PhD Program in Neuroscience

Instituto de Neurociencias (CSIC-UMH)

Impact of intratumoral EMT heterogeneity on breast cancer progression

Doctoral Thesis presented by

Raúl Jiménez Castaño

Thesis Director

Prof. María Ángela Nieto Toledano

Thesis Co-director

Dr. Khalil Kass Youssef

Universidad Miguel Hernández de Elche

-2026-

The doctoral thesis, entitled “Impact of intratumoral EMT heterogeneity on breast cancer progression” is presented **under the conventional thesis format with the following publications:**

“A hormetic transcriptional program coregulates invasion, proliferation and dormancy to define metastatic potential”

Raúl Jiménez-Castaño, Nitin Narwade, Gema Moreno-Bueno, Berta Sánchez-Laorden, Joan Galcerán, Khalil Kass Youssef and M. Angela Nieto

Nature Communications, accepted on 23rd February 2026, Published 4th March 2026

https://doi.org/10.1038/s41467-026-70242-4

“Epithelial–Mesenchymal Transition as a Central Driver of Tumor Cell Plasticity”

Raul Jimenez-Castaño and M. Angela Nieto

Nature Cancer, accepted on 25th February 2026

DOI**:** 10.1038/s43018-026-01154-x

Prof. **Maria Angela Nieto Toledano**, director, and Dr. **Khalil Kass Youssef**, co-director of the doctoral thesis entitled "**Impact of intratumoral EMT heterogeneity on breast cancer progression**".

REPORTS:

That Mr. **Raúl Jiménez Castaño** has carried out, under my supervision, the work entitled "**Impact of intratumoral EMT heterogeneity on breast cancer progression**" in accordance with the terms and conditions defined in her Research Plan and in accordance with the Code of Good Practice of the Universidad Miguel Hernández de Elche, satisfactory fulfilling the objectives foreseen for its public defense as a doctoral thesis.

We sign for appropriate purposes

Thesis director	Thesis co-director
Prof. María Ángela Nieto Toledano	Dr. Khalil Kass Youssef

Dra. Cruz Morenilla Palao, Coordinator of the Neurosciences PhD programme at the Instituto de Neurociencias in Alicante, a joint centre of the Universidad Miguel Hernández de Elche (UMH) and the Centro Superior de Investigaciones Cientificas (CSIC),

REPORTS:

That Mr. **Raúl Jiménez Castaño** has carried out under the supervision of our PhD Programme the work entitled "**Impact of intratumoral EMT heterogeneity on breast cancer progression**" in accordance with the terms and conditions defined in its Research Plan and in accordance with the Code of Good Practice of the Universidad Miguel Hernández de Elche, fulfilling the objectives satisfactorily for its public defense as a doctoral thesis.

Which I sign for the appropriate purposes

Dra. Cruz Morenilla Palao

Coordinator of the PhD Programme in Neurosciences

Funding/Grants/Scholarship:

The realization of this Doctoral Thesis has been carried out thanks to a 'Severo Ochoa Excellence Programme' PhD contract (PRE2020-091888).

This work was supported by grants MCI PID2021-125682NB-I00 to M.A.N., funded by MICIU/AEI/10.13039/501100011033 and by FEDER; UE. Funds were also provided by the AECC Scientific Foundation (FC_AECC PROYE19073NIE) and Generalitat Valenciana (Prometeo Excellence Programme. CIPROM/2021/45) both to M.A.N., who also acknowledges financial support from Centro de Excelencia Severo Ochoa, grant CEX2021-001165-S, funded by MCIN/AEI/10.13039/501100011033, and support from the Scientific Network Conexión Cáncer funded by CSIC.

INDEX

Abbreviations

ACK	Ammonium-Chloride-Potassium
ADAMTS2	A Disintegrin And Metalloproteinase with Thrombospondin motifs type 2
AR-V7	Androgen Receptor Variant 7
AT1	Alveolar Type 1
Atf3	Activating Transcription Factor 3
ATP	Adenosine Triphosphate
BDNF	Brain-Derived Neurotrophic Factor
Bglap3	Bone Gamma-carboxyglutamate protein 3
Birc5	Baculoviral IAP Repeat Containing 5
BSA	Bovine Serum Albumin
C1qc	Complement C1q C chain
CAF	Cancer-Associated Fibroblasts
CCL2/5	Chemokine (C-C motif) ligand 2/5
Ccnd1/2	Cyclin D1/2
CDH1	Cadherin 1
Cdh5	Cadherin 5
CDK4/6	Cyclin-dependent kinase 4/6
Cdkn2a/b/c	Cyclin-Dependent Kinase Inhibitor 2A/B/C
cDNA	complementary DNA
CFH	Complement Factor H
ChIP-Seq	Chromatin Immunoprecipitation Sequencing
CHUS	Hospital Clínico Universitario de Santiago
cKO	Conditional loss of function
CLAHE	Contrast Limited Adaptive Histogram Equalization
Clu	Clusterin
Col1a1	Collagen type I alpha 1 chain
Col3a1	alpha 1 chain of type III collagen
Comp	Cartilage Oligomeric Matrix Protein
CRE	CRE recombinase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSC	Cancer Stem Cell
CTCs	Circulating tumor cells
Cx3cr1	C-X3-C motif chemokine receptor 1
CXCL12	Chemokine (C-X-C motif) ligand 12
Cxcr2	C-X-C motif chemokine receptor 2
DAPI	4',6-diamidino-2-phenylindole
Dcn	Decorin
DDR1	Discoidin Domain Receptor Tyrosine Kinase 1
DEGs	Differentially Expressed Genes
Des	Desmin
Dll1	Delta-like canonical Notch ligand 1
DMEM	Dulbecco's Modified Eagle Medium
DNA	Deoxyribonucleic Acid
Dusp1	Dual Specificity Phosphatase 1
E/M	Epithelial / Mesenchymal
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EH	Shannon’s Equitability Index

EMP1	Epithelial Membrane Protein 1
EMT	Epithelial to mesenchymal transition
EMT-T1/2	Epithelial to mesenchymal transition Trajectory 1/2
EMT-TF	Epithelial to mesenchymal transition-Transcription factor
Epcam	Epithelial Cell Adhesion Molecule
ER	Estrogen Receptor
ERBB2	Erythroblastic Oncogene B 2
ETS-1	E26 Transformation Specific 1
Ex1/2/3	Exon 1/2/3
EZH2	Enhancer of zeste homolog 2
FACS	Fluorescence-Activated Cell Sorting
FBS	Fetal Bovine Serum
FC	Fold change
FFPE	Formalin-Fixed Paraffin-Embedded
FOLR2	Folate Receptor Beta
Fsp1	Ferroptosis Suppressor Protein 1
Gas6	Growth arrest-specific 6
Gata3	GATA Binding Protein 3
G-CSF	granulocyte-colony stimulating factor
GFP	Green Fluorescent Protein
GREM-1	Gremlin-1
HER2	human epidermal growth factor receptor 2
Higd1b	HIG1 hypoxia inducible domain family member 1B
IDC	Invasive Ductal Carcinoma
iDISCO	immunolabeling-enabled three-dimensional imaging of solvent-cleared organs
IF	Immunofluorescence
IFBB	Immunofluorescence Blocking Buffer
Ifit1	Interferon Induced Protein with Tetratricopeptide Repeats 1
IL-1β	Interleukin-1 beta
IL-6	Interleukin-6
iPSC	Induced Pluripotent Stem Cells
ITH	Intratumor heterogeneity
Kdr	Kinase insert Domain Receptor
Klf4	Krüppel-like factor 4
KMT2D	Lysine Methyltransferase 2D
Krt14	Keratin 14
Lama5	Laminin subunit alpha-5 protein
LGR5	Leucine-rich repeat-containing G-protein-coupled receptor 5
Lrrc15	Leucine Rich Repeat Containing 15
Lum A/B	Luminal A/B
M2	alternatively activated (M2) phenotype
M-CSF	Macrophage Colony-Stimulating Factor
MDCK	Madin-Darby Canine Kidney cells
MET	Mesenchymal to epithelial transition
Mfap5	Microfibrillar-Associated Protein 5
Mgp	Matrix Gla Protein
MHC	Major Histocompatibility Complex
MIC	Metastasis-initiating capacity
miR	microRNA
Mme	Membrane Metalloendopeptidase
Mmp2/3	Matrix Metalloproteinase 2/3
MMTV	Mouse Mammary Tumor Virus
Mx1	Myxovirus Resistance

Myl9	Myosin regulatory light polypeptide 9
NaCl	sodium chloride
NALCN	Sodium Leak Channel, Non-Selective
NETs	Neutrophil Extracellular Traps
NF-κB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
NK	Natural Killer
Nr2f1	Nuclear Receptor Subfamily 2, Group F, Member 1 gene
Nr4a1	Nuclear Receptor Subfamily 4, Group A, Member 1
Nrf2	Nuclear factor erythroid 2-related factor 2
ns	not statistically significant
Ogn	Osteoglycin
OS	Overall survival
OSM	Oncostatin M
P2Y2	Purinergic Receptor P2Y2
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PDAC	Pancreatic Ductal Adenocarcinoma
PDGF-BB	Platelet-Derived Growth Factor-BB
Pdgfrb	Platelet-Derived Growth Factor Receptor Beta
PD-L1	Programmed death-ligand 1
Pecam1	Platelet Endothelial Cell Adhesion Molecule
PFA	Paraformaldehyde
p-H3	phosphorylated Histone H3,
Pi16	Peptidase Inhibitor 16
PINGs	Pro-invasive gene signature
POSTN	Periostin
PRC2	Polycomb Repressive Complex 2
Prrx1	Paired Related Homeobox 1
pSmad2	Phospho-Sma- and Mad- Related protein 2
PT	Primary tumor
PyMT	Polyomavirus middle T antigen
RAS	Rat Sarcoma
Rgs5	Regulator Of G Protein Signaling 5
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
RREB1	Ras-Responsive Element-Binding protein 1
RT	Room Temperature
RT-qPCR	Reverse Transcription quantitative Polymerase Chain Reaction
scDNA-seq	Single-cell DNA sequencing
scRNA-seq	single-cell RNA sequencing
SEM	Standard Error of the Mean
Sfrp2	Secreted Frizzled-Related Protein 2
sh	Short-hairpin
Snail1	Snail family transcriptional repressor 1 protein
snATAC-seq	Single-nucleus Assay for Transposase-Accessible Chromatin with Sequencing
SNN	Shared Nearest Neighbor
SPP1	Secreted Phosphoprotein 1
ST	Spatial Transcriptomics
STINS	Spatial Transcriptomics Derived Invasive Signature
SVD	Singular Value Decomposition
TCF4	Transcription Factor 4
Tead2	TEA domain transcription factor 2

Tgfbr3	Transforming Growth Factor Beta Receptor 3
TGFBRIII	Transforming Growth Factor Beta Receptor 3
TGF-β	Transforming Growth Factor-beta
TGFβ1/2	Transforming Growth Factor-beta 1
THBS1	Thrombospondin 1
TMA	Tissue Microarray
TME	Tumor microenvironment
TNBC	Triple-Negative Breast Cancer
Tnc	Tenascin
TNF-α	Tumor Necrosis Factor-alpha
TNM	Tumor Nodes Metastasis
TRACERx	TRAcking Cancer Evolution through Therapy
Tris-HCL	Tris(hydroxymethyl)aminomethane hydrochloride
TrkB	Tropomyosin Receptor Kinase B
Trp63	Transformation Related Protein 63
TSP1	Thrombospondin-1
TSS	Transcription Start Site
UMAP	Uniform Manifold Approximation
UV	Ultraviolet
Wap	Whey Acidic Protein
Wfdc18	WAP four-disulfide core domain 18
WHO	World Health Organization
Xcl1	Chemokine Chemokine (C motif) ligand 1
YAP	Yes-associated protein
Zeb1	Zinc finger E-box binding homeobox 1
ZPF281	Zinc Finger Protein 28

Abstract

Despite notable progress in cancer therapies that have prolonged median survival, metastatic relapse remains the leading cause of cancer-related mortality. Metastasis is a complex, multistep process only activated in a minority of cancer cells. Its timing is markedly heterogeneous, some patients present with metastases at diagnosis, whereas in others, secondary lesions emerge years after initial treatment. Recent discoveries have advanced our understanding of the determinants of metastatic competence. Nonetheless, most studies emphasize the role of the metastatic microenvironment, leaving the contribution of cancer cell-intrinsic states upon arrival largely unexplored.

In this thesis, through the analysis of human breast cancer samples, preclinical *in vivo* models, and culture systems, we demonstrate that (i) cancer cells depart the primary tumor with a pre-established metastatic potential, and (ii) metastatic outgrowths recapitulate the epithelial–mesenchymal transition (EMT) hierarchies observed in the primary tumor.

Specifically, the integration of spatial and single-cell transcriptomics plus chromatin accessibility data reveals that the plasticity factor Prrx1 is a central regulator of metastatic potential. Beyond its established role in promoting invasion, Prrx1 suppresses proliferation by modulating the expression of cell cycle regulators (Ccnd1/2, Cdkn2a/b/c) and induces a dormancy program through targets such as Gas6, Mme, and Ogn. Importantly, intermediate levels of Prrx1 achieve a critical balance between invasiveness and proliferation, resulting in a hormetic (non-linear) relationship between its expression and metastatic burden. As such, Prrx1 expression within the primary tumor emerges as a determinant of metastatic competence, and its intratumoral heterogeneity provides a framework to understand why some invasive cancer cells enter dormancy while others continue to grow at distant tissues.

In addition, analysis of the EMT status across the primary tumor, circulating tumor cells, and metastatic lesions in spontaneous metastatic breast cancer mouse models confirmed the EMT origin of disseminating cells and revealed their high plasticity potential. Despite exhibiting an active, embryonic-like EMT program, metastatic cells retain the ability to recreate the distribution between embryonic and adult-like EMT programs, observed in the primary tumor.

Finally, we show that the combined expression of invasion and proliferation gene signatures enables robust prognostic stratification of breast cancer patients, underscoring the clinical significance of this phenotypic axis.

A pesar del avance en terapias oncológicas que han mejorado la supervivencia, la metástasis sigue siendo la principal causa de muerte por cáncer. Este proceso, altamente complejo y compuesto por múltiples etapas, solo se activa en una minoría de células tumorales. Su dinámica temporal es muy heterogénea: mientras algunos pacientes presentan metástasis desde el diagnóstico, en otros, éstas aparecen años después del tratamiento inicial. El avance de los últimos años en la identificación de factores que determinan la capacidad metastásica se ha centrado fundamentalmente en el microambiente metastático. Aquí hemos explorado en profundidad la contribución de las propiedades intrínsecas de las células tumorales que consiguen metastatizar.

Mediante el análisis de muestras de pacientes con cáncer de mama, modelos preclínicos *in vivo* y sistemas celulares, demostramos que (i) las células tumorales abandonan el tumor primario con una capacidad metastásica ya establecida, y (ii) que las metástasis reproducen los distintos estados celulares controlados por la transición epitelio-mesénquima (EMT) observada en el tumor primario.

La integración de datos de transcriptómica y de su distribución espacial, junto con la accesibilidad cromatínica todos ellos analizados en miles de células únicas identifica al factor de transcripción Prrx1 como un regulador clave de la diseminación metastásica. Además de su papel conocido en la invasión, Prrx1 inhibe la proliferación mediante la regulación de genes del ciclo celular (Ccnd1/2, Cdkn2a/b/c) e induce un programa de letargo a través de genes como *Gas6*, *Mme* y *Ogn*. Niveles intermedios de Prrx1 permiten un equilibrio crítico entre invasión y proliferación, estableciendo una relación hormética entre su expresión y la carga metastásica. Así, su expresión en el tumor primario emerge como un determinante de su potencial metastásico, y su heterogeneidad intratumoral proporciona un marco conceptual para entender por qué ciertas células invasivas entran en letargo mientras otras proliferan en órganos distantes.

Asimismo, el análisis de estados EMT en tumores primarios, células tumorales circulantes y metástasis en modelos murinos confirma el origen EMT de las células invasivas y revela una notable plasticidad. Aunque adoptan un programa pseudo-embrionario para desprenderse del tumor primario, conservan la capacidad de reproducir en las metástasis la misma distribución de estados EMT y por tanto, de heterogeneidad intrametastásica.

Finalmente, demostramos que la combinación de firmas genéticas de invasión y proliferación permite estratificar de forma robusta el pronóstico de pacientes, evidenciando la relevancia clínica del equilibrio entre estas dos propiedades

Introduction

Metastasis is the process by which cancer cells spread from the primary tumor to distant organs, forming new tumor lesions. Notably, the vast majority of cancer-related deaths are associated with metastatic disease rather than the primary tumor itself[1]. Nevertheless, the precise causes of death remain poorly understood and may also result from complex systemic effects[2].

The metastatic process involves a series of biological events known as the metastatic cascade, which includes local invasion, intravasation into the circulatory system, survival in the bloodstream, extravasation into secondary tissues, and colonization of distant organs[3–5] (Figure 1). This cascade is remarkably inefficient, as only a small fraction of disseminated cells successfully give rise to macrometastases[6], due to multiple constraints throughout the journey from the primary tumor to the metastatic site (Figure 1). The origins of metastatic cells and the specific combination of traits required to acquire metastatic competence and complete the entire cascade remain incompletely understood[7]. At the core of the metastatic cascade we found the epithelial-to-mesenchymal transition (EMT), which over the last three decades, has emerged as a central process underlying not only invasion and metastasis but also cancer cell plasticity, immune evasion and therapeutic resistance[8–10].Unravelling the heterogeneity of this biological process and their consequences is crucial for improving cancer prognosis and developing effective antimetatastic therapies.

In this introduction, we provide an in-depth review of the role of EMT at each step of tumor progression and the metastatic cascade, understanding the challenges that cells face along this process and, consequently, identifying how EMT provides the combination of necessary traits for a successful metastatic colonization.

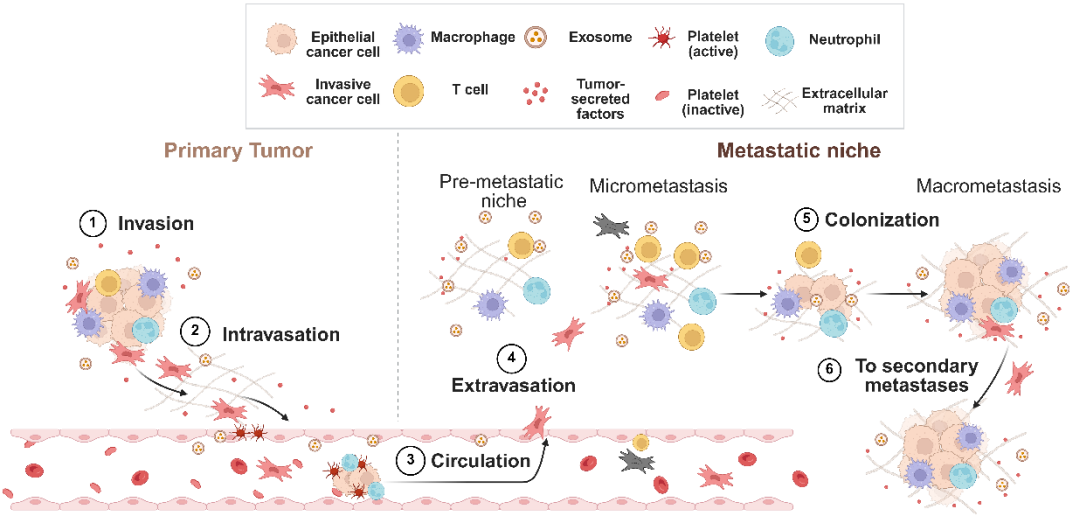


Figure 1. The metastatic cascade. Cancer cells within the primary tumor acquire invasive properties, enabling them to detach from the main tumor mass, break the basement membrane, and invade the surrounding ECM. This process is influenced by the complex tumor microenvironment (TME), including interactions with immune cells (e.g., macrophages, T cells) and stromal components, as well as the release of exosomes and soluble factors. Invasive tumor cells actively penetrate the endothelial wall of nearby blood vessels (or lymphatic vessels), gaining access to the circulatory system. Tumor cells, now designated as circulating tumor cells (CTCs), need to survive transit within the bloodstream to successfully metastasize. They must withstand hemodynamic shear forces and evade immune surveillance. Surviving CTCs arrest at distant capillary beds, adhere to the vascular endothelium, and subsequently migrate across the vessel wall into the parenchyma of the target organ[11]. This stage is often preceded or influenced by the formation of a "pre-metastatic niche", a microenvironment remotely conditioned by factors (e.g., exosomes, secreted proteins) released from the primary tumor, which can recruit immune cells (e.g., macrophages, neutrophils) and modify the local ECM to create a receptive site for incoming CTCs[12]. Successfully extravasated tumor cells must adapt to the foreign microenvironment of the secondary site. This involves surviving initial stresses, potentially entering a period of dormancy, and eventually proliferating to form micrometastases. Continued crosstalk with local stromal and immune cells dictates progression. Growth into clinically detectable macrometastases requires sustained proliferation. Established metastatic lesions can themselves shed cells capable of initiating further metastases at new sites, contributing to widespread disease progression[13].

1. EMT dynamics and heterogeneity at the primary tumor

1.1 EMT overview: regulation, states and programs

Epithelial-to-mesenchymal transition (EMT) is a widespread process in metazoans that enables epithelial cells to acquire mesenchymal features, including front–rear polarity, motility, and resistance to cell death. Originally described in the context of gastrulation and neural crest migration[14–16], EMT has since been implicated in a wide range of physiological and pathological contexts, from organogenesis and tissue repair to fibrosis

and cancer[10,17]. EMT does not always entail a complete transformation into a mesenchymal state, nor does it fully erase epithelial traits, particularly *in vivo*. Rather, it involves an orchestrated, and partial reprogramming of epithelial cells into more plastic and migratory states. In cancer, EMT has emerged as a central mechanism underlying invasion, metastasis, and therapeutic resistance, not merely by enabling motility, but also unlocking a spectrum of cell states endowed with increased adaptability[18].

At the core of the EMT program lies a set of transcription factors, most notably members of the Snail, Twist, Zeb, and Prrx families, first identified for their roles during embryonic development[19–21], and later linked to cancer progression[22–25]. These EMT transcription factors (EMT-TFs) repress the transcription of epithelial genes and activate mesenchymal programs that remodel the cytoskeleton, modify cell–matrix interactions, and disrupt cell–cell adhesion[26,27]. The activity of EMT-TFs is tightly regulated at multiple levels, a topic extensively discussed elsewhere[28,29]. Moreover, the full regulatory landscape has been elegantly summarized recently[30]. EMT-TFs expression is induced by signaling pathways triggered by TGF-β or Wnt receptors, Notch, or receptor tyrosine kinases[28,31] among other, environmental factors such as hypoxia[32,33] and mechanical forces[34,35], or intracellular signals including response to DNA damage[36], oxidative stress[37] or metabolic reprogramming[38] (Figure 2a). In addition, EMT is regulated post-transcriptionally, via microRNAs such as the miR-200 family and miR-34[39], alternative splicing[40], and interactions with chromatin modifiers[41]. Interestingly, systematic analysis of EMT progression across different models, from cell lines treated with TGF-β to mouse embryonic neural crest and breast cancer, reveals a stereotyped sequential activation of EMT-TFs. While Snail1 acts as a pioneer regulator by repressing epithelial features, more potent mesenchymal inducers like Twist and Prrx1 are typically recruited at later EMT stages[42] to induce invasive properties (Figure 2b). In agreement with this, Snail1 has a critical role in the non-invasive partial EMT activated during kidney fibrosis, where “late” EMT-TFs such as Prrx1 are not activated[42]. The sequential recruitment is compatible with the described gene regulatory network involving Snail1 and Prrx1 in development and cancer[43]. Despite the conservation of this sequential program throughout these diverse contexts, it cannot be excluded that a different starting point or temporal hierarchy may occur in different cancer types, including those where the cell of origin is not a *bona fide* epithelial cell, as in pancreatic cancer or melanoma where non-transformed cells already express some EMT-TFs[44–46].

EMT is activated and strictly regulated in both space and time in physiological settings, both in the embryo and in the adult. During embryogenesis, cells undergoing EMT acquire migratory properties and contribute to the formation of the mesoderm, endoderm,

neural crest derivatives, and other lineages, guided by precise developmental cues[26]. In the adult, transient EMT activation plays a critical role in tissue repair for instance, during wound healing[17] (Figure 2c) or heart regeneration[47]. When chronically activated, EMT becomes pathological, promoting disruption of tissue architecture, cell dedifferentiation, stromal remodeling and inflammation, leading to the development of fibrosis, exemplified by renal fibrosis where Snail1 activation in tubular epithelial cells drives a non-invasive inflammatory phenotype[48,49] (Figure 2c).

Although EMT is often depicted as a shift from an epithelial to a mesenchymal state, it is more accurately described as a multidimensional continuum of cell states governed by distinct regulatory modules[27]. Cells can adopt partial EMT configurations, simultaneously expressing epithelial and mesenchymal markers and displaying hybrid phenotypes (Figure 2c). The development of technologies allowing single-cell resolution has led to a better characterization of these hybrid states, which can be stable or metastable within the EMT landscape[27,50]. Furthermore, EMT is often reversible, as cells undergo a mesenchymal-to-epithelial transition (MET). Cells can engage into sequential EMT/MET cycles throughout their history, both during development and in cancer[27], and the downregulation of EMT-TFs is essential for successful metastatic colonization[51,52].

EMT is increasingly recognized not merely as a means for cells to acquire migratory properties, but also as a driver of cellular plasticity, the capacity of a cell to adapt its identity and behavior in response to intrinsic or environmental cues, and now recognized as a hallmark of cancer[53]. EMT confers plasticity to differentiated cells by unlocking terminal transcriptional programs both in cancer and for tissue repair or degeneration[4,42,54–56]. For instance, during reprogramming of fibroblasts to IPSC a MET-like process[57] is required, but a transient activation of the EMT-TF Snail1 increases the efficiency of the process[58]. Furthermore, an ERBB2/YAP-mediated transient EMT activation promotes cardiomyocyte dedifferentiation and cell cycle re-entry leading to heart regeneration in mice[47,59]. In contrast to genetic evolution, which depends on irreversible mutations, phenotypic plasticity enables rapid transitions that enhance cellular fitness under selective pressure. In cancer, EMT-induced plasticity is behind the progression to the metastatic disease, compatible with the absence of metastasis-specific genomic alterations[60–62].

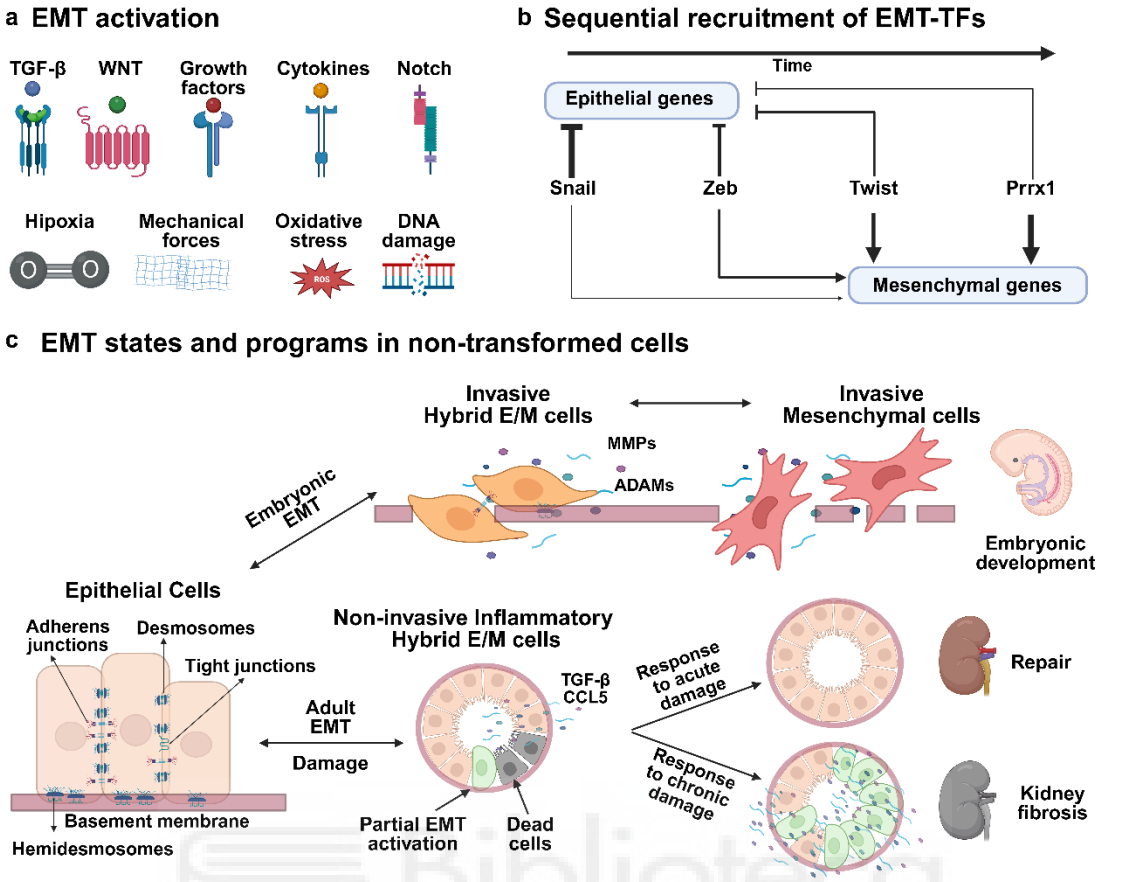


Figure 2. EMT overview: regulation, states and programs. a, EMT is initiated by external cues such as the activation of signaling pathways including those triggered by TGF-β, Wnt, growth factors, cytokines, and Notch ligands[17,27,28], environmental factors such as hypoxia[32,33], mechanical forces[34,35] or intracellular signals such as response to oxidative stress[37] or DNA damage[36]. **b,** The signaling pathways converge in the activation of EMT transcription factors (mainly encoded by *Snail*, *Zeb*, *Twist* and *Prrx* gene families). They are activated in a sequential manner to repress epithelial genes and induce mesenchymal programs. The arrow indicates the direction of EMT-TF recruitment and line width indicates the strength of activation or repression[42,43]. Epithelial repression is mainly driven by Snail and Zeb factors, that cooperate with the main mesenchymal inducers Twist and Prrx. **c,** Non-transformed epithelial cells, with clear apico-basal polarity, strong cell–cell junctions, and attachment to the basal membrane, can undergo two distinct EMT programs. In the embryo, cells can engage into an invasive EMT ranging from different hybrid E/M phenotypes to more mesenchymal states. Adult epithelial cells can activate an inflammatory EMT in response to tissue damage, represented by a non-invasive hybrid E/M state with an inflammatory profile. If the damage is acute, the partial EMT activation is transient and part of the normal repair process, whereas chronic damage leads to organ fibrosis, as exemplified by the kidney[17,48,49].

1.2 EMT in tumor initiation

Since the discovery of the first oncogenes in the 1980's, cancer initiation and evolution has been tightly linked with mutations[63–66]. However, during the last decade, we have learned that mutations are not sufficient to trigger tumor initiation. Indeed, our body can

be understood as a large clonal mosaic[67], and we harbor a multitude of clones with cancer-associated mutations in several tissues[68] (Figure 3). Related to the precise role of driver gene mutations, there is a huge ongoing effort to understand the mechanisms of tumor promotion by many carcinogens, which we now know do not directly cause mutations[69] but rather act as tumor promoters through clonal expansion of cells with pre-existing driver mutations[70] (Figure 3). Surprisingly, a recent study in *Drosophila melanogaster* reveals that tumors can arise after a transient perturbation of the epigenetic machinery promoting an stable cancer fate even in the absence of detectable driver mutations[71].

Lineage tracing studies have shown that not all cell types or cell states are able to give rise to tumors, even when they harbor the same mutation, as exemplified by basal cell carcinoma, where WNT driver mutations will lead to tumor progression only in long-term resident progenitor cells of the interfollicular epidermis and the upper infundibulum[72]. In addition, oncogenes need to coexist with a compatible chromatin cell state to drive tumorigenesis, a state that could result from chronic inflammation or tissue injury, as shown in pancreatic cancer[73]. Even the composition of the ECM can determine the feasibility for transformation, as recently described in skin cancer[74].

Clonal expansion is not enough for tumor formation, as the cells need to bypass additional bottlenecks such as oncogene induced senescence or immune clearance[75] The activation of EMT can endow cancer cells with tumor initiating capacity (TIC)[76,77] while conferring resistance to apoptosis or oncogene-induced senescence[78,79] (Figure 3). Compatible with its pioneer role during EMT activation[42], Snail1 is key in the transition from non-transformed to transformed cells in breast cancer[80]. In agreement with this, cancer cell-specific deletion of Snail1 severely affects tumor initiation and progression in mouse models of breast cancer[42,81,82]. In addition, through multimodal single-cell sequencing of mouse and human samples, a pre-tumoral state has been identified in BRCA-mutated breast cancers characterized by profound epigenomic dysregulation in the absence of driver mutations (Figure 3). This state, marked by Snail1 activation and a partial EMT, acts as a bridge between non-transformed and transformed cells[83].

Considering that EMT activation is an important step during malignization, a question that arises is whether it can provide a permissive or a primed cellular state that cooperates with oncogenic mutations. In fact, a transient epigenetic dysregulation leads to a stable oncogenic program driven by *zfh1*, the fly homologue of *ZEB1*[71]. Recent evidence also suggests that EMT activation contributes to the emergence of genomic instability due to increased chromatin accessibility[84]. Moreover, the ablation of

mesenchymal cancer cell lineages affects genomic evolutionary trajectories, as EMT-induced chromatin remodeling increases genomic instability at very early stages[84] (Figure 3). Altogether, it seems that EMT-induced cell plasticity establishes epigenetic and transcriptional configurations that influence subsequent evolutionary routes. Thus, EMT activation may function as an early "state-setting" event that cooperates with oncogenic mutations and, in some contexts, precedes them creating a permissive cellular landscape that favors the stabilization of tumorigenic programs.

Following that idea, the inflammatory response observed upon EMT activation in response to injury[17] may explain why chronic inflammation is associated with an increase in cancer risk[85] (Figure 3). In addition, environmental factors associated with increased cancer incidence also trigger EMT. For example, nicotine exacerbates the fibrosis induced by silica in a mouse model of pulmonary fibrosis through activation of the STAT3-BDNF-TrkB signaling which leads to the activation of Twist in alveolar type II cells[86] (Figure 3). On the other hand, particulate matters in air pollution induce an EMT-like state through the activation of ETS-1 and NF-κB factors in a lung cancer cell line (Figure 3)[87]. Similarly, systemic conditions such as obesity and natural cycling processes like the estrous cycle, have also been associated with EMT activation[88,89]. Further research in these areas, requiring advanced mouse models and closer integration between oncology and other medical fields, may provide critical insights into tumor initiation and EMT that would help develop novel preventive therapeutic strategies.

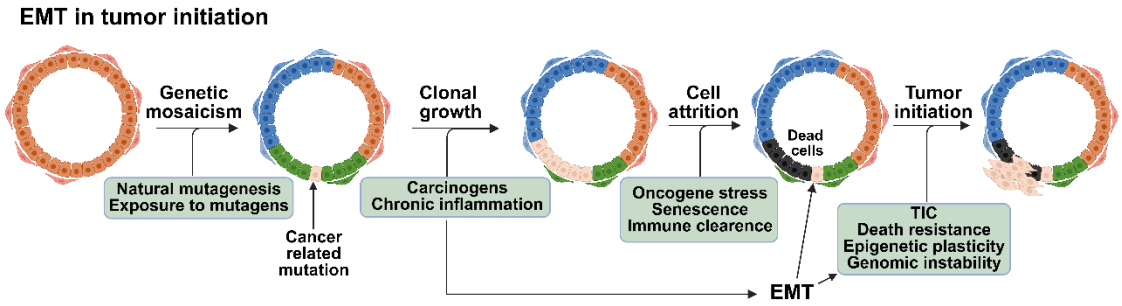


Figure 3. EMT confers tumor initiating properties. a, Tissues are not genetically homogeneous but instead composed of clonal mosaics, resulting from natural mutagenesis during tissue development, homeostasis and ageing, or from exposure to mutagens. Some of these clones may carry cancer-related mutations, but these are not always sufficient to initiate a tumor[67]. Some clones undergo selection and amplification, such as those exposed to carcinogens or inflammation[70]. Many of the amplified cells will die of apoptosis or senescence or will be cleared by the immune system. However, if they activate EMT, acquire resistance to cell death[78,79] together with tumor initiation capacities[76,77] and epigenetic plasticity[83] while increasing genomic instability[84].

1.3 EMT heterogeneity at the primary tumor

Tumors display a high degree of intratumor heterogeneity (ITH), believed to be derived from genetic differences, as distinct clones carrying different mutations coexist within the same tumor[90]. Genetic ITH poses a major challenge to precision oncology[91], as key driver mutations are often not uniformly present across all cancer cells, limiting the efficacy of targeted therapies. The prevailing model of cancer evolution is based on Darwinian selection, where cancer cells accumulate genomic alterations through stochastic processes, undergoing positive or negative selection. Advances in sequencing and biopsy techniques have enabled the use of genetic data to reconstruct tumor phylogenies and track clonal evolution[90]. Consortia like TRACERx are providing new resolution in the description of cancer progression in humans through multiregional sequencing[92]. Yet, despite intensive efforts, no metastasis-specific driver mutations have been identified beyond those already present in the primary tumor[60–62], suggesting that genetic heterogeneity alone cannot fully explain metastatic behavior.

Indeed, an additional layer of ITH arises from cellular plasticity, where genetically identical cells adopt different transcriptomic programs[93], explaining why plasticity was recently included as a new cancer hallmark, given its capacity to enable or modulate other hallmarks[53]. Ongoing efforts to map this transcriptomic heterogeneity include the construction of tumor cell atlases[94,95] and the identification of consensus cellular metaprograms across cancer types[93] (Figure 4). EMT activation in a subset of tumor cells constitutes one such functional source of heterogeneity, long associated with increased invasive properties (Figure 4). Single-cell RNA sequencing (scRNA-seq) has revealed the existence of EMT states across various cancer types, including head and neck cancer[96], melanoma[97], glioblastoma[98], and breast cancer[99,100] confirming its activation *in vivo*, both in animal models and in cancer patients. Notably, intermediate EMT states are associated with increased metastatic potential[99,100]. In breast cancer, a rare basal/stem-like EMT population representing only ~1.5% of tumor cells exhibit disproportionately high metastatic capacity[101] (Figure 4).

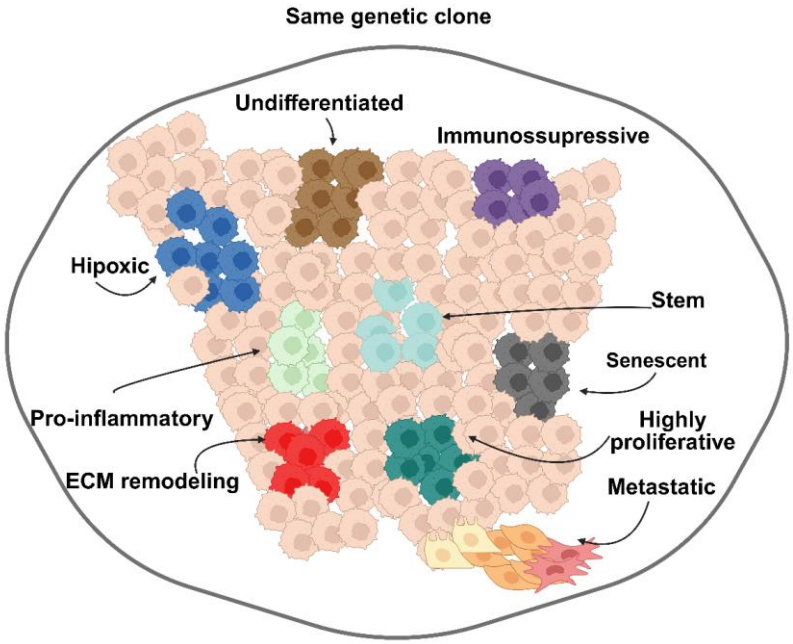


Figure 4. Non genetic intratumor heterogeneity. Recent single-cell studies show that tumors contain diverse cell states not explained by genetic variation. This heterogeneity results from intrinsic and extrinsic factors and reflects cancer cell plasticity. Distinct transcriptomic programs confer different functions, including proliferation, cytokine secretion, extracellular matrix production, hypoxic responses, immune evasion, and, in some cancers, invasion and metastasis often linked to EMT-like processes. These states are not always mutually exclusive, and multiple traits may coexist within the same transcriptomic program.

EMT-driven heterogeneity is observed *in vivo* as a range of cell plastic states distributed along the epithelial-mesenchymal spectrum. Different markers distinguish cell states along the EMT spectrum bearing different metastatic potential in mouse models of squamous cell carcinoma and breast cancer[100]. In addition, different epigenetic regulators like PRC2 and KMT2D govern different steps towards a more mesenchymal state[102].

While those identified trajectories are clearly linked with invasion and metastasis, as part of my PhD research, I contributed to the discovery that, in addition, cancer cells can activate a non-invasive inflammatory EMT, similar to that activated in response to tissue damage in the adult[42] (Figure 5a). The two distinct EMT programs, identified through connectivity (Figure 5b) and trajectory analysis (Figure 5c) of scRNA-seq data, are governed by different EMT transcription factors, and the corresponding cells occupy distinct spatial niches within the primary tumor. Cells engaged in the invasion-related EMT program are enriched at the tumor–stroma interface[42] (Figure 5a). This invasive trajectory is implemented through the sequential activation of EMT-TFs ending with Prrx1 positive cells, described as the metastatic population in melanoma[103] (Figure 5a).

Interestingly, cells in this program dedifferentiate acquiring markers of progenitors of the mammary gland, like those observed during embryonic development (Figure 5a).

In contrast, the EMT inflammatory program, driven by Snail1, is activated in cells that localize internally in the tumor (Figure 5a), and are associated with MHC-II+ macrophages[42], suggesting a link with antitumor immunity. In this context, functional analyses demonstrate that depletion of Prrx1, an EMT-TF specific for the invasive EMT trajectory, not only results in a marked reduction in metastatic burden but also leads to shift toward an inflammatory EMT program, characterized by increased infiltration of anti-tumor macrophages[42]. These findings indicate that the two EMT programs activated in cancer are plastic and interdependent[42], further supported by a common bifurcation point (Figure 5b), and further suggest that they are functionally antagonistic: the invasive program is protumorigenic whereas the inflammatory program exerts antitumor effects. While the precise mechanisms underlying this EMT bifurcation in breast cancer remain poorly understood, emerging evidence suggests that differential integration of TGF-β and RAS signaling pathways determines EMT fate, with RREB1 acting as a key transcriptional hub linking EMT transcription factors, particularly Snail1, to fibrogenic gene expression in carcinoma cells[104,105]. Further studies are required to assess whether this plasticity can be therapeutically exploited to bias EMT toward the antitumor inflammatory trajectory, thereby offering potential clinical benefit[10].

EMT-driven intratumoral heterogeneity adds another layer of complexity to tumor evolution, as cancer cells can occupy distinct positions along the epithelial to mesenchymal spectrum. As mentioned, these states reflect the activation of embryonic-like invasive programs or adult-like inflammatory programs in different tumor cell populations (Figure 5a). Moreover, multiple studies have demonstrated functional interactions between epithelial and mesenchymal cells during tumor progression. For example, grafted mesenchymal cells can induce EMT in neighboring epithelial cells thereby enhancing their invasive and metastatic protential[106–108]. In addition, spontaneous models of pancreatic cancer have revealed a dynamic equilibrium between epithelial and mesenchymal tumor cell populations, maintained by a paracrine signaling loop involving GREM1, an EMT inhibitor, secreted by mesenchymal cancer cells, and SPP1, an EMT inducer secreted by the epithelial compartment[109,110]. Disruption of this loop through genetic deletion of GREM1 shifts the balance toward mesenchymal cells and it is accompanied by increased metastatic burden, whereas deletion of SPP1 reduces both mesenchymal cell abundance and metastasis. Together, these findings underscore the cooperation and functional interdependence of cancer cells at different EMT states within the same tumor.

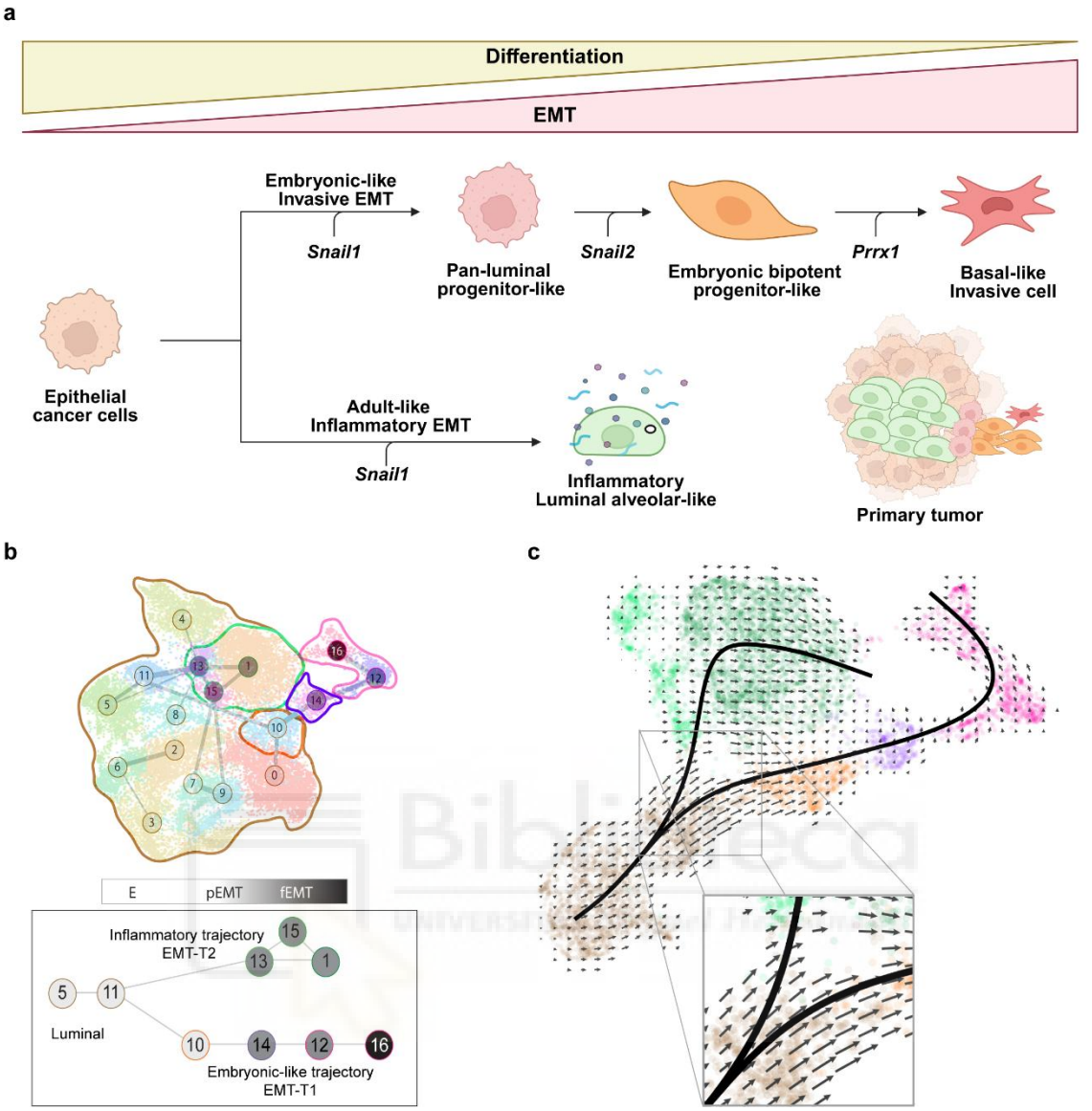


Figure 5. Two distinct epithelial-to-mesenchymal transition programs control invasion and inflammation in segregated tumor cell populations. a, In breast cancer, after oncogenic activation, a subset of cancer cells undergoes EMT that promotes cell dedifferentiation. Dedifferentiated cells acquire progenitor-like phenotypes and engage into one of two distinct EMT trajectories governed by different EMT-TFs[42], each linked to specific functional hallmarks of cancer[53]. In one trajectory, epithelial cancer cells are reprogrammed into an inflammatory-like state linked to antitumor immune infiltration[42]. Other cancer cells continue dedifferentiation into bipotent, embryonic-like mammary progenitors, acquiring a myoepithelial/basal-like fate associated with invasion and metastatic dissemination[42]. **b**, UMAP (cancer cells) and connectivity map showing the two-branch structure of the EMT positive clusters. **c**, RNA velocity analysis of the EMT positive cluster predicting two different EMT trajectories with a common bifurcation point.

1.4 EMT-TME interactions at the primary tumor

Although the EMT is a process undergone by cancer cells, it is highly interconnected with the tumor microenvironment (TME). The activation of EMT is associated with the secretion of multiple molecules that can reprogram the TME. For instance, EMT cancer cells in lung adenocarcinoma secrete platelet-derived growth factor (PDGF)-BB, activating fibroblasts which in turn facilitate cancer cell invasion[111]. Other molecules secreted by cancer cells undergoing EMT such as TGF-β or SPP1[112] have been shown to promote fibroblast activation in cancer and tissue fibrosis[48,49,113].

In turn, cancer-associated fibroblasts (CAFs) can create a pro-EMT feedback loop by secreting EMT-inducing molecules such as IL-6[114] and CXCL12[115] (Figure 6). CAFs can also confer migratory properties to E-cadherin–positive cancer cells through heterotypic E-cadherin/N-cadherin interactions[116]. Additionally, collagen deposition by myofibroblasts may increase matrix stiffness, which further activates EMT through mechanotransduction pathways, primarily via nuclear translocation of Twist[34,35], thus amplifying the EMT–CAF loop.

Interactions with immune populations are also key to understanding cancer cell plasticity and EMT. For instance, interactions between cancer cells and CD163+ (M2) macrophages at the invasive front, involving IL-6 signaling, have been shown to mediate EMT activation[117,118] (Figure 6). In turn, exosomes derived from EMT cancer cells can polarize macrophages toward an M2 phenotype[119]. Direct contact between macrophages and cancer cells was found to be required for the acquisition of an invasive dormant phenotype enabling intravasation[120,121] and Snail1 activation in epithelial cells drives the expression of chemoattractant cytokines such as CCL2/CCL5, inducing the recruitment of macrophages, both in cancer and fibrosis[122,48,49,42]. This creates a positive feedback loop, where EMT cells attract and reprogram immune cells, which enhance EMT in the epithelial compartment. Interestingly, similar EMT- immune cells interactions are also observed during embryonic development. At implantation, the trophoblast induces an M2 phenotype in macrophages, which secrete G-CSF, activating EMT in the trophoblast cells[123].

The capacity of macrophages to induce EMT depends on their origin. In a lung cancer model, tissue-resident macrophages were more efficient in inducing EMT than bone marrow–derived macrophages[124]. However, FOLR2+ tissue-resident macrophages in breast cancer seem to exert the opposite effect, being associated with anti-tumor immunity. Of note, macrophage depletion alters the proportion of intermediate EMT states in mouse models of cancer[100]. In addition, macrophages can also induce

mesenchymal traits in glioblastoma through OSM secretion, directly linking TME interactions to cancer cell plasticity and EMT[125].

Neutrophils also play a prominent role in promoting metastatic traits through EMT promotion, as neutrophil extracellular traps (NETs) can induce a mesenchymal program in breast and colorectal cancer cells[126,127] (Figure 6). Tumor necrosis due to NET accumulation and neutrophil infiltration promotes EMT and metastasis via TGF-β secretion from perinecrotic macrophages[128] (Figure 6). Neutrophils are emerging as key mediators of tumor–host communication, as they have also been linked to enhanced metastasis under stress[129], obesity[130], and post-therapy TME remodelling[131,132].

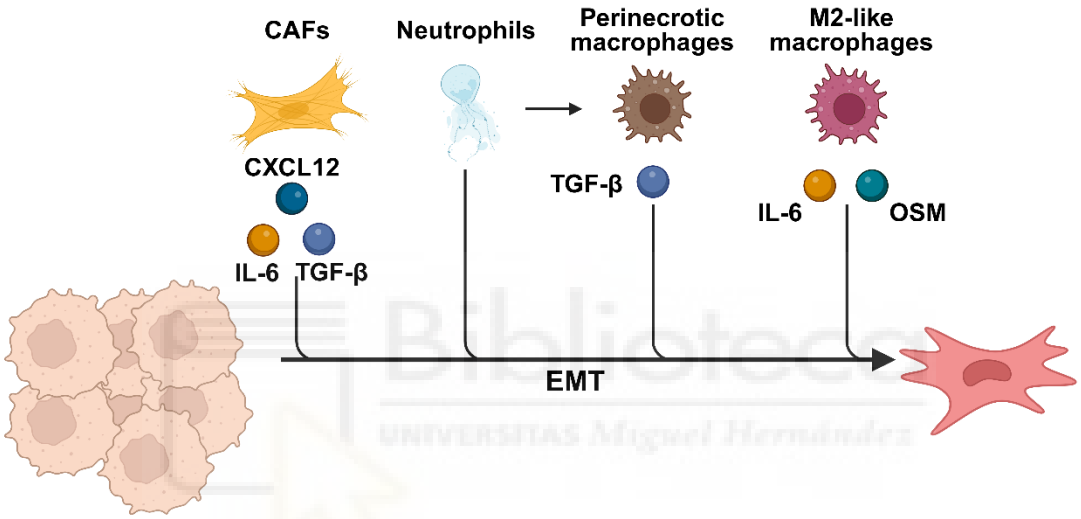


Figure 6. Interactions between cancer cells and the tumor microenvironment during EMT. a, Cancer-associated fibroblasts (CAFs)[48,49,113] promote EMT through the secretion of soluble molecules (CXCL12, IL-6, TGF-β)[114,115]. Neutrophil extracellular traps (NETs) induce EMT in different cancer types[126,127]. In addition, tumor necrosis due to blood vessel occlusion caused by cancer-specific neutrophils, promotes TGF-β secretion by perinecrotic macrophages, also inducing EMT[128]. M2-like macrophages[119] induce EMT in cancer cells via physical interaction or secretion of molecules such as IL-6 and OSM[117,118].

2. Circulating cancer cells

Once cancer cells acquire invasive properties through EMT activation and enter the bloodstream, they are classified as circulating tumor cells (CTCs)[11] (Figure 7a). These cells are widely considered the source of future metastases and represent a valuable resource for understanding the metastatic cascade. As such, CTCs isolated from the pulmonary vein in early-stage non-small cell lung cancer exhibit a higher genomic overlap with future metastases than with the corresponding primary tumor[133].

CTCs are highly heterogeneous, even within a single patient. Subpopulations with distinct metastatic potential have been identified based on surface marker expression[134]. For example, in HER2-negative breast cancer patients, both HER2– and HER2+ CTCs have been detected, and importantly, these subpopulations exhibit plasticity and differential sensitivity to chemotherapy[135]. Whether this represents true plasticity acquired in circulation or merely reflects the presence of a minor HER2+ population within the primary tumor that escaped detection remains an open question with important biological implications for understanding dissemination and drug resistance.

CTCs in the bloodstream need to survive to anoikis, a form of programmed cell death triggered by the loss of cell–matrix interactions (Figure 7a). Beyond its role in promoting invasiveness, EMT also provides resistance to cell death, particularly through the transcription factor SNAIL1[79]. CTCs can be found either as single cells or in clusters, the latter showing a markedly higher metastatic potential (Figure 7a). These clusters are oligoclonal and originate after collective detachment from the primary tumor[136,137]. The increased metastatic potential of clusters is not only due to their multicellular nature, but they also show enhanced resistance to reactive oxygen species[138], which represent a key bottleneck in the metastatic process[139], and benefit from survival advantages conferred by cell–cell adhesion[140] (Figure 7a).

CTC clusters often contain other cell types, forming heterotypic aggregates. Neutrophils within CTC clusters have been associated with increased cancer cell proliferation and enhanced metastatic capacity[141], and interactions with platelets promote metastasis through the activation of EMT[142] and increased vascular permeability mediated by platelet-derived ATP–P2Y2 signalling[143] (Figure 7a).

The existence of epithelial markers such as E-cadherin in CTC clusters, and the observation that E-cadherin depletion impairs CTC survival and metastatic seeding[140], led to the proposal that EMT was not important for cancer cells dissemination. However, CTCs often reside in hybrid or intermediate EMT states, where they co-express epithelial and mesenchymal features, as single cells or in clusters across various tumor types[144–147], and they are associated with enhanced stemness, plasticity, and therapy resistance[148,146,134] (Figure 7a).

The method used to isolate CTCs strongly influences the observed heterogeneity. EpCAM-based approaches such as CellSearch are effective in capturing epithelial-like CTCs but may miss populations undergoing EMT[149]. On the other hand, marker-agnostic systems like Parsortix isolate cells based on physical properties, such as size and deformability, offering a less biased method[150]. However, recent reports suggest that

systems based on specific mechanical properties of the cells may have limited efficiency in capturing mesenchymal CTCs[151], reinforcing the need for complementary isolation strategies.

Finally, recent data suggest that not all epithelial cells found in circulation are necessarily malignant. Mutations in the sodium channel NALCN promote EMT and dissemination of transformed cells but also increase the shedding of non-transformed epithelial cells into the bloodstream[152] (Figure 7a). This observation introduces an additional layer of complexity and raises questions about the cellular identity and clinical relevance of some CTC-like populations.

2.1 CTC Dynamics

We are increasingly understanding how body physiology impacts tumor progression as seen in the effects of chronic stress on increasing metastasis[129] and the differential effects of chemotherapy across the estrous cycle[89]. This extends to CTCs, as their release follows a circadian rhythm[153] (Figure 7b). Disruption of this rhythm has been linked to a modest increase in cancer incidence, mainly based on studies of night-shift workers[154–156], and efforts have been made to incorporate these findings into clinical practice through chronotherapy[157]. In both humans and mouse models, CTC analysis reveals increased release during the rest phase (night for humans and day for rodents) (Figure 7b). Furthermore, CTCs released at night exhibit greater metastatic potential than those released during the active phase (Figure 7b), a difference attributed to the upregulation of mitotic genes in rest-phase CTCs, as detected by scRNA-seq. This effect seems to be mediated by the response to different circadian rhythm hormones such as melatonin, testosterone and glucocorticoids[153].

2.2 Clinical Relevance

CTCs are already established as important biomarkers for precision oncology. Their detection has demonstrated prognostic value in several tumor types, including colorectal, breast, and lung cancer[158,159]. In breast and other cancers, a high number of CTCs correlates with shorter disease-free and overall survival, providing prognostic information regarding therapy response[160–162] (Figure 7c). However, therapeutic patient stratification based on CTC numbers has shown limited success thus far, failing to demonstrate a clear benefit over physician-guided treatment decisions during disease progression[160,163,164].

Regarding the clinical relevance of CTC biological states, some progress has been made. For example, a clinical study associated AR-V7 expression in prostate cancer CTCs with differential outcomes in patients treated with endocrine therapy[165]. However,

a subsequent study using this information to assess the response to cabazitaxel was not informative[166]. There is still substantial development required, particularly in how CTC-derived information can better guide therapeutic decisions. Nevertheless, the clinical importance of CTCs is well recognized, as exemplified by their inclusion in the WHO Classification of Tumors: Breast Tumors[167], and their role in advancing our understanding of the metastatic cascade is undeniable.

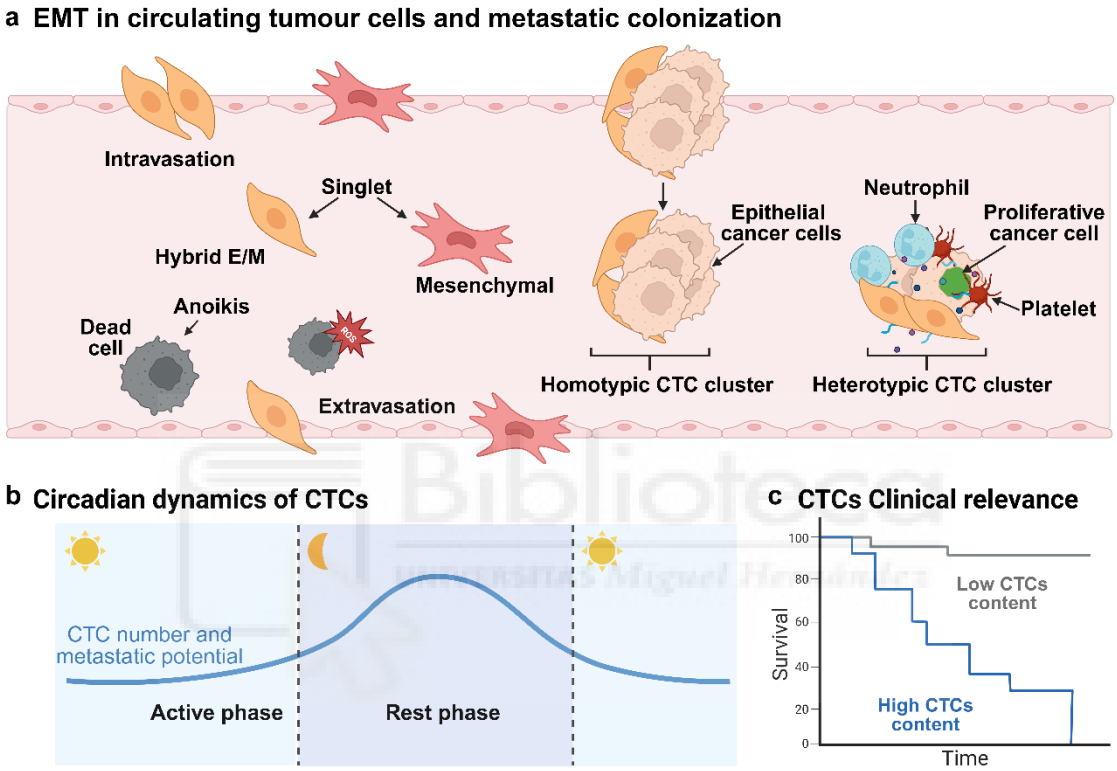


Figure 7. Circulating tumor cells dynamics and clinical relevance. a, Circulating tumor cells (CTCs), shed from primary tumors, must resist anoikis to survive in the bloodstream. Survival is facilitated by epithelial–mesenchymal transition (EMT). CTCs can travel as individual cells or in multicellular clusters. Individual cells have been detected in different EMT states[144–147]. Clusters can contain hybrid E/M cancer cells and exhibit markedly higher metastatic potential[136,137] and often include non-tumor cells such as neutrophils and platelets, which further increase their metastatic capacity by promoting cancer cell proliferation[141,142] (proliferative cancer cell in green). In addition, cells within clusters are more protected than individual cells from reactive oxygen species (ROS)[138] and shear stress, two major bottlenecks encountered in the bloodstream (dead cells in black). **b,** CTC release into the bloodstream follows a cyclic, circadian rhythm, peaking during the organism's rest phase (e.g., nighttime in humans, daytime in rodents). CTCs shed during this rest period possess increased metastatic capability compared to those released during active phases, a phenomenon potentially linked to circadian regulation of gene expression and the effect of circadian hormones on primary tumors**. c**, Clinically, CTC quantification serves as a valuable prognostic biomarker. Patients with higher CTC content (blue line) have significantly shorter overall survival than those with less or undetectable CTCs (grey line), underscoring the prognostic value of CTC burden across multiple cancer types[158,159 160,161,168].

3. Metastatic colonization

3.1 EMT and Metastasis

Following dissemination, cancer cells must reach the metastatic site and resume growth (Figure 1). Upon arrival, they encounter a completely new environment, lacking the protective bulk of the primary tumor and often hostile, as the site has not been fully remodeled to support tumor growth. However, it is now well established that primary tumors can influence future metastatic sites by creating pre-metastatic niches, primarily through the secretion of soluble factors and microvesicles[12], thereby lowering the barriers to metastatic outgrowth (Figure 1). Additionally, systemic physiological changes induced by cancer or other factors such as aging, obesity, or infections can further alter these future metastatic sites[169].

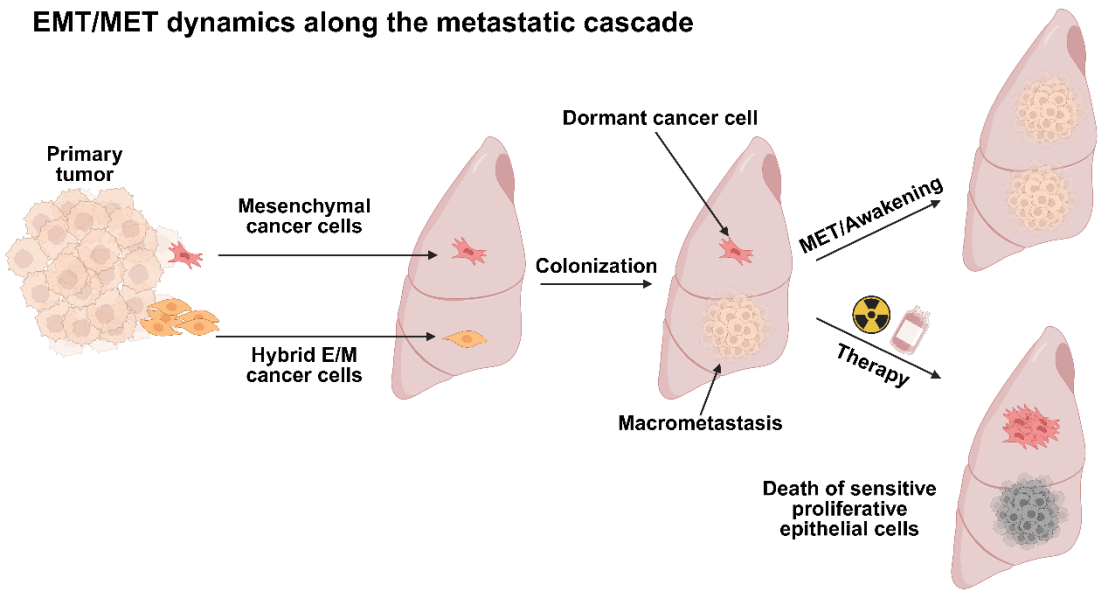
It is now clear that EMT is a major driver of invasion and can also confer cell plasticity and tumor-initiating capacities. However, the correlation between EMT activation and metastasis is not completely linear. Full mesenchymal states, identified by cell surface markers in squamous carcinoma and breast tumors, display high invasive capacities, but also harbor less metastatic potential than E-M hybrids in experimental metastasis assays[51,100,170,171] (Figure 8), and the cells identified as responsible for metastatic relapse, marked by Emp1 expression, also appear to be in a partial EMT state in colorectal cancer[172]. Consistently, low expression or downregulation of EMT-TFs has been identified as a necessary step for metastasis in certain models. For example, PRRX1 downregulation is required for metastatic outgrowth in the lung after tail vein injection with human breast cancer and melanoma cell lines[51,173], and the same applies to Twist in a spontaneous squamous cell carcinoma mouse model[52]. These findings support the idea that MET, or at least a partial reversal of EMT, is required for effective metastatic outgrowth (Figure 8).

Nonetheless, reversion might not be required for cells in a hybrid EMT phenotype, which already confers the highest tumor-initiating properties in breast cancer cells, when compared to epithelial, mesenchymal, or mixed populations[99] (Figure 8). Other cases need plasticity changes, and a clear example is found in Lgr5+ cells in colorectal cancer. While they constitute the cancer stem cell (CSC) pool and are essential for metastatic outgrowth[174], the majority of invasive cells are Lgr5–, only reacquiring Lgr5 expression and tumor-initiating properties upon arrival at the metastatic site[175]. However, lineage-tracing studies suggest that this may be a rare case, as very few metastases arise from highly mesenchymal clones in mouse models[51,176,177], and single-cell DNA sequencing (scDNA-seq) in human tumors to infer clonal lineages suggest that metastases tend to

originate from highly proliferative clones in the primary tumor[178], unlikely to be mesenchymal, which are generally associated with proliferation arrest[79,179] (Figure 8). However, mesenchymal cells contribute more significantly to metastatic outgrowth after treatment[176,180,181] (Figure 8). This difference may result from two non-mutually exclusive mechanisms: either therapy increases the usually small number of fully mesenchymal cells leaving the primary tumor (if still present), or it induces a MET-like process in already-seeded mesenchymal cells that previously lacked outgrowth potential. A detailed analysis of the EMT status in metastases arising from mesenchymal cells after therapy will shed light on why and how these cells can metastasize only after treatment.

Importantly, insights into metastatic colonization mainly come from experimental models using experimental metastasis assays (e.g. tail-vein injection) and homogeneous cell populations (e.g. cancer cell lines), which, although valuable for mechanistic dissection, bypass several steps of the metastatic cascade. These models often fail to fully capture the plasticity and pre-existing heterogeneity of tumor cell populations and may therefore lead to incomplete or even confusing conclusions regarding the role of specific genes or programs in conferring metastatic potential.

Finally, EMT does not only promote invasion or stemness, as it can be considered a driver of somatic cellular plasticity, inducing dedifferentiation, as observed during cancer progression and organ fibrosis[17,42,182,183]. Recent studies in colorectal cancer also highlight how EMT activation is part of a progressive cell plasticity process involving dedifferentiation and the acquisition of non-canonical transcriptional states necessary for metastasis[54,56].



3.2 Dormancy and EMT

In parallel to the changes in EMT status at the metastatic niche, disseminated cells can also engage in a non-proliferative state, commonly termed dormancy[184] (Figure 8). Determining where, when, and how this dormancy program is established, and, more critically, how dormant cells awaken, remains one of the most challenging aspects in metastasis research. A clear clinical example is found in the frequent detection of metastases after surgical removal of the primary tumor. This latency period between primary tumor diagnosis and metastatic relapse is particularly evident in breast cancer patients, where metastases have been detected up to 32 years after initial diagnosis[185]. Notably, this latency varies between breast cancer subtypes, being more common in ER+ breast cancer. ER- breast cancer patients have a higher risk of recurrence and death within the first five years, whereas ER+ cases maintain a relatively steady recurrence rate from five to twenty years[186]. At the other extreme, pancreatic cancer exhibits rapid progression, with approximately 75% of patients developing metastases within the first two years after primary tumor removal[187,188]. However, this does not necessarily mean that pancreatic metastatic cells lack a dormancy phase, as a minimal window for dormancy assessment is missing. Moreover, it is unclear when metastatic cells were seeded, and as previously mentioned, very early dissemination events have been documented in mouse models[189,190], making unclear if short times between diagnostics and metastasis detection are caused by a short dormancy period, very early dissemination or, likely, a combination of both.

Dormancy is primarily characterized by a G1/G0 cell cycle arrest, placing cells in a non- or slow-proliferative state. Most dormant cells exist as single cells or small clusters and are typically negative for cell cycle markers such as Ki67[6,191]. Despite some markers for dormancy have been described, particularly in preclinical studies in breast cancer such as Nr2f1[192] ,Tgfbr3[193], Gas6[194] and Mme[194], we still lack universal methods to detect and identify dormant cells in patients[195]. Whether this non-proliferative state is predetermined before reaching the metastatic site or results from new interactions within the microenvironment remains an open question in metastasis research.

Figure 8. EMT dynamics along the metastatic cascade. Both mesenchymal and hybrid EMT cells leave the primary tumor and arrive at the metastatic site. Hybrid cells are primed for colonization[99,100,177] and can rapidly generate metastasis while mesenchymal cells stay in a low proliferative/dormant state. Additional mechanisms, particularly environmentally driven, can induce a MET in mesenchymal cells and with that an increase in proliferation or “awakening” of dormant cells, giving rise to additional macrometastasis[51,52]. On the other hand, while current therapies can kill proliferating metastatic cells, they can also awaken dormant mesenchymal cells that now can become an important source of metastatic clones[131,176,180,181].

Recent advances in tracing dormant cells reveal that, under naïve conditions, proliferative metastases can arise from cells that never entered dormancy[131], indicating that dormancy is not a required step in the metastatic cascade, and that some cells already possess high metastatic potential before reaching the metastatic niche. Once again, this scenario changes after chemotherapy, when a greater proportion of metastases originate from reawakened dormant cells[131]. These findings raise the question of which pre-established cellular heterogeneity determines whether a cell enters dormancy or not.

Importantly, dormant cancer cells often display a mesenchymal state[189,196], consistent with the previously discussed findings that highly mesenchymal cancer cells have low metastatic potential[51,52,100,177] (Figure 8). Although some studies suggest a link between EMT activation and the acquisition of a dormancy program, including the link between early dissemination and dormancy in breast cancer driven by ZPF281[120,196,197], this relationship remains poorly understood. It is unclear whether dormancy is an EMT-independent program or whether it is tightly linked to the EMT status of the cancer cell upon arrival at the metastatic niche. Moving forward, it will be important to understand how the increased metastatic competence of hybrid E/M states relates to specific dormant states and if there is a connection between MET and dormancy awakening.

3.4 Microenvironmental modulators of EMT at the metastatic site

Despite extensive characterization of microenvironmental EMT-inducing signals at the primary tumor, the mechanisms governing EMT reversion at metastatic sites remain poorly understood[28,29]. The perivascular niche represents the first microenvironment encountered by disseminated cancer cells at secondary sites and serves as a critical hub of interactions. For instance, E-selectin expressed by endothelial cells promotes bone metastatic colonization by inducing MET in breast cancer cells after binding to surface glycoproteins[198] (Figure 9). Interestingly, this interaction is specific to bone metastases, suggesting that the mechanisms that favor MET vary across tissues and may partially explain the well-known cancer-specific organ tropism. In a separate study, responsiveness to WNT signals from lung endothelial cells was found as a key determinant of dormancy in disseminated cancer cells, associated with concurrent EMT activation (Figure 9). Notably, sensitivity to environmental Wnt was dependent on the pre-established epigenetic and EMT cell state[197]. Importantly, the specific state of the vasculature seems to be critical as while endothelial cells secrete Thrombospondin-1

(THSB1) to promote dormancy, the tip cells secrete TGF-β1 and periostin (POSTN) during vascular sprouting to promote tumor growth (Figure 9)[199].

The dichotomy between a predetermined cell state and microenvironment-induced dormancy is an important matter, as it has significant therapeutic implications. Should we target the metastatic or the particular cell state that renders cells dormant niche[195,200]? How appropriate are the experimental approaches we use to model metastasis? A recent study has shown a connection between these dormancy activation mechanisms, showing that the pre-established cell state of a cell arriving at the metastatic site determines how it responds to microenvironmental signals, in this case, endothelial-secreted Wnt[197] (Figure 9). In other words, microenvironmental dormancy inducers were required, but their effects depend on whether the arriving cells are sensitive to them, which appears to be dictated by their previously determined cell state.

Another recent study has shown that interactions between mesenchymal cancer cells and tissue-resident macrophages in the lung through a TGFβII-TGFBRIII axis are key to promoting breast cancer dormancy in the lung[201] (Figure 9). However, this axis appears to operate only in early disseminated cancer cells, not in late disseminated ones, again indicating that the cell state at the time of arrival determines the probability of entering dormancy. All of this brings us back to previously raised questions: How is this heterogeneity created in the first place? Is the differential probability of entering dormancy related to a cell's EMT state? Can this heterogeneity be observed in the primary tumor? If so, could we use that information to better predict patient prognosis?

Immune cells also modulate ECM composition, altering the cues perceived by cancer cells after extravasation. In a breast cancer model, versican deposited by recruited macrophages promoted MET in the lungs through downregulation of pSmad2, thereby enhancing metastatic outgrowth without affecting initial seeding[202] (Figure 9). Intriguingly, cancer cells can also promote their own dormancy through ECM remodeling. They secrete type III collagen which upon interaction with Discoidin domain receptor 1 (DDR1) activates STAT1 signaling in an autocrine manner[203]. Early disseminated cancer cells primed to enter dormancy express high levels of *COL3A1*[196], which has also been implicated in EMT activation in glioma cell lines[204]. These results reinforce the association between dormancy and EMT and suggest that dormancy-inducing signals can be derived from the cancer cells themselves.

Finally, interactions of disseminated cancer cells with non-transformed epithelial cells at the metastatic site also play an important role. Their interaction with AT1+ lung epithelial cells induces Sfrp2 expression in the dormant D2.OR breast cancer line, activating a pro-

survival, non-proliferative partial EMT-like program[205] (Figure 9). Notably, cancer cells also trigger stemness-like programs with EMT features in neighboring epithelial cells, resulting in a supportive niche for metastatic growth, a phenomenon known as “reflected stemness”[206].

The emergence of new technologies, together with advanced experimental models, is key to identifying the diverse mechanisms that induce EMT reversion in different cancer types and organs. A recent *in vivo* CRISPR screen targeting ligand–receptor pairs revealed that PlexinB2 in hepatocytes, after binding to class IV semaphorins expressed by colorectal and pancreatic cancer cells, induces MET through Klf4 activation and drives metastatic colonization[207] (Figure 9). Similar screens combined with spatial transcriptomics, could systematically identify tissue-specific MET drivers while simultaneously profiling microenvironmental changes[208].

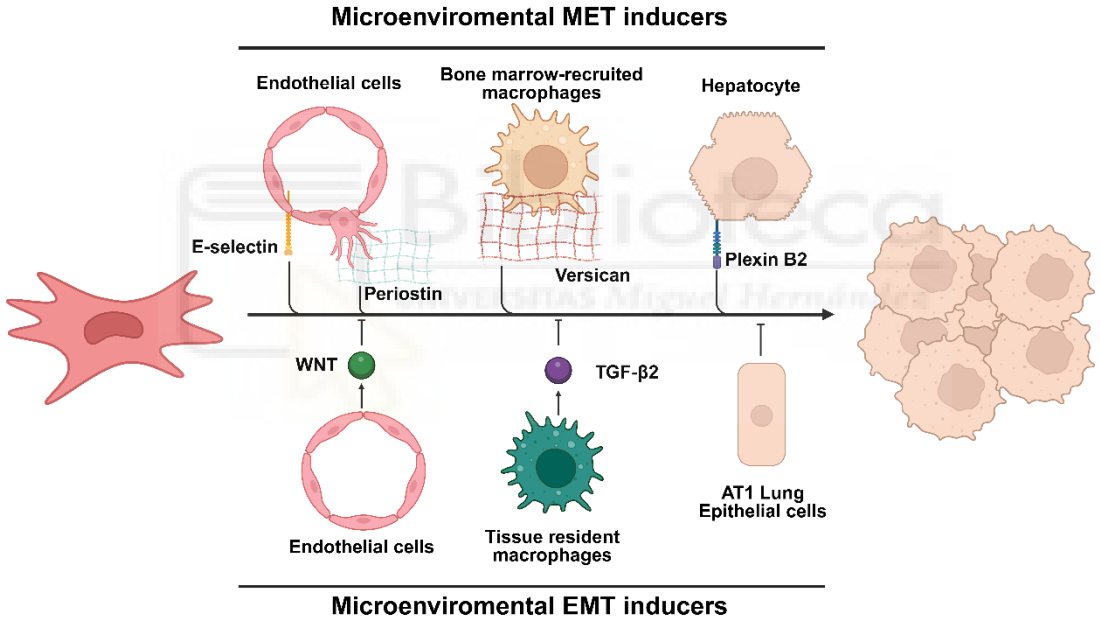


Figure 9. EMT-TME interactions at the metastatic niche. Full mesenchymal cells cannot metastasize and need to revert to an epitheloid phenotype at least partially. E-selectin on the surface of endothelial cells[198] or the secretion of Periostin by tip cells promotes EMT reversion (MET) and metastatic growth, which is prevented by the secretion of WNT and other molecules[197,199]. Bone marrow–recruited macrophages secrete ECM molecules such as versican[202], which can also promote MET and, therefore, metastasis. Simultaneously, lung-resident macrophages can prevent metastatic outgrowth via TGFβ2–TGFBR3 signaling[201]. Non-transformed epithelial cells in the metastatic niche could either favor MET and macrometastasis or retain disseminated cells in an indolent EMT dormant state. Examples are respectively found in hepatocytes expressing PlexinB2 during liver colonization or through the action of AT1 cells after cancer cell dissemination to the lung[205].

3.5 EMT and immune surveillance

The EMT status of cancer cells has a significant impact on immune surveillance and, consequently, on the response to immunotherapy. EMT signatures have been consistently associated with immunosuppression in human datasets[209,210], and several mechanisms have been proposed to explain this connection. For instance, the EMT-TFs Snail1 and Zeb1 directly activate the expression of CD47, a "don't eat me" signal that inhibits macrophage-mediated phagocytosis[211] (Figure 10). These transcription factors also induce the expression of immune checkpoint ligands such as PD-L1, thereby directly impairing T cell function[212,213] (Figure 10). As already mentioned, the loss of Prrx1, a master regulator of the mesenchymal program in breast cancer cells, not only reduces invasion and metastasis but also increases the infiltration by MHC-II$^+$ antitumor macrophages[42].

Another key mechanism of EMT-driven immune evasion involves the suppression of antigen presentation. This can occur through downregulation of the MHC complex[112] or through alterations in the immunoproteasome, the protein complex responsible for generating peptides for MHC presentation[214] (Figure 10). Recently, Tcf4 has been shown to repress the expression of the antigen presentation machinery while simultaneously inducing EMT, that leads to dedifferentiation and mesenchymal transition in melanoma, exemplifying how EMT inducers can coordinate multiple functional axes[215]. Intrinsic features of the mesenchymal state can also render cancer cells more resistant to immune-mediated killing, as EMT confers resistance to pro-death signals such as TNF-α[79] (Figure 10), and the concomitant loss of E-cadherin may impair the formation of effective immune synapses with T cells[216]. In addition, EMT-induced autophagy protects cancer cells from the lysis mediated by T cells[217].

EMT not only confers immunoresistance to cancer cells but also promotes the establishment of an immunosuppressive TME. Snail1 activation enhances the recruitment of regulatory T cells via Tsp1 secretion, leading to the inhibition of cytotoxic CD8+ T cell activity within the tumor[218] (Figure 10). Similarly, mesenchymal cancer cells with high Snail1 expression secrete immunomodulatory factors such as TGF-β1, OPN, and colony stimulating factor M (M-CSF), generating a suppressive niche enriched in M2 macrophages and Tregs, and low for cytotoxic CD8+ T cells[112] (Figure 10). Notably, a small fraction of mesenchymal cells is sufficient to establish this immunosuppressive environment, that also protects neighboring non-EMT cancer cells from immune attack[112]. This EMT-driven immunosuppressive niche is also present during early metastatic colonization, as EMT metastasis-initiating cells secrete IL-1β, which recruits immunosuppressive γδ T cells and reduces CD8+ T cell infiltration (Figure 10), altogether

facilitating the progression to macrometastasis[219]. These data indicate that a link between EMT-associated dormancy at the metastatic niche could be partially explained by immune evasion mechanisms. When exactly these immune evasion mechanisms are activated during EMT, or whether different EMT states display distinct immune evasion trajectories remain open questions in the field.

These immune-related roles of EMT need to be considered in experimental designs, as studies based on immunodeficient mice bypass immune surveillance at early stages and can overestimate the metastatic capacity of epithelial cells. It is still not clear where, along the EMT spectrum, immunosuppressive capabilities are acquired, or whether different hybrid E/M states or EMT trajectories have specific immune-related mechanisms. The use of immunocompetent mouse models capable of recapitulating the full metastatic cascade, from the primary tumor to colonization in distant organs, combined with lineage tracing tools, particularly single-cell DNA barcoding, is essential to understand the dynamics and contributions of distinct EMT states during cancer progression.

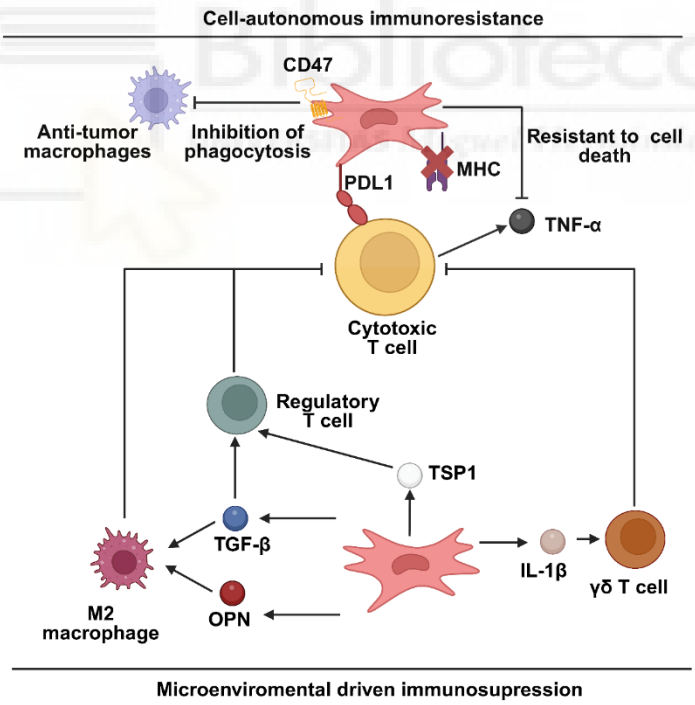


Figure 10. EMT and immunosupression. EMT activation is associated with immunosuppression, acting through both intrinsic and extrinsic mechanisms. Examples of the former include inhibition of phagocytic activity via CD47 expression[211], reduced T cell–mediated killing through expression of PD-L1[212,213], downregulation of antigen-presenting capacities[112], or resistance to cell death induced by TNF-α and other cytotoxic signals[79]. On the other hand, cells undergoing EMT create an immunosuppressive niche through the secretion of multiple molecules that recruit and expand immunosuppressive cells including M2-like macrophages and regulatory T cells[112].

3.5 Awakeness

Despite the low proliferative potential of dormant cancer cells, the dormancy-promoting metastatic environment, and immune surveillance that eliminates cycling cells, many cancer patients eventually develop metastatic outgrowths (Figure 1). However, the mechanisms driving dormancy awakening and the transition to macrometastases remain poorly understood. As previously discussed, awakening often occurs years after diagnosis and, in many cases, following primary tumor removal, making it unlikely that long-term dormancy exit is driven by the emergence of a more metastatic state in the primary tumor. As no clear metastasis-specific mutations have been identified, metastasis outgrowth under these circumstances is more likely linked to global changes in body physiology[2,169].

Current research suggests that external factors affecting global physiology contribute to dormancy reawakening. For example, microenvironmental changes naturally occurring with aging may disrupt dormancy-promoting signals and allow metastatic proliferation[220]. Another major factor is inflammation triggered by injury responses after surgery[221] or obesity[222]. Recent results indicate that viral infections may play a role in dormancy exit[223]. Finally, as discussed above, systemic treatments such as chemotherapy have been shown to awaken dormant cancer cells, thereby promoting metastatic growth[223]. Interestingly, many of these inflammatory effects appear to be mediated by neutrophils, which have recently been described as a link between chronic stress and increased metastasis[129]. A deeper understanding of the interplay between cancer and body physiology will allow combinatorial therapeutical approaches, targeting cancer biology from different angles.

Objectives

Rationale and objectives

Metastasis is a complex, multistep process completed by only a minority of cancer cells. Recent studies have highlighted both intrinsic properties and microenvironmental cues as key determinants of metastatic success. Yet, the specific features that define metastatic cells, and the timing of their acquisition, remain unclear. A plausible scenario is that pre-existing cell states shape future metastatic behavior, either through intrinsic programs or differential sensitivity to cues at distant sites. Epithelial-mesenchymal transition (EMT), a key driver of cellular plasticity and dissemination, is highly dynamic cellular process, through which cancer cells adopt intermediate states along the epithelial-mesenchymal spectrum. These hybrid states often correlate with increased metastatic potential. In our lab, we recently identified EMT-driven intratumoral heterogeneity in breast cancer, uncovering two distinct trajectories, linked to invasion or inflammation, and governed by separate EMT transcription factors. Notably, we found that Prrx1 is essential for delamination from the primary tumor, therefore crucial for the initiation of the metastatic cascade. Yet, high Prrx1 expression prevents metastatic colonization in distant organs in experimental metastasis assays.

General objective

This thesis aims to unveil how invasion and metastatic potential can be uncoupled, and their relationship with Prrx1 levels, with the final aim of uncovering the biological traits underlying successful metastasis and how they are regulated.

Specific Objectives

1. To characterize the expression of PRRX1 in human breast cancer patients and their correlation with metastasis.
2. To generate mouse models with different Prrx1 levels in the cancer cells and study its impact on invasion and metastatic burden.
3. To characterize the EMT status of cancer cells across the metastatic cascade.
4. To assess the clinical relevance of the cellular programs regulated by Prrx1.

Materials and methods

1. Breast Cancer Tissue Microarray (TMA) Analysis

A total of 94 infiltrating grade 3 ductal breast carcinoma (IDC) samples obtained from the archives of the MD Anderson Cancer Center, Madrid, Spain (MD Anderson Foundation Biobank, record number B.0000745, ISCIII National Biobank). Patients underwent surgery between 2010 and 2011, with a mean age at surgery of 57.07 years (range: 44–81 years). According to the TNM Classification staging, 32 tumors were stage I, 31 stage II, and 31 stage III–IV. Histological and immunohistochemical studies were performed on formalin-fixed, paraffin-embedded tissue samples[224]. Two 1-mm cylinders were taken from each tumor block to generate TMA blocks[225]. Inter-observer concordance between pathologists was high, exceeding 90%, in line with standard practice for immunohistochemical evaluation in breast pathology. All cases were independently assessed by two experienced pathologists. In instances where substantial discordance was observed, the immunostaining was repeated and evaluated on full tissue sections rather than on TMA cores, allowing for comprehensive reassessment and consensus scoring. The study followed standard ethical procedures outlined in Spanish regulations (Ley de Investigación Orgánica Biomédica, July 14, 2007, and Real Decreto Biobancos 1716/2011) and was approved by the ethics committee of the MD Anderson Cancer Center Madrid, Spain.

2. Animal Experiments

Mice were fed *ad libitum*. Housing and experimental procedures were conducted in strict compliance with the European Community Council Directive (89/609/EEC) and Spanish legislation. Ethical protocols were approved by the CSIC Ethical Committee and the Animal Welfare Committee of the Institute of Neurosciences. Animals for experiments were selected by genotype, and no randomization or blinding was performed. Animals were housed under SPF conditions at the RMG Animal House (ES-03-119-0001001 SEARMG). Mouse experiments were carried out using the MMTV-PYMT (Tg(MMTV-PyMT)634Mul) model crossed with the Rosa-LSL-tdTomato reporter line (*Gt(ROSA)26Sor*$^{tm14(CAG-tdTomato)Hze}$), expressing tdTomato upon Cre-mediated recombination (Tg(KRT14-cre)1Amc)[42]. The generated line was further crossed with *Prrx1* conditional mutant mice (*Prrx1*^{em1An})[42] to give rise to heterozygous (Tg(MMTV-PyMT)634Mul; Tg(KRT14-cre)1Amc; *Prrx1*^{em1An} /+ named *Prrx1* +/f) and homozygous (Tg(MMTV-PyMT)634Mul; Tg(KRT14-cre)1Amc; *Prrx1*^{em1An} named *Prrx1* f/f) conditional *Prrx1* mutants. Considering that the study focuses on breast cancer, only female PyMT mice were used. Health state, tumor size and burden were monitored weekly. Maximal tumor size/burden was defined as equal to or greater than 1,500 mm^3 per tumor or equal

to or greater than 3,000 mm^3 total tumor burden following the ethical protocols approved by the CSIC Ethical Committee and the Animal Welfare Committee of the Institute of Neurosciences (protocols 2015/VSC/PEA/00211 and 2019/VSC/PEA/0218).

3. Mammary Tumor and Lung Sample and CTCs Collection and Preparation

Tumor and whole lung samples were fixed in 4% PFA overnight at room temperature (RT) for embedding in paraffin or for 4 hours at 4°C for embedding in O.C.T.™ (Sakura). The next day, samples were washed several times and stored in PBS + Azide overnight at 4°C. Pre-fixed samples were embedded in paraffin or O.C.T.™ for further sectioning and collection on SuperFrost Plus microscope slides. Blood samples from 13-14 weeks PyMT mice were obtained via heart puncture in EDTA vacutainers and processed for microfluidic-based CTC capture within 4 h from blood draw. Using the Parsortix Cell Separation System (ANGLE)[150], CTCs were captured in Cell Separation cassettes (GEN3D6.5). Immediately after capture, CTCs were fixed in PFA 4% for 10 minutes in the Parsortix system. For spiking controls, primary culture cells were harvested using trypsin and mixed with 1 mL of blood from control mice. The resulting mixture was processed as described above for blood samples from PyMT animals.

4. Immunofluorescence (IF)

Tumor and Lung Sections: For sample preparation and imaging: 3–5 µm paraffin-embedded sections were dewaxed, and protein epitopes unmasked by immersion in 95°C preheated citrate buffer (pH 6.0) or Tris-EDTA buffer (pH 9.0) for 20 minutes. O.C.T.™ sections (5 µm) were washed three times in PBS for 5 minutes. For IF staining, slides were placed in a humidified chamber and blocked/permeabilized for 1 hour with IF blocking buffer (5% normal goat or donkey serum, 1% BSA, and 0.2% Triton X-100 in PBS). Blocking solution was replaced with the primary antibody diluted in cold IFBB, and the chamber was incubated overnight at 4°C. Slides were washed 3× for 10 minutes in PBS, incubated for 1 hour with secondary antibodies and DAPI, then washed again 3× for 10 minutes in PBS. Slides were mounted using anti-fade mounting medium (DAKO). Primary and secondary antibodies used are shown in Supplementary Table 1 Images were captured using a Leica DMR microscope for brightfield images and an Olympus FV1200 microscope for confocal imaging.

Whole Lungs: TdTomato+ metastases were visualized by IF on whole lungs and cleared following the iDISCO+ protocol[226]. Images were acquired using an UltraMicroscope II (LaVision BioTec), analyzed, and reconstructed in 3D using Vision4D (Arivis) Image Analysis Software. Metastatic burden was assessed using Imaris software (version 9.3.1; BitPlane). The "Surface" function was used for segmentation (tdTomato signal), and volumetric data were extracted. To avoid false positives, a detection cutoff of 90,000 μm^3 was identified as the minimum volume for high-confidence object identification.

Circulating tumor cells and primary cultures: A modified Parsortix protocol was used to stain cassettes containing fixed circulating tumor cells. Cells were blocked/permeabilized for 1 hour with IF blocking buffer (3,5% BSA, and 0.2% Triton X-100 in PBS). Blocking solution was replaced with the primary antibody diluted in cold IFBB, and the chamber was incubated for 4 hours at RT. Cassettes were washed in PBS and incubated for 1 hour with secondary antibodies and DAPI, then washed again in PBS. Images were captured using a Leica DMR microscope.

5. Patient clustering based on PRRX1 expression

To quantify PRRX1 expression of patients, we computed two metrics for each tumor sample: percentage and mean intensity of PRRX1 expression in cancer cells. These values were extracted from single-cell segmentation of IF data using CellProfiler and CellAnalyst softwares[227,228]. The resulting two-dimensional feature space was standardized using StandardScaler from the sklearn.preprocessing module. Unsupervised clustering was then performed using the K-means algorithm (KMeans from sklearn.cluster). The resulting cluster labels were assigned to each patient and used to color-code data points in the scatter plot.

6. *In vivo* scRNAseq lineage tracing data analysis

Preprocessed scRNA-seq files from the M1 primary tumor and corresponding metastatic sites were downloaded from Gene Expression Omnibus (GEO) under accession ID GSE173958[177]. Processed and analyzed lineage data were obtained from Mendeley Data (https://doi.org/10.17632/t98pjcd7t6.1). Single-cell files were imported into the R statistical environment (v4.5.1) and converted into a Seurat object (v5.3.0) using the CreateSeuratObject function; then merged into a single object using the merge function. Only the definitive list of cells provided by the authors for downstream lineage tracing analysis was retained. Pseudo-EMT scores per cell and the cell composition of each

subclone were already provided in the processed lineage tracing files. Pseudo-EMT scores and *Prrx1* expression at the subclone level were calculated as the mean values of the cells comprising each subclone. Unsupervised clustering was performed using the K-means algorithm (KMeans from sklearn.cluster, with n_clusters=3) based on subclonal Prrx1 expression. Finally, the dissemination score, based on Shannon's Equitability Index (EH), was calculated following methodology described[177]. Briefly, Shannon equitability index (E_H) was first calculated as follows:

$$H = -\Sigma\,(pi * ln(pi)).$$

Pi is the sampling normalized proportion at which a subclone is recovered from a harvest site, i.e. if a subclone is only found in the PT, $p_{PT} = 1$, while p = 0 for all other sites. A subclone's H is then used to calculate its E_H.

$$(EH = H\,/\,ln(S))$$

S is the number of distinct harvest sites analyzed (in this case 6).

E_H therefore normalizes H by the number of harvest sites analyzed to exist between 0 and 1, with 1 being completely even dissemination and 0 being no dissemination. For example, a subclone found at only one harvest site is not metastatically aggressive and has an $E_H = 0$.

7. Cancer Cell FACS Sorting and Analysis

Mammary gland carcinomas were collected from 14 to 15-week-old female mice. Harvested tissues were manually minced using sterile scalpels and finely chopped with a McIlwain Tissue Chopper (TED PELLA, INC). Minced samples were incubated in 2.5 ml of digestion buffer (PBS supplemented with 1 Wünsch units of TH Liberase/ml and 25 µg/ml DNase I; Roche) at 37°C for 75 minutes in an Incubator Microplate Shaker (VWR). Tumor fragments were pipetted up and down every 5 minutes to achieve tissue dissociation into single-cell suspensions, and disaggregation was monitored under a microscope. Disaggregated samples were neutralized with 15 mL of breast medium supplemented with 25 µg/ml DNase I and passed through 70 µm and 40 µm filters (BD Falcon). Cells were centrifuged, and pellets were resuspended in 5 mL of red blood cell lysis ACK buffer (Ammonium-Chloride Potassium) for 5 minutes at room temperature (RT) with gentle rotation. After red blood cell removal, cancer cells were neutralized, resuspended in 0.5 ml FACS buffer (D-PBS without calcium and magnesium and 0,1% of FBS) with DAPI per 1M cells, and directly used for FACS to sort Tomato$^+$ cells. Cell sorting was performed using a BD FACSAria III flow cytometer, and 300,000 cancer cells were sorted. The sorted cells were centrifuged at 5,000 rpm for 3 minutes and lysed using the lysis buffer from the *illustra* RNAspin Mini Kit (GE Healthcare) following the

manufacturer's instructions. For immunolabeling, after neutralization of the red blood cell lysis buffer, cells were resuspended in 200 µl of FACS buffer per 1M cells. Fc receptor blocking was performed by adding Palex-BioLegend 156604 (1:50) and incubating for 15 minutes under agitation at 4°C. Fluorescence-labeled antibodies were then added, and cells further incubated for 30 minutes under agitation at 4°C. Finally, cells were resuspended in 200 µl of FACS buffer with DAPI per 1M cells and analyzed using the BD FACSAria III flow cytometer and BD FACSDiva (BD Bioscience) software.

8. Primary Cell Culture

Mammary gland carcinomas were collected from 14 to 15-week-old female mice. Harvested tissues were manually minced using sterile scalpels and finely chopped with a McIlwain Tissue Chopper (TED PELLA, INC). Minced samples were incubated in 2.5 ml of digestion buffer (PBS supplemented with 1 Wünsch units of TH Liberase/ml and 25 µg/ml DNase I; Roche) at 37°C for 75 minutes in an Incubator Microplate Shaker (VWR). Tumor fragments were pipetted up and down every 5 minutes to achieve tissue dissociation into single-cell suspensions, and disaggregation was monitored under a microscope. Disaggregated samples were neutralized with 15 mL of breast medium - DMEM:F12 HAM media (1:1) supplemented with 10% heat inactivated FBS (fetal bovine serum), 10 µg/ml insulin (Roche), 1% Penicillin-Streptomycin (Sigma) and 1% amphotericin (Sigma) supplemented with 25 µg/ml DNase I and passed through 70 µm and 40 µm filters (BD Falcon). Cells were centrifuged, and pellets were resuspended in 5 mL of red blood cell lysis ACK buffer (Ammonium-Chloride Potassium) for 5 minutes at room temperature (RT) with gentle rotation. After red blood cell removal, lysis buffer was neutralized with breast cancer medium without DNase I. Finally, cells were spun down and resuspended in breast cancer medium to wash lysis buffer and cultured in 10-cm culture dishes at 37 °C and 5% CO_2. After 24 hours, medium was refreshed to remove dead cells. Cells were passaged up to a maximum of ten times.

9. Cell Culture

BT-549 (ATCC) and BT-549 shPRRX1 human tumor cell lines were cultured in DMEM (Sigma): F12 HAM (Sigma) media (1:1), supplemented with 10% heat-inactivated FBS (Sigma), 10 µg/ml insulin (Roche), 1% Penicillin-Streptomycin (Sigma), and 1% amphotericin (Sigma). MCDK-NBL2 cells (ATCC) were cultured in DMEM (Sigma) supplemented with 10% heat-inactivated FBS (Sigma), 1% Penicillin-Streptomycin (Sigma) and 1% amphotericin (Sigma). Cells were grown at 37°C and 5% CO_2, with media replaced every two to three days. Cells were passaged when they reached 80–

90% confluency. For RNA extraction, 200K BT-549 or 100K MDCK-NBL2 cells were seeded in six-well plates in complete media and cultured for 48 or 72 hours, respectively.

10. Stable Prrx1 overexpression cell lines

Mouse Prrx1 coding sequences was cloned into the pBABE-neo retroviral vector. Retroviral particles were generated in HEK-293T cells by transfection with Lipofectamine 2000 (Thermo Fisher Scientific) together with the packaging plasmids psPAX2 (RRID:Addgene_12260) and pMD2.G (RRID:Addgene_12259). Viral supernatants were collected 48 h post-transfection and used to infect MDCK cells. Forty-eight hours after infection, MDCK cells expressing either pBABE-neo (empty vector) or pBABE-neo–Prrx1 were selected with G418 (400 µg/mL) for two weeks. Stable clonal cell lines were subsequently generated by limiting dilution and expanded for downstream analyses.

11. Transwell cell migration assay

The top chamber insert (Corning Costar Transwell) was covered with 50 µl of mouse collagen IV (Corning, 50 $\mu g\,ml^{-1}$) and left to dry overnight. The resulting matrix was hydrated with 25 µl of water before seeding the cells. 5×10^4 cells from MDCK-NBL2 clones expressing different levels of *Prrx1* were seeded and allowed to migrate. Cell nuclei at the bottom of the insert were imaged 8 hours after seeding and automatically counted using ImageJ.

12. Growth assay.

A total of 5×10^4 MDCK-NBL2 cells from clones expressing different levels of Prrx1 were seeded into 24-well plates in complete growth medium. Plates were placed in an Incucyte live-cell imaging system (Sartorius), and phase-contrast images were acquired every hour for 72 h. For each condition, three technical replicates were included in three independent experiments. Cell growth was quantified using the Incucyte's built-in confluency analysis pipeline.

13. Total RNA Extraction, cDNA Synthesis, and RT-qPCR

For gene expression analysis, RNA was extracted using the Illustra RNAspin Mini Kit (GE Healthcare). Reverse transcription was performed using the Maxima First-Strand

cDNA Synthesis Kit (ThermoFisher Scientific). RT-qPCR was carried out using Fast SYBR Green Mastermix (Applied Biosystems) on a Step One Plus machine (Applied Biosystems), following the manufacturers' instructions. Relative RNA expression levels (relative fold change) were calculated using the $2^{\wedge-\Delta\Delta Ct}$ formula. Quantitative RT-qPCR primers are listed in Supplementary Table 2.

14. Spatial transcriptomics analysis

Spatial Transcriptomics Panel Design: We designed a MERSCOPE panel of 264 genes (259 genes in the main panel and 6 in the sequential panel) using the Vizgen Gene Panel Design Portal. The gene selection was based on two criteria: Firstly, genes that cover the main cell types and cell states (primarily cancer cell states) identified in previously published scRNA-seq data of PyMT tumors[42]. Secondly, genes involved in general signaling pathways such as Wnt, Hedgehog and Hippo pathways, plus cytokine-cytokine receptor interactions, prioritizing those detected in the scRNA-seq data mentioned above. To avoid optical crowding during transcript decoding, we analyzed public bulk RNA-sequencing data from PyMT tumors and candidate genes with extremely high expression levels were excluded. For each gene, a panel of 30 encoding probes was designed by Vizgen using a proprietary algorithm. The final gene panel also included 30 blank probes serving as negative controls.

Spatial Transcriptomics Sample Preparation and Data Acquisition: Formalin-fixed paraffin-embedded (FFPE) tissue from 3 *Prrx1* +/+ PyMT primary tumor was sectioned to a thickness of 5 µm following standard histology protocols. Sections were floated for 5 minutes in a 45°C water bath, mounted on fiducial bead-coated FFPE MERSCOPE slides (Vizgen, #20400100), dried at room temperature in a fume hood for 1 hour, baked at 55°C for 15 minutes, dried at room temperature for 2 hours, and stored at -20°C overnight. Samples were prepared following the Vizgen MERSCOPE® User Guide for FFPE Tissue Sample Preparation, beginning at the cell boundary staining step. To remove lipids and proteins that interfere with imaging, 5 ml of Clearing Premix (Vizgen #20300003) were mixed with 100 µl of Proteinase K for each sample. The samples were incubated at 47°C in a humidified incubator overnight. Samples in Clearing Solution were photobleached for 3 hours to remove background fluorescence using the MERSCOPE Photobleacher (Vizgen #1010003). Following photobleaching, the samples were incubated overnight at 37°C, with the clearing solution replaced with 200 µL of Clearing Premix containing 50 µL of Proteinase K for an additional 4 hours at 37°C. Finally, the samples were incubated for 48 hours with the customized gene panel described above. Samples were stored in Clearing Solution at 37°C in the incubator prior to imaging for up

to one week. Imaging was performed on the MERSCOPE platform according to the MERSCOPE Instrument User Guide. Each field of view was imaged in seven 1.5 µm-thick z planes at 60× magnification. Images were processed and decoded to identify RNA spots with xyz coordinates and gene assignments using Vizgen's Merlin software.

Cell Segmentation: To improve the MERSCOPE onboard cell segmentation, we re-segmented the data using Cellpose2 v2.2.3[229] implemented in the vizgen post-porcessing tool (VPT) v1.2.2 framework (https://github.com/Vizgen/vizgen-postprocessing). Cyto2 model provided with the Cellpose2 was used for the segmentation. The z_layers was set to 3 and the three image channels i.e. Cellbound2, Cellbound3, and DAPI were mapped to the red, green, and blue channels, respectively. In the preprocessing, gaussian blur (kernel size = 11) was used to smooth the image boundaries. Intensity was normalized using contrast-limited adaptive histogram equalization (CLAHE) with a clip limit of 0.01 and a filter size of 100 × 100 pixels. For the segmentation, DAPI was used as a nuclear channel and object diameter threshold was set to 70 pixels to specify approximate cell size. The diameter threshold was derived from the manual inspection of random areas of MERSCOPE onboard segmentation in cellpose2 visualizer. A flow threshold of 0.95 and a cell probability threshold of -5.5 were applied, with a minimum mask size of 500 pixels to filter the noise. Post-segmentation, polygon smoothing, and cell boundaries simplification were performed with a tolerance of 2 pixels and a smoothing radius of 10 pixels, retaining only polygons with a minimum area of 500 pixels. To harmonize the results, entities were fused with a minimum distance of 1 pixel between them. Finally, partition-transcripts utility from VPT framework was used to obtain the cell by gene matrix using re-segmented data and derive-entity-metadata was used to derive the cell metadata which includes, the cell coordinates, volume, and transcript counts for each cell. The generate-segmentation-metrics was used to generate the report which manually inspected to ensure the quality of re-segmentation of each sample.

Quality control, filtering, and data integration: The cell-by-gene matrix, along with cell metadata, was imported into the Python environment and converted into an AnnData object using Scanpy v1.9.3[230]. Thresholds for gene expression and cell volume were determined by manually inspecting their distributions across each sample. Cells with fewer than 15 gene expression counts, volumes smaller than 50 μm^3, or larger than 5000 μm^3 were removed from downstream analysis which resulted in 1.794 million total number of cells. AnnData objects from three different samples were merged, and library normalization was performed using the scanpy.pp.normalize_total function. Then the

count matrix was log-transformed using the scanpy.pp.log1p function and scaled using the scanpy.pp.scale function (max_value=10). Principal component analysis (PCA) was performed on the scaled data using scanpy.tl.pca function with default parameters. Finally, sample-level data integration was carried out using Harmony[231], through scanpy.external.pp.harmony_integrate function.

Dimensionality reduction, clustering, and cell type annotation: The neighborhood graph was constructed using the top 20 Harmony components with the scanpy.pp.neighbors function (n_neighbors=10). The data were represented in low dimensional space using Uniform Manifold Approximation and Projection (UMAP) using scanpy.tl.umap function (min_dist=0.5 and spread=1.0). The initial round of Leiden clustering was performed using the scanpy.tl.leiden function with a resolution of 0.5. After removing low-quality clusters, a second round of clustering was conducted at a resolution of 1.2, following the same pipeline, starting with re-normalizing the count matrix which resulted in 18 total number of clusters (1.793 million cells). Differential gene expression analysis was carried out using the Wilcoxon test (scanpy.tl.rank_genes_groups), and the expression of *bona fide* marker genes was used to annotate the cell clusters, and expression was represented as dotplots to show their specificity using scanpy.pl.rank_genes_groups_dotplot function.

Identification of invasive edges MERSCOPE data: The cell metadata along with cell type annotation was extracted from the AnnData object and imported into the MERSCOPE visualizer. The invasive areas were drawn manually searching for peripheral tumor edges enriched in *Krt14* expression. Cancer cells were categorized as part of an invasive edge or the bulk of the tumor. The metadata for cancer cells annotated as part of invasive edges or bulk of the tumor was exported and used for the downstream analysis.

Differential gene expression between cancer cells in invasive edge and bulk of the tumor: Differential gene expression analysis was performed to compare the cancer cells located at the invasive edge and the bulk of the tumor. The raw count matrix and associated metadata were imported into the R v.4.3.3 (https://www.r-project.org/) statistical framework and converted into a Seurat (v.5.1.0) object using the CreateSeuratObject function. The cancer cells were subsetted from the Seurat object using subset function and normalized to the library size using the NormalizeData function. The FindMarkers function was used to identify differentially expressed genes between cancer cells at the invasive edge and in the tumor bulk. The resulting statistics were visualized as a volcano plot using ggplot2. Genes upregulated in the invasive edge were identified as the

invasive gene signature. The pathway enrichment analysis was performed for the differentially regulated genes (log2FC > 0.25 and Pvalue < 0.05) using EnrichR[232] against MsigDB and ARCHS4 TFs database. The same gene set was used to create spatial transcriptomics-derived invasive signature (STINS).

15. Single cell RNA-seq analysis

Single-Cell Preparation: Mammary gland carcinomas were collected from 14- to 15-week-old female mice. Harvested tissues were manually minced using sterile scalpels and finely chopped with a McIlwain Tissue Chopper (TED PELLA, INC). Minced samples were incubated in 2.5 mL of digestion buffer (PBS supplemented with 2.5 Wünsch units of TH Liberase/mL and 25 µg/ml DNase I (Roche) at 37°C for 45 minutes in an Incubator Microplate Shaker (VWR). Tumor fragments were pipetted up and down every 5 minutes to achieve tissue dissociation into single-cell suspensions, and disaggregation was monitored under a microscope. Disaggregated samples were neutralized with 15 ml of breast medium supplemented with 25 µg/ml DNase I and passed through 70 µm and 40 µm filters (BD Falcon). Cells were centrifuged, and pellets were resuspended in 5 mL of red blood cell lysis ACK buffer (Ammonium-Chloride Potassium) for 5 minutes at room temperature (RT) with gentle rotation. Following another centrifugation, pellets were resuspended in 5 mL of FACS buffer/5 mM EDTA and passed through a 20 µm filter. To assess single-cell preparation, 20 µl of cell suspension was mixed with an equal volume of Trypan Blue and incubated for 5 minutes to evaluate cell number, viability, and debris content. Cells were adjusted to a concentration of 1,000 cells/µl with cold FACS buffer and used directly for GEM (Gel Bead-In Emulsions) preparation.

Single-Cell GEM and cDNA Library Preparation: 3 *Prrx1 +/+,* 3 *Prrx1 +/f and* 4 *Prrx1 f/f* single-cell preparations were obtained from ten female mice, with each preparation comprising a mix of mammary gland carcinomas from the same mouse. Individual cell encapsulation for single-cell expression profiling and library preparation was performed using the Chromium Single Cell 3' Library & Gel Bead Kit v3.1 (10x Genomics) and the Chromium Controller (10x Genomics). Samples were multiplexed using the Chromium Dual Index Kit TT Set A. Pooled libraries were loaded onto an Illumina Nextseq 2000 machine with a 100 bp read length.

Quality control, filtering, and data integration: scRNA-seq libraries for different genotypes were sequenced to recover 69436 total number of cells. The raw data quality was inspected using FastQC v0.11.9 (Babraham Institute). The reads were aligned to the

mouse genome assembly GRCm38 (mm10; Ensembl reference annotation) and quantified using the 10X Genomics Cell Ranger v6.1.1 pipeline[233] with default parameters. The CellRanger output was imported in the R v.4.3.3 statistical environment and converted to a Seurat (v.5.1.0) object using the CreateSeuratObject function. The low-quality cells were detected based on the number of genes (400-4000) and the percentage of mitochondrial genes (15%) transcripts detected. The putative doublets were detected using scDblFinder v1.8.0[234]. The low-quality cells and doublets were removed from the downstream analysis. The high-quality cells (n=50386) were then subjected to the library normalization using the NormalizeData function with default parameters followed by the high variable gene (HVGs) detection. In this analysis we integrated data from 4 similar stage PyMT tumors *Prrx1* +/+ from[42], which were prepared with 10X single cell 3' v2 chemistry. The top 4000 HVGs were selected for the *Prrx1 f/f* and *Prrx1 +/f* samples, whereas top 6000 HVGs were selected for *Prrx1 +/+* samples. The detection of HVGs was performed separately for each sample. Then the intersection was taken for HVGs calculated for each sample per condition and finally a non-redundant set of genes was taken from the condition specific intersect and used for the sample integration. The gene expression of these HVGs was scaled using the ScaleData function and subjected to the PCA. Finally, the samples were integrated based on the top 30 principal components using Harmony[231] implemented in the RunHarmony Seurat function.

Dimensionality reduction, clustering, and cell type annotation: A Shared Nearest Neighbor (SNN) Graph for the integrated Seurat object was built using FindNeighbors over the top 30 harmony components. The Uniform Manifold Approximation and Projection (UMAP), a nonlinear dimensional reduction was used for the data visualization calculated over the same top 30 harmony components using the RunUMAP function. The single cells were clustered using an SNN modularity optimization-based clustering algorithm implemented in FindClusters function (resolution: 0.035) which resulted in 9 clusters. First the gene counts for all the genes in RNA assay were Log-Normalized using NormalizeData function and the differential gene expression was calculated using FindAllMarkers with default parameters. The expression of *bona fide* markers for each cluster was represented as a dotplot using DotPlot function showing their specificity for a cluster.

Clustering of cancer cell subset: The cancer cell cluster was subsetted and re-clustered to better understand its transcriptional heterogeneity. Sample-specific HVGs were identified using the workflow described above in "Cell quality control, filtering, and data

integration". In the initial round of clustering, we used the top 25 Harmony components with a resolution of 1.0. During this step, clusters with high expression of mitochondrial and ribosomal genes were identified and removed as poor-quality clusters. Subsequently, the cleaned cancer cells were re-integrated based on recalculated HVGs using Harmony. A second round of clustering was performed on the top 25 Harmony components, resulting in the identification of 28 distinct clusters at a resolution of 1.5. To annotate the identified cancer cell clusters, a gene signature scoring approach was applied. Briefly, the top 20 differentially expressed genes for each cancer cell cluster, were taken from [42] and the gene signature score per cell was calculated using the AddModuleScore_UCell function from the UCell v2.0.0 package[235]. These scores, combined with known marker genes describing cancer cell states, were used to annotate the clusters. Scores for dormancy, inflammation, EMT, PINGs and spatial-derived invasive gene signatures were calculated in a similar manner.

Condition-specific differential gene expression analysis: for comparison between conditions (genotypes) and to minimize technical variability introduced by different chemistries, we only used the samples generated with the 10X single cell 3' v3 chemistry. To identify condition-specific differentially expressed genes within target clusters, a pseudobulk approach was employed using the muscat v1.16.0 package[236]. Genes with expression count greater than 5 in more than 3 cells were considered for the comparison. The gene counts for each sample within individual cell states were aggregated using the aggregateData function. Genes with a log2 fold change greater than 0.25 and a p-value less than 0.05 were considered as upregulated. These upregulated genes were then used for pathway enrichment analysis against the MsigDB database using EnrichR[232].

16. Single nucleus ATACseq analysis

Single-Nuclei Preparation: Four mammary gland carcinomas from *Prrx1* +/+ PyMT mice were collected from 14- to 15-week-old female mice and manually minced using sterile scalpels. Minced samples were transferred to a Dounce tissue homogenizer prefilled with Solution D (20 mM Tris-HCl pH 7.5, 0.1% Tween 20, 0.25 M Sucrose, 25 mM KCl, 5 mM $MgCl_2$) and homogenized 8–10 times with rotation. The lysate was passed through 70 µm and 40 µm filters (BD Falcon) and transferred to a 15 mL tube. After centrifugation at 500 × g for 5 minutes at 4°C, the supernatant was removed, and the pellet was resuspended in 4 mL of Solution D. Two milliliters of Optiprep (STEMCELL) were added, and the sample was centrifuged at 1,500 × g for 10 minutes at 4°C. The pellet was resuspended in Wash Buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 2%

BSA, 0.1% Tween 20, 1 mM DTT) and passed through a 40 µm filter. The sample was transferred to a 1.5 mL tube and centrifuged at 500 × g for 5 minutes at 4°C. Nuclei were resuspended in 200 µL of 0.1× lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM $MgCl_2$, 2% BSA, 0.1% Tween 20, 1 mM DTT, 0.1% Nonidet P40, 0.01% Digitonin) and incubated on ice for 5 minutes. Following incubation, 1.3 mL of Wash Buffer was added, and the preparation was passed through a 20 µm filter. After centrifugation at 500 × g for 5 minutes at 4°C, nuclei were resuspended in 100 µL of 1× Nuclei Dilution Buffer (Chromium Next GEM Single Cell ATAC Kit v2). Ten microliters of the preparation were mixed with an equal volume of Trypan Blue, and nuclei number and viability were determined using the automated Countess II cell counter (Invitrogen).

Nucleus Transposition and Single-Nuclei GEM and DNA Library Preparation: Four single-nuclei preparations were obtained from four female mice, with each preparation comprising a mix of mammary gland carcinomas from the same mouse. Prepared nuclei suspensions were transposed following the guidelines of the Chromium Next GEM Single Cell ATAC Kit v2 (10x Genomics) and encapsulated using the 10x Chromium Controller. Samples were multiplexed using the Chromium Single Index Kit N. Pooled libraries were loaded onto an Illumina Nextseq 2000 machine with a 50 bp read length.

Quality control, filtering, and integration process: Libraries were sequenced to obtain 887.61 million reads in total. The quality of the sequenced reads was assessed using FastQC v0.11.9 (Babraham Institute). Sequenced samples were processed using the CellRanger-atac v2.1.0 pipeline (10x Genomics) and aligned to the GRCm38 (mm10) mouse reference genome (Ensembl annotation v99). MACS2 v2.2.9 was used for peak calling with default parameters. We detected a total of 20395 cells with 263346 non-overlapping peaks. Downstream analyses, such as quality control, filtering, normalization, integration, clustering and visualization was performed using the R package Seurat v5.0.1[237] and Signac v1.12.0[238]. The appropriate QC matrices were calculated for each sample separately as per the standard Signac guidelines. The sample-specific putative doublets were predicted using AMULET v1.1[239] and filtered out from the subsequent analysis. We retained only high-quality cells (n=16546) based on number of fragments falling in peak regions (>3500), percentage of reads in peaks (>30%), and TSS enrichment (>2). The selected threshold was derived by manual inspection of the data distribution. To integrate the different samples an integration workflow was followed using high quality cells. Briefly, the four *Prrx1* +/+ samples were merged using merge function provided by Seurat and then Term-Frequency Inverse-Document-Frequency (TFIDF) was calculated for the merged object. FindTopFeatures

was used to select the top features with min.cutoff = 20. The Singular Value Decomposition (SVD) was performed on the selected top features. Harmony was used to perform the integration in SVD space. The first component was not considered for any downstream analysis as it is highly correlated with the sequencing depth in snATAC-seq data[238].

Dimensionality reduction and clustering: A Shared Nearest Neighbor (SNN) Graph was built based on the harmony components using FindNeighbors function with top 25 components, and cells were clustered by a SNN modularity optimization-based clustering algorithm implemented in FindClusters function with resolution 0.15, this resulted in 6 clusters.

Imputation of gene expression and cluster annotation: To impute the gene expression matrix for snATAC-seq assay we employed the reference mapping approach previously described[240]. First, the GeneActivity function was used to calculate gene activity score per cell in the sATAC-seq assay for the high variable genes used in scRNA-seq data. The gene activity score represents the count of fragments located within the 2 kb upstream region from the TSS of a gene. Next, the canonical correlation analysis (CCA) based anchors were identified between the scRNA-seq and snATAC-seq assay using FindTransferAnchors function. The scRNA-seq assay was used as a reference for transferring the gene expression matrix to snATAC-seq assay using TransferData function. The imputed gene expression values were visualized on UMAP using FeaturePlot function to show the specificity of the markers. Additionally, the genomic snapshots of open chromatin region around selected markers were visualized using CoveragePlot function.

Clustering of the cancer cell subset: To explore cancer cell heterogeneity at the open chromatin level, we subset cancer cells based on imputed tdTomato expression (also supported by other cancer cell markers). We re-integrated the data as explained in “Quality control, integration, dimensionality reduction and clustering” of snATAC-seq data analysis. Then, the neighbor search was performed based on the top 20 harmony components and the FindClusters function was used with the resolution 1.0 to cluster the cancer cells into 16 distinct clusters. Next, we used similar gene signature-based approach as described in “Clustering of cancer cell subset” of scRNA-seq data analysis to annotate these clusters. The scores were calculated using imputed gene expression values for in snATAC-seq assay.

Motif enrichment analysis of the invasive cluster: The CORE collection of motifs for the vertebrates deposited in JASPAR database was accessed using JASPAR2024 v0.99.6 R package[241]. The motif information was added to the Signac object using AddMotifs function and the motif activities was calculated per cell using chromVAR[242] implemented in RunChromVAR function. Finally, the differential motif activity was calculated between invasive cluster (cluster15) and the rest of the cancer cell clusters using FindAllMarkers with mean.fxn = rowMeans and max.cells.per.ident = 500.

Prrx1 *binding sites in genes of interest:* To check whether Prrx1 could directly bind the promoter regions of the candidate genes for dormancy, cyclins and cyclin inhibitors in the invasive cluster, we first identified open (accessible) regions in this cluster (cluster 15) using the AccessiblePeaks function. These accessible regions were then annotated to the nearest gene using the annotatePeak function from the ChIPseeker v1.38.0 package[243]. Promotor peaks associated with cyclins, cyclin inhibitors, and dormancy-related genes were analyzed for the presence of Prrx1 binding motif (JASPAR Database motif accession: MA0716.2), using the FIMO v5.5.7[244] motif search tool from the MEME suite. Coverage plots for these binding sites were generated using the CoveragePlot function.

17. *In silico* analysis of human cancer cell lines

The classification of 72 breast cancer cell lines along the epithelial (E) to mesenchymal (M) axis was taken from[42]. Briefly, the enrichment of epithelial (E) and mesenchymal (M) components were computed using epithelial and mesenchymal gene signatures[245] in bulk gene expression data of breast cancer cell lines[246]. The EMT-TFs were excluded from the Mesenchymal signature to avoid biased correlations in the subsequent analyses. Additional scores such as PING[42], Dormancy[194], Proliferation (https://reactome.org/content/detail/R-HSA-69278), and EMT (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp) signatures together with expression of cyclins, cyclin inhibitors, dormancy markers and *PRRX1* were calculated in a similar fashion.

18. PRRX1 ChIP-Seq Data analysis

The raw data for PRRX1 ChIP-Seq in HS578, MDAMB157, and MDAMB436 cell lines were downloaded from Gene Expression Omnibus (GEO) database deposited under GSE202767 accession id. The raw data were filtered using Trimmomatic v0.39 tool with default parameters. The obtained high-quality reads were then aligned to the human

reference genome (hg38) using bowtie v1.0.0 with default parameters. The aligned reads were sorted using samtools v1.11.0, and Picard MarkDuplicates v2.23.0 (https://broadinstitute.github.io/picard/) was used to identify and mark potential duplicated reads. To generate genomic tracks for PRRX1 ChIP-Seq signal for the selected genes, deeptools bamCoverage v3.5.2 was used to convert the bam files into normalized coverage files (BigWig). Finally, the BigWig files were visualized in Integrated Genome Visualizer (IGV) v2.16.0 and genomic snapshots were taken for the selected genes.

19. Clinical analysis of gene signature in METABRIC data

The gene expression data for METABRIC cohort[247–249] along with the clinical parameters was downloaded from cBioPortal data. The median of log normalized values was taken for the proliferation (Cell_Cycle_Mitotic_R-HSA-69278) and PINGs (Pro-Invasive genes) signatures. The patients were categorized based on quartiles, into high (75%) and low (25%) groups combining proliferation (R-HSA-69278) and invasion scores (PINGs). Finally, GraphPad Prism 8 was used to perform overall survival (OS) analysis, and p values were calculated using the Gehan-Breslow-Wilcoxon test.

20. Quantification and statistical analysis

Statistical analyses were performed in R with RStudio, or in GraphPad Prism (GraphPad 8.0 software). Mann-Whitney or unpaired t-tests were used for between-group comparisons, with the level of significance defined as $p < 0.05$. Data are presented as means ± SEM. Boxplots show median, upper and lower quartile; whiskers show upper and lower extremes. Figure asterisks correlate to *p*-value thresholds: ns = > 0.05; * = ≤ 0.05; ** = ≤ 0.01; *** = ≤ 0.001; **** = ≤ 0.0001. For overall survival (OS) curves, *p*-values were obtained by using the Gehan-Breslow-Wilcoxon test.

21. Supplementary tables

Table S1. Antibodies

Primary antibodies

Antibody	Reference, clone	Dilution
IF TMA: Keratin 5/6	DAKO M7237	1_50
IF TMA: Keratin 19	DAKO M0888	1_50
IF Prrx1	Kind gift from Prof. TANAKA (IMP Vienna BioCenter)	1_200
IF Idisco: tdTomato	ROCKLAND 600-401-379	1_4000
IF: tdTomato	Life technologies MA515257, Clone RF5R	1_500
IF: Keratin 14	Palex 454928	1_5000 (Sections) 1_500 (CTCs)
IF: Laminin	Sigma L9393	1_200
IF: p-Histone 3	UPTASE/Millipore-Sigma 06-570	1_200
FACS and IF: Epcam	PALEX-BIOLEGEND - 118225	1_50
FACS: CD90	Invitrogen - 47-0900-82	1_50
IF: cJun	Cell Signaling #9165 Clone 60A8	1_1000 (Sections) 1_500 (CTCs)
IF: p63	Genetex GTX102425, Polyclonal	1_1000
IF: Mafb	Merck HPA005653, Polyclonal	1_500
IF: Mme	Proteintech 10302-1-AP	1_200
IF: CD45- FITC	Invitrogen # 11-0451-82, Clone 30-F11	1_50

Secondary antibodies

Antibody	Reference, clone	Dilution
IF: Goat anti- Rat IgG(H+L) Alexa Fluor 488	Invitrogen A11006	1_500
IF: Goat anti-Rabbit IgG(H+L) Alexa Fluor 488	Invitrogen A-11008	1_500
IF: Goat anti-Rabbit IgG(H+L) Alexa Fluor 568	Invitrogen A-11011	1_500
IF: Goat anti-Mouse IgG(H+L) Alexa Fluor 568	Invitrogen A-11004	1_500
IF: Goat anti-Chicken IgG (H+L) Alexa Fluor 568	Invitrogen A-11041	1_500
IF: Goat anti-Rabbit IgG (H+L) Alexa Fluor 647	Invitrogen A-21244	1_500
IF: Goat anti-Mouse IgG(H+L) Alexa Fluor 647	Invitrogen A-21235	1_500
IF: Goat anti-Chicken IgG (H+L) Alexa Fluor 647	Invitrogen A-32933	1_500

Table S2. RT-qPCR oligonucleotides

	Forward (5´to 3´)	Reverse (3´to 5´)
Mm-*Tbp*	CCTTGTACCCTTCACCAATGAC	ACAGCCAAGATTCACGGTAGA
Mm-*Prrx1*	TCAGCAGGACAATGACCAGT	TGCGAGATCTTCTCGAACAA
Mm-Prrx1 Long (cKO)	TTGTTCGAGAAGATCTCGCA	AAGTAGCCATGGCGCTGTA
Mm-Prrx1 Short (cKO)	TTGTTCGAGAAGATCTCGCA	GCCCCTCGTGTAAACAACAT
Mm-Prrx1 Long (OE)	TGCAGGTGTGGTTTCAGAAC	AAGTAGCCATGGCGCTGTA
Hs-*h18SR*	CACGGACAGGATTGACAGATT	GCCAGAGTCTCGTTCGTTATC
Hs-Prrx1	CTGATGCTTTTGTGCGAGAA	ACTTGGCTCTTCGGTTCTGA
Hs-*CCND1*	GGCGGAGGAGAACAAACAGA	GGAGGGCGGATTGGAAATGA
Hs-*CDKN2A*	ACCAGAGGCAGTAACCATGC	TTCCCGAGGTTTCTCAGAGC
Hs-*CDKN2B*	CAACGGAGTCAACCGTTTCG	GGTGAGAGTGGCAGGGTCTG
Hs-*CDKN2C*	GGATTTGGAAGGACTGCGCT	CGCATCATGAATGACAGCGA
Hs-*CFH*	ACCCTTACAGGAGGAAATGTGT	TGGTCCATCCATCTGTGTCA
Hs-*GAS6*	CATCAACCATGGCATGTGGC	CCGCGATTTTCATGACAGCA
Hs-*MME*	TCCCGTTACGGCAACTTTGA	GCTTTTGCTTTCTGCACTGCT
Hs-*OGN*	CCATAATGCCCTGGAATCCGTG	CAGGCGTATCTCTTCAATGCGG
Dog-TBP	CAACAGCTTGCCGCCTTATG	CTGTGAGTCAGTCCTGTGC
Dog-CCND1	AGCTGTGCATCTACACTGACAA	ATGAAGTCGTGTGGGGTCAT
Dog-CCND2	TTGTCTCTGATCCGCAAGCA	ATCGATGGCGGATACATGGC
Dog-CDKN2B	ATCCCAACGGAGTCAACAGC	GTCGGCACAGTTGGGGTC
Dog-CDKN2C	GGATTTGGAAGGACTGCGCT	TGACAGCGAAACCAGTTCGG
Dog-GAS6	GCTACGCGCTCTCCCATGAT	CCGCAGGCAGCTCCATTT
Dog-MME	GAGACTAGCTCCCGTTACAGT	GTTTTTGCTTTCTGCACTGCT

Table S3. Gene composition of spatial transcriptomics panel

Acta2	Col1a1	H2-Eb1	Krt14	Pdgfa	Tead2
Actn1	Col4a1	H2-K1	Krt15	Pdgfra	Tead3
Adgre1	Comp	H2-Q6	Krt17	Pdgfrb	Tgfb1
Aldh1a3	Cpt1a	H2-T10	Krt5	Pecam1	Tgfb2
Anxa1	Cpt2	Havcr2	Krt7	Pi16	Tgfb3
Arg1	Crabp1	Hes1	Krt8	Plac8	Tgfbr1
Atf3	Crb3	Hey1	Lag3	Pmepa1	Tgfbr2
Axin2	Csf1	Id1	Lama3	Porcn	Tgfbr3
Bglap3	Csf3r	Id2	Lama5	Postn	Tgif1
Bmpr1a	Ctla4	Id3	Lamb3	Pparg	Thbs1
Cav1	Cx3cr1	Ifit1	Lcn2	Ppp1cb	Thy1
Cbx3	Cxcl10	Ifit3	Lef1	Prf1	Tigit
Ccl1	Cxcl11	Ifna2	Lep	Prrx1	Timp1
Ccl2	Cxcl2	Ifnb1	Lgr4	Ptprc	Timp2
Ccl3	Cxcl9	Ifng	Lgr5	Rbpj	Tnc
Ccl4	Dkk3	Ikbkg	Lrp5	Rela	Tnf
Ccl5	Dll1	Il10	Lrp6	Retn	Tnfaip3
Ccl8	Dll3	Il13	Lrrc15	Rgs5	Tnfrsf1b
Ccn3	Dll4	Il17a	Ltbp2	Rnf43	Top2a
Ccnd1	Dpt	Il17b	Ly6c1	Rspo1	Tox

Ccnd2	Dusp1	Il1a	Ly6c2	S100a6	Trdc
Ccr2	Ecrg4	Il1b	Lyve1	S100a8	Trdv4
Ccr3	Egr1	Il1r1	Mafb	S100a9	Trp63
Ccr7	Epcam	Il1rl1	Map3k1	Saa1	Twist1
Ccr8	Ets2	Il1rn	Mgp	Saa2	Vim
Ccrl2	Fabp5	Il2	Mki67	Serpina3g	Wap
Cd14	Fat1	Il33	Mmp13	Serpine1	Wif1
Cd163	Fcer1g	Il4	Mmp14	Sfrp1	Wnt1
Cd274	Flna	Il4ra	Mmp2	Sfrp2	Wnt2
Cd36	Fn1	Il6	Mmp3	Sfrp4	Wnt5a
Cd3g	Folr2	Irf1	Mx1	Slpi	Wnt5b
Cd4	Fos	Irf2	Myd88	Snai1	Wnt6
Cd44	Fosb	Irf3	Myl9	Snai2	Wnt7a
Cd7	Foxc2	Irf7	Mylk	Sox11	Wnt7b
Cd83	Foxp3	Irf8	Ncr1	Sox9	Wwc1
Cd8a	Fzd1	Irf9	Nfkb1	Sparc	Zeb1
Cdh1	Fzd10	Isg15	Nfkbia	Spi1	Zeb2
Cdh11	Fzd2	Jag1	Nos2	Spp1	Zfp36
Cdh5	Gata3	Jag2	Notch1	Tagln	Znrf3
Cdkn1c	Grhl2	Jun	Notch2	Tcf4	
Chad	Gsk3b	Junb	Notch3	Tcf7	
Cldn4	Gzma	Klf4	Nr4a1	Tcf7l1	
Clec3b	Gzmb	Klrb1b	Nr4a2	Tcf7l2	
Clec4e	Gzmk	Klrb1c	Nr4a3	tdTOMATO	
Clu	H2-Aa	Klrb1f	Pdcd1	Tead1	

Table S4. Human data statistics

Pair comparison	Fisher's exact test test (pvalue)	OR	OR-CI (95%)	RR	RR- CI (95%)
Intermediate vs Negative	0,0095	7,729	1,460 to 38,00	3,097	1,231 to 10,96
High vs Negative	0,4269	0,3155	0,02700 to 1,985	0,7866	0,6278 to 1,321
High vs Intermediate	0,0152	0,04082	0,003284 to 0,4974	0,254	0,07101 to 0,7081

Groups	Median survival
PING High_Prol Low	209
PING Low_Prol Low	184,7
PING Low_Prol High	139,6
PING High_Prol High	71,63

Pair comparison	**Gehan-Breslow-Wilcoxon test (pvalue)**	**Hazard Ratio (Mantel-Haenszel) C/D**	**Hazard Ratio (Mantel-Haenszel) CI(95%) C/D**
PING High_Prol Low vs PING Low_Prol Low	0,2338	0,8765	0,5818 to 1,320
PING High_Prol Low vs PING Low_Prol High	<0,0001	0,6026	0,4495 to 0,8077
PING High_Prol Low vs PING High_Prol High	<0,0001	0,3814	0,2278 to 0,6385
PING Low_Prol Low vs PING Low_Prol High	0,0286	0,7816	0,5438 to 1,123
PING Low_Prol Low vs PING High_Prol High	0,0014	0,5445	0,3248 to 0,9128
PING Low_Prol High vs PING High_Prol High	0,0281	0,6971	0,4469 to 1,087

Results and Discussion

1. Non-linear correlation between PRRX1 levels and metastasis.

As mentioned during the introduction, the analysis of Prrx1 function along the metastatic cascade has led to apparently contradictory findings.

While we had observed that an invasive EMT program driven by Prrx1 is necessary for cancer cells to delaminate from PyMT tumors[42], our previous data also showed that its downregulation was required for cancer cells to metastasize[51]. In addition, and consistent with the latter, we had also shown that high levels of *Prrx1* expression in the primary tumor correlated with lack of metastasis in breast cancer patients[51].

To investigate the relationship between PRRX1 expression in cancer cells and metastatic incidence, we analyzed 94 (grade 3) infiltrating ductal carcinoma (IDC) tumors in an adjuvant setting from patients who underwent surgery between 2010 and 2011. PRRX1 was visualized using immunohistochemistry (IHC) in combination with an anti-keratin antibody cocktail to specifically identify cancer cells. Intratumor heterogeneity and intensity of the PRRX1 signal in cancer cells was evaluated by a pathologist. Based on both the intensity and heterogeneity of PRRX1 expression in cancer cells, tumors were classified into three groups: the majority (82%, with no detectable PRRX1-positive cancer cells) as PRRX1-negative; 10% as PRRX1-intermediate, characterized by a sparse presence of cells expressing intermediate PRRX1 levels; and the remaining 8% as PRRX1-high, containing a substantial number of cancer cells strongly positive for PRRX1 (Figure 11a). To substantiate the existence of these three groups in a quantitative manner, a subset of samples containing all groups was analyzed using image-analysis software on a single-cell basis (see Methods). Unsupervised clustering of the patients' scores based on (i) percentage of PRRX1+ and (ii) intensity of PRRX1 expression in cancer cells confirmed the 3 patient groups and individual assignments (Figure 11b). PRRX1-negative tumors were predominantly classified as Luminal A or B (65,1%), with a smaller fraction corresponding to HER2 (11,6%) or Basal (23,3%). In contrast, tumors with PRRX1-intermediate levels were more evenly distributed among Luminal B (22,2%), HER2 (33,3%), and Basal (44,4%) (Figure 11c). Notably, all PRRX1-high tumors belonged to the Basal subtype. This differential distribution aligned with previous observations. *Prrx1*-deficient PyMT tumors are more differentiated[42], and here we find that the PRRX1-negative subgroup is enriched in luminal subtypes.

To assess whether this patient stratification based on PRRX1 expression patterns and levels has clinical relevance, we compared the metastatic incidence among the groups. Surprisingly, we observed a non-linear correlation between PRRX1 levels and metastatic incidence. The PRRX1-intermediate group exhibited the highest metastatic burden, with approximately 78% of patients developing metastases, compared to 31% in the PRRX1-negative group and only 12% in the PRRX1-high group (Figure 11d). These results suggest that PRRX1 levels, rather than its mere presence or absence, may be a critical determinant of the metastatic potential of cancer cells.

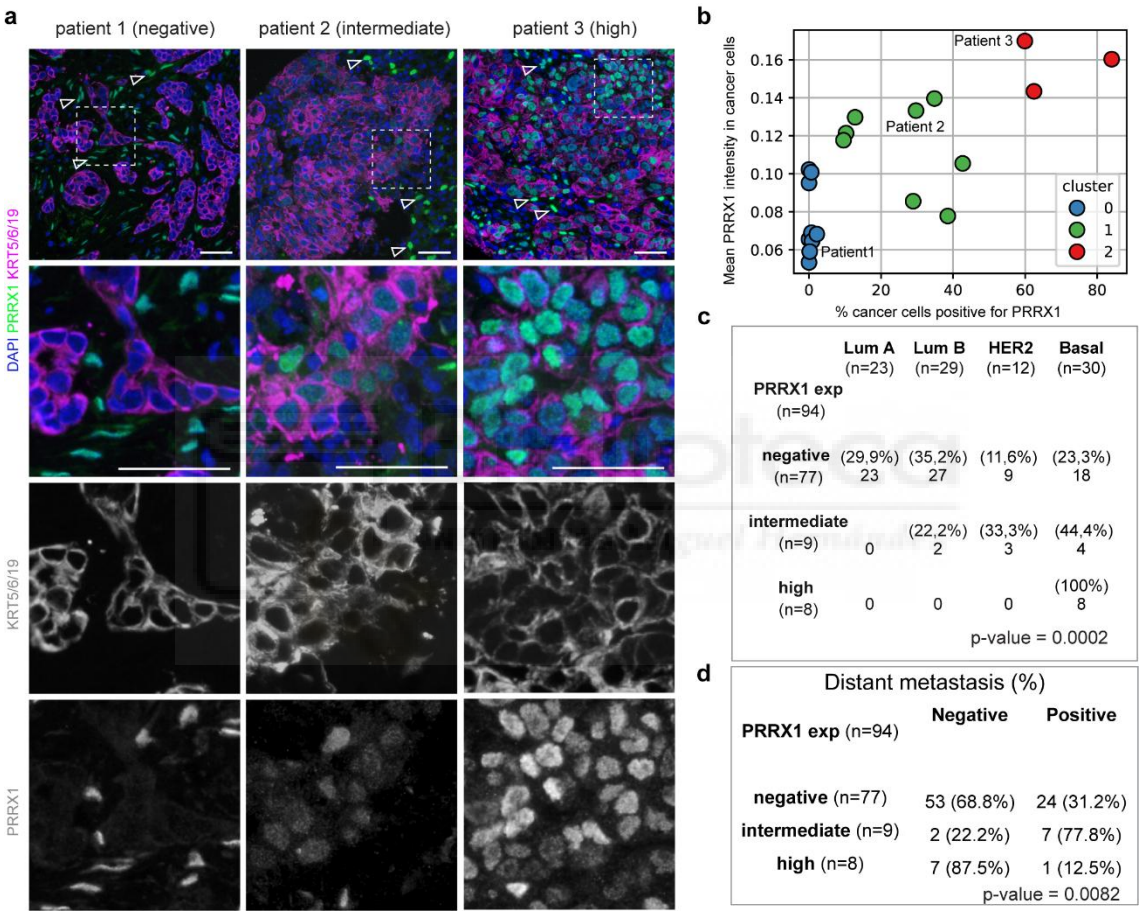


Figure 11. Non-linear correlation between PRRX1 levels and metastatic incidence in breast cancer patients. a, PRRX1 levels in TMAs of breast cancer patients categorized by their expression pattern. Arrowheads indicate stromal PRRX1 expression (not in cancer cells). Scale bars: 25 µm. **b**, Dot plot of patients colored by unsupervised clustering based on PRRX1 positivity and intensity in cancer cells. Labelled patients (1–3) highlight representative cases showed in (a). **c,** Correlation between subgroups of breast patients categorized by PRRX1 expression. Chi-squared test is used to determine statistically significant differences between the groups. **d,** Correlation between the different groups of breast patients identified in (b) and incidence of distant metastasis. Chi-squared test is used to determine statistically significant differences between the groups. See Supplementary Table 4 in Materials and methods for additional statistical analysis.

To better understand the pattern of PRRX1 expression in human breast cancer, we analysed the METABRIC database[247–249]. Interestingly, almost 20% of patients show amplification of the *PRRX1* locus (Figure 12a). Analysis of the Integrative Clusters described in the original METABRIC manuscripts revealed a clear enrichment of IntClust8 among amplified patients (Figure 12b), which is expected since this subgroup is mainly defined by amplification of 1q chromosomal regions where *PRRX1* is located. However, when comparing PRRX1 expression across clusters, IntClust8 was not the highest; instead, IntClust4 showed the highest expression (Figure 12c). Notably, IntClust4 is considered to have a good prognosis[248]. Regarding molecular subtypes, Claudin-low tumors exhibited the highest *PRRX1* expression (Figure 12d), consistent with stronger EMT activation in this subtype compared to others[245,250]. A detailed analysis of the METABRIC database revealed internal heterogeneity inside the Claudin-low patients, including a subgroup with very high EMT levels, with particularly high *ZEB1* expression, and exceptionally good prognosis[250]. It is possible that *PRRX1*-high patients largely correspond to this subgroup, although further validation is required. Regarding METABRIC subgroups, most Claudin-low tumors fall within IntClust4 (Figure 12e), which, even when comparing only Claudin-low cases, remains the cluster with the higher *PRRX1* levels (Figure 12f).

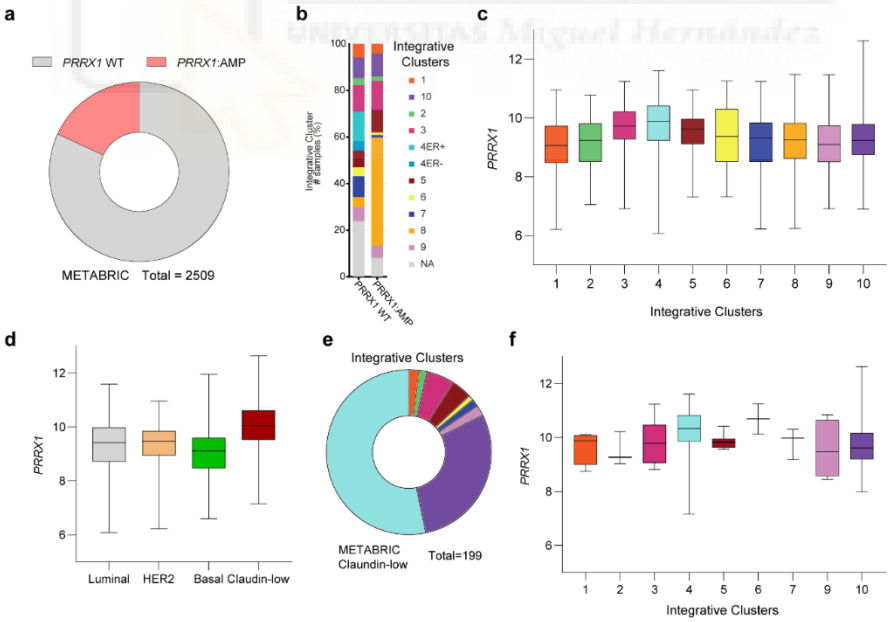


Figure 12. Characterization of PRRX1 amplification and expression patterns in breast cancer. Analyses performed using METABRIC, an integrated database of copy number and gene expression in a discovery and validation set of 997 and 995 primary breast tumours, respectively, with long-term clinical follow-up. **a,** Distribution of breast cancer patients with or without *PRRX1* amplification. **b**, Composition of Integrative clusters in patients bearing breast cancer with *PRRX1* amplified or not. **c**, *PRRX1* expression levels across different Integrative clusters. **d**, *PRRX1* expression levels across different breast cancer subtypes. **e**, Composition of Integrative clusters of claudin-low patients. **f,** *PRRX1* expression levels across different Integrative clusters of claudin-low patients.

The non-linear correlation between tumoral *PRRX1* expression in patients and metastatic incidence was clear and intriguing, but it did not provide information on whether cells with intermediate levels of *PRRX1* were more metastatic than those with low or high expression. To assess this, the ideal approach would be to generate an animal model with different PRRX1 levels in the breast tumors. The prediction would be that cells with intermediate *Prrx1* levels in the primary tumor should be able to generate more metastasis than those negative or high.

We leveraged a recently published single-cell lineage tracing dataset that tracks the metastatic cascade in a PDAC model in the mouse, which revealed that metastases primarily originate from hybrid E/M cells within the primary tumor[177]. This dataset included scRNA-seq and lineage tracing data from the primary tumor, circulating tumor cells (CTCs), and multiple metastatic sites. The authors developed an analytical framework to achieve subclonal lineage resolution, defining subclones as the smallest groups of cells that cannot be further subdivided based on lineage tracing data. The dissemination score of each subclone was quantified using the Shannon Equitability index (see Materials and methods)

Due to the complexity of the data, we first set to reproduce the authors' main finding regarding the high metastatic potential of hybrid E/M subclones, obtaining results very similar to those reported (Figure 13a). Following an approach like the quantitative classification of breast cancer patients based on PRRX1 immunofluorescence, we identified three clusters of subclones according to their mean Prrx1 expression (Figure 13b). Strikingly, subclones in the intermediate group (and therefore with intermediate *Prrx1* expression in this dataset) were more metastatic than those in the first group (negative/very low) or the third group (high) (Figure 13c). Moreover, the first group was more metastatic than the third (Figure 13c), mirroring our previous observations in breast cancer patients (Figure 12). These findings indicate that, within the same tumor, subclones with different *Prrx1* expression levels coexist and exhibit distinct metastatic potentials. Furthermore, they suggest that the non-linear correlation between *Prrx1* and metastasis is not exclusive to breast cancer, as this evidence comes from a pancreatic cancer mouse model.

Overall, these findings suggest that Prrx1 levels, rather than its mere presence or absence, may be a critical determinant of the metastatic potential in cancer cells. This observation reconciled the observation that Prrx1 is necessary for invasion and metastasis[42] while also explaining the increased metastatic potential following its downregulation[51].

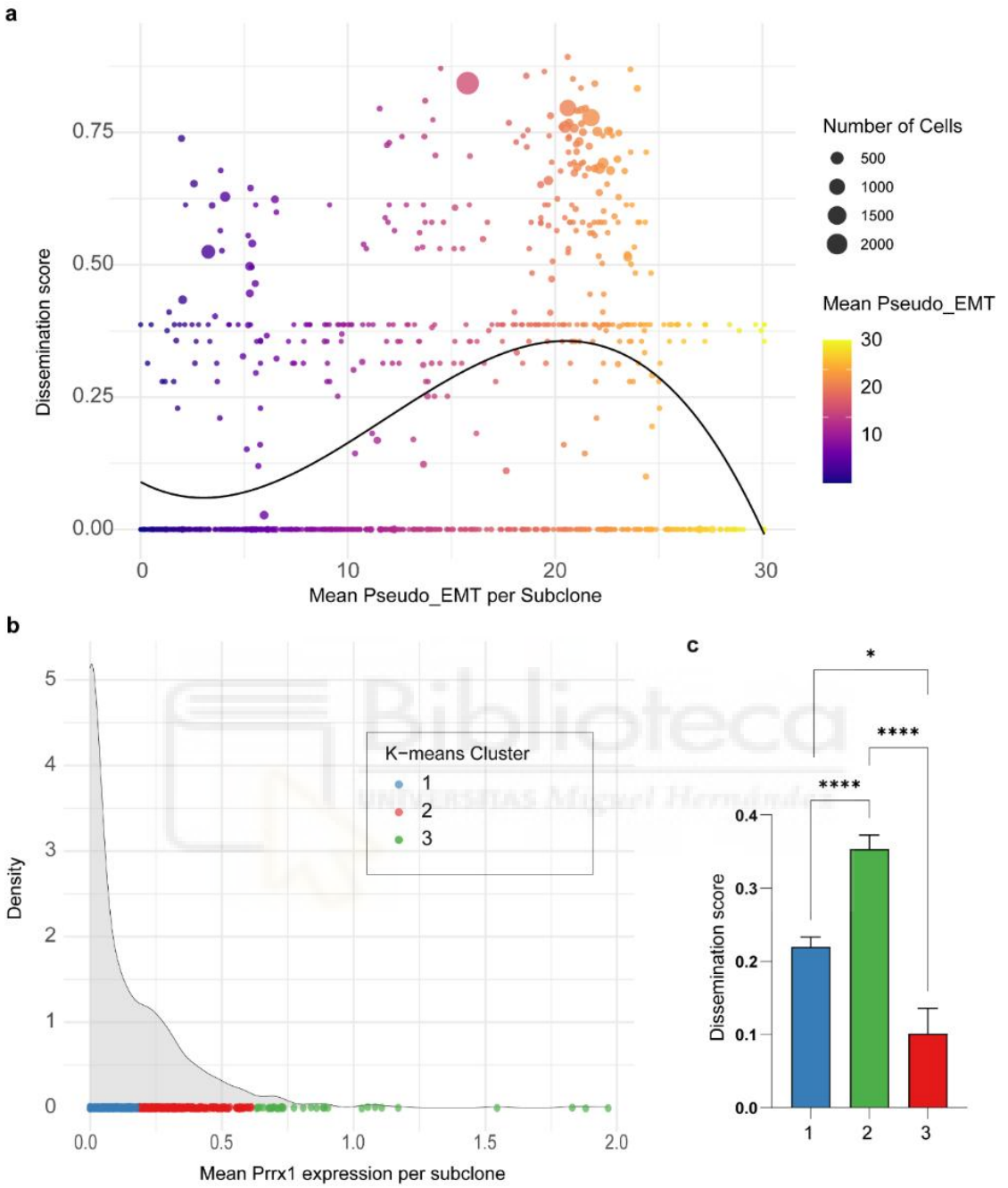


Figure 13. ***In vivo*** **single cell lineage tracing confirms the non-linear relationship between** ***Prrx1*** **expression and metastasis. a,** Relationship between metastatic dissemination (based on Shannon equitability index) and EMT activation for all the subclones in the analyzed tumor. **b,** Density plot of Prrx1 expression of each subclone (calculated as the mean Prrx1 expression of the cells included in the subclone) color by its associated K-means luster. **c,** Comparison of the dissemination scores of the K-means clusters showed in b. Bar represents mean ± SEM. n= 379 for Cluster 1, 219 for Cluster 2, 27 for Cluster 3. Mann-Whitney test. ∗ p <0.05, ∗∗∗∗ p < 0.0001.

2. *In vivo* genetic design of different *Prrx1* levels

The discovery of the previous non-linear correlation between *Prrx1* levels and metastatic incidence in breast cancer patients prompted us to consider that the role of *Prrx1* along the metastatic cascade may be complex and level dependent. To further characterize this, we aimed to recreate the 3 different patient groups based on PRRX1 expression in mouse models, allowing us to investigate *Prrx1* role throughout the metastatic cascade.

First, we needed to select an appropriate cancer model with different *Prrx1* levels. Since we were particularly interested in the impact of heterogeneity within the cancer population and its influence on metastasis, it was essential to use a spontaneous model, which reflects cancer progression[251] more accurately than approaches using established (and much more homogeneous) cell lines. Because the non-linear correlation of *Prrx1* likely reflects its different roles at different stages of the metastatic cascade, it was crucial to monitor all steps of this process. This approach should allow to assess how preexisting heterogeneity among cancer cells in the primary tumor influences metastatic outcomes. Although experimental tail vein injection models are valuable for studying extravasation and metastatic colonization, they bypass the initial invasion and intravasation stages, and are therefore not suitable for investigating cell behavior throughout the entire metastatic cascade[251].

With those criteria in mind, we selected the MMTV-PyMT spontaneous breast cancer mouse model[252] (Figure 14a). This transgenic line expresses the polyoma middle T-antigen, a transmembrane protein and potent oncogene that activates multiple growth pathways[253], under the control of the mammary tumor virus (MMTV) promoter. This promoter is activated by glucocorticoids, progesterone, and dihydrotestosterone and is specifically active in the luminal cells of the mammary gland[254] (Figure 14a). Notably, although the MMTV-PyMT model is considered representative of Luminal B breast cancer[255], it can progress to triple-negative breast cancer (TNBC) through the emergence of basal-like cancer cells following activation of an embryonic-like EMT program[42,256,257]. Indeed, it is now well accepted that basal breast cancers can have a luminal origin, making the PyMT model ideal for studying the role of EMT regulators in breast cancer progression.

Next, we needed to generate different Prrx1 levels. We opted for a genetic approach where varying levels are determined by the number of available *Prrx1* alleles.

Specifically, we used our recently developed *Prrx1* cKO mouse line, which has two loxP sequences flanking *Prrx1* exon 2, an exon essential for *Prrx1* function[42] (Figure 14a). To establish different *Prrx1* levels, we used three genotypes: control (both alleles intact, *Prrx1* +/+) mice, heterozygous *Prrx1* cKO mice (one *floxed* allele, *Prrx1* +/f), and homozygous *Prrx1* cKO mice (both alleles floxed, *Prrx1* f/f) (Figure 14a). This system allowed us to generate three groups: a *Prrx1* high group (bearing tumors with the potential to express high *Prrx1* levels), an *Prrx1* intermediate group (in which expression is limited by the presence of a single allele), and a *Prrx1* negative group (with both alleles deleted) (Figure 14a). A key advantage of this genetic approach is that *Prrx1* expression remains within physiologically relevant levels. In contrast, overexpression-based strategies can introduce artifacts due to unnaturally high levels. Moreover, gene expression remains regulated by its native genomic context rather than being constitutively active. Importantly, this allele-dependent graded system, allows to investigate Prrx1 dosage on cancer progression, an aspect that is difficult to model using conventional overexpression or knockdown approaches.

Finally, to investigate the role of *Prrx1* specifically in cancer cells, we employed the *Krt14*-Cre tdTomato mouse line. The Krt14-Cre system targets mammary gland progenitor cells from early embryonic stages[258] (Figure 14a). Thus, in the adult mammary gland, all epithelial cells (luminal and basal) are Tomato$^+$, regardless of Krt14 expression (Figures 14b and 15a). In PyMT tumors, virtually all cancer cells (PyMT+) are tdTomato+ (≈99% in previous studies[42]), irrespective of Krt14 expression (Figure 15b), while stromal cells remain negative (Figure 15b). Consistent with our recent finding that luminal cells acquire basal-like traits during PyMT tumor progression[42] (see also Figure 5), almost 99% of K14$^+$ cells are PyMT$^+$, ruling out the possibility of them being non-transformed basal-like cells, as the oncogene is active only targets the luminal compartment (Figures 14b and 15c). Altogether, these data confirm that our model is highly sensitive and specific, enabling Prrx1 deletion in all cancer cells before tumor onset without affecting other cell populations.

In summary, our mouse model develops spontaneous metastatic breast cancer tumors, bearing all cancer cells labeled with tdTomato (due to Cre activation during development), and exhibit different *Prrx1* allelic combinations in cancer cells according to their genotype (Figures 14 and 15).

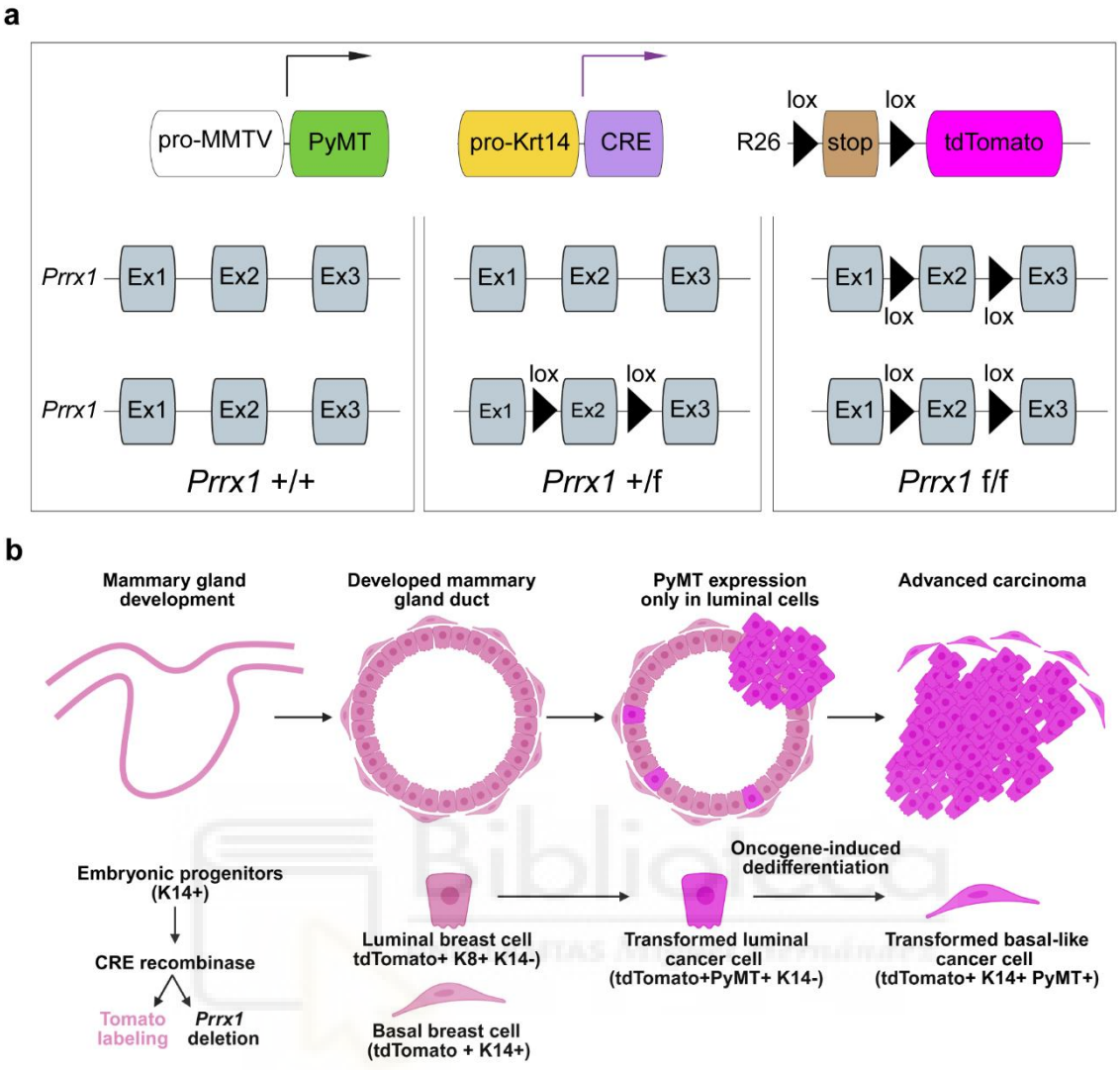


Figure 14. Genetic design and scheme of the PyMT breast cancer mouse model with different *Prrx1* levels in the cancer cells. a, Genetic strategy used to label cancer cells with tdTomato and to create tumors with different Prrx1 levels via genetic dosage. **b,** Scheme of the evolution of the PyMT model together with the dynamics of the K14-reporter and the deletion of *Prrx1* alleles.

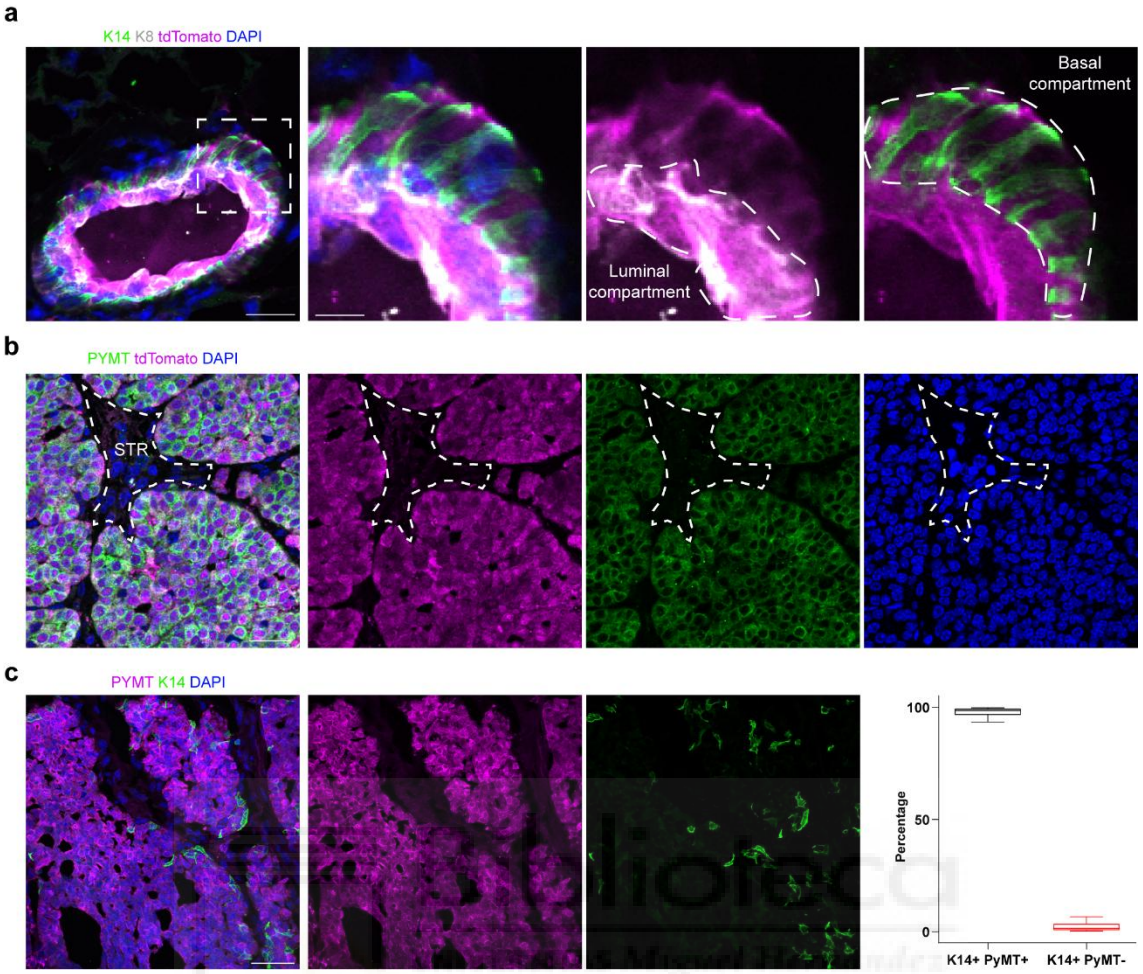


Figure 15. Efficiency and specificity of the final mouse model. a, Representative IF of an adult mammary gland of a mouse harboring K14-CRE and the tdTomato reporter, but negative for MMTV-PYMT cassette. K8 labels luminal cells while K14 identifies the basal myoepithelial mammary gland cells. As the mammary gland embryonic progenitors are K14+, tdTomato labels all the future progeny of the mammary gland irrespective of their lineage. Scale bar: 17 µm and 6 µm. **b,** Representative IF of an advanced PyMT carcinoma (15 weeks). tdTomato and PyMT specifically label all cancer cells. Stromal cells (STR) remain negative for both, confirming the specificity of the genetic approach. Scale bar: 25 µm. **c,** Representative IF and quantification of K14 and PyMT expression in advanced carcinomas (15 weeks). The great majority of K14+ cells present at this stage are PyMT+. As the PyMT oncogene is only activated in the luminal compartment, these K14+ cells derive from transformed luminal cells that acquired a basal-like program. Boxplots show median, upper and lower quartile; whiskers show upper and lower extreme. N= 6 tumors (2 tumors from 3 different mice). Scale bar: 50 µm.

Aware that Prrx1 has different isoforms with different functions, at least in pancreatic cancer[259], our conditional knockout (cKO) design targets exon 2, which is shared by both isoforms and encodes the critical homeobox domain (Figure 14a). We previously demonstrated that this strategy effectively abolishes both isoforms at the protein level in embryonic fibroblasts carrying this genetic design[42]. Furthermore, heterozygous deletion reduces the levels of both isoforms in a comparable manner[42]. To confirm that this remains true when combined with PyMT and to validate that the different alleles generate

distinct Prrx1 levels, we analyzed the expression of the specific isoforms (Long and Short) by qPCR in FACS-sorted cancer cells from our experimental groups. This confirmed that *Prrx1* +/f cancer cells expressed approximately half the Prrx1 levels compared to *Prrx1* +/+ cells and that both isoforms were proportionally reduced according to the number of available alleles (Figure 16). These data rule out the possibility of isoform-dependent differential effects in our model.

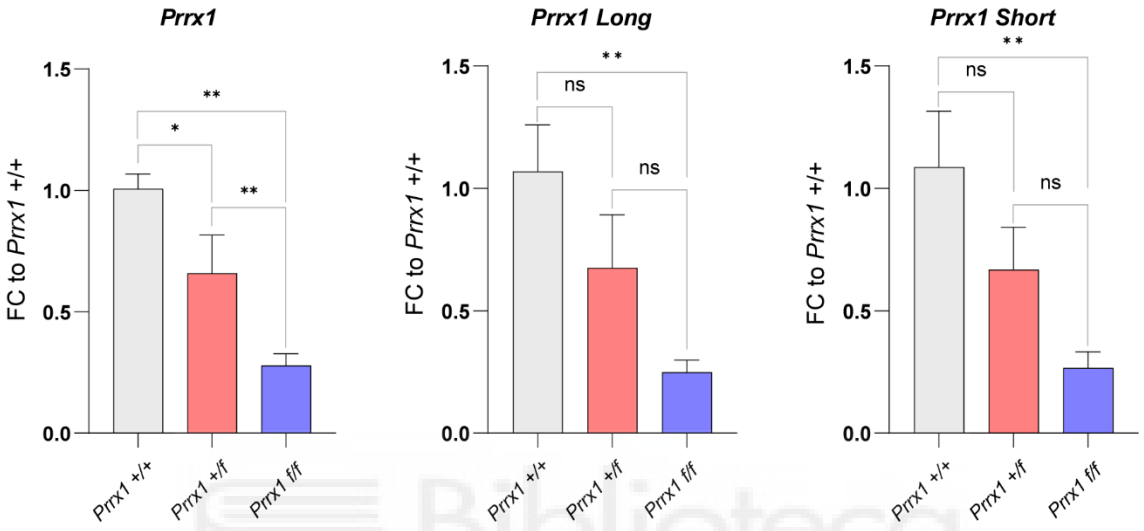


Figure 16. Allele-based genetic strategy generates different Prrx1 levels in cancer cells, RT-qPCR showing the relative transcript levels of total *Prrx1* and isoforms (Long and Short) in FACS-sorted cancer cells per condition. FC is represented as mean ± SEM. *Prrx1* +/+ (n=5); *Prrx1* +/f (n=7); *Prrx1* +/+ (n=8). Mann-Whitney test. ∗ $p < 0.05$; ∗∗ $p < 0.01$.

We had previously demonstrated that *Prrx1* is not required for tumor initiation nor affects tumor growth, as *Prrx1* f/f mice developed similar number and size of tumors compared to those developed in Prrx1 +/+ mice[42]. Here, we confirm that this is also the case for the *Prrx1* +/f mice, as their tumor burden was not significantly different from the other groups (Figure 17a). Consistent with these findings, we did not observe differences in global proliferation, measured by the percentage of mitotic cells across conditions (Figure 17b). It is worth noting here that *Prrx1*-positive cells represent a minority of the total cancer cell population in the PyMT tumors (approximately 1% by scRNAseq)[42] making it unlikely that the different levels of *Prrx1* could have a global impact in tumor growth. In conclusion, we have successfully generated breast cancer mouse models with different levels of Prrx1 in cancer cells, without affecting overall tumor burden, which could be a confounding factor when considering metastatic burden.

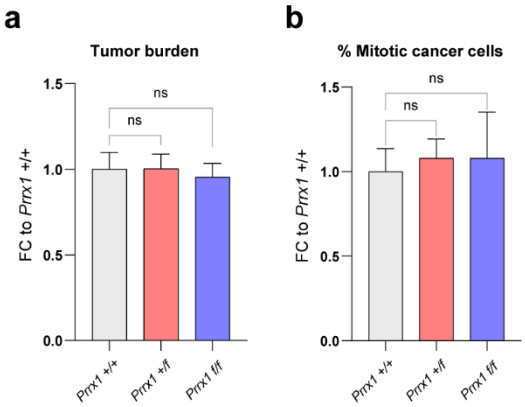


Figure 17. *Prrx1* levels do not affect tumor burden. a, Quantification of total tumor weight per animal in the different *Prrx1* genotypes. FC is represented as mean ± SEM (n=23 mice per group). Two-tailed unpaired test. p > 0.05, not statistically significant (ns). **b,** Mitotic ratio of cancer cells in whole PyMT tumors with different *Prrx1* genotypes. FC is represented as mean ± SEM. n=3. Mann-Whitney test. p > 0.05 were not statistically significant (ns).

3. *In vivo* genetic modeling of different *Prrx1* levels recapitulates the non-linear correlation with metastatic burden

Once we confirmed that our mouse model for different *Prrx1* levels behaved as expected, we investigated the impact of varying *Prrx1* levels on lung metastatic burden to assess whether they recapitulated the non-linear correlation observed in patients and in a PDAC mouse model (Figures 11 and 13). For a more precise quantitative analysis, we employed the iDISCO tissue-clearing protocol, widely used in neuroscience[226] combined with light-sheet imaging technology to reconstruct the lungs of the different experimental conditions in 3D (Figure 18a).

Consistent with our observations in breast cancer patients, mice with *Prrx1* +/f tumors exhibited the highest lung metastatic burden, resulting from an increase in both the number and size of metastatic foci (Figure 18b). In contrast, *Prrx1* f/f mice displayed a dramatic reduction in metastatic burden compared to controls (Figure 18b), aligning with our previous findings[42]. Thus, our mouse models accurately reproduced the non-linear correlation between *Prrx1* levels and metastatic burden observed in patients.

However, it is true that our results did not fully mirror the patient data. While PRRX1-negative patients exhibited metastases, indeed, more commonly than the PRRX1-high group, that seemed contradictory to our mouse model results. That discrepancy may stem from differences in Prrx1 expression dynamics between patients and mice. In the patient group, PRRX1-high tumors were characterized by a substantial number of

PRRX1-positive cells, most of which displayed high expression. In contrast, in the *Prrx1* +/+ mice, we were not artificially driving high expression in all the cancer cells, meaning that these tumors likely consisted of a mixture of negative, intermediate, and high *Prrx1* expressing cells. This heterogeneity may have allowed some cells within the *Prrx1*+/+ group to metastasize if the *Prrx1* level they have are optimal for metastasis. Additionally, it is possible that alternative mechanisms drive invasion and metastasis in certain breast cancer patients, whereas in the PyMT model, metastasis appears to be more dependent on *Prrx1*.

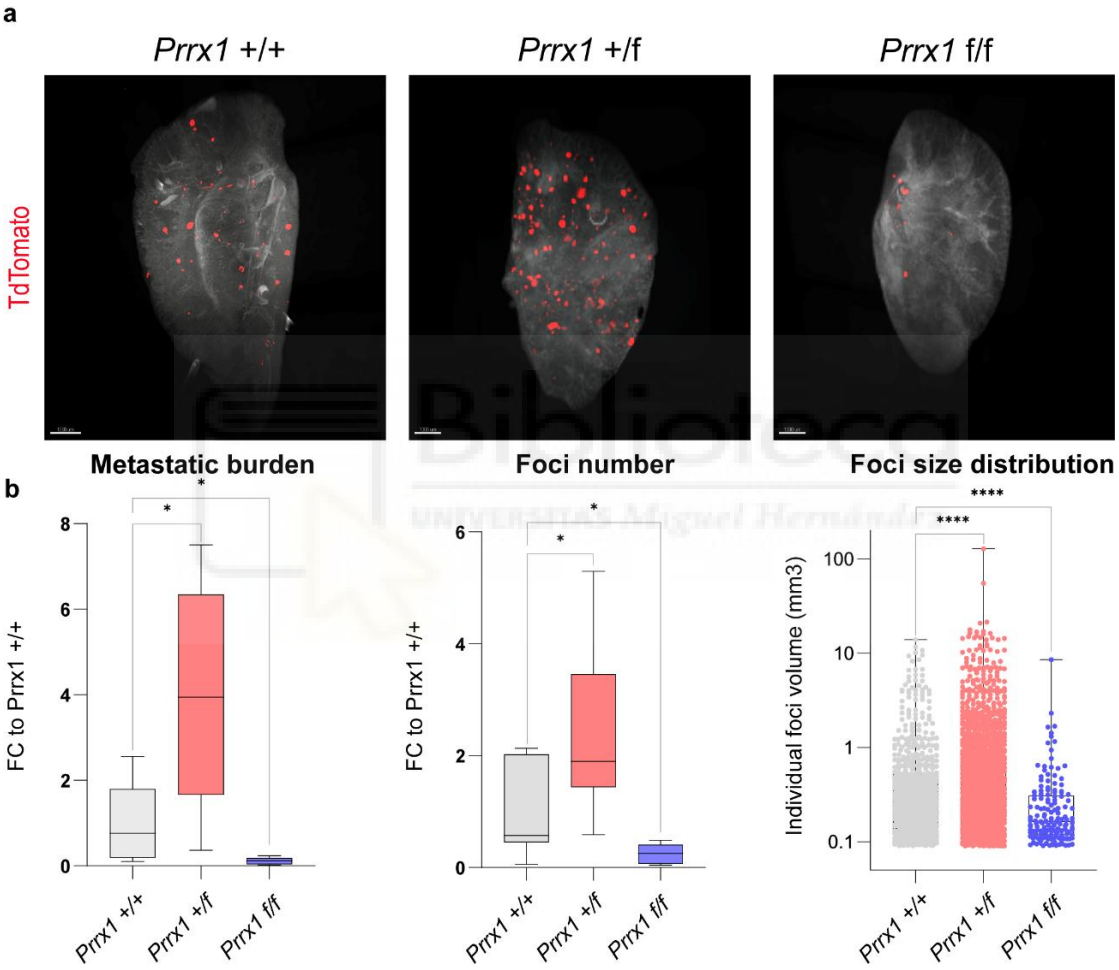


Figure 18. Different *Prrx1* levels in mouse PyMT tumors recapitulate the non-linear correlation between Prrx1 expression and metastatic burden. **a,** 3D reconstitution of whole-mounted clarified lung lobes showing metastatic cells (TdTomato), and the corresponding quantification of metastatic burden, and foci number and size (n=7 mice per group). Boxplots show median, upper, and lower quartile; whiskers show upper and lower extreme. Mann-Whitney test. ∗ p <0.05, ∗∗∗∗ p < 0.0001. Scale bars: 1000 µm.

One possible explanation for the increased metastatic burden observed in the Prrx1 +/f tumors is enhanced invasive potential. Recent evidence in the PyMT model shows that luminal tumor cells progressively acquire basal-like traits, including de novo Krt14

expression, coinciding with activation of an invasive EMT program[42]. Additional studies have demonstrated that Krt14 reliably marks the invasive and disseminative population in this model[260]. Therefore, in line with our previous work and the characterization of the invasive EMT program[42], we combined peripheral enrichment of Krt14 expression with the presence of disorganized laminin borders, a well-established morphological hallmark of invasive behavior, to delineate invasive areas. Consistent with our data on metastatic burden, *Prrx1* +/f tumors displayed a significant increase in invasive areas compared to control tumors (Figure 19a). Importantly, we detect Prrx1/K14 double positive cancer cells in these areas (Figure 19b), indicating that, at least, a subset of the invasive cells expresses Prrx1. An increase in invasive areas can explain the elevated metastatic burden (Figure 18b). However, this is at odds with the previously described direct correlation between *Prrx1* expression and invasion[42,51,261]. This suggested that Prrx1-dependent mechanisms beyond invasion could be playing a critical role in metastatic colonization. In addition, our findings in the genetically engineered breast cancer mouse models confirm the non-linear relationship between *Prrx1* levels and metastasis formation. This is reminiscent of a hormetic behavior, a two-phased dose-response relationship mainly described in toxicology, whereby a low-dose of an agent has a positive effect while a high-dose is inhibitory[262,263]. For Prrx1, low doses promote metastasis, while high doses inhibit their formation.

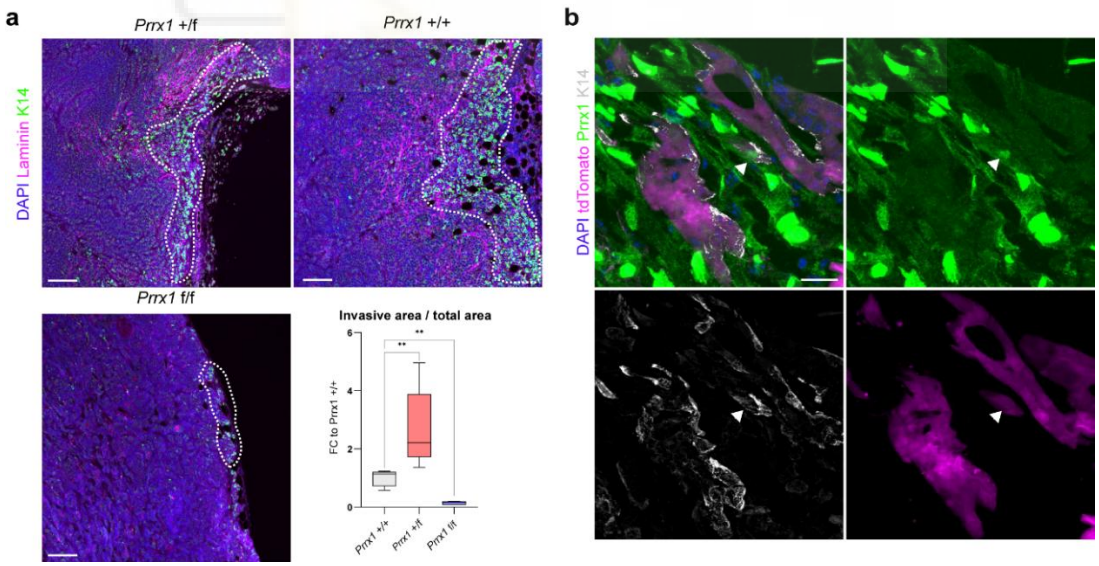


Figure 19. Non-linear correlation between *Prrx1* levels and invasive areas in PyMT tumors. a, Representative images and quantification of invasive areas in PyMT primary tumors. Invasive areas are defined as peripheral areas with disorganized pan-Laminin and enriched in K14-+ cancer cells (dashed lines). FC is represented as mean ± SEM; n=5 mice per group. Boxplots show median, upper and lower quartile; whiskers show upper and lower extreme. Mann-Whitney test. ∗∗ $p < 0.01$. Scale bars: 100μm. **b,** Representative image showing a Prrx1+ K14+ cancer cell (white arrowhead) inside an invasive area. Scale bars: 50 μm.

4. K14 positive cells in the primary tumor are the main source of circulating tumor cells

To confirm if, as expected, the cancer cells that we have identified located at the invasive edge were responsible for metastatic dissemination, we collaborated with the Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS) and the Hospital Clínico Universitario de Santiago (CHUS) to study circulating tumor cells (CTCs) in PyMT tumors.

To isolate CTCs, we used the Parsortix system[150], a marker-agnostic approach that selects cancer cells based on morphological and mechanical properties. This was particularly important because the most widely used technique to isolate CTCs relies on EpCAM positivity[149], which decreases along the epithelial to mesenchymal spectrum and is known to overlook mesenchymal CTC populations. First, we optimized immunofluorescence protocols in the Parsortix system using a newly-generated primary cell line derived from our PyMT tumors and assessed the expression of markers selected to represent the two EMT trajectories that we had previously described in tumor evolution[42] (see also Figure 5). We selected Jun as a marker for the inflammatory-like EMT program and K14 for the invasive EMT program. We mixed cells from our established PyMT primary cultures with control mouse blood and confirmed that immunofluorescence works in this system and that we can detect the presence of cancer cells belonging to the two trajectories (Figure 20a).

To analyze mouse CTCs, we collected blood via cardiac puncture from 13-14 weeks old mice bearing PyMT tumors. The isolated CTCs were stained for EpCAM, K14, and Jun. Importantly, we did not detect any Jun-positive CTCs, indicating that cells belonging to the inflammatory-like trajectory are not invasive as previously inferred by its signature[39], but also that they are not carried into the bloodstream by invasive cells. Interestingly, most CTCs (approximately 90% of the detected CTCs) were K14-positive (Figure 20b-c), supporting the idea that the K14 population in the invasive areas in the primary tumor represents the invasive and disseminative population, consistent with independent data from the same PyMT model[260]. Specifically, the majority of CTCs are EpCAM and K14 positive, which represent hybrid E/M cells already engaged in the invasive EMT program. Additionally, we detected a small fraction of CTCs (around 20%) K14-positive and EpCAM-negative, suggesting that very likely these cells may have progressed along the EMT program to the mesenchymal state (Figure 20b-c). Indeed, in the analysis of the scRNA-seq data we used to identify the two EMT trajectories in the PyMT model[42] (see

also Figure 5), we showed a gradual loss of *Epcam*, reaching particularly low levels in cluster 16 (Figure 20d), while *Krt14* expression was very much decreased but not lost (Figure 20d). This *Epcam*-negative, *Krt14*-low mesenchymal CTC population would be missed using EpCAM-dependent CTC isolation approaches.

Unfortunately, we were unable to confirm visualize Prrx1 expression in this CTCs immunofluorescence assay, very likely due to its low expression levels. However, further analysis of the *in vivo* single-cell lineage tracing dataset described earlier revealed an enrichment in Prrx1 expression in CTCs compared to that in the primary tumor (Figure 20e). Given that the time of residency of cancer cells in the bloodstream is in the order of seconds to minutes[264], the observed differences in expression between primary tumor cells and CTCs are very likely driven by selection rather than by transcriptomic changes.

These data confirm the invasive and disseminative role of cells in EMT-T1, the invasive trajectory, validate our strategy to identify invasive regions in the primary tumor, indicate that the disseminative population expresses Prrx1 and confirm the potential of CTCs analysis to understand cell plasticity dynamics along the metastatic cascade.

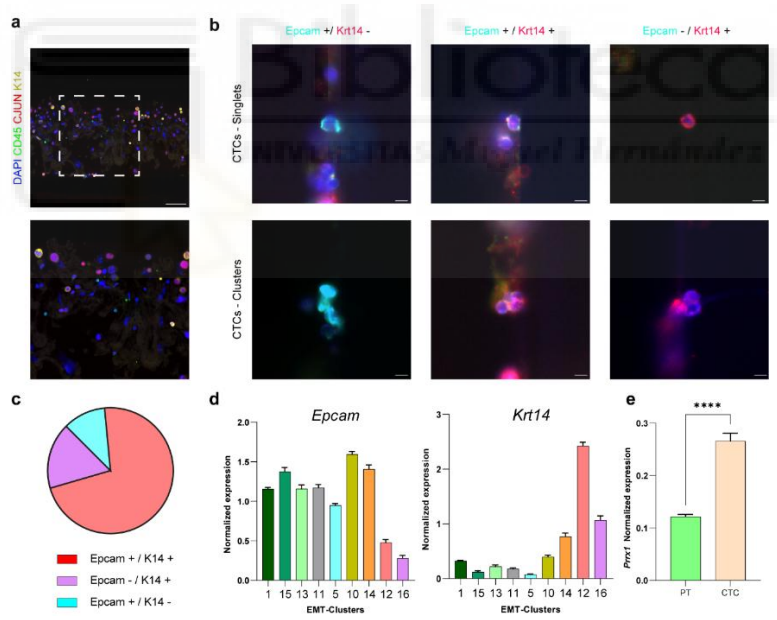


Figure 20. K14 positive primary tumor cells are the main source of circulating tumor cells. a, Spiking control of K14 and Jun Immunofluorescence in cells from a primary cell line derived from PyMT tumors processed though the Parsortix system. Scale bar: 100 µm**. b**, Immunofluorescence images of circulating tumor cells isolated from blood extracted from mice bearing PyMT tumors. No CTCs positive for the T2 marker Jun were found. Scale bar: 5 µm**. c,** Quantification of K14 and Epcam status by immunofluorescence of CTCs disseminated from PyMT tumors. CTCs isolated from two mice were pooled in the same Parsortix cassette. **d,** Expression of *Epcam* (left) and *Krt14* (right) in the different cell clusters along the EMT trajectories inferred from scRNAseq data of PyMT tumors[42]. Bar represents mean ± SEM. **e,** Expression of *Prrx1* in the primary tumor compared to CTCs from the same mouse (PDAC model). Bar represents mean ± SEM. Mann Whitney test. **** $p < 0.0001$.

5. Metastases recapitulate the EMT hierarchy observed in the primary tumor.

After confirming that the vast majority of CTCs belonged to the EMT-T1 (invasive) trajectory (Figure 20b), we next assessed the composition of the lung metastases with respect to the two described EMT trajectories[42] (see also Figure 5). To do this, in addition to the previously analyzed markers, we included p63, a marker of early EMT-T1 , and Mafb, a more advanced EMT-T2 marker compared to Jun (Figure 21a)[42].

Surprisingly, despite the absence of EMT-T2-positive CTCs, we identified Jun and Mafb positive cancer cells in the metastases (Figure 21a). Furthermore, measuring the relative contribution of each trajectory in the lung, we found that metastases almost exactly recapitulate the distribution observed in the primary tumor (Figure 21b-c). Macrometastases harbored a very similar percentage of cancer cells positive for each marker compared to the primary tumor (Figure 21c), even though most cancer cells arriving at the metastatic site were K14+ Jun- (Figure 21c). Nevertheless, we observed an increase in the percentage of K14+ cells in metastases compared to the primary tumor (Figure 21c). We propose three possible hypotheses for this increase: (i) Since the initial metastatic population is predominantly K14+, these cells may revert during metastatic outgrowth and become diluted, yet their cell-of-origin bias remains detectable, (ii) cells that have previously undergone EMT-T1 activation may be primed to reactivate it due to epigenetic memory, (iii) EMT-T1 cells tend to be located near the stroma, and given that metastatic lesions are typically smaller than primary tumors, they exhibit a higher periphery-to-surface ratio, leading to a higher percentage of K14 positive cells relative to the total number of cancer cells compared to larger tumor areas like the primary tumor. Further experiments will be required to clarify the cause of the increased proportion of K14 positive cancer cells in metastases. In conclusion, our findings indicate that cancer cells exhibit high plasticity along the metastatic cascade. The invasive population originates from a minority subpopulation in the primary tumor (EMT-T1 cells, which constitute less than 5% of the cancer cells), but upon reaching the metastatic site, they could re-establish the original cancer cell state composition of the primary tumor.

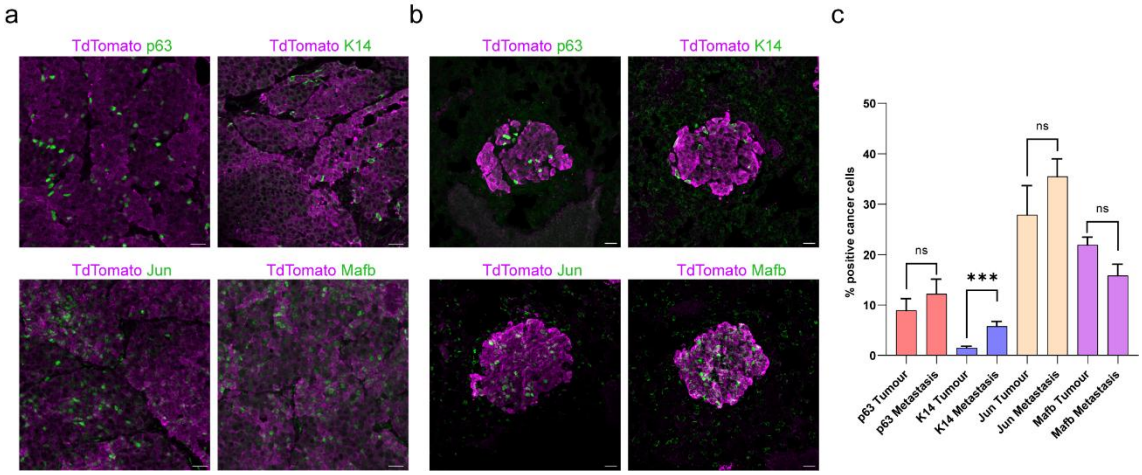


Figure 21. Metastases recapitulatee primary the EMT hierarchy observed in the primary tumor. a, Immunofluorescence images of early T1 (p63), late T1 (K14), early T2 (Jun) and late T2 (Mafb) of PyMT primary tumor. Cancer cells are all positive for TdTomato. Scale bar: 25 μm**. b,** Similar analysis as in (a) in PyMT lung metastases. Scale bar: 20 μm **c,** Quantification of selected trajectory markers (a,b) in PyMT primary tumors compared to lung metastasis. Bar represents mean ± SEM. n=5 tumors, 13 metastases for p63, 20 metastases for K14, 14 metastases for Jun, 11 metastases for Mafb. Mann Whitney test. *** $p < 0.001$, $p > 0.05$, not statistically significant (ns).

6. Spatial transcriptomics gene panel design and implementation

As we confirmed the disseminative nature of the K14 positive cells at the invasive edge, we decided to examine their transcriptomic profile. We performed spatial transcriptomics at the single-cell resolution, which allows the extraction of transcriptomic information while preserving the spatial context, as it is conducted on intact tissue sections[265].

There are two categories of spatial transcriptomic approaches, either based on sequencing or on imaging[266]. Sequencing-based approaches typically rely on a bead system that tags cells with spatial information, which can be retrieved after sequencing. In contrast, imaging-based approaches maintain tissue architecture and use cyclic multiplex in-situ hybridization to identify and decode transcripts through imaging.

The main difference between these two approaches lies in the number of genes detected and the resolution achieved. Sequencing-based methods enable unbiased transcript detection across the genome but lack single-cell resolution. This is because the sequencing data reflects a mixture of transcripts from a small group of cells located within the defined sequencing region, which is determined by the size of the bead. On the other hand, imaging-based approaches provide single-cell resolution (and even subcellular

resolution in some cases) but are limited by the need for a preselected gene panel However, these limitations are rapidly being addressed. For instance, sequencing-based methods are improving in resolution (e.g., Visium HD), while imaging-based platforms are expanding their gene coverage to include thousands of genes (e.g., MERSCOPE Ultra)[266].

For our research question, achieving cellular resolution was crucial for several reasons. First, we were specifically interested in cancer cells located at the invasive edges, where there is extensive tumor-stromal interaction. Without cellular resolution, it is challenging to accurately distinguish transcripts originating from cancer cells versus those from the surrounding stroma. Second, we aimed to understand the EMT status of these cells. EMT cancer cells and certain stromal cells, particularly CAFs, share multiple transcripts, with CAFs exhibiting much higher expression of the same genes. This similarity can make it difficult to accurately identify EMT characteristics in cancer cells. Thus, we decided to use an imaging-based approach, specifically the MERSCOPE platform by Vizgen, based on MERFISH technology[267].

As previously mentioned, this type of technology requires the user to design a gene panel containing the genes to be analyzed. We designed a MERSCOPE panel of 264 genes using the Vizgen Gene Panel Design Portal. Gene selection was based on two criteria: first, the inclusion of genes representing major cell types and cell states (primarily cancer cell states but including different stromal cells) identified in our previously published scRNA-seq data of PyMT tumors[42] (see also Figure 5); second, genes involved in key signaling pathways such as Wnt, Hedgehog, and Hippo pathways, as well as cytokine-cytokine receptor interactions (Figure 22a).

We profiled FFPE sections from three different *Prrx1*+/+ tumors using the described gene panel with the MERSCOPE platform. One of the main challenges in imaging-based spatial transcriptomics is achieving high-quality cell segmentation, as this determines the accurate assignment of detected transcripts to individual cells. We took advantage of a new kit from Vizgen, the Cell Boundary Kit, which stains cell membranes using proprietary agents (Figure 22b), allowing high-quality segmentation in our tissue. This was key due to the high cell density of the PyMT tumors (Figure 22c).

After an initial quality control filter, we identified a total of 1.794 million high-quality cells, which were grouped into 18 transcriptionally distinct clusters (Figure 22d). Importantly, all cell clusters displayed a consistent distribution across the three independent samples (Figure 22e), indicating both the robustness and reproducibility of the PyMT model, of the technology used and a successful data integration.

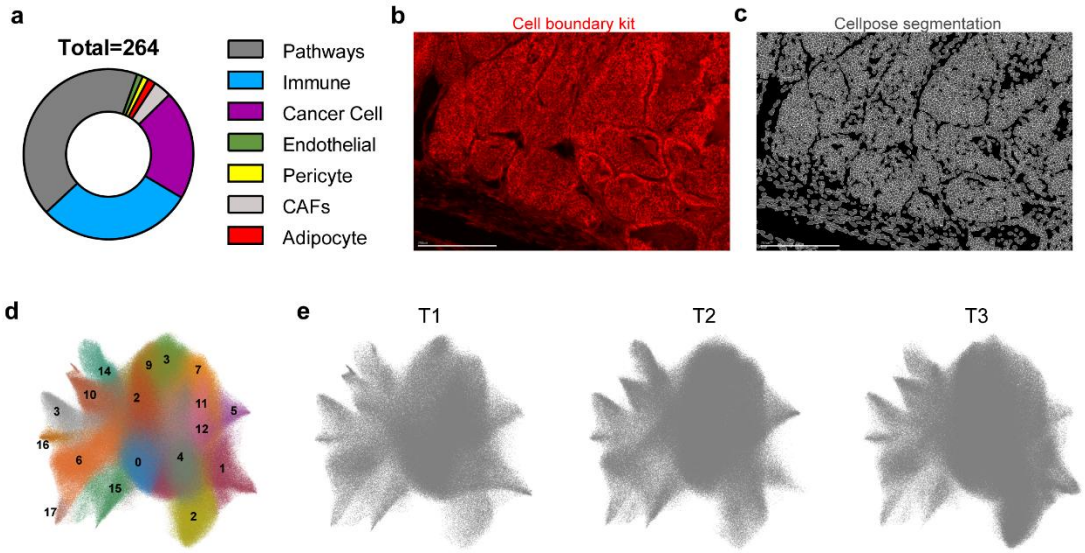


Figure 22. Spatial transcriptomics gene panel design and implementation. a, Distribution of the 264 genes in the MERSCOPE Custom Pannel designed to detect the main cell types and EMT activation in PyMT tumors. **b**, Representative image of a section obtained from a PyMT tumor showing the staining with MERSCOPE Cell Boundary Kit. Scale bar: 250 µm. **c**, Representative image of cell segmentation using Cellpose 2 algorithm. Scale bar: 250 µm. **d**, UMAP projection of ST analysis of three PyMT tumors.

7. Spatial transcriptomics identifies a molecular signature of cancer cells in the invasive areas

To classify the different cell clusters into cell types, we used a supervised annotation approach based on domain knowledge and previous scRNA-seq analyses of PyMT tumors[42,268,269]. Briefly, we identified the differentially expressed genes in each cluster and used this information to determine their corresponding cell types. We successfully annotated spatial transcriptomics data into five major cell types: cancer cells, CAFs, macrophages, endothelial cells, and pericytes (Figure 23a). Each of these populations exhibited differential expression of bona fide cell type markers, including *Gata3* (cancer cells), *Lrrc15* (CAFs), *Rgs5* (pericytes), *Pecam1* (endothelial cells), and *Cx3cr1* (macrophages) (Figure 23b). Additional not identified cell populations, such as lymphocytes and neutrophils, are present in PyMT tumors but represent a small fraction of the total cells and the custom designed panel only contained a few genes for their identification. It is possible that a more detailed analysis of stromal populations could identify these rarer cell types, which, at this stage of the analysis, may be intermingled with other cell clusters.

Visual exploration of the spatial transcriptomics data allowed us to examine the spatial distribution of these different cell types within the tumor (Figure 23c). As expected, cancer cells were arranged in dense islands surrounded by tumor microenvironment populations. Some of these populations infiltrated the tumor, particularly macrophages, which appeared to be located within the tumor islands (Figure 23c). Those macrophages could correspond to a pro-inflammatory M1 profile, as we previously demonstrated that this subtype tends to infiltrate PyMT tumors, in contrast to the anti-inflammatory population[42]. Other observations, such as the expected colocalization of endothelial cells and pericytes and the enrichment of stromal cells, mainly macrophages and fibroblasts, at the tumor margins (Figure 23c), further supported the high quality of our data and the accuracy of our cell type annotations.

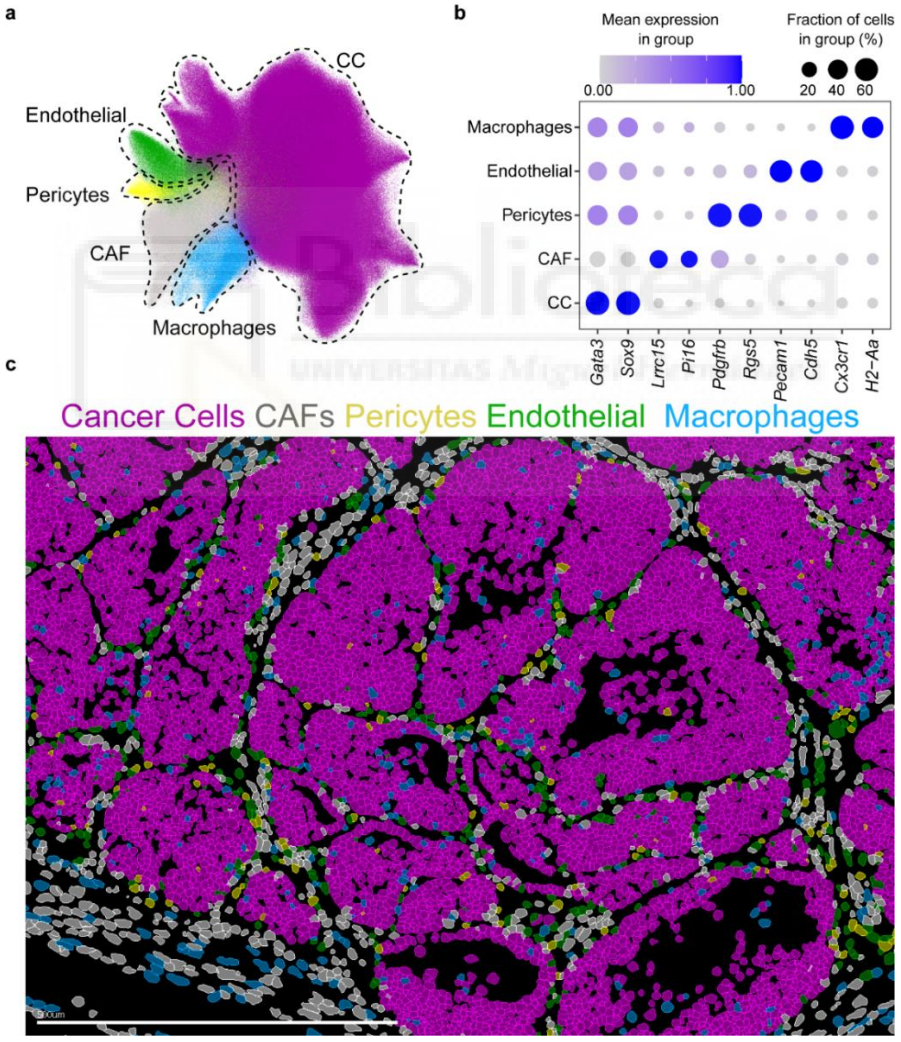


Figure 23. Spatial transcriptomics identify major cell types in PyMT tumors. a, UMAP projection of Spatial transcriptomics (ST) analysis of PyMT *Prrx1* +/+ tumors (n=3) identifying different cell types. **b**, Dot plot showing the expression of representative marker genes that identify different cell types as in (b). **c**, Representative image of MERSCOPE based single-cell spatial transcriptomics (ST) of a section obtained from PyMT tumor showing different cell types. See color code in the figure. Scale bar: 500 µm.

Once we classified the different cell types, we focused our analysis on characterizing the invasive regions. We manually identified areas of high keratin 14 (*Krt14*) expression at tumor margins (Figure 24a), applying a strategy like the one we used for the quantification of invasive areas by immunofluorescence (Figure 19a). Comparing the expression profiles of cancer cells at invasive edges (Figure 24b) with those in cells located elsewhere, we observed that cells within the invasive regions were significantly enriched in invasion-related genes (*Postn, Tnc, Mmp2, Mmp3*) and exhibited a basal-like expression program (*Krt5, Trp63*) (Figure 24c), while showing low levels of injury response and interferon-related gene transcripts (*Fos, Klf4, Mx1, Ifit1*) (Figure 24c). One upregulated gene that drew our attention was Cyclin D2 (*Ccnd2*) (Figure 24c), a crucial regulator of the most critical checkpoint of the cell cycle, the G1/S transition[270]. Observations in human cancer samples associated Cyclin D2 expression with invasive tumor margins and correlated with greater invasion and metastasis in colorectal cancer[271]. Interestingly, Cyclin D2 also correlates with gastric cancer progression, while it is not the case for Cyclin D1 (CCND1)[272]. Furthermore, overexpression of *CCND2* was enough to increase invasive properties of human squamous carcinoma cells[273]. Whether invasive cancer cells utilize a distinct set of cell cycle regulators compared to other cancer cells, and if this difference impacts their biology or response to therapy, could be an interesting avenue for future research. This is particularly relevant given that frontline treatments for many luminal breast cancers are based on CDK4/CDK6 inhibitors, which are regulated by Cyclin D family members[274].

Pathway enrichment analysis revealed that cancer cells residing in invasive regions exhibited an EMT transcriptional profile and activation of the Wnt signaling—an important EMT-promoting pathway[275] (Figure 24d). This was associated with transcription factors such as *Snail2*, a key regulator of the basal program[80], Fibrillin-1 (*Fbn1*), and, importantly, *Prrx1* (Figure 24e). Visual inspection of the spatial transcriptomics data confirmed that this gene program is activated in cancer cells. As several genes, such as *Col1a1*, *Postn* and *Tnc*, are typically highly expressed in stromal populations, we carefully verified that these results were not biased by transcript misassignment or imperfect cell segmentation at tumor borders (Figure 24f). Overall, our spatial transcriptomics analysis identified a distinct molecular signature in cancer cells located at the invasive areas, highlighting an invasive program consistent with our recently identified *Prrx1*-dependent invasive EMT trajectory[42] (see also Figure 5).

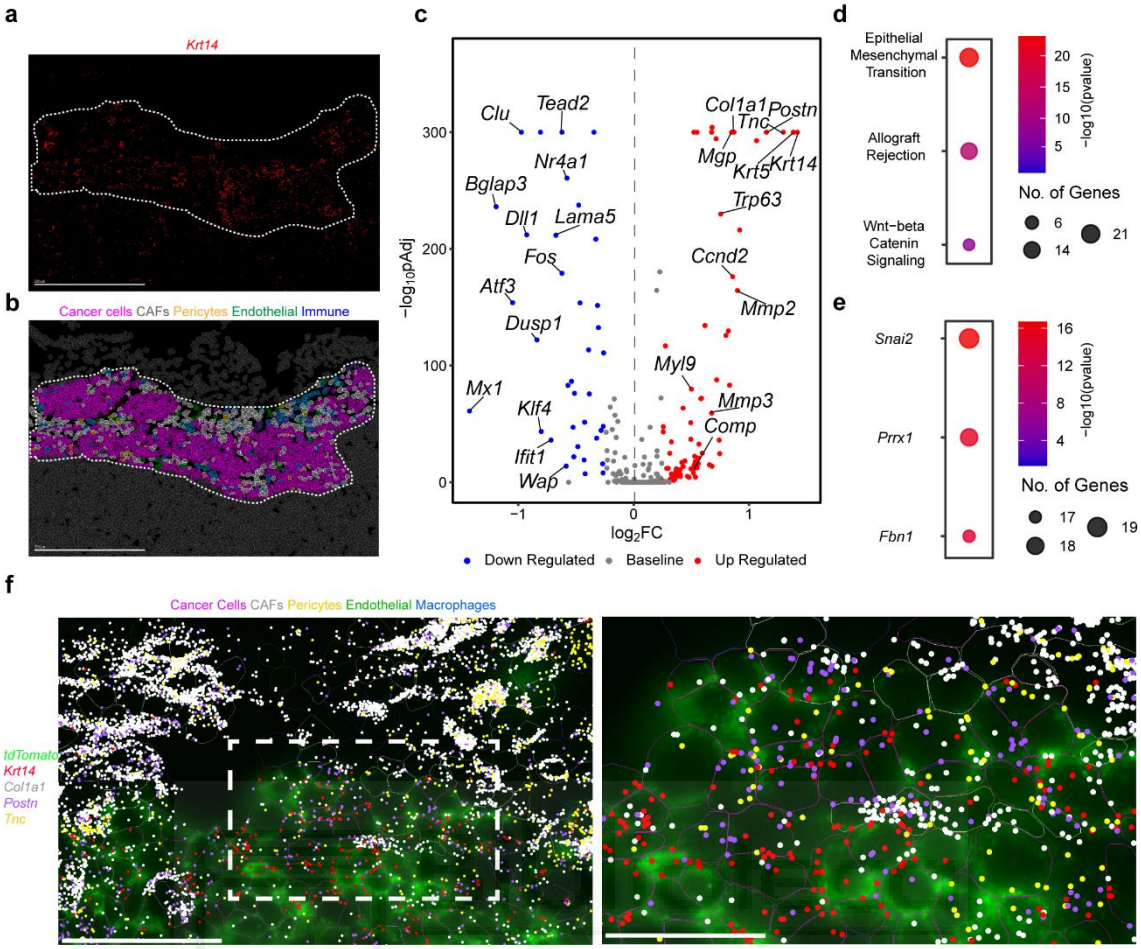


Figure 24. Spatial Transcriptomics identifies an Invasive EMT molecular signature in cancer cells at invasive areas. a, Representative example of a selected invasive edge (outlined) based on *Krt14* expression. Scale bar: 250 µm. **b,** Color-coded cell types in the selected invasive edge. Scale bar: 250 µm. **c,** Volcano plot of differentially expressed genes (DEGs) in MERSCOPE ST data for cancer cells located at invasive edges vs the rest of the tumor (n=3 tumors). DEGs with *p*-value <0.05 and log2 fold change >0.25 labeled as increased (red) or log2 fold change <-0.25 labeled as decreased (blue). **d,** Dot plot showing pathway enrichment (MsigDB_Hallmarks) of differentially upregulated genes (p< 0.05, log2 fold change >0.25) in cells in invasive areas. **e,** Dot plot showing transcription factor association (ARCHS4) of differentially upregulated genes (p< 0.05, log2 fold change >0.25) in cells in invasive areas. **f,** Representative image of a selected invasive area showing the expression of mesenchymal / invasive markers (dots) and tdTomato expression (green labeling), the latter specific to cancer cells. The colors of the cell borders indicate their annotated cell type. See color code in the figure. Scale bar: 500 µm.

8. Single cell transcriptomics analysis of cell populations in PyMT primary tumors

To further dissect how different *Prrx1* levels shape the transcriptional landscape of invasive cells, we performed scRNA-seq (10X Genomics V3) from 3 *Prrx1* +/+, 3 *Prrx1* +/f, and 4 *Prrx1* f/f tumors (see Methods). Additionally, we integrated this new data with our previously published scRNA-seq dataset from 4 *Prrx1* +/+ tumors (10X Genomics V2)[42] (see Methods) to increase the number of cells, allowing for greater resolution of cancer cell states. Since these prior samples have already been annotated and analyzed, this integration also enables us to cross-check potential issues in the analysis of the new data.

Unsupervised clustering identified nine major cell types and cancer cell states (Figure 25a) without any obvious difference in their distribution in the different genotypes (Figure 25b). Each cell cluster was defined by *bona fide* lineage markers (Figure 25c-d) Compared to our previously published scRNA-seq[42], which cells are included in this analysis, the addition of more cells and the improved 10X Genomics V3 chemistry (yielding a higher number of genes and transcripts detected per cell) allowed us to better distinguish between CAFs and pericytes (Figure 25c-d). In addition, we were also able to detect a new cell population composed of neutrophils (Figure 25c-d), an important immune cell type that has recently gained significant attention for its role in cancer progression and for its systemic effects[276].

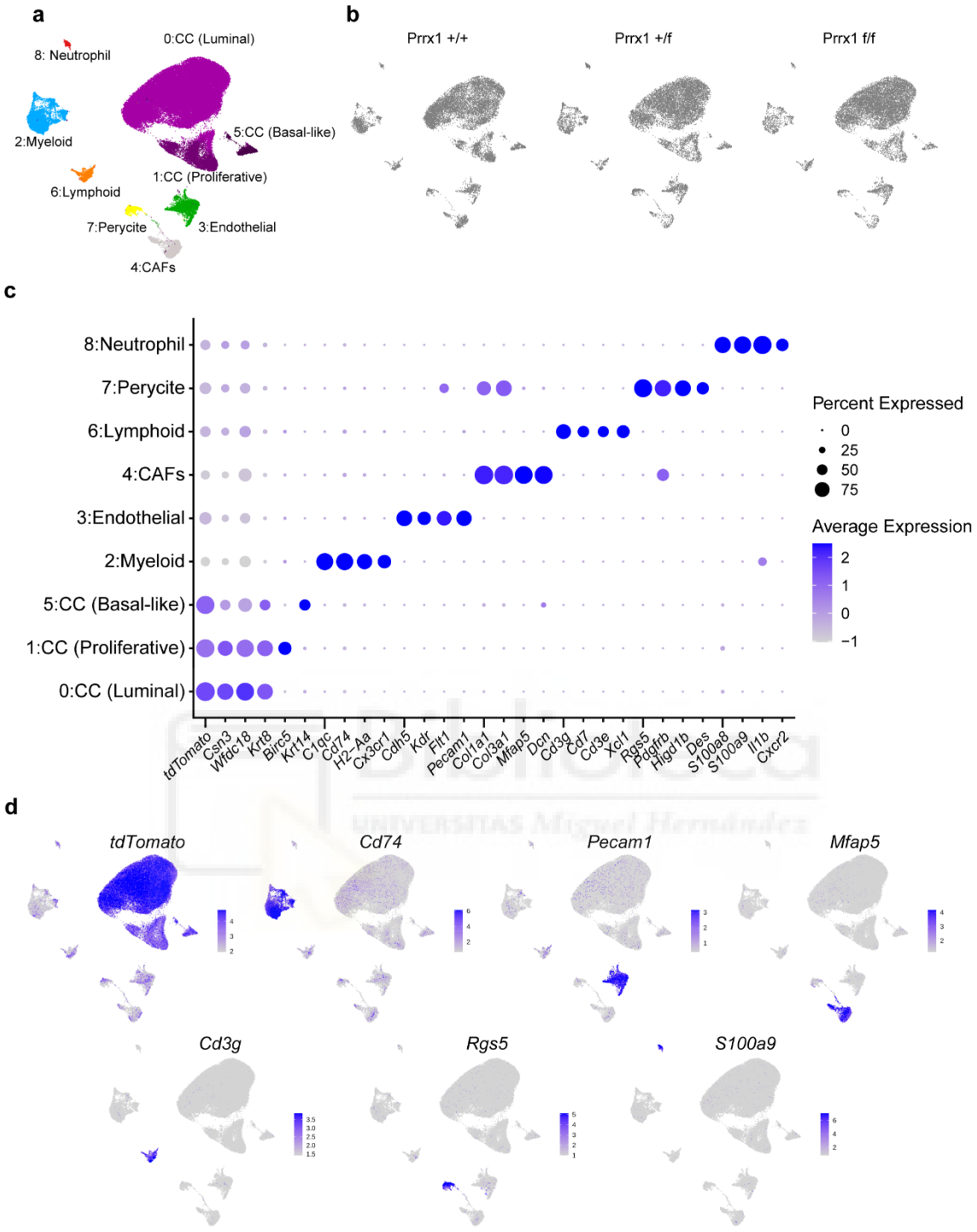


Figure 25. Single cell transcriptomics analysis of cell populations in PyMT primary tumors. a, Combined UMAP projection of scRNA-seq analysis of PyMT tumors of the following genotypes: *Prrx1* +/+ (n=7), *Prrx1 +/f* (n=3) and Prrx1 f/f (n=4) and identification of the different cell types. **b**, UMAP projection of the same tumors showing the contribution of the different genotypes with the same number of cells per condition. **c**, Dot plot showing genes specifically expressed in the different cell types. **d**, UMAPs showing the expression of cell type specific markers (see panel c).

9. Intermediate *Prrx1* levels promote an invasive EMT cell state

Focusing on cancer cells, we subset them and identified 28 subpopulations (Figure 26a). As our aim was to study the mechanism behind the non-linear correlation between different *Prrx1* levels and metastatic burden, and since *Prrx1* expression in cancer cells wasl confined to the invasive regions in tumors, we focused our comparative analysis on those cells. To identify the invasive population, we developed a Spatial Transcriptomics-derived Invasive Signature (STINS) using the differentially regulated genes (log2FC > 0.25 and Pvalue < 0.05) in the cancer cells residing in the invasive areas compared to those located elsewhere (Figure 24f; See methods) and applied it to the scRNA-seq data, identifying clusters 20 and 25 as enriched for invasive features (Figure 26b). These invasive cell populations present the highest levels of EMT activation in the tumor (Figure 26c). In addition, Clusters 20 and 25, are enriched in the pro-invasive gene signature (PINGs)[42] that we described recently. This signature can discriminate between invasive and non-invasive human breast cancer cell lines (Figure 26d). Critically, *Prrx1* expression was only detected in cluster 25 (Figure 26e). This cluster has the lowest score for epithelial and highest for mesenchymal signature (Figure 26e), exemplified by *Epcam* and *Vim,* respectively (Figure 26e). This is further supported by the detection of Vimentin protein expression in K14+ invasive cells (Figure 26g).

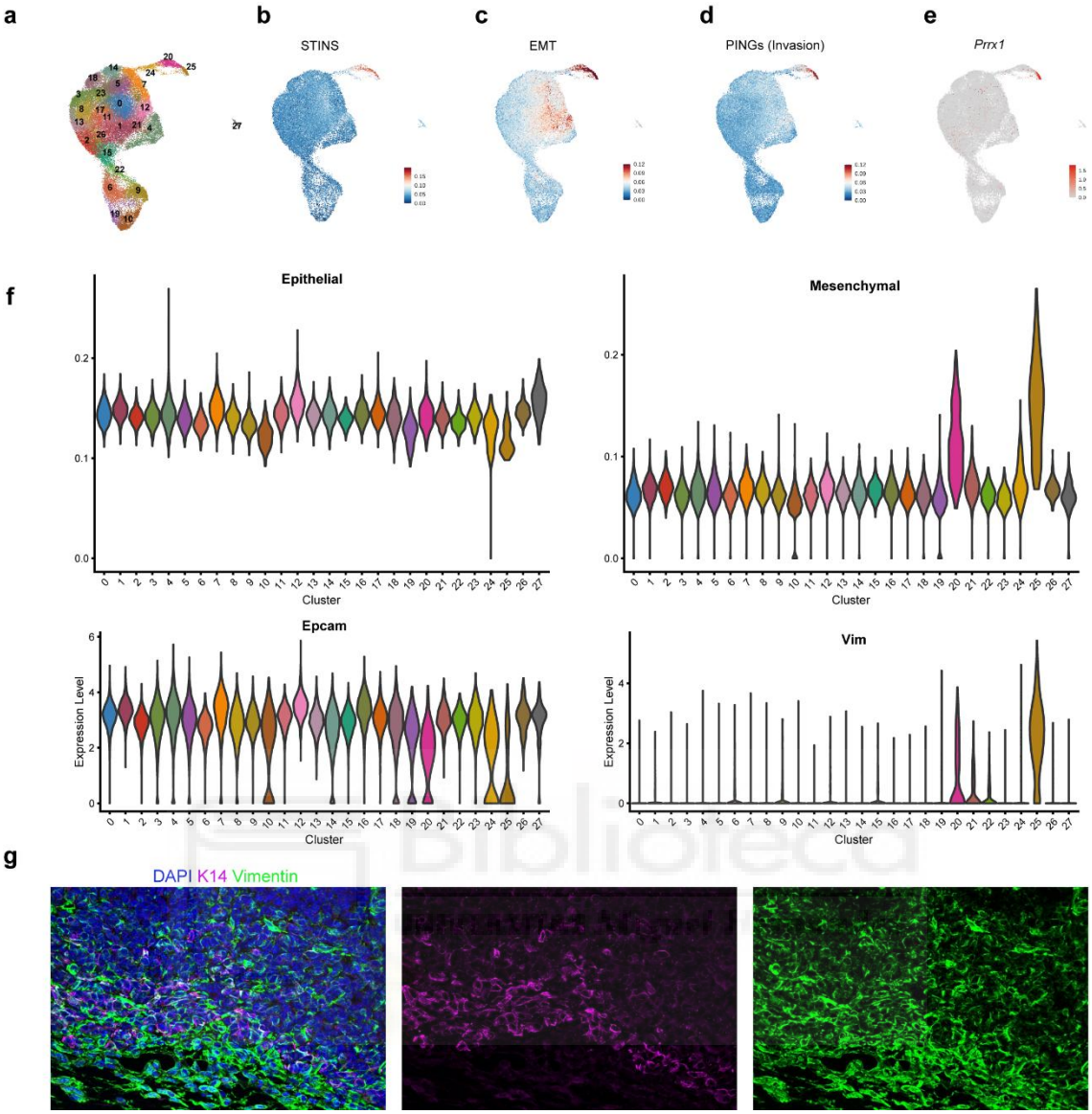


Figure 26. Single cell transcriptomics analysis of cancer cell states in PyMT primary tumors. a, UMAP projection of scRNA-seq analysis of cancer cell subset for *Prrx1* +/+ (n=7); *Prrx1* +/f (n=3), and *Prrx1* f/f (n=4) tumors. **b**, Enrichment of STINS (spatial transcriptomic-derived invasive signature) represented over UMAP. **c**, Enrichment of EMT hallmark represented over UMAP **d**, UMAPs projection showing the enrichment in PING (pro-invasive) signature[19]. **e**, Normalized *Prrx1* expression represented over UMAP. **f,** Violin plots showing the expression of Epithelial signature, *Epcam*, Mesenchymal signature and *Vim* in all cancer cell clusters. Cluster 25 is the lowest for *Epcam* and highest for *Vim* consistent with its EMT phenotype. **g,** Representative image of an invasive edge showing the expression of K14 and Vimentin together and in single channels. Scale bars: 50 µm

We next examined how the identified invasive cell populations, clusters 20 and 25, aligned with our recently described EMT trajectories in PyMT tumors[42]. To this end, we generated gene signatures specific to each previously identified cell state along the trajectory (Figure 27a) (see Methods) and applied them to the new data, using an approach similar to that used for STINS (Figure 27b). We confirmed the distribution of

the two programs, the inflammatory adult-like EMT and the embryonic-like EMT, in distinct cancer cell populations, separate from differentiated cell states (Figure 27b). Notably, we verified that the invasive cell populations identified with STINS closely aligned with the final stages of the invasive embryonic-like EMT trajectory[42].

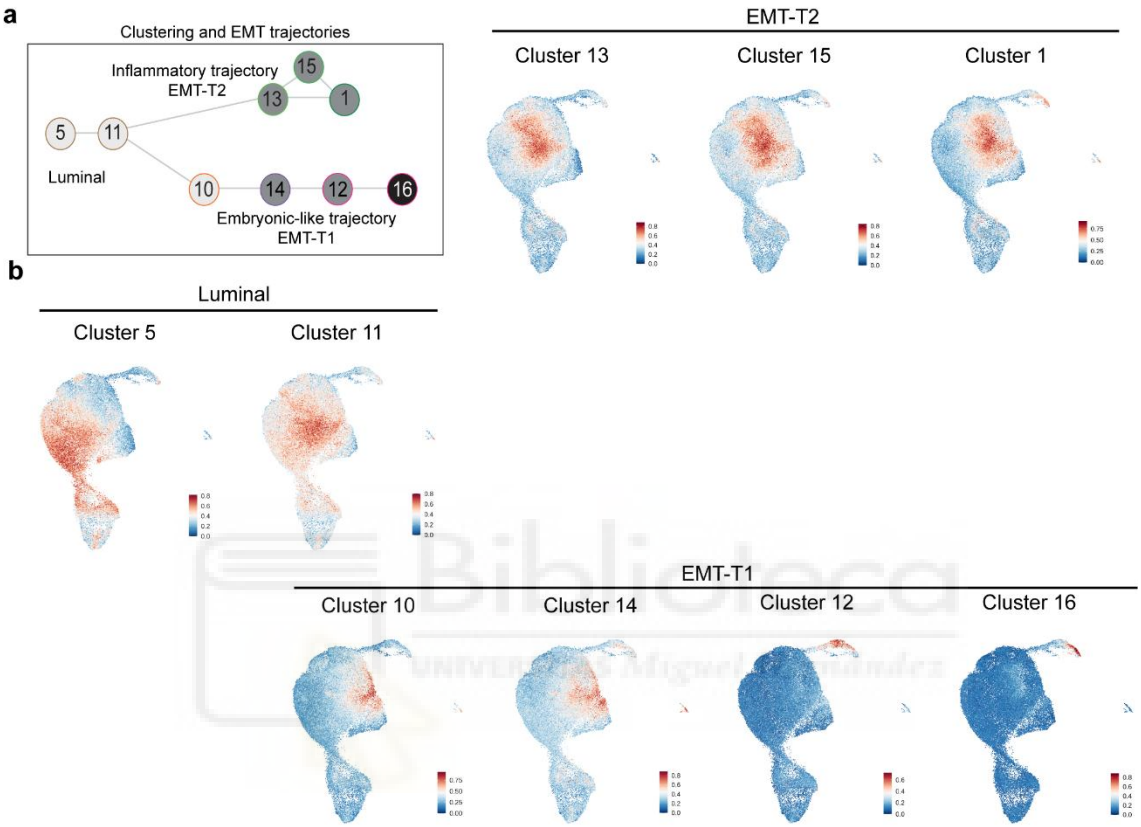


Figure 27. Invasive clusters correspond to the late stages of the embryonic-like EMT trajectory. a, Structure of previously identified EMT trajectories including cluster numbers[42]. **b**, UMAPs projections showing the enrichment of signatures corresponding to the clusters previously identified along the EMT trajectories in PyMT tumors. Besides using a new chemistry for single cell library preparation, we can identify the corresponding clusters. Specially relevant for this work, clusters 20 and 25, annotated as invasive using spatial transcriptomics-based signature (Figure 4b), correspond to the late embryonic-like EMT-T1 clusters 12 and 16 in the figure[42].

As we did not see differences in the distribution of the clusters (Figure 28a-b), and given the established link between *Prrx1* and EMT, we hypothesized that *Prrx1* levels could modulate the transcriptome of the invasive cells, influencing their metastatic potential. To explore this hypothesis, we used published epithelial and mesenchymal gene signatures[245], which we previously used to classify 71 human breast cancer cell lines based on their EMT status[42]. Comparative analysis of cluster 25 cells revealed that while the complete loss of *Prrx1* (*Prrx1* f/f) significantly reduced EMT features (significantly

increasing the epithelial score) and invasive signatures (Figure 28c), *Prrx1* +/f cells retained an invasive EMT phenotype despite their lower *Prrx1* expression levels (Figure 28c). These findings suggested that intermediate *Prrx1* levels are sufficient to sustain an EMT-driven invasive program, compatible with a partial and invasive EMT phenotype previously described as highly metastatic[99,100,171]. This was also consistent with prior observations showing that early stages of EMT induction, with low *Prrx1* levels, already confer invasive properties, as demonstrated in TGFβ-treated MDCK cells[42]. However, this did not completely align with the non-linear correlation between *Prrx1* levels and the size of invasive areas in the primary tumor (Figure 19a), an observation we have previously discussed, as *Prrx1* +/f cells did not show an increased invasion score, at least with the current analysis (Figure 28c). Therefore, while EMT status or invasion programs could explain the reduced metastatic competence of *Prrx1* negative tumors, they did not fully account for the enhanced metastatic competence of tumors with intermediate *Prrx1* levels.

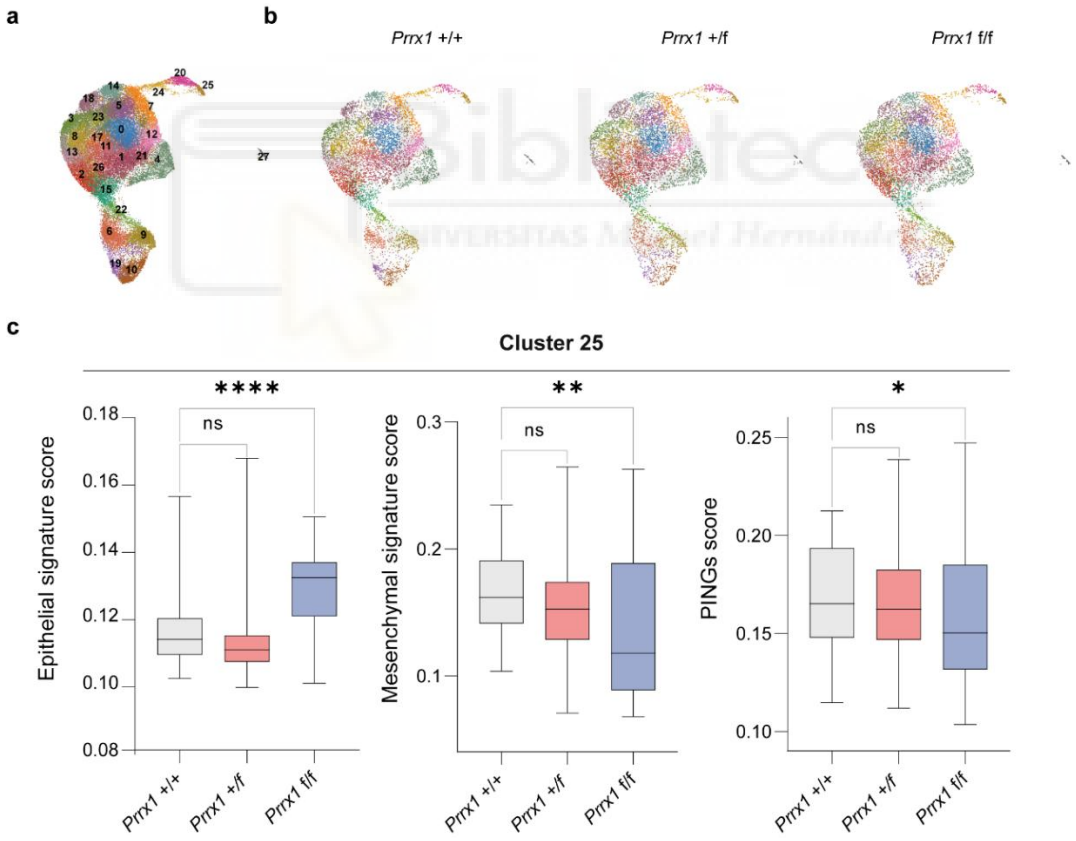


Figure 28. Intermediate *Prrx1* levels promote an invasive EMT cell state. a, UMAP projection of scRNA-seq analysis of cancer cell subset of *Prrx1* +/+ (n=7); *Prrx1* +/f (n=3), and *Prrx1* f/f (n=4) tumors. **b,** Combined UMAP projection of scRNA-seq analysis of PyMT tumors of the following genotypes: *Prrx1 +/+* (n=7), *Prrx1 +/f* (n=3) and Prrx1 f/f (n=4). **c,** Score of Epithelial, Mesenchymal and PING signatures of cells in cluster 25 for the different tumor genotypes. Boxplots show median, upper and lower quartile, whiskers show upper and lower extreme. Mann Whitney test. ∗ p <0.05; ∗∗ p < 0.01; ∗∗∗∗ p < 0.0001; p > 0.05, not statistically significant (ns).

10. Intermediate *Prrx1* levels promote a highly proliferative state in invasive cancer cells

As EMT or invasive features did not seem to fully explain the increase in metastatic burden of the *Prrx1* +/f mice, we sought to investigate additional mechanisms that *Prrx1* could be playing a role. To do so, we compared the transcriptomic profile of cluster 25 cells between *Prrx1* +/+ and *Prrx1* +/f tumors using *muscat*[236], an optimized tool for identifying differentially expressed genes in inter-conditional multi-sample comparisons. Pathway enrichment analysis revealed that genes upregulated in *Prrx1* +/f cells were primarily associated with proliferation and cell cycle regulation (Figure 29a). This finding was further supported by cell cycle scoring (Reactome) using the whole transcriptome of the cells, which indicated that *Prrx1* +/f invasive cells are significantly more proliferative than their *Prrx1* +/+ counterparts (Figure 29b). To confirm that the differences detected at the RNA level reflect actual changes in cell proliferation, we performed immunofluorescence labelling of mitotic cells using p-H3 and quantified the ratio of dividing cancer cells at the invasive edge in *Prrx1* +/+ and *Prrx1* +/f tumors. This analysis confirmed that invasive *Prrx1* +/f cells are more proliferative than their *Prrx1* +/+ counterparts (Figure 29c). Importantly, this observation provided a plausible explanation for the enlarged invasive areas observed in *Prrx1* +/f tumors (Figure 19a).

We propose that enlargement of invasive areas results from the expansion of cells residing within these areas, rather than an increased invasive potential of the cells themselves. However, we would have expected to see an obvious increase in the size of cluster 25 in the *Prrx1* +/f tumors. Due to the small size of this cell population in the tumor (2-3%) and their localization (not homogeneous distribution, therefore highly affected by sampling), a higher number of cells per sample and more samples are required to be able to measure a difference in a robust manner. Consistent with an effect dependent on *Prrx1* dose, the proliferative advantage is restricted to the invasive cell population, as we did not detect any difference in total tumor burden or global proliferation of cancer cells between in the different genotypes (Figure 17a-b).

Intrigued by this differential proliferative capacity, we further investigated the transcriptomic profile of cluster 25 to identify potential cell cycle regulators expressed in this population. We found a notable enrichment of *Ccnd2* and *Cdkn2* cell cycle inhibitors (Figure 29d). Interestingly, *Ccnd2* was among the upregulated genes in the cells within invasive areas in our spatial transcriptomics analysis (Figure 24c), further supporting the idea of a specific Cyclin D code for invasive cells. On the other hand, *Cdkn2a/b/c* are

fundamental cell cycle regulators. Specifically, they bind Cdk4/6+Cyclin Ds complexes and inhibit their kinase activity, thereby preventing progression through the G1/S checkpoint and inducing cell cycle arrest in G0/G1[270,277]. Indeed, *Cdkn2a/b* are powerful tumor suppressors, and there is strong negative selection against them during tumor evolution. Evidence of this is reflected in the fact that *CDKN2A* and *CDKN2B* deletions are respectively the most and third most common genes depicting copy number alterations in a recent pan-cancer genomic study[278]. Overall, these findings suggest that while high *Prrx1* expression maintained invasive cells in a low-proliferative state, intermediate levels sustained invasive potential while allowing proliferation, facilitating metastatic colonization. This mechanism could explain the higher metastatic incidence observed in our genetic mouse models for the *Prrx1* heterozygous condition and in breast cancer patients bearing tumors with intermediate *Prrx1* levels.

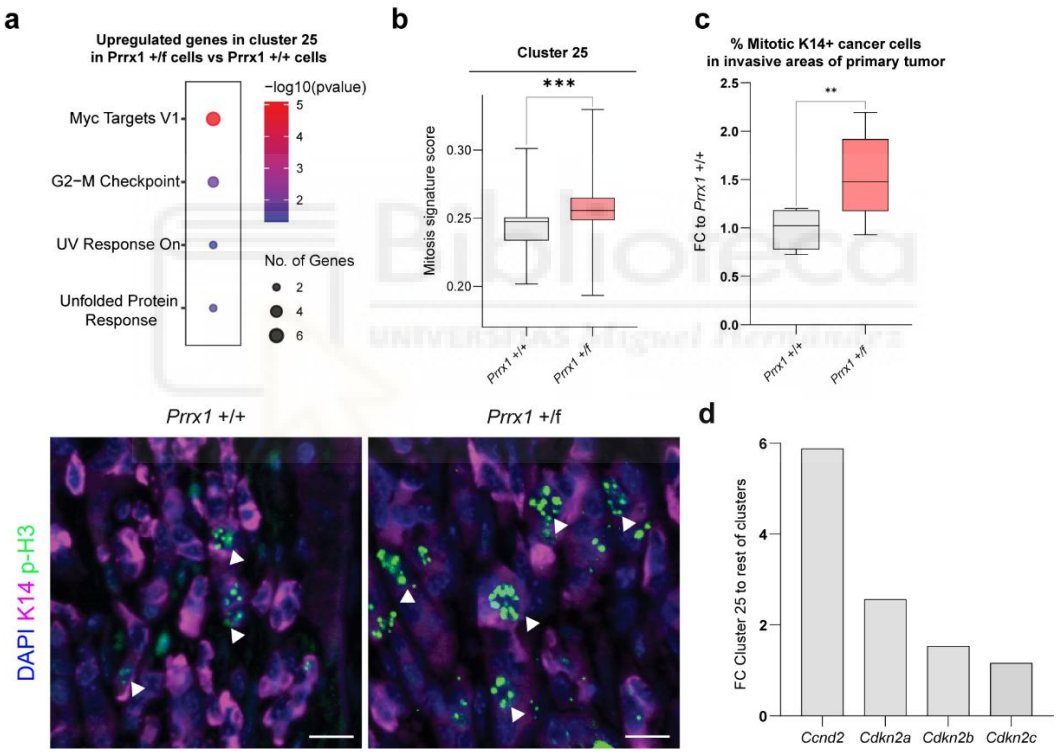


Figure 29. Intermediate *Prrx1* levels promote an invasive and highly proliferative state in cancer cells. **a,** Dot plot showing pathway enrichment (MsigDB_Hallmarks) analysis of differentially upregulated genes ($p < 0.05$) determined by MUSCAT of *Prrx1* +/f vs *Prrx1* +/+ cells in cluster 25. **b,** Score of cell cycle signature (Cell_Cycle_Mitotic_R-HAS 69278) on cluster 25 cells for the different tumor genotypes. Boxplots show median, upper and lower quartile; whiskers show upper and lower extreme. Mann Whitney test. *** $p < 0.001$. **c**, Representative images and quantification of mitotic rate of K14+ cancer cells at invasive edges of primary tumors. n= 9 edges of *Prrx1* +/+ tumors from 3 mice; n= 11 edges of *Prrx1* +/f tumors from 3 mice. Boxplots show median, upper and lower quartile, whiskers show upper and lower extreme. Mann-Whitney test. ** $p < 0.01$. Scale bar: 50µm. **d**, Enrichment in the expression of cell cycle regulators in cluster 25 (scRNA-seq) relative to all the other clusters.

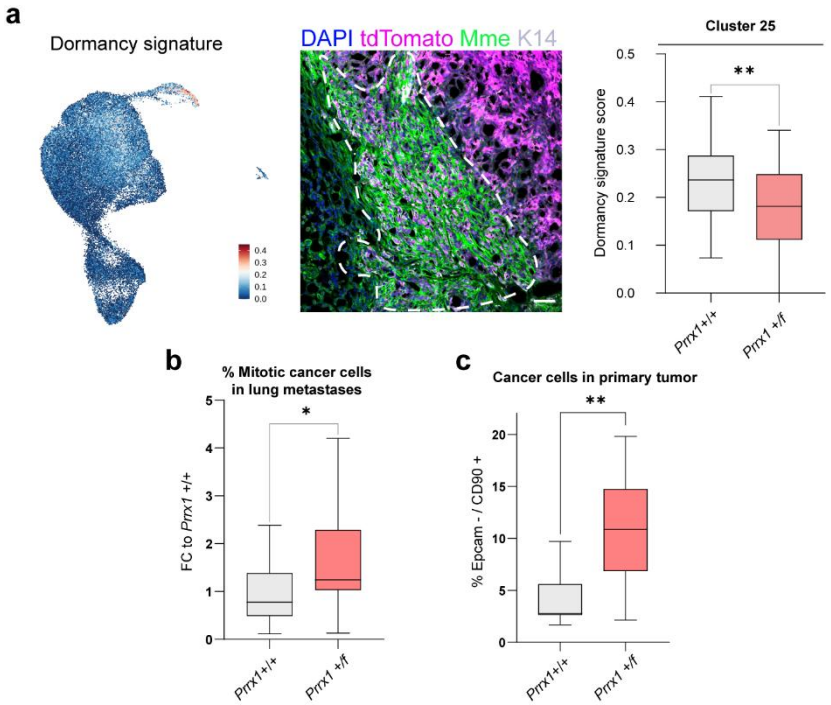
11. Intermediate *Prrx1* levels reduce dormancy in invasive cancer cells

Given that cluster 25 invasive cancer cells were enriched in cell cycle inhibitors, and cell cycle arrest is one of the main characteristics of dormancy[279], we wondered if they also expressed dormancy-associated genes. Using a breast cancer dormancy signature recently derived after comparing cycling and non-cycling metastatic PyMT cells, and that can discriminate between the dormant and non-dormant versions of the same cell line (D2.A1 vs D2.OR cell lines)[194], we found that the invasive clusters (20 and 25) had the highest dormancy scores in the tumor, and cluster 25 in particular (Figure 30a). Immunofluorescence analysis of Mme expression, a marker of the dormancy program[194], confirmed high expression in invasive areas (Figure 30a), indicating that dormancy and invasion are coregulated in tumors. It is important to highlight that this dormancy signature does not contain *Prrx1* or any cell cycle regulators, so our gene of interest or cell cycle genes were not biasing the analysis toward our population of interest.

Notably, cluster 25 from *Prrx1* +/f tumors exhibited significantly lower levels of the dormancy score than that in *Prrx1* +/+ tumors (Figure 30a), suggesting that, independently of cell cycle regulation, intermediate *Prrx1* levels could allow invasive cells to avoid dormancy. Dormancy typically refers to a state found in disseminated tumor cells at metastatic sites rather than in the primary tumor. To assess whether the described transcriptomic differences in dormancy observed in the invasive cells of the primary tumor impacted metastasis, we quantified the proportion of dividing cancer cells in lung metastases as an indirect proxy of dormancy, a common approach in the field[201,220]. Immunofluorescence analysis revealed that metastases derived from *Prrx1* +/f tumors contained a higher fraction of mitotic cells compared to those from *Prrx1* +/+ tumors (Figure 30b). This finding aligned with the increased size of metastatic foci observed in mice harboring *Prrx1* +/f tumors (Figure 18) and suggested that the proliferative status observed in the invasive regions (Figure 29c) influenced metastatic growth. However, the difference in proliferation between the two genotypes was less pronounced in metastases than at the invasive edge of the primary tumor. We interpreted this as a consequence of our analyses of metastases at an advanced stage, where the initial proliferative priming may already be diluted. Indeed, previously discussed results have shown that metastases recapitulated the EMT hierarchy of the primary tumor (Figure 21 a-c) and that different *Prrx1* levels did not affect global primary tumor proliferation (Figure 17a-b). It is possible that, at very early stages of metastatic colonization, especially at the level of disseminated single cells, the difference in metastatic proliferation between *Prrx1* +/f and *Prrx1* +/+ cells would be more pronounced. However, the PyMT model

does not allow enough temporal resolution to assess metastasis initiation, due to the asynchronous growth of several tumors in the same mouse. Thus, we could not find an optimal timepoint with enough disseminated cancer cells and prior to macrometases formation. Together, these results suggest that the specific proliferative and dormant status of invasive cells within the primary tumor determines their future metastatic competence.

Finally, we wanted to determine whether, apart from differences in proliferation and dormancy, *Prrx1* levels also affected the tumor-initiating capacity of this invasive population. To investigate this, we analyzed CD90 expression, a well-characterized marker of metastatic-initiating cells in the PyMT model[280]. Previous studies have characterized this CD90+ population, showing that it was particularly enriched for mesenchymal markers, with significant co-expression of K14, the marker we used to identify invasive regions and was enriched in the CTC population, and Vimentin, a mesenchymal marker[281]. All of this suggests that our *Prrx1*+ invasive population could be an important source of these metastatic-initiating cells. We performed FACS analysis on freshly dissociated cancer cells and confirmed that EpCAM- cancer cells, which represent an advanced EMT state in the tumors (Figure 20d) and coincide with invasive clusters 20 and 25 (Figure 26f), presented a higher fraction of CD90+ cells in *Prrx1* +/f tumors when compared to control *Prrx1* +/+ tumors (Figure 30c). This indicates that intermediate *Prrx1* levels promote metastasis-initiating capabilities, compatible with previous findings in the lab in cultured human cell lines, where *PRRX1* downregulation increased the $CD44^{+}/CD24^{-}$ ratio, enhancing sphere formation[51].



12. Chromatin accessibility in PyMT tumors

To get more insight into the regulatory landscape of the invasive population, specifically how *Prrx1* regulates cell cycle progression and dormancy, we profiled chromatin accessibility using the single nucleus assay for transposase-accessible chromatin sequencing (snATAC-seq) and analyzed chromatin dynamics in four *Prrx1* +/+ control tumors (Figures 31 and 32). This assay reveals which chromatin regions are accessible (targeted by a transposase) at a given time and at the single-cell level. This enables the study of different epigenetic states within the same tumor and allows advanced analysis of gene regulation[282].

Clustering of the generated open chromatin data identified six main clusters (Figure 32a), with a consistent cell distribution across samples (Figure 32b). To annotate these clusters into major cell types, we used the gene activity of *bona fide* cell type markers (Figure 32c). Gene activity serves as a proxy for gene transcription by quantifying the number of reads (and thus, accessible regions) in gene promoters. While this is a simplified approach to studying gene regulation (other genomic regions, such as enhancers, play crucial roles in initiating transcription), it is expected that when comparing clearly distinct cell populations or cell types, lineage-specific genes will also be heavily regulated by promoter accessibility, as it can be clearly observed in Figure 32d. We found that the composition and contribution of different cell types in the tumors differed significantly from those found using scRNA-seq (Figure 32a). This discrepancy likely stemmed from the differing dissociation protocols used in each experiment. In the case of scRNA-seq, the final product was intact cells, whereas snATAC-seq isolated nuclei were obtained through a mechanical process. Different cell types exhibit varying mechanical properties and thus respond differently to mechanical stress. Since our protocol was optimized for cancer cell nuclei extraction, other cell types may have been more severely affected, resulting in the isolation of fewer nuclei from those populations.

Figure 30. High Prrx1 levels activates a dormancy programme. a, Dormancy signature[194] represented over the UMAP. Representative image showing Mme (part of the dormancy signature) expression in tumors, showing specificity for K14+ invasive areas. Scale bars: 50 μm. Score of Dormancy signature of cells in cluster 25 from *Prrx1* +/+ and *Prrx1* +/f tumors. Boxplots show median, upper and lower quartile; whiskers show upper and lower extreme. Mann Whitney test. ∗∗ p< 0.01. **b,** Mitotic rate of cancer cells in lung metastases. *Prrx1* +/+ n= 24 metastases from 7 mice; *Prrx1* +/f n= 19 metastases from 6 mice. Boxplots show median, upper and lower quartile; whiskers show upper and lower extreme. Two-tailed unpaired test. ∗ p <0.05. **c,** Quantification by FACs of CD90 status in Epcam- cancer cells from *Prrx1* +/+ (n=9 tumors from 3 mice) and *Prrx1* +/f (n=14 tumors from 4 mice) tumors. Boxplots show median, upper, and lower quartile; whiskers show upper and lower extreme. Mann Whitney test. ∗ p< 0.05, ∗∗ p< 0.01.

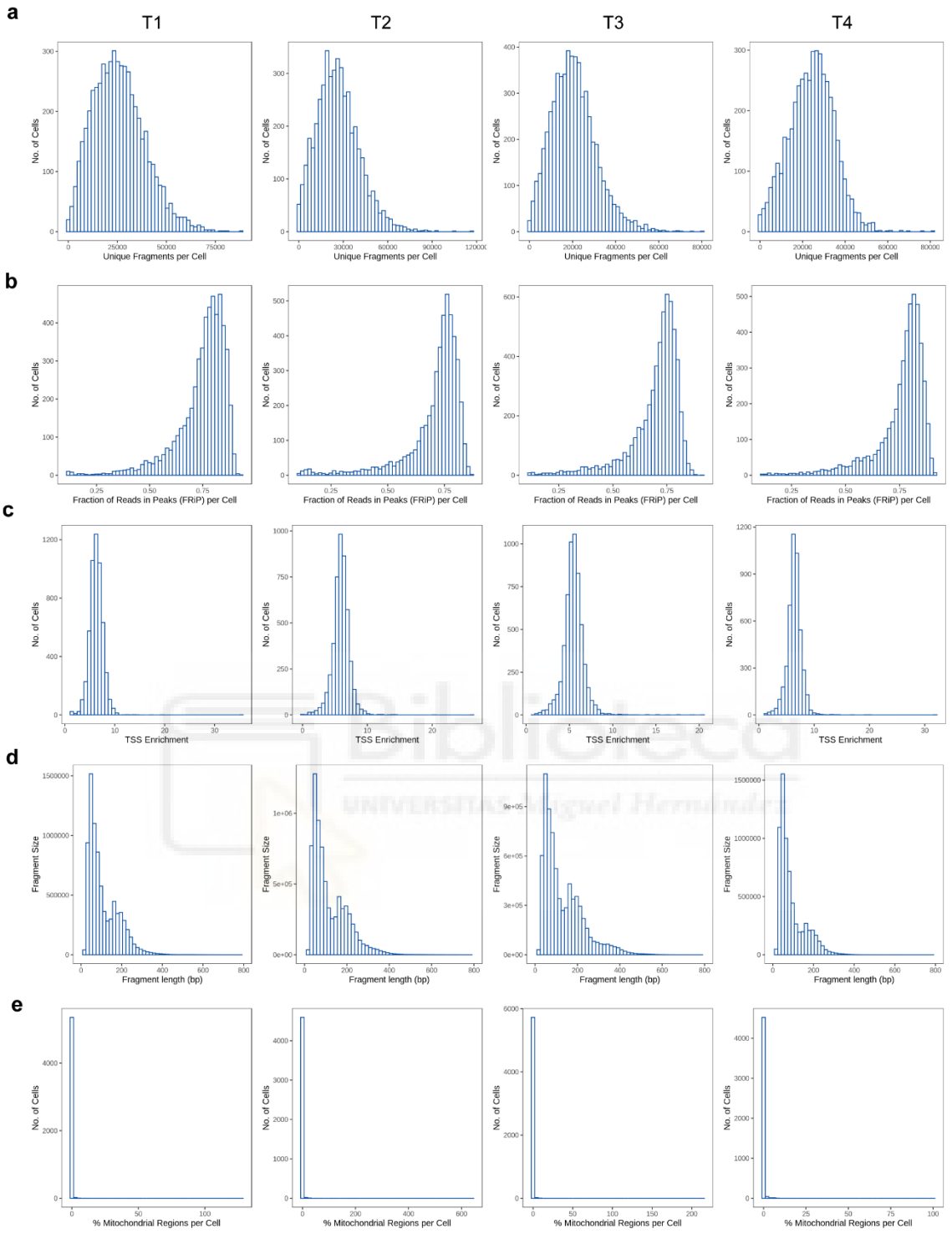


Figure 31. Quality control of single cell ATAC-seq data from PyMT tumors. Distribution of different parameters in each tumor sample (T1-T4). **a,** Number of unique fragments per cell. **b,** Fraction of reads in Peaks (FRiP) per cell. **c,** TSS enrichment score. **d,** Fragments length (bp). **e,** Percentage of mitochondrial regions per cell.

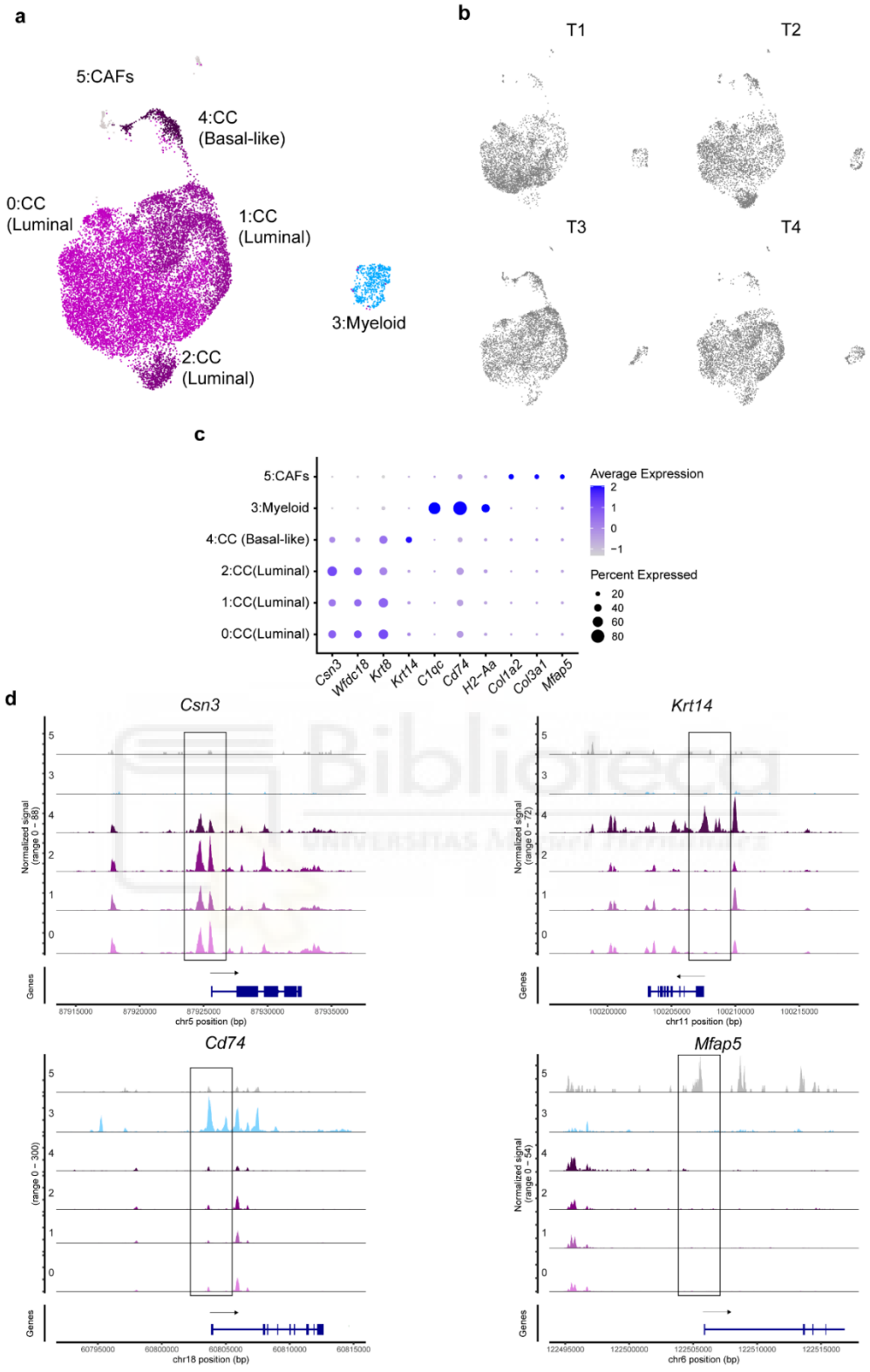


Figure 32. Single nucleus ATAC-seq analysis to assess chromatin accessibility in PyMT tumors. a, UMAP projection of scATAC-seq analysis of PyMT tumors (n=4), with annotation of the different cell clusters. **b**, Contribution of the different genotypes to the combined UMAP shown in (a) with the same number of cells per condition. **c**, Dot plot showing genes specifically expressed in different cell populations. **d**, Coverage plots showing open chromatin regions around the promoters of representative genes for each annotated cell population.

13. Chromatin accessibility analysis establishes *Prrx1* as a direct regulator of proliferation and dormancy in the invasive population.

As for the scRNA-seq analysis, we focused on the cancer cell population and identified 16 distinct clusters (Figure 33a). To identify the invasive cells, we first imputed gene expression into the snATAC-seq cells using our scRNA-seq data as a reference (see Methods). This approach allowed us to predict the transcriptomic profile of the cells based on their chromatin accessibility. We then applied the STINS signature and observed that two clusters, 14 and 15, emerged as the invasive populations (Figure 33b), reminiscent of the invasive clusters 20 and 25 found when applying STINS to our scRNA-seq data (Figure 26b)Figure 33. Importantly, cluster 15 exhibited the highest *Prrx1* imputed expression (Figure 33c), mirroring cluster 25 in scRNA-seq (Figure 26d). As *Prrx1* is a transcription factor, we examined if it could be acting as a master regulator this invasive cell population. We analyzed the accessibility of *Prrx1* binding sites across the genome and the different cancer cell clusters and found that the sites were significantly more accessible in cluster 15 compared to the rest (Figure 33d). This suggested that *Prrx1* expression in the invasive cells profoundly impacted their cell state, as its target genes (where *Prrx1* binding sites are located) became more accessible and, therefore, predicted to be more active than in other cancer cells.

Since we were particularly interested in the regulatory role of *Prrx1* on cell cycle regulators and dormancy-associated genes, we further explored these aspects using the snATAC-seq data. Further analysis of the chromatin state of cluster 15 revealed that several accessible regions in cluster 15 located near the promoters of genes encoding cell-cycle regulators (cyclins and cyclin inhibitors) and dormancy-related genes[194] (*Gas6*, *Mme* (Figure 33e,f).

These results indicate that Prrx1 is not only specifically expressed in the invasive population but also that it can play a master regulatory role, as in addition to its role in regulating invasion[39], it can directly bind promoters of genes associated with cell-cycle control and dormancy, providing mechanistic evidence for the differences in metastatic potential observed at different Prrx1 expression levels.

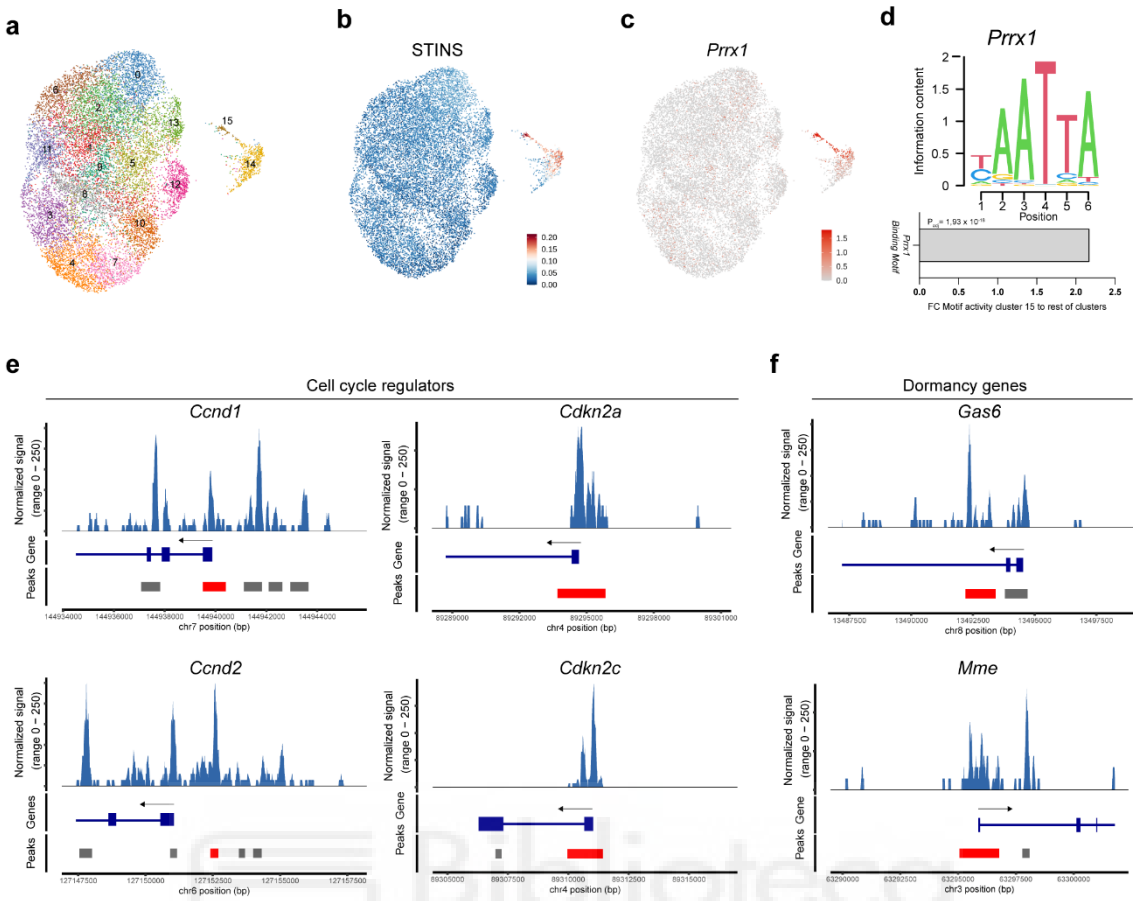


Figure 33. Chromatin accessibility analysis establishes *Prrx1* as a direct regulator of proliferation and dormancy in the invasive population. a, UMAP of cancer cells from *Prrx1* +/+ tumors (n=4) as assessed by snATAC-seq analysis. **b**, Enrichment of STINS (spatial transcriptomic-derived invasive signature) based on imputed expression represented over the UMAP (see Methods). **c**, Normalized imputed expression of Prrx1 over the UMAP. **d**, Prrx1 binding motif and relative enrichment of motif activity in cluster 15 compared to the rest of the clusters. **e**, Coverage plots of identified peaks in cells from cluster 15 linked to genes encoding regulators of cell proliferation: cyclins (*Ccnd1*/*2*) and cyclin Inhibitors (*Cdkn2a*/*2c*). Red lines, peaks with Prrx1 binding sites. **f**, Similar analysis as that in (e) but with dormancy associated genes.

We next checked whether our recently described EMT trajectories in PyMT tumors[42] had a correlate at the chromatin level. To this end, we applied our previously identified cell state specific gene signatures along the trajectory to the snATAC-seq data based on the calculated imputed expression (Figure 34a; see Methods). We found a similar distribution, with inflammatory adult-like and embryonic-like EMTs, suggesting that distinct transcriptional EMT states are supported by specific chromatin profiles (Figure 34b). Therefore, this snATAC-seq data can be used to further characterize these EMT trajectories. They can be used to identify key regulators of the transitions, as chromatin analysis is the gold standard to generate gene regulatory networks[283]. In addition, this analysis revealed that invasive cell populations had chromatin profiles that closely aligned with the final stages of the invasive embryonic-like EMT trajectory (Figure 34b)[42].

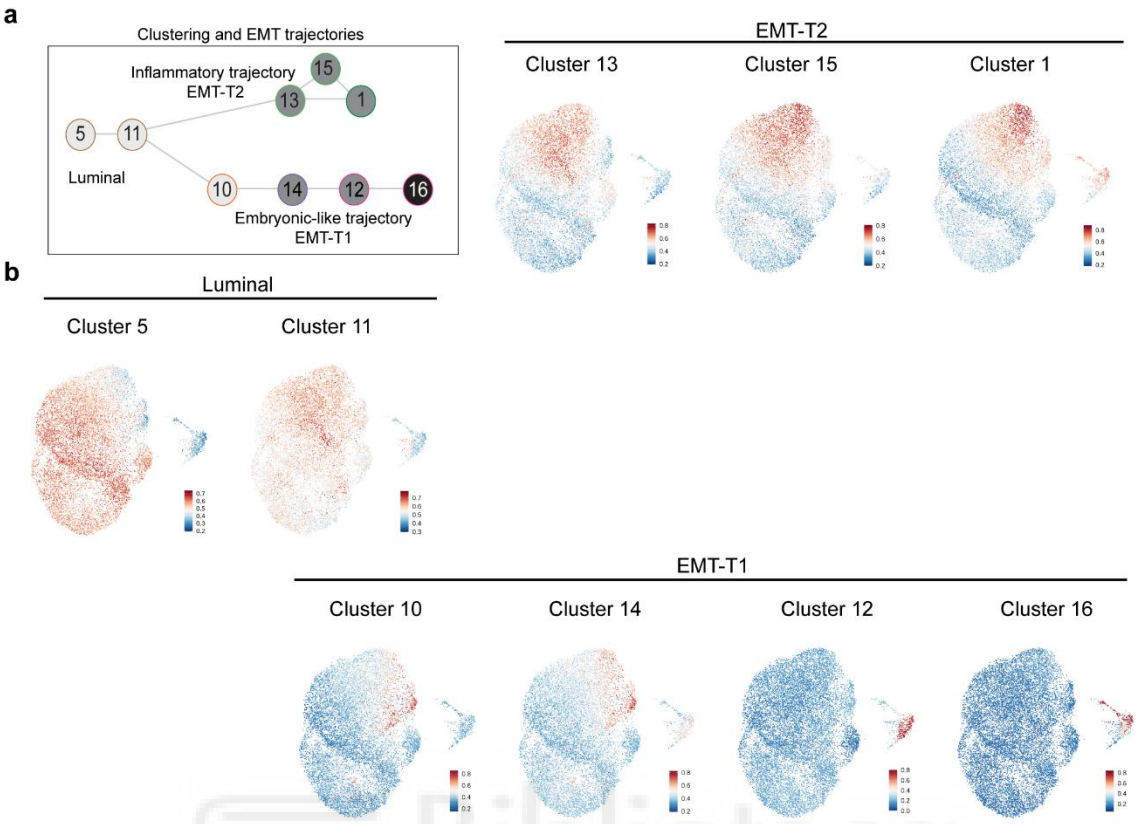


Figure 34. Single cell chromatin accessibility analysis of EMT trajectories. a, Structure of previously identified EMT trajectories[39]. **b**, UMAP plots showing the enrichment of signatures derived from cell states along the EMT trajectories based on imputed expression. Chromatin states of the invasive clusters (14 and 15), identified using spatial transcriptomics-based gene signature (Figure 23c), correspond to the late embryonic-like EMT- T1 clusters 12 and 16 in the figure[42].Our data indicate that distinct transcriptional EMT states are supported by specific chromatin profiles.

14. Engineered cell line models reveal dose dependent Prrx1 regulation of invasion, proliferation, and dormancy

To examine whether this regulatory pattern was conserved in human, we analyzed whole-genome transcriptomes from 71 different human breast cancer cell lines[246], grouped by EMT status according to epithelial and mesenchymal gene signatures previously described[42,245] (Figure 35a). We observed a clear correlation between *PRRX1* levels, invasion (PING), and dormancy signatures (Figure 35a). This observation further supported the co-regulation of an invasive and dormancy program in cancer cells, as we observed by scRNA-seq in tumors, where cells with the highest enrichment for PING (invasion) also displayed the highest levels of the dormancy score. In addition, the high-EMT cell group also showed higher levels of the *CDKN2* cell cycle inhibitors, reminiscent

of our results in tumors, where these genes were highly enriched in the *Prrx1*+ invasive population.

To specifically assess if PRRX1 could be regulating these genes in a human setting, we selected the BT549 cell line, as it resembles cluster 25 in our scRNA-seq data, with high levels of PRRX*1*, dormancy-associated genes, and cell cycle inhibitors (Figure 35a). Furthermore, we had previously shown that BT549 cells were not metastatic in experimental tail vein metastasis assays, but that became highly metastatic upon *PRRX1* downregulation[51]. We took advantage of a BT549 clone previously generated in the lab that contained a short hairpin RNA interference construct targeting *PRRX1* (sh*PRRX1*)[51] and confirmed that *PRRX1* downregulation led to an increase in cyclin (*CCND1)* transcription and a reduction in the expression of genes encoding both cyclin inhibitors of the CDKN2 family and dormancy-associated markers (Figure 35b).

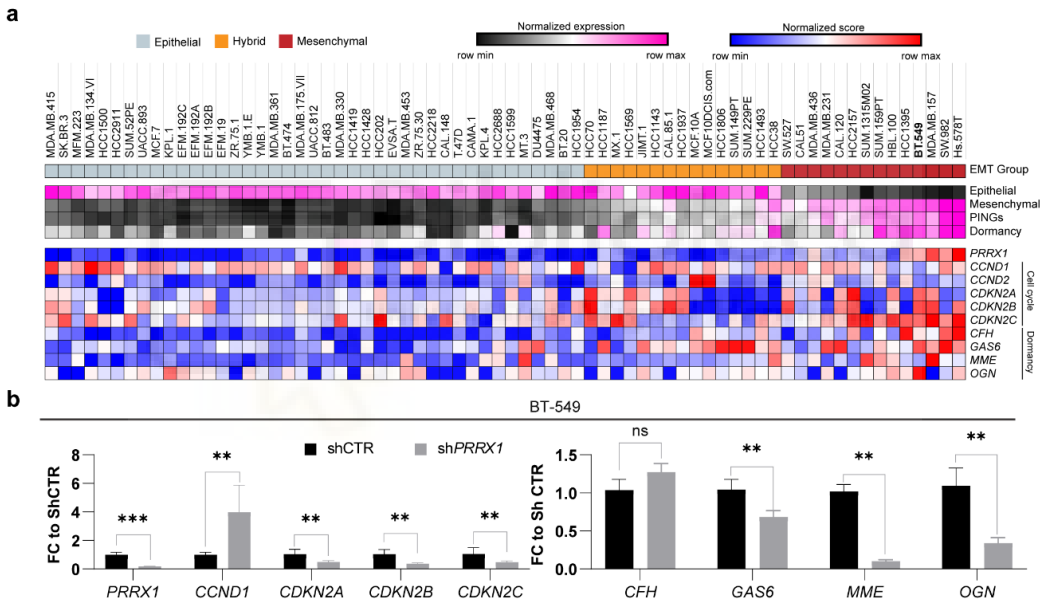


Figure 35. *PRRX1* activates the expression of cell cycle inhibitors and dormancy associated genes in human breast cancer cell lines. a, Expression analysis of signatures and genes of interest in 71 breast cancer cell lines ordered according to their epithelial-to-mesenchymal score[42]. **b,** RT-PCR of cell cycle regulators (left) and dormancy-related genes (right) in shCTR vs *shPRRX1 BT549* cells. FC is represented as mean ± SEM (n= 6). Mann-Whitney test. ∗ p <0.05; ∗∗ p < 0.01; p > 0.05, not statistically significant (ns).

Finally, the analysis of Chip-seq dataset in different human breast cancer cell lines[284] showed that PRRX1 can directly bind the promoters of cell cycle regulators, dormancy and invasion genes in human breast cancer cell lines (Figure 36). These findings confirmed that *PRRX1* was an activator of cell cycle inhibitors and dormancy-associated genes in human breast cancer cells and provided a mechanistic explanation for the requirement of *PRRX1* downregulation for metastatic outgrowth.

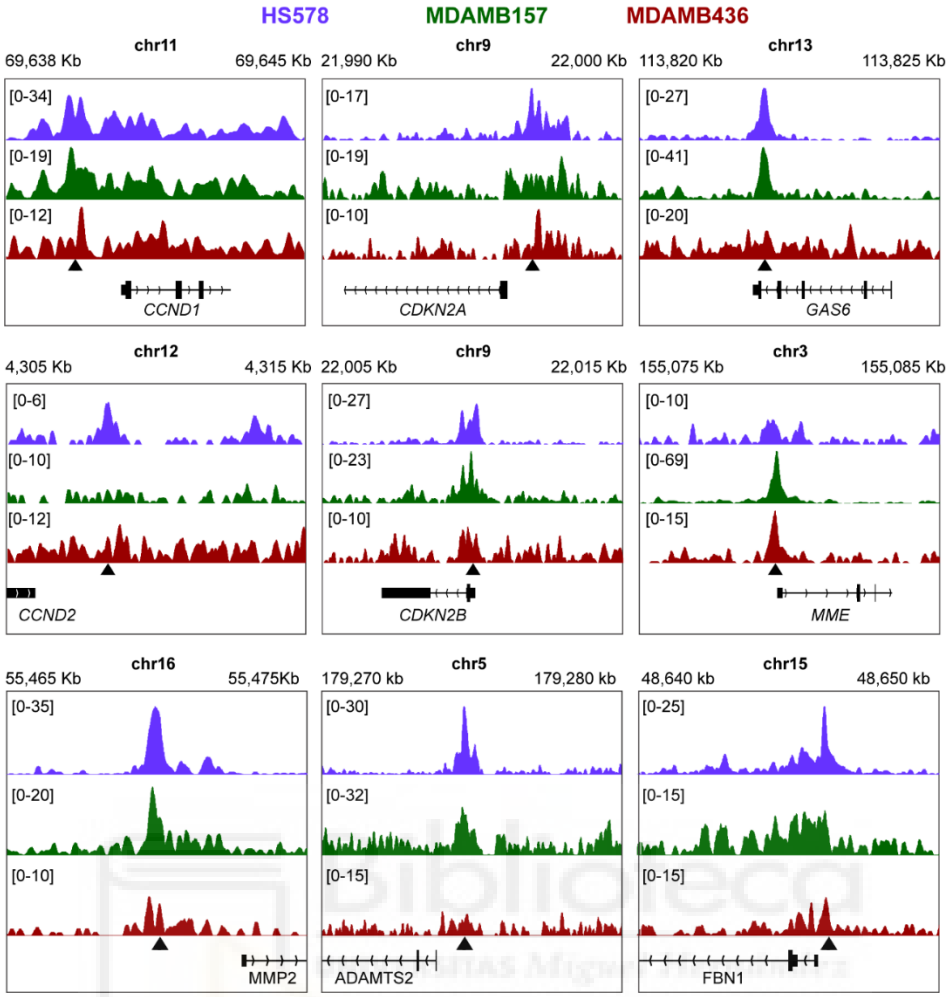


Figure 36. PRRX1 directly binds promoters of cell cycle regulators, dormancy and invasion genes. Chip-seq peaks representing PRRX1 binding at the genomic loci of cell cycle regulators, dormancy and invasion related genes in 3 human breast cancer cell lines: HS578 (blue), MDAMB157 (green), and MDAMB436 (red). Data from Ref [284].

To mechanistically model graded Prrx1 expression in an epithelial setting, we used MDCK-NBL2 cells. They bear strong phenotypic plasticity, undergo a well-characterized EMT response to TGF-β[20,23,42,285,286], and reproduce the sequential activation of EMT transcription factors seen in neural crest cells and in the PyMT model, including the late induction of Prrx1 expression[42]. Importantly, MDCK cells are *bona fide* epithelial cells, allowing to mimic *PRRX1* induction during EMT and to assess how increasing Prrx1 levels influence cell-state transitions. To explore the link between Prrx1 levels, invasion, and proliferation, we generated MCDK-NBL2 clones expressing low (CTR), intermediate, or high Prrx1 levels (Figure 37a), with a ~2-fold difference between intermediate and high, comparable to the existing allelic differences in our mouse model. Both intermediate and high Prrx1 clones were equally and significantly more invasive than CTR cells (Figure 37b). In contrast, growth assays showed that Prrx1-high clones had

markedly reduced proliferation, while intermediate clones grew significantly more than high clone (Figure 37c). Gene expression analysis mirrored this functional property: Prrx1-high clones exhibited lower *CCND2* and higher *CDKN2B/C* levels, whereas intermediate clones were closer to CTR cells (Figure 37d). Dormancy-associated genes (*GAS6*, *MME*) were strongly upregulated in Prrx1-high clones (~100-fold) but downregulated in intermediate clones (Figure 37d). In summary, high Prrx1 expression promotes invasion but suppresses proliferation and induces dormancy, while intermediate levels enhance invasion without decreasing proliferation, therefore maximizing metastatic potential. These findings recapitulate our results in tumors and support a dose-dependent, hormetic model of Prrx1 function regarding metastatic potential.

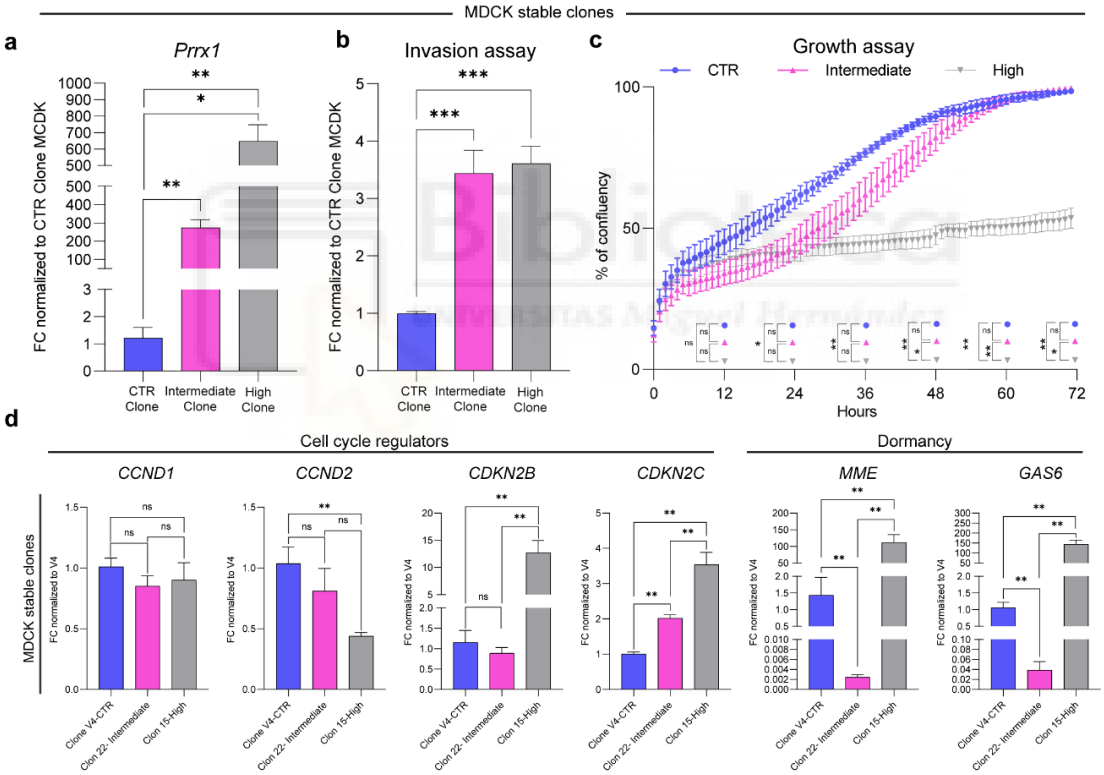


Figure 37. **Engineered cell models reveal dose-dependent Prrx1 regulation of invasion, proliferation and dormancy. a,** RT-qPCR showing the relative *Prrx1* transcript levels in different MCDK-NBL2 clones. FC is represented as mean ± SEM (n=5). Mann-Whitney test. ∗ p <0.05; ∗∗ p < 0.01; p > 0.05, not statistically significant (ns). **b,** Quantification of invasive properties in Transwell assays in the *Prrx1* expressing clones**.** FC is represented as mean ± SEM (n=3). Unpaired two-tailed t test. ∗∗∗ p < 0.001. **c,** Growth analysis of the different *Prrx1* expressing clones. Each timepoint is represented as mean ± SEM (n=3, 3 technical replicates each). Two-way ANOVA test. ∗ p <0.05; ∗∗ p < 0.01; p > 0.05, not statistically significant (ns). **d,** RT-PCR analysis of transcripts for cell cycle regulators (left) and dormancy-related genes (right) in the different *Prrx1* expressing clones. FC is represented as mean ± SEM (n= 5). Mann-Whitney test. ∗∗ p < 0.01; p > 0.05, not statistically significant (ns).

15. Patient stratification by combining invasion and proliferation signatures reveals a high-risk breast cancer subgroup with intermediate *PRRX1* levels

We next wanted to assess whether the relationship between *PRRX1* levels and invasion, proliferation, and dormancy observed in the mouse tumors and human cell lines was also observed in human breast cancer. We analyzed the human breast cancer METABRIC database[248,249] and found that *PRRX1* expression positively correlates with signatures associated with invasion (PINGs) and dormancy, and negatively correlated with proliferation (Figure 38a). Notably, this relationship extends to genes identified as putative PRRX1 targets in our previous analyses, including *MMP2, ADAMTS2, FBN1, COL1A1, CCND1, CDKN2B/C, CFH, GAS6, OGN, MME* (Figure 38b).

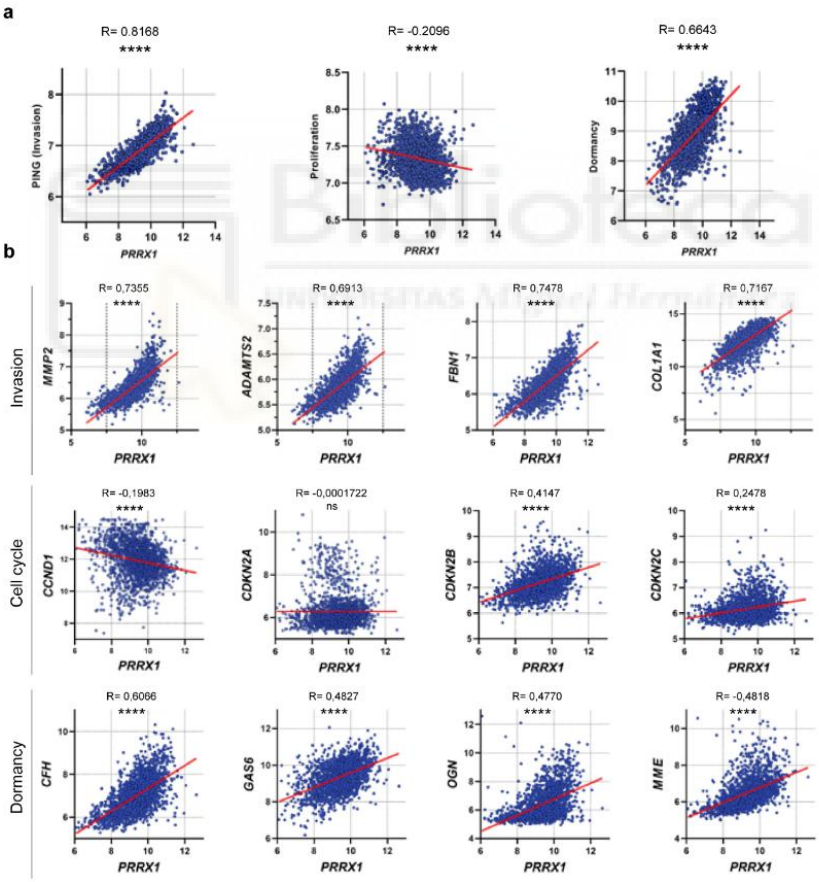


Figure 38. Correlation between *PRRX1* expression and that of cell-cycle and dormancy-related genes in human breast cancer. a, Correlation analysis between *PRRX1* expression and PING, Proliferation and Dormancy signatures on METABRIC human breast cancer data (n=1980). The red line represents the regression line. Two-tale Pearson correlation test. **** $p < 0.0001$. **b,** Correlation analysis between the expression of *PRRX1* and genes associated with invasion, cell cycle regulation and dormancy. Selected from METABRIC breast cancer human data (n=1980). Red line represents regression line. Two-tale Pearson correlation test. **** $p < 0.0001$; $p > 0.05$, not statistically significant (ns).

To evaluate whether the interplay between invasion and proliferation as a determining factor of metastatic competence could have clinical relevance, we used the METABRIC database and grouped patients using a combinatorial approach. To fully integrate invasion and proliferation into the grouping strategy, we used a combination of the gene signatures employed throughout the project as proxies for invasion (PING) and proliferation (Cell_Cycle_Mitotic_R-HSA-69278), both of which have been found to be dependent on *Prrx1* levels in our mouse model. We categorized the patients into four groups based on quartiles: (i) Invasion High_Proliferation High, (ii) Invasion High_ Proliferation Low, (iii) Invasion Low_Proliferation Low, and (iv) Invasion Low_Proliferation High, and extracted their survival data (Figure 39a, see Methods).

Interestingly, the patient group with the best prognosis was the Invasion High_ Proliferation Low, closely followed by the Invasion Low_Proliferation Low. This was particularly intriguing, as it suggests that the impact of high invasion on survival was contingent on proliferation, with almost no difference between invasive and non-invasive tumors when they were non-proliferative (Figure 39a). More importantly, this classification revealed that patients with tumors scoring high for both invasion and proliferation exhibited significantly worse overall survival (OS), particularly during the first months after diagnosis (Figure 39a). These results further supported our proposal that metastatic competence arises from the coexistence of both programs, invasion and proliferation. Moreover, these findings aligned with our mouse models with different *Prrx1* levels, where the condition combining invasion with high proliferation, as in the *Prrx1* +/f invasive cells, presented a higher metastatic burden compared to conditions where invasion was compromised, as in *Prrx1* f/f cells, or where there was invasion but low proliferation, as in *Prrx1* +/+ invasive cells.

Given the established link between invasion and proliferation, we next investigated whether our patient stratification reflected differences in dormancy scores. As expected, patients from the group with the poorest survival, bearing highly invasive and highly proliferative tumors, showed a markedly lower dormancy score (Figure 39b). Interestingly, this group of aggressive tumors also exhibited lower *PRRX1* expression levels than the invasive, low-proliferative group (Figure 39c). Both results were compatible with the phenotype observed in *Prrx1* +/f mouse tumors compared to the *Prrx1* +/+controls, as they presented lower dormancy and higher proliferation scores. Furthermore, these findings aligned with the non-linear metastatic burden observed in TMAs from patients with varying PRRX1 expression levels.

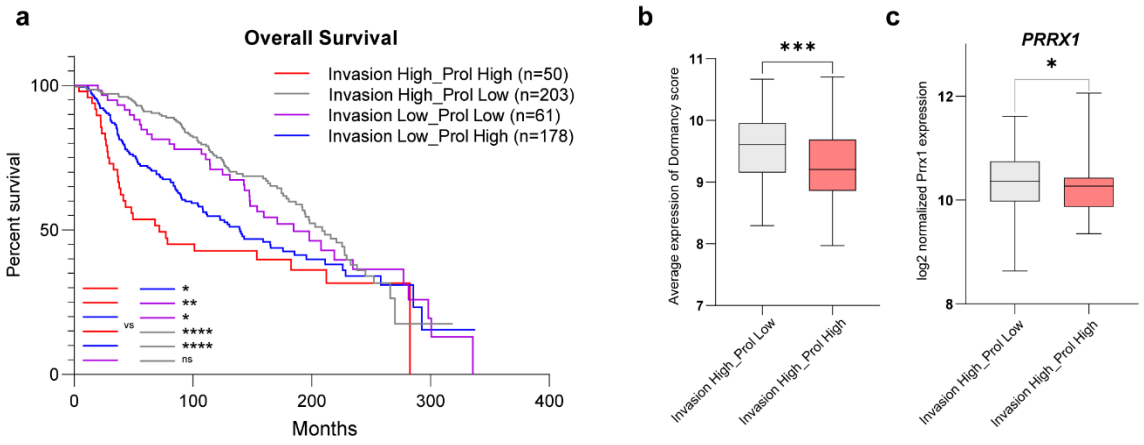


Figure 39. Patient stratification by combining invasion and proliferation signatures reveals a high-risk breast cancer patients' subgroup with intermediate *Prrx1* levels. a, Kaplan-Meier overall survival (OS) analysis of METABRIC breast cancer patients stratified by high and low expression of invasion (PINGs) and proliferation (Cell_Cycle_Mitotic_R-HSA-69278) signatures (n=492) (see Methods). Gehan-Breslow-Wilcoxon test. * p <0.05; ** p < 0.01; **** p < 0.0001; p > 0.05, not statistically significant (ns). See Supplementary Table 4 (Materials and methods) for additional statistical analysis **b,** Dormancy score of the groups defined in (a). Boxplots show median, upper, and lower quartiles; whiskers show upper and lower extremes. Mann Whitney test. *** p < 0.001; **** p < 0.0001; p > 0.05, not statistically significant (ns). **c,** Comparison of *PRRX1* levels in the PINGs High groups described in (a). Boxplots show median, upper and lower quartiles; whiskers show upper and lower extremes. Mann Whitney test. * p <0.05; *** p < 0.001; **** p < 0.0001; p > 0.05, not statistically significant (ns).

Importantly, even when stratified by treatment, a very similar pattern emerged across all conditions (Figure 40). Taken together, these findings indicate that human breast cancer patients with tumors bearing intermediate Prrx1 levels in cancer cells exhibited a unique combination of invasive and proliferative traits, defining a particularly aggressive subset of tumors. Furthermore, our stratification approach could serve as a valuable tool to identify high-risk patients who may benefit from therapeutic strategies targeting both invasion and proliferation.

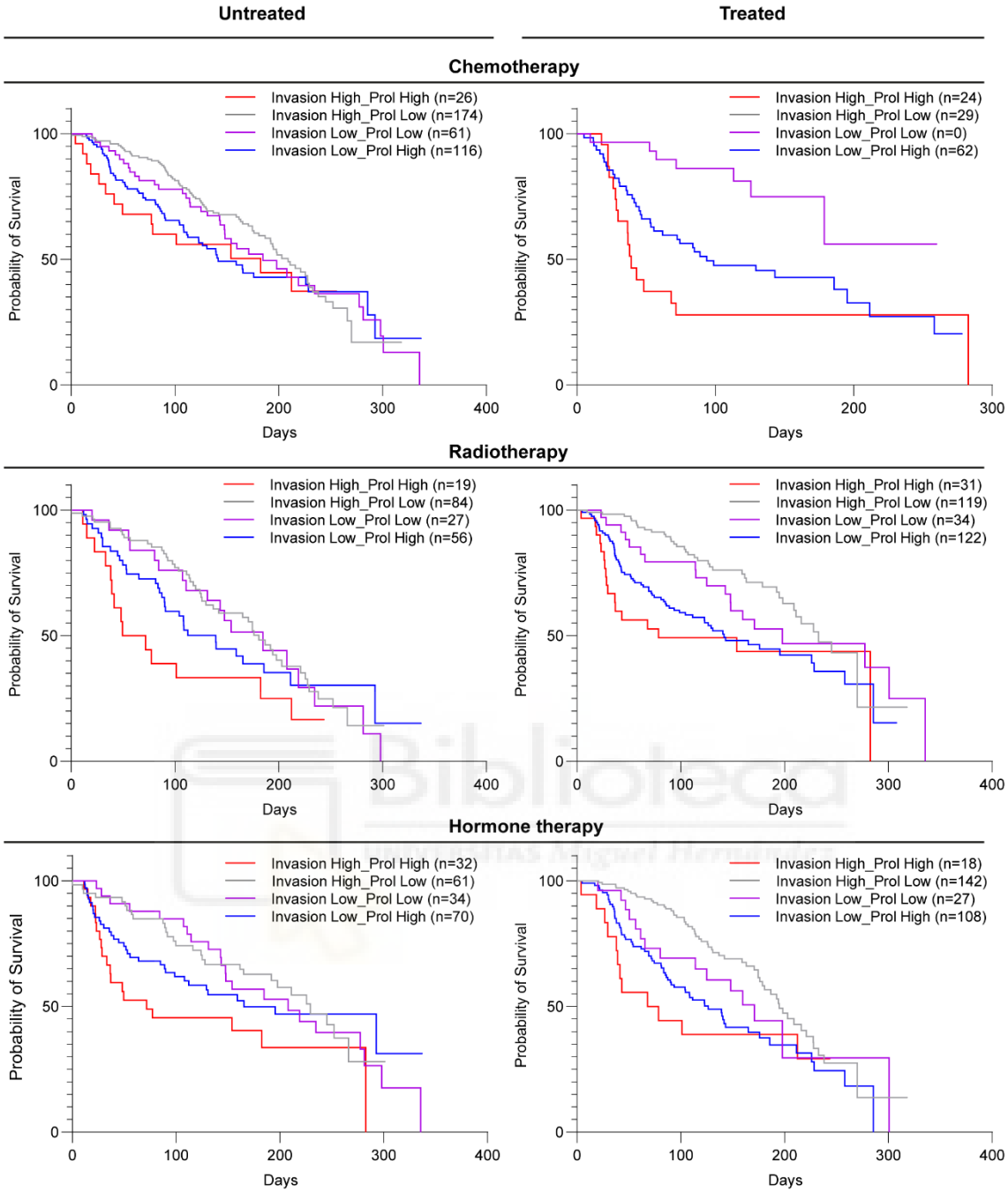


Figure 40. Signature-based breast cancer patients' stratification grouped by therapy. Kaplan-Meier overall survival (OS) analysis of METABRIC breast cancer patients stratified by high and low expression of invasion (PINGs) and proliferation (Cell_Cycle_Mitotic_R-HSA-69278) signatures and grouped by the treatment received.

General Discussion

Understanding and effectively combating metastasis remain among the greatest clinical challenges in oncology, as it is the leading cause of cancer-related mortality and lacks specific therapeutic options. Moreover, metastatic cells often resist treatments developed for early-stage tumors. Throughout the metastatic cascade, cancer cells must invade, intravasate, survive transit, extravasate, and ultimately resume growth. This complex and multi-step journey makes identifying key metastatic determinants particularly challenging.

In this study, we integrate human breast cancer data, genetic mouse models, and multi-omics analyses to identify a specific cell state in the primary tumor that predicts high metastatic outgrowth, determined by Prrx1 levels. Our findings provide insight into how the epithelial-to-mesenchymal (EMT) status of invasive cells influences their metastatic potential.

Previously, we had demonstrated that EMT drives cell plasticity in breast cancer, adding another layer of tumor heterogeneity by engaging cells in two distinct EMT trajectories: an embryonic-like EMT associated with invasion, and an adult-like EMT linked to inflammation and immune regulation[42] (see also Figure 5). In this thesis, we dissect the role of Prrx1 as a master regulator of invasive cells, controlling not only invasion but also cell cycle progression and dormancy. Intermediate Prrx1 expression levels are sufficient to confer invasive properties, comparable to those observed at high expression levels. However, because high Prrx1 levels also activate the expression of the CDKN2 family of cell cycle inhibitors and of dormancy-related genes, increased invasion and dissemination do not necessarily translate into metastatic outgrowth. This dual regulation creates a non-linear correlation between Prrx1 levels and metastatic burden, as the latter is maximized at intermediate Prrx1 levels (Figure 41, page 123).

In this discussion, I will contextualize our main findings within the broader cancer research landscape, highlighting their contribution to our understanding of metastatic biology and potential clinical implications.

1. Non-Linear Dose-Response in Cancer Progression: the importance of considering levels

One of the most notable findings of this work is the identification of a non-linear correlation between PRRX1 levels and the incidence of metastasis in breast cancer, observed in both patients and mouse models. This biphasic dose–response, known as

hormesis, has been primarily studied in the fields of toxicology and cellular stress[262,287]. Several proteins, including the transcription factors Nrf-2 and NF-κB, have been identified as exhibiting hormetic behavior in response to cellular stress[288]. Interestingly, a recent study described how different Nrf2 levels can have opposing effects on lung cancer initiation and progression, may extend the concept of hormesis to cancer[289]. However, in that case, the underlying mechanisms driving this dual role were not identified. In our study, we propose that Prrx1 hormetic behavior arises from the simultaneous regulation of cellular programs with opposing roles, invasion, and dormancy.

Several examples in the literature illustrate how cells can respond to different levels of a signal or factor by activating distinct programs both in embryonic development and in cancer. For instance, Sox2 levels dictate the developmental switch between stemness and differentiation[290], and varying levels of Twist regulate stemness versus cancer cell invasiveness[291]. However, these cases do not fully align with hormetic behavior. In the case of Sox2, the response is better described as a binary switch between two distinct states rather than a non-linear dose-response. Nevertheless, the underlying mechanism is intriguing, as Sox2 binds different co-factors when it is expressed at different levels, leading to the activation of distinct gene sets.

For Twist, the data align more closely with a requirement for high expression levels to fully execute its biological functions, enabling both tumor progression and invasion, whereas lower levels are only sufficient for tumor initiation. Notably, in that study, different Twist levels were modeled using a genetic approach based on allelic variation, like the strategy we have employed in this project. Interestingly, our lab also demonstrated that Prrx1 downregulation is required for metastatic outgrowth, and a similar role for Twist was described in a mouse model of squamous cell carcinomas[52]. More recently, increasing attention has been given to the relationship between RAS levels and tumor progression. Intermediate RAS levels appear to provide an optimal balance for tumor growth, while high RAS expression can induce senescence, impairing tumor progression[292]. Ongoing efforts are focused on dissecting the molecular mechanisms underlying this continuum of phenotypes.

Altogether, these findings have important clinical implications. First, the existence of distinct biological states at different protein levels, whether through non-linear responses or binary switches, complicates therapeutic strategies. Given that completely abolishing the expression of a protein is virtually impossible, targeting proteins that exhibit hormetic behavior could not only lead to suboptimal responses but also to paradoxical effects.

This highlights the need for a more nuanced approach when selecting therapeutic targets and designing treatment strategies.

Additionally, these findings have important implications for biomarker discovery and patient stratification. As we have shown, patients with intermediate Prrx1 levels have the worst prognosis, highlighting the need to consider expression levels when correlating gene expression with clinical variables. A clear example of the clinical impact of accounting for intermediate levels is the case of the HER2-low breast cancer subtype. Traditionally, breast cancer was classified as either HER2-positive or HER2-negative. However, it is now known that nearly 55% of breast cancer patients fall into the HER2-low category, including some previously classified as triple-negative breast cancer (TNBC). Until recently, only HER2-positive patients received HER2-targeted therapy, such as trastuzumab deruxtecan (Enhertu®). However, the DESTINY-Breast04 clinical trial demonstrated that in HER2-low patients who had already undergone multiple lines of chemotherapy, treatment with trastuzumab deruxtecan reduced the risk of disease progression by approximately 50% and the risk of death by 36% compared to chemotherapy alone, fundamentally changing breast cancer treatment[293]. Our data indicates that intermediate PRRX1 expression could be a marker of metastasis in breast cancer independent of treatment modality. Nevertheless, large, prospective, multicenter studies specifically designed to assess treatment response and long-term survival are needed to assess whether PRRX1 expression levels can be a clinical decision-making criterium or a predictive biomarker.

2. Cell plasticity and Metastatic Progression: Reversible States and Therapeutic Implications

Initial observations of metastasis biology revealed that metastatic cells often exhibit an epithelial phenotype like, or even more pronounced than, that of the primary tumor. This was evident in breast cancer studies using E-cadherin staining[294]. These findings led to skepticism about the role of EMT in metastasis, as no clear signs of EMT were detected in metastatic lesions. However, in 2012, two seminal studies demonstrated the necessity of EMT reversion at metastatic sites, explaining why fully developed metastases appear epithelial[51,52].

In this work, by examining EMT heterogeneity throughout the metastatic cascade, we reveal that cell plasticity not only involves shifts between epithelial and mesenchymal markers but also encompasses the hierarchical organization of distinct EMT programs[42]

(see also Figure 5). Through immunophenotyping of circulating tumor cells (CTCs) in the PyMT model, we observed an almost 40-fold enrichment of cells exhibiting an embryonic-like invasive EMT program compared to the primary tumor, where these cells constitute only ~2% of the total cancer cells population. This strong selection of the invasive EMT program supports the necessity of EMT activation for invasion. Notably, our findings align with previous reports of K14 enrichment in CTCs[137].

More strikingly, we found no evidence of cancer cells expressing Jun, a broad marker of the adult-like inflammatory EMT trajectory. This confirms that this type of partial EMT lacks invasive capacity, similar to what has been described in adult tissue fibrosis[48,49]. However, when analyzing these two EMT programs in metastases, we found their distribution to be remarkably like that of the primary tumor, yet significantly different from that of CTCs, which are considered the seeds of metastases. Until now, it was known that the same tumor cell lineage could generate populations belonging to different EMT trajectories, as seen in the PyMT model, where cells of luminal origin can give rise to both EMT programs[42]. Our data further indicate that even cells already engaged in a particular EMT trajectory, at least in the case of invasive EMT cells, retain plasticity and can revert to recapitulate the EMT distribution observed in the primary tumor.

This plasticity poses a major challenge for therapeutic strategies aimed at targeting specific cell populations, as eliminated cells could be replenished once treatment ceases. However, the same plasticity could be exploited to induce redifferentiation into drug-sensitive states, thereby facilitating tumor regression. For example, Schade et al. demonstrated that inhibiting the AKT and EZH2 pathways can drive triple-negative breast cancer cells toward a differentiated luminal-like phenotype, achieving a synergistic therapeutic effect[295]. Similarly, highly plastic EMT cells could be transdifferentiated into more indolent cell types, such as adipocytes[296] or even into dendritic cells with anti-tumor properties[297].

In our case, the therapeutic implications of modulating EMT trajectories are particularly relevant. We previously showed that Prrx1 loss in the primary tumor reduces the proportion of cells engaging in the invasive EMT trajectory while increasing those in the inflammatory one. This shift correlates with immune landscape reprogramming, characterized by enhanced infiltration of anti-tumoral macrophages[42]. Given the plasticity of EMT states and the persistence of EMT inflammatory-like cells in metastases, a potential strategy would be to identify factors that expand this population, thereby fostering a tumor microenvironment more conducive to immunotherapy, not only in primary tumors but also in metastatic settings.

Another critical implication of EMT recapitulation in metastases is the potential for metastatic self-seeding, where metastases can reseed the primary tumor from which they originated[13] or even generate secondary metastases from pre-existing metastatic lesions. This phenomenon has been observed in clinical settings through multi-regional sequencing of paired primary tumors and metastases[298,299]. The clinical significance of this is substantial: even if the primary tumor is eradicated, metastases remain a dual threat, not only do they sustain disease progression, but they can also seed new metastatic outgrowths. Interestingly, such sequential EMT-MET cycles are integral to the formation of multiple embryonic structures, including the valves of the vertebrate heart, suggesting that cancer cells do not merely hijack developmental programs like EMT but also their underlying dynamics.

3. Leveraging Omics Technologies to Dissect Tumor Heterogeneity and Metastatic Potential

Advancements in technology are increasing the granularity of biological analysis, enabling the identification of rare cell types, specific biomarkers, and multimodal data integration. Currently, scientists are exploring ways to transition from descriptive approaches to uncovering mechanisms that drive complex biological processes.

In this study, we first employed spatial transcriptomics to profile the gene expression of cells located in specific tumor regions, particularly the invasive edges. However, image-based spatial transcriptomics currently has limitations in the number of genes that can be profiled. To overcome this, we leveraged single-cell RNA sequencing (scRNA-seq) to identify cell states that resemble those found at invasive edges. This experimental pipeline compensates for the limitations of both techniques: the low gene depth in spatial transcriptomics and the lack of spatial context in scRNA-seq. We believe this approach could be widely applied in contexts where profiling cells within specific tissue structures is of interest.

Similar strategies have been used to dissect cellular contributions in various diseases. For instance, spatial transcriptomics analysis of glioblastoma post-treatment identified perivascular-derived fibroblast-like cells as key drivers of scar formation and therapeutic resistance[300]. The spatial information was further refined using scRNA-seq to study cell-cell interactions. Another common approach involves analyzing transcriptomic changes in relation to cell proximity. For example, in colorectal cancer, macrophages near T cells show increased expression of antigen-presenting genes, which has significant clinical

implications for immunotherapy response[301]. Recently, a compilation of studies published in *Nature* journals (30th October 2024) have explored multiomics analyses in cancer, one of which examined EMT heterogeneity using spatial transcriptomics[302]. This study highlighted distinct cell cycle gene expression patterns in EMT-high versus EMT-low cells, findings that align with our observation of differential cyclin expression in invasive cells compared to the rest of the tumor.

Importantly, spatial analysis of human samples reveals critical correlations between spatial features and clinical variables, particularly in therapy response. For example, in breast cancer, the spatial distribution of T cells relative to cancer cells is a stronger predictor of immunotherapy response than the total T cell count[303]. Further analysis of our spatial transcriptomics data, focusing on cell composition and interactions at the invasive front, could provide valuable insights into the biology of invasive cells and the origins of Prrx1 heterogeneity.

Our study also highlights the power of chromatin dynamics analysis in identifying key regulators of specific cell types and states. Transcription factors are often expressed at low levels, making them difficult to detect with scRNA-seq alone. However, epigenetic data can reveal their activity within specific subpopulations even when their expression is below detection thresholds. In our case, this approach was instrumental in confirming that Prrx1 is not only a marker of invasive cells but also a master regulator of their state, with direct access to promoters of key genes involved not only in invasion, but also in cell cycle progression and dormancy. Similar methodologies have been used to uncover key regulators of metastatic colonization in the liver, identifying Klf4 as a master regulator in liver metastases through snATAC-seq[207]. Additionally, chromatin analysis has improved the characterization of cell states during pancreatic cancer progression by predicting their plasticity potential based on chromatin accessibility[304].

In addition, we show that different EMT programs also correlate with distinct epigenetic states. Notably, the transition to the later stages of the invasive trajectory appears to require the most dramatic chromatin changes. Moreover, our previous analysis[42] (see also Figure 5). did not fully resolve the sequence of cell states in T2, the inflammatory trajectory. However, chromatin data now suggest a clearer progression along the trajectory (from cluster 13 to 15 to 1). Additionally, the availability of chromatin data enables the construction of detailed gene regulatory networks, facilitating the use of computational tools such as *in silico* perturbation with CellOracle[283]. This type of analysis would allow the virtual screening of hundreds of transcription factors to identify master regulators of key transitions during PyMT progression, particularly at the bifurcation point

where dedifferentiated cells choose between the two EMT trajectories. Identifying candidates which modulation could alter tumor progression represents an exciting avenue for further research.

4. The Dual Role of Prrx1 in Metastatic Fitness: Balancing Invasion, proliferation, and dormancy

Our data provides a clearer understanding of the dichotomous role of Prrx1 in metastasis, which has been described as both essential and antagonistic to the process. This paradox arises from the dual regulatory role of Prrx1, as it induces invasion while simultaneously inhibiting proliferation.

Regarding invasion, it is evident that without invasive properties, no matter how proliferative or stem-like a cancer cell is, it will never establish new tumor lesions in distant tissues. Some degree of EMT, the primary driver of invasive behavior, is therefore required for metastasis. However, the extent of EMT necessary for sufficient invasion remains unclear. Not all EMT programs confer invasive properties; for instance, EMT activation during injury response in adults does not endow cells with migratory abilities, as demonstrated in studies on kidney fibrosis[48,49]. Interestingly, despite exhibiting EMT markers and a partial EMT state, considered more metastatic in cancer, these cells retain their original location and do not migrate. Notably, Prrx1 expression is absent in this context[42,48], which may explain the lack of invasion. Compatible with this,, we have shown that even early stages of EMT induction, characterized by low Prrx1 levels, already enhance invasive properties in epithelial cell lines[42]. This aligns with our findings that intermediate Prrx1 levels are sufficient to confer the minimal invasive capacity required for metastasis. Although we observed a positive linear correlation between Prrx1 levels and invasive signatures, literature reports indicate that invasive partial EMT cells exhibit greater disseminative capacity than fully mesenchymal cells in both *in vitro* assays and intravital imaging, suggesting that excessive mesenchymalization may even hinder invasion[176]. Indeed, unpublished data from our laboratory shows that clones with extremely high levels of Prrx1 (forced overexpression) show a marked downregulation of their migratory properties.

In terms of proliferation, we and others have shown that EMT-TFs such as Snail1 and Zeb1 suppress cell cycle activators including Ccnd1 and Ccnd2, thereby inhibiting proliferation[79,179]. Along with the downregulation of tumor-initiating abilities[51,52], this has been proposed as the reason for the metastasis-suppressive effect of high EMT levels.

While the inverse correlation between EMT and cell cycle progression remains valid, our findings reveal that intermediate Prrx1 levels (and thus intermediate EMT activation) can simultaneously sustain both invasion and proliferation. This provides mechanistic insight into why hybrid E/M cells exhibit high metastatic potential *in vivo*[99,100,171,176].

Importantly, our data offer a new perspective on recent studies linking metastasis with proliferation in the primary tumor. Single-cell lineage reconstructions in mouse models suggest that metastatic clones predominantly arise from hybrid EMT states with high proliferation scores[177]. In addition, our analysis showed that subclones with intermediate Prrx1 levels are the most metastatic, further linking intermediate *Prrx1* expression and sustained proliferative capacity. Similarly, a new scDNA-seq-based tool for characterizing cell cycle stages in human samples through multiregion sequencing has revealed that metastatic clones originate from highly proliferative clones in the primary tumor[178,305]. The authors themselves acknowledged that this finding was unexpected, given that disseminating cells undergo EMT, which is typically associated with an invasive yet less proliferative phenotype. They proposed that proliferative clones undergo EMT and subsequently revert to a proliferative state in distant organs. However, this explanation does not fully resolve the issue, as EMT should “presumably” suppress proliferation, making the identification of highly proliferative metastatic clones in the primary tumor paradoxical. Instead, we propose that their results are better explained by the presence of rare clones that are both highly proliferative and highly invasive, serving as the main source of metastases.

Additionally, recent studies on circulating tumor cell (CTC) dynamics have shown that CTCs released during the rest phase (night) are more metastatic than those released during the active phase (day)[153]. The authors attributed this difference to the proliferative status of the CTCs, as those released at night show upregulation of mitotic genes. These findings further challenge the idea that invasive cells are non-proliferative and support our proposal that the proliferative state of invasive cells in the primary tumor prior to dissemination influences their metastatic potential. This circadian effect also raises interesting questions. Is the difference in CTC proliferation due to a direct circadian regulation of the CTCs themselves, or does it reflect a broader circadian control of proliferation in the primary tumor? Do Prrx1 levels fluctuate in a circadian manner, potentially contributing to Prrx1 heterogeneity? Are the CTCs released during the active phase in Prrx1 intermediate tumors more similar to the rest-phase CTCs than those from Prrx1 high tumors? Addressing these questions could provide a better understanding of CTC biology and improve its interpretation in clinical setting.

5. Heterogeneity at the Invasive Front: Linking EMT States to Metastatic Potential and Dormancy

Most studies rely on experimental metastasis models using cell lines, which cannot reproduce the intratumoral phenotypic heterogeneity observed *in vivo* and its impact on different steps of the metastatic cascade. Here, we show that heterogeneity at the invasive edge of the primary tumor, driven by different Prrx1 levels, is sufficient to influence metastatic behavior. This suggests that cancer cells leave the primary tumor with intrinsic information that primes their metastatic potential. Our data in breast tumors and analysis of lineage tracing data from pancreatic cancer support a model in which invasive cells exist in distinct states: some (Prrx1 intermediate) maintain proliferation while invading and disseminating, enabling rapid metastatic outgrowth, while others (Prrx1 high), though being invasive, enter dormancy and rely more on additional signals from the metastatic niche to attenuate their EMT (mesenchymal) state and revert to a more epithelial phenotype to resume growth. This heterogeneity may underlie the undefined chromatin states proposed to account for differential sensitivity to the pro-dormancy signal Wnt upon arrival at the metastatic site[197], may explain the origin of the differential metastatic potential between early and late disseminated cancer cells [201] or even determine the probability to enter or not dormancy[131]. Our results could also help to understand recent dormancy-tracing data showing that, under naïve conditions, most metastases originate from cells that never entered dormancy[131]. This suggests the existence of a specific population within the primary tumor that is primed for direct metastasis, our hypothesis being that Prrx1-intermediate cells represent this population. Overall, our data provide a new framework to understand why some cancer cells enter dormancy while others rapidly resume growth at the same metastatic microenvironment.

Although biological processes do not always align with "why" questions, it is difficult not to wonder why tumors generate cancer cells with low metastatic ability if they are also capable of generating cells already primed for metastasis. One possible explanation is probabilistic: based on our model, invasion, proliferation, and dormancy are simultaneously regulated by the same transcription factor, Prrx1 in this case. Different upstream mechanisms likely regulate Prrx1 levels, creating a continuum of cells with varying metastatic potential. In other words, since increasing invasiveness comes at the cost of reduced proliferation, tumors can only consistently generate highly metastatic cells in scenarios where Prrx1 levels are exquisitely regulated, as observed in embryonic development[290,306], assumption unlikely to be true considering the chaotic nature of tumors.

Another, non-exclusive explanation is that despite their lower metastatic potential, high Prrx1 levels may confer certain selective advantages. For example, quiescent cancer cells, which have some commonalities with dormant cells, exhibit increased resistance to immune killing, primarily by promoting an immunosuppressive niche[307]. We have also discussed in the introduction the link between EMT activation and immunosuppression, both through the upregulation of immune checkpoint inhibitors like PD-L1[308] and the secretion of immunosuppressive molecules[112]. Notably, even a minor presence of EMT cells is sufficient to establish this immunosuppressive microenvironment, thereby protecting non-EMT cancer cells from immune killing[112]. It remains unclear where along the EMT spectrum these immune-related properties are acquired. If high EMT activation is required, the presence of a few highly mesenchymal cancer cells in the metastatic niche, despite their intrinsically low metastatic potential, could enhance the success rate of more primed cancer cells by shielding them from immune attack.

The existence of these heterogeneity has important clinical implications because dormant cells are more resistant to chemotherapy than proliferative ones, and form a pool of cells that, while not highly metastatic under naïve conditions, can survive treatments that eliminate highly metastatic cells[309–312]. Data on the differential contributions of EMT states support this idea: in untreated mice, metastases predominantly originate from cells activating a partial invasive EMT program, with very few deriving from fully EMT-activated cells. However, after treatment, this balance shifts, and metastases are largely composed of cells that have undergone an advanced EMT state[131,176,180,181]. Investigating whether, despite the higher metastatic burden of *Prrx1* +/f tumors, their metastatic potential is more susceptible to chemotherapy than *Prrx1* +/+ tumors due to their reduced capacity to enter dormancy could be an interesting future research direction with important clinical implications.

6. Combinatorial Strategies and Clinical Implications in Metastasis and Dormancy

Patient stratification involves categorizing patients into subgroups based on specific disease characteristics, reflecting the underlying pathology to personalize treatment strategies and improve outcomes. Traditionally, patient stratification based on gene signatures is done by comparing high-expression to low-expression groups. However, since metastasis is the leading cause of cancer-related deaths, it significantly impacts survival data. Our findings show that metastasis is influenced by various, even

anticorrelated, biological properties, suggesting that traditional approaches may not be effective for prognostic purposes.

We propose that the coexistence of invasion and proliferation is crucial for effective metastatic colonization, and we explore the potential clinical benefits of combinatorial stratification strategies. Using this approach, we identified a subgroup of breast cancer patients with an extremely poor prognosis, particularly in the first few months after diagnosis. Interestingly, a high invasion score in a low proliferative context does not correlate with worse survival compared to when both invasion and proliferation are low. However, invasion clearly has a negative impact in patients when proliferation is high, highlighting how considering multiple variables together can change clinical interpretation and prevent masking the potential clinical value of signatures. If only invasion scores were used, they would likely represent intermediate values.

Our results also shed light on the inverse correlation between proliferation scores and clinical outcome across breast, lung, and prostate cancers[313–315]. This suggests that proliferation scores could inform not only about primary tumor dynamics but also the metastatic potential of cancer cells. The invasive signatures developed from breast cancer cell lines may not be directly transferable to other tumor types, but this opens the possibility of generating similar signatures for other cancers. Our results highlight the potential of multi-omics strategies, with spatial transcriptomics offering a clinically tractable platform due to its compatibility with FFPE specimens, the standard method for pathological archives, and its conceptual alignment with histopathological workflows. Defining improved signatures derived from invasive areas in human samples across cancer types, may enable an optimized patient stratification through combinatorial molecular approaches.

As mentioned in the introduction, circulating tumor cells (CTCs) are increasingly used for patient stratification in cancers like breast cancer. However, beyond enumeration, it remains unclear if further characterization of their cell state offers clinical benefits. We propose that metastatic competence is determined in the primary tumor based on the balance between invasion and proliferation, and other studies have shown that CTCs with higher cycling activity have increased metastatic ability. We suggest that characterizing the cycling status of CTCs in patients could provide valuable insights, with a higher ratio of proliferating CTCs indicating higher metastatic risk. Recent data showing that patients with CTCs high in MKI67 expression have significantly reduced progression-free survival supports this idea[316].

The linked behavior of invasion, proliferation, and dormancy could guide new therapeutic approaches. Since cancer is often diagnosed after dissemination has started, targeting primary tumor dissemination alone may be insufficient. Finding the best strategies to prevent and target metastatic disease is crucial[317]. Despite some efficacy of drugs used for primary tumors in metastatic settings, large macrometastases are considered virtually incurable. To date, no agents have been approved exclusively for adjuvant use without indication in the macrometastatic setting, although dormancy-focused clinical trials in the adjuvant setting are in progress. The research community is increasingly focusing on approaches targeting micrometastases and dormant cancer cells, which are proposed to be the source of macrometastases[318]. However, clinical targeting of dormant cancer cells presents two major challenges: their intrinsic resistance to many therapies, and their detection. Initially, there were proposals to purposefully awaken dormant cells to make them more vulnerable, but this is considered a risky approach, as expecting that all proliferative cancer cells could be killed it unrealistic. Currently, efforts are focused on maintaining cells in a dormant state or, ideally, eliminating them[195]. Our findings on the relationship between Prrx1 levels and dormancy make it a promising target for strategies aimed at maintaining dormancy and preventing metastatic outgrowth, particularly in patients who have undergone primary tumor resection, as increasing Prrx1 levels could also enhance the invasive capacity of tumor cells.

Our results may also help identify specific vulnerabilities within the dormant cancer cell population. For example, the observed switch to cyclin D2 in the invasive and dormancy-primed population may offer insights. Since frontline treatments for many luminal breast cancers rely on CDK4/CDK6 inhibitors, which are regulated by Cyclin D family members[274], it would be interesting to determine if the efficacy of these treatments depends on the main Cyclin D expressed in the cells. Targeting Cyclin D2 specifically could reduce metastatic proliferation without the toxic systemic effects seen with CDK4/CDK6 inhibitors, potentially enabling its prolonged use as a preventive strategy. In addition, as we propose that EMT and dormancy are tightly linked, anti-EMT therapeutical strategies could be used to target dormant cancer cells, one of the biggest clinical challenges in cancer[195].

In summary, our study reveals that metastatic competence is established at the primary tumor, and that metastases recapitulate primary tumor biology thanks to high cancer cell plasticity. We find that different levels of Prrx1 create cancer cell heterogeneity at the invasive edge of the primary tumor, determining metastatic competence in a non-linear manner by regulating invasion, proliferation, dormancy and tumor initiating capacity (Figure 42). This integrates previous findings about the role of Prrx1 as an EMT-TF

throughout the metastatic cascade, required for invasion but acting as an inhibitor of metastatic colonization. Our data also provide mechanistic insights into the high metastatic potential of the hybrid E/M state and reveal the combination of invasive and proliferative signatures as a promising strategy for patient stratification.

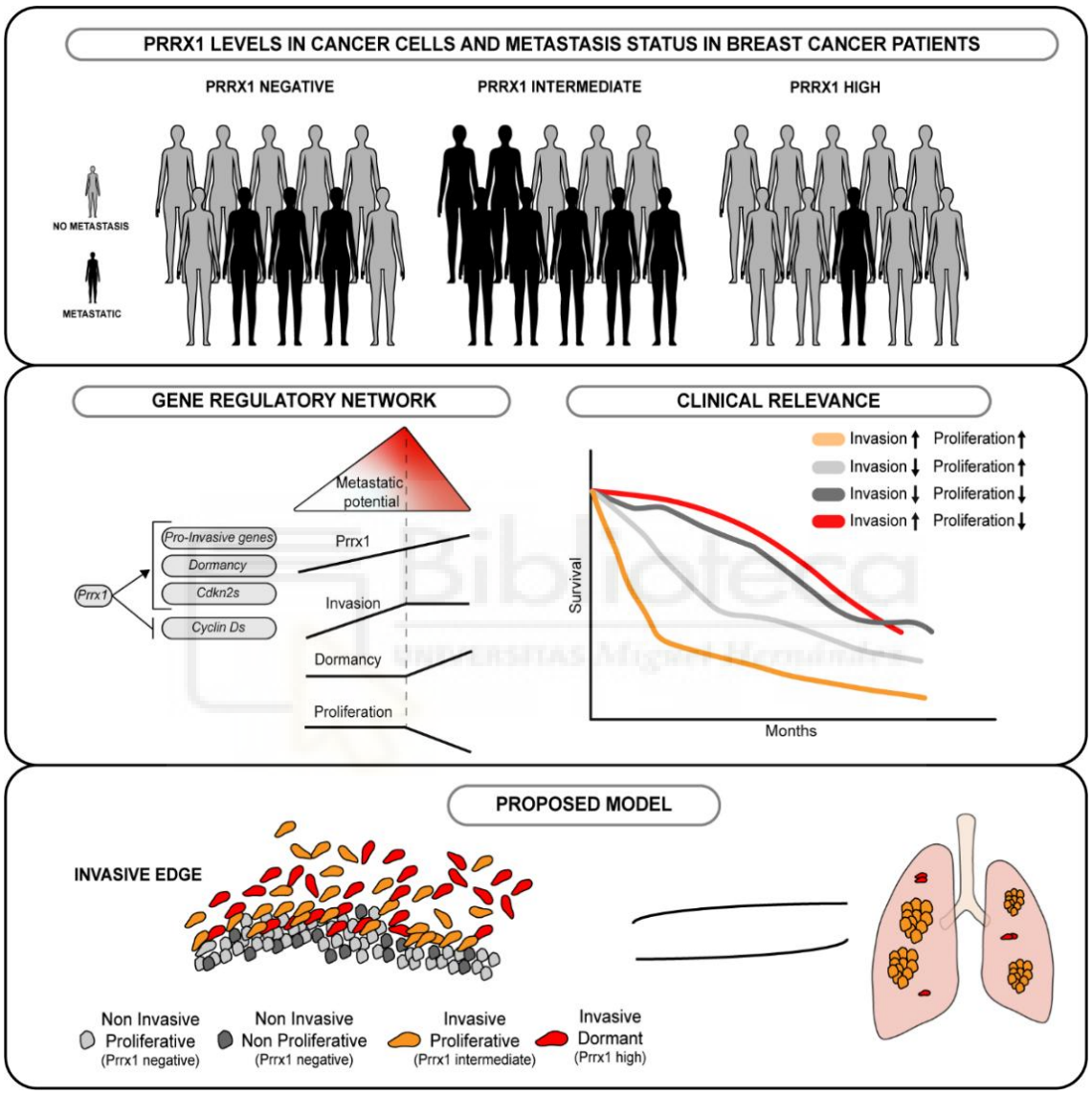


Figure 41. A hormetic transcriptional program coregulates invasion, proliferation and dormancy to define metastatic potential. Patients with breast tumors bearing intermediate levels of Prrx1 have a higher metastatic incidence than those with negative or high levels. Mechanistically, Prrx1 activates a pro-invasive gene program while promoting cell cycle arrest and dormancy through the activation of the cell cycle inhibitors of the CDKN2 family members and the repression of Cyclin Ds. The combination of invasion plus proliferation/dormancy index leads to a hormetic response of Prrx1 levels with respect to metastatic potential. The combination of Prrx1-regulated invasion and proliferation programs allows breast cancer patient stratification, identifying a subgroup with the poorest prognosis. We propose that Prrx1 heterogeneity in the primary tumor defines the metastatic potential based on cell invasion and proliferation capabilities.

Conclusions

1. Breast cancer can be classified based on PRRX1 expression in patients. There is a non-linear relationship between PRRX1 levels and metastatic burden. Tumors with intermediate PRRX1 expression show a higher metastatic incidence than those that are either PRRX1-negative or highly expressing PRRX1.

2. Preclinical mouse models bearing different *Prrx1* levels recapitulate the non-linear relationship between its expression and metastatic burden.

3. CTCs characterization reveals that K14+ cells are the main source of metastasis in the PyMT model.

4. Lung metastases recreate the EMT hierarchy and heterogeneity observed in the primary tumor.

5. The EMT plasticity factor Prrx1 is a master regulator of metastatic programs, coregulating invasion, proliferation, and dormancy.
    a. It is required for invasion.
    b. Intermediate levels provide invasive and proliferative capacities, maximizing metastatic potential.
    c. High levels induce cell cycle arrest and dormancy.

6. Mechanistically, Prrx1 activates promoters of genes encoding cell cycle inhibitors, inhibits those for cell cycle progression and induces a dormancy programme.

7. Acquisition of metastatic potential and dormancy traits starts before the cell leaves the primary tumor.

8. The combination of invasive and proliferative gene signatures enables robust prognostic stratification of breast cancer patients.

1. Los tumores humanos de cáncer de mama se pueden clasificar por la expresión de PRRX1. No hay una relación lineal entre expresión y carga metastásica. Los pacientes con tumores con niveles intermedios de PRRX1 tienen una incidencia metastásica superior a los que tienen tumores negativos o que expresan niveles altos.

2. Modelos murinos preclínicos con diferentes niveles de Prrx1 recapitulan la relación no lineal entre su expresión y la carga metastásica.

3. La caracterización de CTCs revela que las células K14 positivas son la principal fuente de metástasis en el modelo PyMT.

4. Las metástasis pulmonares reproducen la jerarquía y heterogeneidad de los estados de EMT observados en el tumor primario.

5. El factor de EMT y plasticidad Prrx1 es un controlador clave de los programas metastásicos, coregulando invasión, proliferación y latencia.
    a. Es necesario para invasión.
    b. Niveles intermedios proporcionan capacidades invasivas y proliferativas, maximizando el potencial metastásico.
    c. Niveles altos inducen parada del ciclo celular y latencia.

6. A nivel mecanístico, Prrx1 activa promotores de genes codificantes de inhibidores del ciclo celular, bloquea aquellos necesarios para su progresión del ciclo celular e induce un programa de latencia.

7. El potencial metastásico y las características asociadas a latencia celular aparecen antes de que la célula tumoral abandone el tumor primario.

8. La combinación de firmas génicas de invasión y proliferación permite una estratificación prognóstica robusta de pacientes de cáncer de mama.

Bibliography

1. Dillekås, H., Rogers, M. S. & Straume, O. Are 90% of deaths from cancer caused by metastases? *Cancer Med.* **8**, 5574–5576 (2019).
2. Boire, A., Burke, K., Cox, T. R., Guise, T., Jamal-Hanjani, M., Janowitz, T., Kaplan, R., Lee, R., Swanton, C., Vander Heiden, M. G. & Sahai, E. Why do patients with cancer die? *Nat. Rev. Cancer* **24**, 578–589 (2024).
3. Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. EMERGING BIOLOGICAL PRINCIPLES OF METASTASIS. *Cell* **168**, 670–691 (2017).
4. Majidpoor, J. & Mortezaee, K. Steps in metastasis: an updated review. *Med. Oncol. Northwood Lond. Engl.* **38**, 3 (2021).
5. Gerstberger, S., Jiang, Q. & Ganesh, K. Metastasis. *Cell* **186**, 1564–1579 (2023).
6. Luzzi, K. J., MacDonald, I. C., Schmidt, E. E., Kerkvliet, N., Morris, V. L., Chambers, A. F. & Groom, A. C. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am. J. Pathol.* **153**, 865–873 (1998).
7. Massagué, J. & Ganesh, K. Metastasis-Initiating Cells and Ecosystems. *Cancer Discov.* **11**, 971–994 (2021).
8. Thompson, E. W., Redfern, A. D., Brabletz, S., Berx, G., Agarwal, V., Ganesh, K., Huang, R. Y., Ishay-Ronen, D., Savagner, P., Sheng, G., Stemmler, M. P. & Brabletz, T. EMT and cancer: what clinicians should know. *Nat. Rev. Clin. Oncol.* https://doi.org/10.1038/s41571-025-01058-2 (2025) doi:10.1038/s41571-025-01058-2.
9. Dong, A. & Blanpain, C. Identification, functional insights and therapeutic targeting of EMT tumour states. *Nat. Rev. Cancer* https://doi.org/10.1038/s41568-025-00873-0 (2025) doi:10.1038/s41568-025-00873-0.
10. Jiménez-Castaño, R. & Nieto, M. A. Epithelial–Mesenchymal Transition as a Central Driver of Tumor Cell Plasticity. *Nat. Cancer* (2026) doi: 10.1038/s43018-026-01154-x .
11. Dai, C. S., Mishra, A., Edd, J., Toner, M., Maheswaran, S. & Haber, D. A. Circulating tumor cells: Blood-based detection, molecular biology, and clinical applications. *Cancer Cell* **43**, 1399–1422 (2025).
12. Peinado, H., Zhang, H., Matei, I. R., Costa-Silva, B., Hoshino, A., Rodrigues, G., Psaila, B., Kaplan, R. N., Bromberg, J. F., Kang, Y., Bissell, M. J., Cox, T. R., Giaccia, A. J., Erler, J. T., Hiratsuka, S., Ghajar, C. M. & Lyden, D. Pre-metastatic niches: organ-specific homes for metastases. *Nat. Rev. Cancer* **17**, 302–317 (2017).
13. Kim, M.-Y., Oskarsson, T., Acharyya, S., Nguyen, D. X., Zhang, X. H.-F., Norton, L. & Massagué, J. Tumor Self-Seeding by Circulating Cancer Cells. *Cell* **139**, 1315–1326 (2009).
14. Hay, E. D., Fleischmajer, R., & Billingham, R. E. Epithelial-mesenchymal interactions. *Proc. 18th Hahnemann Symp.* (1968).
15. Bellairs, R. in *Developmental and Evolutionary Aspects of the Neural Crest* 123–145 (Wiley, New York).
16. Hay, E. D. & Zuk, A. Transformations between epithelium and mesenchyme: Normal, pathological, and experimentally induced. *Am. J. Kidney Dis.* **26**, 678–690 (1995).
17. Youssef, K. K. & Nieto, M. A. Epithelial–mesenchymal transition in tissue repair and degeneration. *Nat. Rev. Mol. Cell Biol.* 1–20 (2024) doi:10.1038/s41580-024-00733-z.
18. Brabletz, T., Kalluri, R., Nieto, M. A. & Weinberg, R. A. EMT in cancer. *Nat. Rev. Cancer* **18**, 128–134 (2018).
19. Chen, Z. F. & Behringer, R. R. twist is required in head mesenchyme for cranial neural tube morphogenesis. *Genes Dev.* **9**, 686–699 (1995).

20. Comijn, J., Berx, G., Vermassen, P., Verschueren, K., van Grunsven, L., Bruyneel, E., Mareel, M., Huylebroeck, D. & van Roy, F. The two-handed E box binding zinc finger protein SIP1 downregulates E-cadherin and induces invasion. *Mol. Cell* **7**, 1267–1278 (2001).
21. Nieto, M. A., Bennett, M. F., Sargent, M. G. & Wilkinson, D. G. Cloning and developmental expression of Sna, a murine homologue of the Drosophila snail gene. *Dev. Camb. Engl.* **116**, 227–237 (1992).
22. Batlle, E., Sancho, E., Francí, C., Domínguez, D., Monfar, M., Baulida, J. & García De Herreros, A. The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. *Nat. Cell Biol.* **2**, 84–89 (2000).
23. Cano, A., Pérez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., Portillo, F. & Nieto, M. A. The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. *Nat. Cell Biol.* **2**, 76–83 (2000).
24. Spaderna, S., Schmalhofer, O., Hlubek, F., Berx, G., Eger, A., Merkel, S., Jung, A., Kirchner, T. & Brabletz, T. A transient, EMT-linked loss of basement membranes indicates metastasis and poor survival in colorectal cancer. *Gastroenterology* **131**, 830–840 (2006).
25. Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A. & Weinberg, R. A. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**, 927–939 (2004).
26. Thiery, J. P., Acloque, H., Huang, R. Y. J. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871–890 (2009).
27. Nieto, M. A., Huang, R. Y.-J., Jackson, R. A. & Thiery, J. P. EMT: 2016. *Cell* **166**, 21–45 (2016).
28. Gonzalez, D. M. & Medici, D. Signaling mechanisms of the epithelial-mesenchymal transition. *Sci. Signal.* **7**, re8 (2014).
29. Zhang, J., Hu, Z., Horta, C. A. & Yang, J. Regulation of epithelial-mesenchymal transition by tumor microenvironmental signals and its implication in cancer therapeutics. *Semin. Cancer Biol.* **88**, 46–66 (2023).
30. Haerinck, J., Goossens, S. & Berx, G. The epithelial-mesenchymal plasticity landscape: principles of design and mechanisms of regulation. *Nat. Rev. Genet.* **24**, 590–609 (2023).
31. Lee, J. H. & Massagué, J. TGF-β in developmental and fibrogenic EMTs. *Semin. Cancer Biol.* **86**, 136–145 (2022).
32. Krishnamachary, B., Zagzag, D., Nagasawa, H., Rainey, K., Okuyama, H., Baek, J. H. & Semenza, G. L. Hypoxia-Inducible Factor-1-Dependent Repression of E-cadherin in von Hippel-Lindau Tumor Suppressor–Null Renal Cell Carcinoma Mediated by TCF3, ZFHX1A, and ZFHX1B. *Cancer Res.* **66**, 2725–2731 (2006).
33. Yang, M.-H., Wu, M.-Z., Chiou, S.-H., Chen, P.-M., Chang, S.-Y., Liu, C.-J., Teng, S.-C. & Wu, K.-J. Direct regulation of TWIST by HIF-1alpha promotes metastasis. *Nat. Cell Biol.* **10**, 295–305 (2008).
34. Wei, S. C., Fattet, L., Tsai, J. H., Guo, Y., Pai, V. H., Majeski, H. E., Chen, A. C., Sah, R. L., Taylor, S. S., Engler, A. J. & Yang, J. Matrix stiffness drives epithelial-mesenchymal transition and tumour metastasis through a TWIST1-G3BP2 mechanotransduction pathway. *Nat. Cell Biol.* **17**, 678–688 (2015).
35. Fattet, L., Jung, H.-Y., Matsumoto, M. W., Aubol, B. E., Kumar, A., Adams, J. A., Chen, A. C., Sah, R. L., Engler, A. J., Pasquale, E. B. & Yang, J. Matrix Rigidity Controls Epithelial-Mesenchymal Plasticity and Tumor Metastasis via a Mechanoresponsive EPHA2/LYN Complex. *Dev. Cell* **54**, 302-316.e7 (2020).
36. Zhang, P., Wei, Y., Wang, L., Debeb, B. G., Yuan, Y., Zhang, J., Yuan, J., Wang, M., Chen, D., Sun, Y., Woodward, W. A., Liu, Y., Dean, D. C., Liang, H., Hu, Y., Ang, K. K., Hung, M.-C.,

Chen, J. & Ma, L. ATM-mediated stabilization of ZEB1 promotes DNA damage response and radioresistance through CHK1. *Nat. Cell Biol.* **16**, 864–875 (2014).
37. Hayes, J. D., Dinkova-Kostova, A. T. & Tew, K. D. Oxidative Stress in Cancer. *Cancer Cell* **38**, 167–197 (2020).
38. Bhattacharya, D., Azambuja, A. P. & Simoes-Costa, M. Metabolic Reprogramming Promotes Neural Crest Migration via Yap/Tead Signaling. *Dev. Cell* **53**, 199-211.e6 (2020).
39. Díaz-López, A., Moreno-Bueno, G. & Cano, A. Role of microRNA in epithelial to mesenchymal transition and metastasis and clinical perspectives. *Cancer Manag. Res.* **6**, 205–216 (2014).
40. Neumann, D. P., Goodall, G. J. & Gregory, P. A. Regulation of splicing and circularisation of RNA in epithelial mesenchymal plasticity. *Semin. Cell Dev. Biol.* **75**, 50–60 (2018).
41. Skrypek, N., Goossens, S., De Smedt, E., Vandamme, N. & Berx, G. Epithelial-to-Mesenchymal Transition: Epigenetic Reprogramming Driving Cellular Plasticity. *Trends Genet. TIG* **33**, 943–959 (2017).
42. Youssef, K. K., Narwade, N., Arcas, A., Marquez-Galera, A., Jiménez-Castaño, R., Lopez-Blau, C., Fazilaty, H., García-Gutierrez, D., Cano, A., Galcerán, J., Moreno-Bueno, G., Lopez-Atalaya, J. P. & Nieto, M. A. Two distinct epithelial-to-mesenchymal transition programs control invasion and inflammation in segregated tumor cell populations. *Nat. Cancer* **5**, 1660–1680 (2024).
43. Fazilaty, H., Rago, L., Kass Youssef, K., Ocaña, O. H., Garcia-Asencio, F., Arcas, A., Galceran, J. & Nieto, M. A. A gene regulatory network to control EMT programs in development and disease. *Nat. Commun.* **10**, 5115 (2019).
44. Rukstalis, J. M. & Habener, J. F. Snail2, a mediator of epithelial-mesenchymal transitions, expressed in progenitor cells of the developing endocrine pancreas. *Gene Expr. Patterns GEP* **7**, 471–479 (2007).
45. Lasierra Losada, M., Pauler, M., Vandamme, N., Goossens, S., Berx, G., Leppkes, M., Schuhwerk, H., Brabletz, S., Brabletz, T. & Stemmler, M. P. Pancreas morphogenesis and homeostasis depends on tightly regulated Zeb1 levels in epithelial cells. *Cell Death Discov.* **7**, 138 (2021).
46. Caramel, J., Papadogeorgakis, E., Hill, L., Browne, G. J., Richard, G., Wierinckx, A., Saldanha, G., Osborne, J., Hutchinson, P., Tse, G., Lachuer, J., Puisieux, A., Pringle, J. H., Ansieau, S. & Tulchinsky, E. A switch in the expression of embryonic EMT-inducers drives the development of malignant melanoma. *Cancer Cell* **24**, 466–480 (2013).
47. Aharonov, A., Shakked, A., Umansky, K. B., Savidor, A., Genzelinakh, A., Kain, D., Lendengolts, D., Revach, O.-Y., Morikawa, Y., Dong, J., Levin, Y., Geiger, B., Martin, J. F. & Tzahor, E. ERBB2 drives YAP activation and EMT-like processes during cardiac regeneration. *Nat. Cell Biol.* **22**, 1346–1356 (2020).
48. Grande, M. T., Sánchez-Laorden, B., López-Blau, C., De Frutos, C. A., Boutet, A., Arévalo, M., Rowe, R. G., Weiss, S. J., López-Novoa, J. M. & Nieto, M. A. Snail1-induced partial epithelial-to-mesenchymal transition drives renal fibrosis in mice and can be targeted to reverse established disease. *Nat. Med.* **21**, 989–997 (2015).
49. Lovisa, S., LeBleu, V. S., Tampe, B., Sugimoto, H., Vadnagara, K., Carstens, J. L., Wu, C.-C., Hagos, Y., Burckhardt, B. C., Pentcheva-Hoang, T., Nischal, H., Allison, J. P., Zeisberg, M. & Kalluri, R. Epithelial-to-mesenchymal transition induces cell cycle arrest and parenchymal damage in renal fibrosis. *Nat. Med.* **21**, 998–1009 (2015).
50. Sinha, D., Saha, P., Samanta, A. & Bishayee, A. Emerging Concepts of Hybrid Epithelial-to-Mesenchymal Transition in Cancer Progression. *Biomolecules* **10**, 1561 (2020).

51. Ocaña, O. H., Córcoles, R., Fabra, A., Moreno-Bueno, G., Acloque, H., Vega, S., Barrallo-Gimeno, A., Cano, A. & Nieto, M. A. Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell* **22**, 709–724 (2012).
52. Tsai, J. H., Donaher, J. L., Murphy, D. A., Chau, S. & Yang, J. Spatiotemporal Regulation of Epithelial-Mesenchymal Transition is Essential for Squamous Cell Carcinoma Metastasis. *Cancer Cell* **22**, 725–736 (2012).
53. Hanahan, D. Hallmarks of cancer-Then and now, and beyond. *Cell* S0092-8674(25)01498–9 (2026) doi:10.1016/j.cell.2025.12.049.
54. Cammareri, P., Raponi, M., Hong, Y., Billard, C. V., Peckett, N., Zhu, Y., Velez-Bravo, F. D., Younger, N. T., Dunican, D. S., Pohl, S. Ö.-G., Bastem Akan, A., Doleschall, N. J., Falconer, J., White, M., Quinn, J., Pennel, K., Garau, R., Malla, S. B., Dunne, P. D., Meehan, R. R., Sansom, O. J., Edwards, J., Dunlop, M. G., Din, F. V. N., Tejpar, S., Steele, C. W. & Myant, K. B. Loss of colonic fidelity enables multilineage plasticity and metastasis. *Nature* https://doi.org/10.1038/s41586-025-09125-5 (2025) doi:10.1038/s41586-025-09125-5.
55. Hanahan, D. Hallmarks of Cancer: New Dimensions. *Cancer Discov.* **12**, 31–46 (2022).
56. Moorman, A., Benitez, E. K., Cambulli, F., Jiang, Q., Mahmoud, A., Lumish, M., Hartner, S., Balkaran, S., Bermeo, J., Asawa, S., Firat, C., Saxena, A., Wu, F., Luthra, A., Burdziak, C., Xie, Y., Sgambati, V., Luckett, K., Li, Y., Yi, Z., Masilionis, I., Soares, K., Pappou, E., Yaeger, R., Kingham, T. P., Jarnagin, W., Paty, P. B., Weiser, M. R., Mazutis, L., D'Angelica, M., Shia, J., Garcia-Aguilar, J., Nawy, T., Hollmann, T. J., Chaligné, R., Sanchez-Vega, F., Sharma, R., Pe'er, D. & Ganesh, K. Progressive plasticity during colorectal cancer metastasis. *Nature* **637**, 947–954 (2025).
57. Samavarchi-Tehrani, P., Golipour, A., David, L., Sung, H.-K., Beyer, T. A., Datti, A., Woltjen, K., Nagy, A. & Wrana, J. L. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010).
58. Jj, U., R, Z., K, K., M, C., Jt, P., S, R., T, O., T, S., Ra, W. & Gq, D. The epithelial-mesenchymal transition factor SNAIL paradoxically enhances reprogramming. *Stem Cell Rep.* **3**, (2014).
59. D'Uva, G., Aharonov, A., Lauriola, M., Kain, D., Yahalom-Ronen, Y., Carvalho, S., Weisinger, K., Bassat, E., Rajchman, D., Yifa, O., Lysenko, M., Konfino, T., Hegesh, J., Brenner, O., Neeman, M., Yarden, Y., Leor, J., Sarig, R., Harvey, R. P. & Tzahor, E. ERBB2 triggers mammalian heart regeneration by promoting cardiomyocyte dedifferentiation and proliferation. *Nat. Cell Biol.* **17**, 627–638 (2015).
60. Yaeger, R., Chatila, W. K., Lipsyc, M. D., Hechtman, J. F., Cercek, A., Sanchez-Vega, F., Jayakumaran, G., Middha, S., Zehir, A., Donoghue, M. T. A., You, D., Viale, A., Kemeny, N., Segal, N. H., Stadler, Z. K., Varghese, A. M., Kundra, R., Gao, J., Syed, A., Hyman, D. M., Vakiani, E., Rosen, N., Taylor, B. S., Ladanyi, M., Berger, M. F., Solit, D. B., Shia, J., Saltz, L. & Schultz, N. Clinical Sequencing Defines the Genomic Landscape of Metastatic Colorectal Cancer. *Cancer Cell* **33**, 125-136.e3 (2018).
61. Martínez-Jiménez, F., Movasati, A., Brunner, S. R., Nguyen, L., Priestley, P., Cuppen, E. & Van Hoeck, A. Pan-cancer whole-genome comparison of primary and metastatic solid tumours. *Nature* **618**, 333–341 (2023).
62. Nguyen, B., Fong, C., Luthra, A., Smith, S. A., DiNatale, R. G., Nandakumar, S., Walch, H., Chatila, W. K., Madupuri, R., Kundra, R., Bielski, C. M., Mastrogiacomo, B., Donoghue, M. T. A., Boire, A., Chandarlapaty, S., Ganesh, K., Harding, J. J., Iacobuzio-Donahue, C. A., Razavi, P., Reznik, E., Rudin, C. M., Zamarin, D., Abida, W., Abou-Alfa, G. K., Aghajanian, C., Cercek, A., Chi, P., Feldman, D., Ho, A. L., Iyer, G., Janjigian, Y. Y., Morris, M., Motzer, R.

J., O'Reilly, E. M., Postow, M. A., Raj, N. P., Riely, G. J., Robson, M. E., Rosenberg, J. E., Safonov, A., Shoushtari, A. N., Tap, W., Teo, M. Y., Varghese, A. M., Voss, M., Yaeger, R., Zauderer, M. G., Abu-Rustum, N., Garcia-Aguilar, J., Bochner, B., Hakimi, A., Jarnagin, W. R., Jones, D. R., Molena, D., Morris, L., Rios-Doria, E., Russo, P., Singer, S., Strong, V. E., Chakravarty, D., Ellenson, L. H., Gopalan, A., Reis-Filho, J. S., Weigelt, B., Ladanyi, M., Gonen, M., Shah, S. P., Massague, J., Gao, J., Zehir, A., Berger, M. F., Solit, D. B., Bakhoum, S. F., Sanchez-Vega, F. & Schultz, N. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185**, 563-575.e11 (2022).

63. Stehelin, D., Varmus, H. E., Bishop, J. M. & Vogt, P. K. DNA related to the transforming gene(s) of avian sarcoma viruses is present in normal avian DNA. *Nature* **260**, 170–173 (1976).
64. Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R. C. & Croce, C. M. Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 7824–7827 (1982).
65. Santos, E., Tronick, S. R., Aaronson, S. A., Pulciani, S. & Barbacid, M. T24 human bladder carcinoma oncogene is an activated form of the normal human homologue of BALB- and Harvey-MSV transforming genes. *Nature* **298**, 343–347 (1982).
66. Shih, C. & Weinberg, R. A. Isolation of a transforming sequence from a human bladder carcinoma cell line. *Cell* **29**, 161–169 (1982).
67. Coorens, T. H. H., Oh, J. W., Choi, Y. A., Lim, N. S., Zhao, B., Voshall, A., Abyzov, A., Antonacci-Fulton, L., Aparicio, S., Ardlie, K. G., Bell, T. J., Bennett, J. T., Bernstein, B. E., Blanchard, T. G., Boyle, A. P., Buenrostro, J. D., Burns, K. H., Chen, F., Chen, R., Choudhury, S., Doddapaneni, H. V., Eichler, E. E., Evrony, G. D., Faith, M. A., Fazzio, T. G., Fulton, R. S., Garber, M., Gehlenborg, N., Germer, S., Getz, G., Gibbs, R. A., Hernandez, R. G., Jin, F., Korbel, J. O., Landau, D. A., Lawson, H. A., Lennon, N. J., Li, H., Li, Y., Loh, P.-R., Marth, G., McConnell, M. J., Mills, R. E., Montgomery, S. B., Natarajan, P., Park, P. J., Satija, R., Sedlazeck, F. J., Shao, D. D., Shen, H., Stergachis, A. B., Underhill, H. R., Urban, A. E., VonDran, M. W., Walsh, C. A., Wang, T., Wu, T. P., Zong, C., Lee, E. A., Vaccarino, F. M., & Somatic Mosaicism across Human Tissues Network. The Somatic Mosaicism across Human Tissues Network. *Nature* **643**, 47–59 (2025).
68. Kakiuchi, N. & Ogawa, S. Clonal expansion in non-cancer tissues. *Nat. Rev. Cancer* **21**, 239–256 (2021).
69. Riva, L., Pandiri, A. R., Li, Y. R., Droop, A., Hewinson, J., Quail, M. A., Iyer, V., Shepherd, R., Herbert, R. A., Campbell, P. J., Sills, R. C., Alexandrov, L. B., Balmain, A. & Adams, D. J. The mutational signature profile of known and suspected human carcinogens in mice. *Nat. Genet.* **52**, 1189–1197 (2020).
70. Lopez-Bigas, N. & Gonzalez-Perez, A. Are carcinogens direct mutagens? *Nat. Genet.* **52**, 1137–1138 (2020).
71. Parreno, V., Loubiere, V., Schuettengruber, B., Fritsch, L., Rawal, C. C., Erokhin, M., Győrffy, B., Normanno, D., Di Stefano, M., Moreaux, J., Butova, N. L., Chiolo, I., Chetverina, D., Martinez, A.-M. & Cavalli, G. Transient loss of Polycomb components induces an epigenetic cancer fate. *Nature* **629**, 688–696 (2024).
72. Youssef, K. K., Van Keymeulen, A., Lapouge, G., Beck, B., Michaux, C., Achouri, Y., Sotiropoulou, P. A. & Blanpain, C. Identification of the cell lineage at the origin of basal cell carcinoma. *Nat. Cell Biol.* **12**, 299–305 (2010).
73. Alonso-Curbelo, D., Ho, Y.-J., Burdziak, C., Maag, J. L. V., Morris, J. P., Chandwani, R., Chen, H.-A., Tsanov, K. M., Barriga, F. M., Luan, W., Tasdemir, N., Livshits, G., Azizi, E., Chun, J.,

Wilkinson, J. E., Mazutis, L., Leach, S. D., Koche, R., Pe'er, D. & Lowe, S. W. A gene–environment-induced epigenetic program initiates tumorigenesis. *Nature* **590**, 642–648 (2021).

74. Bansaccal, N., Vieugue, P., Sarate, R., Song, Y., Minguijon, E., Miroshnikova, Y. A., Zeuschner, D., Collin, A., Allard, J., Engelman, D., Delaunois, A.-L., Liagre, M., de Groote, L., Timmerman, E., Van Haver, D., Impens, F., Salmon, I., Wickström, S. A., Sifrim, A. & Blanpain, C. The extracellular matrix dictates regional competence for tumour initiation. *Nature* **623**, 828–835 (2023).
75. Zhang, S., Xiao, X., Yi, Y., Wang, X., Zhu, L., Shen, Y., Lin, D. & Wu, C. Tumor initiation and early tumorigenesis: molecular mechanisms and interventional targets. *Signal Transduct. Target. Ther.* **9**, 149 (2024).
76. Mani, S. A., Guo, W., Liao, M.-J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shipitsin, M., Campbell, L. L., Polyak, K., Brisken, C., Yang, J. & Weinberg, R. A. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715 (2008).
77. Morel, A.-P., Lièvre, M., Thomas, C., Hinkal, G., Ansieau, S. & Puisieux, A. Generation of breast cancer stem cells through epithelial-mesenchymal transition. *PloS One* **3**, e2888 (2008).
78. Ohashi, S., Natsuizaka, M., Wong, G. S., Michaylira, C. Z., Grugan, K. D., Stairs, D. B., Kalabis, J., Vega, M. E., Kalman, R. A., Nakagawa, M., Klein-Szanto, A. J., Herlyn, M., Diehl, J. A., Rustgi, A. K. & Nakagawa, H. Epidermal growth factor receptor and mutant p53 expand an esophageal cellular subpopulation capable of epithelial-to-mesenchymal transition through ZEB transcription factors. *Cancer Res.* **70**, 4174–4184 (2010).
79. Vega, S., Morales, A. V., Ocaña, O. H., Valdés, F., Fabregat, I. & Nieto, M. A. Snail blocks the cell cycle and confers resistance to cell death. *Genes Dev.* **18**, 1131–1143 (2004).
80. Ye, X., Tam, W. L., Shibue, T., Kaygusuz, Y., Reinhardt, F., Ng Eaton, E. & Weinberg, R. A. Distinct EMT programs control normal mammary stem cells and tumour-initiating cells. *Nature* **525**, 256–260 (2015).
81. Ni, T., Li, X.-Y., Lu, N., An, T., Liu, Z.-P., Fu, R., Lv, W.-C., Zhang, Y.-W., Xu, X.-J., Grant Rowe, R., Lin, Y.-S., Scherer, A., Feinberg, T., Zheng, X.-Q., Chen, B.-A., Liu, X. S., Guo, Q.-L., Wu, Z.-Q. & Weiss, S. J. Snail1-dependent p53 repression regulates expansion and activity of tumour-initiating cells in breast cancer. *Nat. Cell Biol.* **18**, 1221–1232 (2016).
82. Tran, H. D., Luitel, K., Kim, M., Zhang, K., Longmore, G. D. & Tran, D. D. Transient SNAIL1 expression is necessary for metastatic competence in breast cancer. *Cancer Res.* **74**, 6330–6340 (2014).
83. Landragin, C., Saichi, M., Laisné, M., Durand, A., Prompsy, P., Leclere, R., Mesple, J., Borgman, K., Trouchet, A., Faraldo, M. M., Chiche, A., Vincent-Salomon, A., Salmon, H. & Vallot, C. Epigenomic disorder and partial EMT impair luminal progenitor integrity in Brca1-associated breast tumorigenesis. *Mol. Cancer* **24**, 127 (2025).
84. Perelli, L., Zhang, L., Mangiameli, S., Giannese, F., Mahadevan, K. K., Peng, F., Citron, F., Khan, H., Le, C., Gurreri, E., Carbone, F., Russell, A. J. C., Soeung, M., Lam, T. N. A., Lundgren, S., Marisetty, S., Zhu, C., Catania, D., Mohamed, A. M. T., Feng, N., Augustine, J. J., Sgambato, A., Tortora, G., Draetta, G. F., Tonon, G., Futreal, A., Giuliani, V., Carugo, A., Viale, A., Kim, M. P., Heffernan, T. P., Wang, L., Kalluri, R., Cittaro, D., Chen, F. & Genovese, G. Evolutionary fingerprints of epithelial-to-mesenchymal transition. *Nature* **640**, 1083–1092 (2025).

85. Greten, F. R. & Grivennikov, S. I. Inflammation and Cancer: Triggers, Mechanisms, and Consequences. *Immunity* **51**, 27–41 (2019).
86. Chen, H., Tao, X., Cao, H., Li, B., Sun, Q., Wang, W., Zou, Y., Mu, M., Tao, H., Zhao, Y. & Ge, D. Nicotine exposure exacerbates silica-induced pulmonary fibrosis via STAT3-BDNF-TrkB-mediated epithelial-mesenchymal transition in alveolar type II cells. *Food Chem. Toxicol. Int. J. Publ. Br. Ind. Biol. Res. Assoc.* **175**, 113694 (2023).
87. Chen, Y.-C., Chuang, T.-Y., Liu, C.-W., Liu, C.-W., Lee, T.-L., Lai, T.-C. & Chen, Y.-L. Particulate matters increase epithelial-mesenchymal transition and lung fibrosis through the ETS-1/NF-κB-dependent pathway in lung epithelial cells. *Part. Fibre Toxicol.* **17**, 41 (2020).
88. Bowers, L. W., Rossi, E. L., McDonell, S. B., Doerstling, S. S., Khatib, S. A., Lineberger, C. G., Albright, J. E., Tang, X., deGraffenried, L. A. & Hursting, S. D. Leptin Signaling Mediates Obesity-Associated CSC Enrichment and EMT in Preclinical TNBC Models. *Mol. Cancer Res. MCR* **16**, 869–879 (2018).
89. Bornes, L., van Winden, L. J., Geurts, V. C. M., de Bruijn, B., Azarang, L., Lanfermeijer, M., Caruso, M., Proost, N., Boeije, M., Lohuis, J. O., Belthier, G., Noguera Delgado, E., de Gruil, N., Kroep, J. R., van de Ven, M., Menezes, R., Wesseling, J., Kok, M., Linn, S., Broeks, A., van Rossum, H. H., Scheele, C. L. G. J. & van Rheenen, J. The oestrous cycle stage affects mammary tumour sensitivity to chemotherapy. *Nature* **637**, 195–204 (2025).
90. Gerlinger, M., Rowan, A. J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N. Q., Butler, A., Jones, D., Raine, K., Latimer, C., Santos, C. R., Nohadani, M., Eklund, A. C., Spencer-Dene, B., Clark, G., Pickering, L., Stamp, G., Gore, M., Szallasi, Z., Downward, J., Futreal, P. A. & Swanton, C. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
91. Ramón Y Cajal, S., Sesé, M., Capdevila, C., Aasen, T., De Mattos-Arruda, L., Diaz-Cano, S. J., Hernández-Losa, J. & Castellví, J. Clinical implications of intratumor heterogeneity: challenges and opportunities. *J. Mol. Med. Berl. Ger.* **98**, 161–177 (2020).
92. Al Bakir, M., Huebner, A., Martínez-Ruiz, C., Grigoriadis, K., Watkins, T. B. K., Pich, O., Moore, D. A., Veeriah, S., Ward, S., Laycock, J., Johnson, D., Rowan, A., Razaq, M., Akther, M., Naceur-Lombardelli, C., Prymas, P., Toncheva, A., Hessey, S., Dietzen, M., Colliver, E., Frankell, A. M., Bunkum, A., Lim, E. L., Karasaki, T., Abbosh, C., Hiley, C. T., Hill, M. S., Cook, D. E., Wilson, G. A., Salgado, R., Nye, E., Stone, R. K., Fennell, D. A., Price, G., Kerr, K. M., Naidu, B., Middleton, G., Summers, Y., Lindsay, C. R., Blackhall, F. H., Cave, J., Blyth, K. G., Nair, A., Ahmed, A., Taylor, M. N., Procter, A. J., Falzon, M., Lawrence, D., Navani, N., Thakrar, R. M., Janes, S. M., Papadatos-Pastos, D., Forster, M. D., Lee, S. M., Ahmad, T., Quezada, S. A., Peggs, K. S., Van Loo, P., Dive, C., Hackshaw, A., Birkbak, N. J., Zaccaria, S., TRACERx Consortium, Jamal-Hanjani, M., McGranahan, N. & Swanton, C. The evolution of non-small cell lung cancer metastases in TRACERx. *Nature* **616**, 534–542 (2023).
93. Gavish, A., Tyler, M., Greenwald, A. C., Hoefflin, R., Simkin, D., Tschernichovsky, R., Galili Darnell, N., Somech, E., Barbolin, C., Antman, T., Kovarsky, D., Barrett, T., Gonzalez Castro, L. N., Halder, D., Chanoch-Myers, R., Laffy, J., Mints, M., Wider, A., Tal, R., Spitzer, A., Hara, T., Raitses-Gurevich, M., Stossel, C., Golan, T., Tirosh, A., Suvà, M. L., Puram, S. V. & Tirosh, I. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature* **618**, 598–606 (2023).
94. Regev, A., Teichmann, S. A., Lander, E. S., Amit, I., Benoist, C., Birney, E., Bodenmiller, B., Campbell, P., Carninci, P., Clatworthy, M., Clevers, H., Deplancke, B., Dunham, I.,

Eberwine, J., Eils, R., Enard, W., Farmer, A., Fugger, L., Göttgens, B., Hacohen, N., Haniffa, M., Hemberg, M., Kim, S., Klenerman, P., Kriegstein, A., Lein, E., Linnarsson, S., Lundberg, E., Lundeberg, J., Majumder, P., Marioni, J. C., Merad, M., Mhlanga, M., Nawijn, M., Netea, M., Nolan, G., Pe'er, D., Phillipakis, A., Ponting, C. P., Quake, S., Reik, W., Rozenblatt-Rosen, O., Sanes, J., Satija, R., Schumacher, T. N., Shalek, A., Shapiro, E., Sharma, P., Shin, J. W., Stegle, O., Stratton, M., Stubbington, M. J. T., Theis, F. J., Uhlen, M., van Oudenaarden, A., Wagner, A., Watt, F., Weissman, J., Wold, B., Xavier, R., Yosef, N., & Human Cell Atlas Meeting Participants. The Human Cell Atlas. *eLife* **6**, e27041 (2017).

95. Rozenblatt-Rosen, O., Regev, A., Oberdoerffer, P., Nawy, T., Hupalowska, A., Rood, J. E., Ashenberg, O., Cerami, E., Coffey, R. J., Demir, E., Ding, L., Esplin, E. D., Ford, J. M., Goecks, J., Ghosh, S., Gray, J. W., Guinney, J., Hanlon, S. E., Hughes, S. K., Hwang, E. S., Iacobuzio-Donahue, C. A., Jané-Valbuena, J., Johnson, B. E., Lau, K. S., Lively, T., Mazzilli, S. A., Pe'er, D., Santagata, S., Shalek, A. K., Schapiro, D., Snyder, M. P., Sorger, P. K., Spira, A. E., Srivastava, S., Tan, K., West, R. B., Williams, E. H., & Human Tumor Atlas Network. The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. *Cell* **181**, 236–249 (2020).
96. 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., Deschler, D. G., Varvares, M. A., Mylvaganam, R., Rozenblatt-Rosen, O., Rocco, J. W., Faquin, W. C., Lin, D. T., Regev, A. & Bernstein, B. E. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell* **171**, 1611-1624.e24 (2017).
97. Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.-R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G. K., Kazer, S. W., Gaillard, A., Kolb, K. E., Villani, A.-C., Johannessen, C. M., Andreev, A. Y., Van Allen, E. M., Bertagnolli, M., Sorger, P. K., Sullivan, R. J., Flaherty, K. T., Frederick, D. T., Jané-Valbuena, J., Yoon, C. H., Rozenblatt-Rosen, O., Shalek, A. K., Regev, A. & Garraway, L. A. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
98. Neftel, C., Laffy, J., Filbin, M. G., Hara, T., Shore, M. E., Rahme, G. J., Richman, A. R., Silverbush, D., Shaw, M. L., Hebert, C. M., Dewitt, J., Gritsch, S., Perez, E. M., Gonzalez Castro, L. N., Lan, X., Druck, N., Rodman, C., Dionne, D., Kaplan, A., Bertalan, M. S., Small, J., Pelton, K., Becker, S., Bonal, D., Nguyen, Q.-D., Servis, R. L., Fung, J. M., Mylvaganam, R., Mayr, L., Gojo, J., Haberler, C., Geyeregger, R., Czech, T., Slavc, I., Nahed, B. V., Curry, W. T., Carter, B. S., Wakimoto, H., Brastianos, P. K., Batchelor, T. T., Stemmer-Rachamimov, A., Martinez-Lage, M., Frosch, M. P., Stamenkovic, I., Riggi, N., Rheinbay, E., Monje, M., Rozenblatt-Rosen, O., Cahill, D. P., Patel, A. P., Hunter, T., Verma, I. M., Ligon, K. L., Louis, D. N., Regev, A., Bernstein, B. E., Tirosh, I. & Suvà, M. L. An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell* **178**, 835-849.e21 (2019).
99. Kröger, C., Afeyan, A., Mraz, J., Eaton, E. N., Reinhardt, F., Khodor, Y. L., Thiru, P., Bierie, B., Ye, X., Burge, C. B. & Weinberg, R. A. Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 7353–7362 (2019).
100. Pastushenko, I., Brisebarre, A., Sifrim, A., Fioramonti, M., Revenco, T., Boumahdi, S., Van Keymeulen, A., Brown, D., Moers, V., Lemaire, S., De Clercq, S., Minguijón, E., Balsat, C., Sokolow, Y., Dubois, C., De Cock, F., Scozzaro, S., Sopena, F., Lanas, A., D'Haene, N.,

Salmon, I., Marine, J.-C., Voet, T., Sotiropoulou, P. A. & Blanpain, C. Identification of the tumour transition states occurring during EMT. *Nature* **556**, 463–468 (2018).

101. Lawson, D. A., Bhakta, N. R., Kessenbrock, K., Prummel, K. D., Yu, Y., Takai, K., Zhou, A., Eyob, H., Balakrishnan, S., Wang, C.-Y., Yaswen, P., Goga, A. & Werb, Z. Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526**, 131–135 (2015).
102. Zhang, Y., Donaher, J. L., Das, S., Li, X., Reinhardt, F., Krall, J. A., Lambert, A. W., Thiru, P., Keys, H. R., Khan, M., Hofree, M., Wilson, M. M., Yedier-Bayram, O., Lack, N. A., Onder, T. T., Bagci-Onder, T., Tyler, M., Tirosh, I., Regev, A., Lees, J. A. & Weinberg, R. A. Genome-wide CRISPR screen identifies PRC2 and KMT2D-COMPASS as regulators of distinct EMT trajectories that contribute differentially to metastasis. *Nat. Cell Biol.* **24**, 554–564 (2022).
103. Karras, P., Bordeu, I., Pozniak, J., Nowosad, A., Pazzi, C., Van Raemdonck, N., Landeloos, E., Van Herck, Y., Pedri, D., Bervoets, G., Makhzami, S., Khoo, J. H., Pavie, B., Lamote, J., Marin-Bejar, O., Dewaele, M., Liang, H., Zhang, X., Hua, Y., Wouters, J., Browaeys, R., Bergers, G., Saeys, Y., Bosisio, F., van den Oord, J., Lambrechts, D., Rustgi, A. K., Bechter, O., Blanpain, C., Simons, B. D., Rambow, F. & Marine, J.-C. A cellular hierarchy in melanoma uncouples growth and metastasis. *Nature* **610**, 190–198 (2022).
104. Su, J., Morgani, S. M., David, C. J., Wang, Q., Er, E. E., Huang, Y.-H., Basnet, H., Zou, Y., Shu, W., Soni, R. K., Hendrickson, R. C., Hadjantonakis, A.-K. & Massagué, J. TGF-β orchestrates fibrogenic and developmental EMTs via the RAS effector RREB1. *Nature* **577**, 566–571 (2020).
105. Lee, J. H., Sánchez-Rivera, F. J., He, L., Basnet, H., Chen, F. X., Spina, E., Li, L., Torner, C., Chan, J. E., Yarlagadda, D. V. K., Park, J. S., Sussman, C., Rudin, C. M., Lowe, S. W., Tammela, T., Macias, M. J., Koche, R. P. & Massagué, J. TGF-β and RAS jointly unmask primed enhancers to drive metastasis. *Cell* **187**, 6182-6199.e29 (2024).
106. Celià-Terrassa, T., Meca-Cortés, O., Mateo, F., Martínez de Paz, A., Rubio, N., Arnal-Estapé, A., Ell, B. J., Bermudo, R., Díaz, A., Guerra-Rebollo, M., Lozano, J. J., Estarás, C., Ulloa, C., Álvarez-Simón, D., Milà, J., Vilella, R., Paciucci, R., Martínez-Balbás, M., de Herreros, A. G., Gomis, R. R., Kang, Y., Blanco, J., Fernández, P. L. & Thomson, T. M. Epithelial-mesenchymal transition can suppress major attributes of human epithelial tumor-initiating cells. *J. Clin. Invest.* **122**, 1849–1868 (2012).
107. Neelakantan, D., Zhou, H., Oliphant, M. U. J., Zhang, X., Simon, L. M., Henke, D. M., Shaw, C. A., Wu, M.-F., Hilsenbeck, S. G., White, L. D., Lewis, M. T. & Ford, H. L. EMT cells increase breast cancer metastasis via paracrine GLI activation in neighbouring tumour cells. *Nat. Commun.* **8**, 15773 (2017).
108. Tsuji, T., Ibaragi, S., Shima, K., Hu, M. G., Katsurano, M., Sasaki, A. & Hu, G. Epithelial-mesenchymal transition induced by growth suppressor p12CDK2-AP1 promotes tumor cell local invasion but suppresses distant colony growth. *Cancer Res.* **68**, 10377–10386 (2008).
109. Lan, L., Evan, T., Li, H., Hussain, A., Ruiz, E. J., Zaw Thin, M., Ferreira, R. M. M., Ps, H., Riising, E. M., Zen, Y., Almagro, J., Ng, K. W., Soro-Barrio, P., Nelson, J., Koifman, G., Carvalho, J., Nye, E. L., He, Y., Zhang, C., Sadanandam, A. & Behrens, A. GREM1 is required to maintain cellular heterogeneity in pancreatic cancer. *Nature* **607**, 163–168 (2022).
110. Li, H., Lan, L., Chen, H., Zaw Thin, M., Ps, H., Nelson, J. K., Evans, I. M., Ruiz, E. J., Cheng, R., Tran, L., Allen, M., Ma, J., Yi, T., Wang, C., He, Y., Guppy, N., Sadanandam, A., Lin, S.-Z., Zhang, C. & Behrens, A. SPP1 is required for maintaining mesenchymal cell fate in pancreatic cancer. *Nature* 1–7 (2025) doi:10.1038/s41586-025-09574-y.

111. Neri, S., Miyashita, T., Hashimoto, H., Suda, Y., Ishibashi, M., Kii, H., Watanabe, H., Kuwata, T., Tsuboi, M., Goto, K., Menju, T., Sonobe, M., Date, H., Ochiai, A. & Ishii, G. Fibroblast-led cancer cell invasion is activated by epithelial-mesenchymal transition through platelet-derived growth factor BB secretion of lung adenocarcinoma. *Cancer Lett*. **395**, 20–30 (2017).
112. Dongre, A., Rashidian, M., Eaton, E. N., Reinhardt, F., Thiru, P., Zagorulya, M., Nepal, S., Banaz, T., Martner, A., Spranger, S. & Weinberg, R. A. Direct and Indirect Regulators of Epithelial-Mesenchymal Transition-Mediated Immunosuppression in Breast Carcinomas. *Cancer Discov.* **11**, 1286–1305 (2021).
113. Sharon, Y., Raz, Y., Cohen, N., Ben-Shmuel, A., Schwartz, H., Geiger, T. & Erez, N. Tumor-derived osteopontin reprograms normal mammary fibroblasts to promote inflammation and tumor growth in breast cancer. *Cancer Res*. **75**, 963–973 (2015).
114. Goulet, C. R., Champagne, A., Bernard, G., Vandal, D., Chabaud, S., Pouliot, F. & Bolduc, S. Cancer-associated fibroblasts induce epithelial-mesenchymal transition of bladder cancer cells through paracrine IL-6 signalling. *BMC Cancer* **19**, 137 (2019).
115. Orimo, A., Gupta, P. B., Sgroi, D. C., Arenzana-Seisdedos, F., Delaunay, T., Naeem, R., Carey, V. J., Richardson, A. L. & Weinberg, R. A. Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**, 335–348 (2005).
116. Labernadie, A., Kato, T., Brugués, A., Serra-Picamal, X., Derzsi, S., Arwert, E., Weston, A., González-Tarragó, V., Elosegui-Artola, A., Albertazzi, L., Alcaraz, J., Roca-Cusachs, P., Sahai, E. & Trepat, X. A mechanically active heterotypic E-cadherin/N-cadherin adhesion enables fibroblasts to drive cancer cell invasion. *Nat. Cell Biol.* **19**, 224–237 (2017).
117. Che, D., Zhang, S., Jing, Z., Shang, L., Jin, S., Liu, F., Shen, J., Li, Y., Hu, J., Meng, Q. & Yu, Y. Macrophages induce EMT to promote invasion of lung cancer cells through the IL-6-mediated COX-2/PGE2/β-catenin signalling pathway. *Mol. Immunol.* **90**, 197–210 (2017).
118. Wei, C., Yang, C., Wang, S., Shi, D., Zhang, C., Lin, X., Liu, Q., Dou, R. & Xiong, B. Crosstalk between cancer cells and tumor associated macrophages is required for mesenchymal circulating tumor cell-mediated colorectal cancer metastasis. *Mol. Cancer* **18**, 64 (2019).
119. Yang, C., Dou, R., Wei, C., Liu, K., Shi, D., Zhang, C., Liu, Q., Wang, S. & Xiong, B. Tumor-derived exosomal microRNA-106b-5p activates EMT-cancer cell and M2-subtype TAM interaction to facilitate CRC metastasis. *Mol. Ther. J. Am. Soc. Gene Ther.* **29**, 2088–2107 (2021).
120. Borriello, L., Coste, A., Traub, B., Sharma, V. P., Karagiannis, G. S., Lin, Y., Wang, Y., Ye, X., Duran, C. L., Chen, X., Friedman, M., Sosa, M. S., Sun, D., Dalla, E., Singh, D. K., Oktay, M. H., Aguirre-Ghiso, J. A., Condeelis, J. S. & Entenberg, D. Primary tumor associated macrophages activate programs of invasion and dormancy in disseminating tumor cells. *Nat. Commun.* **13**, 626 (2022).
121. Pignatelli, J., Bravo-Cordero, J. J., Roh-Johnson, M., Gandhi, S. J., Wang, Y., Chen, X., Eddy, R. J., Xue, A., Singer, R. H., Hodgson, L., Oktay, M. H. & Condeelis, J. S. Macrophage-dependent tumor cell transendothelial migration is mediated by Notch1/MenaINV-initiated invadopodium formation. *Sci. Rep.* **6**, 37874 (2016).
122. Hsu, D. S.-S., Wang, H.-J., Tai, S.-K., Chou, C.-H., Hsieh, C.-H., Chiu, P.-H., Chen, N.-J. & Yang, M.-H. Acetylation of snail modulates the cytokinome of cancer cells to enhance the recruitment of macrophages. *Cancer Cell* **26**, 534–548 (2014).
123. Ding, J., Yang, C., Zhang, Y., Wang, J., Zhang, S., Guo, D., Yin, T. & Yang, J. M2 macrophage-derived G-CSF promotes trophoblasts EMT, invasion and migration via activating

PI3K/Akt/Erk1/2 pathway to mediate normal pregnancy. *J. Cell. Mol. Med.* **25**, 2136–2147 (2021).

124. Casanova-Acebes, M., Dalla, E., Leader, A. M., LeBerichel, J., Nikolic, J., Morales, B. M., Brown, M., Chang, C., Troncoso, L., Chen, S. T., Sastre-Perona, A., Park, M. D., Tabachnikova, A., Dhainaut, M., Hamon, P., Maier, B., Sawai, C. M., Agulló-Pascual, E., Schober, M., Brown, B. D., Reizis, B., Marron, T., Kenigsberg, E., Moussion, C., Benaroch, P., Aguirre-Ghiso, J. A. & Merad, M. Tissue-resident macrophages provide a pro-tumorigenic niche to early NSCLC cells. *Nature* **595**, 578–584 (2021).
125. Hara, T., Chanoch-Myers, R., Mathewson, N. D., Myskiw, C., Atta, L., Bussema, L., Eichhorn, S. W., Greenwald, A. C., Kinker, G. S., Rodman, C., Gonzalez Castro, L. N., Wakimoto, H., Rozenblatt-Rosen, O., Zhuang, X., Fan, J., Hunter, T., Verma, I. M., Wucherpfennig, K. W., Regev, A., Suvà, M. L. & Tirosh, I. Interactions between cancer cells and immune cells drive transitions to mesenchymal-like states in glioblastoma. *Cancer Cell* **39**, 779-792.e11 (2021).
126. Martins-Cardoso, K., Almeida, V. H., Bagri, K. M., Rossi, M. I. D., Mermelstein, C. S., König, S. & Monteiro, R. Q. Neutrophil Extracellular Traps (NETs) Promote Pro-Metastatic Phenotype in Human Breast Cancer Cells through Epithelial-Mesenchymal Transition. *Cancers* **12**, 1542 (2020).
127. Stehr, A. M., Wang, G., Demmler, R., Stemmler, M. P., Krug, J., Tripal, P., Schmid, B., Geppert, C. I., Hartmann, A., Muñoz, L. E., Schoen, J., Völkl, S., Merkel, S., Becker, C., Schett, G., Grützmann, R., Naschberger, E., Herrmann, M. & Stürzl, M. Neutrophil extracellular traps drive epithelial-mesenchymal transition of human colon cancer. *J. Pathol.* **256**, 455–467 (2022).
128. Adrover, J. M., Han, X., Sun, L., Fujii, T., Sivetz, N., Daßler-Plenker, J., Evans, C., Peters, J., He, X.-Y., Cannon, C. D., Ho, W. J., Raptis, G., Powers, R. S. & Egeblad, M. Neutrophils drive vascular occlusion, tumour necrosis and metastasis. *Nature* https://doi.org/10.1038/s41586-025-09278-3 (2025) doi:10.1038/s41586-025-09278-3.
129. He, X.-Y., Gao, Y., Ng, D., Michalopoulou, E., George, S., Adrover, J. M., Sun, L., Albrengues, J., Daßler-Plenker, J., Han, X., Wan, L., Wu, X. S., Shui, L. S., Huang, Y.-H., Liu, B., Su, C., Spector, D. L., Vakoc, C. R., Van Aelst, L. & Egeblad, M. Chronic stress increases metastasis via neutrophil-mediated changes to the microenvironment. *Cancer Cell* **42**, 474-486.e12 (2024).
130. McDowell, S. A. C., Luo, R. B. E., Arabzadeh, A., Doré, S., Bennett, N. C., Breton, V., Karimi, E., Rezanejad, M., Yang, R. R., Lach, K. D., Issac, M. S. M., Samborska, B., Perus, L. J. M., Moldoveanu, D., Wei, Y., Fiset, B., Rayes, R. F., Watson, I. R., Kazak, L., Guiot, M.-C., Fiset, P. O., Spicer, J. D., Dannenberg, A. J., Walsh, L. A. & Quail, D. F. Neutrophil oxidative stress mediates obesity-associated vascular dysfunction and metastatic transmigration. *Nat. Cancer* **2**, 545–562 (2021).
131. He, D., Wu, Q., Tian, P., Liu, Y., Jia, Z., Li, Z., Wang, Y., Jin, Y., Luo, W., Li, L., Zhang, P., Jin, Q., Zhao, W., Hu, W., Liang, Y., Zhou, B., Yang, Q., Jiang, Y.-Z., Shao, Z.-M. & Hu, G. Chemotherapy awakens dormant cancer cells in lung by inducing neutrophil extracellular traps. *Cancer Cell* S1535-6108(25)00257–0 (2025) doi:10.1016/j.ccell.2025.06.007.
132. Nolan, E., Bridgeman, V. L., Ombrato, L., Karoutas, A., Rabas, N., Sewnath, C. A. N., Vasquez, M., Rodrigues, F. S., Horswell, S., Faull, P., Carter, R. & Malanchi, I. Radiation exposure elicits a neutrophil-driven response in healthy lung tissue that enhances metastatic colonization. *Nat. Cancer* **3**, 173–187 (2022).

133. Chemi, F., Rothwell, D. G., McGranahan, N., Gulati, S., Abbosh, C., Pearce, S. P., Zhou, C., Wilson, G. A., Jamal-Hanjani, M., Birkbak, N., Pierce, J., Kim, C. S., Ferdous, S., Burt, D. J., Slane-Tan, D., Gomes, F., Moore, D., Shah, R., Al Bakir, M., Hiley, C., Veeriah, S., Summers, Y., Crosbie, P., Ward, S., Mesquita, B., Dynowski, M., Biswas, D., Tugwood, J., Blackhall, F., Miller, C., Hackshaw, A., Brady, G., Swanton, C., Dive, C., & TRACERx Consortium. Publisher Correction: Pulmonary venous circulating tumor cell dissemination before tumor resection and disease relapse. *Nat. Med.* **26**, 1147 (2020).
134. Baccelli, I., Schneeweiss, A., Riethdorf, S., Stenzinger, A., Schillert, A., Vogel, V., Klein, C., Saini, M., Bäuerle, T., Wallwiener, M., Holland-Letz, T., Höfner, T., Sprick, M., Scharpff, M., Marmé, F., Sinn, H. P., Pantel, K., Weichert, W. & Trumpp, A. Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat. Biotechnol.* **31**, 539–544 (2013).
135. Jordan, N. V., Bardia, A., Wittner, B. S., Benes, C., Ligorio, M., Zheng, Y., Yu, M., Sundaresan, T. K., Licausi, J. A., Desai, R., O'Keefe, R. M., Ebright, R. Y., Boukhali, M., Sil, S., Onozato, M. L., Iafrate, A. J., Kapur, R., Sgroi, D., Ting, D. T., Toner, M., Ramaswamy, S., Haas, W., Maheswaran, S. & Haber, D. A. HER2 expression identifies dynamic functional states within circulating breast cancer cells. *Nature* **537**, 102–106 (2016).
136. Aceto, N., Bardia, A., Miyamoto, D. T., Donaldson, M. C., Wittner, B. S., Spencer, J. A., Yu, M., Pely, A., Engstrom, A., Zhu, H., Brannigan, B. W., Kapur, R., Stott, S. L., Shioda, T., Ramaswamy, S., Ting, D. T., Lin, C. P., Toner, M., Haber, D. A. & Maheswaran, S. Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110–1122 (2014).
137. Cheung, K. J., Padmanaban, V., Silvestri, V., Schipper, K., Cohen, J. D., Fairchild, A. N., Gorin, M. A., Verdone, J. E., Pienta, K. J., Bader, J. S. & Ewald, A. J. Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E854-863 (2016).
138. Labuschagne, C. F., Cheung, E. C., Blagih, J., Domart, M.-C. & Vousden, K. H. Cell Clustering Promotes a Metabolic Switch that Supports Metastatic Colonization. *Cell Metab.* **30**, 720-734.e5 (2019).
139. Piskounova, E., Agathocleous, M., Murphy, M. M., Hu, Z., Huddlestun, S. E., Zhao, Z., Leitch, A. M., Johnson, T. M., DeBerardinis, R. J. & Morrison, S. J. Oxidative stress inhibits distant metastasis by human melanoma cells. *Nature* **527**, 186–191 (2015).
140. Padmanaban, V., Krol, I., Suhail, Y., Szczerba, B. M., Aceto, N., Bader, J. S. & Ewald, A. J. E-cadherin is required for metastasis in multiple models of breast cancer. *Nature* **573**, 439–444 (2019).
141. Szczerba, B. M., Castro-Giner, F., Vetter, M., Krol, I., Gkountela, S., Landin, J., Scheidmann, M. C., Donato, C., Scherrer, R., Singer, J., Beisel, C., Kurzeder, C., Heinzelmann-Schwarz, V., Rochlitz, C., Weber, W. P., Beerenwinkel, N. & Aceto, N. Neutrophils escort circulating tumour cells to enable cell cycle progression. *Nature* **566**, 553–557 (2019).
142. Labelle, M., Begum, S. & Hynes, R. O. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. *Cancer Cell* **20**, 576–590 (2011).
143. Schumacher, D., Strilic, B., Sivaraj, K. K., Wettschureck, N. & Offermanns, S. Platelet-derived nucleotides promote tumor-cell transendothelial migration and metastasis via P2Y2 receptor. *Cancer Cell* **24**, 130–137 (2013).
144. Armstrong, A. J., Marengo, M. S., Oltean, S., Kemeny, G., Bitting, R. L., Turnbull, J. D., Herold, C. I., Marcom, P. K., George, D. J. & Garcia-Blanco, M. A. Circulating tumor cells

from patients with advanced prostate and breast cancer display both epithelial and mesenchymal markers. *Mol. Cancer Res. MCR* **9**, 997–1007 (2011).
145. Lecharpentier, A., Vielh, P., Perez-Moreno, P., Planchard, D., Soria, J. C. & Farace, F. Detection of circulating tumour cells with a hybrid (epithelial/mesenchymal) phenotype in patients with metastatic non-small cell lung cancer. *Br. J. Cancer* **105**, 1338–1341 (2011).
146. Yu, M., Bardia, A., Wittner, B. S., Stott, S. L., Smas, M. E., Ting, D. T., Isakoff, S. J., Ciciliano, J. C., Wells, M. N., Shah, A. M., Concannon, K. F., Donaldson, M. C., Sequist, L. V., Brachtel, E., Sgroi, D., Baselga, J., Ramaswamy, S., Toner, M., Haber, D. A. & Maheswaran, S. Circulating breast tumor cells exhibit dynamic changes in epithelial and mesenchymal composition. *Science* **339**, 580–584 (2013).
147. Balcik-Ercin, P., Cayrefourcq, L., Soundararajan, R., Mani, S. A. & Alix-Panabières, C. Epithelial-to-Mesenchymal Plasticity in Circulating Tumor Cell Lines Sequentially Derived from a Patient with Colorectal Cancer. *Cancers* **13**, 5408 (2021).
148. Giordano, A., Gao, H., Anfossi, S., Cohen, E., Mego, M., Lee, B.-N., Tin, S., De Laurentiis, M., Parker, C. A., Alvarez, R. H., Valero, V., Ueno, N. T., De Placido, S., Mani, S. A., Esteva, F. J., Cristofanilli, M. & Reuben, J. M. Epithelial-mesenchymal transition and stem cell markers in patients with HER2-positive metastatic breast cancer. *Mol. Cancer Ther.* **11**, 2526–2534 (2012).
149. Coumans, F. & Terstappen, L. Detection and Characterization of Circulating Tumor Cells by the CellSearch Approach. *Methods Mol. Biol. Clifton NJ* **1347**, 263–278 (2015).
150. Miller, M. C., Robinson, P. S., Wagner, C. & O'Shannessy, D. J. The Parsortix™ Cell Separation System—A versatile liquid biopsy platform. *Cytometry* **93**, 1234–1239 (2018).
151. Vandenbussche, N., Imschoot, R., Lintermans, B., Denolf, L., Taminau, J., Fieuws, C., Berx, G., Gevaert, K. & Claes, K. B. M. EMT-associated bias in the Parsortix® system observed with pancreatic cancer cell lines. *Mol. Oncol.* https://doi.org/10.1002/1878-0261.70066 (2025) doi:10.1002/1878-0261.70066.
152. Rahrmann, E. P., Shorthouse, D., Jassim, A., Hu, L. P., Ortiz, M., Mahler-Araujo, B., Vogel, P., Paez-Ribes, M., Fatemi, A., Hannon, G. J., Iyer, R., Blundon, J. A., Lourenço, F. C., Kay, J., Nazarian, R. M., Hall, B. A., Zakharenko, S. S., Winton, D. J., Zhu, L. & Gilbertson, R. J. The NALCN channel regulates metastasis and nonmalignant cell dissemination. *Nat. Genet.* **54**, 1827–1838 (2022).
153. Diamantopoulou, Z., Castro-Giner, F., Schwab, F. D., Foerster, C., Saini, M., Budinjas, S., Strittmatter, K., Krol, I., Seifert, B., Heinzelmann-Schwarz, V., Kurzeder, C., Rochlitz, C., Vetter, M., Weber, W. P. & Aceto, N. The metastatic spread of breast cancer accelerates during sleep. *Nature* **607**, 156–162 (2022).
154. Schernhammer, E. S., Laden, F., Speizer, F. E., Willett, W. C., Hunter, D. J., Kawachi, I. & Colditz, G. A. Rotating night shifts and risk of breast cancer in women participating in the nurses' health study. *J. Natl. Cancer Inst.* **93**, 1563–1568 (2001).
155. Schernhammer, E. S., Laden, F., Speizer, F. E., Willett, W. C., Hunter, D. J., Kawachi, I., Fuchs, C. S. & Colditz, G. A. Night-shift work and risk of colorectal cancer in the nurses' health study. *J. Natl. Cancer Inst.* **95**, 825–828 (2003).
156. Straif, K., Baan, R., Grosse, Y., Secretan, B., Ghissassi, F. E., Bouvard, V., Altieri, A., Benbrahim-Tallaa, L. & Cogliano, V. Carcinogenicity of shift-work, painting, and fire-fighting. *Lancet Oncol.* **8**, 1065–1066 (2007).
157. Lévi, F. Circadian chronotherapy for human cancers. *Lancet Oncol.* **2**, 307–315 (2001).
158. Cristofanilli, M., Budd, G. T., Ellis, M. J., Stopeck, A., Matera, J., Miller, M. C., Reuben, J. M., Doyle, G. V., Allard, W. J., Terstappen, L. W. M. M. & Hayes, D. F. Circulating tumor cells,

disease progression, and survival in metastatic breast cancer. *N. Engl. J. Med.* **351**, 781–791 (2004).

159. Hou, J.-M., Krebs, M. G., Lancashire, L., Sloane, R., Backen, A., Swain, R. K., Priest, L. J. C., Greystoke, A., Zhou, C., Morris, K., Ward, T., Blackhall, F. H. & Dive, C. Clinical significance and molecular characteristics of circulating tumor cells and circulating tumor microemboli in patients with small-cell lung cancer. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **30**, 525–532 (2012).
160. Smerage, J. B., Barlow, W. E., Hortobagyi, G. N., Winer, E. P., Leyland-Jones, B., Srkalovic, G., Tejwani, S., Schott, A. F., O'Rourke, M. A., Lew, D. L., Doyle, G. V., Gralow, J. R., Livingston, R. B. & Hayes, D. F. Circulating tumor cells and response to chemotherapy in metastatic breast cancer: SWOG S0500. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **32**, 3483–3489 (2014).
161. Wallwiener, M., Riethdorf, S., Hartkopf, A. D., Modugno, C., Nees, J., Madhavan, D., Sprick, M. R., Schott, S., Domschke, C., Baccelli, I., Schönfisch, B., Burwinkel, B., Marmé, F., Heil, J., Sohn, C., Pantel, K., Trumpp, A. & Schneeweiss, A. Serial enumeration of circulating tumor cells predicts treatment response and prognosis in metastatic breast cancer: a prospective study in 393 patients. *BMC Cancer* **14**, 512 (2014).
162. Larsson, A.-M., Jansson, S., Bendahl, P.-O., Levin Tykjaer Jörgensen, C., Loman, N., Graffman, C., Lundgren, L., Aaltonen, K. & Rydén, L. Longitudinal enumeration and cluster evaluation of circulating tumor cells improve prognostication for patients with newly diagnosed metastatic breast cancer in a prospective observational trial. *Breast Cancer Res. BCR* **20**, 48 (2018).
163. Bidard, F.-C., Jacot, W., Kiavue, N., Dureau, S., Kadi, A., Brain, E., Bachelot, T., Bourgeois, H., Gonçalves, A., Ladoire, S., Naman, H., Dalenc, F., Gligorov, J., Espié, M., Emile, G., Ferrero, J.-M., Loirat, D., Frank, S., Cabel, L., Diéras, V., Cayrefourcq, L., Simondi, C., Berger, F., Alix-Panabières, C. & Pierga, J.-Y. Efficacy of Circulating Tumor Cell Count–Driven vs Clinician-Driven First-line Therapy Choice in Hormone Receptor–Positive, ERBB2-Negative Metastatic Breast Cancer. *JAMA Oncol.* **7**, 1–8 (2021).
164. Cabel, L., Berger, F., Cottu, P., Loirat, D., Rampanou, A., Brain, E., Cyrille, S., Bourgeois, H., Kiavue, N., Deluche, E., Ladoire, S., Campone, M., Pierga, J.-Y. & Bidard, F.-C. Clinical utility of circulating tumour cell-based monitoring of late-line chemotherapy for metastatic breast cancer: the randomised CirCe01 trial. *Br. J. Cancer* **124**, 1207–1213 (2021).
165. Antonarakis, E. S., Lu, C., Wang, H., Luber, B., Nakazawa, M., Roeser, J. C., Chen, Y., Mohammad, T. A., Chen, Y., Fedor, H. L., Lotan, T. L., Zheng, Q., Marzo, A. M. D., Isaacs, J. T., Isaacs, W. B., Nadal, R., Paller, C. J., Denmeade, S. R., Carducci, M. A., Eisenberger, M. A. & Luo, J. AR-V7 and Resistance to Enzalutamide and Abiraterone in Prostate Cancer. *N. Engl. J. Med.* **371**, 1028–1038 (2014).
166. Belderbos, B. P. S., Sieuwerts, A. M., Hoop, E. O., Mostert, B., Kraan, J., Hamberg, P., Van, M. N., Beaufort, C. M., Onstenk, W., van Soest, R. J., Martens, J., Sleijfer, S., de Wit, R., Mathijssen, R. H. J. & Lolkema, M. P. Associations between AR-V7 status in circulating tumour cells, circulating tumour cell count and survival in men with metastatic castration-resistant prostate cancer. *Eur. J. Cancer Oxf. Engl. 1990* **121**, 48–54 (2019).
167. Tan, P. H., Ellis, I., Allison, K., Brogi, E., Fox, S. B., Lakhani, S., Lazar, A. J., Morris, E. A., Sahin, A., Salgado, R., Sapino, A., Sasano, H., Schnitt, S., Sotiriou, C., van Diest, P., White, V. A., Lokuhetty, D., Cree, I. A., & WHO Classification of Tumours Editorial Board. The 2019 World Health Organization classification of tumours of the breast. *Histopathology* **77**, 181–185 (2020).

168. Costa, C., Muinelo-Romay, L., Cebey-López, V., Pereira-Veiga, T., Martínez-Pena, I., Abreu, M., Abalo, A., Lago-Lestón, R. M., Abuín, C., Palacios, P., Cueva, J., Piñeiro, R. & López-López, R. Analysis of a Real-World Cohort of Metastatic Breast Cancer Patients Shows Circulating Tumor Cell Clusters (CTC-clusters) as Predictors of Patient Outcomes. *Cancers* **12**, 1111 (2020).
169. Swanton, C., Bernard, E., Abbosh, C., André, F., Auwerx, J., Balmain, A., Bar-Sagi, D., Bernards, R., Bullman, S., DeGregori, J., Elliott, C., Erez, A., Evan, G., Febbraio, M. A., Hidalgo, A., Jamal-Hanjani, M., Joyce, J. A., Kaiser, M., Lamia, K., Locasale, J. W., Loi, S., Malanchi, I., Merad, M., Musgrave, K., Patel, K. J., Quezada, S., Wargo, J. A., Weeraratna, A., White, E., Winkler, F., Wood, J. N., Vousden, K. H. & Hanahan, D. Embracing cancer complexity: Hallmarks of systemic disease. *Cell* **187**, 1589–1616 (2024).
170. Pastushenko, I., Mauri, F., Song, Y., de Cock, F., Meeusen, B., Swedlund, B., Impens, F., Van Haver, D., Opitz, M., Thery, M., Bareche, Y., Lapouge, G., Vermeersch, M., Van Eycke, Y.-R., Balsat, C., Decaestecker, C., Sokolow, Y., Hassid, S., Perez-Bustillo, A., Agreda-Moreno, B., Rios-Buceta, L., Jaen, P., Redondo, P., Sieira-Gil, R., Millan-Cayetano, J. F., Sanmatrtin, O., D'Haene, N., Moers, V., Rozzi, M., Blondeau, J., Lemaire, S., Scozzaro, S., Janssens, V., De Troya, M., Dubois, C., Pérez-Morga, D., Salmon, I., Sotiriou, C., Helmbacher, F. & Blanpain, C. Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. *Nature* **589**, 448–455 (2021).
171. Cui, J., Zhang, C., Lee, J.-E., Bartholdy, B. A., Yang, D., Liu, Y., Erler, P., Galbo, P. M., Hodge, D. Q., Huangfu, D., Zheng, D., Ge, K. & Guo, W. MLL3 loss drives metastasis by promoting a hybrid epithelial-mesenchymal transition state. *Nat. Cell Biol.* **25**, 145–158 (2023).
172. Cañellas-Socias, A., Cortina, C., Hernando-Momblona, X., Palomo-Ponce, S., Mulholland, E. J., Turon, G., Mateo, L., Conti, S., Roman, O., Sevillano, M., Slebe, F., Stork, D., Caballé-Mestres, A., Berenguer-Llergo, A., Álvarez-Varela, A., Fenderico, N., Novellasdemunt, L., Jiménez-Gracia, L., Sipka, T., Bardia, L., Lorden, P., Colombelli, J., Heyn, H., Trepat, X., Tejpar, S., Sancho, E., Tauriello, D. V. F., Leedham, S., Attolini, C. S.-O. & Batlle, E. Metastatic recurrence in colorectal cancer arises from residual EMP1+ cells. *Nature* **611**, 603–613 (2022).
173. Ferreres, J. R., Vinyals, A., Campos-Martin, R., Espín, R., Podlipnik, S., Ramos, R., Bertran, E., Carrera, C., Marcoval, J., Malvehy, J., Fabregat, I., Puig, S. & Fabra, À. PRRX1 silencing is required for metastatic outgrowth in melanoma and is an independent prognostic of reduced survival in patients. *Mol. Oncol.* **18**, 2471–2494 (2024).
174. de Sousa e Melo, F., Kurtova, A. V., Harnoss, J. M., Kljavin, N., Hoeck, J. D., Hung, J., Anderson, J. E., Storm, E. E., Modrusan, Z., Koeppen, H., Dijkgraaf, G. J. P., Piskol, R. & de Sauvage, F. J. A distinct role for Lgr5+ stem cells in primary and metastatic colon cancer. *Nature* **543**, 676–680 (2017).
175. Fumagalli, A., Oost, K. C., Kester, L., Morgner, J., Bornes, L., Bruens, L., Spaargaren, L., Azkanaz, M., Schelfhorst, T., Beerling, E., Heinz, M. C., Postrach, D., Seinstra, D., Sieuwerts, A. M., Martens, J. W. M., van der Elst, S., van Baalen, M., Bhowmick, D., Vrisekoop, N., Ellenbroek, S. I. J., Suijkerbuijk, S. J. E., Snippert, H. J. & van Rheenen, J. Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer. *Cell Stem Cell* **26**, 569-578.e7 (2020).
176. Lüönd, F., Sugiyama, N., Bill, R., Bornes, L., Hager, C., Tang, F., Santacroce, N., Beisel, C., Ivanek, R., Bürglin, T., Tiede, S., van Rheenen, J. & Christofori, G. Distinct contributions of partial and full EMT to breast cancer malignancy. *Dev. Cell* **56**, 3203-3221.e11 (2021).

177. Simeonov, K. P., Byrns, C. N., Clark, M. L., Norgard, R. J., Martin, B., Stanger, B. Z., Shendure, J., McKenna, A. & Lengner, C. J. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell* **39**, 1150-1162.e9 (2021).
178. Lucas, O., Ward, S., Zaidi, R., Bunkum, A., Frankell, A. M., Moore, D. A., Hill, M. S., Liu, W. K., Marinelli, D., Lim, E. L., Hessey, S., Naceur-Lombardelli, C., Rowan, A., Purewal-Mann, S. K., Zhai, H., Dietzen, M., Ding, B., Royle, G., Aparicio, S., TRACERx Consortium, PEACE Consortium, McGranahan, N., Jamal-Hanjani, M., Kanu, N., Swanton, C. & Zaccaria, S. Characterizing the evolutionary dynamics of cancer proliferation in single-cell clones with SPRINTER. *Nat. Genet.* **57**, 103–114 (2025).
179. Mejlvang, J., Kriajevska, M., Vandewalle, C., Chernova, T., Sayan, A. E., Berx, G., Mellon, J. K. & Tulchinsky, E. Direct repression of cyclin D1 by SIP1 attenuates cell cycle progression in cells undergoing an epithelial mesenchymal transition. *Mol. Biol. Cell* **18**, 4615–4624 (2007).
180. Fischer, K. R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S. T. C., Choi, H., El Rayes, T., Ryu, S., Troeger, J., Schwabe, R. F., Vahdat, L. T., Altorki, N. K., Mittal, V. & Gao, D. Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. *Nature* **527**, 472–476 (2015).
181. Zheng, X., Carstens, J. L., Kim, J., Scheible, M., Kaye, J., Sugimoto, H., Wu, C.-C., LeBleu, V. S. & Kalluri, R. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. *Nature* **527**, 525–530 (2015).
182. Ge, Y., Gomez, N. C., Adam, R. C., Nikolova, M., Yang, H., Verma, A., Lu, C. P.-J., Polak, L., Yuan, S., Elemento, O. & Fuchs, E. Stem Cell Lineage Infidelity Drives Wound Repair and Cancer. *Cell* **169**, 636-650.e14 (2017).
183. Marjanovic, N. D., Hofree, M., Chan, J. E., Canner, D., Wu, K., Trakala, M., Hartmann, G. G., Smith, O. C., Kim, J. Y., Evans, K. V., Hudson, A., Ashenberg, O., Porter, C. B. M., Bejnood, A., Subramanian, A., Pitter, K., Yan, Y., Delorey, T., Phillips, D. R., Shah, N., Chaudhary, O., Tsankov, A., Hollmann, T., Rekhtman, N., Massion, P. P., Poirier, J. T., Mazutis, L., Li, R., Lee, J.-H., Amon, A., Rudin, C. M., Jacks, T., Regev, A. & Tammela, T. Emergence of a High-Plasticity Cell State during Lung Cancer Evolution. *Cancer Cell* **38**, 229-246.e13 (2020).
184. Tri Giang Phan, Phan, T. G., Peter I. Croucher & Croucher, P. I. The dormant cancer cell life cycle. *Nat. Rev. Cancer* **20**, 398–411 (2020).
185. Pedersen, R. N., Esen, B. Ö., Mellemkjær, L., Christiansen, P., Ejlertsen, B., Lash, T. L., Nørgaard, M. & Cronin-Fenton, D. The Incidence of Breast Cancer Recurrence 10-32 Years After Primary Diagnosis. *J. Natl. Cancer Inst.* **114**, 391–399 (2022).
186. Pan, H., Gray, R., Braybrooke, J., Davies, C., Taylor, C., McGale, P., Peto, R., Pritchard, K. I., Bergh, J., Dowsett, M., Hayes, D. F., & EBCTCG. 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. *N. Engl. J. Med.* **377**, 1836–1846 (2017).
187. Groot, V. P., Rezaee, N., Wu, W., Cameron, J. L., Fishman, E. K., Hruban, R. H., Weiss, M. J., Zheng, L., Wolfgang, C. L. & He, J. Patterns, Timing, and Predictors of Recurrence Following Pancreatectomy for Pancreatic Ductal Adenocarcinoma. *Ann. Surg.* **267**, 936–945 (2018).
188. Van den Broeck, A., Sergeant, G., Ectors, N., Van Steenbergen, W., Aerts, R. & Topal, B. Patterns of recurrence after curative resection of pancreatic ductal adenocarcinoma. *Eur. J. Surg. Oncol. J. Eur. Soc. Surg. Oncol. Br. Assoc. Surg. Oncol.* **35**, 600–604 (2009).
189. Harper, K. L., Sosa, M. S., Entenberg, D., Hosseini, H., Cheung, J. F., Nobre, R., Avivar-Valderas, A., Nagi, C., Girnius, N., Davis, R. J., Farias, E. F., Condeelis, J., Klein, C. A. & Aguirre-Ghiso, J. A. Mechanism of early dissemination and metastasis in Her2+ mammary cancer. *Nature* **540**, 588–592 (2016).

190. Hosseini, H., Obradović, M. M. S., Hoffmann, M., Harper, K. L., Sosa, M. S., Werner-Klein, M., Nanduri, L. K., Werno, C., Ehrl, C., Maneck, M., Patwary, N., Haunschild, G., Gužvić, M., Reimelt, C., Grauvogl, M., Eichner, N., Weber, F., Hartkopf, A. D., Taran, F.-A., Brucker, S. Y., Fehm, T., Rack, B., Buchholz, S., Spang, R., Meister, G., Aguirre-Ghiso, J. A. & Klein, C. A. Early dissemination seeds metastasis in breast cancer. *Nature* **540**, 552–558 (2016).
191. Pantel, K., Schlimok, G., Braun, S., Kutter, D., Lindemann, F., Schaller, G., Funke, I., Izbicki, J. R. & Riethmüller, G. Differential expression of proliferation-associated molecules in individual micrometastatic carcinoma cells. *J. Natl. Cancer Inst.* **85**, 1419–1424 (1993).
192. Sosa, M. S., Parikh, F., Maia, A. G., Estrada, Y., Bosch, A., Bragado, P., Ekpin, E., George, A., Zheng, Y., Lam, H.-M., Morrissey, C., Chung, C.-Y., Farias, E. F., Bernstein, E. & Aguirre-Ghiso, J. A. NR2F1 controls tumour cell dormancy via SOX9- and RARβ-driven quiescence programmes. *Nat. Commun.* **6**, 6170 (2015).
193. Bragado, P., Estrada, Y., Parikh, F., Krause, S., Capobianco, C., Farina, H. G., Schewe, D. M. & Aguirre-Ghiso, J. A. TGF-β2 dictates disseminated tumour cell fate in target organs through TGF-β-RIII and p38α/β signalling. *Nat. Cell Biol.* **15**, 1351–1361 (2013).
194. Ren, Q., Khoo, W. H., Corr, A. P., Phan, T. G., Croucher, P. I. & Stewart, S. A. Gene expression predicts dormant metastatic breast cancer cell phenotype. *Breast Cancer Res.* **24**, 10 (2022).
195. Agudo, J., Aguirre-Ghiso, J. A., Bhatia, M., Chodosh, L. A., Correia, A. L. & Klein, C. A. Targeting cancer cell dormancy. *Nat. Rev. Cancer* **24**, 97–104 (2024).
196. Nobre, A. R., Dalla, E., Yang, J., Huang, X., Wullkopf, L., Risson, E., Razghandi, P., Anton, M. L., Zheng, W., Seoane, J. A., Curtis, C., Kenigsberg, E., Wang, J. & Aguirre-Ghiso, J. A. ZFP281 drives a mesenchymal-like dormancy program in early disseminated breast cancer cells that prevents metastatic outgrowth in the lung. *Nat. Cancer* **3**, 1165–1180 (2022).
197. Jakab, M., Lee, K. H., Uvarovskii, A., Ovchinnikova, S., Kulkarni, S. R., Jakab, S., Rostalski, T., Spegg, C., Anders, S. & Augustin, H. G. Lung endothelium exploits susceptible tumor cell states to instruct metastatic latency. *Nat. Cancer* **5**, 716–730 (2024).
198. Esposito, M., Mondal, N., Greco, T. M., Wei, Y., Spadazzi, C., Lin, S.-C., Zheng, H., Cheung, C., Magnani, J. L., Lin, S.-H., Cristea, I. M., Sackstein, R. & Kang, Y. Bone vascular niche E-selectin induces mesenchymal-epithelial transition and Wnt activation in cancer cells to promote bone metastasis. *Nat. Cell Biol.* **21**, 627–639 (2019).
199. Ghajar, C. M., Peinado, H., Mori, H., Matei, I. R., Evason, K. J., Brazier, H., Almeida, D., Koller, A., Hajjar, K. A., Stainier, D. Y. R., Chen, E. I., Lyden, D. & Bissell, M. J. The perivascular niche regulates breast tumour dormancy. *Nat. Cell Biol.* **15**, 807–817 (2013).
200. Ghajar, C. M. Metastasis prevention by targeting the dormant niche. *Nat. Rev. Cancer* **15**, 238–247 (2015).
201. Dalla, E., Papanicolaou, M., Park, M. D., Barth, N., Hou, R., Segura-Villalobos, D., Salazar, L. V., Sun, D., Forrest, A. R. R., Casanova-Acebes, M., Entenberg, D., Merad, M. & Aguirre-Ghiso, J. A. Lung-resident alveolar macrophages regulate the timing of breast cancer metastasis. *Cell* **187**, 6631-6648.e20 (2024).
202. Gao, D., Joshi, N., Choi, H., Ryu, S., Hahn, M., Catena, R., Sadik, H., Argani, P., Wagner, P., Vahdat, L. T., Port, J. L., Stiles, B., Sukumar, S., Altorki, N. K., Rafii, S. & Mittal, V. Myeloid progenitor cells in the premetastatic lung promote metastases by inducing mesenchymal to epithelial transition. *Cancer Res.* **72**, 1384–1394 (2012).

203. Di Martino, J. S., Nobre, A. R., Mondal, C., Taha, I., Farias, E. F., Fertig, E. J., Naba, A., Aguirre-Ghiso, J. A. & Bravo-Cordero, J. J. A tumor-derived type III collagen-rich ECM niche regulates tumor cell dormancy. *Nat. Cancer* **3**, 90–107 (2022).
204. Yin, W., Zhu, H., Tan, J., Xin, Z., Zhou, Q., Cao, Y., Wu, Z., Wang, L., Zhao, M., Jiang, X., Ren, C. & Tang, G. Identification of collagen genes related to immune infiltration and epithelial-mesenchymal transition in glioma. *Cancer Cell Int.* **21**, 276 (2021).
205. Montagner, M., Bhome, R., Hooper, S., Chakravarty, P., Qin, X., Sufi, J., Bhargava, A., Ratcliffe, C. D. H., Naito, Y., Pocaterra, A., Tape, C. J. & Sahai, E. Crosstalk with lung epithelial cells regulates Sfrp2-mediated latency in breast cancer dissemination. *Nat. Cell Biol.* **22**, 289–296 (2020).
206. Rodrigues, F. S., Karoutas, A., Ruhland, S., Rabas, N., Rizou, T., Di Blasio, S., Ferreira, R. M. M., Bridgeman, V. L., Goldstone, R., Sopena, M. L., Lee, J.-H., Ombrato, L. & Malanchi, I. Bidirectional activation of stem-like programs between metastatic cancer and alveolar type 2 cells within the niche. *Dev. Cell* **59**, 2398-2413.e8 (2024).
207. Borrelli, C., Roberts, M., Eletto, D., Hussherr, M.-D., Fazilaty, H., Valenta, T., Lafzi, A., Kretz, J. A., Guido Vinzoni, E., Karakatsani, A., Adivarahan, S., Mannhart, A., Kimura, S., Meijs, A., Baccouche Mhamedi, F., Acar, I. E., Handler, K., Ficht, X., Platt, R. J., Piscuoglio, S. & Moor, A. E. In vivo interaction screening reveals liver-derived constraints to metastasis. *Nature* **632**, 411–418 (2024).
208. Binan, L., Jiang, A., Danquah, S. A., Valakh, V., Simonton, B., Bezney, J., Manguso, R. T., Yates, K. B., Nehme, R., Cleary, B. & Farhi, S. L. Simultaneous CRISPR screening and spatial transcriptomics reveal intracellular, intercellular, and functional transcriptional circuits. *Cell* **188**, 2141-2158.e18 (2025).
209. Hugo, W., Zaretsky, J. M., Sun, L., Song, C., Moreno, B. H., Hu-Lieskovan, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., Seja, E., Lomeli, S., Kong, X., Kelley, M. C., Sosman, J. A., Johnson, D. B., Ribas, A. & Lo, R. S. Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **165**, 35–44 (2016).
210. Wang, G., Xu, D., Zhang, Z., Li, X., Shi, J., Sun, J., Liu, H.-Z., Li, X., Zhou, M. & Zheng, T. The pan-cancer landscape of crosstalk between epithelial-mesenchymal transition and immune evasion relevant to prognosis and immunotherapy response. *NPJ Precis. Oncol.* **5**, 56 (2021).
211. Noman, M. Z., Van Moer, K., Marani, V., Gemmill, R. M., Tranchevent, L.-C., Azuaje, F., Muller, A., Chouaib, S., Thiery, J. P., Berchem, G. & Janji, B. CD47 is a direct target of SNAI1 and ZEB1 and its blockade activates the phagocytosis of breast cancer cells undergoing EMT. *Oncoimmunology* **7**, e1345415 (2018).
212. Dongre, A., Rashidian, M., Reinhardt, F., Bagnato, A., Keckesova, Z., Ploegh, H. L. & Weinberg, R. A. Epithelial-to-Mesenchymal Transition Contributes to Immunosuppression in Breast Carcinomas. *Cancer Res.* **77**, 3982–3989 (2017).
213. Guo, Y., Lu, X., Chen, Y., Rendon, B., Mitchell, R. A., Cuatrecasas, M., Cortés, M., Postigo, A., Liu, Y. & Dean, D. C. Zeb1 induces immune checkpoints to form an immunosuppressive envelope around invading cancer cells. *Sci. Adv.* **7**, eabd7455 (2021).
214. Tripathi, S. C., Peters, H. L., Taguchi, A., Katayama, H., Wang, H., Momin, A., Jolly, M. K., Celiktas, M., Rodriguez-Canales, J., Liu, H., Behrens, C., Wistuba, I. I., Ben-Jacob, E., Levine, H., Molldrem, J. J., Hanash, S. M. & Ostrin, E. J. Immunoproteasome deficiency is a feature of non-small cell lung cancer with a mesenchymal phenotype and is associated with a poor outcome. *Proc. Natl. Acad. Sci. U. S. A.* **113**, E1555-1564 (2016).

215. Pozniak, J., Pedri, D., Landeloos, E., Van Herck, Y., Antoranz, A., Vanwynsberghe, L., Nowosad, A., Roda, N., Makhzami, S., Bervoets, G., Maciel, L. F., Pulido-Vicuña, C. A., Pollaris, L., Seurinck, R., Zhao, F., Flem-Karlsen, K., Damsky, W., Chen, L., Karagianni, D., Cinque, S., Kint, S., Vandereyken, K., Rombaut, B., Voet, T., Vernaillen, F., Annaert, W., Lambrechts, D., Boecxstaens, V., Saeys, Y., van den Oord, J., Bosisio, F., Karras, P., Shain, A. H., Bosenberg, M., Leucci, E., Paschen, A., Rambow, F., Bechter, O. & Marine, J.-C. A TCF4-dependent gene regulatory network confers resistance to immunotherapy in melanoma. *Cell* **187**, 166-183.e25 (2024).
216. Le Floc'h, A., Jalil, A., Vergnon, I., Le Maux Chansac, B., Lazar, V., Bismuth, G., Chouaib, S. & Mami-Chouaib, F. Alpha E beta 7 integrin interaction with E-cadherin promotes antitumor CTL activity by triggering lytic granule polarization and exocytosis. *J. Exp. Med.* **204**, 559–570 (2007).
217. Akalay, I., Janji, B., Hasmim, M., Noman, M. Z., André, F., De Cremoux, P., Bertheau, P., Badoual, C., Vielh, P., Larsen, A. K., Sabbah, M., Tan, T. Z., Keira, J. H., Hung, N. T. Y., Thiery, J. P., Mami-Chouaib, F. & Chouaib, S. Epithelial-to-mesenchymal transition and autophagy induction in breast carcinoma promote escape from T-cell-mediated lysis. *Cancer Res.* **73**, 2418–2427 (2013).
218. Kudo-Saito, C., Shirako, H., Takeuchi, T. & Kawakami, Y. Cancer metastasis is accelerated through immunosuppression during Snail-induced EMT of cancer cells. *Cancer Cell* **15**, 195–206 (2009).
219. Rozalén, C., Sangrador, I., Avalle, S., Blasco-Benito, S., Tzortzi, P., Sanz-Flores, M., Palomeque, J. Á., Torren-Duran, P., Dalmau, M., Brunel, H., Coll-Manzano, A., Pérez-Núñez, I., Martos, T., Servitja, S., Pérez-Buira, S., Chacón, J. I., Guerrero-Zotano, Á., Martínez de Dueñas, E., Guillén, Y., Comerma, L., Bermejo, B., Bigas, A., Casanova-Acebes, M., Alemany, A., Rojo, F., Albanell, J. & Celià-Terrassa, T. TIM3+ breast cancer cells license immune evasion during micrometastasis outbreak. *Cancer Cell* S1535-6108(25)00265-X (2025) doi:10.1016/j.ccell.2025.06.015.
220. Fane, M. E., Chhabra, Y., Alicea, G. M., Maranto, D. A., Douglass, S. M., Webster, M. R., Rebecca, V. W., Marino, G. E., Almeida, F., Ecker, B. L., Zabransky, D. J., Hüser, L., Beer, T., Tang, H.-Y., Kossenkov, A., Herlyn, M., Speicher, D. W., Xu, W., Xu, X., Jaffee, E. M., Aguirre-Ghiso, J. A. & Weeraratna, A. T. Stromal changes in the aged lung induce an emergence from melanoma dormancy. *Nature* **606**, 396–405 (2022).
221. Krall, J. A., Reinhardt, F., Mercury, O. A., Pattabiraman, D. R., Brooks, M. W., Dougan, M., Lambert, A. W., Bierie, B., Ploegh, H. L., Dougan, S. K. & Weinberg, R. A. The systemic response to surgery triggers the outgrowth of distant immune-controlled tumors in mouse models of dormancy. *Sci. Transl. Med.* **10**, eaan3464 (2018).
222. Quail, D. F., Olson, O. C., Bhardwaj, P., Walsh, L. A., Akkari, L., Quick, M. L., Chen, I.-C., Wendel, N., Ben-Chetrit, N., Walker, J., Holt, P. R., Dannenberg, A. J. & Joyce, J. A. Obesity alters the lung myeloid cell landscape to enhance breast cancer metastasis through IL5 and GM-CSF. *Nat. Cell Biol.* **19**, 974–987 (2017).
223. Chia, S. B., Johnson, B. J., Hu, J., Valença-Pereira, F., Chadeau-Hyam, M., Guntoro, F., Montgomery, H., Boorgula, M. P., Sreekanth, V., Goodspeed, A., Davenport, B., De Dominici, M., Zaberezhnyy, V., Schleicher, W. E., Gao, D., Cadar, A. N., Petriz-Otaño, L., Papanicolaou, M., Beheshti, A., Baylin, S. B., Guarnieri, J. W., Wallace, D. C., Costello, J. C., Bartley, J. M., Morrison, T. E., Vermeulen, R., Aguirre-Ghiso, J. A., Rincon, M. & DeGregori, J. Respiratory viral infections awaken metastatic breast cancer cells in lungs.

Nature https://doi.org/10.1038/s41586-025-09332-0 (2025) doi:10.1038/s41586-025-09332-0.

224. Cheang, M. C. U., Chia, S. K., Voduc, D., Gao, D., Leung, S., Snider, J., Watson, M., Davies, S., Bernard, P. S., Parker, J. S., Perou, C. M., Ellis, M. J. & Nielsen, T. O. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J. Natl. Cancer Inst.* **101**, 736–750 (2009).
225. Abd El-Rehim, D. M., Ball, G., Pinder, S. E., Rakha, E., Paish, C., Robertson, J. F. R., Macmillan, D., Blamey, R. W. & Ellis, I. O. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. *Int. J. Cancer* **116**, 340–350 (2005).
226. Renier, N., Wu, Z., Simon, D. J., Yang, J., Ariel, P. & Tessier-Lavigne, M. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. *Cell* **159**, 896–910 (2014).
227. Dao, D., Fraser, A. N., Hung, J., Ljosa, V., Singh, S. & Carpenter, A. E. CellProfiler Analyst: interactive data exploration, analysis and classification of large biological image sets. *Bioinforma. Oxf. Engl.* **32**, 3210–3212 (2016).
228. Stirling, D. R., Swain-Bowden, M. J., Lucas, A. M., Carpenter, A. E., Cimini, B. A. & Goodman, A. CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 433 (2021).
229. Pachitariu, M. & Stringer, C. Cellpose 2.0: how to train your own model. *Nat. Methods* **19**, 1634–1641 (2022).
230. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).
231. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P. & Raychaudhuri, S. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
232. Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R. & Ma'ayan, A. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128 (2013).
233. Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga, J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., Bharadwaj, R., Wong, A., Ness, K. D., Beppu, L. W., Deeg, H. J., McFarland, C., Loeb, K. R., Valente, W. J., Ericson, N. G., Stevens, E. A., Radich, J. P., Mikkelsen, T. S., Hindson, B. J. & Bielas, J. H. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
234. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Research* **10**, 979 (2021).
235. Andreatta, M. & Carmona, S. J. UCell: Robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).
236. Crowell, H. L., Soneson, C., Germain, P.-L., Calini, D., Collin, L., Raposo, C., Malhotra, D. & Robinson, M. D. muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat. Commun.* **11**, 6077 (2020).
237. Hao, Y., Stuart, T., Kowalski, M. H., Choudhary, S., Hoffman, P., Hartman, A., Srivastava, A., Molla, G., Madad, S., Fernandez-Granda, C. & Satija, R. Dictionary learning for

integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).

238. Stuart, T., Srivastava, A., Madad, S., Lareau, C. A. & Satija, R. Single-cell chromatin state analysis with Signac. *Nat. Methods* **18**, 1333–1341 (2021).
239. Thibodeau, A., Eroglu, A., McGinnis, C. S., Lawlor, N., Nehar-Belaid, D., Kursawe, R., Marches, R., Conrad, D. N., Kuchel, G. A., Gartner, Z. J., Banchereau, J., Stitzel, M. L., Cicek, A. E. & Ucar, D. AMULET: a novel read count-based method for effective multiplet detection from single nucleus ATAC-seq data. *Genome Biol.* **22**, 252 (2021).
240. Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P. & Satija, R. Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
241. Rauluseviciute, I., Riudavets-Puig, R., Blanc-Mathieu, R., Castro-Mondragon, J. A., Ferenc, K., Kumar, V., Lemma, R. B., Lucas, J., Chèneby, J., Baranasic, D., Khan, A., Fornes, O., Gundersen, S., Johansen, M., Hovig, E., Lenhard, B., Sandelin, A., Wasserman, W. W., Parcy, F. & Mathelier, A. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **52**, D174–D182 (2024).
242. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
243. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinforma. Oxf. Engl.* **31**, 2382–2383 (2015).
244. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinforma. Oxf. Engl.* **27**, 1017–1018 (2011).
245. Taube, J. H., Herschkowitz, J. I., Komurov, K., Zhou, A. Y., Gupta, S., Yang, J., Hartwell, K., Onder, T. T., Gupta, P. B., Evans, K. W., Hollier, B. G., Ram, P. T., Lander, E. S., Rosen, J. M., Weinberg, R. A. & Mani, S. A. Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 15449–15454 (2010).
246. Klijn, C., Durinck, S., Stawiski, E. W., Haverty, P. M., Jiang, Z., Liu, H., Degenhardt, J., Mayba, O., Gnad, F., Liu, J., Pau, G., Reeder, J., Cao, Y., Mukhyala, K., Selvaraj, S. K., Yu, M., Zynda, G. J., Brauer, M. J., Wu, T. D., Gentleman, R. C., Manning, G., Yauch, R. L., Bourgon, R., Stokoe, D., Modrusan, Z., Neve, R. M., de Sauvage, F. J., Settleman, J., Seshagiri, S. & Zhang, Z. A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.* **33**, 306–312 (2015).
247. Rueda, O. M., Sammut, S.-J., Seoane, J. A., Chin, S.-F., Caswell-Jin, J. L., Callari, M., Batra, R., Pereira, B., Bruna, A., Ali, H. R., Provenzano, E., Liu, B., Parisien, M., Gillett, C., McKinney, S., Green, A. R., Murphy, L., Purushotham, A., Ellis, I. O., Pharoah, P. D., Rueda, C., Aparicio, S., Caldas, C. & Curtis, C. Dynamics of breast-cancer relapse reveal late-recurring ER-positive genomic subgroups. *Nature* **567**, 399–404 (2019).
248. 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., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., METABRIC Group, Langerød, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A.-L., Brenton, J. D., Tavaré, S., Caldas, C. & Aparicio, S. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).

249. Pereira, B., Chin, S.-F., Rueda, O. M., Vollan, H.-K. M., Provenzano, E., Bardwell, H. A., Pugh, M., Jones, L., Russell, R., Sammut, S.-J., Tsui, D. W. Y., Liu, B., Dawson, S.-J., Abraham, J., Northen, H., Peden, J. F., Mukherjee, A., Turashvili, G., Green, A. R., McKinney, S., Oloumi, A., Shah, S., Rosenfeld, N., Murphy, L., Bentley, D. R., Ellis, I. O., Purushotham, A., Pinder, S. E., Børresen-Dale, A.-L., Earl, H. M., Pharoah, P. D., Ross, M. T., Aparicio, S. & Caldas, C. The somatic mutation profiles of 2,433 breast cancers refine their genomic and transcriptomic landscapes. *Nat. Commun.* **7**, 11479 (2016).
250. Pommier, R. M., Sanlaville, A., Tonon, L., Kielbassa, J., Thomas, E., Ferrari, A., Sertier, A.-S., Hollande, F., Martinez, P., Tissier, A., Morel, A.-P., Ouzounova, M. & Puisieux, A. Comprehensive characterization of claudin-low breast tumors reflects the impact of the cell-of-origin on cancer evolution. *Nat. Commun.* **11**, 3431 (2020).
251. Khanna, C. & Hunter, K. Modeling metastasis in vivo. *Carcinogenesis* **26**, 513–523 (2005).
252. Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol. Cell. Biol.* **12**, 954–961 (1992).
253. Zhou, A. Y., Ichaso, N., Adamarek, A., Zila, V., Forstova, J., Dibb, N. J. & Dilworth, S. M. Polyomavirus Middle T-Antigen Is a Transmembrane Protein That Binds Signaling Proteins in Discrete Subcellular Membrane Sites. *J. Virol.* **85**, 3046–3054 (2011).
254. Otten, A. D., Sanders, M. M. & McKnight, G. S. The MMTV LTR promoter is induced by progesterone and dihydrotestosterone but not by estrogen. *Mol. Endocrinol. Baltim. Md* **2**, 143–147 (1988).
255. Attalla, S., Taifour, T., Bui, T. & Muller, W. Insights from transgenic mouse models of PyMT-induced breast cancer: recapitulating human breast cancer progression in vivo. *Oncogene* **40**, 475–491 (2021).
256. Koren, S., Reavie, L., Couto, J. P., De Silva, D., Stadler, M. B., Roloff, T., Britschgi, A., Eichlisberger, T., Kohler, H., Aina, O., Cardiff, R. D. & Bentires-Alj, M. PIK3CA(H1047R) induces multipotency and multi-lineage mammary tumours. *Nature* **525**, 114–118 (2015).
257. Van Keymeulen, A., Lee, M. Y., Ousset, M., Brohée, S., Rorive, S., Giraddi, R. R., Wuidart, A., Bouvencourt, G., Dubois, C., Salmon, I., Sotiriou, C., Phillips, W. A. & Blanpain, C. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* **525**, 119–123 (2015).
258. Wuidart, A., Sifrim, A., Fioramonti, M., Matsumura, S., Brisebarre, A., Brown, D., Centonze, A., Dannau, A., Dubois, C., Van Keymeulen, A., Voet, T. & Blanpain, C. Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat. Cell Biol.* **20**, 666–676 (2018).
259. Takano, S., Reichert, M., Bakir, B., Das, K. K., Nishida, T., Miyazaki, M., Heeg, S., Collins, M. A., Marchand, B., Hicks, P. D., Maitra, A. & Rustgi, A. K. Prrx1 isoform switching regulates pancreatic cancer invasion and metastatic colonization. *Genes Dev.* **30**, 233–247 (2016).
260. Cheung, K. J., Gabrielson, E., Werb, Z. & Ewald, A. J. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* **155**, 1639–1651 (2013).
261. Sugiyama, M., Hasegawa, H., Ito, S., Sugiyama, K., Maeda, M., Aoki, K., Wakabayashi, T., Hamaguchi, M., Natsume, A. & Senga, T. Paired related homeobox 1 is associated with the invasive properties of glioblastoma cells. *Oncol. Rep.* **33**, 1123–1130 (2015).
262. Calabrese, E. J. & Mattson, M. P. How does hormesis impact biology, toxicology, and medicine? *Npj Aging Mech. Dis.* **3**, 1–8 (2017).
263. Mattson, M. P. Hormesis defined. *Ageing Res. Rev.* **7**, 1–7 (2008).

264. Hamza, B., Miller, A. B., Meier, L., Stockslager, M., Ng, S. R., King, E. M., Lin, L., DeGouveia, K. L., Mulugeta, N., Calistri, N. L., Strouf, H., Bray, C., Rodriguez, F., Freed-Pastor, W. A., Chin, C. R., Jaramillo, G. C., Burger, M. L., Weinberg, R. A., Shalek, A. K., Jacks, T. & Manalis, S. R. Measuring kinetics and metastatic propensity of CTCs by blood exchange between mice. *Nat. Commun.* **12**, 5680 (2021).
265. Rao, A., Barkley, D., França, G. S. & Yanai, I. Exploring tissue architecture using spatial transcriptomics. *Nature* **596**, 211–220 (2021).
266. Nagasawa, S., Zenkoh, J., Suzuki, Y. & Suzuki, A. Spatial omics technologies for understanding molecular status associated with cancer progression. *Cancer Sci.* **115**, 3208–3217 (2024).
267. Xia, C., Fan, J., Emanuel, G., Hao, J. & Zhuang, X. Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 19490–19499 (2019).
268. Jiang, G., Tu, J., Zhou, L., Dong, M., Fan, J., Chang, Z., Zhang, L., Bian, X. & Liu, S. Single-cell transcriptomics reveal the heterogeneity and dynamic of cancer stem-like cells during breast tumor progression. *Cell Death Dis.* **12**, 1–13 (2021).
269. Valdés-Mora, F., Salomon, R., Gloss, B. S., Law, A. M. K., Venhuizen, J., Castillo, L., Murphy, K. J., Magenau, A., Papanicolaou, M., Rodriguez de la Fuente, L., Roden, D. L., Colino-Sanguino, Y., Kikhtyak, Z., Farbehi, N., Conway, J. R. W., Sikta, N., Oakes, S. R., Cox, T. R., O'Donoghue, S. I., Timpson, P., Ormandy, C. J. & Gallego-Ortega, D. Single-cell transcriptomics reveals involution mimicry during the specification of the basal breast cancer subtype. *Cell Rep.* **35**, 108945 (2021).
270. Vermeulen, K., Van Bockstaele, D. R. & Berneman, Z. N. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif.* **36**, 131–149 (2003).
271. Sarkar, R., Hunter, I. A., Rajaganeshan, R., Perry, S. L., Guillou, P. & Jayne, D. G. Expression of cyclin D2 is an independent predictor of the development of hepatic metastasis in colorectal cancer. *Colorectal Dis. Off. J. Assoc. Coloproctology G. B. Irel.* **12**, 316–323 (2010).
272. Takano, Y., Kato, Y., Masuda, M., Ohshima, Y. & Okayasu, I. Cyclin D2, but not cyclin D1, overexpression closely correlates with gastric cancer progression and prognosis. *J. Pathol.* **189**, 194–200 (1999).
273. Liu, S. C., Bassi, D. E., Zhang, S. Y., Holoran, D., Conti, C. J. & Klein-Szanto, A. J. P. Overexpression of cyclin D2 is associated with increased in vivo invasiveness of human squamous carcinoma cells. *Mol. Carcinog.* **34**, 131–139 (2002).
274. Wang, X., Zhao, S., Xin, Q., Zhang, Y., Wang, K. & Li, M. Recent progress of CDK4/6 inhibitors' current practice in breast cancer. *Cancer Gene Ther.* **31**, 1283–1291 (2024).
275. Xue, W., Yang, L., Chen, C., Ashrafizadeh, M., Tian, Y. & Sun, R. Wnt/β-catenin-driven EMT regulation in human cancers. *Cell. Mol. Life Sci. CMLS* **81**, 79 (2024).
276. Coffelt, S. B., Wellenstein, M. D. & de Visser, K. E. Neutrophils in cancer: neutral no more. *Nat. Rev. Cancer* **16**, 431–446 (2016).
277. Sherr, C. J. & Roberts, J. M. Inhibitors of mammalian G1 cyclin-dependent kinases. *Genes Dev.* **9**, 1149–1163 (1995).
278. Jee, J., Fong, C., Pichotta, K., Tran, T. N., Luthra, A., Waters, M., Fu, C., Altoe, M., Liu, S.-Y., Maron, S. B., Ahmed, M., Kim, S., Pirun, M., Chatila, W. K., de Bruijn, I., Pasha, A., Kundra, R., Gross, B., Mastrogiacomo, B., Aprati, T. J., Liu, D., Gao, J., Capelletti, M., Pekala, K., Loudon, L., Perry, M., Bandlamudi, C., Donoghue, M., Satravada, B. A., Martin, A., Shen, R., Chen, Y., Brannon, A. R., Chang, J., Braunstein, L., Li, A., Safonov, A., Stonestrom, A.,

Sanchez-Vela, P., Wilhelm, C., Robson, M., Scher, H., Ladanyi, M., Reis-Filho, J. S., Solit, D. B., Jones, D. R., Gomez, D., Yu, H., Chakravarty, D., Yaeger, R., Abida, W., Park, W., O'Reilly, E. M., Garcia-Aguilar, J., Socci, N., Sanchez-Vega, F., Carrot-Zhang, J., Stetson, P. D., Levine, R., Rudin, C. M., Berger, M. F., Shah, S. P., Schrag, D., Razavi, P., Kehl, K. L., Li, B. T., Riely, G. J. & Schultz, N. Automated real-world data integration improves cancer outcome prediction. *Nature* **636**, 728–736 (2024).

279. Julio A. Aguirre-Ghiso, Aguirre-Ghiso, J. A., María Soledad Sosa & Sosa, M. S. Emerging Topics on Disseminated Cancer Cell Dormancy and the Paradigm of Metastasis. **2**, 377–393 (2018).
280. Malanchi, I., Santamaria-Martínez, A., Susanto, E., Peng, H., Lehr, H.-A., Delaloye, J.-F. & Huelsken, J. Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* **481**, 85–89 (2012).
281. Fico, F., Bousquenaud, M., Rüegg, C. & Santamaria-Martínez, A. Breast Cancer Stem Cells with Tumor- versus Metastasis-Initiating Capacities Are Modulated by TGFBR1 Inhibition. *Stem Cell Rep.* **13**, 1–9 (2019).
282. Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., Chang, H. Y. & Greenleaf, W. J. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
283. Kamimoto, K., Stringa, B., Hoffmann, C. M., Jindal, K., Solnica-Krezel, L. & Morris, S. A. Dissecting cell identity via network inference and in silico gene perturbation. *Nature* **614**, 742–751 (2023).
284. Jovanović, B., Temko, D., Stevens, L. E., Seehawer, M., Fassl, A., Murphy, K., Anand, J., Garza, K., Gulvady, A., Qiu, X., Harper, N. W., Daniels, V. W., Xiao-Yun, H., Ge, J. Y., Alečković, M., Pyrdol, J., Hinohara, K., Egri, S. B., Papanastasiou, M., Vadhi, R., Font-Tello, A., Witwicki, R., Peluffo, G., Trinh, A., Shu, S., Diciaccio, B., Ekram, M. B., Subedee, A., Herbert, Z. T., Wucherpfennig, K. W., Letai, A. G., Jaffe, J. D., Sicinski, P., Brown, M., Dillon, D., Long, H. W., Michor, F. & Polyak, K. Heterogeneity and transcriptional drivers of triple-negative breast cancer. *Cell Rep.* **42**, 113564 (2023).
285. Nicolás, F. J., Lehmann, K., Warne, P. H., Hill, C. S. & Downward, J. Epithelial to mesenchymal transition in Madin-Darby canine kidney cells is accompanied by down-regulation of Smad3 expression, leading to resistance to transforming growth factor-beta-induced growth arrest. *J. Biol. Chem.* **278**, 3251–3256 (2003).
286. Medici, D., Hay, E. D. & Goodenough, D. A. Cooperation between snail and LEF-1 transcription factors is essential for TGF-beta1-induced epithelial-mesenchymal transition. *Mol. Biol. Cell* **17**, 1871–1879 (2006).
287. Calabrese, E. J. Hormetic mechanisms. *Crit. Rev. Toxicol.* **43**, 580–606 (2013).
288. Calabrese, E. J. & Kozumbo, W. J. The hormetic dose-response mechanism: Nrf2 activation. *Pharmacol. Res.* **167**, 105526 (2021).
289. DeBlasi, J. M., Falzone, A., Caldwell, S., Prieto-Farigua, N., Prigge, J. R., Schmidt, E. E., Chio, I. I. C., Karreth, F. A. & DeNicola, G. M. Distinct Nrf2 Signaling Thresholds Mediate Lung Tumor Initiation and Progression. *Cancer Res.* **83**, 1953–1967 (2023).
290. Blassberg, R., Patel, H., Watson, T., Gouti, M., Metzis, V., Delás, M. J. & Briscoe, J. Sox2 levels regulate the chromatin occupancy of WNT mediators in epiblast progenitors responsible for vertebrate body formation. *Nat. Cell Biol.* **24**, 633–644 (2022).
291. Beck, B., Lapouge, G., Rorive, S., Drogat, B., Desaedelaere, K., Delafaille, S., Dubois, C., Salmon, I., Willekens, K., Marine, J.-C. & Blanpain, C. Different levels of Twist1 regulate skin tumor initiation, stemness, and progression. *Cell Stem Cell* **16**, 67–79 (2015).

292. Chan, A. S. L., Zhu, H., Narita, M., Cassidy, L. D., Young, A. R. J., Bermejo-Rodriguez, C., Janowska, A. T., Chen, H.-C., Gough, S., Oshimori, N., Zender, L., Aitken, S. J., Hoare, M. & Narita, M. Titration of RAS alters senescent state and influences tumour initiation. *Nature* **633**, 678–685 (2024).
293. Modi, S., Jacot, W., Yamashita, T., Sohn, J., Vidal, M., Tokunaga, E., Tsurutani, J., Ueno, N. T., Prat, A., Chae, Y. S., Lee, K. S., Niikura, N., Park, Y. H., Xu, B., Wang, X., Gil-Gil, M., Li, W., Pierga, J.-Y., Im, S.-A., Moore, H. C. F., Rugo, H. S., Yerushalmi, R., Zagouri, F., Gombos, A., Kim, S.-B., Liu, Q., Luo, T., Saura, C., Schmid, P., Sun, T., Gambhire, D., Yung, L., Wang, Y., Singh, J., Vitazka, P., Meinhardt, G., Harbeck, N. & Cameron, D. A. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. *N. Engl. J. Med.* **387**, 9–20 (2022).
294. Kowalski, P. J., Rubin, M. A. & Kleer, C. G. E-cadherin expression in primary carcinomas of the breast and its distant metastases. *Breast Cancer Res. BCR* **5**, R217-222 (2003).
295. Schade, A. E., Perurena, N., Yang, Y., Rodriguez, C. L., Krishnan, A., Gardner, A., Loi, P., Xu, Y., Nguyen, V. T. M., Mastellone, G. M., Pilla, N. F., Watanabe, M., Ota, K., Davis, R. A., Mattioli, K., Xiang, D., Zoeller, J. J., Lin, J.-R., Morganti, S., Garrido-Castro, A. C., Tolaney, S. M., Li, Z., Barbie, D. A., Sorger, P. K., Helin, K., Santagata, S., Knott, S. R. V. & Cichowski, K. AKT and EZH2 inhibitors kill TNBCs by hijacking mechanisms of involution. *Nature* **635**, 755–763 (2024).
296. Ishay-Ronen, D., Diepenbruck, M., Kalathur, R. K. R., Sugiyama, N., Tiede, S., Ivanek, R., Bantug, G., Morini, M. F., Wang, J., Hess, C. & Christofori, G. Gain Fat—Lose Metastasis: Converting Invasive Breast Cancer Cells into Adipocytes Inhibits Cancer Metastasis. *Cancer Cell* **35**, 17-32.e6 (2019).
297. Ascic, E., Åkerström, F., Sreekumar Nair, M., Rosa, A., Kurochkin, I., Zimmermannova, O., Catena, X., Rotankova, N., Veser, C., Rudnik, M., Ballocci, T., Schärer, T., Huang, X., de Rosa Torres, M., Renaud, E., Velasco Santiago, M., Met, Ö., Askmyr, D., Lindstedt, M., Greiff, L., Ligeon, L.-A., Agarkova, I., Svane, I. M., Pires, C. F., Rosa, F. F. & Pereira, C.-F. In vivo dendritic cell reprogramming for cancer immunotherapy. *Science* **386**, eadn9083 (2024).
298. Andrew McPherson, McPherson, A., Andrew Roth, Roth, A., Emma Laks, Laks, E., Tehmina Masud, Masud, T., Ali Bashashati, Bashashati, A., Allen W. Zhang, Zhang, A. W., Gavin Ha, Ha, G., Justina Biele, Biele, J., Biele, J., Damian Yap, Yap, D., Adrian Wan, Wan, A., Leah M Prentice, Prentice, L. M., Jaswinder Khattra, Jaswinder Khattra, Khattra, J., Maia A. Smith, Smith, M. A., Cydney B. Nielsen, Nielsen, C. B., Sarah C. Mullaly, Mullaly, S. C., Steve E. Kalloger, Kalloger, S. E., Anthony N. Karnezis, Karnezis, A. N., Karey Shumansky, Shumansky, K., Celia Siu, Siu, C., Jamie Rosner, Jamie Rosner, Rosner, J., Hector Li Chan, Chan, H. L., Julie Ho, Ho, J., Nataliya Melnyk, Melnyk, N., Janine Senz, Senz, J., Winnie Yang, Yang, W., Richard A. Moore, Moore, R. A., Richard Moore, Andrew J. Mungall, Mungall, A. J., Marco A. Marra, Marra, M. A., Alexandre Bouchard-Côté, Bouchard-Côté, A., C. Blake Gilks, Gilks, C. B., C. Blake Gilks, David G. Huntsman, Huntsman, D. G., Jessica N. McAlpine, McAlpine, J. N., Samuel Aparicio, Aparicio, S., Sohrab P. Shah & Shah, S. P. Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat. Genet.* **48**, 758–767 (2016).
299. Gunes Gundem, Gundem, G., Gundem, G., Peter Van Loo, Van Loo, P., Bárbara Kremeyer, Kremeyer, B., Ludmil B. Alexandrov, Alexandrov, L. B., Tubio Jmc., Jmc., T., Elli Papaemmanuil, Papaemmanuil, E., Papaemmanuil, E., Elli Papaemmanuil, Daniel Brewer, Brewer, D., Heini Kallio, Kallio, H., Gunilla Högnäs, Högnäs, G., Matti Annala, Annala, M.,

Kati Kivinummi, Kivinummi, K., Victoria Goody, Goody, V., Calli Latimer, Latimer, C., Sarah O'Meara, O'Meara, S., Kevin J. Dawson, Dawson, K. J., William B. Isaacs, Isaacs, W. B., M. R. Emmert-Buck, Emmert-Buck, M. R., Matti Nykter, Nykter, M., Chris Foster, Christopher S. Foster, Foster, C. S., Foster, C. S., Foster, C., Z Kote–Jarai, Kote-Jarai, Z., Kote-Jarai, Z., Easton, D. F., Douglas Easton, Easton, D., Easton, DouglasF., Hayley C. Whitaker, Whitaker, H. C., DE Neal, Neal, D. E., Colin S. Cooper, Colin S. Cooper, Cooper, C., Ros Eeles, Eeles, R. A., Tapio Visakorpi, Visakorpi, T., Visakorpi, T., P J Campbell, Campbell, P. J., McDermott, M. W., Ultan McDermott, McDermott, U., David C. Wedge, Wedge, D. C., G Bova, G. S. Bova, G. S. Bova & Bova, G. S. The evolutionary history of lethal metastatic prostate cancer. *Cancer Res.* **75**, (2015).

300. Watson, S. S., Zomer, A., Fournier, N., Lourenco, J., Quadroni, M., Chryplewicz, A., Nassiri, S., Aubel, P., Avanthay, S., Croci, D., Abels, E., Broekman, M. L. D., Hanahan, D., Huse, J. T., Daniel, R. T., Hegi, M. E., Homicsko, K., Cossu, G., Hottinger, A. F. & Joyce, J. A. Fibrotic response to anti-CSF-1R therapy potentiates glioblastoma recurrence. *Cancer Cell* **42**, 1507-1527.e11 (2024).
301. Acha-Sagredo, A., Andrei, P., Clayton, K., Taggart, E., Antoniotti, C., Woodman, C. A., Afrache, H., Fourny, C., Armero, M., Moinudeen, H. K., Green, M., Bhardwaj, N., Mikolajczak, A., Rodriguez-Lopez, M., Crawford, M., Connick, E., Lim, S., Hobson, P., Linares, J., Ignatova, E., Pelka, D., Smyth, E. C., Diamantis, N., Sosnowska, D., Carullo, M., Ciraci, P., Bergamo, F., Intini, R., Nye, E., Barral, P., Mishto, M., Arnold, J. N., Lonardi, S., Cremolini, C., Fontana, E., Rodriguez-Justo, M. & Ciccarelli, F. D. A constitutive interferon-high immunophenotype defines response to immunotherapy in colorectal cancer. *Cancer Cell* **43**, 292-307.e7 (2025).
302. Klughammer, J., Abravanel, D. L., Segerstolpe, Å., Blosser, T. R., Goltsev, Y., Cui, Y., Goodwin, D. R., Sinha, A., Ashenberg, O., Slyper, M., Vigneau, S., Jané-Valbuena, J., Alon, S., Caraccio, C., Chen, J., Cohen, O., Cullen, N., DelloStritto, L. K., Dionne, D., Files, J., Frangieh, A., Helvie, K., Hughes, M. E., Inga, S., Kanodia, A., Lako, A., MacKichan, C., Mages, S., Moriel, N., Murray, E., Napolitano, S., Nguyen, K., Nitzan, M., Ortiz, R., Patel, M., Pfaff, K. L., Porter, C. B. M., Rotem, A., Strauss, S., Strasser, R., Thorner, A. R., Turner, M., Wakiro, I., Waldman, J., Wu, J., Gómez Tejeda Zañudo, J., Zhang, D., Lin, N. U., Tolaney, S. M., Winer, E. P., Boyden, E. S., Chen, F., Nolan, G. P., Rodig, S. J., Zhuang, X., Rozenblatt-Rosen, O., Johnson, B. E., Regev, A. & Wagle, N. A multi-modal single-cell and spatial expression map of metastatic breast cancer biopsies across clinicopathological features. *Nat. Med.* **30**, 3236–3249 (2024).
303. Wang, X. Q., Danenberg, E., Huang, C.-S., Egle, D., Callari, M., Bermejo, B., Dugo, M., Zamagni, C., Thill, M., Anton, A., Zambelli, S., Russo, S., Ciruelos, E. M., Greil, R., Győrffy, B., Semiglazov, V., Colleoni, M., Kelly, C. M., Mariani, G., Del Mastro, L., Biasi, O., Seitz, R. S., Valagussa, P., Viale, G., Gianni, L., Bianchini, G. & Ali, H. R. Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature* **621**, 868–876 (2023).
304. Burdziak, C., Alonso-Curbelo, D., Walle, T., Reyes, J., Barriga, F. M., Haviv, D., Xie, Y., Zhao, Z., Zhao, C. J., Chen, H.-A., Chaudhary, O., Masilionis, I., Choo, Z.-N., Gao, V., Luan, W., Wuest, A., Ho, Y.-J., Wei, Y., Quail, D. F., Koche, R., Mazutis, L., Chaligné, R., Nawy, T., Lowe, S. W. & Pe'er, D. Epigenetic plasticity cooperates with cell-cell interactions to direct pancreatic tumorigenesis. *Science* **380**, eadd5327 (2023).
305. Blohmer, M., Cheek, D. M., Hung, W.-T., Kessler, M., Chatzidimitriou, F., Wang, J., Hung, W., Lee, I.-H., Gorelick, A. N., Wassenaar, E. C., Yang, C.-Y., Yeh, Y.-C., Ho, H.-L., Speiser, D., Karsten, M. M., Lanuti, M., Pai, S. I., Kranenburg, O., Lennerz, J. K., Chou, T.-Y., Kloor, M. &

Naxerova, K. Quantifying cell divisions along evolutionary lineages in cancer. *Nat. Genet.* 1–12 (2025) doi:10.1038/s41588-025-02078-5.

306. Lagha, M., Bothma, J. P. & Levine, M. Mechanisms of transcriptional precision in animal development. *Trends Genet. TIG* **28**, 409–416 (2012).
307. Baldominos, P., Barbera-Mourelle, A., Barreiro, O., Huang, Y., Wight, A., Cho, J.-W., Zhao, X., Estivill, G., Adam, I., Sanchez, X., McCarthy, S., Schaller, J., Khan, Z., Ruzo, A., Pastorello, R., Richardson, E. T., Dillon, D., Montero-Llopis, P., Barroso-Sousa, R., Forman, J., Shukla, S. A., Tolaney, S. M., Mittendorf, E. A., von Andrian, U. H., Wucherpfennig, K. W., Hemberg, M. & Agudo, J. Quiescent cancer cells resist T cell attack by forming an immunosuppressive niche. *Cell* **185**, 1694-1708.e19 (2022).
308. Chen, L., Gibbons, D. L., Goswami, S., Cortez, M. A., Ahn, Y.-H., Byers, L. A., Zhang, X., Yi, X., Dwyer, D., Lin, W., Diao, L., Wang, J., Roybal, J. D., Patel, M., Ungewiss, C., Peng, D., Antonia, S., Mediavilla-Varela, M., Robertson, G., Jones, S., Suraokar, M., Welsh, J. W., Erez, B., Wistuba, I. I., Chen, L., Peng, D., Wang, S., Ullrich, S. E., Heymach, J. V., Kurie, J. M. & Qin, F. X.-F. Metastasis is regulated via microRNA-200/ZEB1 axis control of tumour cell PD-L1 expression and intratumoral immunosuppression. *Nat. Commun.* **5**, 5241 (2014).
309. Dhimolea, E., de Matos Simoes, R., Kansara, D., Al'Khafaji, A., Bouyssou, J., Weng, X., Sharma, S., Raja, J., Awate, P., Shirasaki, R., Tang, H., Glassner, B. J., Liu, Z., Gao, D., Bryan, J., Bender, S., Roth, J., Scheffer, M., Jeselsohn, R., Gray, N. S., Georgakoudi, I., Vazquez, F., Tsherniak, A., Chen, Y., Welm, A., Duy, C., Melnick, A., Bartholdy, B., Brown, M., Culhane, A. C. & Mitsiades, C. S. An Embryonic Diapause-like Adaptation with Suppressed Myc Activity Enables Tumor Treatment Persistence. *Cancer Cell* **39**, 240-256.e11 (2021).
310. Naumov, G. N., Townson, J. L., MacDonald, I. C., Wilson, S. M., Bramwell, V. H. C., Groom, A. C. & Chambers, A. F. Ineffectiveness of doxorubicin treatment on solitary dormant mammary carcinoma cells or late-developing metastases. *Breast Cancer Res. Treat.* **82**, 199–206 (2003).
311. Ranganathan, A. C., Zhang, L., Adam, A. P. & Aguirre-Ghiso, J. A. Functional coupling of p38-induced up-regulation of BiP and activation of RNA-dependent protein kinase-like endoplasmic reticulum kinase to drug resistance of dormant carcinoma cells. *Cancer Res.* **66**, 1702–1711 (2006).
312. Sumaiyah K. Rehman, Rehman, S. K., Jennifer Haynes, Jennifer Haynes, Haynes, J., Evelyne Collignon, Collignon, E., Kevin R. Brown, Brown, K. R., Yadong Wang, Wang, Y., Allison M.L. Nixon, Nixon, A. M. L., Jeffrey P. Bruce, Bruce, J., Jeffrey Wintersinger, Wintersinger, J., Wintersinger, J., Arvind Singh Mer, Mer, A. S., E. Lo, Lo, E. B. L., Clement H. C. Leung, Cherry Leung, Leung, C., Cherry Leung, Evelyne Lima-Fernandes, Lima-Fernandes, E., Evelyne Lima-Fernandes, Evelyne Lima-Fernandes, Nicholas Pedley, Pedley, N. M., Nicholas M. Pedley, Nicholas M. Pedley, Fraser Soares, Fraser Soares, Fraser Soares, Soares, F., Sophie McGibbon, McGibbon, S., Housheng Hansen He, He, H. H., Pollett, A., Aaron Pollet, Pollet, A., Trevor J. Pugh, Pugh, T. J., Benjamin Haibe-Kains, Haibe-Kains, B., Quaid Morris, Morris, Q., Miguel Ramalho-Santos, Miguel Ramalho-Santos, Ramalho-Santos, M., Ramalho-Santos, M., Sidhartha Goyal, Goyal, S., Jason Moffat, Jason Moffat, Jason Moffat, Moffat, J., Catherine O'Brien & O'Brien, C. A. Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive Chemotherapy. *Cell* **184**, (2021).
313. Cuzick, J., Swanson, G. P., Fisher, G., Brothman, A. R., Berney, D. M., Reid, J. E., Mesher, D., Speights, V. O., Stankiewicz, E., Foster, C. S., Møller, H., Scardino, P., Warren, J. D., Park, J., Younus, A., Flake, D. D., Wagner, S., Gutin, A., Lanchbury, J. S., Stone, S., & Transatlantic

Prostate Group. Prognostic value of an RNA expression signature derived from cell cycle proliferation genes in patients with prostate cancer: a retrospective study. *Lancet Oncol.* **12**, 245–255 (2011).

314. van Diest, P. J., van der Wall, E. & Baak, J. P. A. Prognostic value of proliferation in invasive breast cancer: a review. *J. Clin. Pathol.* **57**, 675–681 (2004).
315. Wistuba, I. I., Behrens, C., Lombardi, F., Wagner, S., Fujimoto, J., Raso, M. G., Spaggiari, L., Galetta, D., Riley, R., Hughes, E., Reid, J., Sangale, Z., Swisher, S. G., Kalhor, N., Moran, C. A., Gutin, A., Lanchbury, J. S., Barberis, M. & Kim, E. S. Validation of a proliferation-based expression signature as prognostic marker in early stage lung adenocarcinoma. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **19**, 6261–6271 (2013).
316. Magbanua, M. J. M., Rugo, H. S., Wolf, D. M., Hauranieh, L., Roy, R., Pendyala, P., Sosa, E. V., Scott, J. H., Lee, J. S., Pitcher, B., Hyslop, T., Barry, W. T., Isakoff, S. J., Dickler, M., Van't Veer, L. & Park, J. W. Expanded Genomic Profiling of Circulating Tumor Cells in Metastatic Breast Cancer Patients to Assess Biomarker Status and Biology Over Time (CALGB 40502 and CALGB 40503, Alliance). *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **24**, 1486–1499 (2018).
317. Ganesh, K. & Massagué, J. Targeting metastatic cancer. *Nat. Med.* **27**, 34–44 (2021).
318. Risson, E., Nobre, A. R., Maguer-Satta, V. & Aguirre-Ghiso, J. A. The current paradigm and challenges ahead for the dormancy of disseminated tumor cells. *Nat. Cancer* **1**, 672–680 (2020).