across cancers may shed light on cancer's self-assembling principles and yield a new *taxonomy* of malignancy with a corresponding menu of architecture-targeted interventions. We are aware that this remains a challenging road. Comprehensive matched human primary-metastasis scRNA-seq remains largely uncommon in BC and many other tumor types. We mitigated this by triangulating the first discovery in scRNA-seq with matched mouse models and analyses of clinical cohorts, including curated public datasets; however, the data richness one could leverage for tumors other than BC remains limited. Looking forward, progress will hinge on coordinated biobanking of archived and fresh specimens with longitudinal outcomes, expansion of single-cell and high-plex spatial/transcriptomic analyses, and development of a suite of editing-tolerant patient-derived models and branching permissive organoid media. All patients were female of European ancestry, and this should be considered a limitation when interpreting the generalizability of the findings.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to the lead contact, Stefano Piccolo (stefano.piccolo@unipd.it).

Materials availability

All unique reagents generated in this study are available from the lead contact with a completed materials transfer agreement.

Data and code availability

- All scRNA-seq, bulk RNA-seq, and ChIP-seq raw data generated for this study, along with count matrices and metadata for each sample, have been entered in the Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) as GEO: GSE281492. Previously published scRNA-seq data of primary BC tumors and liver BC metastases were downloaded from the GEO (GEO: GSE161529) and the Broad DUOS (https://duos.broadinstitute.org), respectively.
- All code was performed using publicly available packages explicitly cited in the manuscript and deposited in https://github.com/labPiccolo/Caire_2026.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

We thank the patients and their families for their generosity in contributing to this study. We thank Vincenzo Vindigni and Franco Bassetto for reduction mastoplasty surgical samples; Alexander Gruhl from Miltenyi Biotech for light sheet imaging with the Ultramicroscope Blaze; the Nanostring/Bruker CosMx Spatial Imager R&D team; Annamaria Lia for intracranial injection of glioma cells; F. Miller, Y. Kato, and M. Montagner for the gift of mouse BC cell lines; and A. Rosato for the gift of GFP-Luc plasmid. We also thank Ylenia Silvestri, Luca Lazzari, Smeralda Rapisarda, and Paolo Luraghi for managing the Metamech platform. This work was supported by the following agencies/charities: FONDAZIONE AIRC, Italy, under the 5 per Mille 2019 - ID. 22759 program to S.P., M.P., S.M., and G.B., under the IG 2019 - ID. 23307 project to S.P., and under the IG 2022 - ID. 27883 project to M. Cordenonsi; the European Research Council Executive Agency (ERCEA) under the ERC-2022-ADG grant agreement no. 101098074-CHARTAGING to S.P.; the European Union-NextGenerationEU and STARS@UNIPD, GF-MET-Growth factor signaling in metastatic organotropism to F.Z.; the National Center for Gene Therapy and Drugs Based on RNA Technology (CUP B93D21010860004), funded in the framework of the National Recovery and Resilience Plan (PNRR-CN3), Mission 4, Component 2, Investment 1.4, funded by the European Union - Next Generation EU, Project CN00000041, CUP C93C22002780006, Spoke n.2 (“Cancer”) to F.Z., T.P., and S.P.; and Spoke n.4 (“Metabolic and cardiovascular diseases”) to M. Cordenonsi. R.C. was the recipient of a European Molecular Biology Organization “EMBO Postdoctoral Fellowship 2022, ALTF 8592022” and is the recipient of a European Commission Marie Skłodowska-Curie Fellowship under Horizon Europe Call: HORIZON-MSCA-2022-PF-01, ID. 101105584.

AUTHOR CONTRIBUTIONS

R.C. carried out most of the *in vivo* and *in vitro* experiments. E.A. carried out the *in vivo* experiments. M. Cordenonsi, S.B., M.M., M. Forcato, and O.R. contributed to bioinformatic analyses. F.Z., A.S., R. Bordo, R. Bason, and T.P. contributed to single-cell experiments and metastatic organoids. P.-H.W., T.P., and D.W. helped with CODA. G. Battilana and E.A. helped with tissue culture and *in vitro* experiments. M. Fakiola and R. Bason carried out ChIP-seq analyses. S.H. coordinated our initial work with the CosMx platform. V. Guzzardo provided histology sections. S.D., M.V.D., G.G., M. Carosi, M. Fassan, A.P.D.T., S.M., C.C., G. Blandino, F.V., and V. Guarneri provided clinical samples. P.C. contributed with transgenic mouse models. M.P. coordinated part of our single-cell collection, ChIP-seq experiments, and collection of metastatic organoids. The idea was originally conceived by S.P. and M. Cordenonsi. R.C., M. Cordenonsi, and S.P. wrote the paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Patient sample collection
  - Mice
  - Cell cultures
- METHOD DETAILS
  - scRNA-seq of human samples
  - Analysis of scRNA-seq data from human mammary gland and primary and metastatic breast cancer samples
  - Definition of Fetal Mammary Gland signatures
  - Functional enrichment analysis
  - Analysis of scRNA-seq data from PyMT and HER2 BC models
  - Definition and quantification of gene signatures from single-cell RNA-seq data
  - Analysis of the FMM signature expression in breast cancer primary-metastasis pairs
  - Survival analysis of breast cancer primary tumors
  - 2D topological analyses of human TNBC samples
  - Spatial transcriptomics of human and mouse BC tissues
  - Derivation of patient-derived organoids from metastasis tissues
  - Generation and analysis of ChIP-seq data from brain metastasis PDOs
  - Derivation of epithelial mammary gland cells from human healthy tissue
  - 3D cell culture in collagen 1 gels
  - Mammosphere assay
  - Plasmids and lentiviral transduction
  - Sorting of fluorescent cells
  - Cell editing with CRISPR-cas9
  - Western Blot
  - qRT-PCR
  - Wound healing assay
  - Analysis of bulk RNA-seq of mouse and human cell lines
  - *In vivo* metastasis and primary tumor formation assays
  - Immunohistochemistry on paraffin sections
  - 3D reconstruction of metastasis based on cytokeratin IHC staining
  - *In vivo* bioluminescence imaging
  - Imaging and 3D reconstruction
  - Image analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2026.03.009.

Received: December 3, 2024
Revised: November 17, 2025
Accepted: March 6, 2026

REFERENCES

1. Gerstberger, S., Jiang, Q., and Ganesh, K. (2023). Metastasis. Cell *186*, 1564–1579. https://doi.org/10.1016/j.cell.2023.03.003.
2. Garcia-Recio, S., Hinoue, T., Wheeler, G.L., Kelly, B.J., Garrido-Castro, A.C., Pascual, T., De Cubas, A.A., Xia, Y., Felsheim, B.M., McClure, M.B., et al. (2023). Multiomics in primary and metastatic breast tumors from the AURORA US network finds microenvironment and epigenetic drivers of metastasis. Nat. Cancer *4*, 128–147. https://doi.org/10.1038/s43018-022-00491-x.
3. Nolan, E., Lindeman, G.J., and Visvader, J.E. (2023). Deciphering breast cancer: from biology to the clinic. Cell *186*, 1708–1728. https://doi.org/10.1016/j.cell.2023.01.040.
4. Pal, B., Chen, Y., Vaillant, F., Capaldo, B.D., Joyce, R., Song, X., Bryant, V.L., Penington, J.S., Di Stefano, L., Tubau Ribera, N., et al. (2021). A

single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. EMBO J. *40*, e107333. https://doi.org/10.15252/embj.2020107333.

5. Wu, S.Z., Al-Eryani, G., Roden, D.L., Junankar, S., Harvey, K., Andersson, A., Thennavan, A., Wang, C., Torpy, J.R., Bartonicek, N., et al. (2021). A single-cell and spatially resolved atlas of human breast cancers. Nat. Genet. *53*, 1334–1347. https://doi.org/10.1038/s41588-021-00911-1.

6. Gonzalez, H., Mei, W., Robles, I., Hagerling, C., Allen, B.M., Hauge Okholm, T.L., Nanjaraj, A., Verbeek, T., Kalavacherla, S., van Gogh, M., et al. (2022). Cellular architecture of human brain metastases. Cell *185*, 729–745.e20. https://doi.org/10.1016/j.cell.2021.12.043.

7. Klughammer, J., Abravanel, D.L., Segerstolpe, Å., Blosser, T.R., Goltsev, Y., Cui, Y., Goodwin, D.R., Sinha, A., Ashenberg, O., Slyper, M., et al. (2024). A multi-modal single-cell and spatial expression map of metastatic breast cancer biopsies across clinicopathological features. Nat. Med. *30*, 3236–3249. https://doi.org/10.1038/s41591-024-03215-z.

8. Slyper, M., Porter, C.B.M., Ashenberg, O., Waldman, J., Drokhlyansky, E., Wakiro, I., Smillie, C., Smith-Rosario, G., Wu, J., Dionne, D., et al. (2020). A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. Nat. Med. *26*, 792–802. https://doi.org/10.1038/s41591-020-0844-1.

9. Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. J. Clin. Oncol. *27*, 1160–1167. https://doi.org/10.1200/JCO.2008.18.1370.

10. Squair, J.W., Gautier, M., Kathe, C., Anderson, M.A., James, N.D., Hutson, T.H., Hudelle, R., Qaiser, T., Matson, K.J.E., Barraud, Q., et al. (2021). Confronting false discoveries in single-cell differential expression. Nat. Commun. *12*, 5692. https://doi.org/10.1038/s41467-021-25960-2.

11. Kiemen, A.L., Braxton, A.M., Grahn, M.P., Han, K.S., Babu, J.M., Reichel, R., Jiang, A.C., Kim, B., Hsu, J., Amoa, F., et al. (2022). CODA: quantitative 3D reconstruction of large tissues at cellular resolution. Nat. Methods *19*, 1490–1499. https://doi.org/10.1038/s41592-022-01650-9.

12. Affolter, M., Zeller, R., and Caussinus, E. (2009). Tissue remodelling through branching morphogenesis. Nat. Rev. Mol. Cell Biol. *10*, 831–842. https://doi.org/10.1038/nrm2797.

13. Wallingford, J.B. (2012). Planar cell polarity and the developmental control of cell behavior in vertebrate embryos. Annu. Rev. Cell Dev. Biol. *28*, 627–653. https://doi.org/10.1146/annurev-cellbio-092910-154208.

14. Kim, J., Lee, S.W., and Park, K. (2021). CXCR4 Regulates Temporal Differentiation via PRC1 Complex in Organogenesis of Epithelial Glands. Int. J. Mol. Sci. *22*, 619. https://doi.org/10.3390/ijms22020619.

15. Nichol, D., and Stuhlmann, H. (2012). EGFL7: a unique angiogenic signaling factor in vascular development and disease. Blood *119*, 1345–1352. https://doi.org/10.1182/blood-2011-10-322446.

16. Ting, S.B., Caddy, J., Hislop, N., Wilanowski, T., Auden, A., Zhao, L.L., Ellis, S., Kaur, P., Uchida, Y., Holleran, W.M., et al. (2005). A homolog of Drosophila grainy head is essential for epidermal integrity in mice. Science *308*, 411–413. https://doi.org/10.1126/science.1107511.

17. Gu, B., Sun, P., Yuan, Y., Moraes, R.C., Li, A., Teng, A., Agrawal, A., Rhéaume, C., Bilanchone, V., Veltmaat, J.M., et al. (2009). Pygo2 expands mammary progenitor cells by facilitating histone H3 K4 methylation. J. Cell Biol. *185*, 811–826. https://doi.org/10.1083/jcb.200810133.

18. Liu, Y., and Guo, W. (2021). SOX factors as cell-state regulators in the mammary gland and breast cancer. Semin. Cell Dev. Biol. *114*, 126–133. https://doi.org/10.1016/j.semcdb.2021.01.002.

19. Lawson, D.A., Bhakta, N.R., Kessenbrock, K., Prummel, K.D., Yu, Y., Takai, K., Zhou, A., Eyob, H., Balakrishnan, S., Wang, C.Y., et al. (2015). Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. Nature *526*, 131–135. https://doi.org/10.1038/nature15260.

20. Fischer, K.R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S.T.C., Choi, H., El Rayes, T., Ryu, S., Troeger, J., et al. (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature *527*, 472–476. https://doi.org/10.1038/nature15748.

21. Aiello, N.M., and Kang, Y. (2019). Context-dependent EMT programs in cancer metastasis. J. Exp. Med. *216*, 1016–1026. https://doi.org/10.1084/jem.20181827.

22. Saini, M., Schmidleitner, L., Moreno, H.D., Donato, E., Falcone, M., Bartsch, J.M., Klein, C., Vogel, V., Würth, R., Pfarr, N., et al. (2023). Resistance to mesenchymal reprogramming sustains clonal propagation in metastatic breast cancer. Cell Rep. *42*, 112533. https://doi.org/10.1016/j.celrep.2023.112533.

23. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352. https://doi.org/10.1038/nature10983.

24. Leon-Ferre, R.A., and Goetz, M.P. (2023). Advances in systemic therapies for triple negative breast cancer. BMJ *381*, e071674. https://doi.org/10.1136/bmj-2022-071674.

25. He, S., Bhatt, R., Brown, C., Brown, E.A., Buhr, D.L., Chantranuvatana, K., Danaher, P., Dunaway, D., Garrison, R.G., Geiss, G., et al. (2022). High-plex imaging of RNA and proteins at subcellular resolution in fixed tissue by spatial molecular imaging. Nat. Biotechnol. *40*, 1794–1806. https://doi.org/10.1038/s41587-022-01483-z.

26. Della Chiara, G., Gervasoni, F., Fakiola, M., Godano, C., D'Oria, C., Azzolin, L., Bonnal, R.J.P., Moreni, G., Drufuca, L., Rossetti, G., et al. (2021). Epigenomic landscape of human colorectal cancer unveils an aberrant core of pan-cancer enhancers orchestrated by YAP/TAZ. Nat. Commun. *12*, 2340. https://doi.org/10.1038/s41467-021-22544-y.

27. Nasser, J., Bergman, D.T., Fulco, C.P., Guckelberger, P., Doughty, B.R., Patwardhan, T.A., Jones, T.R., Nguyen, T.H., Ulirsch, J.C., Lekschas, F., et al. (2021). Genome-wide enhancer maps link risk variants to disease genes. Nature *593*, 238–243. https://doi.org/10.1038/s41586-021-03446-x.

28. Herriges, J.C., Verheyden, J.M., Zhang, Z., Sui, P., Zhang, Y., Anderson, M.J., Swing, D.A., Zhang, Y., Lewandoski, M., and Sun, X. (2015). FGF-Regulated ETV Transcription Factors Control FGF-SHH Feedback Loop in Lung Branching. Dev. Cell *35*, 322–332. https://doi.org/10.1016/j.devcel.2015.10.006.

29. Linnemann, J.R., Miura, H., Meixner, L.K., Irmler, M., Kloos, U.J., Hirschi, B., Bartsch, H.S., Sass, S., Beckers, J., Theis, F.J., et al. (2015). Quantification of regenerative potential in primary human mammary epithelial cells. Development *142*, 3239–3251. https://doi.org/10.1242/dev.123554.

30. Rauner, G., Traugh, N.C., Trepicchio, C.J., Parrish, M.E., Mushayandebvu, K., and Kuperwasser, C. (2025). Single-cell organogenesis captures complex breast tissue formation in three dimensions. Development *152*, dev204813. https://doi.org/10.1242/dev.204813.

31. Morris, V.L., Tuck, A.B., Wilson, S.M., Percy, D., and Chambers, A.F. (1993). Tumor progression and metastasis in murine D2 hyperplastic alveolar nodule mammary tumor cell lines. Clin. Exp. Metastasis *11*, 103–112. https://doi.org/10.1007/BF00880071.

32. Danaher, P., Zhao, E., Yang, Z., Ross, D., Gregory, M., Reitz, Z., Kim, T.K., Baxter, S., Jackson, S., He, S., et al. (2022). Insitutype: likelihood-based cell typing for single cell spatial transcriptomics. Preprint at bioRxiv, 2022.2010.2019.512902. https://doi.org/10.1101/2022.10.19.512902.

33. Jin, S., Plikus, M.V., and Nie, Q. (2024). CellChat for systematic analysis of cell-cell communication from single-cell transcriptomics. Nat. Protoc. *20*, 180–219. https://doi.org/10.1038/s41596-024-01045-4.

34. Sootome, H., Fujita, H., Ito, K., Ochiiwa, H., Fujioka, Y., Ito, K., Miura, A., Sagara, T., Ito, S., Ohsawa, H., et al. (2020). Futibatinib Is a Novel Irreversible FGFR 1–4 Inhibitor That Shows Selective Antitumor Activity against FGFR-Deregulated Tumors. Cancer Res. *80*, 4986–4997. https://doi.org/10.1158/0008-5472.CAN-19-2568.

35. Hannezo, E., Scheele, C.L.G.J., Moad, M., Drogo, N., Heer, R., Sampogna, R.V., van Rheenen, J., and Simons, B.D. (2017). A Unifying Theory of Branching Morphogenesis. Cell *171*, 242–255.e27. https://doi.org/10.1016/j.cell.2017.08.026.

36. Bordeu, I., Chatzeli, L., and Simons, B.D. (2023). Inflationary theory of branching morphogenesis in the mouse salivary gland. Nat. Commun. *14*, 3422. https://doi.org/10.1038/s41467-023-39124-x.

37. Montagner, M., Bhome, R., Hooper, S., Chakravarty, P., Qin, X., Sufi, J., Bhargava, A., Ratcliffe, C.D.H., Naito, Y., Pocaterra, A., et al. (2020). Crosstalk with lung epithelial cells regulates Sfrp2-mediated latency in breast cancer dissemination. Nat. Cell Biol. *22*, 289–296. https://doi.org/10.1038/s41556-020-0474-3.

38. Sagara, A., Igarashi, K., Otsuka, M., Kodama, A., Yamashita, M., Sugiura, R., Karasawa, T., Arakawa, K., Narita, M., Kuzumaki, N., et al. (2017). Endocan as a prognostic biomarker of triple-negative breast cancer. Breast Cancer Res. Treat. *161*, 269–278. https://doi.org/10.1007/s10549-016-4057-8.

39. Castellan, M., Guarnieri, A., Fujimura, A., Zanconato, F., Battilana, G., Panciera, T., Sladitschek, H.L., Contessotto, P., Citron, A., Grilli, A., et al. (2021). Single-cell analyses reveal YAP/TAZ as regulators of stemness and cell plasticity in Glioblastoma. Nat. Cancer *2*, 174–188. https://doi.org/10.1038/s43018-020-00150-z.

40. Dieci, M.V., Tsvetkova, V., Griguolo, G., Miglietta, F., Tasca, G., Giorgi, C.A., Cumerlato, E., Massa, D., Lo Mele, M., Orvieto, E., et al. (2020). Integration of tumour infiltrating lymphocytes, programmed cell-death ligand-1, CD8 and FOXP3 in prognostic models for triple-negative breast cancer: Analysis of 244 stage I-III patients treated with standard therapy. Eur. J. Cancer *136*, 7–15. https://doi.org/10.1016/j.ejca.2020.05.014.

41. Tolaney, S.M., Garrett-Mayer, E., White, J., Blinder, V.S., Foster, J.C., Amiri-Kordestani, L., Hwang, E.S., Bliss, J.M., Rakovitch, E., Perlmutter, J., et al. (2021). Updated Standardized Definitions for Efficacy End Points (STEEP) in Adjuvant Breast Cancer Clinical Trials: STEEP Version 2.0. J. Clin. Oncol. *39*, 2720–2731. https://doi.org/10.1200/JCO.20.03613.

42. Cordenonsi, M., Zanconato, F., Azzolin, L., Forcato, M., Rosato, A., Frasson, C., Inui, M., Montagner, M., Parenti, A.R., Poletti, A., et al. (2011). The Hippo transducer TAZ confers cancer stem cell-related traits on breast cancer cells. Cell *147*, 759–772. https://doi.org/10.1016/j.cell.2011.09.048.

43. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888–1902.e21. https://doi.org/10.1016/j.cell.2019.05.031.

44. Grandi, F., Caroli, J., Romano, O., Marchionni, M., Forcato, M., and Bicciato, S. (2022). popsicleR: A R Package for Pre-processing and Quality Control Analysis of Single Cell RNA-seq Data. J. Mol. Biol. *434*, 167560. https://doi.org/10.1016/j.jmb.2022.167560.

45. Wolock, S.L., Lopez, R., and Klein, A.M. (2019). Scrublet: Computational Identification of Cell Doublets in Single-Cell Transcriptomic Data. Cell Syst. *8*, 281–291.e9. https://doi.org/10.1016/j.cels.2018.11.005.

46. Aran, D., Looney, A.P., Liu, L., Wu, E., Fong, V., Hsu, A., Chak, S., Naikawadi, R.P., Wolters, P.J., Abate, A.R., et al. (2019). Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. Nat. Immunol. *20*, 163–172. https://doi.org/10.1038/s41590-018-0276-y.

47. Puram, S.V., Tirosh, I., Parikh, A.S., Patel, A.P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C.L., Mroz, E.A., Emerick, K.S., et al. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell *171*, 1611–1624.e24. https://doi.org/10.1016/j.cell.2017.10.044.

48. Chen, Y., Lun, A.T.L., and Smyth, G.K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. F1000Res *5*, 1438. https://doi.org/10.12688/f1000research.8987.2.

49. Giraddi, R.R., Chung, C.Y., Heinz, R.E., Balcioglu, O., Novotny, M., Trejo, C.L., Dravis, C., Hagos, B.M., Mehrabad, E.M., Rodewald, L.W., et al. (2018). Single-Cell Transcriptomes Distinguish Stem Cell State Changes and Lineage Specification Programs in Early Mammary Gland Development. Cell Rep. *24*, 1653–1666.e7. https://doi.org/10.1016/j.celrep.2018.07.025.

50. Pal, B., Chen, Y., Vaillant, F., Jamieson, P., Gordon, L., Rios, A.C., Wilcox, S., Fu, N., Liu, K.H., Jackling, F.C., et al. (2017). Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. Nat. Commun. *8*, 1627. https://doi.org/10.1038/s41467-017-01560-x.

51. Sun, H., Zhou, Y., Fei, L., Chen, H., and Guo, G. (2019). scMCA: A Tool to Define Mouse Cell Types Based on Single-Cell Digital Expression. Methods Mol. Biol. *1935*, 91–96. https://doi.org/10.1007/978-1-4939-9057-3_6.

52. Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinform. *14*, 128. https://doi.org/10.1186/1471-2105-14-128.

53. Dravis, C., Chung, C.Y., Lytle, N.K., Herrera-Valdez, J., Luna, G., Trejo, C.L., Reya, T., and Wahl, G.M. (2018). Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. Cancer Cell *34*, 466–482.e6. https://doi.org/10.1016/j.ccell.2018.08.001.

54. Dai, X., Wang, X., Yang, C., Huang, M., Zhou, Z., Qu, Y., Cui, X., Liu, R., and Chen, C. (2022). Human fibroblasts facilitate the generation of iPSCs-derived mammary-like organoids. Stem Cell Res. Ther. *13*, 377. https://doi.org/10.1186/s13287-022-03023-7.

55. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

56. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.

57. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

58. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

59. Reynolds, G., Vegh, P., Fletcher, J., Poyner, E.F.M., Stephenson, E., Goh, I., Botting, R.A., Huang, N., Olabi, B., Dubois, A., et al. (2021). Developmental cell programs are co-opted in inflammatory skin disease. Science *371*, eaba6500. https://doi.org/10.1126/science.aba6500.

60. Liu, N., Matsumura, H., Kato, T., Ichinose, S., Takada, A., Namiki, T., Asakawa, K., Morinaga, H., Mohri, Y., De Arcangelis, A., et al. (2019). Stem cell competition orchestrates skin homeostasis and ageing. Nature *568*, 344–350. https://doi.org/10.1038/s41586-019-1085-7.

61. Wang, S., Drummond, M.L., Guerrero-Juarez, C.F., Tarapore, E., MacLean, A.L., Stabell, A.R., Wu, S.C., Gutierrez, G., That, B.T., Benavente, C.A., et al. (2020). Single cell transcriptomics of human epidermis identifies basal stem cell transition states. Nat. Commun. *11*, 4239. https://doi.org/10.1038/s41467-020-18075-7.

62. Xu, Q., Halle, L., Hediyeh-Zadeh, S., Kuijs, M., Riedweg, R., Kilik, U., Recaldin, T., Yu, Q., Rall, I., Frum, T., et al. (2025). An integrated transcriptomic cell atlas of human endoderm-derived organoids. Nat. Genet. *57*, 1201–1212. https://doi.org/10.1038/s41588-025-02182-6.

63. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.

64. Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science *352*, 189–196. https://doi.org/10.1126/science.aad0501.

65. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. Sci. Rep. *7*, 16878. https://doi.org/10.1038/s41598-017-17204-5.

66. Schmidt, U., Weigert, M., Broaddus, C., and Myers, G. (2018). Cell Detection with Star-Convex Polygons. Lect Notes. Comput. Sci. *11071*, 265–273. https://doi.org/10.1007/978-3-030-00934-2_30.

67. Dekkers, J.F., van Vliet, E.J., Sachs, N., Rosenbluth, J.M., Kopper, O., Rebel, H.G., Wehrens, E.J., Piani, C., Visvader, J.E., Verissimo, C.S., et al. (2021). Long-term culture, genetic manipulation and xenotransplantation of human normal and breast cancer organoids. Nat. Protoc. *16*, 1936–1965. https://doi.org/10.1038/s41596-020-00474-1.

68. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

69. Ernst, J., and Kellis, M. (2012). ChromHMM: automating chromatin-state discovery and characterization. Nat. Methods *9*, 215–216. https://doi.org/10.1038/nmeth.1906.

70. Roadmap; Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., et al. (2015). Integrative analysis of 111 reference human epigenomes. Nature *518*, 317–330. https://doi.org/10.1038/nature14248.

71. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. Nature *481*, 389–393. https://doi.org/10.1038/nature10730.

72. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

73. Miller, D.H., Sokol, E.S., and Gupta, P.B. (2017). 3D Primary Culture Model to Study Human Mammary Development. Methods Mol. Biol. *1612*, 139–147. https://doi.org/10.1007/978-1-4939-7021-6_10.

74. Vitari, A.C., Leong, K.G., Newton, K., Yee, C., O'Rourke, K., Liu, J., Phu, L., Vij, R., Ferrando, R., Couto, S.S., et al. (2011). COP1 is a tumour suppressor that causes degradation of ETS transcription factors. Nature *474*, 403–406. https://doi.org/10.1038/nature10005.

75. Panciera, T., Azzolin, L., Fujimura, A., Di Biagio, D., Frasson, C., Bresolin, S., Soligo, S., Basso, G., Bicciato, S., Rosato, A., et al. (2016). Induction of Expandable Tissue-Specific Stem/Progenitor Cells through Transient Expression of YAP/TAZ. Cell Stem Cell *19*, 725–737. https://doi.org/10.1016/j.stem.2016.08.009.

76. Kubota, S.I., Takahashi, K., Nishida, J., Morishita, Y., Ehata, S., Tainaka, K., Miyazono, K., and Ueda, H.R. (2017). Whole-Body Profiling of Cancer Metastasis with Single-Cell Resolution. Cell Rep. *20*, 236–250. https://doi.org/10.1016/j.celrep.2017.06.010.

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
H3K27Ac	Abcam	Cat# ab4729, RRID: AB_2118291
H3K4Me1	Diagenode	Cat# C15410194, RRID: AB_2637078
H3K4Me3	Millipore	Cat# 07-473, RRID: AB_1977252
H3K36Me3	Diagenode	Cat# C15410192, RRID: AB_2744515
H3K27Me3	Millipore	Cat# 07-449, RRID: AB_310624
H3K9Me3	Abcam	Cat# ab8898, RRID: AB_306848
ETV1	Thermo Fisher Scientific	Cat# PA5-77975, RRID: AB_2735767
ETV5	Thermo Fisher Scientific	Cat# PA5-30023, RRID: AB_2547497
ETV4	Santa Cruz Biotechnology	Cat# sc-113, RRID: AB_2100996
FGFR2	Cell Signaling Technologies	Cat# 23328, RRID: AB_2798862
GAPDH	Millipore	Cat# MAB347, RRID: AB_94881
YAP1	Proteintech	Cat# 13584-1-AP, RRID: AB_2218915
CK8	Abcam	Cat# ab53280, RRID: AB_869901
MNF116 (pan-cytokeratin)	Agilent	Cat# M0821, RRID: AB_2858276
Ki67	Novus	Cat# NB600-1252, RRID: AB_2142376
EGFP vioR667	Miltenyi Biotec	Cat# 130-131-646, RRID: AB_3664460
CD31 vioR667	Miltenyi Biotec	Cat# 130-128-736, RRID: AB_2904940
Biological samples		
Fresh and FFPE human breast cancer tissues, patient-derived organoids	This study	N/A
Chemicals, peptides, and recombinant proteins		
Collagenase I	Invitrogen	Cat# 17100-017
Hyaluronidase	Sigma	Cat# H3506
RNase free DNase set	Qiagen	Cat# 79254
TrypLE	Gibco	Cat# 12604-013
Dispase I	Gibco	Cat# 17105-041
BD Pharm Lyse Lysing Buffer	BD Biosciences	Cat# 555899
Matrigel GFR Phenol Red-Free	Corning	Cat# 356231
Cell Recovery solution	Corning	Cat# 354253
Protein G Dynabeads	ThermoFisher	Cat# 10003D
Proteinase K	NEB	Cat# P8107S
SPRIselect Beads	Beckman Coulter	Cat# B23318
Bovine pituitary extract (BPE)	Thermo Scientific	Cat# 13028014
Human EGF	Peprotech	Cat# AF-100-15
Forskolin	Merck	Cat# F6886
Insulin	Merck	Cat# 91077C
Hydrocortisone	Merck	Cat# H0396
ultra-low-adhesion plates for collagen droplets formation	Greiner	Cat# 662102
Tubes containing beads for protein extraction	Fisher	Cat# 15565799
Polyethyleneimine	Sigma	Cat# P3143
Phalloidin AF647	ThermoFisher	Cat# A22287

(*Continued on next page*)

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
SYTO16	ThermoFisher	Cat# S7578
Hoechst33342	ThermoFisher	Cat# 62249
FGF10	Peprotech	Cat# 100-26
bFGF	Peprotech	Cat# 100-18
Cas9-GFP	IDT	Cat# 10008100
TracrRNA	IDT	Cat# 1072533
Electroporator Enhancer	IDT	Cat# 1075916
Gelatin	Sigma	Cat# G6144
Futibatinib	MedChemExpress	Cat# HY-100818
IVISbrite D-luciferin	Revvity	Cat# 122799
Propidium Iodide	ThermoFisher	Cat# P3566
Silicon Oil	Lesker	Cat# 704BB
Mineral Oil	Sigma	Cat# M8410
N-buthyldiethanolamine	Tokyo Chemical	Cat# B0725
Antypirin	Tokyo Chemical	Cat# D1876
Nicotinamide	Tokyo Chemical	Cat# N0078
Critical commercial assays		
Tumor dissociation kit	Miltenyi	Cat# 130-095-929
Dead cell removal kit	Miltenyi	Cat# 130-090-101
RNAscope 2.5HD Duplex Assay	ACD bio	Cat# 322500
Mouse CosMX Universal cell Characterization RNA panel	Nanostring	Cat# CMX-M-USCP-1KP-R
Human CosMX Universal cell Characterization RNA panel	Nanostring	Cat# CMX-H-USCP-1KP-R
CosMX RNA Imaging Tray 1000plex 4-slide run	Nanostring	Cat# CMX-RIMGT-1KP-R
Watchmaker DNA Library Prep Kit for double-stranded DNA	Watchmaker Genomics	Cat# 7K0102-096
Collagen I kit	Fujifilm	Cat# 638-00781
BCA Protein Assay Kit	Thermo Scientific	Cat# A55865
SE Cell Line 4D-Nucleofector® X Kit S	Lonza	Cat# V4XC-1032
P3 Primary Cell 4D-Nucleofector® X Kit S	Lonza	Cat# V4XP-3032
MACS clearing kit	Miltenyi	Cat# 130-126-719
Deposited data		
scRNAseq, bulk RNA-seq, ChiP-seq	This paper, GEO: GSE281492	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE281492
scRNAseq of additional primary tumors (*#9281, #9282, #9289, #9290, #9291, #9293, #9294, #9299, #9306, #9315, #9317*)	Previously published,[4] GEO: GSE161529	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE161529
scRNA-seq from liver BC metastasis (BLM285, BLM963)	Previously published[8]	https://duos.broadinstitute.org
METABRIC dataset	Previously published,[23] EGAD00010000210	http://www.ebi.ac.uk/ega/
Experimental models: Cell lines		
Human breast cancer cells MDA-MB-231	ATCC	Cat# HTB-26; RRID: CVCL_0062
Human breast cancer cells HCC1954	ATCC	Cat# CRL-2338; RRID: CVCL_1259
Human fibrosarcoma cells HOS143-B	ATCC	Cat# VR-295
Human osteosarcoma cells HT1080	ATCC	Cat# CCL-121; RRID: CVCL_0317
Mouse mammary cancer cells TUBO	Merck	Cat# SCC222; RRID: CVCL_2A33

(*Continued on next page*)

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse glioma cells CT-2A	Merck	Cat# SCC194; RRID: CVCL_ZJ44
Mouse mammary cancer cells 4T1	ATCC	Cat# CRL-2539; RRID: CVCL_0125
Mouse mammary cancer cells D2.OR	Montagner et al.[37]	N/A
Human breast cancer cells Brain-trained MDA-MB-231	Y. Kato[38]	N/A
Experimental models: Organisms/strains		
Mouse: FVB/N-Tg (MMTV-PyMT)634Mul/J	The Jackson Laboratory	Cat# 002374; RRID: IMSR_JAX:002374
Mouse: B6.FVB-Tg(MMTV-PyMT)634Mul/LellJ	The Jackson Laboratory	Cat# 022974; RRID: IMSR_JAX:022974
Mouse: FVB/N-Tg(MMTVneu)202Mul/J	The Jackson Laboratory	Cat# 002376; RRID: IMSR_JAX:002376
Oligonucleotides		
sgRNAs against ETV1, ETV4, ETV5 and FGFR2, see Table S1M	This study	N/A
CTRL sgRNA (control crRNA)	IDT	Cat# 1072544
PCR primers, see Table S1N	This paper	N/A
Recombinant DNA		
GFP/luciferase lentiviral construct	From A. Rosato (Castellan et al.[39])	N/A
pLV-hPGK-hETV1AAA-FLAG-T2A-EGFP	Vectorbuilder	Cat# VB231006-1546ayh
pLV-hPGK-EGFP	Vectorbuilder	Cat# VB900088-2268pnh
pLV-hPGK-mScarlet3	Vectorbuilder	Cat# VB250305-1042hus
Software and algorithms		
Aivia (14.1.0)	Leica microsystems	https://www.leica-microsystems.com/products/microscope-software/p/aivia/?utm_source=google&utm_medium=cpc&utm_term=aivia&utm_campaign=23-EM-LSR-L3-AIVI-GOOG-PP-EU-Aivia-Search&gad_source=1&gad_campaignid=19434775813&gclid=EAIaIQobChMI0JmG2ub9kAMV2a2DBx1gTDb-EAAYASAAEgI7sPD_BwE
Fiji (2.9.0)	ImageJ2	https://imagej.net/software/fiji/downloads
Qupath (0.5.0)	QuPath	https://qupath.github.io/
Imaris viewer (10.2.0)	Oxford Instruments Group	https://imaris.oxinst.com/imaris-viewer
LasX (4.1.0.23081)	Leica microsystems	https://www.leica-microsystems.com/products/microscope-software/p/leica-las-x-ls/
R (3.6.3 and 4.4.3)	R Core Team	https://www.r-project.org/
Prism (10.6.1)	GraphPad Software	https://www.graphpad.com/

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Patient sample collection

Clinical samples were merged under the Master Observational Trial METAMECH (NCT04625023) (IFOM-CPO007/2019/PO006). Fresh surgical metastatic BC samples (see Table S1A) were collected from: Azienda Ospedaliera di Padova, with informed consent according to University of Padua's institutional guidelines and the Azienda Ospedaliera di Padova Ethics Committee (CESC); Regina Elena National Cancer Institute (Rome), with informed consent, as approved by the ethics committee and institutional review board of Regina Elena National Cancer Institute (Rome). Primary breast cancer samples were obtained from Istituto Nazionale dei Tumori, Milan, and brain metastatic samples were also from Istituto Neurologico "Carlo Besta", Milan, each following ethical approval from their Institutional Review Boards and acquisition of informed consent from all patients. Normal mammary gland samples were from anonymized healthy women undergoing reduction mastoplasty collected from Azienda Ospedaliera di Padova, with informed consent as above, in the context of the CancerOrg protocol [0045505 (3989/AO/16)].

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

Archival FFPE samples were from a clinical cohort of 132 patients diagnosed with TNBC (non-metastatic at diagnosis) between 2000 and 2017.[40] All patients underwent surgery of the primary tumor and received adjuvant systemic chemotherapy, and they were referred to the Istituto Oncologico Veneto in Padua. TNBC was defined as estrogen and progesterone receptor expression in <10% of tumor cells by immunohistochemistry, and HER2 0/1+ by immunohistochemistry and/or fluorescent in situ hybridization not amplified. Clinicopathological, treatment and follow-up data were collected from medical records and are summarized in Table S1H. Median age at diagnosis was 59 years (min 26 - max 83).

We used the distant relapse-free interval (DRFI) as the survival endpoint to compare patients with high versus low topo-morphometric parameters. DRFI was defined, according to the Standardized Definitions for Efficacy End Points (STEEP[41]) in Adjuvant Breast Cancer Clinical Trials 2.0, as the time from surgery to distant relapse (or death from breast cancer in the absence of information on the date of distant relapse). Patients without event were censored at the time of last follow up/non breast cancer-related death. At a median follow up of 8.3 years (95% CI 7.0-9.6), 24 patients experienced a DRFI event, all DRFI events were distant relapses.

Human TNBC surgical samples were fixed and archived as FFPE tissues according to institutional protocols. Histological staining and molecular analyses were performed on 5μm-thick tissue sections. All the samples in the cohort were processed for YAP IHC (Proteintech, 13584-1-AP, following the protocol used in Ref.[42]) to facilitate histopathological definition of tumor areas; 8 of these archival samples were analysed by spatial transcriptomics (Table S1I).

All patients were female of European ancestry. Data on participants' race, and ethnicity were not collected, as these variables were not included in the informed consent process and could not be obtained under the applicable ethical guidance; this should be considered a limitation when interpreting the generalizability of the findings.

Mice

Animal experiments were performed adhering to our institutional guidelines as approved by the OPBA (University of Padua) and the Ministry of Health of Italy. The housing conditions comprised a diet with 28% protein. A maximum of five adult mice weighing up to 20 g were housed in a single cage, maintaining the ambient temperature at 19–23 °C, the humidity at 55% ± 10% and a 12-h light/12-h dark cycle. The well-being of experimental animals was monitored by a licensed veterinarian, and mice were euthanized as per veterinary advice as soon as they reached human endpoints, including signs of suffering, general aspect and behavior, maximum tumor size, clinical signs of poor health associated to tumor or metastases growth.

Mouse models of metastatic BC, MMTV-PyMT (FVB, 002374; C57BL/6J, 022974) and MMTV-Neu (FVB, 002376) were purchased from Jackson Laboratories.

Immunocompromized (NOD/SCID and NSG) and BALB/C mice were purchased from Envigo; for injection of tumor cells, 8-12 weeks-old mice were used.

Cell cultures

MDA-MB-231, HCC1954, HOS143-B, HT1080 and 4T1 cells were purchased from ATCC. CT-2A (SCC194) and TUBO (SCC222) cells were purchased from Merck. D2.OR cells were a gift from M. Montagner.[37] Brain-trained MDA-MB-231 were from Y. Kato.[38]

4T1, MDA-MB-231, HCC1954 and TUBO cells were cultured in DMEM/F12 with 10%FBS, 2% PS and 1% L-glutamine. D2.OR, HT1080, HOS143B and CT-2A cells were cultured in DMEM with 10%FBS, 2% PS and 1% L-glutamine.

All cell lines were routinely tested for mycoplasma contamination and confirmed negative. All human cell lines were authenticated by BMR genomics.

METHOD DETAILS

scRNA-seq of human samples

After surgery, human samples were preserved in RPMI supplemented with antibiotics. Tissue dissociation was performed according to either of the following procedures, as indicated in Table S1A.

In Protocol 1, tissues were minced into small pieces, discarding necrotic areas; the tissue was then incubated in hemolytic solution for 2 min at 4°C to remove red blood cells and washed in Wash Medium (Advanced DMEM:F12, 1mM Hepes, 1% GlutaMAX, 1% PenStrep). Tissue was digested in 600 U/ml Collagenase I, 200 U/ml Hyaluronidase and 1μg/m DNAse in Wash Medium at 37°C, for 30 min for tumor tissues and overnight for mammary gland tissues. After one wash in Wash Medium, digested tissue was further incubated in hemolytic solution for 5 min at 4°C to complete red blood cell lysis. Dissociation of tumor clumps to single cells was performed with TrypLE at 37°C, whereas dissociation of clumps of MG cells was performed using 0.25%Tripsin/EDTA in HBSS for 5 min, followed with addition of 10 U/ml dispase/1 mg/ml DNAse. Digestion was stop when, by visual inspection, most cells appeared as single cells. Cells were washed once in Wash Medium and twice in HBSS with 0.1% BSA, and filtered through a 40μm cell strainer.

In Protocol 2, tissues were trimmed into smaller pieces (~3-5mm of size) and digested in a gentleMACS C tube with 5ml of RPMI supplemented with antibiotics and enzymes A, H and R from Tumor Dissociation kit (Miltenyi). h_tumor_01 and 37C_NTDK_1 digestion programs from gentleMACS Octo Dissociator were used for primary BCs and brain metastases, respectively. Cell suspensions were filtered with a 100 μm cell strainer to remove undigested material and then pelleted at 300*g* for 10 min. When necessary, cells were incubated with BD Pharm Lyse Lysing Buffer (BD Biosciences) for 3 minutes on ice to lysate red blood cells.

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

For all samples, after dissociation, cells were counted and viability assessed by Trypan Blue exclusion. In case of excessive cell death, Dead Cell Removal Kit (Miltenyi) was used to improve the overall viability of the suspension. Only suspensions being at least 70% viable cells were used for this study. Concentration was then adjusted to 10^6 cells/ml in HBSS with 0.1% BSA.

Single-cell collection was performed using the Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1 and Chip G in a Chromium Controller (10X Genomics) targeting 10,000 cells, following manufacturer's instructions. Sequencing of single-indexed libraries was performed on Illumina HiSeq or NovaSeq platforms (2x150bp; 40 to 50 thousand reads per cell).

Analysis of scRNA-seq data from human mammary gland and primary and metastatic breast cancer samples

We analyzed single-cell RNA-seq data of the normal human mammary gland (hMG) from 3 healthy donors and of 14 primary and 10 metastatic human breast cancer (BC) samples. Tumor samples included 3 primary BCs (BC3, BC4, and BC16), 7 brain metastases (Breast Brain Metastasis: BBM0, BBM5, BBM7, BBM8, BBM9, BBM10, BBM13), and one skin metastasis (Breast Skin Metastasis: BSM1) obtained from our institutions (see Table S1A).

We also obtained single-cell RNA-seq data from 11 additional primary breast cancer tumors (samples #9281, #9282, #9289, #9290, #9291, #9293, #9294, #9299, #9306, #9315, #9317) from Pal et al.[4] and from 2 liver BC metastases (BLM285, BLM963) from Slyper et al.[8] The single-cell expression matrices of the BC tumors and raw reads for the liver BC metastases were downloaded from Gene Expression Omnibus: GSE161529 and the Broad DUOS (https://duos.broadinstitute.org), respectively.

For all samples, raw reads have been aligned to the human reference transcriptome (GRCh38, 10x Genomics pre-built reference version 3.0.0) and UMI counts calculated using the 10x Cell Ranger Count pipeline (version 3.1.0). The expression data were imported in R (version 3.6.3) and analyzed with *Seurat*[43] (version 3.1.5) and *popsicleR*[44] (version 0.2.1) R packages. For each sample, low quality cells were identified as outliers based on the number of genes, UMI counts, and the percentage of reads mapping to mitochondrial genes per cell, and these were subsequently discarded. Doublets were detected and removed using *Scrublet*[45] (version 0.2.1). Cell cycle scores were assigned to each cell using *Seurat*, with automated identification of cells in S or in G2/M phases of the cell cycle. Regression was performed on individual samples after evaluating the distributions of gene counts, UMI counts, mitochondrial gene percentages, and cell cycle scores. Following regression, principal component analysis (PCA) was used to reduce dataset dimensionality. Louvain clustering and UMAP visualization were performed on PCA reduced embedding, using Seurat *FindClusters* and *RunUMAP* functions respectively, after selecting an appropriate number of principal components for each sample. Upon the evaluation of an optimal clustering resolution, we used BlueprintEncode reference dataset from *SingleR*[46] (version 1.0.6) and known marker genes to annotate clusters as epithelial, endothelial, stromal, myeloid, T/NK cell, B/Plasma cell, and for metastases, cells from the host tissue (e.g., glial cells in the brain). Low-quality clusters were filtered out after inspecting the number of genes, UMI counts, and MALAT1 expression levels.

We merged the single-cell RNA-seq data of the 3 normal mammary glands with the Seurat *merge* function to establish a reference for the normal breast tissue. Using the Seurat integration strategy, we applied the *FindIntegrationAnchors* and *IntegrateData* functions to integrate all epithelial cell clusters from each normal mammary gland sample, creating a unified reference for normal epithelial breast cells. After integration, we identified and removed clusters containing contaminant cells (i.e., clusters with a mix of immune and epithelial cells). The cleaned epithelial clusters were then manually annotated as Basal, Luminal Type 1 (L1), and Luminal Type 2 (L2) cells based on the expression levels of known marker genes.[4] Genes differentially expressed between the Basal, L1, and L2 populations were identified using the Seurat *FindMarkers* function, with min.pct = 0.05, logfc.threshold = 0 and pseudocount.use = 0.01. Lists of genes specifically enriched in each MG epithelial cell population were then defined as those expressed in at least 10% of the cells (pct $\geq$0.1), and being significantly upregulated compared to both of the other two populations (logFC $\geq$ 1and adjusted p-value < 0.05).

We analyzed all human BC samples with *InferCNV*[47] (version 1.2.1) to profile copy number alterations in single cells from epithelial clusters, using all non-epithelial clusters as the reference for non-malignant diploid cells. Genome-wide copy number alterations were used to distinguish clusters of normal epithelial cells (Normal MG cell) from clusters of malignant cells (BC cell). Genetic amplifications of the *ERBB2* locus were inferred on BC cells using the CNV levels predicted by the Hidden Markov Model of *InferCNV*.

UMAP projections of Figures 1B, 1C, S1C and S1F were obtained after merging 139,628 cancer and TME cells from primary and metastatic BC tissues.

For each sample, only malignant clusters were retained and re-clustered using *Seurat*, and, to enhance the purity of tumor cells, clusters containing few cells of both tumor and non-tumor types were removed. The tumor cells were then used for differential expression analysis and the extraction of gene signatures.

Differential expression analysis between BC cells of metastases and primary tumors was performed using the *edgeR* package[48] (version 3.28.1) on pseudo-bulk data. Pseudo-bulk datasets were generated for each tumor sample by aggregating the expression matrix of tumor cells using the *aggregate.Matrix* function from the *Matrix.utils* R package (version 0.9.8), with 'sample' as the grouping variable and 'sum' as the aggregation function.

To classify primary and metastatic BC samples according to their PAM50 subtype, we first used the R function *rowMeans* to calculate in each BC cell of each sample the expression values of three signatures included in the PAM50 gene list,[9] identifying Luminal-like, Her2 and Basal-like BCs, respectively (listed in Table S1L). Average expression values of each signature were then calculated for each sample. Lesions were classified as Her2 if the level of the corresponding signature exceeded the average expression of that signature in all the samples, irrespectively of the expression of the Luminal-like and the Basal-like signature. As a control, all the

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

samples classified as Her2 BCs contained a large proportion (>80%) of BC cells bearing amplifications of the *ERBB2* locus. The remaining cases were classified either as Luminal-like or Basal-like if the corresponding signature was expressed more than the other two signatures.

For visualization purposes, we merged all BC samples using the Seurat *merge* function.

To create a reference matrix of expected cell type expression profiles for cell type annotation of human BC spatial transcriptomic data, we separately integrated immune, vascular and stromal clusters from both normal human mammary gland and BC samples using the *Seurat* integration strategy with the *FindIntegrationAnchors* and *IntegrateData* functions. Low-quality cells and contaminants were identified and removed. The resulting well-characterized subpopulations of the tumor microenvironment were then merged with the previously identified tumor cells and Host Tissue cells.

Definition of Fetal Mammary Gland signatures

Fetal Mammary Morphogenesis and Metastatic Trabecular Morphogenesis signatures

To define the Fetal Mammary Morphogenesis (FMM) signature we first downloaded single-cell RNA-seq data of the mouse MG development from Ref. [49] (all samples) and ref. [50] (puberty sample replicate 1; Gene Expression Omnibus: GSM2759554) as raw reads from Gene Expression Omnibus: GSE111113 and GSE103272, respectively. Raw reads have been aligned to the mouse reference transcriptome (mm10, version 3.0.0) and UMI counts calculated using the 10x Cell Ranger Count pipeline (version 3.1.0). Expression data were imported in R (version 3.6.3) and analyzed using *Seurat* (version 3.1.5) and *popsicleR* (version 0.2.1) R packages. For each sample, low quality cells were identified as outliers within the distribution of the number of genes, UMI counts, and percent of reads mapping on mitochondrial genes per cell, and subsequently discarded. Doublets were detected and removed using *Scrublet* (version 0.2.1). In each sample, cell cycle scores were assigned to each cell and regressed out before performing dimensional reduction. Following regression, principal component analysis (PCA) was used to reduce dataset dimensionality. Louvain clustering and UMAP visualization were performed on PCA reduced embedding, using Seurat *FindClusters* and *RunUMAP* functions respectively, after selecting an appropriate number of principal components for each sample. Upon the evaluation of an optimal clustering resolution, clusters were annotated with the *scMCA* R package[51] and according to known markers for different epithelial cell populations of the developing MG.[49] As expected, the adult MG contained three well-defined cell populations (Basal cells, Luminal Progenitors and Mature Luminal cells), whereas the Fetal (E16) MG contained cells in a less differentiated state.

Epithelial clusters of each sample were integrated using the *FindIntegrationAnchors* and *IntegrateData* functions of the *Seurat* integration strategy. The transcriptome from the whole Fetal MG was compared with the transcriptomes of each of the three cell populations of the adult MG using the *FindMarkers* function of the *Seurat* package with min.pct = 0.05, logfc.threshold = 0 and pseudocount.use = 0.01.

The Fetal Mammary Morphogenesis signature (Table S1D) was defined as a list of genes specifically enriched in the epithelial cells of the fetal mouse MG, depleted of genes related to cell proliferation and of genes enriched in the three cell populations of the adult human MG. For this, we first defined a list of genes expressed in at least 10% of the cells of the Fetal MG (pct $\geq$0.1) and in less than 10% (pct <0.1) of each of the three epithelial cell populations composing the adult mouse MG, also displaying a significant upregulation (logFC $\geq$ 0.5 and adjusted p-value < 0.05) in the Fetal MG compared to each of the epithelial cell populations of the adult mouse MG. This list was then depleted of genes related to cell proliferation, including those related to mitosis, cell cycle progression and DNA repair, identified through *Enrichr*.[52] Symbols of the resulting gene list were then converted into the corresponding human homologous genes using the HUGO Gene Nomenclature Committee (HGNC) database (https://www.genenames.org/cgi-bin/hcop) and then depleted of all genes enriched in the three human MG epithelial cell populations (see above).

The Metastatic Trabecular Morphogenesis signature (MTM) is the FMM depleted of genes not enriched in metastatic samples of our cohort, as determined by GSEA, and not significantly upregulated (FC > 1; FDR < 0.05) in the metastatic samples of both the AURORA US and the RAP datasets, resulting in a list of 523 out of 645 genes (Table S1G).

Published mouse Fetal Mammary Gland signature

For the alternative mouse Fetal MG signature (Dravis Fetal MaSCs), we used the fMaSC signature from Table S1 of Dravis et al.,[53] converted into the corresponding human homologous genes using the HUGO Gene Nomenclature Committee (HGNC) database (https://www.genenames.org/cgi-bin/hcop) and then depleted of genes related to cell proliferation (identified through *Enrichr*[52]), and of all genes enriched in the three human MG epithelial cell populations, as described for the FMM signature.

Fetal-like human Mammary Gland organoids signature

For the Fetal-like human Mammary Gland organoids signature (Fetal-like hMG organoids signature), we used RNA-seq data from a study in which mammary differentiation was induced in embryonic bodies obtained from human iPSCs.[54] That study profiled the differentiation process at three culture stages: 10 days, when organoids stop to express pluripotent markers and acquire fetal MG specification, as visualized by expression of fetal MG genes and co-expression of basal and luminal cell markers in the same cells; 20 days, when, under the stimulation of PTHrP, proliferation decreases and the basal and luminal cell lineages start to split; 30 days, representing fully-differentiated post-natal MG gland organoids, with no mixing of the basal and luminal lineages.

Bulk RNA-seq data were obtained from the Sequence Read Archive: SRP276825 (BioProject PRJNA656090). Read quality was assessed using FastQC (v0.11.9; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), and reads were aligned to the human reference genome (hg38) with STAR[55] (v2.7.9a). Gene-level counts were generated in R (v4.1.1) using the *featureCounts* function from the *Rsubread* package[56] (v2.8.1), based on the GENCODE v44 annotation. Raw counts were normalized to counts per million

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

mapped reads (CPM) with *edgeR* package[57] (v3.36.0). Genes with CPM > 1 in at least one sample were retained, and differential expression analysis was performed using the *exactTest* function in *edgeR*. The Fetal-like hMG organoids signature was defined as a list of genes specifically upregulated at 10 days of culture (Fetal-like MG stage, as explained above), depleted of genes related to cell proliferation and of genes enriched in the three cell populations of the adult human MG. For this, we first defined a list of genes significantly upregulated (logFC $\geq$ 3 and adjusted p-value < 0.0001) at 10 days compared to 20 and 30 days of culture. This list was then depleted of all genes enriched in the three human MG epithelial cell populations and genes related to cell proliferation, including those related to mitosis, cell cycle progression and DNA repair, identified through *Enrichr*.[52]

Functional enrichment analysis

Results from the analysis of differential expression between metastatic and primary BC samples were functionally annotated using Gene Set Enrichment Analysis (GSEA). The GSEA software (http://software.broadinstitute.org/gsea/index.jsp) was applied in pre-ranked mode to the gene list ranked on $\log_2$ fold change. Gene sets were considered significantly enriched at FDR $\leq$ 0.05 when using classic enrichment statistics and 1,000 permutations of gene sets.

For Table S1C, GSEA was performed using the Biological Processes gene sets of the Gene Ontology collections from the Broad Institute Molecular Signatures Database (http://software.broadinstitute.org/gsea/msigdb). To simplify the enriched results and enhance interpretability, we applied hierarchical clustering to the enriched terms using the *pairwise_termsim* and *treeplot* functions from the *enrichplot* R package[58] (version 1.25.6). In *pairwise_termsim*, we utilized the Jaccard similarity index (JC) as the pairwise similarity metric for the enriched terms. The *treeplot* function was executed with the average agglomeration method and the default number of clusters. Cluster labels were assigned based on high-frequency terms.

For Figure 1E and Table S1D, GSEA was performed using signatures depicting different differentiation steps of the epithelial cells of the normal MG. To capture the full complexity of the cells composing the adult human MG, we used signatures identifying specific cell populations previously defined by other scRNA-seq studies, as listed in Table S1D. Since human MG development has not been so far molecularly characterized, we also exploited gene expression signatures obtained from studies on the fetal and puberal murine MG organogenesis. As the main MG branching events occur at puberty, thanks to formation of the Terminal End Buds (TEBs), we included signatures derived from studies characterizing the gene expression program of mouse TEBs. The FMM signature was obtained by re-analysis of the scRNA-seq data of Ref. [49], as described above.

For results presented in Tables S1E and S1F, GSEA was performed using the same signatures used for Figure 1E, with the exclusion of the FMM signature, and the inclusion of the Dravis Fetal MaSCs and the Fetal-like hMG organoids signatures (see above) for Table S1E, or signatures derived from fetal human endodermal organoids and from epidermal stem cells[3,59–62] for Table S1F.

Analysis of scRNA-seq data from PyMT and HER2 BC models

Primary tumors were dissected from 3 months-old MMTV-PyVT female mice, and from 6 months-old MMTV-Neu female mice. For each sample, tumors from 2 donors were pooled and processed together. Metastasis-containing lungs were harvested from 3 months-old MMTV-PyVT female mice, and from 7-8 months-old MMTV-Neu female mice; lungs were processed after confirming the presence of visible macroscopic metastases; extrapulmonary primary bronchi and mediastinal lymph nodes were discarded; for each sample, lungs from two mice were pooled and processed together.

Primary tumors and metastasis-containing lungs were finely minced with blades and incubated in dissociation medium 1 (600U/ml collagenase I (ThermoFisher), 400U/ml hyaluronidase (SIGMA), 5U/ml dispase (ThermoFisher), 10µg/ml DNAse (Roche) in DMEM:F12) for 30 min at 37°C. The tissue was then incubated in Haemolytic solution (2 x 3min at 4°C). Tissue pellets were then washed twice in HBSS containing 0.2% EDTA before incubation in dissociation medium 2 (10U/ml dispase and 10µg/ml DNAse in HBSS without Ca^{2+}/Mg^{2+}) at 37°C until complete dissociation of tissue clumps to single cells. Cell suspensions were then washed in HBSS with 0.1% BSA, filtered through a 40µm cell strainer, counted and diluted to $\sim 10^6$ cells/ml.

Cell suspensions were processed with the Chromium Next GEM Single Cell 3' GEM, Library & Gel Bead Kit v3.1 and Chip G in a Chromium Controller (10X Genomics) targeting 10,000 cells, following manufacturer's instructions. Each sample was subjected to two cell captures, thus obtaining two replicas of about 10,000 cells each per sample. Sequencing of single-indexed libraries was performed on Illumina HiSeq or NovaSeq platforms (2x150bp; 40 to 50 thousand reads per cell).

Raw reads have been aligned to the mouse reference transcriptome (mm10, version 3.0.0) and UMI counts calculated using the 10x Cell Ranger Count pipeline (version 3.1.0). The expression data were imported into R (version 3.6.3) and analyzed with *Seurat* (version 3.1.5) and *popsicleR* (version 0.2.1) R packages. For each replicate, low quality cells were identified as outliers based on the distribution of the number of genes, UMI counts, and percentage of reads mapping on mitochondrial genes per cell and were subsequently discarded. Doublets were detected and removed using *Scrublet* (version 0.2.1). After doublet removal, the two replicates of each sample were merged. Cell cycle scores were assigned to each cell and regressed out prior to dimensional reduction, with automated identification of cells in S or in G2/M phases of the cell cycle. Following regression, principal component analysis (PCA) was used to reduce dataset dimensionality. Louvain clustering and UMAP visualization were performed on PCA reduced embedding, using Seurat *FindClusters* and *RunUMAP* functions respectively, after selecting an appropriate number of principal components for each sample. Upon the evaluation of an optimal clustering resolution, clusters were annotated with the *scMCA* R package.

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

Tumor cells of PyMT and HER2/Neu lung metastasis were integrated with the tumor cells of the corresponding primary tumors using the *FindIntegrationAnchors* and *IntegrateData* functions of the *Seurat* integration strategy.

To create a reference matrix of expected cell type expression profiles for cell type annotation of mouse BC spatial transcriptomic data, we separately integrated epithelial, immune, vascular and stromal cells from the two mouse metastatic samples using the same procedure outlined above. After identification and removal of low-quality cells and contaminants, we used the remaining well-characterized subpopulations to generate a reference matrix of expected cell type expression profiles for cell phenotyping of mouse cell spatial transcriptomic data.

Definition and quantification of gene signatures from single-cell RNA-seq data

Specific BC cell states were defined on the expression of gene signatures, listed in Tables S1B, S1D and S1G. Signature expression was extracted averaging the expression values of the signature genes in each cell with R function *rowMeans*. For each signature, genes expressed in none of the cells were filtered out.

The FMM (Table S1E) and the MTM signatures (Table S1G) were obtained as described above; Luminal and Basal signatures (Table S1B) are from a prior scRNA-seq study of the normal human MG[19]; the EMT signature (Table S1G) is from the HALLMARK gene-set collection.[63] The Cycling score was calculated as the sum of the S and G2/M scores, as determined by *Seurat*[64].

MTM^{HIGH} BC cells were defined as those in which the MTM signature exceeded the 95th percentile of the distribution of the MTM signature expression in the L2 cells of the normal human MG. The rationale for this is that if the MTM signature reflects a Fetal cell state, then the expression of that signature in adult mammary gland cells should be consider background noise.

Primary tumors were classified as ''MTM-Low'' or ''MTM-High'' if their average MTM expression was $\leq$ 95th percentile or > 95th percentile (respectively) of the distribution of MTM expression in the L2 cells of the normal MG.

Analysis of the FMM signature expression in breast cancer primary-metastasis pairs

Raw bulk RNA-seq gene counts from primary-metastasis breast cancer pairs in the AURORA US and RAP datasets[2] were used to assess the MTM signature levels in primary and metastatic samples. From the original 129 samples of the AURORA US dataset (Gene Expression Omnibus: GSE209998), we selected 54 non-regional lymph node metastases paired with their corresponding primary tumors. From the original 125 samples of the RAP dataset (Gene Expression Omnibus: GSE193103), we selected 67 non-regional lymph node metastases paired with their corresponding primary tumors. For both datasets, raw counts were normalized to counts per million mapped reads (CPM) using the *edgeR* R package[57] (version 3.36.0). Genes with a CPM greater than 1 in at least 10% of the samples were retained for further analysis. The median expression of the MTM signature was calculated for each sample, based on the expression levels of its constituent genes. The distribution and statistical significance of the MTM signature median expression in primary versus metastatic samples were analyzed using the *ggwithinstats* function from the *ggstatsplot* (version 0.12.3) package in R, applying Welch's t-test.

Primary tumors of the AURORA US were classified as having either low or high MTM signature expression relative to the median expression in the metastatic samples from the same dataset. As shown in Figure 3C, the expression of the MTM signature in primary Low BC samples is significantly lower compared to their matched metastases, whereas no difference is observed between BC primaries already displaying high representation of the MTM program and their matched metastatic samples. Subsequently, primary breast cancers with high and low MTM signature expression were compared using Welch's t-test to evaluate the time (in years) required to develop clinically detectable metastases.

Genes differentially expressed between metastatic and primary paired samples were identified using a paired design and the *glmQLFit* and *glmQLFTest* functions of *edgeR* package with the following parameters FC >= and FDR < 0.05.

Collection and processing of human metastases and CODA Deep-learning workflow for 3D reconstruction

Breast cancer metastases in liver (TNBC), brain (HER2+) and soft tissue (ER+) were collected from three different patients undergoing metastasectomy. Primary breast cancer (TNBC, HER2+ and ER+ specimens were collected from patients undergoing breast cancer surgery before receiving any treatment. Triple-negative and ER+ primary tumors were from the same two patients that later developed liver and soft tissue metastases, respectively. Samples were formalin-fixed, paraffin-embedded and sectioned every 4 μm. Every third tissue section was stained using H&E, with two sections every three held out for the brain metastasis. For liver metastasis, every second tissue section was stained with H&E with one section every two held out. For ER+ metastases and primary tumors, all section were stained with H&E. Tumors were scanned at ×20 using a Nanozoomer Scanner 2.0RS (Hamamatsu), equipped with the NDPscan software (v. 3.1). CODA deep learning pipeline[11] was then used to perform image registration, deep learning tissue multi-labeling (based on manual annotations of tissue type defined as cancer, stroma, blood vessels, necrosis, muscle, adipocytes and hepatocytes) and segmentation. After segmentation, each tissue types were used to perform 3D meshes reconstruction allowing a straightforward 3D visualization of the structures of interest. 3D reconstruction was performed using the Aivia software (version 14.1.0).

Survival analysis of breast cancer primary tumors

To test the association of MTM signature with metastatic proclivity, we downloaded the METABRIC collection from the European Genome-Phenome Archive (EGA, http://www.ebi.ac.uk/ega/) under accession number EGAD00010000210.[23] Original Illumina

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

probe identifiers have been mapped to Entrez gene IDs using the Bioconductor *illuminaHumanv3.db* annotation package for Illumina HT-12 v3 arrays obtaining log2 intensity values for a total of 19,422 genes.

To evaluate the prognostic value of the MTM signature, we first stratified the n=997 BC patients of the METABRIC Discovery dataset according to high (more than the third quartile, >Q3), intermediate (less or equal to the third quartile and greater than the first quartile, <Q3&>Q1) or low (less or equal to the first quartile, <Q1) expression of the MTM signature. Then, we estimated, using the Kaplan– Meier method, the probabilities that patients would remain free of metastatis. To confirm these findings, the Kaplan–Meier curves were compared using the log-rank (Mantel–Cox) test, reflecting the significance of the association between low MTM signature expression and longer distant metastasis-free survival. P-values were calculated according to the standard normal asymptotic distribution. Survival analysis was performed in GraphPad Prism.

Multivariate analysis of the association of BC subtypes, Lymph node status, histological grade, age, Hormonal Therapy (Hormonal), Chemotherapy (Chemio), Radiotherapy (Radio) and expression of the MTM signature with disease-free survival has been performed on the 997 BC patients of the METABRIC discovery dataset using the Cox regression model of the *analyse_multivariate* function of the *survivalAnalysis* R package (version 0.3.0). Patients were stratified according to the expression value of the MTM signature in the primary tumor being lower (Low) or higher (High) than the median. For subtypes based on histo-pathological data, we divided patients into two categories, separating those bearing primary tumors positive for hormone receptors but negative for HER2 expression ($HR^+/HER2^-$), from patients carrying the more aggressive HER2-positive or triple-negative BCs (other). For Lymph node status, we stratified according to the absence (0) or presence (>0) of Lymph node positivity. For histological grade, we combined patients carrying grade1 and grade2 BCs (1-2), opposed to patients bearing grade 3 BCs. For the age, we used a 50-year threshold, usually linked to menopause and increased risk of BC-related death. Therapeutic interventions were coded as either occurring to the patient (HT, CT or RT, respectfully), or not occurring (NO_HT, NO_CT and NO_RT, respectfully). Forest plot representation of the multivariate analysis results has been obtained using the *forest* function of the *forestplotter* R package (version 1.1.2).

To assess whether specific breast cancer subtypes were unevenly represented in the high- versus low-MTM signature groups, we visualized the distribution of breast cancer subtypes across the MTM signature groups using a Sankey diagram. This diagram, conceptually similar to an alluvial plot but emphasizing the quantitative flow between categories, was generated with the *sankeyNetwork* function from the R package *networkD3*.

2D topological analyses of human TNBC samples

Brightfield whole-slide images (WSIs) of immunohistochemical staining were acquired using a Nanozoomer Scanner 2.0RS (Hamamatsu), operated with NDPscan 3.1 software. Immunohistochemical staining were segmented to obtain polygons representing single cells using QuPath[65] (v. 0.5.0), extended through custom Groovy scripting to enable advanced annotation, segmentation, and batch processing workflows through a StarDist segmentation model.[66] This custom-trained model was integrated into QuPath through scripted pipelines for automated segmentation of complex nuclear and cellular shapes. Regions of interest (ROIs) within WSIs were defined using a custom Groovy-based contrast-enhanced tissue detection algorithm. Segmentation parameters (e.g., nuclear size thresholds, shape descriptors) were optimized to minimize false detections. Following segmentation, a weakly supervised machine learning object classifier was trained within Stardist. The classifier leveraged morphological and intensity-based features to distinguish tumor cells from stromal and other non-neoplastic cells. The model was iteratively refined using manually annotated training sets and validated by testing on independent ROIs for a total 68,408 cell objects (24,296 Tumor cell objects and 44,112 Stromal cell objects) obtaining a Precision score of 0.95 and an F_1 score of 0.89. R package sf (version 1.0-21) was employed to obtain tumor tissue polygons (ttp) by merging of contiguous Tumor cell polygons after simplification, expansion and erosion. Starting from ttps global topological descriptors (Tumor perimeter/area ratio and tumor islands/area ratio) were derived for each patient as described above for 2D topological analyses of tumor and metastases outgrowths.

Patient samples were then categorized as Expansile or Trabecular based on Tumor islands/area ratio Z-score (below and above Z-score=0, respectively). Kaplan-Meier analyses of Figure 3H was then conducted based on this categorization.

Spatial transcriptomics of human and mouse BC tissues

Data acquisition

Samples were 5 μm FFPE sections from archival surgical specimens (listed in Table S1I) derived from 8 primary TNBCs and from a HER2+ breast cancer brain metastasis from our institution, and from a metastasis-bearing lung, freshly-harvested from a female MMTV-PyMT C57BL/6J mouse at 5.5 months of age. Two sections were obtained from each sample, one stained with H&E, and the other used for the CosMX protocol (Nanostring). To control for RNA stability, an additional section was taken from the human samples and tested for the expression of *UBC* by ISH with RNAscope 2.5HD Duplex Assay (ACD bio) following manufacturer's instructions.

The CosMX procedure was carried out following manufacturer's instructions, using either the Human or the Mouse CosMX Universal cell Characterization RNA panels (Nanostring), both including antibodies for B2M/CD298, Pan-CK and CD45.

Data preprocessing and cell phenotyping

Data preprocessing was conducted using the AtoMx Spatial Informatics Platform. Images were segmented to extract the transcript count matrix, cell boundaries (polygons), and corresponding cell metadata. The expression matrices, polygon coordinates, field of

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

view (FOV) coordinates, and cell metadata were exported and loaded into R using *LoadNanostring* function of *Seurat* (version 5.0.3) in R 4.3.2. Each sample's raw data was normalized individually using the regularized negative binomial regression model implemented in Seurat *SCTransform* function. Before normalization, cells with low total transcript counts - based on the distribution of total counts across the dataset - were removed. This filtering resulted in a total of 999,262 cells for the 8 primary TNBC samples, 145,871 cells for the brain metastasis, and 37,896 cells for the lungs of the PyMT mouse model.

Cell typing was performed using Insitutype[32] and scRNA-seq data obtained from cell populations of human and mouse tumor tissues as references to perform the label transfer onto the CosMx dataset. The Insitutype classifier assumes that the number of transcripts observed for any gene in individual cells follows a negative binomial distribution and employs an Expectation-Maximization (EM) algorithm to assign cells to specific cell types, while iteratively updating the expected expression profiles for each type[32]

For the human HER2+ breast cancer brain metastasis sample, the reference matrix was constructed using cells from BBM0, BBM5, BBM7, and BBM9 HER2+ brain metastases. This reference included 40 well-characterized cell subpopulations from the tumor microenvironment. These were defined by integrating immune and stromal clusters from both normal human mammary gland and breast cancer samples, along with tumor and host tissue cells, as indicated above. Similarly, the reference matrix for the PyMT mouse lung metastasis, comprising 38 distinct cell types, was constructed using cells from PyMT metastases, derived by integrating the subpopulations from two mouse lung metastases, as indicated above. For the primary TNBC samples, we used two custom reference matrices: one was used to classify cell types on the basis of their expected expression profiles and was constructed using cells of the normal hMG, 4 TNBC samples BC16, #9282, #9281 and BBM8. The second custom profile matrix was used to classify BC cells into MTM^{HIGH} and MTM^{LOW} cells, and it was constructed using BC cells of BC16, #9282, #928, BBM8 and BLM963 stratified according to the expression of the MTM signature.

To create the custom profile matrices, we employed the *create_profile_matrix* function from the *SpatialDecon* R package (version 1.12.3), which generates average gene expression profiles for each cell type based on a scRNA-seq count matrix and cell annotations. Cell typing was performed in supervised mode using the *insitutypeML* function from the *InSituType* R package (v1.0.0; https://github.com/Nanostring-Biostats/InSituType). As inputs, we used scRNA-seq reference data, raw expression profiles, spatial context, and morphological features (e.g., cell size, shape, and immunofluorescence intensity for PanCK and CD45) of individual spatial transcriptomic samples. The cell type labels provided by *insitutypeML* were then manually refined, incorporating prior knowledge of marker genes specific of each cell type.

Analysis of cell-cell communications

To quantify cell-cell communication and identify key interaction patterns among distinct cell populations, we utilized the *CellChat* R package[33] (version 2.1.2), using normalized gene expression data, cell spatial coordinates (in microns), and cell labels as inputs. We followed the recommended processing parameters from the official CellChat guide (https://github.com/jinworks/CellChat) for inferring spatially proximal cell-cell communication from spatially resolved transcriptomics. Specifically, we set the conversion factor for spatial coordinates from pixels to micrometers to 0.12, calculated the minimum cell center-to-center distance using the *computeCellDistance* function, and used half of this value as the tolerance factor. The interaction range, or contact range, was set to 10 microns to restrict contact-dependent signaling. Communication and signaling networks for individual ligand-receptor pairs were inferred using the *computeCommunProb* function, along with the human and mouse CellChatDB v2 databases. The *computeCommunProb* function assigns a communication probability (or strength) to each interaction by calculating probability values and conducting permutation tests to assess statistical significance. The CellChatDB v2 databases are manually curated collections of literature-supported ligand-receptor interactions for both human and mouse. Finally, we used the *computeCommunProbPathway* function to calculate communication probabilities at the signaling pathway level by summarizing the probabilities of all ligand-receptor pairs with significant interactions (p-value ≤ 0.05) associated with each pathway.

Visualization of spatial information

To visualize the spatial distribution of cell types and communication probabilities, we utilized cell boundaries exported from AtoMx, along with functions from the *sf* (version 1.0-16) and *tmap* (version 3.3-4) R packages.

Derivation of patient-derived organoids from metastasis tissues

To generate organoids, single cell suspensions obtained as described above were resuspended in GFR Phenol Red-Free Matrigel (Corning) and 40µl drops containing ~20.000 cells were plated in 24-well plates. As reported in Dekkers et al.[67], the optimal growth medium was defined by seeding organoids in both Type 1 and Type 2 expansion media. BBM7 and BBM17 outgrowth was achieved in Type 2 medium. Organoids were split by enzymatic dissociation with TrypLE Express Enzyme (Thermo Fisher) as soon as they started to become dark. Actively growing, well-formed, healthy organoids were used for downstream applications.

Generation and analysis of ChIP-seq data from brain metastasis PDOs

ChIP-seq of Met-Orgs

Organoids were isolated from Matrigel through incubation with Cell Recovery solution (Corning) for 30 min on ice. After washing with cold PBS, intact organoids were fixed with 1% Formaldehyde stabilized with Methanol (Sigma) in PBS for 10 minutes at room temperature with rotation and then quenched with 125mM Glycine (VWR) for 5 min. Following 3 washes with cold PBS at 500 x *g* for 5 min, pellets were snap frozen in dry ice and stored at -80°C.

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

Organoid pellets were lysed with sonication buffer (10mM Tris-HCl pH8, 0.25% SDS, 2mM EDTA, supplemented with protease inhibitors) for 30 min on ice and then chromatin was fragmented with Covaris E220 Evolution sonicator to achieve a desired shearing size ranging from ~200bp to 700bp (settings: Peak Incidence Power 105 Watts, Duty Factor 10%, Cycles per Burst 200, duration 7 min). Lysates were diluted with 1.5 volumes of equilibration buffer (10mM Tris-HCl pH8, 233mM NaCl, 1.66% Triton X-100, 0.166% sodium deoxycholate, 1mM EDTA, supplemented with protease inhibitors) and centrifuged at 14.000 x *g* for 10 min at 4°C to pellet insoluble material.

Immunoprecipitation was performed by incubating 1-1.5μg of sheared chromatin with an equal amount of antibody targeting the following histone modifications: H3K27Ac (Abcam ab4729), H3K4Me1 (Diagenode C15410194), H3K4Me3 (Millipore 07473), H3K36Me3 (Diagenode C15410192), H3K27Me3 (Millipore 07449), H3K9Me3 (Abcam ab8898). After an overnight incubation at 4°C on a wheel, Protein G Dynabeads (Thermofisher) blocked with PBS + 0.1% BSA were used to isolate antibody-bound chromatin fragments by incubation for 2 hours at 4°C on a wheel. Washes of progressive stringency were then performed in the presence of protease inhibitors: twice with RIPA-low salt (10mM Tris-HCl pH8, 140mM NaCl, 1mM EDTA, 0.1% SDS, 0.1% sodium deoxycholate, 1% Triton X-100), twice with RIPA-high salt (10mM Tris-HCl pH8, 1mM EDTA, 500mM NaCl, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate), twice with RIPA-LiCl (10mM Tris-HCl pH8, 1mM EDTA, 250mM LiCl, 0.5% NP-40, 0.5% sodium deoxycholate), once with 1X TE. Protein G-bound immunocomplexes were then resuspended in ChIP elution buffer (10mM Tris-HCl pH8, 5mM EDTA, 300mM NaCl and 0.4% SDS) with Proteinase K (NEB) and reverse-crosslinked for 1 hour at 55°C and overnight at 65°C. Finally, immunoprecipitated DNA was purified with SPRIselect Beads (Beckman Coulter).

Libraries for Next Generation Sequencing were prepared using the Watchmaker DNA Library Prep Kit for double-stranded DNA (Watchmaker Genomics), following manufacturer's instructions. Single-end sequencing was performed with Illumina NextSeq 550Dx to obtain ~25 million 75bp-long reads per sample.

ChIP-seq data processing

Read quality control was performed with FastQC v0.11.9 and MultiQC v1.9. The reads were aligned to the human hg38 reference (GENCODE Release 25 basic gene annotation) using BWA- mem v.0.7.17, sorted using SAMtools[68] v1.10 and converted into binary files (BAM). PCR duplicates, multi-mapped and low-quality reads were marked and removed using Picard v2.23.1 and SAMtools v1.10. Reads overlapping ENCODE blacklisted regions hg38 (i.e. regions in the human genome with signal artefacts in next generation sequencing experiments) were removed. For the visualization of ChIP-seq tracks, bedGraph tracks were generated using the GenomeCoverageBed function from bedtools v2.29.2, scaling the coverage by a constant factor calculated as 1 million divided by the number of mapped reads. The tracks were converted into bigwig using ENCODE bedGraphToBigWig v357.

De-novo chromatin state characterization

De novo chromatin state characterization was performed using a multivariate Hidden Markov Model approach (ChromHMM[69] v1.12) considering five histone modifications (H3K4me3, H3K27ac, H3K4me1, H3K36me3, and H3K27me3) across the MetPDOs and including additional public available ChIP-seq data, using default parameters. The reads count for all the considered samples were computed in non-overlapping 200-bp bins across the entire genome. The binarization was performed comparing ChIP-seq read count to corresponding input DNA as control to reduce the technical noise. Several models were trained in parallel and the 12-state model was selected for downstream analysis since it captured the key interaction between histone marks with minimal redundancy. Active chromatin states were defined as genomic regions with combined enrichment of at least two of the active histone marks H3K27ac, H3K4me3, and H3Kme1 and the depletion of the repressive histone mark H3K27me3, according to the Roadmap Epigenomics Consortium guidelines.[70]

Analysis of active regions

Accessible regions within the active chromatin states were annotated to target genes by performing an overlap of active chromatin states with enhancer-promoter interacting regions with an Activity By Contact (ABC) score >0.02, retrieved from 3D chromosome conformation data on BC cell lines. A consensus peakset (metBC-Org Consensus) was generated using *DiffBind*[71] v2.10.0 by merging together active chromatin states detected in all three metBC-Orgs.

Motif binding discovery

Motif discovery was performed within the accessible regions identified within the active chromatin states. The regions were further filtered to identify those annotated to genes of the MTM signature listed in Table S1G (Target regions); the remaining active regions were defined as Background regions. The HOMER[72] *findMotifsGenome* function was used to evaluate the enrichment of known motifs in the exact size of the Target regions (setting region size parameter to "given") compared to the Background regions. Enriched motifs with a FDR <0.0001 were then subjected to a Bland-Altman analysis to identify TF motifs whose frequency in the regulatory regions of MTM genes far exceeded their frequency in the Background regions.

Derivation of epithelial mammary gland cells from human healthy tissue

Human mammary epithelial cells (hMEC) were collected from reduction mammoplasties. Surgical samples were washed in HBSS supplemented with 2% PS and pre-dissected to retain epithelial tissue. Partial digestion was performed as reported in Miller et al.[73]. Briefly, dissected specimens were digested overnight in dissociation solution containing Advanced DMEM/F12 (Life Technologies), 10 mM HEPES (Life Technologies), 1% GlutaMAX™ (Life Technologies), 1% P/S, 400 U/ml hyaluronidase (Sigma) and 600 U/ml collagenase I (Life Technologies). After sedimentation of mammary gland ductal tree portions samples were washed with HBSS 2% P/S; red blood cells were then removed using a buffered solution containing 0.64% NH4Cl (Sigma). Cell pellets

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

were centrifuged at 250 × g for 5 min and washed three times with PBS containing 5% FBS. The resulting pellets were resuspended in fibroblast depletion medium (DMEM supplemented with 10% FBS; Life Technologies), plated in 10 cm tissue culture dishes, and incubated for 90 min to allow fibroblast attachment. The medium containing non-adherent mammary epithelial cells was then collected, and the plates were gently rinsed with PBS to recover remaining cells. The combined suspensions were centrifuged at 250 × g for 5 min, and the supernatant was discarded. The cell pellets were resuspended in 1 ml of 3D culture medium consisting of Advanced DMEM/F12 supplemented with 10 mM HEPES (Life Technologies), 1% GlutaMAX™ (Life Technologies), 1% penicillin-streptomycin, 2% FBS, 64 ug/ml bovine pituitary extract (BPE)(Thermo Scientific, 13028014), 10 ng/ml human EGF (Peprotech, AF-100-15), 10 μM forskolin (Merck, F6886), 10 μg/ml insulin (Merck, 91077C), and 0.5 ug/ml hydrocortisone (Merck, H0396). In parallel, collagen I (Fujifilm, 638-00781) was prepared by mixing MEM 10x, collagen I, and compound C in a 1:1:8 ratio. Equal volumes (1 ml each) of the collagen I solution and cell suspension were combined, and 80 μl drops of the mixture were dispensed into ultra-low-adhesion plates (Greiner, 662102). After polymerization for 30 min at 37°C, 1 ml of 3D medium was added to each well. After 7 days of culture, the collagen drops were examined and only structures exhibiting tube-like morphology were selected and expanded.

To recover and expand the cultures, collagen gels were digested with 1 ml per well of collagenase I (600 U/ml) for 30 min at 37°C. Following digestion, the released mammary gland cells were collected and washed twice with PBS. The cell pellets were then dissociated using TrypLE (Thermo Fisher) for 5 min at 37°C to obtain single-cell suspensions, which were subsequently resuspended in collagen I gels and re-embedded as described above. For protein extraction, four drops were pooled per condition. Samples were centrifuged and resuspended in 200 μl of RIPA buffer (Thermo Scientific) supplemented with protease and phosphatase inhibitor cocktails. The suspension was transferred to tubes containing beads (Fisher, 15565799) and homogenized using a Fisherbrand™ Bead Mill 24 Homogenizer (Fisher Scientific) at speed 5 with a run/stop cycle of 30 s/15 s for three cycles. The homogenate was centrifuged briefly, and the supernatant was transferred to a new microcentrifuge tube. Following centrifugation at maximum speed for 10 min, the clarified lysate was collected. DNA was precipitated by adding polyethyleneimine (PEI; 0.02% w/v final concentration from a 1 mg/ml [0.1% w/v] stock solution) and incubating the mixture on ice for 1 h. The samples were then centrifuged, and protein concentrations were determined using a BCA Protein Assay Kit (Thermo Scientific) according to the manufacturer's instructions.

3D cell culture in collagen 1 gels

Sub-confluent cells were trypsinized using TrypLE (ThermoFisher) and washed with PBS. Cells were resuspended at $1x10^5$ cells/ml in collagen1 at 1.2 or 2.4 mg/ml. Collagen was prepared as described for hMEC culture. 80 μl droplets were let to jellify for 30 minutes at 37°C in 24 well Ultra Low Attachment plates (Greiner, 662102). Complete media was then added to the wells. Gels were used for RNA extraction (see below) or fixed with 4% PFA for 15 minutes and labeled with phalloidin (AlexaFluor 647, ThermoFisher) 1/200 and SYTO16 (ThermoFisher) 1/1000 or Hoechst (Thermofischer) 1/1000 in 0.1% Triton solution in PBS for 2 hours at 37°C. For FGF experiments, after seeding, cells were cultured in 10% FBS containing FGF10 at 25ng/ml (Peprotech, 100-26) and basic FGF (FGF2) at 40ng/ml (Peprotech, 100-18). The next day, cells were washed twice in FBS free medium and grown in 2% FBS containing FGF10 and FGF2 for 4 days. Medium change with fresh FGF supplementation was performed every 2 days.

Mammosphere assay

Subconfluent monolayers of 4T1 cells were trypsinized, counted, filtered with a 40 μm cell strainer and plated as single-cell suspensions (4000 cells/cm^2) on ultra-low attachment plates (Corning) in media containing 5% matrigel. Colonies were imaged with an inverted light Microscope (Axiovert 200M, Leica) using a 10x objective at day 3 and day 7.

Plasmids and lentiviral transduction

The dual GFP/Luciferase-expressing lentiviral construct is a gift from Antonio Rosato.

pLV-hPGK-hETV1AAA-FLAG-T2A-EGFP, pLV-hPGK-EGFP and pLV-hPGK-mScarlet3 were produced on demand by Vectorbuilder. All plasmids were validated by sequencing. ETV1AAA refers to ETV1 human sequence with mutation in two phosphodegron motifs: V63P64D65 to A63A64A65 and V71P72D73 to A71A72A73. These mutations are made to prevent ubiquitination and proteasomal degradation thus promoting protein stability.[74]

Lentiviral particles were prepared and delivered to cells as previously described in Panciera et al.[75]. In brief, lentiviral particles were prepared by transient liposome-mediated transfection with the specified lentiviral vectors and packaging vectors. 48 hr post-transfection supernatant was collected, filtered through 0.45 micrometers and used for transduction.

Sorting of fluorescent cells

For isolating EGFP or mScarlet3 expressing cells, cells were collected using TrypLE and resuspended in their respective culture medium. Cells were then washed, resuspended in sorting buffer (PBS with 1 mM EDTA, 25 mM HEPES, 0.1% BSA) and sorted using a FACS Aria sorter (BD Biosciences). Non fluoresecent cells were used to define fluorescence background. After FACS purification, cells were seeded in plates coated with collagen I.

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

Cell editing with CRISPR-cas9

Cells were detached using TrypLE, washed 3 times in PBS and nucleofected with the SE-kit (Lonza) using Nucleofector (Lonza) following manufacturer's instructions. MDA-MB-231 parental and brain trained were nucleofected at $2x10^5$ cells with DS-138 program. TUBO cells were nucleofected at $2x10^5$ with CM-150 program. 4T1 cells were nucleofected at $1x10^5$ cells with CM-150 program. All reagents required for nucleofection (Cas9-GFP, tracrRNA, and Electroporator enhancer) were from IDT. crRNA design was done using CCTop and IDT online predictors, strictly excluding crRNA with high off-target probability. At least four independent crRNA were tested for each target and the best was selected based on protein decrease in WB. For KO of human ETV1 we combined three different crRNA to obtain a high KO efficiency. The AltR-crRNA (IDT) used in this study are listed in Table S1M.

Control (CTRL) cells were nucleofected with the same protocol using a control crRNA (#1072544) from IDT. Cell editing efficiency was validated by WB. For TUBO, 4T1 FGFR2 KO and MDA-MB-231 cell lines, the KO efficiency was high, thus cells were used at the population level after nucleofection avoiding a clonal selection step.

For 4T1 ETV4 KO, after nucleofection, cells were seeded at single-cell dilution in a 96 well plate, allowing to obtain individual clones. Clones were then amplified and selected for Etv4 expression by immunoblot. Two ETV4 KO clones were validated and then used individually in the subsequent experiments. CTRL 4T1 cells were instead made of three sgCTRL-nucleofected clones mixed together.

For hMEC, collagen gels were digested, and 10^5 dissociated cells were nucleofected using the P3 kit (Lonza) with EL-110 program. After nucleofection cells were embedded in fresh collagen gels for 7 days. Collagen gels were fixed for confocal imaging or used for protein extraction in order to validate ETV KO efficiency by Immunoblotting.

Western Blot

Immunoblots were performed as previously described.[42] The following primary antibodies were used: ETV1 (Invitrogen, PA5-77975), ETV5 (Invitrogen, PA5-30023), ETV4 (PEA3, SantaCruz, sc-113), FGFR2 (Cell Signaling Technologies, 23328), GAPDH (Millipore, MAB347).

qRT-PCR

Cells were collected using the RNeasy Mini Kit (Qiagen) for total RNA extraction, and contaminant DNA was removed by DNase treatment. The qRT-PCR analyses were carried out on reverse-transcribed cDNAs with QuantStudio 5 (Applied Biosystems, Thermo Fisher Scientific) and analysed with QuantStudio Design and Analysis software (version 1.4.3). Expression levels were always normalized to *GAPDH* expression. Primers are listed in Table S1N.

Wound healing assay

4T1 cells were seeded at a density of 3×10^5 cells/ml (70 µl per well) into culture insert 2-well chambers placed in 35-mm µ-Dishes (Ibidi, 81176) one day prior to the experiment. On the following day, the inserts were gently removed, and each dish was filled with 2 ml of pre-warmed culture medium. Phase-contrast images were acquired every hour for 10 h using an inverted light Microscope (Leica, Axiovert 200M) using a 4x objective.

Analysis of bulk RNA-seq of mouse and human cell lines

RNA was extracted from mouse 4T1, TUBO or human MDA-MB-231 BC cells cultured on classic 2D culture conditions or in 3D collagen I gels. Harvesting of cells growing in 2D condition was performed as described in Castellan et al.[39]. For harvesting of the 3D cultures, collagen 1 gels were washed once with HBSS and digested with 1ml collagenase I (Invitrogen, 17100-017) at 300U/ml for 10 minutes at 37°C. After 10 minutes, mechanical disruption of the gels was performed by pipetting. Additional 5 minutes incubation at 37°C was performed. The suspension was then washed with HBSS by centrifugation and cell pellet was frozen at -80°C. RNA extraction was performed using RNeasy Mini Kit (Qiagen); DNA was removed by DNase treatment.

RNA-seq libraries for deep-sequencing were prepared with TruSeq Stranded mRNA (Illumina), and sequencing was performed with NextSeq500 System (Ilumina). About 20-40 million reads per sample were obtained. Read quality was assessed using FastQC (version 0.11.9; https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Adapter sequences and low-quality bases ($Q < 20$) were trimmed from raw reads with BBDuk (version 38.92; https://sourceforge.net/projects/bbmap/) and reads shorter than 35 bp post-trimming were discarded. Trimmed reads were then aligned using STAR[55] (version 2.7.9a). Specifically, reads of the MDA-MB-231 samples were aligned to the human reference genome (hg38), and reads of the 4T1 samples were aligned to the mouse reference (mm10). Raw gene counts were generated in R 4.1.1 using the *featureCounts* function from the *Rsubread* package[56] (version 2.8.1) with gene annotation provided by GENCODE v44 for hg38 human genome and the Cell Ranger v3.0.0 annotation for mm10 mouse genome. Raw counts were normalized to counts per million mapped reads (CPM) and to fragments per kilobase of exon per million mapped reads (FPKM) using the *edgeR* R package[57] (version 3.36.0). Only genes with a CPM greater than 1 in at least two samples were retained for further analysis. Differential gene expression was analyzed using the *exactTest* function of the *edgeR* R package.

In vivo metastasis and primary tumor formation assays

For injections in mice, subconfluent monolayers of cells were trypsinized and washed 3 times in PBS. For tail vein and intracardiac injections, cells were resuspended in HBSS without calcium and magnesium (ThermoFisher) at the desired concentration. Before

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

injection, cells were filtered with a 40 μm cell strainer to avoid injection of cell aggregates. Mice were warmed to allow vasodilatation and 100 μl were injected in the tail vein. For intracardiac injections, mice were anesthetized, and 100 μl were injected in the left ventricle of the heart. For histological examination and analysis, organs were freshly collected, fixed in 4% PFA (Sigma-Aldrich) in PBS overnight and processed for paraffin embedding and H&E staining.

For mammary fatpad injections, cells were resuspended in 100% Matrigel (Corning) and 100 μl were injected in the inguinal mammary fatpad. In the experiment performed in Figures 6B–6D, 10^4 4T1 CTRL cells expressing EGFP and 10^4 4T1 ETV4 KO cells expressing mScarlet3 were mixed together in 100% Matrigel and injected in the right inguinal mammary fat pad. After 4 weeks, primary tumors and lungs from the same animal were collected. After fixation in 4% PFA, primary tumors were cryoprotected in sucrose 30%, overnight at 4°C. Tumors were then embedded in a gelatin 100 G (7.5%)-sucrose (10%) solution and frozen in an isopentane bath at −50 °C for 2 min before storage at −80 °C.

For metPDO subcutaneous injection, organoids were isolated from Matrigel through incubation with Cell Recovery solution (Corning) for 30 min on ice, washed in ice-cold PBS and dissociated to single cells by incubation in 10 U/ml dispase. After extensive washing, cells were counted and resuspended in GFR Phenol Red-Free Matrigel at a concentration of 10^7 cells/ml. 100 μl of cell suspension were subcutaneously injected in the flank of 8-10 weeks old female NSG mice. After 12 weeks masses were harvested and processed for paraffin embedding and serial sectioning.

Intracranial injection of glioma (CT-2A) cells was performed as previously described.[39] Briefly, $3x10^5$ cells in 2 μl were stereotaxically injected into mouse brain at the following coordinates: 2.2ML; - 0.5AP; 2.5DV. After 7 days, brains were collected and processed for CUBIC protocol.

For FGFR inhibitor treatment experiment, mice were treated with futibatinib (MedChemExpress, HY-100818) administered by oral gavage, 5 times a week for two consecutive weeks. The compound was prepared as a 20 mg/ml solution in DMSO and then diluted at 2 mg/ml in corn oil (Sigma). Each animal received 100 μl per day, corresponding to a daily dose of 0.2 mg Futibatinib. Control mice were given the vehicle alone following the same dosing schedule.

Immunohistochemistry on paraffin sections

Immunohistochemical staining was performed on PFA-fixed, paraffin-embedded tissue sections as previously described.[42] Brightfield images were obtained with a Nanozoomer Scanner 2.0RS (Hamamatsu), equipped with the NDPscan software (v. 3.1). Primary antibodies used in immunostaining were: CK8 (Abcam, ab53280), pan-cytokeratin (MNF116, Agilent, M0821), Ki67 (Novus, NB600-1252).

3D reconstruction of metastasis based on cytokeratin IHC staining

For the data presented in Figures 5M and S3D, 3D reconstruction of cancer cells based on cytokeratin staining was used. In both cases, samples were fixed in 4% PFA and embedded in paraffin. Samples were entirely and serially sectioned at 4 μm. In Figure 5M sections with lungs bearing D2.OR ETV1 metastasis were stained with CK8 antibody. In Figure S3D sections with metPDO subcutaneous tumors were stained with MNF116 antibody. For both experiments, Registration of 2D sections were performed by HistoDigital UG (www.histodigital.de). Sections were extracted and first rigidly registered to restore general alignment. This was refined in a second, non-rigid registration step, to compensate for the non-linear deformations due to histological preparation. After registration, CK-positive cancer cells were segmented by training and applying a machine learning based pixel classifier algorithm using the Aivia software (v14.1.0). The cancer segmented channel was then used for applying the "3D Object Analysis-Meshes" recipe allowing 3D reconstruction of the structure.

In vivo bioluminescence imaging

Mice were administered 3mg of IVISbrite D-luciferin (Revvity) via intraperitoneal (IP) injection and anesthetized. *In vivo* bioluminescent images were acquired 10 minutes after IP injections using IVIS Lumina S5 instrument with Living Imagesoftware (Revvity) in automatic mode. Bioluminescent signal from both a ventral and a dorsal view was acquired. Living Image software was used to analyze bioluminescent signal by applying an identical ROI in the head area of each mouse. Bioluminescence signal was expressed as radiance (photons/sec) allowing comparison between different acquisitions. Data are presented as the sum of total flux from the ventral and dorsal view.

Imaging and 3D reconstruction

Metastatic cells expressing fluorescent reporters were injected in the left ventricle of the heart (intracardiac) to obtain brain metastasis, or injected in tail vein to obtain lung metastasis. Primary tumors were obtained by injection in the mammary fat pad. The cell number and timing for each cell line are indicated in the figure legend. Entire brain hemispheres, lungs and primary tumors were collected and optically cleared with CUBIC protocol[76] and labelled with Propidium iodide (PI) (Figures 2D, 2F, 6A, 6D, S3B, and S3E) or with MACS protocol and labelled with an anti-EGFP vioR667 antibody (MiltenyBiotec, 130-131-646) (Figures 2E, 2G, 5D, 5F, 7H, 7J, S3A, S3C, and S4F). For CUBIC protocol, mice were perfused with 10ml PBS. Organs were collected, washed in PBS, and fixed overnight in PFA 4% at 4°C. Organs were washed 3 times in PBS for 2 hours under gentle agitation at RT. After immersion in 50% CUBIC-L for 6 hours, samples were placed in 100% CUBIC-L containing PI at 5 μg/ml (only for EGFP fluorescent samples) for at least 3 days at 37°C under slow rotation. After 3 washes in PBS for 2 hours, samples were immersed in 50%

CUBIC-R for 6 hours. Samples were then immersed in CUBIC-R 100% for 1 day at RT under slow rotation. Samples were then transferred in a 1:1 mix of silicon (Lesker, 704BB) and mineral oil (Sigma, M8410) to obtain a final refractive index of 1.52 and kept in oil until imaging. CUBIC-L was made of 10w% Triton X100 (Sigma) and 10w% N-buthyldiethanolamine (Tokyo chemical) in H2O. CUBIC-R was made of 45w% Antypirin (Tokyo chemicals) and 30w% of nicotinamide (Tokyo chemicals) in H2O to obtain a refractive index of 1.52. MACS clearing protocol was done using clearing kit (MiltenyBiotec, 130-126-719) following manufacturer's protocols for brain and lungs tissue. When indicated, tissues were also stained with anti-CD31 vioR667 (MiltenyBiotec, 130-128-736) to label blood vessels.

Entire cleared organs were scanned using the Blaze light-sheet microscope (Miltenyi biotec) with a 1x objective. Single metastases were imaged at higher resolution using a 12x objective. Entire organs/metastasis were visualized by using 3D view of the reconstructed fluorescent z-stack using Imaris viewer (version 10.2.0) or Aivia (version 14.1.0) software. 3D meshes reconstructions were made using the Aivia software. For this, fluorescence signal was segmented by training and applying a machine learning based pixel classifier algorithm. The metastasis segmented channel was then used for applying the ''3D Object Analysis-Meshes'' recipe allowing a straightforward 3D visualization of the metastatic structure faithful to the original signal. See Video S6 for a compendium of the fluorescent images used for 3D reconstructions depicted in Figures 2, 5, 6, 7, and S3.

Collagen1 gels were imaged by confocal microscopy using Leica STELLARIS 5 microscope, equipped with LASX software (version 4.1.0.23081) with a 10x objective. 3D reconstruction was performed based on the segmented phalloidin signal.

Image analysis

Quantification mammosphere and wound healing assays

Quantifications were performed using Fiji software (version 2.9.0). For mammosphere assay, images were binarized and the *analyze particles* function was used to measure the area of each structure. Results are presented as the result of 3 independent experiments merged together.

For wound healing assay, time-lapse images were analyzed. The initial image acquired immediately after insert removal was defined as representing 100% wound area. Subsequent images were processed through threshold-based binarization, hole filling, and background filtering. The wound area at each time point was quantified using the *Analyze Particles* function, and results were expressed as the percentage of remaining gap area relative to the initial wound area. Results are presented with 2 biological replicates in each condition and is representative of 3 independent experiments.

Histological analyses

For lung metastasis quantifications, histological analyses were performed using QuPath (version 0.5.0) software allowing semi-automatic metastasis selection. All quantifications in lungs were the results of the analysis of 4 sections spanning across the entire lungs. The following parameters were calculated: Total metastasis area, representing the sum of the area of all the detected metastasis for one animal. Metastasis average area, representing total metastasis area divided by metastasis number.

For primary tumors quantification presented in Figure 6C, cryosections of 30 μm were mounted and acquired using Pannoramic 250 Flash III digital scanner (3dhistech). The analysis was performed with Fiji software. After background removal, EGFP and mScarlet3 signals were binarized and their respective surface area was measured. The representation of EGFP vs mScarlet3 signal was plotted as percentage of area.

3D quantification of metastasis volume and number

The analysis was performed using Aivia software. Whole lungs containing 4T1-EGFP$^+$ metastatic nodules were imaged by light-sheet microscopy. The EGFP fluorescence signal was segmented by training and applying a machine learning–based pixel classifier algorithm. The segmented metastasis channel was then used to generate 3D mesh reconstructions. For each 3D mesh (corresponding to an individual metastasis), the volume (μm^3) was extracted and used for quantification. Metastases arising inside soft tissues were excluded. Only metastases larger than 5×10^3 μm^3 were retained, corresponding to metastatic nodules ranging from small clumps of tumor cells to full-blown macrometastases. Smaller meshes were discarded, as they likely represented background or autofluorescent signals that could not be reliably distinguished from clusters of a few cancer cells. In Figure 6D, the number of macrometastases per lungs was measured in entire lungs acquired by light-sheet microscopy.

3D quantification of morphometric parameters

Light-sheet images of brain and lung metastases were used. EGFP-positive cancer cell signals were segmented by training and applying a machine learning–based pixel classifier. The segmented cancer channel was then processed with the ''3D Object Analysis – Meshes'' recipe. Sphericity and equivalent spherical diameter measurements were directly obtained from the reconstructed 3D meshes. For branching analysis, a custom skeletonization script (SkeletonizeWithNodesDetection_3D_1_00.py https://github.com/AiviaCommunity/PythonForAivia/blob/master/PythonEnvForAivia/Recipes_NoAutomatedTests/ProcessImages/SkeletonizeWithNodesDetection_3D.py) was applied to the segmented cancer channel. To ensure consistent skeletonization results across different metastases, the segmented channel was rescaled in XY to a 2 μm/pixel ratio using the ''ScaleImage.py'' Python recipe. Skeletonization produced a branches channel, on which the ''3D Object Analysis – Meshes'' recipe was applied to measure in 3D the number and length of branches within each metastasis. The total branch length was calculated as the sum of the lengths of all branches per metastasis. For size normalization, branch parameters were divided by the equivalent spherical

Please cite this article in press as: Caire et al., A 3D morphogenetic blueprint for metastatic outgrowth in breast cancer, Cell (2026), https://doi.org/10.1016/j.cell.2026.03.009

diameter, yielding branch number per metastasis size and total branch length per metastasis size. To generate a single branching parameter, z-scores were calculated for both normalized variables, and a branching composite index was obtained by averaging these z-scores.

For *in vitro* collagen gel branching quantifications, a similar analysis pipeline was used. In this case, phalloidin staining was employed to train the pixel classifier. To account for potential differences in initial cell density between experiments, the total branch length was divided by the number of independent (non-connected in 3D) structures in each image. Results are presented with at least nine biological replicates obtained across three independent experiments.

2D topological analyses of experimental metastases outgrowths

Tumor tissue annotations were generated from light-sheet microscopy images using Aivia image analysis software. Three lung metastases were analyzed per group. For each sample, ten 2D images encompassing the entire metastatic lesion were extracted and analyzed. These images were exported as raster files and converted into binary masks using the Otsu thresholding algorithm (implemented in the EBImage package, version 4.48.0) in R (version 4.4.3). Binary masks were subsequently converted into tumor tissue polygons (TTPs) using the raster package (version 3.6-32). Each individual TTP morphology was quantitatively assessed for area and perimeter using the sf package (version 1.0-21). For each tumor sample, two global topological descriptors were computed: Tumor Perimeter/Area (PA), defined as the ratio between the total perimeter and total area of all TTPs within a sample ($PA = \Sigma P_n / \Sigma A_n$) and Tumor Islands/Area (nI), defined as the ratio between the number of TTPs and their total area within a sample ($nI = n / \Sigma A_n$).

Ki67 positive cells quantification

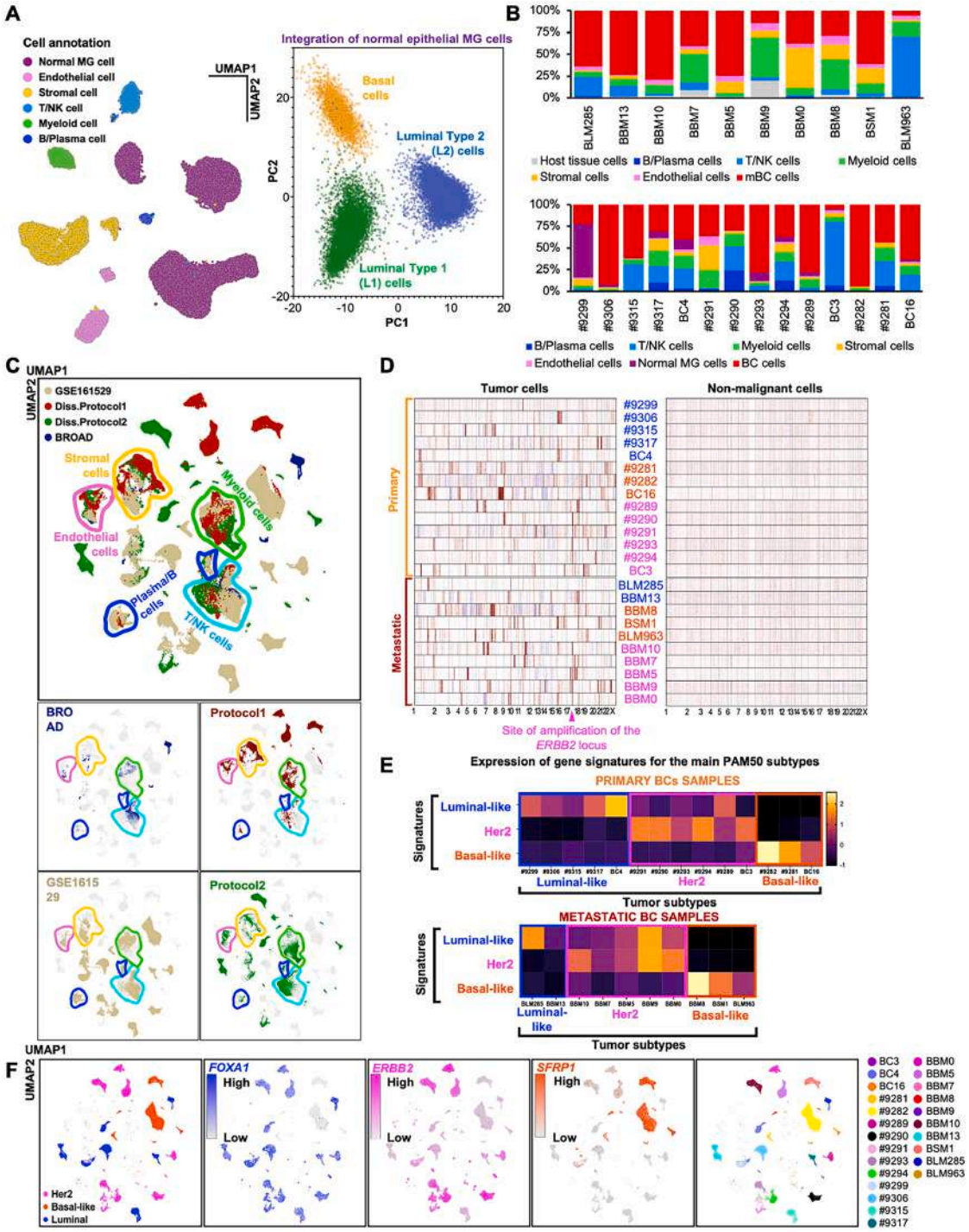
Following image acquisition, Ki67 immunostaining in 4T1 and MDA-MB-231 primary tumors was quantified using QuPath (version 0.5.0). Cell segmentation was performed with the "Cell Detection" function to identify all cells in 3 non-necrotic regions of interest for each tumor. "Positive Cell Detection" function was then applied to identify nuclei exhibiting Ki67 staining. Analysis of Ki67 immunostaining in lung metastases was performed using AIVIA software (version 14.1.0). Metastatic regions were manually selected and defined as regions of interest (ROIs). Ki67-positive nuclei and all nuclei were independently segmented by training and applying two different machine learning based pixel classifier algorithms. For each metastatic ROI, the "Nuclei Count" recipe was applied to measure the total number of nuclei and the number of Ki67-positive nuclei. 3 lungs per group were analyzed. All results (relative to primary tumors and metastasis) are expressed as the percentage of Ki67-positive cells relative to the total number of detected cells.

QUANTIFICATION AND STATISTICAL ANALYSIS

The number of biological and technical replicates and the number of animals is indicated in the Figure Legends, Main text and Methods section. All tested animals were included. For all studies only female mice were considered. The animal ages are specified in the Methods section. The sample size was not predetermined. Randomization was not applicable to our experiments with cell lines. Methods of statistical analysis are indicated in the Figure Legends; unless otherwise indicated, analyses were carried out with GraphPad Prism v.8.0.2 for MAC software.

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

Figure S1. Cell composition of metastatic and primary BCs, related to Figure 1

(A) scRNA-seq analysis of the normal MG tissues from three healthy donors. Left panel: UMAP projection of the cells obtained after integration of the three scRNA samples, colored according to the main cell types. Right panel: result of the principal-component analysis (PCA) of MG epithelial cells, showing that they form three independent clusters, each corresponding to one of the known main epithelial cell populations of the MG.

(B) Cell composition of each metastatic (upper graphs) and primary (lower graphs) BC lesion according to scRNA-seq analyses. Normal MG cells are epithelial cells lacking CNV alterations. Host tissue cells are glial cells (astrocytes and oligodendrocytes) for brain metastases, hepatocytes and biliary duct cells for liver metastases, and keratinocytes for skin metastases. Myeloid cells include monocytes, macrophages, dendritic cells (DCs), and, for brain metastases, microglia. Stromal cells include fibroblasts, pericytes, and smooth muscle cells.

(C) UMAP projections of primary and metastatic BC tissues, showing cells colored according to dissociation protocol (see STAR Methods) or data source. Note how cells of the TME (endothelial, stromal, myeloid, T/natural killer [NK], and B/plasma cells) form separate clusters according to each cell type (highlighted with the same color code as in Figure 1B), but irrespective of the dissociation protocol or data source.

(D) Identification of BC cells by inferring genomic alterations from scRNA-seq data. Panels are the results of InferCNV analyses comparing BC cells (red cells in Figure 1B) with TME cells (all non-red cells of Figure 1B) of each lesion. BC and TME cells are organized in rows, whereas genes are aligned in columns and ordered along each chromosome (from 1 to X, as indicated below the graphs). Inferred gene amplifications and deletions are shown in red and blue, respectively. As expected, HER2+ tumors display the amplification of the ERBB2 locus on chromosome 17.

(E and F) PAM50 classification of the analyzed tumors.

(E) Heatmaps showing the expression of the gene lists depicting the three main PAM50 subtypes.

(F) Excerpts of the UMAP projection are shown in Figure 1B, featuring only BC cell clusters. Cells are colored according to the PAM50 classification of each lesion (left panel), the expression of selected markers highly enriched in each BC subtype (middle panels, *FOXA1* for luminal-like BCs, *ERBB2* for Her2 BCs, and *SFRP1* for basal-like BCs), or the sample of origin (right panel).

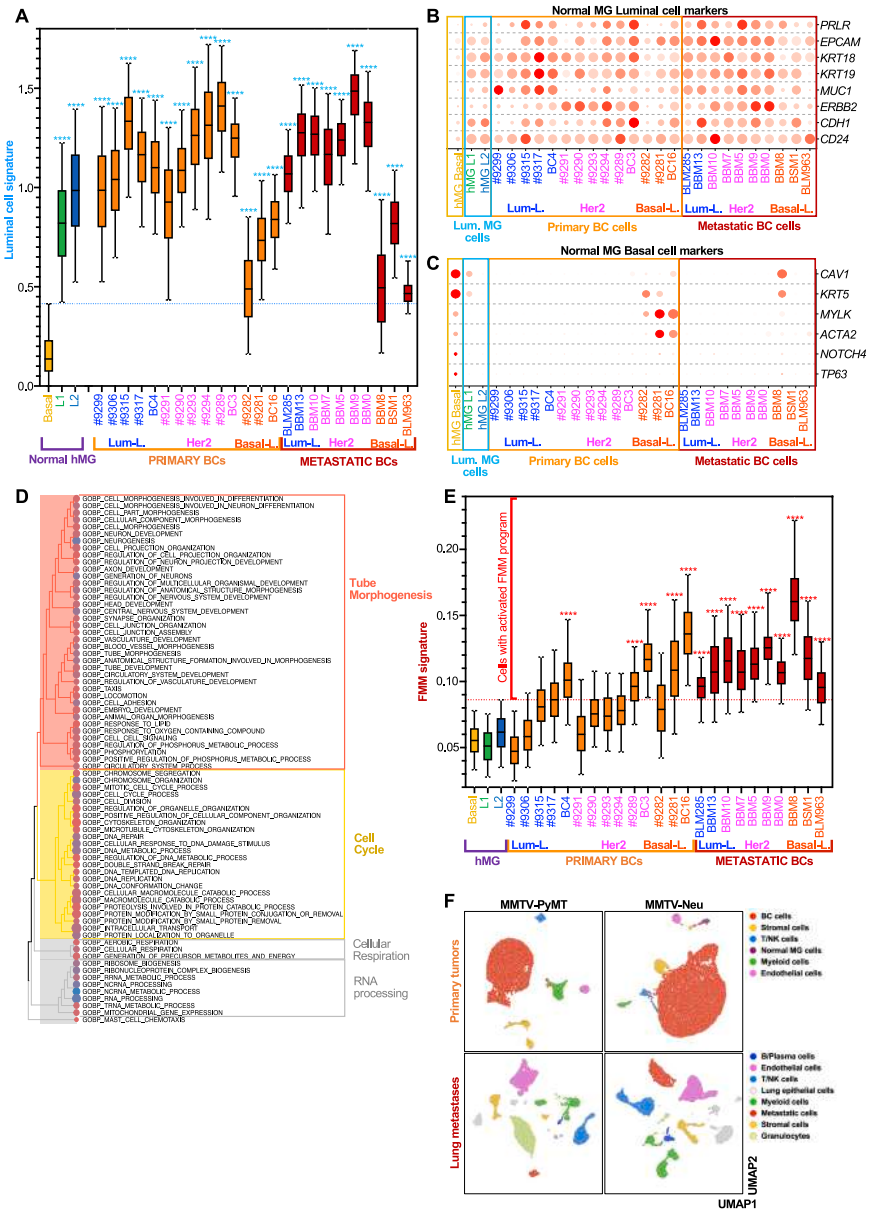


Figure S2. Characterization of the metastatic and primary BC cell states, related to Figure 1

(A) Distribution of the expression of a signature earmarking the luminal epithelial cells of the MG (Table S1B) in the three epithelial cell populations of the normal MG and in tumor cells from primary and metastatic BCs. Data are shown as box-and-whiskers plots (whiskers extend from the 5th to the 95th percentile, the box extends from the 25th to the 75th percentile, and the line within the box represents the median). As expected by BC having a luminal origin, the expression of the signature is negligible in basal cells of the normal MG, whereas it is very high in the two normal luminal cell populations. Of note, the large majority of tumor cells composing primary and metastatic BCs express the luminal signature at levels similar to those of normal mammary luminal cells, as the signature expression significantly (****p value < 0.0001 by two-tailed one-sample t test) exceeds the 5th percentile of the expression of the signature in the L1 luminal cells of the normal MG (blue dotted line).

(B and C) Dot-plot of luminal cell (B) and basal cell (C) markers in the three normal MG epithelial cell populations and in the tumor cells of each primary and metastatic BC lesion, organized according to their inferred PAM50 subtype. Basal and luminal cells of the normal mammary gland serve as controls of the specificity of the markers for the luminal and basal cell states, respectively. Note the widespread expression of luminal markers in primary and metastatic tumor samples.

(D) Unsupervised hierarchical clustering of the gene signatures that were highly enriched (FDR < 0.001) in metastases after GSEA using the gene lists of the GO Biological Program dataset (see Figure 1D for a schematic of the procedure and Table S1C for the full GSEA results). Major clusters were labeled based on high-frequency words in the GO-term subtrees. Two major clusters emerged, one containing GO terms mostly related to the morphogenesis of tubular organs and the other related to cell-cycle progression.

(legend continued on next page)

(E) Distribution of the expression of the FMM signature in the three epithelial cell populations of the normal MG and in tumor cells of each primary and metastatic BC lesion, organized according to their inferred PAM50 subtype. Data are shown as box-and-whiskers plots (whiskers extend from the 5th to the 95th percentile, the box extends from the 25th to the 75th percentile, and the line within the box represents the median). Expression of the signature is very low in the epithelial cells of the normal adult MG (as in Figure 1F), as expected for a signature depicting a fetal cell state. Using those cells as a reference for low-level/background expression of the FMM signature, the FMM program is then considered activated in BC cells with FMM expression above the 95th percentile of its distribution in L2 cells of the normal MG (red dotted line). All metastases are composed mainly by cells with activated FMM programs, as supported by statistical analysis (****p value < 0.0001 by two-tailed one-sample t test). Primary BCs instead contain variable amounts of cells with the activated FMM program.

(F) Cell composition of primary tumors and lung metastases from the indicated mouse models of metastatic BC, as identified by scRNA-seq analyses. Panels depict UMAP projections of the cells, colored according to their cell type annotation. Lung epithelial cells include airway epithelial cells and alveolar type 1 and type 2 cells. Myeloid cells include dendritic cells and, for lung samples, alveolar macrophages and interstitial macrophages.

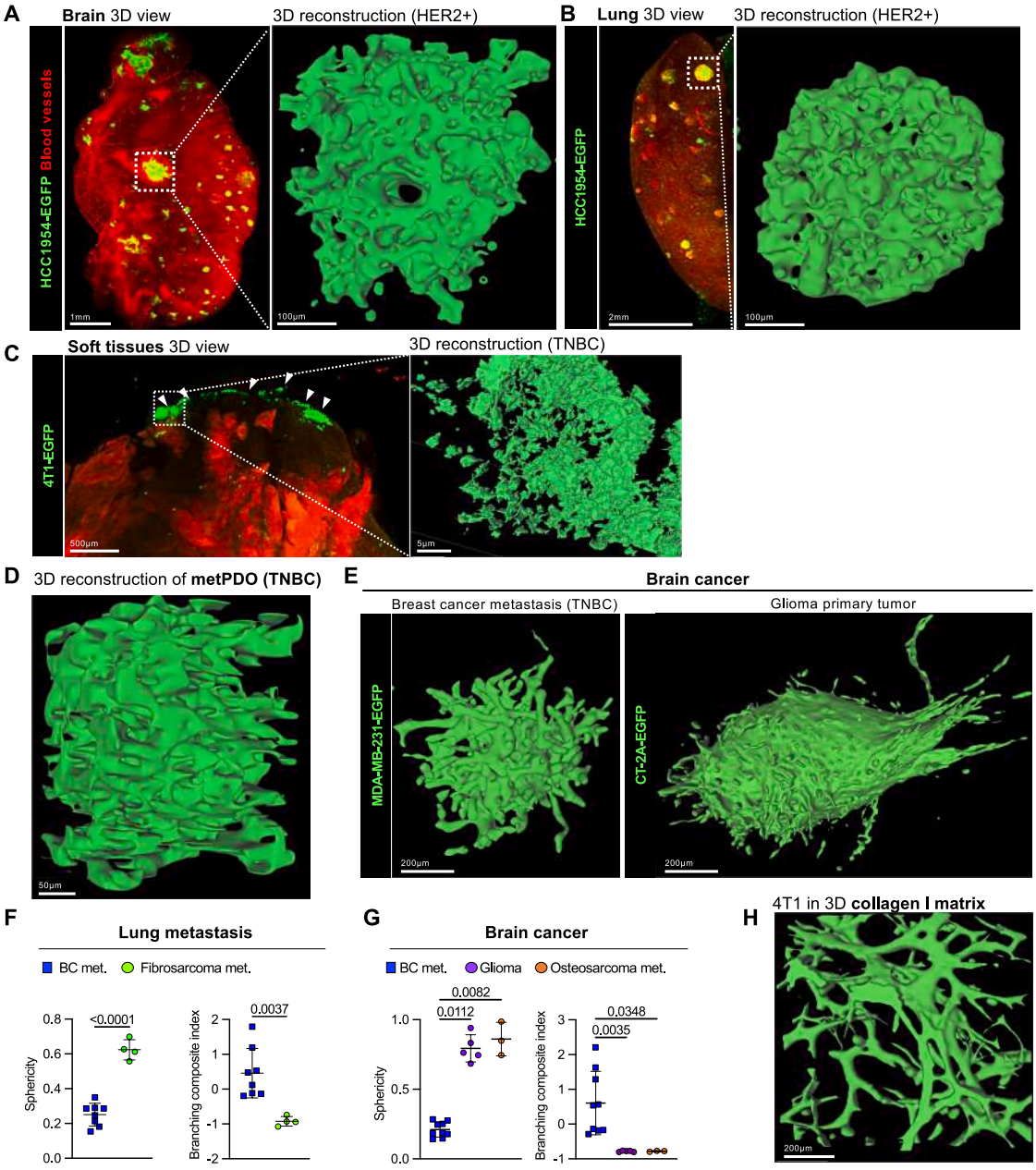


Figure S3. 3D structure of BC metastases, related to Figure 2

(A and B) Representative 3D views and 3D reconstructions of experimental metastases obtained after injection in NOD-SCID mice of EGFP-expressing human metastatic HCC1954 cells (an HER2+ human cell line).

(A) Left panel: 3D imaging of a brain hemisphere of a mouse with intracardiac injection of HCC1954 cells. The entire brain was collected, optically cleared with the MACS protocol, labeled with an anti-EGFP antibody to visualize cancer cells, and then scanned using light-sheet microscopy, as depicted in Figure 2C (scale bar, 1 mm; red are blood vessels stained by CD31 antibody). Right panel: 3D reconstruction of the structures formed by BC cells in a representative metastatic nodule (scale bar, 100 µm).

(B) Left panel: 3D imaging of a lung lobe of a mouse receiving tail-vein injection of HCC1954 cells, and lungs were collected, optically cleared with CUBIC protocol to visualize EGFP+ metastatic cells, and then scanned using light-sheet microscopy, as depicted in Figure 2C (scale bar, 2 mm; red: autofluorescence). Right panel: enlargement of a single metastasis reconstructed in 3D (scale bar, 100 µm).

(C) 3D view and reconstruction of soft-tissue metastases obtained after tail vein injection of EGFP-expressing 4T1 cells in BALB/c mice. Left panel: 3D view (scale bar, 500 µm) of the lungs and surrounding soft tissues optically cleared with the MACS protocol, labeled with an anti-EGFP antibody to visualize cancer cells, and then scanned, as depicted in Figure 2C. Soft-tissue metastases (white arrowheads in the left panel) can be seen as those not embedded in the surrounding tissues, which are characterized by a marked autofluorescence (red). Right panel: 3D reconstruction of a single soft-tissue metastasis (scale bar, 5 µm).

(legend continued on next page)

(D) 3D reconstruction of a subcutaneous lesion formed in NSG mice by injection of metastatic BC cells of the metPDO BBM2, based on serial sections stained for cytokeratin (see STAR Methods).

(E) Comparison between the architectures of similar-sized lesions formed by glioma cells or metastatic BC cells in brains from NOD/SCID mice. Panels are 3D reconstructions of experimental brain lesions obtained after intracardiac injection of EGFP-expressing MDA-MB-231 human TNBC cells or intracranial injection of EGFP-expressing CT-2A mouse glioma cells (scale bar, 200 μm).

(F and G) Quantification of 3D topo-morphometric parameters for the tumor outgrowths formed in mouse lungs by (F) or brain (G) by metastatic BC cells (MDA-MB-231 and HCC1954 cells) or, as a control, by metastatic sarcoma cells (HT1080 in F and HOS143-B in G) or by glioma cells (CT-2A). The branching composite index (left panels) combines the number of branches with their length, whereas the sphericity index (right panels) measures how much the architecture of tumor nodules approximates a sphere. The low sphericity and high branching composite indices of BC metastases support the notion that they display branched, trabecular architecture. Graphs are shown as scatterplot, mean, and SD. *p* values are obtained by a two-tailed Mann-Whitney test in (F) and the Kruskal-Wallis test with Dunn's post hoc multiple comparison test in (G).

(H) Representative 3D reconstruction of branching structures formed by 4T1 cells in a 3D collagen I matrix. 4T1 cells were seeded in large gel droplets of collagen I, and after 4 days, the droplets were fixed and stained with phalloidin. 3D reconstruction was based on segmentation of the fluorescent signal. Scale bar, 200 μm.

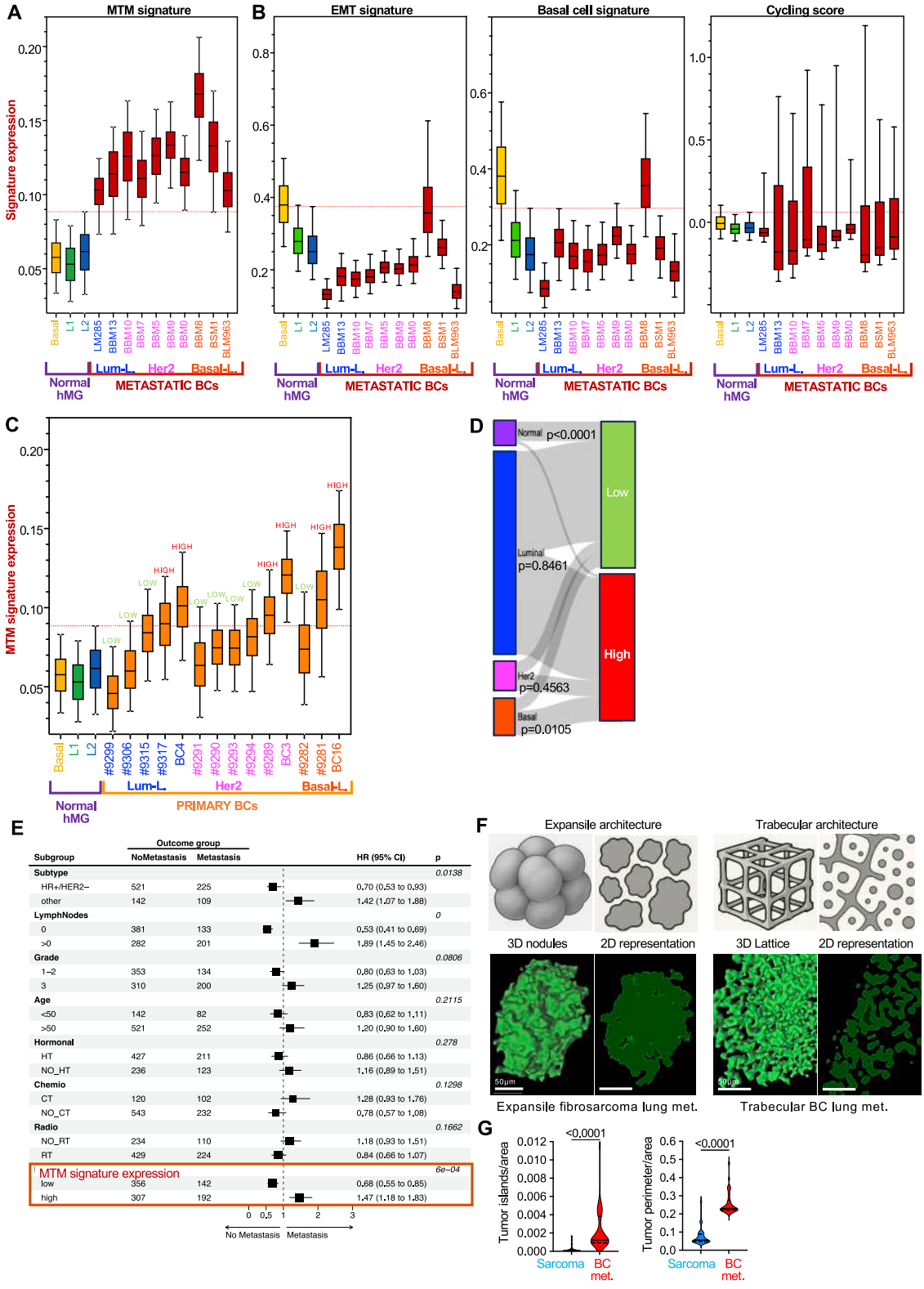


Figure S4. High expression of the MTM signature correlates with metastatic proclivity in primary tumors, related to Figure 3

(A and B) Distribution of the expression of the MTM signature (A) or of signatures depicting EMT (B, left), basal mammary (B, middle), or cycling (B, right) cell states in the three epithelial cell populations of the normal MG and in tumor cells of each metastatic BC lesion, organized according to their inferred PAM50 subtype.

(legend continued on next page)

Data are shown as box-and-whiskers plots (whiskers extend from the 5th to the 95th percentile, the box extends from the 25th to the 75th percentile, and the line within the box represents the median). The dotted red lines indicate the thresholds for background levels of expression of each signature, defined as the 95th percentile of the distribution of each signature in L2 cells of the normal mammary gland.

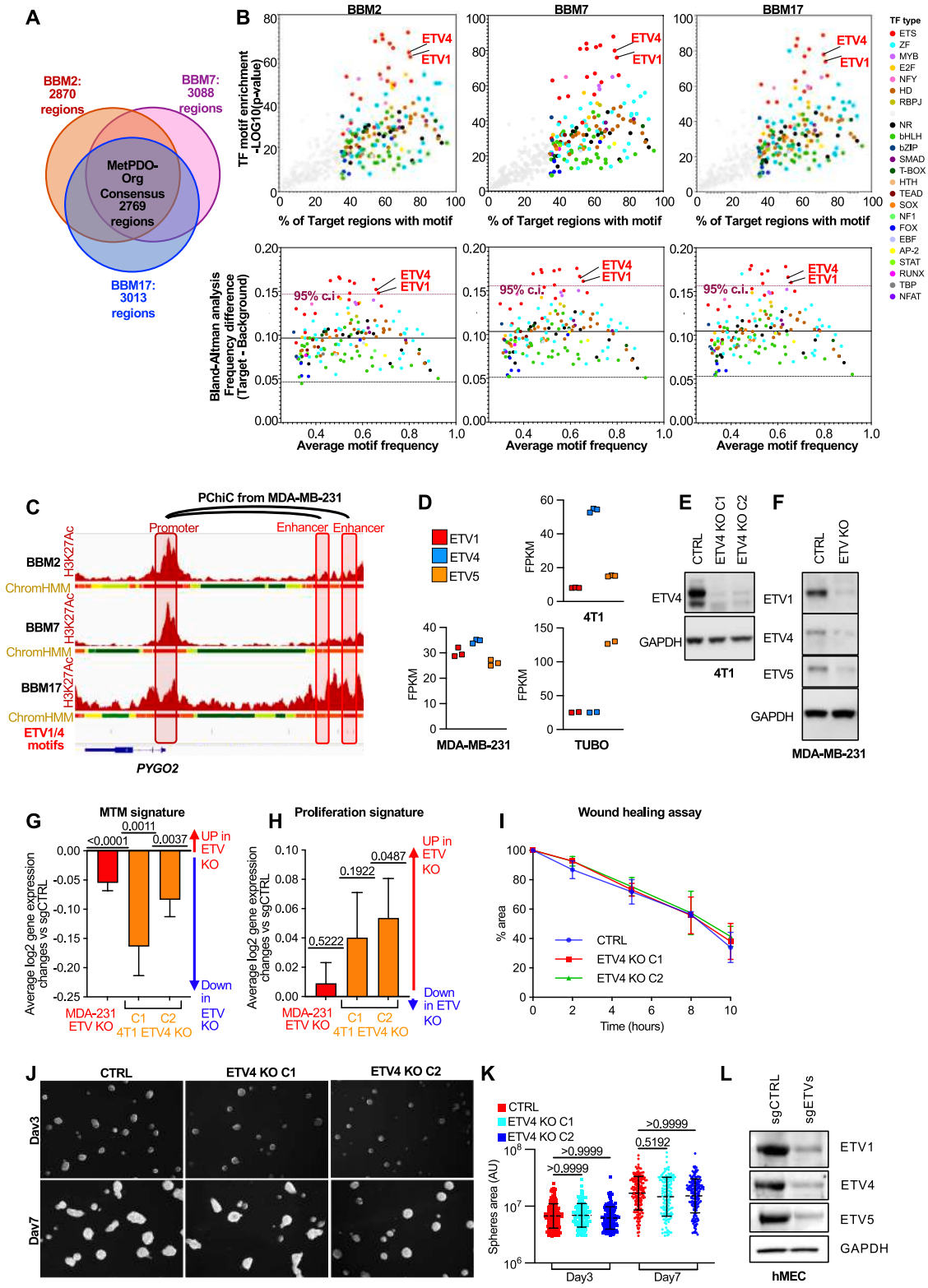
(C) Distribution of the expression of the MTM signature in the three epithelial cell populations of the normal MG and in tumor cells of each primary BC lesion, organized according to their inferred PAM50 subtype. Data are shown as box-and-whiskers plots (whiskers extend from the 5th to the 95th percentile, the box extends from the 25th to the 75th percentile, and the line within the box represents the median). The dotted red line shows the 95th percentile of the distribution of MTM expression in L2 cells, which is used as a threshold to define "MTM low" (average MTM expression $\leq$ threshold) and "MTM high" (average MTM expression > threshold) primary tumors.

(D) Sankey diagram depicting the distribution of PAM50 BC subtypes across the low and high MTM signature groups of the METABRIC dataset. *p* values for enrichment of each subtype in the two categories were calculated using Fisher's tests.

(E) Multivariate analysis and forest plot of the association of BC subtypes, lymph node status, histological grade, age, hormonal therapy (hormonal), chemotherapy (chemio), radiotherapy (radio), and expression of the MTM signature with metastasis-free survival on the 997 BC patients of the METABRIC Discovery dataset. Results show that high expression of the MTM signature is a risk factor for metastasis formation, independent of the indicated clinical variables. Patients were stratified according to the expression value of the MTM signature in the primary tumor being lower (low) or higher (high) than the median. For subtypes based on histopathological data, we divided patients into two categories, separating those bearing primary tumors positive for hormone receptors but negative for HER2 expression (HR+/HER2−) from patients carrying the more aggressive HER2-positive or triple-negative BCs (other). For lymph node status, we stratified according to the absence (0) or presence (>0) of lymph node positivity. For histological grade, we combined patients carrying grade 1 and grade 2 BCs (1–2), as opposed to patients bearing grade 3 BCs. For the age, we used a 50 years threshold, usually linked to menopause and increased risk of BC-related death. Therapeutic interventions were coded as either occurring to the patient (HT, CT, or RT, respectively) or not occurring (NO_HT, NO_CT, and NO_RT, respectively). Shown are the Hazard Ratio (HR) and its 95% confidence intervals. The *p* value of the Cox regression model reflects the significance of the association between each variable and metastasis-free survival.

(F) Examples of expansile and trabecular architectures in lung metastases obtained by injection of EGFP-expressing HT1080 (fibrosarcoma, left) or, as a comparison, MDA-MB-231 (TNBC, right) cells in NOD/SCID mice. The top panels are the same as Figure 3E and are schematics showing how tumors with different 3D architecture (left, expansile; right, trabecular) would appear in 2D sections. The bottom panels are 3D reconstructions and their 2D sections of representative lung metastases, showing how the 3D architecture reflects on 2D shapes, as predicted.

(G) Violin plots showing quantification of the indicated topo-morphometric parameters in 2D sections of 3D reconstructions of the fibrosarcoma (n = 30 sections) or TNBC (n = 30 sections) lung metastases, as depicted in (F). Both parameters (tumor islands/area and perimeter/area) are significantly higher in the trabecular TNBC metastases than in the expansile fibrosarcoma metastasis, validating their usage to infer the 3D architecture of a tumor lesion. The black line represents the median, whereas the dotted lines represent the first and third quartiles of each distribution, respectively. *p* values are determined by the two-tailed Mann-Whitney test.



(legend on next page)

Figure S5. Identification of master TFs of the MTM^{HIGH} cell state by analyzing the epigenetic profiles of metastatic BC organoids, related to Figure 4

(A) Venn diagram representing the number of active chromatin regions associated with the genes of the MTM signature in each metPDO. The area of overlap (metPDO consensus) represents the regions found associated with the MTM genes in all metPDOs.

(B) Enrichment analyses of TF motifs in the regulatory regions associated with the genes of the MTM signature. Upper panels are scatterplots for TF motifs (represented by dots) found in the regulatory regions of the MTM genes in each individual metPDO, as described in Figure 4C for metPDO consensus.

(C) Representative tracks of H3K27ac and ChromHMM profiles of an example of the genes of the MTM signature containing motifs predicted to bind ETV4 and ETV1 (red vertical lines) in the regulatory regions. Interactions between enhancers and promoter regions based on PCHiC data from human MDA-MB-231 BC cells are shown above the graphs.

(D) Normalized expression (features per thousand million, FPKM) of the ETV1/4/5 family members in the indicated mouse (4T1 and TUBO) and human (MDA-MB-231) BC cell lines as derived from RNA-seq analyses.

(E and F) Immunoblot for the indicated ETV proteins in CTRL and ETV-KO 4T1 (E) and MDA-MB-231 (F) cells. For 4T1 cells, two ETV4 KO clones (C1 and C2) were selected for the absence of Etv4 expression, while ETV-KO MDA-MB-231 cells were used as a whole population. GAPDH serves as a loading control.

(G and H) Average log2 gene-expression changes of the MTM signature (G) and of a signature for cell proliferation[64] (H) in ETV-KO MDA-MB-231 cells (red bars) or ETV4 KO 4T1 cells (orange bars) compared with their matching CTRL cells. Results are shown as bar plots, mean, and SEM. *p* values are calculated using two-tailed one-sample *t* tests.

(I) Results of wound-healing assays carried out on CTRL and the two clones (C1 and C2) of ETV4 KO 4T1 cells. The graph depicts the decrease in width (measured as % of the cell area) of a gap made on a monolayer of the indicated cells. No differences are observed between the curves of CTRL and ETV4 KO cells. Results are representative of 3 independent experiments with two biological replicas each.

(J and K) Live bright-field microscopy images (J) and quantifications (K) of 4T1 cells cultured for 3 or 7 days as mammospheres in suspension. Note that the size of the spheres formed by ETV4 KO cells remains the same as that of those formed by CTRL cells at both time frames. Graphs are shown as scatterplot, mean, and SD, and *p* values are calculated with the Kruskal-Wallis test with Dunn's post hoc multiple comparison test. Results are from 2 independent experiments.

(L) Immunoblot for the indicated ETV proteins in CTRL and ETV-KO primary human mammary epithelial cells (hMECs) used for the experiments depicted in Figure 4I. GAPDH serves as a loading control.

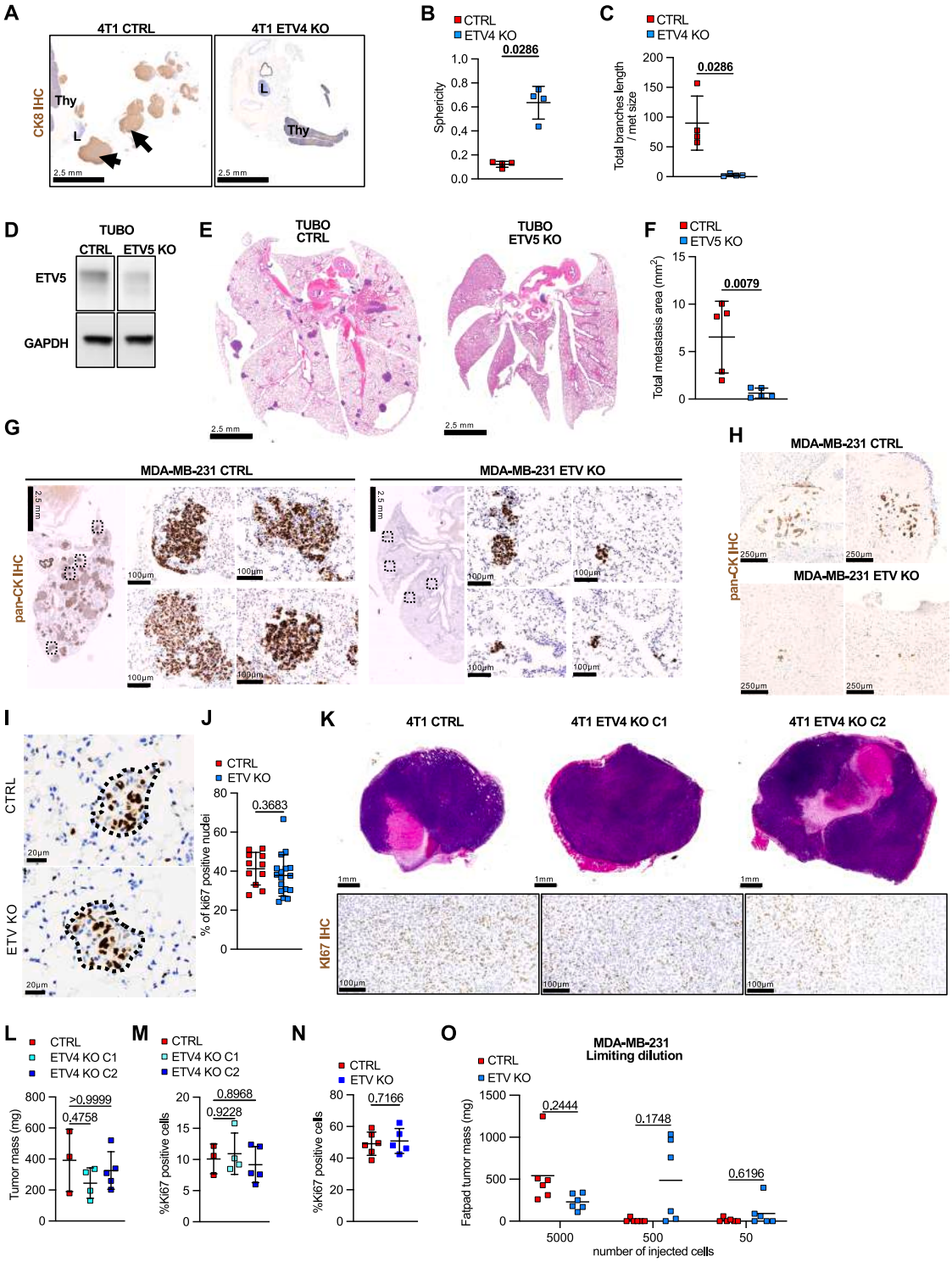


Figure S6. ETV1/4/5 are master genes required for metastatic 3D morphogenesis and outgrowth but dispensable for primary tumor growth, related to Figures 5 and 6

(A) Representative sections of the mediastina of BALB/c mice injected in the tail vein with 5×10^4 CTRL or ETV4 KO 4T1 cells and killed after 3 weeks. Sections were stained with CK8 antibody to highlight soft-tissue metastases and counterstained with hematoxylin (scale bars, 2.5 mm). This is representative of 2 ETV4 KO clones and 2 independent experiments. Thy, thymus; L, lymph node.

(legend continued on next page)

(B and C) Quantification of the 3D topo-morphometric parameters sphericity (B) and length of branches (C) for lung metastases depicted in Figure 5D, showing that the small outgrowths formed by ETV4 KO 4T1 cells ($n = 4$) display a more spherical and less branched architecture than those formed by CTRL cells ($n = 4$). Graphs are shown as scatterplot, mean, and SD. p values are obtained by the two-tailed Mann-Whitney test.

(D–F) Effects of Etv5 KO on the outgrowth of lung metastases from the HER2+ mouse BC cell line TUBO cells.

(D) Immunoblot results for ETV5 protein expression in control and ETV5 KO TUBO cells.

(E) Representative H&E-stained sections of lungs from BALB/c mice ($n = 5$ per group) injected in the tail vein with 1×10^4 CTRL or ETV5 KO TUBO cells. Mice were sacrificed after 3 weeks.

(F) Quantification of the section area (mm^2) covered by metastases, shown as a scatterplot, mean, and SD. p value is determined by the two-tailed Mann-Whitney test.

(G and H) Representative sections of (G) lungs (scale bar, 2.5 mm for left panels and 100 μm for the corresponding magnification) or (H) brains (scale bar, 250 μm) of mice injected with CTRL or ETV-KO MDA-MB-231 cells as indicated in Figures 5E–5H, respectively. Sections are stained with pan-cytokeratin (pan-CK) antibody and counterstained with hematoxylin.

(I and J) ETV1/4/5 loss does not interfere with the proliferation of nascent metastatic nodules. Panels in (I) are representative sections (scale bars, 20 μm) of incipient (<100 cells) metastatic nodules from lungs of mice injected with CTRL or ETV-KO MDA-MB-231 cells, stained with Ki67 antibody and counterstained with hematoxylin. Quantifications in (J) show no significant changes in proliferation rate between CTRL ($n = 11$) and ETV-KO ($n = 17$) lesions at this stage of metastatization. Graphs are shown as scatterplot, mean, and SD. p values are obtained by two-tailed unpaired t test. Data are obtained from the lungs of 3 independent mice.

(K–M) ETV4 KO does not impair primary tumor formation and growth by 4T1 cells. Panels are representative sections (K) and quantifications (shown as scatterplot, mean, and SD) of the weight (mg) (L) and percentage of Ki67-positive cells (M) of primary tumors formed by injection of 1×10^4 4T1 cells in the mammary fat pad of BALB/c mice and collected 3 weeks after injection (CTRL cells, $n = 3$ mice; ETV4 KO C1 cells, $n = 4$ mice; ETV4 KO C2 cells, $n = 5$ mice). p values are obtained by the Kruskal-Wallis test with Dunn's multiple comparison correction post hoc test for (L) and one-way ANOVA with Sidak's multiple comparison post hoc test for (M). Upper panels in (K) are the whole sections stained with H&E (scale bars, 1 mm). Lower panels are higher magnifications of sections stained with Ki67 antibody to highlight proliferating cells (scale bars, 100 μm). Note that both the size of the tumors and the relative amount of Ki67 proliferating cells are not significantly different between CTRL and ETV4 KO conditions.

(N and O) Loss of ETV1/4/5 does not impair primary tumor seeding and growth by MDA-MB-231 cells. Graphs are quantifications (shown as scatterplot, mean, and SD) of the percentage of Ki67-positive cells (N) and of the weight (O) of primary tumors formed by injection of the indicated number of CTRL ($n = 6$ per group) or ETV-KO ($n = 5$ in N and $n = 6$ per group in O) MDA-MB-231 cells in the mammary fat pad of NOD-SCID mice. Data are shown as scatterplot, mean, and SD. p values are obtained by a two-tailed unpaired t test with Welch's corrections for (N) and a two-way ANOVA with Sidak's multiple comparison post hoc test for (O).

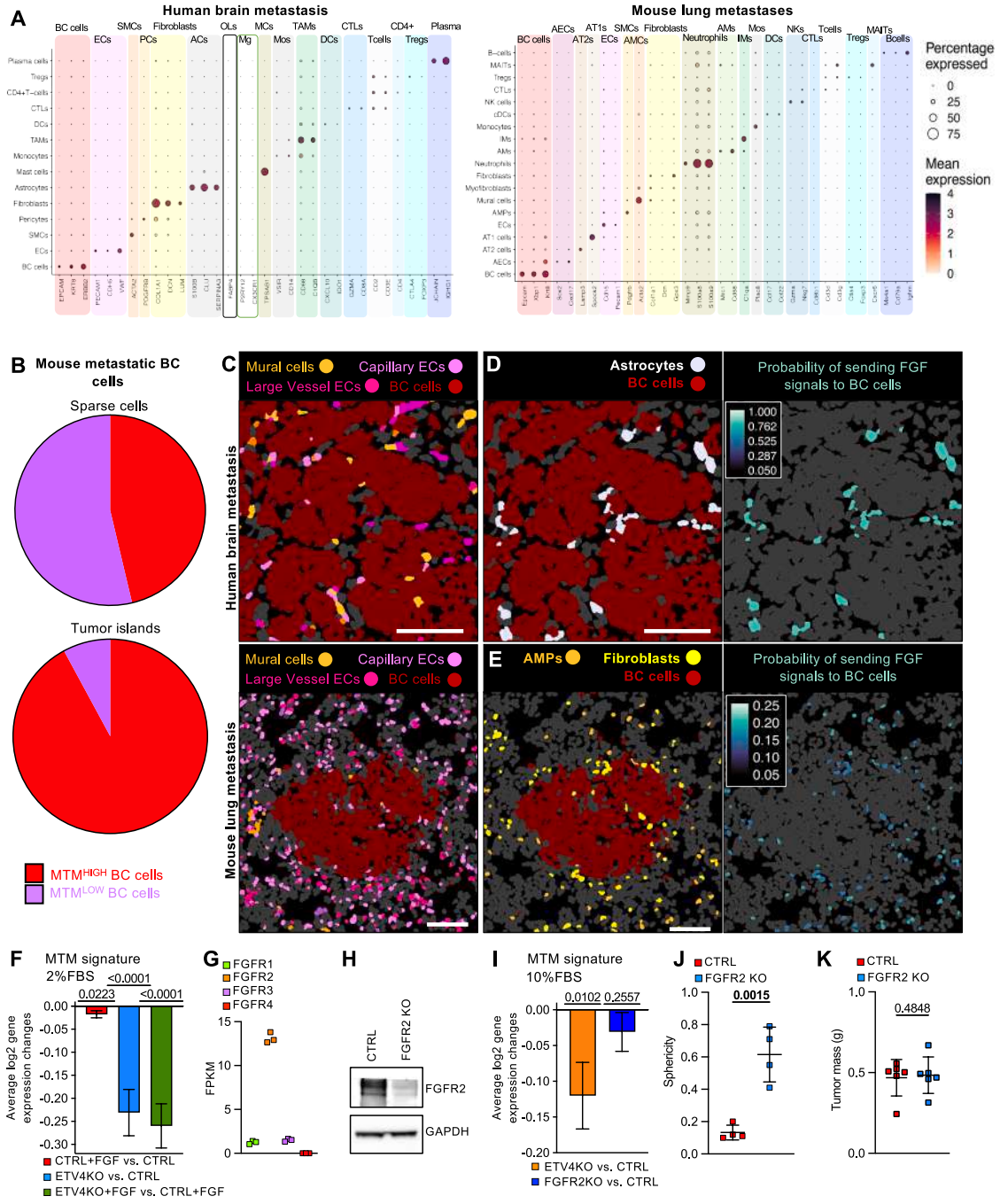


Figure S7. ST and functional analyses on FGF signaling, related to Figure 7

(A) Dot plot of markers for the indicated cell types in the cell types identified through ST analyses of the sections from the human BC brain metastasis and the mouse lung metastasis shown in Figures 7A and 7B, respectively. Cell-type abbreviations: ECs, endothelial cells; SMCs, smooth muscle cells; PCs, pericytes; ACs, astrocytes; MCs, mast cells; Mos, monocytes; TAMs, tumor-associated macrophages; DCs, dendritic cells; CTLs, cytotoxic T-lymphocytes; plasma, plasma cells; AECs, airway epithelial cells; AT2s, alveolar type 2 cells; AT1s, alveolar type 1 cells; AMPs, Axin2+ myofibrogenic progenitor cells; AMs, alveolar macrophages; IMs, interstitial macrophages; NKs, natural killer cells; MAITs, mucosal-associated invariant T cells.

(B) Pie charts showing the percentage of BC cells annotated by ST as MTM^{HIGH} and MTM^{LOW} in small aggregations (sparse cells, composed by 1–3 cells; number of cells = 108) or in major outgrowths (tumor islands; number of cells = 7,586) in the lung parenchyma of the mouse metastatic model depicted in Figure 7B.

(C) Compendium of Figures 7A and 7B. Panels are *in silico* reconstructions depicting the relative spatial distribution of BC and vascular cells in the same FOVs of the human brain BC metastasis (top panel) and the mouse lung metastasis (bottom panel), shown in Figures 7A and 7B, respectively.

(D and E) Panels are *in silico* reconstructions of the same FOVs of the human brain BC metastasis (D) and the mouse lung metastasis (E) shown in Figures 7C and 7D, respectively, depicting the relative spatial distribution of BC cells and cells of the TME that are the source of FGF signals (left panels) and the distribution of the

(legend continued on next page)

probability of producing FGF signals by TME cells (right panels). Cells not producing FGF ligands are in dark gray. Only astrocytes are predicted to produce FGF ligands in the microenvironment of the brain metastasis, whereas FGF signals are produced by stromal cells (fibroblasts and fibroblast-like AMP cells) in the mouse lung microenvironment (compare left and right panels).

(F) Average log2 gene-expression changes of the MTM signature in 4T1 cells cultured in 3D collagen I gels in the presence or absence of FGF in the medium, as depicted in Figure 7E. Bars depict the changes of the expression of MTM signature genes induced by FGF treatment in CTRL cells (red bar), by ETV4 KO in untreated cells (blue bar), and by ETV4 KO in FGF-treated cells (green bar). Results are shown as bar plots, mean, and SEM. *p* values are calculated using two-tailed one-sample *t* tests.

(G) Normalized expression (features per thousand million, FPKM) of FGF receptors in 4T1 cells, as derived from RNA-seq analysis.

(H) Immunoblot for FGFR2 in CTRL and FGFR2 KO 4T1 cells. GAPDH serves as a loading control.

(I) Average log2 gene-expression changes of the MTM signature in FGFR2 KO 4T1 cells (blue bar) compared with their matching CTRL cells. Expression changes of the MTM signature in ETV4 KO 4T1 cells (orange bar) are shown as a positive control. Results are shown as bar plots, mean, and SEM. *p* values are calculated using two-tailed one-sample *t* tests.

(J) Quantification of the 3D topo-morphometric parameter sphericity for metastatic nodules formed in mouse lungs by CTRL ($n = 4$) or FGFR2 KO ($n = 4$) 4T1 cells, as in Figures 7H and 7I. Graphs are shown as scatterplot, mean, and SD, and the *p* value is obtained by a two-tailed unpaired *t* test.

(K) FGFR2 KO does not impair primary tumor formation and growth by 4T1 cells. Quantification (shown as scatterplot, mean, and SD) of the weight (mg) of primary tumors formed by injection of 2×10^4 4T1 cells in the mammary fat pad of NOD/SCID mice and collected 3 weeks after injection ($n = 6$ mice each). *p* value is obtained by the Mann-Whitney test. Results are representative of 2 independent experiments.