Identification and functional analysis of molecular mechanisms involved in the latency of ER positive breast cancer

Sylwia Henryka Gawrzak

UNIVERSITAT DE BARCELONA

IRB BARCELONA
INSTITUTE FOR RESEARCH IN BIOMEDICINE

Identification and functional analysis of molecular mechanisms involved in the latency of ER positive breast cancer

Sylwia Henryka Gawrzak

Memòria presentada per Sylwia Henryka Gawrzak per optar al grau de doctora per la Universitat de Barcelona

Doctoral thesis in Biomedicine

Universitat de Barcelona, 2016

Thesis director: Roger Gomis Cabré, Ph.D

Institute for Research in Biomedicine (IRB) Barcelona

Laboratory of growth control and metastasis

Roger Gomis Cabré

director and tutor

Sylwia Henryka Gawrzak

Ph.D candidate

Precę dedykuję mojej mamie oraz
wszystkim kobietom.

ACKNOWLEDGEMENTS

This thesis owes its existence to the help, support and inspiration of several people. I would like to express my sincere appreciation and gratitude to:

Roger Gomis, the thesis director, tutor and the head of the laboratory for his guidance during my research, supervision and thesis corrections. I highly appreciate the trust, scientific freedom and honest and critical feedback that he gave me.

Jelena Urošević, who has been a constant source of encouragement, and a mentor not only for this thesis project but also for my life.

Marc Guiu for his superb conduction of animal experiments with a particular empathy and awareness of animal welfare.

Angel Nebreda, **Raúl Mendéz** and **Francesc Viñals**, my thesis advisory committee members for their valuable advices.

Enrique Arenas and **Esther Fernández** for helping me with many experimental procedures and applying their expertise into this project.

Salvador Aznar Benitah and **Lorenzo Rinaldi** for the collaboration and their expertise in epigenetics.

Met Labmates: Anna Arnal, Anna Bellmunt, Alicia Llorente, Milica Pavlović, Fernando Salvador, Ivan Burkov, Juan Miguel Cejalvo, Daniela Kalafatović and those mentioned abovefor creating a vibrant and plausible niche. the core of the metlab

Master´s students Georgina Florez, Barbara Müllauer and Rianne Groeneveld for the enthusiastic assistance.

Facilities at IRB and PCB (Advanced Digital Microscopy, Biostatistics/ Bioinformatics, Functional Genomics, Histopathology, Flow Cytometry), the administration, and members of the IRB Student Council for the contribution to my research and support.

Tanya Yates and **Konstantin Slobodnyuk** for the editorial support and text corrections

.

Peter Sicinski, **Junko Odajima**, **Maria Ciemerych-Litwinienko** and **Iwona Grabowska** for introducing me into science and setting the high standards.

I am also indebted to my **friends** from all around the word. You were fundamental in creating joyful and memorable moments and supporting me during stressful and difficult time.

Finally, my deepest gratitude goes to **my beloved mum and my family** for their unflagging love and unconditional support throughout my life. You helped me develop, provided the best education, and accepted choices that I have made. I could not ask for more.

TABLE OF CONTENTS

ABSTRACT

Breast cancer is the most frequently diagnosed cancer and remains the second leading cause of death among women in Europe and United States. In this malignancy, metastasis remains to be an incurable condition, and therefore the major cause of death. Metastatic lesions can appear within a wide time ranging from months to years or decades after primary tumor resection. In particular, in the estrogen receptor (ER) positive breast cancer subgroup metastatic latency continues to be a major challenge for the researchers, clinicians and patients.

This thesis reports the identification and functional analysis of molecular mechanisms involved in the latency of ER positive breast cancer. For that purpose we based our research on a comprehensive approach that relies on genetically engineered human breast cancer cells, experimental mouse models, unbiased whole-genome screen and clinical data. The first part of the thesis describes a novel mouse model of breast cancer dormancy. We showed that metastatic cells home the bone and enter the latency phase as micrometastatic lesions where tumor growth is restricted mainly due to the equilibrated ratios of cell proliferation and cell death. This experimental mouse model was used to identify genes relevant for long-latent relapse. To this end, we performed *in vivo* loss-of-function shRNA screening. In the screening we challenged a whole-genome library of shRNA to uncover genes whose depletion negatively regulates dormancy. Among the candidate genes revealed by the screen we focused on MSK1 as a long-latent metastasis regulator. The *in vivo* and *in vitro* validation results indicate that MSK1 plays a role in homing and differentiation of metastatic cells. We showed that MSK1 promotes the expression luminal tran-

scription factors - FOXA1 and GATA-3. Therefore, MSK1 depletion is beneficial for metastatic cells leading to a partial phenotype shift towards a more aggressive and poorly differentiated basal population. Furthermore, our data suggest that MSK1 may be involved in metastatic cell plasticity by remodeling the chromatin. Importantly, low MSK1 gene expression levels associate with early metastasis in ER positive breast cancer.

RESUMEN DE LA THESIS

El cáncer de mama es el tipo de cáncer más frecuentemente diagnosticado, siendo la segunda causa de muerte entre las mujeres de Europa y Estados Unidos. En esta enfermedad, la metástasis sigue siendo incurable, y por ello es la principal causa de muerte. Las lesiones metastásicas pueden aparecer dentro de un amplio periodo de tiempo que va desde meses hasta años o incluso décadas después de la extirpación del tumor primario. Concretamente, en el subgrupo de cáncer de mama receptor de estrógeno (RE) positivo, este largo periodo de latencia es el principal desafío para investigadores, médicos y pacientes.

En esta tesis se muestra la identificación y el análisis funcional de mecanismos moleculares implicados en la latencia del cáncer de mama RE positivo. Para este propósito, nuestros estudios se han llevado a cabo mediante una estrategia experimental basada en líneas celulares de cáncer de mama genéticamente modificadas, modelos experimentales de ratón, análisis global del genoma y datos clínicos. La primera parte de la tesis describe un novedoso modelo de ratón de dormancia de cáncer de mama. Observamos que, en nuestro modelo, las células metastásicas llegan al hueso y entran en una fase de latencia en forma de lesiones micrometastásicas en la que el crecimiento del tumor se ve impedido, principalmente debido a que la tasa de proliferación celular se iguala a la tasa de muerte celular. Este modelo experimental de ratón se usó para identificar genes relevantes en el proceso de latencia y por tanto en la recurrencia a largo plazo. Para ello, llevamos a cabo un análisis in vivo de pérdida de función con shRNA. En este análisis utilizamos una amplia librería de shRNA para descubrir genes cuya eliminación regula la dormancia de manera negativa. Entre los

genes candidatos identificados en este análisis nos focalizamos en MSK1 como un regulador de la metástasis latente. La validación *in vitro* e *in vivo* indica que MSK1 juega un papel en el anidamiento y la diferenciación de las células metastásicas. Mostramos que MSK1 promueve la expresión de factores de transcripción de tipo luminal – FOXA1 y GATA-3. Por ello, la deleción de MSK1 es beneficiosa para las células metastáticas y las convierte en una población basal más agresiva y menos diferenciada. Además, nuestros datos sugieren que MSK1 podría estar implicada en la plasticidad de las células metastásicas a través de la remodelación de la cromatina. Por último, es importante destacar que la baja expresión del gen MSK1 está asociada con metastásis temprana en cáncer de mama RE positivo.

LIST OF ABBREVIATIONS

A

ANGLPTL4 angiopoietin-like 4 *5*

ATF-1 activating transcription factor 1 *29*

B

BLI bioluminescence imaging *40*

BMP bone morphogenetic proteins *10*

BRCA-1 breast cancer 1 *98*

BrdU bromodeoxyuridine *48*

C

CGH comparative genomic hybridization *42*

CREB cAMP response element binding *29*

CRISPR clustered regularly interspaced short palindromic repeats *68*

CSCs cancer stem cells *23*

CSF-1 colony-stimulating factor 1 *4*

CTCs circulating tumor cells *4*

CTKD C-terminal kinase domain *26*

CXCR4 C-X-C chemokine receptor 4 *21*

D

DBM dormant bone metastatic derivative *42*

DTCs disseminated tumor cells *5*

DYRK dual specificity tyrosine phosphorylation-regulated kinases *11*

E

ECM extracellular matrix *3*

EdU deoxyethynyluridine *51*

EGF epidermal growth factor *4*

EGFP enhanced GFP *70*

EMT epithelial-to-mesenchymal transition *4*

ER estrogen receptor *14*

ER- estrogen receptor negative *15*

ER+ estrogen receptor positive *14*

ERK1/2 extracellular signal-regulated kinase 1/2 *11*

F

FACS fluorescence-activated cell sorting *57*

FOXA1 forkhead box A1 *81*

FOXO3a forkhead box 3a *32*

G

GAS6 growth arrest-specific 6 *10*

GATA2 GATA binding protein 2 *81*

GATA3 GATA binding protein 3 *81*

gDNA genomic DNA *58*

GEMM genetically engineered mouse models *38*

GFP+ green fluorescent protein positive *48*

GO gene ontology *78*

gRNA guide RNA *68*

GSEA gene set enrichment analysis *78*

H

H3K27ac acetylation of histone H3 at lysine 27 *97*

H3K27me3 trimethylation of histone H3 lysine 27 *30*

H3S10 phosphorylation of histone H3 at serine 10 *30*

H3S28A substitution of serine 28 with alanine at histone H3 *97*

H3S28 phosphorylation of histone H3 at serine 28 *30*

Her2 human epidermal growth factor receptor-2 *15*

HMGN1 high mobility group nucleosome binding domain 1 *30*

HSCs hematopoietic stem cells *20*

HSP27 heat shock 27 kDa protein *12*

I

IEG immediate-early genes *29*

IGFs insulin-like growth factors *22*

IL interleukin *22*

IVIS-μCT live-imaging with micro-computed tomography *46*

J

JNK including c-Jun N-terminal kinase *29*

K

KO knock-out *29*

L

LRCs label-retaining cells *48*

M

MEK mitogen-activated protein kinase kinase *27*

MET mesenchymal-to-epithelial transition *6*

MIC metastasis-initiating cancer cell *23*

MMP matrix metalloproteinases *22*

X

MMTV mouse mammary tumor virus *33*

MMTV-PyMT MMTV-polyoma middle T *95*

MOI multiplicity of infection *41*

mRNA messenger RNA *91*

MSK mitogen- and stress-activated kinase *26*

N

NES normalized enrichment score *78*

NFκB nuclear factor kappa-light-chain-enhancer of activated B cells *29*

NGS next generation sequencing *57*

NHEJ non-homologous end-joining *68*

NTKD N-terminal kinase domain *26*

O

OPG osteoprotegerin *22*

P

p38 MAPK mitogen-activated protein kinase p38 *11*

PAM protospacer adjacent motifs *68*

PcG Polycomb group *30*

PRC Polycomb repressive complexes *97*

PRK2 double-stranded-RNA-dependent protein kinase 2 *27*

PR progesterone receptor *14*

PTHLH parathyroid hormone-like hormone *5*

PTHrP parathyroid hormone-like protein *22*

R

RANKL receptor activator of nuclear factor kappa-B ligand *22*

RNAi RNA interference *40*

ROCKII Rho-associated protein kinase 2 *27*

RSK 90-kDa ribosomal S6 kinases *26*

S

SDF-1 stromal cell-derived factor 1 (also known as CXCL12 *21*

shRNA short hairpin RNA *40*

siRNA small interfering RNA *91*

STAT3 signal transducer and activator of transcription 3 *29*

T

TGFß transforming growth-factor beta *4*

TICs tumor-initiating cells *23*

TNF-α tumor necrosis factor *22*

TSP-1 thrombospondin-1 *12*

U

uPAR urokinase plasminogen activator receptor *11*

UPR unfolded protein response *11*

V

VEGF vascular endothelial growth factor *12*

W

WT-E wild type-edited *70*

WT wilde-type *70*

INTRODUCTION

1. Latency in metastatic cancer progression

Despite the progress in clinical oncology and basic cancer research, metastasis continues to be a major cause of cancer-related death. In metastasis, malignant cells spread from the primary tumor to distant sites, where they may resist conventional therapies, proliferate, and cause failure of a vital organ. Through the systemic dissection of the molecular, cellular, genetic, and clinical mechanisms underlying metastatic progression cancer research field aims to develop new strategies to prevent and treat metastases. The main limitations of metastasis research include the biological heterogeneity of cancer types, clonal heterogeneity of primary tumors, genetic heterogeneity of cancer cells in the primary and secondary sites, and complex interactions between cancer cells and the microenvironment. In line with this, different cancer types show distinct metastatic organ-tropism. In addition, although steps in the metastatic cascade are part of continuous biological sequence, their acquisition in time may vary from one tumor type to another.

1.1 The metastatic cascade

The metastatic cascade is a series of stochastic events that collectively lead to the formation of overt metastases in a distant organ. It involves seven steps: invasion, intravasation, dissemination in the circulation and survival, arrest at a distant site, extravasation, tumor initiation, and, finally, outgrowth and clinical manifestation (Valastyan and Weinberg 2011; Obenauf and Massague 2015) (Figure 1).

Metastasis is a highly inefficient process in which each step of the cascade is a bottleneck for cancer cells and drives clonal selection. By the end of this process, only small fraction of thousands of daily seeded cells reinitiates a tumor in a distant site. Studies based

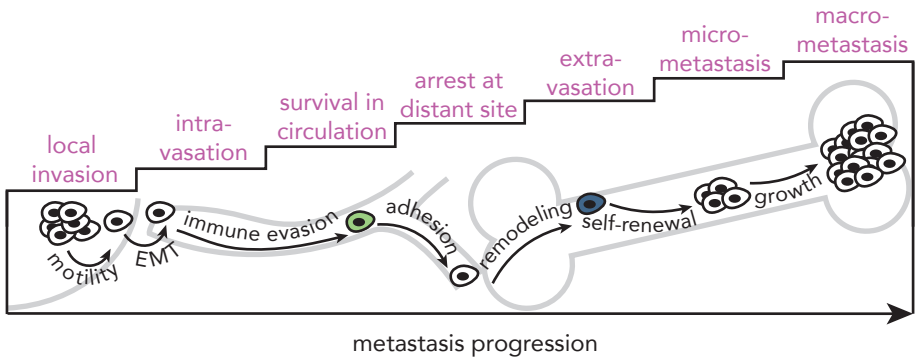

Figure 1. The metastatic cascade. Metastasis progresses through the sequence of steps that promote malignant cells, from primary tumor, to disseminate and colonize a distant organ. Acquisition of each step is driven by specific cellular functions. Cascade steps are indicated in pink, cell autonomous functions in black, circulating tumor cell in green, and disseminated tumor cell in blue.

on experimental models estimate that 0.02% of cells succeed in colonizing distant organs (Luzzi, MacDonald et al. 1998; Cameron, Schmidt et al. 2000; Chambers, Groom et al. 2002). In order to metastasize, cancer cells need to orchestrate diverse cellular functions to overcome the difficulties of the metastatic cascade. These functions are not only limited to cell-autonomous traits, but also highly depend on the interaction of the metastatic cell with the tumor and host stroma. In some cases, several functions are required to implement a single step, while others may influence multiple steps. From a mechanistic perspective, genetic, epigenetic and translational traits alter the expression of promoter and suppressor genes to facilitate metastasis

Metastasis originates in the primary tumor invasive front, where cancer cells migrate toward surrounding tissues. To achieve this movement, cellular motility is altered by cytoskeleton reorganization and the secretion of extracellular matrix (ECM) remodelers, mainly proteases (Kessenbrock, Plaks et al. 2010). Tumor stroma composed of tumor-associated macrophages and fibroblasts sup-

ports the invasion of metastatic cells by secreting pro-migratory factors (Joyce, Baruch et al. 2004; Qian and Pollard 2010).

In order to intravasate, metastatic cells undergo an epithelial-to-mesenchymal transition (EMT). The loss of epithelial features, like adhesion or polarization, followed by gain of invasiveness greatly contributes to metastasis. In this regard, the downregulation of epithelial protein E-cadherin is a well-established prognostic marker for metastasis (Beavon 2000). Successful intravasation also requires the formation of a leaky vasculature in the primary tumor. The tumor vasculature is a physical barrier to cancer cells and in order to cross it they secrete factors that increase vessel permeability, thereby allowing their entry into the circulation. Various mediators are involved in this process, including cytokines and growth factors: transforming growth-factor beta (TGFß), or those molecules produced by supportive tumor stroma, namely epidermal growth factor (EGF) and colony-stimulating factor 1 (CSF-1) (Wyckoff, Wang et al. 2007; Giampieri, Manning et al. 2009).

The bloodstream or lymphatic system is a hostile environment for cancer cells, and transition through vessels results in massive cell death. On the one hand, cells are challenged by innate immune natural killer cells and, on the other, they die from mechanical damage (Nieswandt, Hafner et al. 1999; Massague and Obenauf 2016). In order to enhance survival in the circulation, cancer cells associate with blood platelets or adhere to the endothelium at the destination site (Joyce and Pollard 2009). The isolation of circulating tumor cells (CTCs) from the blood of cancer patients offers valuable information about disease progression and treatment design. Results from 300 clinical trials have revealed the prognostic relevance of CTC counts with respect to metastatic progression. In addition to the counts, the analysis of surface markers expressed by CTCs can

be used to monitor response to therapy and treatment-driven clonal selection (Mitra, Mishra et al. 2015).

After reaching the secondary site, metastatic cells are arrested in the microvasculature of the host organ prior to extravasation. Adhesion and interaction between CTCs and the host stroma facilitates microvasculature trapping (Labelle and Hynes 2012). Extravasation in bone or liver is facilitated by extrinsic factors such as the permeability of capillaries. In other organs, such as the lungs, cancer cells acquire new functions in order to cross the vessel wall, composed of endothelial cells, basement membrane and tissue-specific cells, and enter the parenchyma. Vessel remodeling can be achieved by cancer cell-secreted factors that increase the permeability of the endothelium. In the case of lung metastasis, angiopoietin-like 4 (ANGLPTL4) disrupts cell junctions in the vascular endothelium (Padua, Zhang et al. 2008), and parathyroid hormone-like hormone (PTHLH) induces endothelial cell death (Urosevic, Garcia-Albeniz et al. 2014).

Once metastatic cells extravasate and settle in the secondary site as disseminated tumor cells (DTCs), they must adapt to the microenvironment of the host organ in order to achieve homing to a distant location. Organ-specific extrinsic factors, including stroma, ECM, cytokines, and growth factors, compromise the survival of DTCs in the tissues. To overcome these obstacles, metastatic cells use cell-autonomous traits that facilitate homing and survival by altering SRC tyrosine kinase signaling (Zhang, Wang et al. 2009). These cells also improve homing and micrometastasis formation by creating pre-metastatic niches at the destination. In this scenario, the primary tumor secretes systemic factors to prime tissues at the secondary site. Consequently, cells extravasate to more a permissive microenvironment. In this regard, the enzyme lysyl ox-

idase is a potent pre-metastatic niche regulator (Erler, Bennewith et al. 2009). Moreover, exosomes have recently been shown to promote pre-metastatic niche formation (Peinado, Aleckovic et al. 2012). While EMT (de-differentiation) greatly contributes to invasion, disseminated cells revert to their mesenchymal phenotype to reinitiate the tumor in a process of mesenchymal-to-epithelial transition (MET), also called re-differentiation. This process explains why carcinoma-derived metastases show epithelial characteristics and resemble, to certain extent, the primary tumor (Brabletz 2012). Whereas some lesions expand rapidly, in many tumor types DTCs are arrested and remain dormant for many years (see sections 1.2 and 1.3).

The last step in the metastatic cascade is the overgrowth of micrometastases into full-blown symptomatic lesions that are clinically detectable. Metastatic cells extensively proliferate, causing the failure of vital organs. This metastatic virulence is driven in organ-specific manner and depends on a wide range of intrinsic and extrinsic mechanisms. For example, breast cancer cells create a permissive metastatic niche in lungs via stimulation of stromal fibroblasts to produce periostin. This ECM protein stimulates WNT signaling pathway in a stem cell subpopulation of metastatic cells promoting secondary tumor growth (Malanchi, Santamaria-Martinez et al. 2012). Periostin also interacts with tanescin C which can be secreted by tumor cells or stomal fibroblasts. Its high expression associates with poor prognosis for breast cancer patients and lung metastasis, since tanescin C promotes cancer cell survival and proliferation (Oskarsson, Acharyya et al. 2011).

1.2 The temporal course of metastasis

Although the steps of the metastatic cascade are, to certain extent, uniform for most types of carcinoma, the kinetics of metastasis is highly dependent on primary tumor location. Clinically detectable distant metastasis can occur simultaneously with primary tumor diagnosis or within a time ranging from weeks to decades (Figure 2). The period between primary tumor detection and metastatic relapse is often defined as latency. In metastatic latency, malignant cells that survived treatment and are neither detectable by conventional tests nor manifest symptoms contribute to minimal residual disease. Therefore, CTCs and DTCs in patients' blood or bone marrow are direct evidence of minimal residual disease

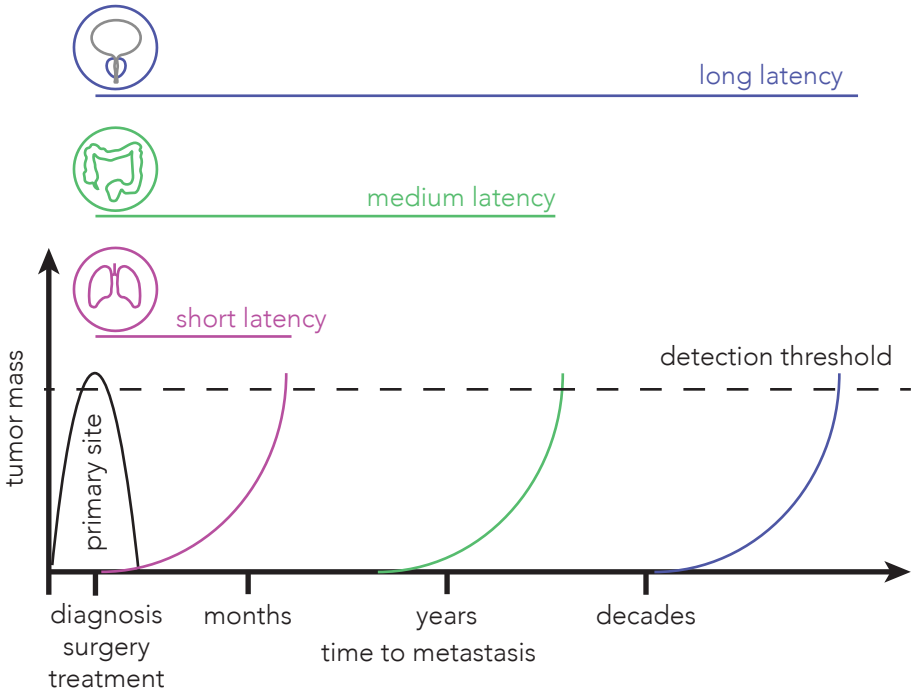

Figure 2. The temporal course of metastasis. Metastatic relapse may occur within months, years or decades after primary tumor diagnosis, removal, and systemic treatment. Different cancer types exhibit variability in length of the latency: short for lung cancer (pink), middle for colon cancer (green), and long for prostate cancer (blue). Figure adapted and modified from Aguirre-Ghiso 2007.

in metastatic latency and risk factors for recurrence (Aguirre-Ghiso 2007). Strikingly, metastatic cells from minimal residual disease can be transferred through organ transplants. Organs from donors diagnosed with melanoma, but successfully treated and clinically disease-free for over 10 years develop metastases after transplantation (Stephens, Everson et al. 2000; Strauss and Thomas 2010).

The duration of metastatic latency varies between cancer types, and for the most aggressive ones it is very short, resulting in high relapse and mortality rates following diagnosis. In lung cancer, the metastatic latency interval usually lasts only a few weeks, thus 5-year survival rates estimate around 17% (Goss and Chambers 2010; Howlader N 2016). The relapse rate is lower, reaching 30-40% in stage I lung adenocarcinoma patients (Nesbitt, Putnam et al. 1995). In this type of cancer, malignant cells acquire metastatic traits for rapid and massive cell dissemination, followed by colonization of multiple secondary organs. Sequential metastasis to liver and lungs is often observed in colorectal cancer progression from stage III to IV, and more than 85% of recurrences are detected within the first 3 years of follow-up (Nguyen, Bos et al. 2009). Therefore, this particular type of cancer shows medium latency and aggressiveness, resulting in 65% survival at 5 years for stage III. A well-known example of a tumor type with very long latency is prostate cancer. According to statistics from the National Cancer Institute, nearly 100% of diagnosed patients survive 5 years, and 82% are still alive 15 years after diagnosis (the most recent statistics report 15 year survival rate for 94% of patients diagnosed after 1994) (Howlader N 2016). The short latency in lung cancer implies that malignant cells in the primary tumor acquire most of the metastatic traits, thus enabling them to overtake organs immediately after infiltration. However, in long latent metastasis, an interval is

needed for CTCs and DTCs to alter the functions required for tumor initiation and expansion in the secondary site. In this case, the microenvironment of the host organ plays a key role in the acquisition of these functions.

At the cellular level, latency is often considered as dormancy. The cell that enters a state of dormancy is inactive in the proliferation, whereas the size of the dormant micrometastatic lesion is unchanged for a period of time. Therefore, dormancy is a crucial trait that allows DTCs and micrometastases to survive, adapt, and colonize a distant organ in the interval of long-latent metastatic progression (Nguyen, Bos et al. 2009).

1.3 Mechanisms of metastatic dormancy

Broadly defined, tumor dormancy is an arrest in tumor growth, which may occur during the formation of primary tumors or after dissemination to distant organs. However, primary tumor dormancy and metastatic dormancy appear to be distinct processes. The latter is often explained as a result of delayed adaption of DTCs to new microenvironments (Giancotti 2013). At least three distinct mechanisms have been proposed to maintain dormancy, including cellular, angiogenic and immunological processes. All of these contribute to the dormant period and involve various factors, such as genetic traits, tumor microenvironment components, and cancer therapeutics (Osisami and Keller 2013) (Figure 3).

During metastatic dormancy, a single DTC can undergo growth arrest, which is called cellular dormancy or solitary cell dormancy. In contrast, the expansion of a dividing tumor cell population in micrometastatic lesions is antagonized by a process termed tumor mass dormancy. Cellular dormancy occurs when a DTC enters a state of quiescence accompanied by decreased expression

of proliferation marker Ki67. In contrast to mostly irreversible senescence, G0/G1 cell cycle arrest in the quiescent phase is likely to be responsible for cellular dormancy, hence cells are able to leave a dormant state and proliferation is re-activated. Cell cycle arrest can be induced in response to mitogens, stress factors or other factors present in the host organ microenvironment (Osisami and Keller 2013). To ensure their survival in the arrest phase, DTCs alter signaling pathways that coordinate metabolic homeostasis. The inhibition of the PI3K-AKT pathway is correlated with the dormancy phenotype in DTCs from breast cancer patients and dormant tumors from head and neck squamous cell carcinoma (Giancotti 2013). Factors secreted by the microenvironment, such as mesenchymal cell-derived bone morphogenetic proteins (BMP) and growth arrest-specific 6 (GAS6) produced by osteoblasts, can directly inhibit DTC proliferation. In prostate cancer bone metastasis, the secretion of BMP7 activates the metastasis suppressor gene

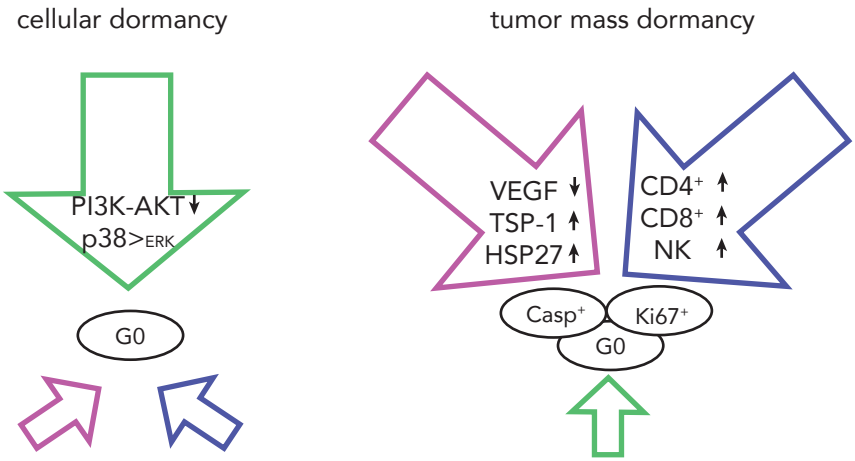

Figure 3. The mechanisms of cancer dormancy. Metastatic dormancy is induced and maintained by cellular (green), angiogenic (pink) and immune (blue) mechanisms, which contribute to dormancy in different proportions (arrow size). Solitary cell dormancy (cellular dormancy), defined as arrest in the cell cycle, is mediated by different signaling pathways. In tumor mass dormancy (right) proliferation is balanced by cell death due to lack of blood supply and immunesurveillance. Both intrinsic tumor cell traits and factors from microenviorment plays role in dormancy. G0 indicates an arrested cell, Casp+ depicts an apoptotic cell, Ki67+ states for a proliferating cell.

N-myc downstream-regulated gene 1 (*NDGR1*), thereby inducing dormancy. This subsequently leads to an increase in mitogen-activated protein kinase p38 (p38 MAPK) activation, cell cycle inhibitor p21 expression, and cell cycle arrest (Kobayashi, Okuda et al. 2011). Leukemia and prostate cancer cells often reside in the bone marrow and they are therefore sensitive to GAS6-driven dormancy (Shiozawa, Pedersen et al. 2010). Cross-talk between mitogen- and stress-induced signaling is also important for the induction of solitary cell dormancy. The extracellular signal-regulated kinase (ERK1/2) to p38 MAPK ratio regulates the cell cycle since high levels of ERK1/2 activity favor proliferation. Upon downregulation of urokinase plasminogen activator receptor (uPAR), squamous carcinoma cells enter cellular dormancy as a result of a higher ratio of the p38 MAPK over the ERK1/2. Increased p38 MAPK activity triggers the activation of the unfolded protein response (UPR), which upregulates the transcription factor ATF6, thus promoting cell arrest and survival (Aguirre-Ghiso, Liu et al. 2001; Aguirre-Ghiso, Estrada et al. 2003). In head and neck squamous cell carcinoma TGFß from stroma regulates through p38 MAPK signaling pathway the expression of genes that maintain the cell dormancy (Bragado, Estrada et al. 2013). Stromal signals from lung microenvironment also induce dormancy in metastatic breast cancer. To exit form cellular dormancy cancer cells produce Coco protein that enables them to overcome inhibitory signals from lung microenvironment and outgrowth into metastatic lesions (Gao, Chakraborty et al. 2012) Other kinases implicated to quiescence and dormancy in cultured colon, pancreatic and ovarian cancer cells are members of dual specificity tyrosine phosphorylation-regulated kinases (DYRK). In particular, DYRK1B activity support G0 arrest of cancer cells and is frequently

amplified or upregulated in ovarian cancer cells (Ewton, Hu et al. 2011; Hu, Deng et al. 2013).

In contrast to cellular dormancy caused by the arrest in the proliferation of solitary DTCs, the expansion of micrometastatic lesions can be inhibited by similar rates of proliferation and apoptosis. To a certain extent, cell growth arrest takes part in tumor mass dormancy; however, tumor cells in micrometastatic lesions usually divide (Figure 3). The proliferation to apoptosis balance is caused by restricted blood supply or an active immune system. Both processes are tightly regulated by the tumor microenvironment.

In order to grow beyond 1-2 mm, micrometastatic lesions induce vessel formation by secreting angiogenic factors, including vascular endothelial growth factor (VEGF), that attract endothelial and immune pro-angiogenic cells. However, tumor mass dormancy can be maintained by the high expression of angiogenic suppressors or the downregulation of pro-angiogenic chemokines (Osisami and Keller 2013). One well-known angiogenic inhibitor is thrombospondin-1 (TSP-1). The upregulation of this molecule in cancer leads to poor vascularization and dormancy in *in vivo* models of breast cancer, gliobastoma, osteosarcoma and liposarcoma (Lawler 2002). Moreover, TSP-1 secretion by the mature endothelium induces cellular dormancy in DTCs, thereby indicating that this factor can promote dormancy through various mechanisms (Ghajar, Peinado et al. 2013). In addition to secreted factors, chaperons, including heat shock 27 kDa protein (HSP27), can also regulate angiogenesis directly and by inducing pro-angiogenic factors. The downregulation of HSP27 protein expression in angiogenic human breast cancer cells triggers long-term *in vivo* dormancy, whereas its upregulation induces exit from dormancy and increases vascular density. Furthermore, HSP27 was shown to upregulate the secre-

tion of the angiogenic factors belonging to the VEGF family (Straume, Shimamura et al. 2012).

The third mechanism of dormancy includes the role of immune system in the clearance of tumor cells. Immunosuppressed patients develop tumors more often than healthy individuals. In line with this, tumor formation and progression is higher in immunodeficient mice than in immunocompetent counterparts (Shankaran, Ikeda et al. 2001). An intact immune system recognizes and removes tumor cells by cytolysis performed by adaptive immune cells, mainly cytotoxic T lymphocytes. During immunoediting, low immunogenic tumor cells exist in a balance with immunological clearance. The depletion of $CD4^+$ and $CD8^+$ T cells in mouse models results in escape from dormancy. These results have been supported by clinical studies showing that a lower proportion of memory T cells among the $CD4^+$ and $CD8^+$ cell populations in the bone marrow of breast cancer patients correlates with larger tumors (Feuerer, Rocha et al. 2001). In addition to immunosurveillance in primary tumors, the immune system also regulates DTC numbers and the size of micrometastatic lesions. The bone marrow of patients with breast cancer that contain dormant DTCs also shows high levels of several subpopulations of immune system cells, including natural killer cells, macrophages, and T lymphocytes. Therefore, the immune system recognizes these DTCs, and memory T lymphocytes migrate to the bone marrow to control metastatic spread (Romero, Garrido et al. 2014). Indeed, depletion of these immune cell populations increases overt metastasis and inhibition of a negative regulator and - tyrosine kinase Mer, in natural killer cells, suppresses metastasis (Bidwell, Slaney et al. 2012; Malladi, Macalinao et al. 2016).

2. Latency in ER+ breast cancer bone metastasis

The last two decades have witnessed improvements in the detection of breast cancer at early stage. Such improvements are attributed, in part, to the decrease in breast cancer-related deaths in the developed world. The detection of early-stage primary tumors is associated with good prognosis. Unfortunately, a subset of breast cancer patients is initially diagnosed with distant metastasis that is nearly always incurable. Other patients, without detectable metastasis at the time of diagnosis, will eventually recur with disease in distant organs despite receiving adjuvant therapy. This implies that nearly 30% of women initially diagnosed with early-stage disease will eventually develop metastatic lesions, often months or even years later (Redig and McAllister 2013).

2.1 Molecular portraits of breast cancer

Breast cancer is a heterogeneous set of mammary malignances, each with a distinct etiology, progression course, and outcome. Breast cancer is divided into six molecular subtypes: luminal A, luminal B, Her2-enriched, basal-like, claudin-low, and normal breast-like; however, clinical division into 4 subtypes is predominantly used (Perou, Sorlie et al. 2000; Prat, Parker et al. 2010; Cancer Genome Atlas 2012). This classification is based on molecular profiling through the gene expression analysis of distinct gene clusters: luminal, basal, proliferative, and Her2, amongst others (Table 1).

Luminal A and B subtypes express hormone receptors such estrogen (ER) and progesterone (PR) and are therefore classified in a subgroup of ER positive (ER+) breast tumors. Luminal B subtype is associated with poorer outcome and distinguished from luminal A by higher proliferative status and lower expression of PR. The gene expression pattern (luminal cluster), including cytokeratins 7,

8, and 18, of both luminal subtypes is similar to that of luminal epithelial cells in the mammary gland.

Generally, the Her2-enriched subgroup consists of Her2 positive (Her2+) tumors that overexpress human epidermal growth factor receptor-2 (Her2) through the amplification of *ERBB2* gene. Molecular profiling results in the enrichment in the Her2 and proliferation gene cluster, as well as low expression of luminal and basal clusters. In immunohistochemical analysis, these tumors are ER- or ER+ and Her2+.

High expression of basal and proliferation cluster genes and low expression of luminal and Her2 ones distinguish basal-like subtypes. On the basis of negative immunochistochemical staining of three receptors (ER, PR, Her2), these tumors are called triple negative. The expression of genes from the basal cluster is similar to that of basal epithelial cells in the mammary gland, which includes basal epithelial cytokeratins 5, 14, and 17.

The other members of ER- subgroup of breast tumors are normal breast-like and claudin-low. As these names indicate, normal breast-like subtypes have a similar gene expression pattern to normal breast tissue and adipose tissue, while claudin-low tumors express low levels of tight junction and cell-cell adhesion proteins, namely claudins.

In addition to gene expression analysis, genomic studies have identified frequently mutated genes in each breast cancer subtype. *TP53* and *PIK3CA* are mutated in different proportions in each subtype. A greater number of ER+ subtypes harbor a mutation in *PIK3CA*, while in ER- tumors *TP53* mutates more frequently. *MAP3K1* is mutated exclusively in luminal A and B subtypes, whereas *GATA3* mutation is found only in luminal A subtype. In addition, 8% of tumors classified in the Her2-enriched group show mutated *PIK3R1*

(Cancer Genome Atlas 2012) (Table 1). Molecular profiling in combination with classical clinical markers, including age, node status, tumor size, and histological grade, helps clinicians to stratify breast cancer patients and select optimal treatment.

Table 1. Summary of breast cancer subtypes characteristics

	Luminal A	Luminal B	Basal-like	Her2-enriched
mRNA Expression	High ER cluster Low proliferation	Lower ER cluster High proliferation	Basal signature High proliferation	*HER2* amplicon signature High proliferation
Copy Number	Most diploid Many with quiet genomes 1q, 8q, 8p11 gain 8p, 16q loss 11q13.3 amp (24%)	Most aneuploid Many with focal amps 1q, 8q, 8p11 gain 8p, 16q loss 11q13.3 amp (51%) 8p11.23 amp (28%)	Most aneuploid High genomic instability 1q, 10p gain 8p, 5q loss *MYC* focal gain (40%)	Most aneuploid High genomic instability 1q, 8q gain 8p loss 17q12 focal *ERRB2* amp (71%)
DNA Mutations	*PIK3CA* (49%) *TP53* (12%) *GATA3* (14%) *MAP3K1* (14%)	*TP53* (32%) *PIK3CA* (32%) *MAP3K1* (5%)	*TP53* (84%) *PIK3CA* (7%)	*TP53* (75%) *PIK3CA* (42%) *PIK3R1* (8%)
DNA Methylation	n/d	Hyper-methylated phenotype for subset	Hypo-methylated	n/d
Protein Expression	High Estrogen signaling High cMYB RPPA reactive subtypes	Less Estrogen signaling High FOXM1 and cMYC RPPA reactive subtypes	High expression of DNA repair proteins, PTEN and INPP4B loss signature (p-AKT)	High protein and phos-protein expression of HER1 and HER2

Abbreviations: ER, estrogen receptor; HER2, human epidermal growth factor receptor-2; MYC, V-Myc avian myelocytomatosis viral oncogene homolog; ERBB2, Erb-B2 receptor tyrosine kinase 2; PIK3CA, phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha; TP53, tumor protein P53; GATA3, GATA binding protein 3; MAP3K1, mitogen-activated protein kinase kinase kinase 1, E3 ubiquitin protein ligase; PIK3R1, phosphoinositide-3-kinase, regulatory subunit 1 (alpha); cMYB, V-myb avian myeloblastosis viral oncogene homolog; RPPA, reverse phase protein array; FOXM1, forkhead box M1; PTEN, phosphatase and tensin homolog; INPP4B, inositol polyphosphate-4-phosphatase, type II, 105kDa; pAKT, phospho-AKT kinase; n/d, no data.
Table adapted and modified from Cancer Genome Atlas 2012

2.2 Recurrence patterns in metastatic breast cancer

Breast cancer dissemination is a non-random, organotropic process; therefore subtypes show distinct patterns in terms of temporal course and site of relapse. Typically, ER- (basal-like, Her2-enriched, normal breast-like, and claudin-low) patients relapse within 5 years after diagnosis and disease spreads to bone and visceral organs in balanced proportions. Luminal A and B tumors metastasize predominantly to the bone in a wide time range. Post-recurrence survival depends on the site of relapse, but on average it is between 2 and 3 years (Hess, Pusztai et al. 2003).

The ER+ subtype presents strong bone-tropism since skeleton metastasis is predominant in luminal A tumors and luminal B metastases (accounting for 66% and 70% of metastases, respectively) in comparison to visceral tissues. This subtype is associat-

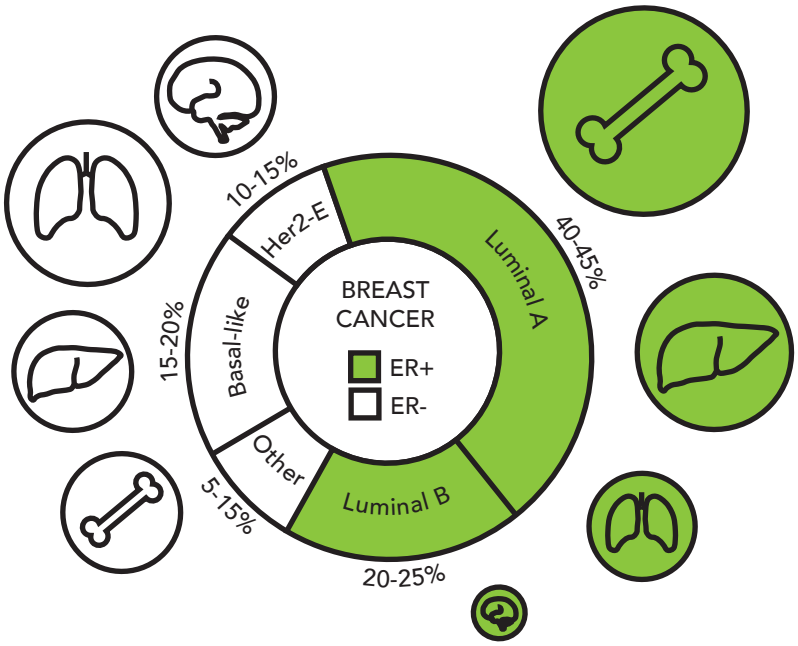

Figure 4. The organ-tropism of breast cancer metastasis. ER+ breast cancer subtypes metastases typically occur in the bone, while less frequently in the liver, lung, and brain. Metastatic spread of ER- subtypes affects more equally visceral organs (the lung, liver, and brain) and the bone.

ed with less than 5% frequency of brain metastasis, while disease spreads to the lung or liver in 25-30% of ER+ cases. In general, ER- tumors present more diverse recurrence patterns without a predominant site. Visceral metastasis occurs more frequently than in the ER+ tumors, with a significant increase in the brain metastasis, ranging from 22 to 28% of cases. Relapse in the lung is detected in 45% of ER- patients, while in the liver frequencies depend on the subtype; 41% for Her2-enriched and 21% for basal-like. ER- subtypes also vary, with bone metastasis rates being the highest for the Her2-enriched subtype (60%) and the lowest for the basal-like subtype (39%). Overall, despite the differences between subtypes, bone is the predominant site of metastasis in breast cancer (Hess, Pusztai et al. 2003; Kennecke, Yerushalmi et al. 2010).

Metastasis in breast cancer usually manifests asynchronously with the primary tumor and shows variable time to become clinically detected. This lag time depends on the volume, stage, and molecular subtype of the primary tumor. In addition to tumor stage, ER status is related to the recurrence time (Figure 5). ER- tumors are characterized by a more aggressive spread, thus recurrence peaks at around 2 years after diagnosis. However, the relapse rate diminishes to a low level 5 years after diagnosis thereforeER- subtypes are classified as short or medium latent cancer types (Hess, Pusztai et al. 2003; Early Breast Cancer Trialists' Collaborative 2005; Zhang, Giuliano et al. 2013). In contrast, the ER+ subtype has a lower risk of recurrence than the ER- subtype in the initial 5 years after diagnosis, but thereafter has a greater chronic annual risk of recurrence than the latter. Thus, more than half the metastases of ER+ tumors occur 5 years or longer after diagnosis and surgical removal of the primary tumor. Moreover, some patients suffer recurrence after more than 20 years (Hess, Pusztai et al. 2003; Early Breast Cancer

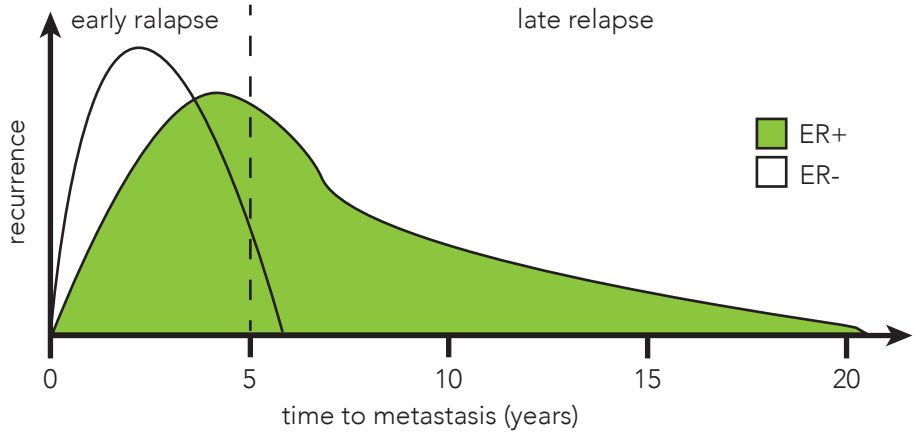

Figure 5. The temporal course of breast cancer metastasis. ER- breast cancer subtypes metastases typically occur within 5 years after primary tumor diagnosis whereas ER+ can relapse early (before 5 years) or late, up to decades after initial diagnosis. Dashed line indicates clinical threshold for early and late relapse.

Trialists' Collaborative 2005). In addition, luminal subtypes show higher rates of heterogeneity during the course of metastasis. In this regard, some patients will develop metastasis shortly after diagnosis and others after long latency. Strikingly, 15-year recurrence and mortality rates for ER- and ER+ subtypes are similar, reaching up to 45% recurrence and 30% mortality in patients diagnosed at early stage (Goss and Chambers 2010). Late recurrence decades after the initial diagnosis indicates a long latency in ER+ breast cancer metastatic progression. However, metastasis in some ER+ patients progresses rapidly, implying broad heterogeneity in recurrence patterns.

2.3 Mechanisms of breast cancer metastatic colonization of the bone

Once detected, bone metastasis is incurable and severely reduces the quality of life of breast cancer patients. The most common symptoms of bone metastasis include bone pain, pathological fractures, hypercalcemia, and nerve compression. Metastatic can-

cer cells disrupt bone physiology, hematopoiesis, and the immune system, thus compromising general homeostasis and resulting in patient death. Generally, metastatic lesions are detected in the heavily vascularized areas of the skeleton, such as the red marrow of the long bones, sternum, pelvis, ribs, and vertebrae. Bone lesions are formed as a result of alterations in physiological bone remodeling processes, namely bone destruction and formation. Most breast cancer patients with metastatic disease form osteolytic lesions as a result of increased bone resorption, while osteoblastic lesions, characterized by excess bone formation, occur less frequently (Kozlow and Guise 2005; Weilbaecher, Guise et al. 2011)

The process of bone colonization starts by pre-metastatic niche formation, before the arrival of metastatic CTCs. The primary tumor conditions the bone marrow by secreting soluble factors that target cells in the bone microenvironment (Weilbaecher, Guise et al. 2011). Molecules such heparanase, osteopontin, and lysyl oxidase facilitate the invasion, survival, and proliferation of metastatic breast cancer cells (Kelly, Suva et al. 2005; Anborgh, Mutrie et al. 2010; Cox, Rumney et al. 2015).

After extravasation, metastatic DTCs may occupy various native niches in the bone, including hematopoietic stem cell (HSCs), osteogenic and perivascular to benefit from physiological signals promoting cell survival in the new environment (Figure 6).

The perivascular niche is localized around blood capillaries and, depending on the activity of endothelium, it secrets tumor-suppressive or -promoting signals. DTCs localized near mature vessels are usually maintained in a quiescent state by endothelium-derived TSP-1, which is a potent tumor suppressor (Ghajar, Peinado et al. 2013). As a result of neovascular sprouting, which disrupts vessel homeostasis, endothelial cells release more tumor-promoting sig-

nals such as ECM molecules and growth factors, including periostin and active TGFß, which drive micrometastatic outgrowth (Ghajar, Peinado et al. 2013) (Figure 6)

Since the bone marrow is permissive for the homing and residence of HSCs, the HSC niche seems to be a protective environment for DTCs because it provides pro-survival chemokines that sustain the viability of metastatic cells (Yoneda 2000). For example, secretion of stromal cell-derived factor 1 (SDF1; also known as CXCL12) by mesenchymal cells in the bone marrow promotes the survival of metastatic cells that express C-X-C chemokine receptor 4 (CXCR4) (Kang, Siegel et al. 2003). Moreover, overexpression of the tyrosine kinase SRC in DTCs amplifies PI3K-AKT pathway activation induced by CXCR4 (Zhang, Wang et al. 2009) (Figure 6)

The third niche that DTCs occupy is the osteogenic niche, where interactions with the stoma enhance mTOR activity and drive

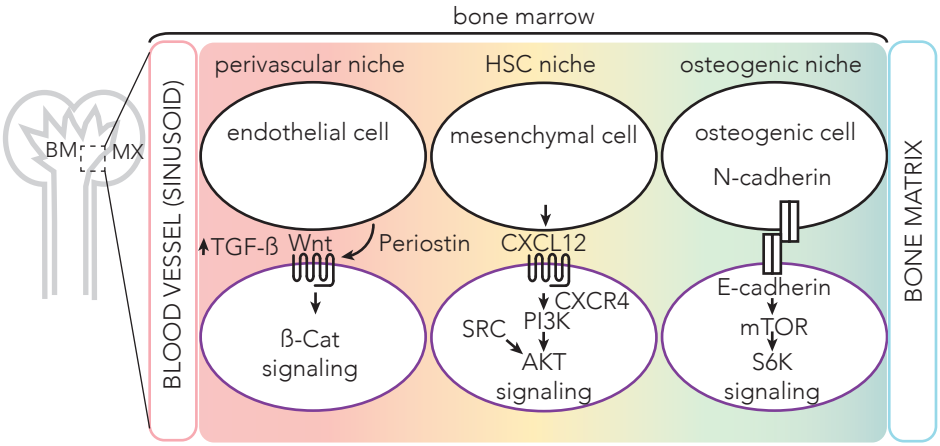

Figure 6. The metastatic niches in the bone marrow. In order to survive, DTC reside in different compartments in the bone marrow, so called niches. Niche occupancy depends of the pro-survival interactions between cancer cells (purple) and host stroma (black). In perivascular niche, sprouting neovasculature promotes micrometastatic outgrowth by secretion of periostin and Wnt signaling activation. Mesenchymal cells in HSC niche release chemokine CXCL12 that, by CXCR4 receptor binding in cancer cells, signals through PI3K-AKT pathway. SRC kinase alterations in DTC amplify AKT pathway activation. DTC can also interact with osteogenic cells by formation of heterotypic adherens junctions between N-cadherin and expressed by cancer cells E-cadherin. Therefore, mTOR signaling is activated promoting tumor growth. Figure modified from (ref). Abbreviations are explained in the main text, except from BM, bone marrow; BX, bone matrix.

progression from single cells to micrometastases prior to osteolysis (Wang, Yu et al. 2015) (Figure 6). These distinct mechanisms that metastatic cells use to survive in the bone microenvironment and to exit dormancy reflect the heterogeneity of metastatic populations. Niche occupancy depends on the traits that cells acquire during the metastatic cascade, followed by the interactions between tumor and host cells. Therefore, DTCs home to the bone, start to proliferate (often after a period of dormancy), form micrometastatic lesions, and finally induce vicious cycles of bone lysis and tumor growth.

In the final phase of metastatic colonization of breast cancer, cancer cells control the bone microenvironment to activate osteoclasts and suppress bone formation. This is achieved by paracrine crosstalk among cancer cells, osteoblasts, osteoclasts, and the bone matrix (Figure 7). Cancer cells secrete osteolytic factors that activate bone-resorbing osteoclasts. To activate osteoblasts, metastatic cells produce cytokines and growth factors, including parathyroid hormone-like protein (PTHrP), interleukin (IL)-11, IL-6, IL-8, VEGF, and tumor necrosis factor (TNF-α). As a result, osteoblasts release soluble receptor activator of nuclear factor kappa-B ligand (RANKL) and inactivate its antagonist osteoprotegerin (OPG). The ratio of RANKL to OPG is critical for osteoclast activation since OPG prevents RANKL from binding to its receptor, RANK. Once activated upon ligand binding, the multinucleated osteoclasts attach to the bone surface and release acid and proteolytic enzymes, such as cathepsin K and matrix metalloproteinases (MMP) to resorb the bone matrix. Osteolysis releases growth factors stored in the matrix, including TGFß, insulin-like growth factors (IGFs), and BMPs, as well as calcium ions, into the bone microenvironment. In addition to tumor growth enhancement, in metastatic cells, TGFß

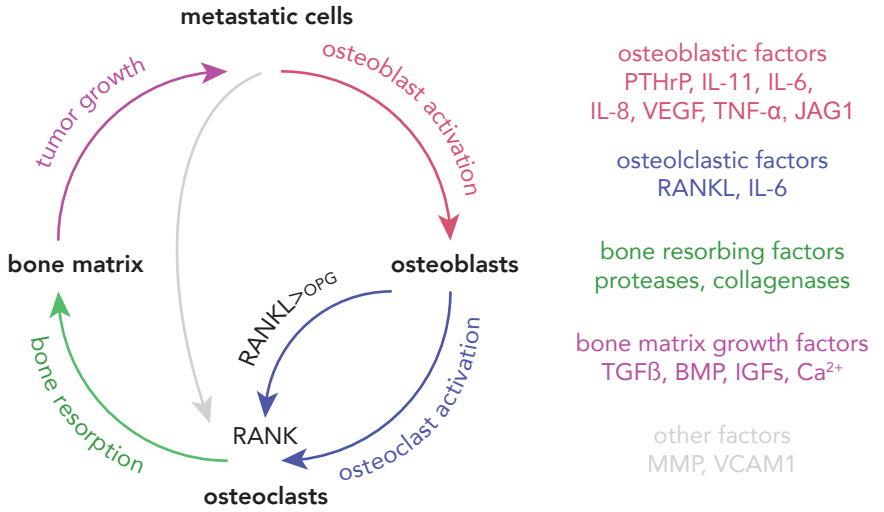

Figure 7. The vicious cycle of bone metastasis. During the osteolitic cycle, tumor cells in the bone microenvironment produce molecules that stimulate osteoclastic bone resorption directly, or indirectly through osteoblast activation by osteoblastic factors (red). Activated osteoblasts secret osteoclastic factors (blue) mainly RANKL to promote bone degradation (green). The consequence of increased resorption is the release of growth factors from the bone matrix (pink) that feed back to the tumor cells, further stimulating their growth. Abbreviations are explained in the main text, except from $Ca2^{+}$ that stands for calcium ions.

activates both Smad-dependent and Smad-independent signal pathways to induce PTHrP. Therefore, tumor growth is stimulated, leading to the production of additional osteolytic and osteoblastic factors and resulting in the vicious cycle of bone metastasis (Kozlow and Guise 2005; Weilbaecher, Guise et al. 2011). In addition, bone resorption can be promoted by the Notch signaling pathway, which results in IL-6 secretion upon binding of tumor-derived JAGGED-1 (JAG-1) to osteoblasts (Sethi, Dai et al. 2011)

2.4 Origins of breast cancer metastasis-initiating cells

Breast cancer metastases originate from DTCs that homed to the bone and activated cycle of bone lysis and tumor growth. Therefore, a DTC capable of reinitiating a macroscopic tumor in a distant tissue is called a metastasis-initiating cancer cell (MICs). MICs can be considered tumor-initiating cells (TICs) or cancer stem

cells (CSCs) due to their ability to undergo self-renewal and differentiation, thus initiating a new tumor and giving rise to all types of tumor cells. However, although MICs and CSCs exhibit stem cell properties, they do not necessarily originate from the transformation of normal tissue stem cells. Therefore, the cell from which they derive may also be a more mature cell or progenitor that has undergone transformation. Although the frequency of TICs or CSCs may vary among tumor types, these cells often represent a minor subset of tumor cells endowed exclusively with tumor-initiating capacity (Visvader and Lindeman 2012).

The tumor-initiating cells of breast cancer were identified using cell-surface markers EpCAM, CD24, and CD44 from human solid tumors and pleural effusion samples (Al-Hajj, Wicha et al. 2003). The EpCAM+/$CD24^{-}$/$CD44^{+}$ breast cancer cells show self-renewal ability and primary tumor-initiating capacity when injected into the mammary fat pad of immunodeficient mice, resulting in mammary tumor formation and progression. Furthermore, tumors generated from this population recapitulated the heterogeneity of the initial tumor with a small fraction of cells expressing the same stem-like signature and a majority of differentiated, non-tumorigenic progeny (Al-Hajj, Wicha et al. 2003). In ER- breast cancer subtypes, tumor-initiating cells can be found in both $CD24^{-}$/$CD44^{+}$ and $CD24^{+}$/$CD44^{+}$ cell populations (Meyer, Fleming et al. 2010). Another widely used breast cancer stem cell marker is CD49f. Due to its high expression on TICs, CD49f is associated with poor survival of breast cancer patients (Friedrichs, Ruiz et al. 1995) and, in combination with the $CD24^{-}$/$CD44^{+}$ cell population, enhances tumor-initiating capacity (Meyer, Fleming et al. 2010).

The $CD24^{-}$/$CD44^{+}$ signature enhances not only primary tumor-initiating capacities, but also cancer cell ability to metastasize,

since metastatic cells isolated from the bone marrow of early-stage breast cancer patients are highly enriched for this signature (Balic, Lin et al. 2006). Noteworthy, the relative percentage of tumor-initiating cells observed in bone marrow metastases is higher than in primary tumors (65% vs. 10%). TICs isolated from breast tumors can also spontaneously form metastases when implanted in immumodeficient mice (Sheridan, Kishimoto et al. 2006; Liu, Patel et al. 2010). In addition, metastasis-free and overall survival of breast cancer patients was predicted using a 186-gene invasiveness signature generated from $CD44^{+}/CD24^{-}$ cell population (Liu, Wang et al. 2007). Therefore CD24, CD44 and CD49f may serve as markers for MICs in breast cancer.

At the molecular level, adhesion molecule and EMC receptor, CD44, modulates diverse cellular signaling pathways to regulate cell proliferation, migration, and invasion (Luo, Clouthier et al. 2015). The expression of CD44 on breast cancer cells promotes EMT, cancer cell adhesion to bone marrow endothelial cells, and bone metastasis. Moreover, high CD24 expression has been shown to suppress CXCR4 activity, which is involved in breast cancer cell metastasis to bone. Therefore, low CD24 expression enhances CXCL12/CXCR4 signaling activity and facilitates bone homing. The $CD24^{-}/CD44^{+}$ population expresses high levels of factors associated with invasion and bone metastasis, including IL-6, IL-8, and urokinase plasminogen activator (Luo, Clouthier et al. 2015).

3. Mitogen and stress-activated kinase signaling

Cells respond to changes in the environment using signaling cascades that begin with signal recognition by a transmembrane receptor, follow through kinase cascade and end in the nucleus, where gene transcription is activated or repressed. The ERK and p38 MAPK kinases regulate the complex intracellular signaling events contributing to gene expression in a cell-type and stimulus-specific manner. Mitogen- and stress-activated kinase (MSK) is a common downstream target of these two signaling pathways that contributes to the regulation of gene expression in the nucleus.

3.1 Characterization and activity of MSKs

Mammalian cells express two isoforms of MSKs, termed MSK1 and MSK2, encoded by *RPS6KA5* and *RPS6KA4* genes, respectively. These belong to the AGC subfamily of protein kinases that comprises 60 members. MSKs are regulatory enzymes that change the properties of a substrate by attaching a phosphate group from ATP to serine or threonine residues (Pearce, Komander et al. 2010).

Similar to their structural homologs 90-kDa ribosomal S6 kinases (RSK), MSKs contain two kinase domains in a single polypeptide, namely the N-terminal and C-terminal kinase domain (NTKD) and (CTKD) respectively, connected by a linker. The inactive conformation of MSK prevents ATP binding and substrate phosphorylation. The MSK activation cascade starts with ERK- or p38 MAPK-mediated phosphorylation of three residues of CTKD, leading to its activation. The CTKD then phosphorylates and activates the N-terminal kinase domain. Following autophosphorylation, the MSK conformation allows ATP binding and alters catalytic activity through NTKD, resulting in the phosphorylation of downstream tar-

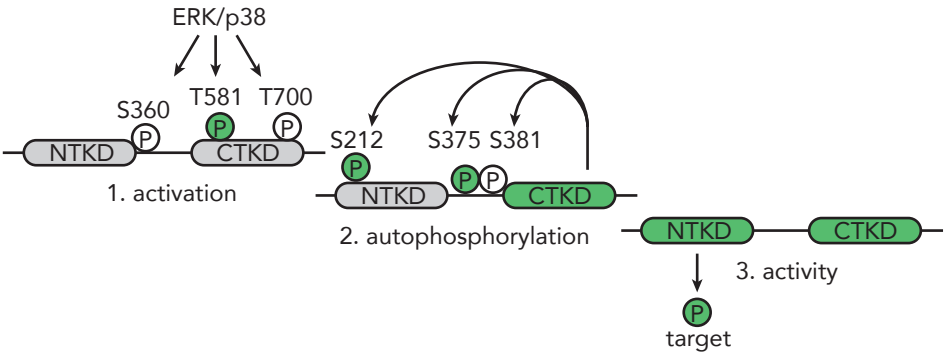

Figure 8. The MSK activation cascade. A C-terminal docking domain assures binding of the activating ERK and p38 kinases and phosphorylation of serine 360, threonine 581 and 700 residues. Phosphorylation of Thr581, which is located within the C-terminal kinase domain (CTKD), is essential in the activation of MSK1 (green), whereas mutation of Ser360 only partly reduces MSK1 activity. The activated CTKD phosphorylates serines 212, 376 and 381. Among these, S212 and S376 are essential for MSK1 activity (green). Finally, the active N-terminal kinase domain (NTKD) is responsible for phosphorylation of MSK substrates. Figure modified from Pearce, Komander et al. 2010.

gets. (Vermeulen, Vanden Berghe et al. 2009; Pearce, Komander et al. 2010) (Figure 8).

Because of the dual activation mode of ERK and p38MAPK, MSKs are able to integrate mitogenic signals with pro-inflammatory cytokines and cellular stress (Anjum and Blenis 2008). In general, growth factors like EGF, steroid hormones (estrogen, progestin), chemokines and some neurotransmitters activate MSK through the ERK pathway, whereas stress factors (hydrogen peroxide, anisomycin and UV-irradiation) signal through the p38 MAPK pathway. Other agents, including the cytokine TNFα, use both MAPK pathways (Anjum and Blenis 2008) (Figure 9). MSKs are effectively inhibited by SB747651A inhibitor (Naqvi, Macdonald et al. 2012). Unfortunately, this molecule inhibits four other kinases, namely PRK2 (double-stranded-RNA-dependent protein kinase 2), RSK1, p70S6 kinase 1, and ROCKII (Rho-associated protein kinase 2), with a similar potency to that shown over MSK. Another way to inhibit MSK activity is to block the upstream kinases. The inhibitor of mitogen-activated protein kinase kinase (MEK) U0126 prevents ERK

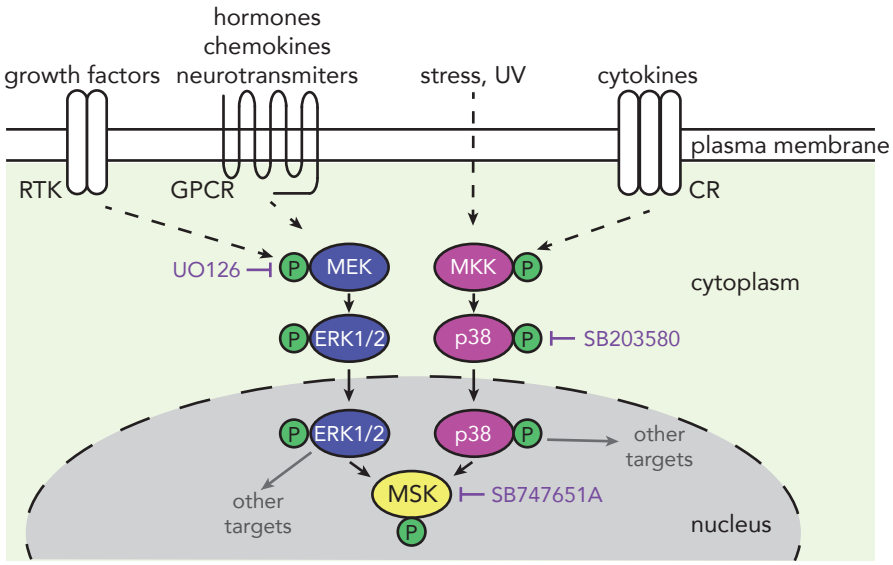

Figure 9. The MSK activation pathways. Extracellular ligans mainly mitogens bind to their receptors including receptor tyrosine kinase (RTK) and G protein-coupled receptors (GPCR). Ligand-binding initiates activation cascade leading to sequential phosphorylation of MEK and ERK kinases. External stress and cytokines induce signaling by phosphorylation of MKK and p38. Both ERK and p38 must be translocated to the nucleus, where they activate MSK and other targets. MSK activity can be inhibited using direct inhibitor or combination of compounds (violet). Dashed line indicades indirect interactions. Abbreviations: MEK, mitogen-activated protein kinase kinase; ERK, extracellular signal-regulated kinase; MSK, mitogen and stress-activated protein kinase; CR, cytokine receptor. Figure modified from Anjum and Blenis 2008.

activation, whereas stress-induced signaling is blocked by the p38 MAPK inhibitor SB203580. Therefore, both inhibitors are required to block MSK activity by adding a number of additionally inhibited MSK-unspecific targets (Anjum and Blenis 2008; Naqvi, Macdonald et al. 2012).

3.2 Regulation of cellular processes by MSKs

MSK 1 and 2 regulate various cellular functions by phosphorylating and activating downstream targets. A number of MSK functions can be deduced from the nature of its substrates, pointing towards roles in transcriptional regulation, cell cycle regulation, and cell survival in response to extracellular mitogenic and stress signals (Table 2).

Transcriptional regulation is the most prominent role of MSK, since it directly phosphorylates and activates transcription factors or modulates chromatin status. Transcription factor cAMP response element binding (CREB) was the first MSK substrate identified. Following activation, phosphorylation at serine 133 residue of CREB promotes the transcription of several genes including *c-FOS* and neuropeptides (Vermeulen, Vanden Berghe et al. 2009). Although MSK1 was shown to be the most efficient molecule in activating CREB, RSK can also phosphorylate S133 in response to mitogens since in Msk1/2 double knock-out (KO) fibroblasts CREB phosphorylation is partially reduced (Wiggin, Soloaga et al. 2002). In addition, MSKs are also required for the phosphorylation of activating transcription factor 1 (ATF1) - a member of the CRE-binding factors - at serine 63. This phosphorylation is necessary for full transcriptional activation of ATF1 (Wiggin, Soloaga et al. 2002). Another MSK target is p65, a member of nuclear factor kappa-light-chain-enhancer of activated B cells (NFκB) protein complex. This transcription complex regulates many important genes for cell survival, proliferation, synaptic plasticity and immune response. Signal transducer and activator of transcription 3 (STAT3) is also phosphorylated by MSK1 and other kinases, including c-Jun N-terminal kinase (JNK), SRC and other MAPK signaling pathway members (Zhang, Liu et al. 2001). Although MSK1-mediated phosphorylation of STAT3 at serine 727 is important, tyrosine 705 phosphorylation is required for its full activation (Vermeulen, Vanden Berghe et al. 2009). These examples show that MSK does not exclusively promote activation of transcription factors in response to mitogens and stress, thereby allowing cells to adjust to environmental conditions.

Target genes of MSK-activated transcription factors include immediate-early genes (IEGs) that, upon extracellular stimulation,

are induced rapidly but transiently. Once expressed, genes such *c-FOS* and *JUN* contribute to the expression of late genes and modulate cellular response. IEG activation depends on STAT3; however, rapid transcription occurs mainly as a result of chromatin modification in promoter region named nucleosomal response (Sawicka and Seiser 2012). This term describes the rapid and transient phosphorylation of histone H3 at serine 10 (H3S10) and high mobility group nucleosome binding domain 1 (HMGN1) at serine 6, which accompanies the transcriptional activation of immediate early genes upon stimulation with growth factors. This specific modification is restricted to the gene promoters. Studies using fibroblasts derived from Msk1/Msk2 double KO embryos showed that these are the major active kinases phosphorylating histone H3 at S10 and S28 outside mitosis (Soloaga, Thomson et al. 2003; Dyson, Thomson et al. 2005). Therefore, MSKs drive nucleosomal response to mitogens and stress by phosphorylating HMGN1 and histone H3 (Louie, Gloor et al. 2000).

In addition to IEG regulation, histone 3 phosphorylation, especially at S28, is involved in transcription activation of Polycomb group (PcG)-repressed genes. The proteins of PcG family are epigenetic regulators of gene expression that are crucial for cell-fate decisions. Through binding to gene promoter followed by trimethylation of histone H3 lysine 27 (H3K27), PcG proteins repress transcription (Simon and Kingston 2009). MSK1/2-mediated H3S28 phosphorylation displaces silencing PcG complexes, thereby revealing another example of MSK involvement in gene transcription at the epigenetic level (Gehani, Agrawal-Singh et al. 2010; Lau and Cheung 2011).

Despite the regulatory role of histone H3 phosphorylation during interphase, global phosphorylation is required for chroma-

tin condensation and proper chromosome segregation in mitosis (Sawicka and Seiser 2012). These processes are driven mainly by the Aurora kinases, and the activity of these enzymes plays an essential role in cell division (Crosio, Fimia et al. 2002).

Table 2. MSK1/2 targets

target	site	function	other kinases	reference to MSK function
Transcription factors				
CREB	S133	promotes cell growth and survival	RSK, PKA (De Cesare, Jacquot et al. 1998)	(Wiggin, Soloaga et al. 2002)
ATF1	S63	promotes cell growth and survival	CDK3 (Zheng, Cho et al. 2008)	(Wiggin, Soloaga et al. 2002)
P65	S276	activation of NF-kB, proliferation	PKA (Zhong, SuYang et al. 1997)	(Vermeulen, De Wilde et al. 2003)
STAT3	S727	proliferation, differentiation	JNK, SRC, MAPK (Zhang, Liu et al. 2001)	(Zhang, Liu et al. 2001)
Chromatin components				
H3	S10	mitosis, transcriptional activation	Aurora A, Aurora B (Crosio, Fimia et al. 2002)	(Soloaga, Thomson et al. 2003)
	S28	transcriptional activation, mitosis	Aurora B (Goto, Yasui et al. 2002)	(Soloaga, Thomson et al. 2003)
HMGN1	S6	transcriptional activation		(Louie, Gloor et al. 2000)
others				
Bad	S112	apoptosis suppression	RSK (Shimamura, Ballif et al. 2000)	(Clark, McDade et al. 2007)

Abbreviations: CREB, cAMP response element-binding; ATF1, activating transcription factor 1; p65, transcription factor p65; STAT3, signal transducer and activator of transcription 3; H3, histone H3; HMGN1, high mobility group nucleosome binding domain 1; Bad, BCL2-associated agonist of cell death; RSK, ribosomal S6 kinase; PKA, protein kinase A; CDK3, cyclin-dependent kinase 3; JNK, c-Jun N-terminal kinase; MAPK, mitogen-activated protein kinase;
Table adapted and modified from Vermeulen, Vanden Berghe et al. 2009

3.3 The role of MSK in cancer progression

MSK1 is infrequently altered at the genomic level across different cancer types. In breast cancer, genomic alteration frequency ranges between 1-3% (cBioPortal). However, MSK1 activity can be deregulated by alterations in upstream kinases belonging to RAS-MAPK signaling pathway. HER2/EGFR overexpression leads to the constitutive activation of the RAS-MAPK signaling pathway, which subsequently results in the phosphorylation and activation of the MSK proteins.

In particular, alterations in the activity of two main activators of MSK, ERK1/2 and p38 MAPK, are implicated in breast cancer. ERK1/2 is hyper-activated in a large subset of breast tumors, and its high activity correlates with poor patient survival (Mueller, Flury et al. 2000). ERK1/2 also promotes mammary tumor progression in xenograft models modulating phosphorylation and degradation of tumor suppressor - forkhead box 3a (FOXO3a)(Yang, Zong et al. 2008). During liver colonization increased activity of ERK2 drives colorectal cancer cell growth, however activity of ERK2 is dispensable for lung metastases. In order to colonize lungs p38 MAPK activity need to be reduced in colon cancer cells (Urosevic, Garcia-Albeniz et al. 2014). Interestingly, p38 MAPK function in breast cancer is very complex. On the one hand, p38 MAPK has tumor-suppressive functions since the inactivation of p38 facilitates HER2-induced mammary tumorigenesis *in vivo* and 18% of human primary breast tumors display amplification in p38 MAPK phosphatase - Wip1/PPM1D (Bulavin, Phillips et al. 2004). Nevertheless, p38 MAPK activity increases in tamoxifen-resistant xenograft and patient breast tumors (Gutierrez, Detre et al. 2005). Moreover, the cross-talk between activity of ERK1/2 and p38 MAPK has been pro-

posed to play role in cancer dissemination and cellular dormancy (Aguirre-Ghiso, Estrada et al. 2003).

Remarkably, MSK has been involved in the nuclear receptors signaling such as steroid hormone receptors ER and PR. One of the well-characterized example of transcriptional control by nuclear receptors is the promoter of the mouse mammary tumor virus (MMTV). PR forms a complex with ERK1/2 and MSK1, and this complex upon phosphorylation is rapidly recruited to the MMTV promoter. Crucially, PR/ERK/MSK1 complex is phosphorylated by ERK1/2 and not by MSK1 (Vicent, Ballare et al. 2006).

OBJECTIVES

1. Hypothesis

Long latency in ER+ breast cancer to bone metastasis remains a poorly understood phenomenon. We hypothesize that certain genes endow dormancy in metastatic cells. The exit from dormancy followed by metastatic relapse are consequences of reduced expression or inactivation of these genes. Moreover, alterations in gene expression or signaling pathway activity of dormancy-controllers associate with risk of early or late relapse in breast cancer patients. Therefore, they could be used to identify a subset of ER+ breast cancer patients with high risk of early or late relapse. In addition, identifying molecular mechanisms responsible for long-latency could provide new therapeutic opportunities.

2. Aims

To develop and characterize an animal model that mimics metastatic patterns of long-latent ER+ breast cancer patients.

To identify genes that contribute to long latency using loss-of-function shRNA *in vivo* screening system.

To select and validate dormancy-controlling genes through bioinformatics analysis of breast cancer patient datasets.

To functionally validate a candidate gene using shRNA-independent method of gene depletion.

To characterize at cellular and biochemical level the role of the newly identified gene in the control of dormancy.

To dissect the molecular mechanisms involved in the metastatic dormancy.

RESULTS

1. Experimental design to identify molecular mechanisms involved in latency of ER+ breast cancer

To elucidate mechanisms of long-latent ER+ breast cancer metastasis to the bone we aimed to design systemic and unbiased experimental workflow which will allow to identify dormancy controlling genes. Indeed, the project consisted of five main steps (i) characterization of model of dormancy, (ii) set-up of the screen, (iii) selection of screening target gene, (iv) validation of the target gene, and (v) identification of molecular mechanism involved in dormancy (Figure 10). Each step of this process was designed, based on several premises, to increase the robustness and relevance of the whole study. This chapter will briefly describe the principles of each phase of the project, whereas all detailed results are presented in following sections.

The first step was to identify an animal model that mimic complexity of human disease and allow the experimental manipulations necessary for the screening (Figure 10 step 1). Crucially, the experimental system had to be suitable for the whole genome screening, reproducible, relatively short, effective, and affordable. Currently, metastasis researchers lack genetically engineered mouse models (GEMM) of ER+ breast cancer that metastasizes to the bone, an in addition no model of latency is available (Eckhardt, Francis et al. 2012). Although in GEMM primary tumors develop with high penetrance this process takes months leaving often little time for metastasis development (Eckhardt, Francis et al. 2012). Moreover, *in vivo* genome-wide screening on primary tumors is less feasible in comparison to screening on cell lines and the xenograft models were used for large-scale screenings that identified breast cancer metastasis suppressor and enhancers (Gao, Chakraborty et

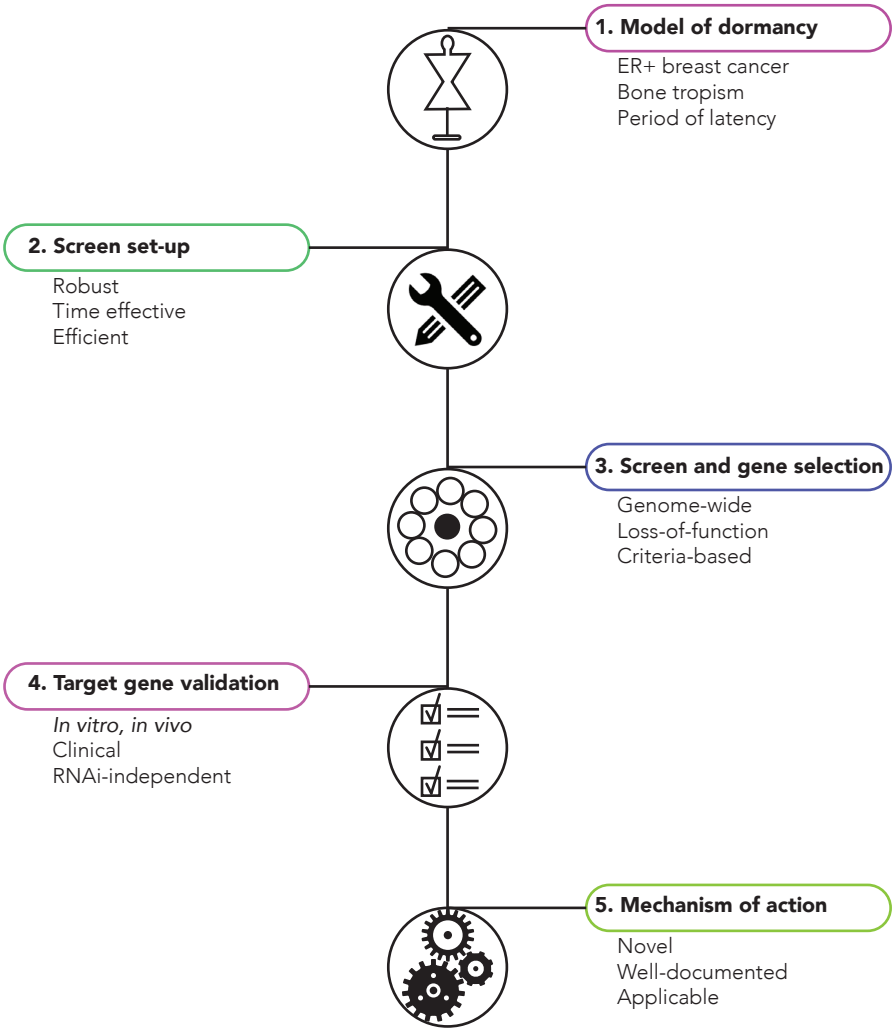

Figure 10. An overview of the project. Experimental workflow consist of five major parts, which have to be completed sequentially. (**1**) Part one focuses on the characterizing a model of long-latent ER+ breast cancer. (**2**) The main aim of the second part is optimization of shRNA loss-of-function screen. (**3**) Part three consists of genome-wide screen conducting and selecting the candidate gene. (**4**) Validation of a candidate gene is an aim of part four. (**5**) Part five will provide information about the mechanisms that are involved in the control of dormancy by the gene revealed in the screen.

al. 2012; Gao, Chakraborty et al. 2014; Murugaesu, Iravani et al. 2014). Lastly, both GEMM and transplantable mouse tumor system (syngeneic) are limited to study biology of mouse metastatic cells. Therefore, we decided to use the T47D human ER+ cell line derived from pleural fluids of patients with advanced metastatic disease (described in section 2 on page 42).

Next, we aimed to integrate RNAi technology and massive parallel sequencing in a our model of dormant metastasis to discover and validate genes involved in ER+ breast cancer metastasis to bone (Figure 10 step 2). To this end, we designed and optimized the screening platform based on several attributes. For example, a large number of cells can be transduced with pooled shRNA lentiviral library allowing whole genome screen. Moreover, intracardiac injection is time effective procedure in which we could inject more than 100 mice in 2 days. We used a bioluminescence imaging (BLI) and cell-tagging to facilitate monitoring of metastatic progression and cancer cell purification. Finally, we optimized the next-generation sequencing conditions that provide sensitive tool to deconvolute shRNA enriched by screening pressure in large number of samples (described in section 3.1 on page 54).

The main aim of the screen was to identify genes that positively regulate dormancy of micrometastatic lesions (Figure 10 step 3). The screening strategy was designed based on several principals that allow identifying dormancy enhancers or metastasis suppressors in an unbiased manner. First, the screen was genome-wide. The lentiviral library consists of shRNA targeting 16019 of human genes, representing about 85% of protein-coding human genome (Ezkurdia, Juan et al. 2014). Second, screen was based on the loss-of-function approach due to the use of RNAi-mediated gene silencing to identify targets. In this case the gene knockdown leads

to a detectable phenotype manifested as metastatic outgrowth. Next, screen was executed in the pooled format, in which the library was randomly introduced into cells at high representation of each shRNA and a relatively low multiplicity of infection (MOI). The high representation ensures that every shRNA present in the library has a chance to integrate, and the low MOI implies that the majority of cells receive only one shRNA. Following, our approach used positive and specific selection as a tool to identify targets. Metastatic bone colonization process served as a strong selective pressure resulting in a small number of cells that are able to survive, home and proliferate in the bone. Additionally, by next generation sequencing we were looking for the enrichment in shRNAs relative proportion of different integrated constructs in the cell population. Lastly, screening was design to use *in vivo* approach, because the libraries were screened for shRNAs that become depleted in mice during latent metastasis to bone. Moreover, in order to select the target gene from the list of hits we created a list of criteria (described in section 3.2 on page 58).

The fourth step of the project was to validate the relevance of the target gene in the dormancy of ER+ breast cancer (Figure 10 step 4). This step was difficult to plan and design ahead because of unknown identity of the target gene. However, we considered to functionally validate the target *in vitro* (cell culture) and *in vivo* (animals), as well as, using publicly available databases of gene expression of breast cancer tumors supplemented with clinical annotations about time and site of metastasis. As well we planned to use alternative to RNAi method of protein depletion (described in sections 4.1 on page 64 and 4.2 on page 68).

The last step was fully dependent on all previous ones and as well the most time-consuming (Figure 10 step 5). Nevertheless, we

aimed to find novel molecular mechanisms implicated in dormancy in the long-latent ER+ breast cancer with potential of clinical application (described in sections 4.3 on page 71 and 4.4 on page 74).

2. Model of latency in ER+ breast cancer to bone metastasis

2.1 Isolation and characterization of latent bone metastasis cell line

Due to the lack of animal models for studying dormancy in a long-latent ER+ breast cancer to bone metastasis we have developed our own model. To this end, we isolated (from human ER+ breast cancer cell line - T47D) a dormant bone metastatic derivative (DBM) cell subpopulation. We achieved this by two rounds of *in vivo* selection in mice (Figure 11a). This method is based on the use of a bioluminescence imaging to track cancer cells in real time in whole animal in order to monitor metastasis progression. Intracardiac inoculation of the cells introduces them to the circulation and allows spontaneous formation of metastases in different organs. Because the T47D cells are poorly metastatic, only 14% of animals developed symptomatic bone metastasis 205 days post intracardiac injection (Figure 11a). Such indolent kinetics in combination with an infrequent formation of overt metastatic lesions decreases the robustness of this model. Therefore, by *in vivo* selection, we intended increasing it.

Next, we characterized the novel cell line DBM and compared it to the parental cells T47D at genomic and transcriptomic level. Comparative genomic hybridization (CGH) analysis showed few and insignificant genomic losses and no gains in DBM cell line (Figure 11b). Gene expression analysis indicates that about 7% of

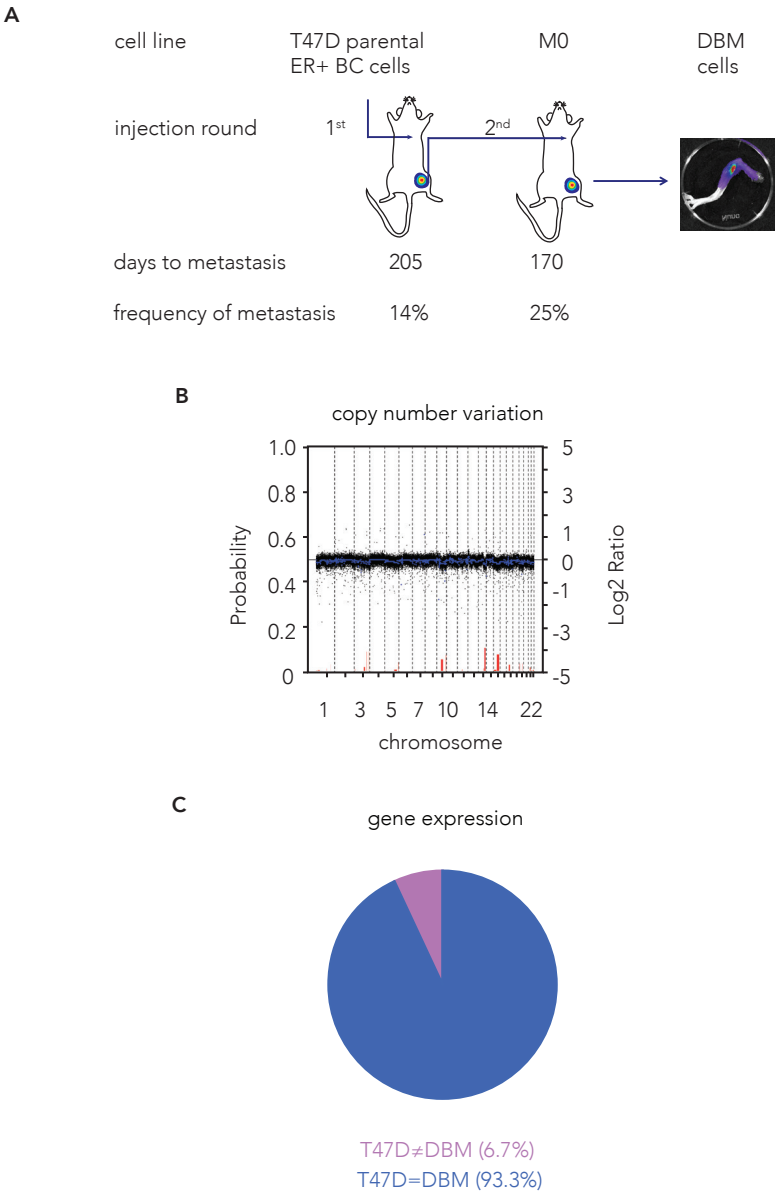

Figure 11. The Isolation of dormant bone metastatic (DBM) cell line. (**a**) Schematic of *in vivo* selection of DBM cell line. ER + breast cancer cell line T47D was injected intracardiacly into immuno-deficient mice and metastasis in hind limbs was determined by bioluminescence. (**b**) CGH analysis of DBM and parental cells. Genetic gains and losses in DBM cells, for each chromosome, are depicted with red and green color bars respectively. (**c**) Pie graph representing differentially expressed genes between parental and DBM derivative cell line.

genes are differentially expressed in DBM cells in comparison to T47D (Figure 11c). Therefore, DBM cells acquired modifications in the transcriptomic profile during *in vivo* selection process, but not as a result of gross genomic alterations.

The main aim of the first step of our study was to model the latent relapse during metastasis to the bone. Indeed, DBM cells intracardiacly injected into mice display a stable and reproducible phase of latency during the course of bone metastasis (Figure 12). This is in a sharp contrast with the other well defined ER+ models of bone-tropic breast cancer cells - BoM2 (Pavlovic, Arnal-Estape et al. 2015). We observed that, the majority (10/18) of animals injected with aggressive BoM2 cell line harbor bone metastasis 40 days post injection, while DBM-injected animals were free of symptomatic macrometastatic lesions at this time point (Figure 12a). Unfortunately, due to the poor health condition of the animals injected with BoM2 cells (7) we could not follow them up for longer than 65 days. Strikingly, the first bone metastasis from DBM cells appears 50 days after injection, already suggesting the dormancy period in the metastatic process. Interestingly, when we analyzed the growth kinetics of 8 lesions that formed overt metastasis from DBM cells, we noticed that it divides into 3 phases: bone homing, latency and metastasis (Figure 12b). During homing (first 21 days after injection) cells disseminate to the bone and proliferate, so BLI signal increases. This phase is followed by dormancy when previously formed micrometastatic lesions size is stable. The phase of latency visualized by plateau of the growth curve, is flanked by entry and the exit (reactivation) period when BLI kinetics changes rapidly. Finally, dormant lesions exit the latent phase and form large and symptomatic metastases. On the contrary, BoM2 metastatic progression can be

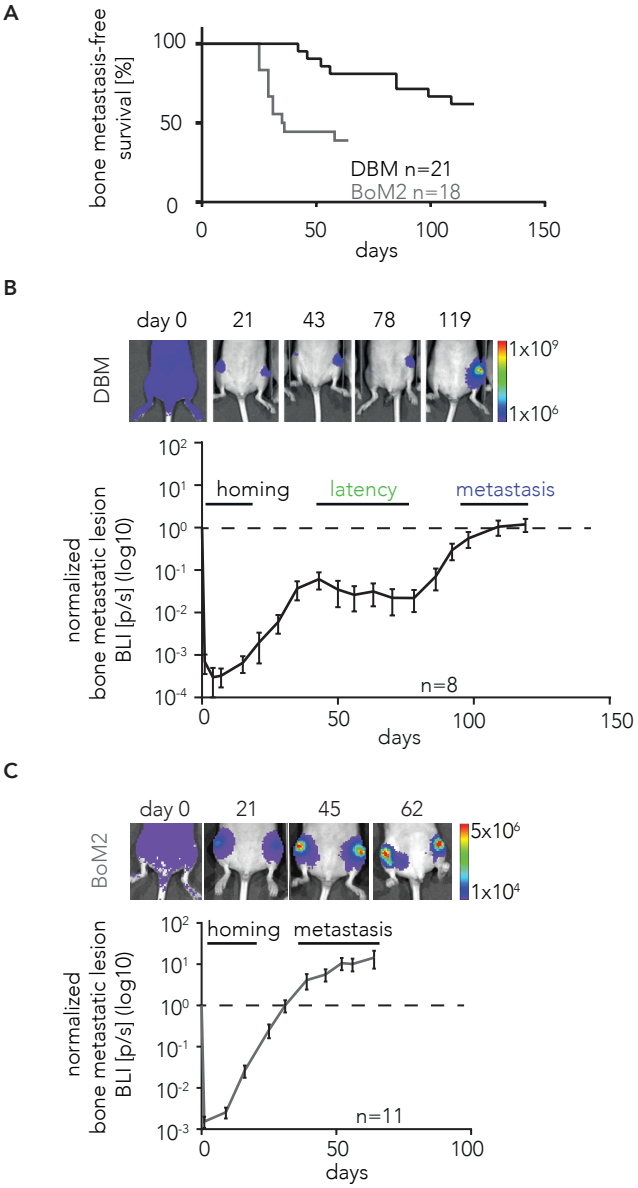

Figure 12. DBM cell line shows latency in bone metastasis. (**a**) Kaplan-Meier analysis of bone metastasis-free survival in a xenograft mouse model. (**b**) (top) Representative bioluminescence images showing bone metastasis progression in DBM xenograft mouse model (bottom) quantification of bone metastatic lesion BLI signal. (**c**) (top) Representative bioluminescence images showing bone metastasis progression in BoM2 xenograft mouse model (bottom) quantification of bone metastatic lesion BLI signal. Panels b and c show data as mean ± SEM, dashed line represents day 0 threshold.

divided into 2 parts where BLI signal increases: bone homing and metastasis (Figure 12c).

DBM cells have the potential to serve as a novel and unique model to study latency in ER+ breast cancer bone metastasis. To this end, each phase of the metastatic progression was characterized. We used a combination of live-imaging techniques with micro–computed tomography (IVIS-μCT) to visualize metastatic cells in the mouse skeleton (Figure 13). High-resolution X-ray tomography showed that disseminated DBM cells in the phase of bone homing are clinically asymptomatic since no bone loss was observed. Dormant micrometastatases in the latent phase had capacity to induce osteolysis. Nevertheless, the formation of large osteolythic lesions accelerated during metastatic outgrowth resulting in sever bone loss (Figure 13a). Importantly, throughout all metastatic process DBM cells retain high levels of expression of estrogen receptor (Figure 13b).

Taken together, DBM cells present unique prolonged bone dormancy phenotype that is opposite to the standard dynamics reported for aggressive breast to bone metastasis models BoM2, MDA-MB-231 or 4T1 (Kang, Siegel et al. 2003; Murugaesu, Iravani et al. 2014; Pavlovic, Arnal-Estape et al. 2015). Therefore, this novel model system can be used for studying mechanisms of dormancy in ER+ breast cancer.

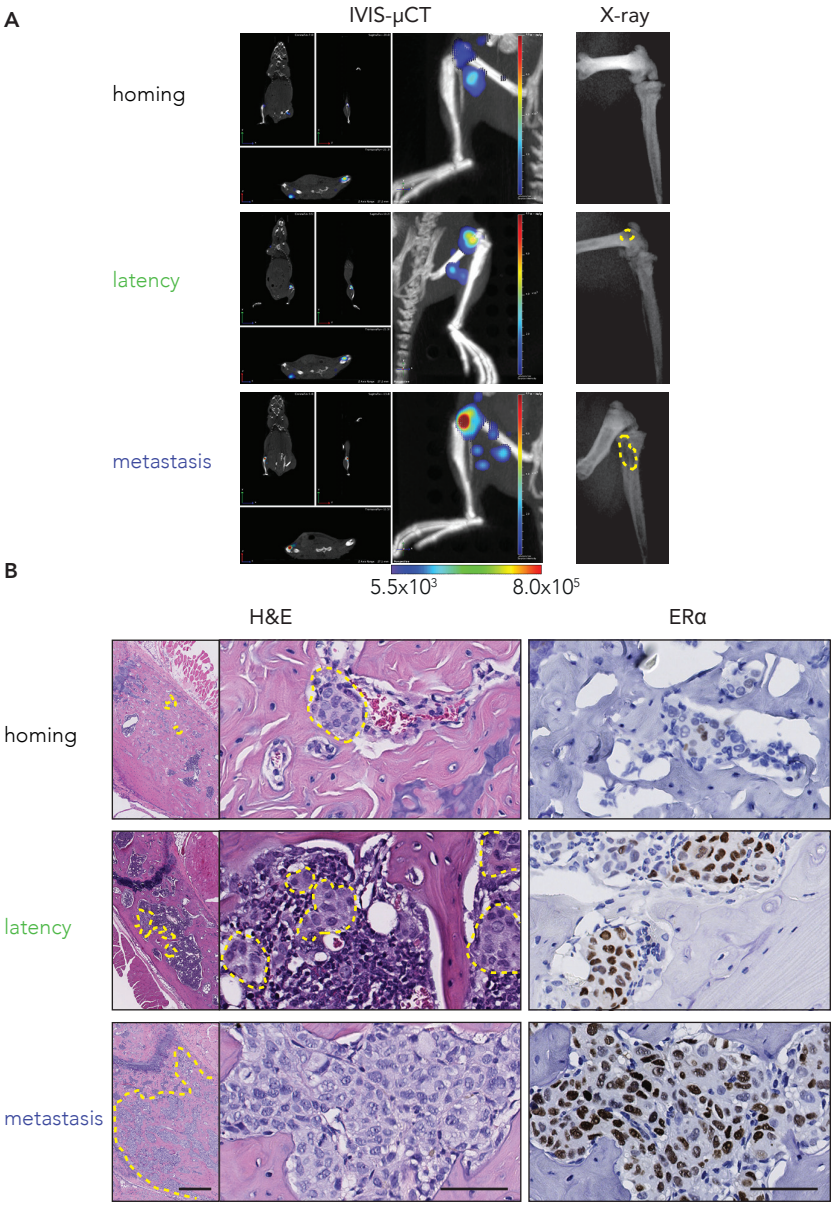

Figure 13. Characterization of DBM long-latent metastasis progression. (**a**) IVIS-CT (left) and X-ray (right) representative pictures of bone lesions in different phases of metastatic progression. (**b**) H&E and ER staining of hind limbs bones different phases of metastatic progression. Dashed line – metastatic DBM cells. Scale bars: 500 μm (small picture), 50 μm (large picture).

2.2 Cell and tumor mass dormancy in model of ER+ breast cancer to bone metastasis

As mentioned in the introduction, metastatic latency can be explained by different mechanisms of dormancy (cellular and tumor mass amongst others) with each making important contribution to the latent state. Therefore, to check how latency is regulated in our model, we investigated proliferation and cell death in lesions during the latent phase and metastasis.

First, we were interested whether during bone homing all xenografted cells proliferate or if there are dormant cells arrested in the cell cycle. In order to test this we performed label retention assay using lipophilic fluorescent dye - DiD that amount inside the cell decreases only upon cell division (Figure 14a). Figure 14b shows that all cells were labeled with DiD dye at the time of the injection (day 0). 21 days post inoculation, label-retaining cells (LRCs) were 0.2% fraction of all GFP+ cancer cells residing in the bone (Figure 14c). This result confirms previous observation from BLI signal quantification and suggest that the majority of cells during homing are actively proliferating, thus the population of arrested cells is small.

Next, we focused our attention on the differences in proliferation between cells in dormant and metastatic lesions. We could assess the cells in S-phase by injecting mice harboring dormant micrometastatic or full-blown metastatic lesions with thymidine analogue - BrdU. Note that BrdU pulse of 4 hours allows detection of fast proliferating cells. Injected cancer cells were co-stained with anti-BrdU and anti-GFP antibodies (Figure 15a). For immunochistochemical analysis we used bone marrow as a positive control since this tissue contains fast proliferating hematopoietic stem cells. BrdU staining results show 2-fold increase of proliferating cells in meta-

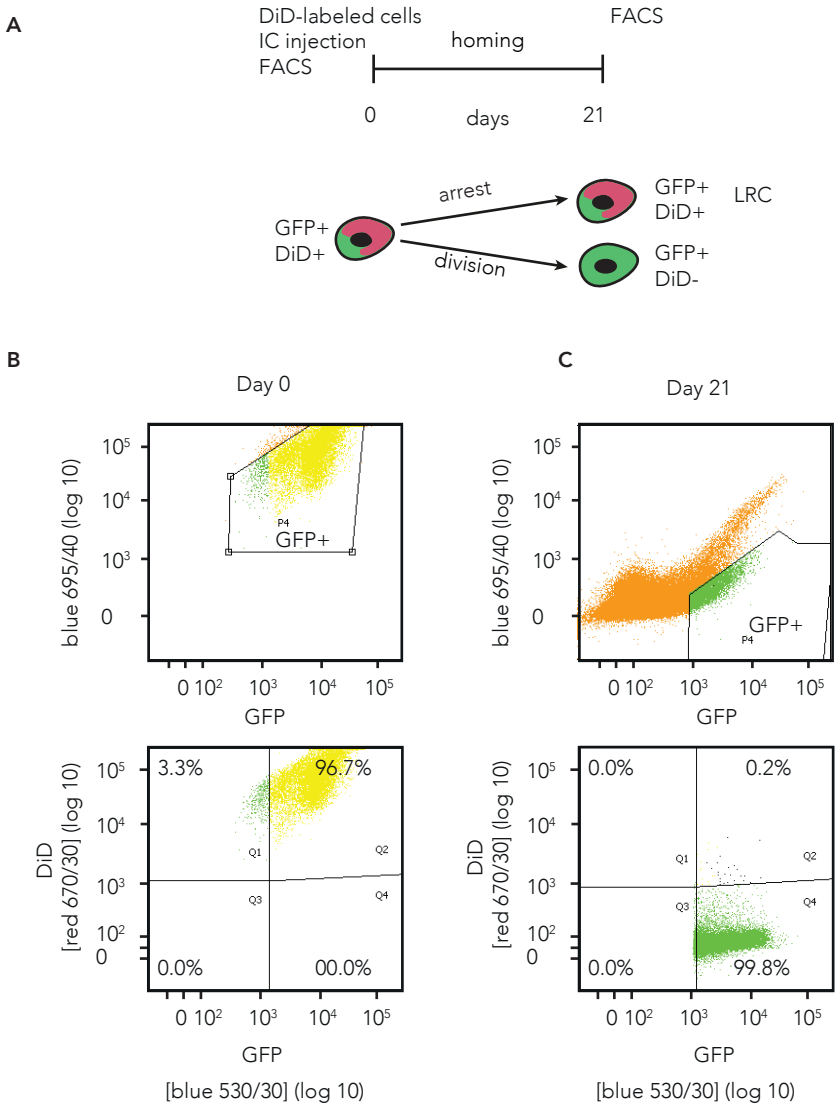

Figure 14. Cell dormancy in bone homing. (**a**) Schematic of label retaining cell (LRC) detection in bone homing. DBM cells were labeled with DiD for 30 minutes before injection and BLI detected bone resident cells were isolated 21 day post inoculation and analyzed on flow cytometer. (**b**) (top) Flow cytometric analysis of GFP+ cells after labeling and (bottom) DiD+ cells at day 0 (**c**) Flow cytometric analysis of cells extracted from micrometastatic lesion (day 21). (top panel) Metastatic cells were gated based on GFP expression and (bottom) analyzed in Q2 gate for DiD label retention. For all FACS analysis percentage of gated cells is indicated.

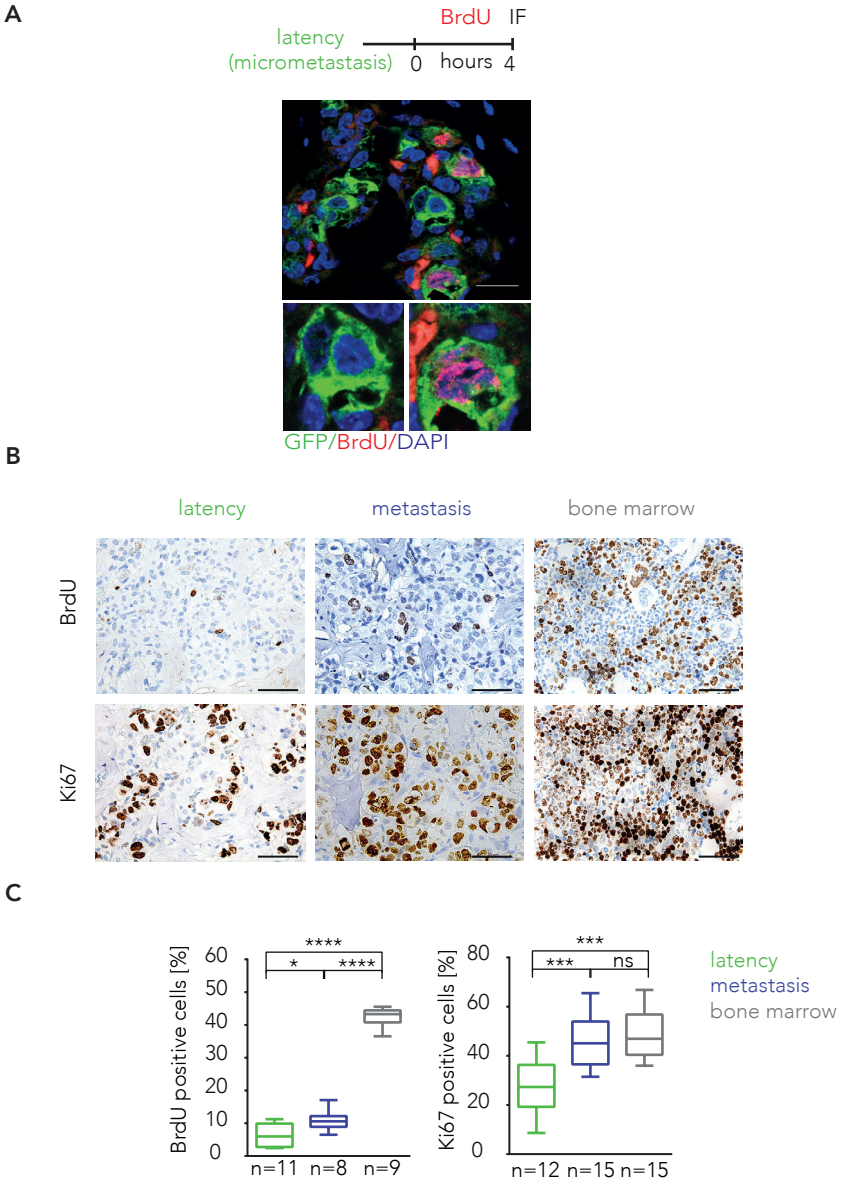

Figure 15. Impaired proliferation during latency. (**a**) (top) Schematic of BrdU incorporation experiment. Animals bearing latent micrometastatic or metastatic lesions were injected with BrdU 4 hours before sample collection. (bottom) Representative immunofluorescent image of bone lesion stained for GFP (green), BrdU (red) nuclei (blue). Scale bar: 20 μm (**b**) (top) Representative images of BrdU and (bottom) Ki67 immunohistochemical staining performed on latent micrometastatic (latency), metastatic (metastasis) lesion and bone marrow cells. Scale bar: 50 μm. (**c**) (left) Quantification of BrdU positive and (right) Ki67 positive cells. Panel c shows data as whisker plots: mid-line, median; box, 25 – 75 percentile; whisker, minimum to maximum. Statistical significance: ns, not significant $p>0.05$; *, $p\leq0.05$; **, $p\leq0.01$; ***, $p\leq0.001$; ****, $p\leq0.0001$.

static lesions in comparison to dormant micrometastasis (11% vs 6%) (Figure 15b and c). These values were much lower than in the bone marrow. Complementary results were obtained using Ki67 antibody that is a marker for all active phases of the cell cycle (G1, S, G2 and mitosis) (Figure 15b and c). Using this staining we confirmed an increase of cycling cells in metastatic lesions (from 27% to 45%). In addition, the difference in the number of Ki67 positive cells between metastatic lesions and bone marrow cells is not significant. It means that the majority of cells in the metastatic lesions are in active cell cycle, but proliferation turnover is slower than in the bone marrow. Moreover, as shown on Figure 15b and c, a high number of Ki67 negative cells indicates that they were arrested in G0 phase in the dormant lesions.

To further describe the cell dormancy in our model we performed pulse-chase experiment to detect LRCs in dormant and metastatic lesions. In this setting animals harboring either dormant micrometastases or metastatic lesions were pulsed for 10 days with fluorescent thymidine analogue – EdU. We assumed that during the pulse phase any proliferating cell that completed the S-phase would incorporate EdU resulting in slow and fast proliferating cells labeling. 12-day chase phase allows fast proliferating cells to dilute the dye below detection level. In result cells that retain the fluorescence are slow proliferating LRCs. Figure 16a and b shows that the percentage of LRCs is higher in the latent lesion in comparison to the metastatic.

Growth dynamics of the lesion does not rely only on proliferation rate but also is balanced by a cell death. Therefore, we tested the apoptosis in latent and metastatic setting. We used an *in vivo* approach based on caspase 3/7-substrate activity quantified by BLI. The results showed that latent micrometastatic lesions have

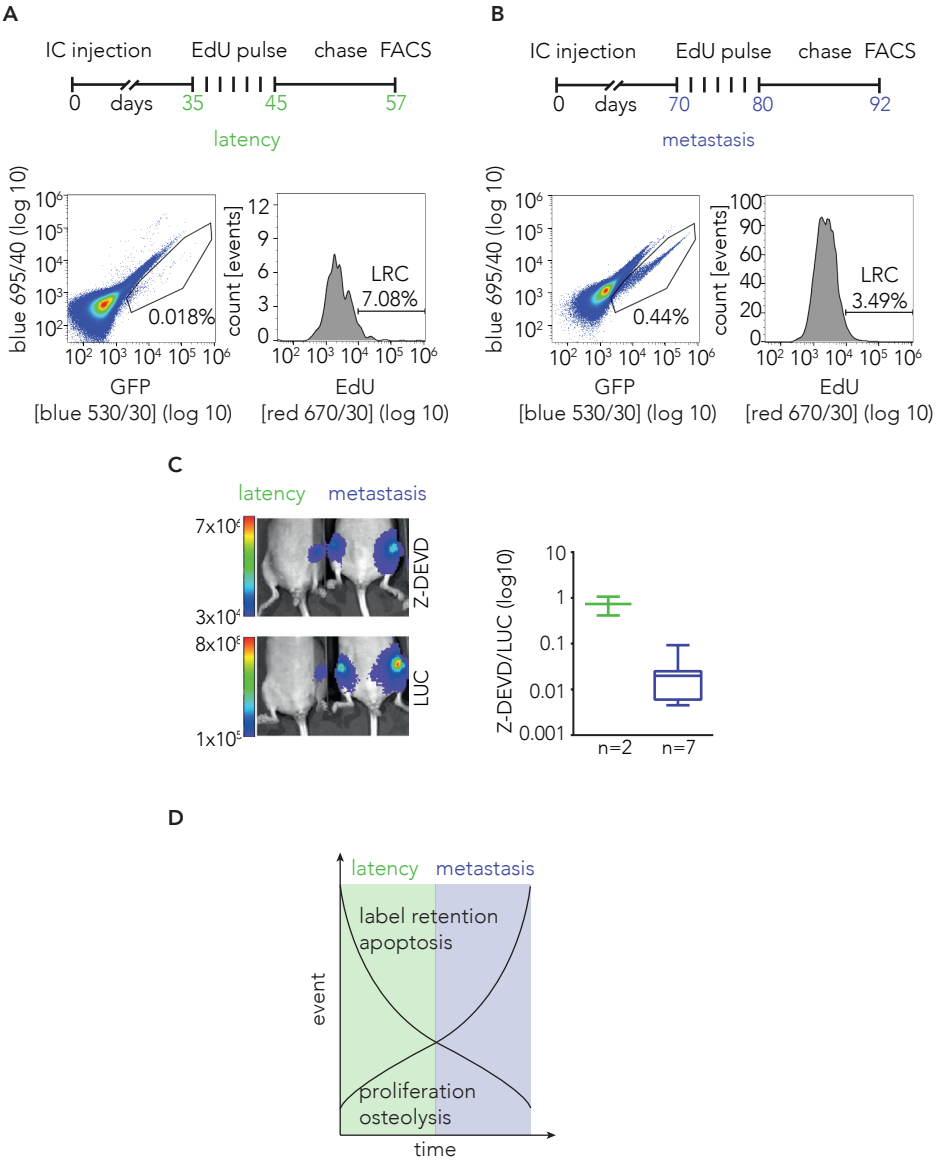

Figure 16. Tumor mass dormancy in the ER+ breast cancer to bone metastasis. (**a,b**) (top) Schematic of label retaining cell detection in latent and metastatic lesions. DBM cells were injected and animals presenting latent micrometastatic (latency) or metastatic (metastasis) lesions were pulsed with EdU for 10 days followed by 12 days of chase. Isolated cells were gated based on GFP expression (left) and EdU positive cell number was quantified (right). (**c**) (top) Animals were injected with caspase 3/7 substrate (Z-DEVD) and BLI signal was measured. (middle) Luciferin (LUC) BLI activity was assessed 4 hours later. (bottom) Quantification of apoptotic signal (Z-DEVD) normalized to lesion size (LUC). (**d**) Summary showing main alterations in proliferation, osteolysis, apoptosis and LRC number between latent micrometastatic and metastatic lesions. Panel b shows data as whisker plots: mid-line, median; box, 25 – 75 percentile; whisker, minimum to maximum.

higher apoptotic/total cell number ratio in comparison to overt metastasis (Figure 16c).

Collectively, our data indicates that, in our model, the latency state is controlled by both mechanisms of dormancy: cellular and tumor mass. The micorometastatic lesions in the latent phase are composed of balanced number of proliferating and apoptotic cells. Moreover, a fraction of arrested cells and LRCs is significantly higher in the latent lesions. In metastatic lesions, the number of cells in G0, LRCs and apoptotic cells significantly decrease and the proliferation to cell death balance is disrupted enforcing a rapid growth of metastatic lesion (Figure 16d).

3. Genome-wide loss-of-function shRNA screen identifies MSK1 as a latency regulator.

3.1 Optimization of screen conditions

We used our novel model of latency to perform unbiased screen to identify new dormancy regulators in ER+ breast cancer. We decided to perform the genome-wide screen that prevents focus only on well-known metastasis signaling pathways. To this end, by using RNAi technology we screened for genes which, upon silencing, enhance metastatic capacities of dormant cells. Such *in vivo* screens were used by others to identify novel targets in cancer, metastasis and tumor suppressor genes (Zender, Xue et al. 2008; Murugaesu, Iravani et al. 2014). Nevertheless, this technology was never used to reveal dormancy-controllers in ER+ breast cancer and this screen challenges the library to identify shRNAs that become enriched during metastasis to the bone.

For genome-wide loss-of-function screen we used the TRC library of pooled shRNA plasmids packed in the lentiviral vectors (Figure 17a) that was created by The RNAi Consortium (Moffat, Grueneberg et al. 2006; Root, Hacohen et al. 2006). This library consists of 80717 shRNA constructs (5 different per gene) down-regulating 16019 human genes. Main steps of the screening were: (i) DBM cell transduction with the library, (ii) intracardiac injections, (iii) metastatic lesion formation, (iv) genomic DNA isolation, (v) bioinformatics analysis of the data, and (vi) candidate gene selection (Figure 17b).

Each step of the screening was optimized and preformed independently to improve the process. First, we verified cell transduction conditions to ensure the single insertion of each shRNA. To this point, GFP+ cells were transduced with an increasing number

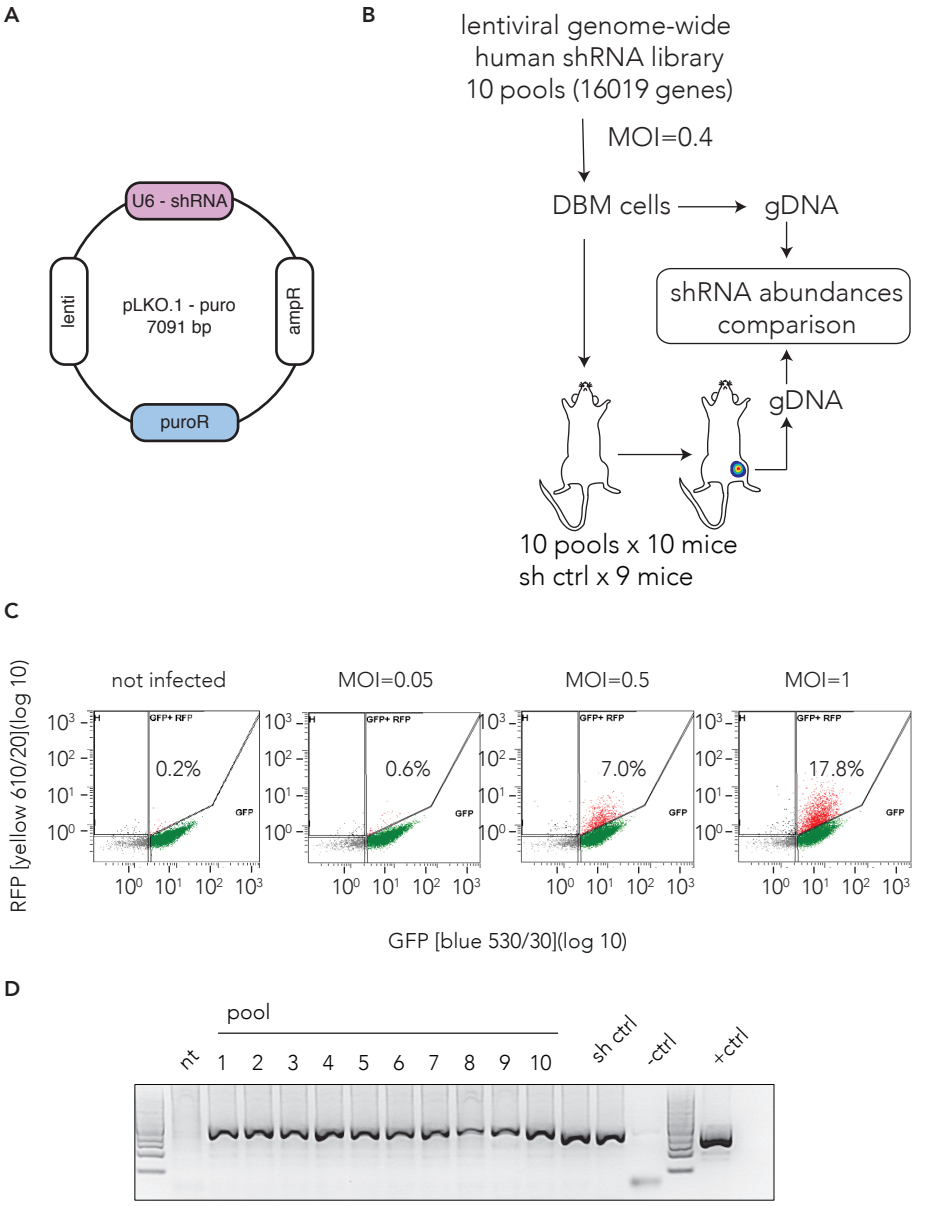

Figure 17. Genome-wide loss-of-function screening for dormancy regulators in ER+ breast cancer. (**a**) Schematic of pLKO vector. Main features: shRNA sequence, puromycin resistance, ampicillin resistance and lentiviral packaging sequences. (**b**) Schematic of an *in vivo* shRNA screening strategy. DBM cells were infected with a human shRNA library and injected intracardiaclly into mice. Genomic DNA was isolated from pre-inoculated DBM cells and cells isolated from metastatic lesions. shRNA inserts were sequenced using massive parallel methods to compare insert abundances between samples. (**c**) Determination of the best MOI. DBM cells were infected with increasing number of viral particles. Transduced RFP positive cells were assessed by flow cytometer. (**d**) Agarose gel electrophoresis of gDNA isolated from DBM cells. The band indicates integration of shRNA construct.

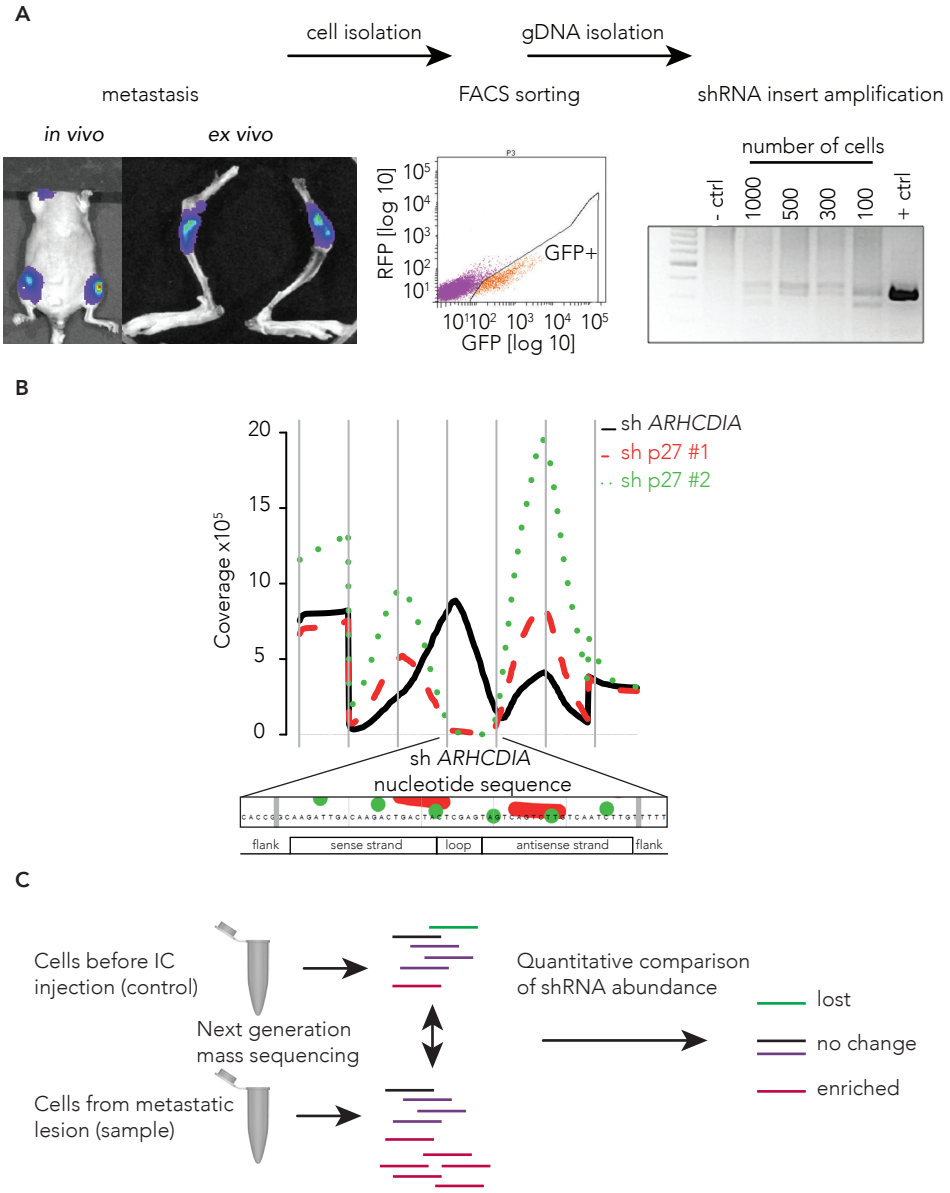

Figure 18. Genomic DNA isolation and shRNA sequence deconvolution. (**a**) Schematic of sample processing and isolation of gDNA. Cells were isolated from metastatic lesions, sorted based on GFP and submitted to gDNA isolation followed by shRNA insert amplification. (**b**) Alignment to sh control sequence of NGS reads of gDNA for sh control (black) and sh p27 infected cells (green and red). (**c**) Schematic of deconvolution strategy. gDNA from pre-inoculation (control) samples and from metastatic lesions was sequenced and quantitative comparison of shRNA sequence number revealed sequences enriched after metastatic process.

of RFP viral particles corresponding to multiplicity of infection (MOI) range from 0.05 to 1. As shown on Figure 17c infection efficiency increases proportionally to the MOI reaching 17.8% for 1 viral particle per cell (MOI=1). Next, based on this data and the literature, we chose MOI<0.5 as optimal to ensure single shRNA insertion (Chen, Heller et al. 2012). In the preliminary setting MOI=0.5 corresponds to 7% of infected cells and is lower than recommended limit of 10% that favors single shRNA delivery. Following the infection, PCR analysis confirms the integration of shRNA of library pools and control into the genome of transduced cells (Figure 17d).

In addition, using small-scale trials we optimized procedures of metastatic cell isolation, FACS sorting, genomic DNA extraction and shRNA insert amplification (Figure 18a). The first two procedures were potential limitations for the screening. Since we could isolate only up to 2000 cells from a bone lesion, we were unsure if such small amount of cells is sufficient for DNA isolation, amplification and next generation sequencing (NGS). Therefore, to check the feasibility of our approach we isolated DNA from different number of cells (1000-100) with integrated sh ARHCDIA and two sh p27#1 and sh p27#2. PCR amplification of 100 cells showed unspecific band indicating that the quality or the quantity of DNA was suboptimal (Figure 18a). Next, the PCR products amplified from 500 cells were submitted to sequencing and the reads were aligned to the known nucleotide templates. Figure 18b shows that only shRNA ARHCDIA sample reads align to the template with shRNA ARHCDIA sequence. On the contrary, for shRNA p27 reads coverage to the shRNA ARHCDIA region dramatically decreases. Importantly, results of NGS serve as a proof-of-concept for the specificity of shRNA construct quantification. Based on this results we confirmed the feasibility of the screen and we selected the optimal conditions.

3.2 Analysis of screen results and candidate gene selection

To perforem the screen, we intracardiaclly injected control and library infected DBM cells into 109 animals. Library was divided into 10 pools and each pool was injected into 10 mice while 9 animals were injected with sh control vector. Animals were monitored for over 4 months and during this time mice that developed overt metastatic lesions were sacrificed to subsequently collect genomic DNA (gDNA) samples from the metastatic lesions. Next, the gDNA samples were analyzed and subjected to the NGS. We compared the abundance of each shRNA in metastatic lesions with shRNA represented in the initially injected cells to identify the shRNAs that were enriched in the metastatic process (Figure 18c). In this setting, high shRNA sequence abundance fold change corresponds to better downregulation of gene transcription.

The results showed that DBM cells infected with the pooled shRNA library enriched metastatic properties *in vivo*. While bone homing incidence was similar between control and shRNA library group (Figure 19a), metastasis incidence increases in mice injected with library by 4 fold (Figure 19b).

Based on gDNA quality and quantity we submitted to the NGS analysis a total of 30 samples (7 pre-inoculation controls and 23 from metastatic lesions) therefore, we covered 70% of genes (Table 3). Next, we applied 5 criteria in order to select dormancy-enforcing candidate genes. First, we selected from control samples (initially injected cells) shRNAs which representation was between 10 and 300 in a pool (Figure 19c and Table 3). This step was necessary to omit genes whose silencing enhances cell death or proliferation *in vitro*, before the injection. Next, we applied the abundance threshold to select for genes enriched in the metastatic samples. Figure 19d shows 322 genes chosen for further analysis

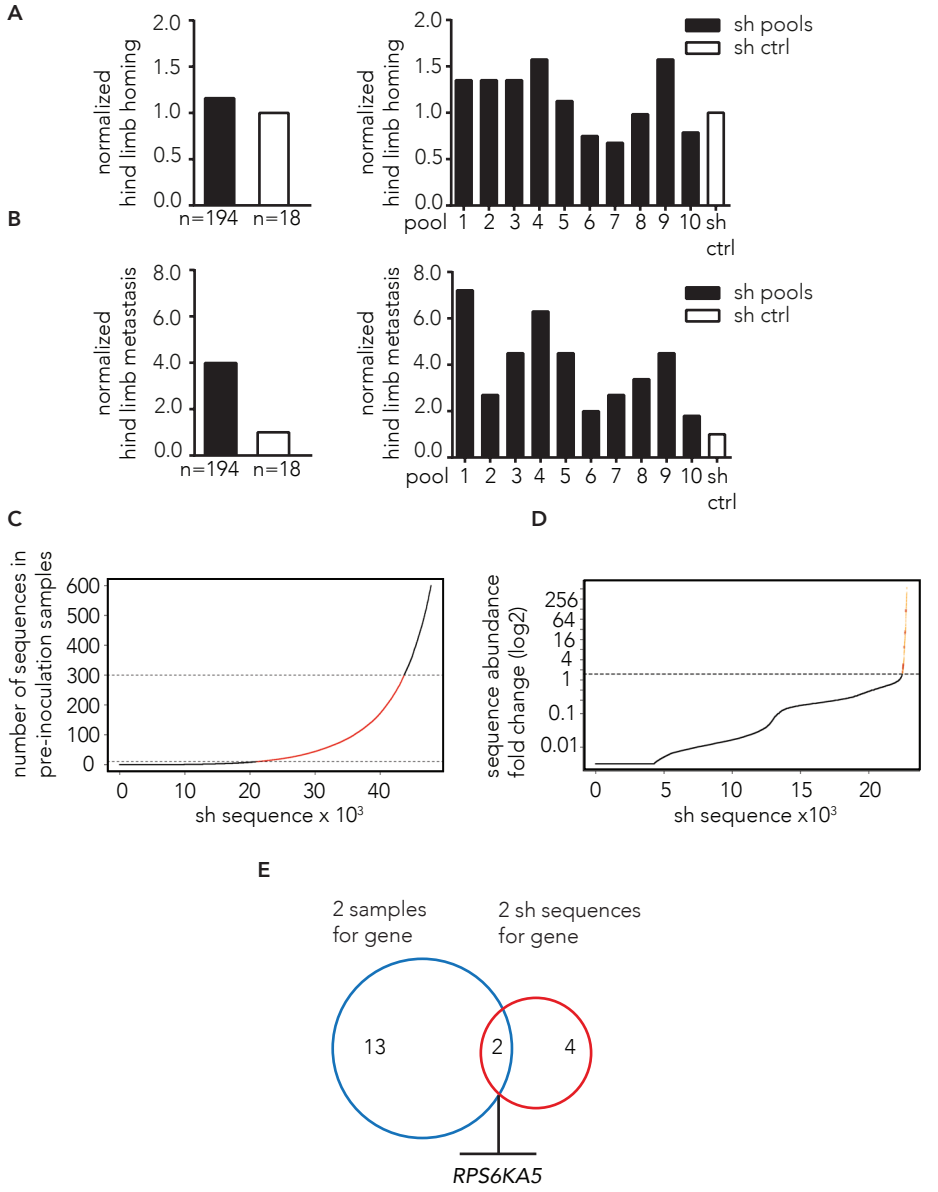

Figure 19. The *in vivo* shRNA screen and candidate gene selection. (**a**) (left) Hind limb homing of library (black) and sh control (white) infected cells. (**b**) (left) Hind limb metastasis of library (black) and sh control (white) infected cells. (**a** and **b**) (right) Bone homing and metastasis results are represented for each pool separately. (**c**) Distribution of shRNA sequence number in pre-inoculation samples. shRNAs between 10 and 300 repeats are labeled in red. (**d**) Distribution of shRNA sequence abundance fold change between pre-inoculated and bone metastasis derived samples. Highly abundant shRNA sequences are marked in yellow and red, the dash line marks 1.5 hit selection threshold. (**e**) Venn diagram showing the overlap between shRNA sequences highly abundant in 2 or more independent samples (blue) and targeted by 2 different shRNA sequences (red).

based on 1.5 sequence enrichment cut-off. From these, we selected genes downregulated in 2 or more metastatic lesions (15 genes) or targeted by 2 or more independent shRNAs (6 genes). Among those, 2 genes, *RPS6KA5* and *CELA3A,* fulfilled both criteria meaning that they were found in 2 independent samples and silenced by 2 shRNAs (Figure 19 and Table 3). *RPS6KA5* gene encodes a mitogen and stress-activated kinase 1 (MSK1), a downstream effector of MAPK signaling pathway. Whereas, *CELA3A* gene product codes for a secreted protein. Throughout the thesis the protein names instead of genes will be used for the simplification reason.

Table 3. Screening top hit selection

Analysis step	Description	Deconvoluted shRNAs		Deconvoluted genes	
		Number	Percent. [%]	Number	Percent. [%]
All library	Pools 1-10	80,717	100.000	16,019	100.000
Samples analyzed	Pools 1-5, 9, 10	55,754	69.073	11,297	70.523
1st criterion	Control sample 10-300	22,808	28.257	8,880	55.434
2nd criterion	FC>1.5	328	0.406	322	2.010
3rd criterion	2 samples for gene	17	0.021	15	0.094
4th criterion	2 shRNA for gene	n/a	n/a	6	0.037
5th criterion	criteria 3 and 4	n/a	n/a	2	0.012

Abbreviations: percent., percentage; n/a, not applicable

In the final step of dormancy-controller selection, we used clinical samples to validate a candidate gene by bioinformatics analysis on patient cohort with annotation for time and site of metastasis. To this end, we used publically avaliable datasets composed of transcriptomic profiles of 560 primary breast tumors with annotated clinical follow-up. In this collection of tumors 349 are ER+ and 211 are ER-, based on the mRNA expression of *ESR1* gene

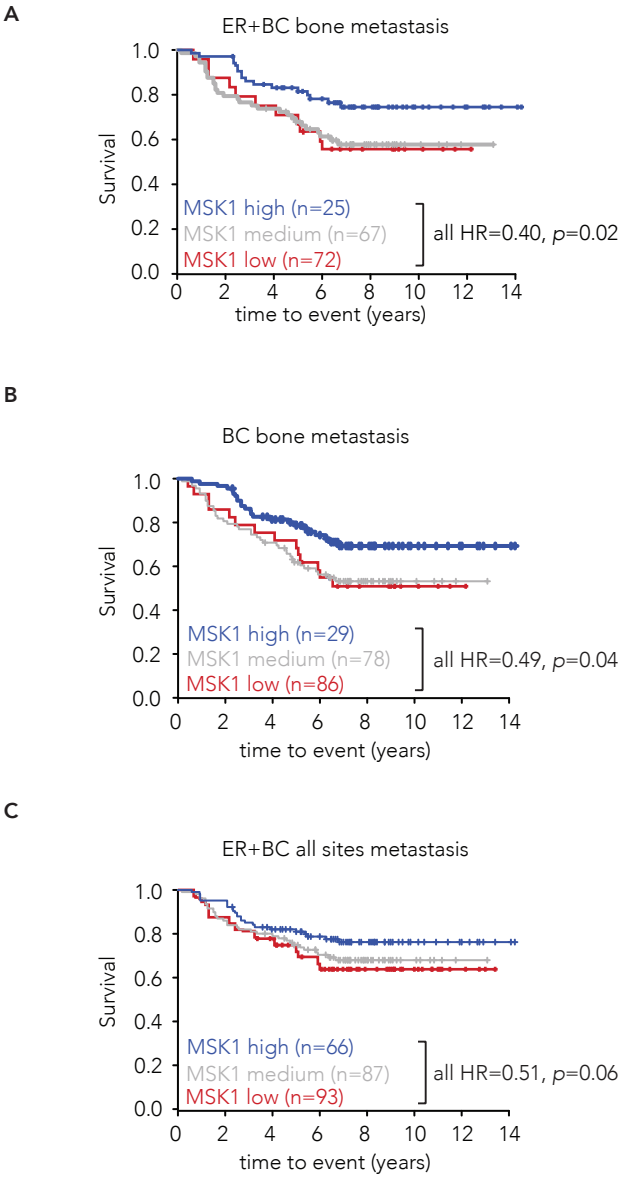

Figure 20. MSK1 low expression in primary tumors correlates with high risk of bone metastasis
(**a**) Survival analysis representing the proportion of bone metastasis recurrence-free patients stratified according to MSK1 mRNA levels in ER+ breast cancer patient samples. (**b**) Survival analysis representing the proportion of bone metastasis recurrence-free patients stratified according to MSK1 mRNA levels in breast cancer patient samples. (**c**) Survival analysis representing the proportion of all sites of metastasis recurrence-free patients stratified according to MSK1 mRNA levels in ER+ breast cancer patient samples. HR=$H^{MSK\ high}/H^{MSK\ low}$.

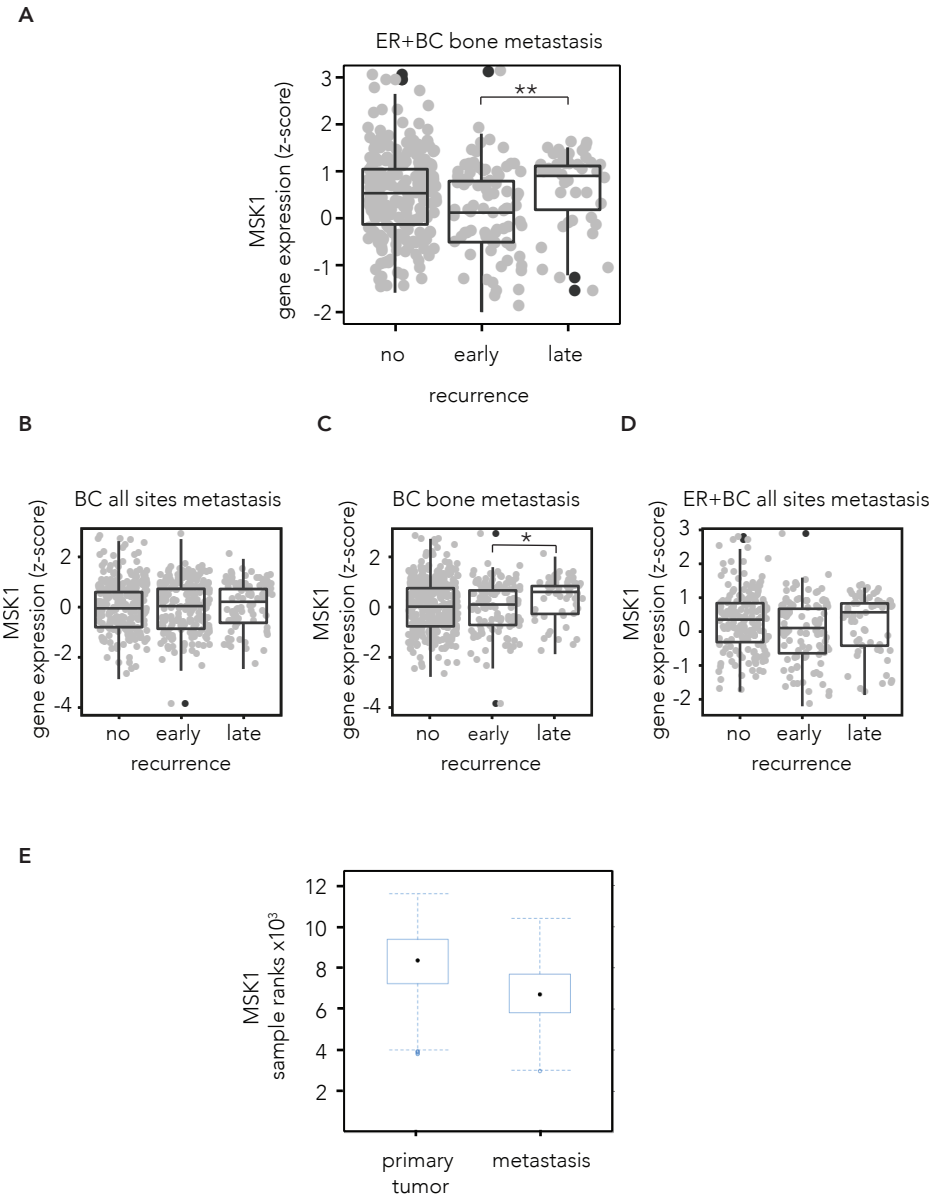

Figure 21. MSK1 expression in primary tumors inversely correlates with early bone metastasis. (**a**) MSK1 gene expression of ER+ breast cancer patients stratified according to bone metastatic relapse: none, early (before 3 years after primary tumor diagnosis) or late (more than 3 years after diagnosis). (**b**)MSK1 gene expression of breast cancer patients stratified according to metastatic relapse in all sites. (**c**) MSK1 gene expression of breast cancer patients stratified according to bone metastatic relapse. (**d**) MSK1 gene expression of ER+ breast cancer patients stratified according to metastatic relapse in all sites. (**e**) MSK1 gene expression in primary tumors and metastatic samples from breast cancer patients. Panels show data as whisker plots: mid-line, median; box, 25 – 75 percentile; whisker, 10-90 percentile. Statistical significance: *, $p \leq 0.05$; **, $p \leq 0.01$.

expression. Notably, Kalan-Meier analysis showed that in ER+ primary tumors, MSK1 mRNA expression levels are inversely associated with a high probability of bone relapse in the patients (Figure 20a-c).In line with this, the MSK1 gene expression is significantly downregulated in primary tumors of patients that suffer early bone relapse defined as recurrence before 3 years from primary tumor diagnosis (Figure 20a). Strikingly, associations of both expression analyses are highly significant for bone metastasis in ER+ breast cancer subtype. Moreover, Figure 20e demonstrates that MSK1 is downregulated in metastatic samples in comparison to primary tumors, suggesting its suppressive role in metastasis (Figure 21e). Thus, we selected MSK1 for further studies as a metastatic latency promoter.

4. MSK1 controls metastatic latency in ER+ breast cancer bone metastasis

4.1 Phenotypic validation of MSK1 as a mediator of metastatic latency

The candidate gene must be validated in the conditions analogical to those used in the screen. To this end, we silenced MSK1 protein levels by two individual shRNA constructs that belong to the TRC library (Figure 22a). Figure 22b shows that MSK1 downregulation represses the expression of its known target genes such as *CFOS*, *EGR1*, *COX2*, *JUNB* (Figure 22c and d). Nevertheless, it has no major impact on cell growth *in vitro* (Figure 22c and d).

Intracardiac injections of DBM cells infected with sh control and two different sh MSK1 sequences phenocopied the observations previously described in the screening and validated MSK1 as dormancy-controller in ER+ breast cancer bone (Figure 23). In detail, MSK1 downregulation increases metastatic lesions number (33% and 52% vs 13%) in intracardiac xenograft assay (Figure 23a). Strikingly, BLI quantification shows also differences in growth kinetics dependent on MSK1 levels. Control cells grew rapidly and formed micrometastatic lesions, that stayed stable in size (latency) and progressed to macrometastasis (lesion size > day 0 threshold). MSK1 downregulated cells proliferate intensively after injection and progress to metastatic lesions without latent phase (Figure 23b and c). The immunohistochemical analysis confirmed MSK1 silencing in the lesions derived from MSK1 downregulated cells (Figure 24).

These results suggest that MSK1 downregulation promotes metastatic incidence and prevents the micrometastatic dormancy in the ER+ breast cancer model of latent bone metastasis.

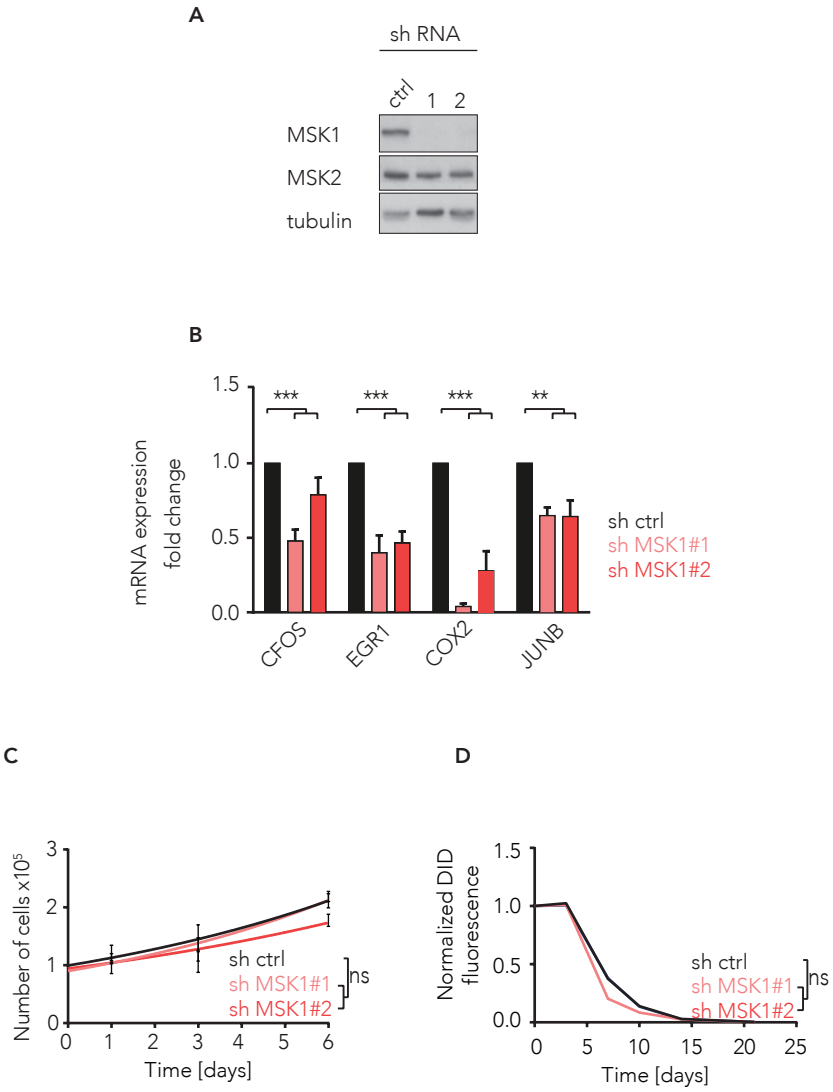

Figure 22. MSK1 downregulation by shRNA. (**a**) Western blot analysis of MSK1, MSK2 and tubulin in MSK1 downregulated cell lines. (**b**) Gene expression levels in sh control and sh MSK1 infected cells. (**c**) *In vitro* cell growth curve of sh control and sh MSK1 infected DBM cells. (**d**) Label retention assay of sh control and sh MSK1 infected DBM cells. Panels b and c show data as mean ± SD from 3 biological replicates. Statistical significance: ns, not significant $p>0.05$; *, $p\leq0.05$; **, $p\leq0.01$; ***, $p\leq0.001$.

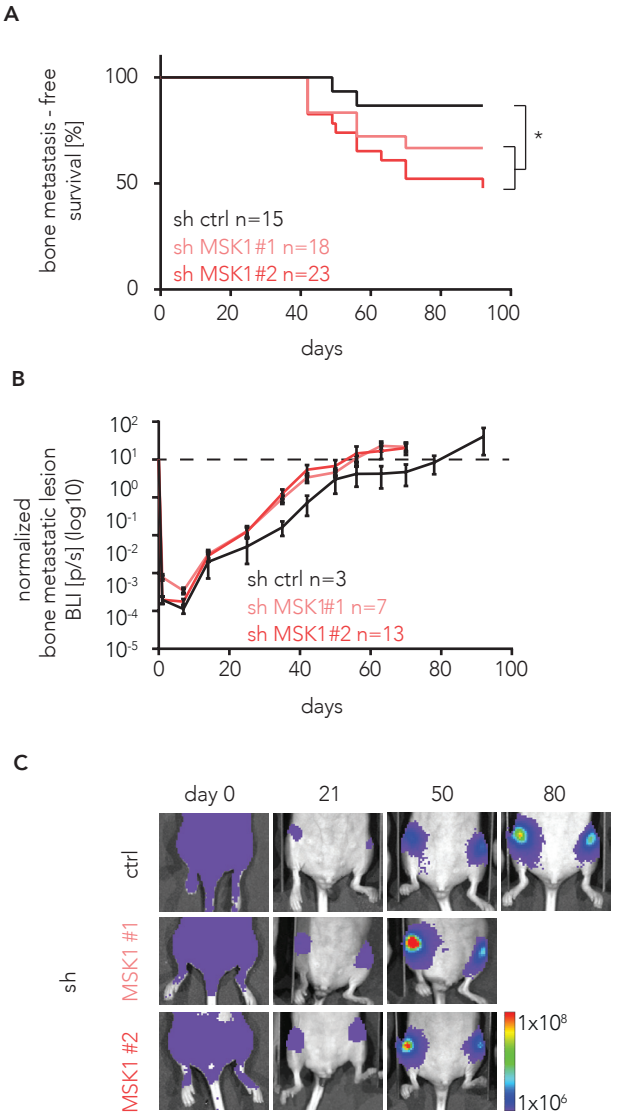

Figure 23. MSK1 downregulation increases bone metastasis. (**a**) Kaplan-Meier analysis of bone metastasis-free survival between sh control and sh MSK1 infected cells. (**b**) Quantification of bone metastatic lesion BLI signal. Dashed line represents day 0 threshold. (**c**) Representative bioluminescence images showing bone metastasis progression of sh control and sh MSK1 infected DBM cells. Panel b shows data as mean ± SEM. Statistical significance: *, $p \leq 0.05$.

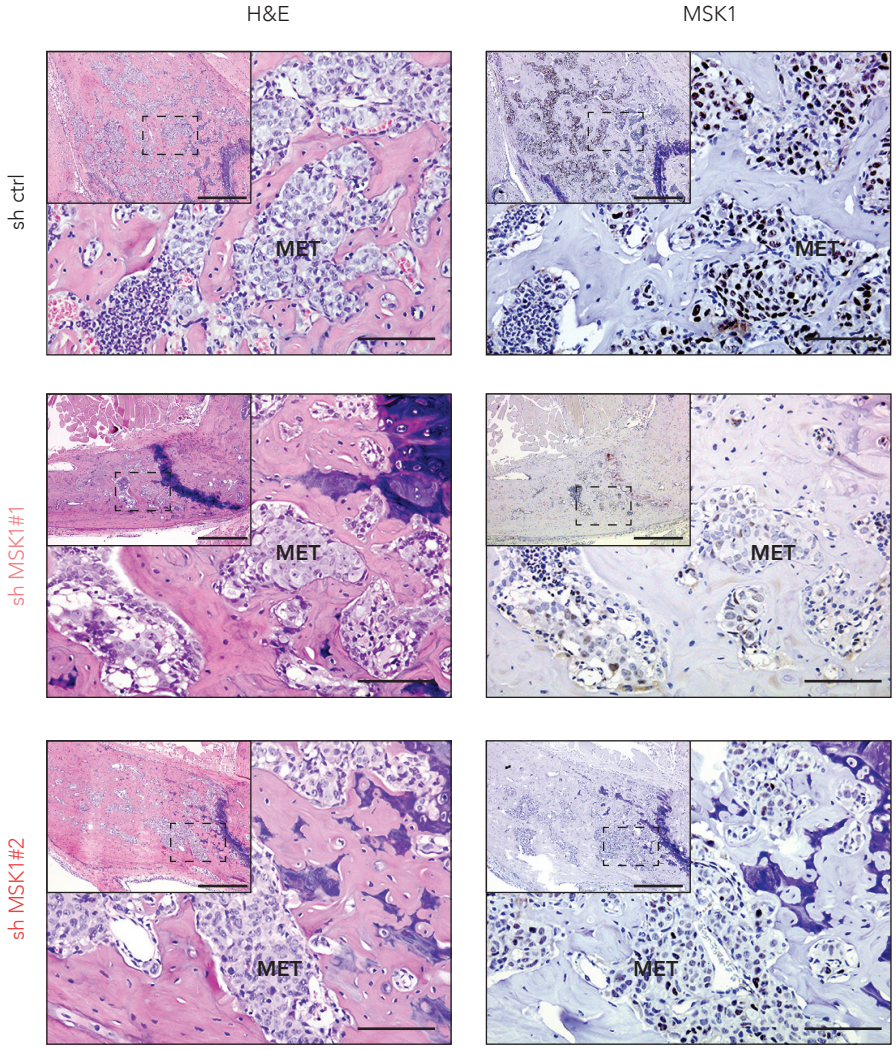

Figure 24. MSK1 levels validation in bone metastatic lesions. (**left**) H&E staining of hind limbs bones from mice injected with DBM sh control (top), sh MSK1#1 (middle) and sh MSK1#2 (bottom). (**right**) MSK1 staining of hind limbs bones. MET- metastatic lesion; dashed square – magnified region. Scale bars: 500 μm (small picture), 100 μm (large picture).

4.2 Generation of a MSK1 knockout DBM cells by genome editing

To further validate, in shRNA independent manner, the role of MSK1 loss-of-function and its contribution to bone metastasis we used recently developed technique of genome editing (Cong, Ran et al. 2013; Hsu, Scott et al. 2013). The CRISPR/Cas9 genome editing is a cleavage of DNA by Cas9 nuclease that is guided to the genomic site by 20-nt guide RNA (gRNA). Created double strand break is repaired by non-homologous end-joining (NHEJ) that often results in indel mutations and gene disruption. Gene targeting serves as an additional method to validate the candidate gene because it eliminates viral infections, possible shRNA off-target effects and generates permanent ablation of the protein.

To this end, we designed 5 different targeting pX330 vectors for MSK1 knockout (KO) generation and one control vector without gRNA sequence (Figure 25a and b). MSK1 KO vectors contain gRNA sequences complementary to DNA target sequences located in exons 5, 6, 8, 9, and 13, respectively. These target loci overlap with the location of residues critical for catalytic and kinase activity of the protein (Ser 212, Ser 376 and Thr 581) (Figure 25a). We used for targeting only gRNAs that were predicted by the sequence designing software (Montague, Cruz et al. 2014) to be free of potential off-targets (see Table 6 in Material and Methods section). This complex and alternative strategy was used to increased the success rate of MSK1 KO cells generation. Figure 25a shows an example of targeting strategy for MSK1 exon 5 with the highlighted 20-nt target sequence and 3-nt PAM sequence located on the minus DNA strand. Guide RNA composed of complementary target sequence and RNA scaffold anneals to DNA and facilitates Cas9 mediated DNA cleavage between 3rd and 4th nucleotide up-

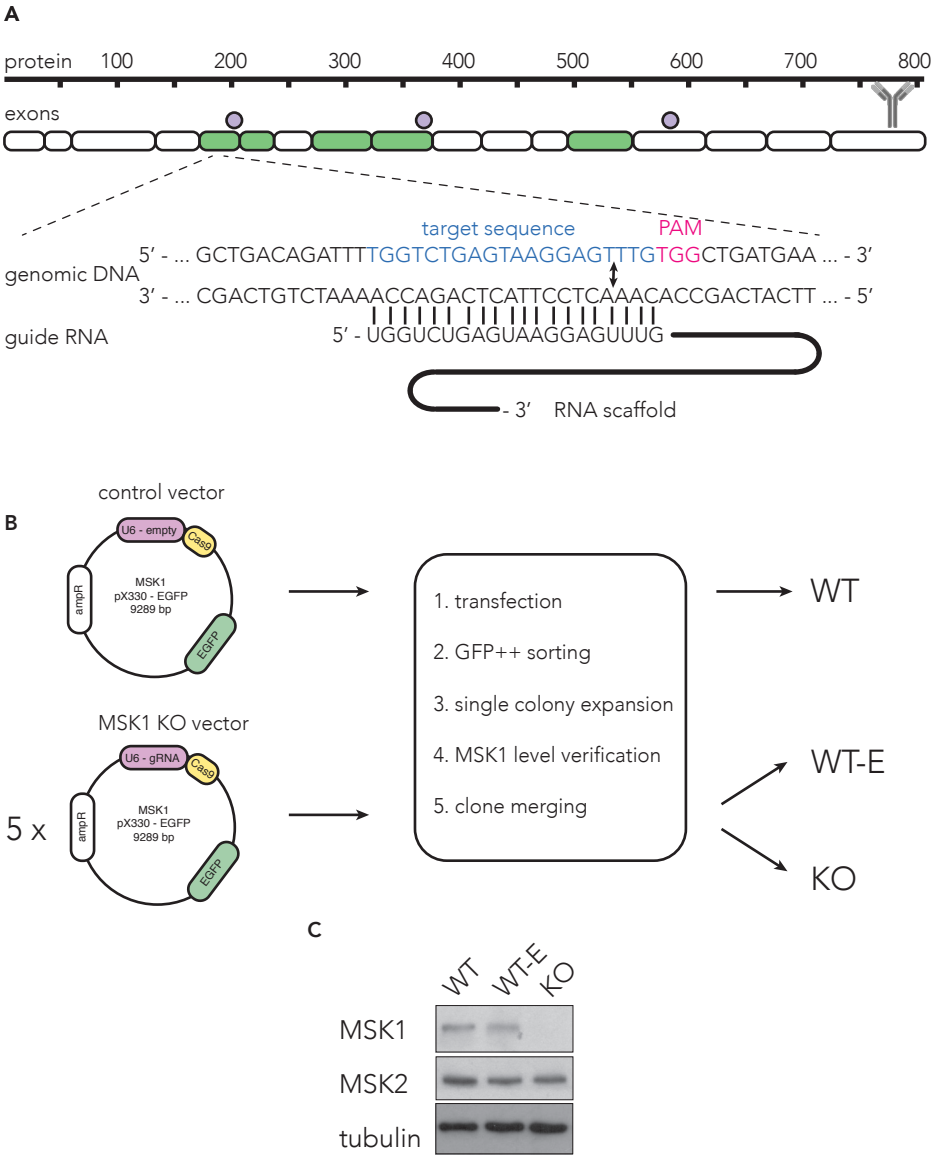

Figure 25. MSK1 knockout generation by CRISPR/Cas9 gene editing. (**a**) Targeting strategy. Simplified amino acid sequence of MSK1 protein (top). Five target sequences located in exons 5, 6, 8, 9, and 13, respectively, (green) that overlap with residues critical for MSK1 catalytic and kinase activity (violet circles). 20-nt target sequence in exon 5 (blue), 3-nt PAM sequence (red) and guide RNA composed of complementary target sequence and RNA scaffold. Cleavage site locates between 3rd and 4th nucleotide upstream from PAM (arrows). Anti-MSK1 antibody binds to the C-terminal end. (**b**) Schematic of MSK1 WT, WT-E and KO cell line generation. Main features of pX330 include: gRNA (or empty) sequence under U6 promoter, Cas9 endonuclease, ampicillin resistance and EGFP. (**c**) Western blot analysis of MSK1, MSK2 and tubulin in pooled cell lines.

stream from PAM. Analogical strategy was applied to target other loci in MSK1 gene.

DBM cells were transfected with control or MSK1 KO vectors containing EGFP selection marker (Figure 25b). This transfection resulted in low efficiency of GFP++ cells (0.2%-0.6%), but it was sufficient to isolate 38 cell lines derived from single-cell colonies (Table 4).

Table 4. Genome targeting efficiency

gRNA	Transfection [%]	Clonal expansion	MSK1 status			KO efficiency [%]
			Absent	Present	Inconclusive	
1	0.3	7/96	2	4	1	28.6
2	0.6	10/96	6	2	2	60.0
3	0.4	7/96	2	3	2	28.6
4	0.4	3/96	0	1	2	0.0
5	0.4	9/96	4	1	4	44.4
1-5	n/a	36/480	14 ^{KO}	11 $^{WT-E}$	11	38.9
empty	0.2	2/96	0	2 ^{WT}	0	n/a
total	n/a	38/576	14	13	11	36.8

Abbreviations: n/a, not applicable; KO, knockout pool; WT-E, wild type edited pool; WT, wild type pool

These include 14 MSK1 KO cells lines, 11 wild type-edited (WT-E), derived from targeting vector transfected cells but expressing MSK1 protein, and 2 of control wilde-type (WT) cells (Table 4). In order to restore the heterogeneity of cell populations we pooled separately MSK1 KO derived from different targeting vectors, WT-E and WT clones and created 3 new cell lines (Table 4, Figure 25b). Finally, we confirmed that genome editing was specific only for MSK1, whereas MSK2 level was intact (Table 4, Figure 25c).

4.3 MSK1 genetic and translational depletion promotes bone homing

We tested DBM WT, WT-E and KO cells in intracardiac injection assay of bone metastasis. First, we focused on the bone homing incidence since it is an crucial factor determining metastatic efficiency of the intracardiac assay (Figure 26a). We saw that bone homing incidence was significantly altered upon MSK1 depletion by two independent techniques. We also noticed that genome-editing increased homing capacities of DBM cells by 2 folds (Figure 26b and c).

Bone homing can be driven by multiple cell functions including survival in hypoxia and anoikis, alterations in adhesion, migration and invasion as well as in the ability to initiate new lesion in the secondary site (Figure 27a). In order to identify which cell functions are altered by MSK1 silencing *in vitro*, we used DBM shMSK1 cells. We focused our attention on cell survival, adhesion, migration and invasion assays. Data shows that survival in the hypoxic condition was intact upon MSK1 downregulation (Figure 27b). MSK1 silencing was insignificant for cell adhesion to different matrixes including collagen, fibronectin and matrigel (Figure 27c) as well for cell migration (Figure 27d). Interestingly, cell invasion was significantly promoted by MSK1 silencing (Figure 27e).

Next, we interrogated the effect of MSK1 depletion on tumor initiation properties using an oncosphere formation assay. Strikingly, MSK1 depletion increased oncosphere formation by 50% (Figure 28a and b). To validate the suppressive effect of high MSK1 levels on the formation of oncospheres we used a gain-of- function approach. To this end, ectopic expression of MSK1 moderately but significantly attenuates tumor initiating potential of DBM cells reducing the number of tumor initiating cells (Figure 28c).

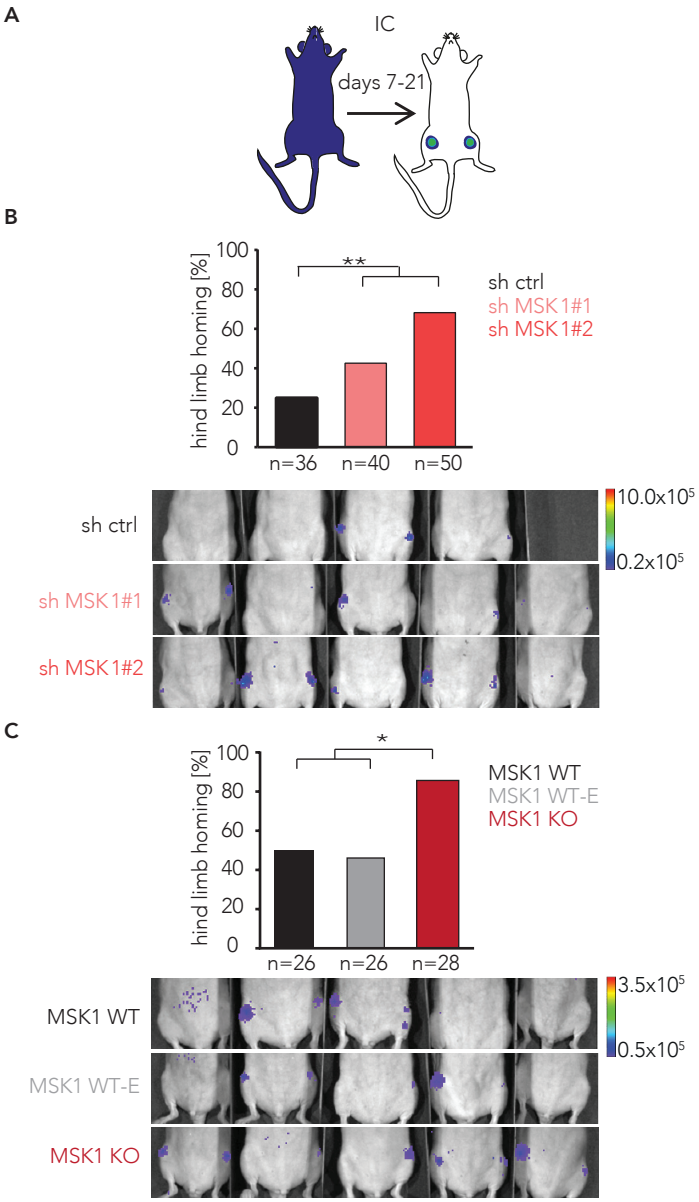

Figure 26. MSK1 depletion supports bone homing. (**a**) Schematic of intra-cardiac (IC) injection experiment. (**b**) Hind limb bone homing of sh control and sh MSK1 infected cells. (**c**) Hind limb bone homing of MSK1 WT, WT-E and KO cells.

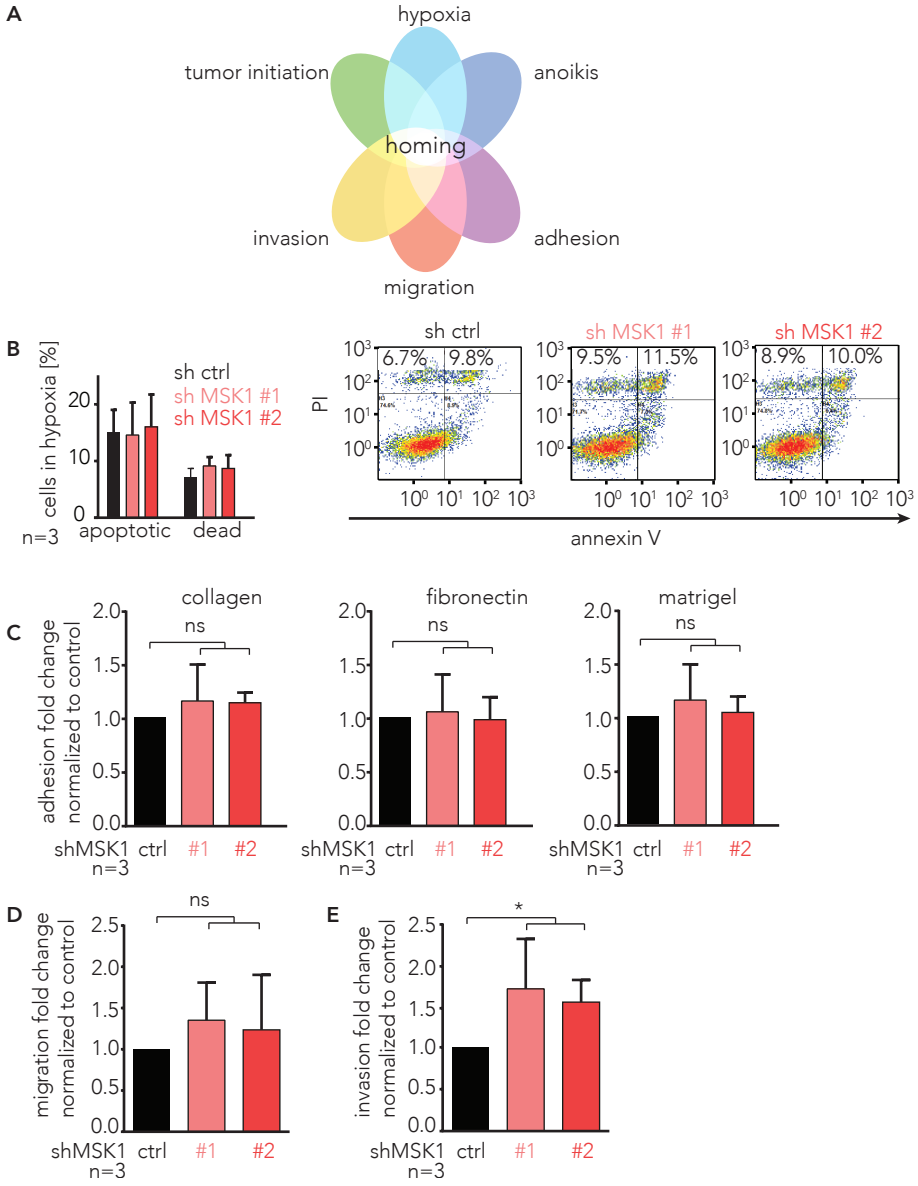

Figure 27. MSK1 suppresses cell invasion. (**a**) Schematic of cell functions implicated into the bone homing. (**b**)Flow-cytometry analysis quantification of apoptosis and cell death upon 72 h of hypoxic conditions (left). Representative plots of flow cytometric analysis of annexin V and PI staining in sh control and sh MSK1 infected cells (right). Apoptotic cells are in the top-right quadrants and dead cells in the top-left quadrants. (**c**) Cell adhesion to collagen (left), fibronectin (middle) and matrigel (right). (**d**) Cell migration of sh control and sh MSK1 infected DBM cells. (**e**) Cell invasion of sh control and sh MSK1 infected DBM cells. Panels b-e show data as mean ± SD. Statistical significance: ns, not significant $p>0.05$; *, $p\leq0.05$.

In summary, MSK1 depletion increased the number of invasive and tumor-initiating cells which may promote bone homing *in vivo*. In contrast, high levels of MSK1 attenuated tumor-initiating capacities validating MSK1 as a potent bone homing suppressor.

4.4 MSK1regulates bone homing by increasing the number of tumor-initiating cells and modulating cancer cell differentiation

Tumor-initiating properties of a cancer cell population can be enhanced by the increase of breast cancer stem cell number (Al-Hajj, Wicha et al. 2003). This cell population is characterized by $CD44^{+}$ $CD24^{-/low}$ cell surface antigens profile. Another widely used marker of breast cancer stem cells is CD49f. FACS analysis of breast cancer stem cell markers shows the increase of stem cell population upon MSK1 depletion (Figure 28d and e), whereas the over-expression of MSK1 reduces the number of $CD44^{+}$ $CD24^{-/low}$ and $CD49f^{high}$ cells (Figure 28f).

MSK1-dependent expansion of breast cancer stem cell population suggests a role of MSK1 in the control of cancer cell plasticity and differentiation. To test this hypothesis we decided to investigate the proportions between differentiated (luminal) and poorly differentiated (basal) cells upon MSK1 depletion in the ER+ breast cancer cell population. To do so, we monitored gene expression of keratins. These proteins are fundamental markers of epithelial differentiation and can serve as an independent and complimentary assay to functionally explain tumor initiation capacities (Abd El-Rehim, Pinder et al. 2004). Keratins 7 (*KRT7*) and 18 (*KRT18*) high expression in primary tumors is associated with luminal breast cancer subtype and good prognosis (Abd El-Rehim, Pinder et al. 2004; Ha, Lee et al. 2011). While high expression of basal-type keratin 14 (*KRT14*) has been seen in poorly differentiated aggressive tumors

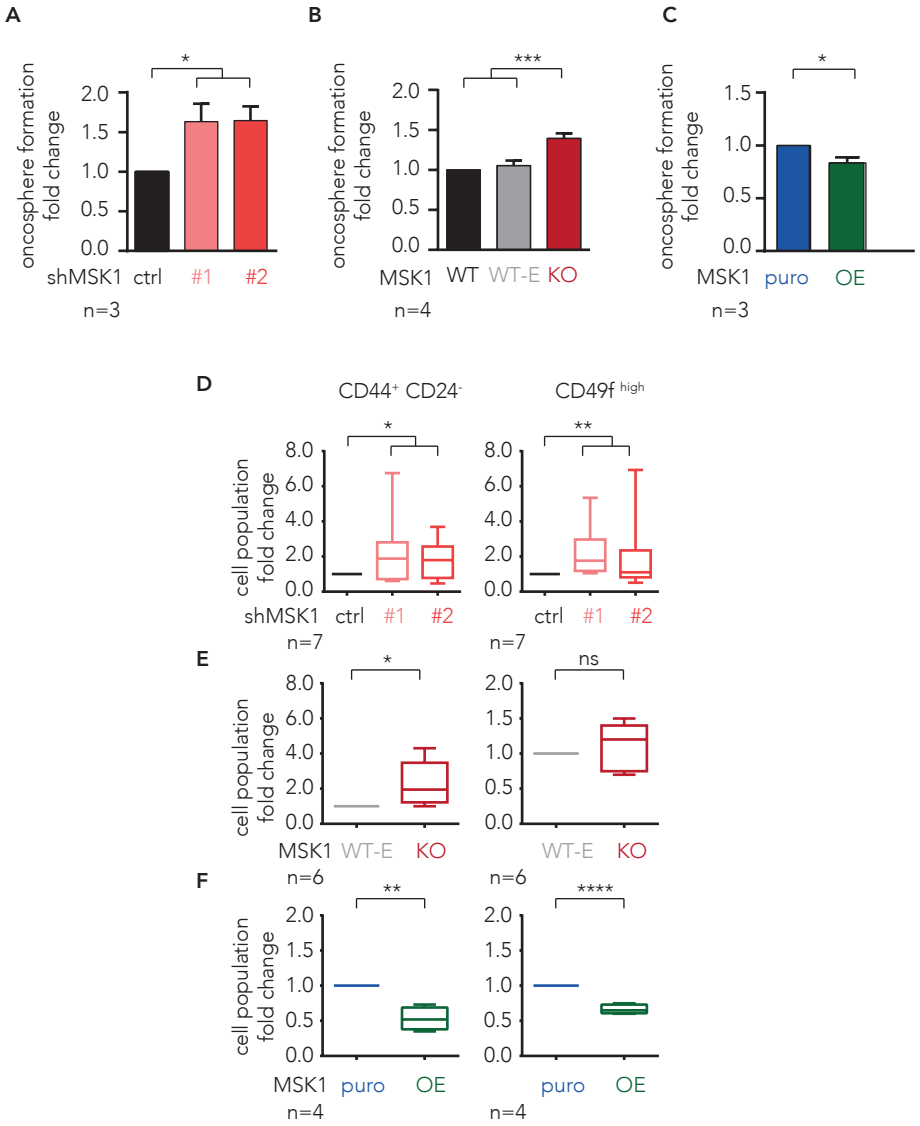

Figure 28. MSK1 silencing promotes tumor initiation through breast cancer stem cells population. (**a-c**) 2D Oncosphere formation assay of MSK1sh, MSK1 KO cells and MSK1 overexpressing cells, respectively. (**d-f**) Quantification of $CD44^{+}CD24^{-}$ cells (left) and $CD49f^{high}$ (right) in the MSK1sh, MSK1 KO cells and MSK1 overexpressing cells, respectively. Panels a-c show data as mean ± SEM and panels d-f as whisker plots: mid-line, median; box, 25 – 75 percentile; whisker, minimum to maximum. Statistical significance: ns, not significant $p>0.05$; *, $p\leq0.05$; **, $p\leq0.01$; ***, $p\leq0.001$; ****, $p\leq0.0001$.

(Jones, Ford et al. 2004). Quantitative PCR showed that upon MSK1 depletion expression of luminal keratins 7 and 18 was significantly reduced and basal keratin 14 mRNA level increased (Figure 29a and b). Next, we extended this analysis to MCF 7 cell line that is another model of ER+ breast cancer. These experiments confirmed previous results obtained with DBM cells (Figure 29c).

Altogether, these results show that MSK1 controls broad range of genes including *CD24*, *CD44*, *CD49f* and keratins *KRT7*, *KRT14*, *KRT18* playing a key role in the expansion of tumor-initiating cells and cancer cell differentiation. Cells, upon MSK1 depletion, become poorly differentiated and more prone to initiate new lesions in the bone.

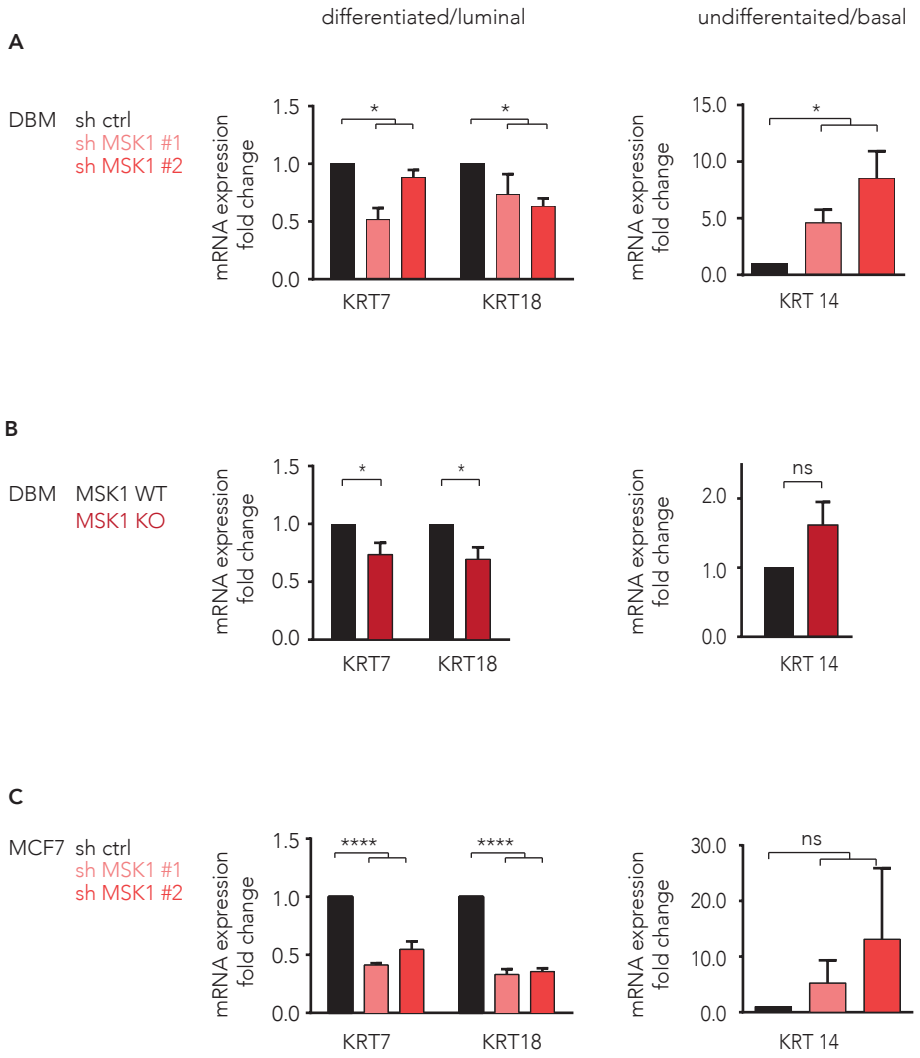

Figure 29. MSK1 controls keratins expression profile in ER+ breast cancer cells. *KRT7*, *KRT18* (left) and *KRT14* (right) mRNA expression in (**a**) sh control and sh MSK1 infected DBM cells. (**b**) MSK1 WT and KO DBM cells. (**c**) sh control and sh MSK1 infected MCF7 cells. . All panels show data as mean ± SEM from 3 biological replicates. Statistical significance: ns, not significant $p>0.05$; *, $p\leq0.05$; ****, $p\leq0.0001$.

4.5 MSK1 modulates Polycomb group target genes

To investigate the mechanisms responsible for the MSK1 dependent bone metastasis suppression in ER+ breast cancer patients we performed an expression correlation analysis followed by a gene set enrichment analysis (GSEA). In this analysis we used transcriptional profiles from ER+ primary breast tumor speciments to correlate MSK1 gene expression with the expression of other genes. To this end, genes positively correlated with MSK1 expression were classified by gene ontology (GO) terms, ordered using normalized enrichment score (NES) and assigned to four categories. Results show that MSK1 expression positively correlates mainly with genes belonging into two categories: chromatin modifications - transcription and splicing - translation (Figure 30a). In detail, data of cellular compartments GO indicate that MSK1 co-expresses with genes belonging to transcriptional repressor, histone deacetylase and Polycomb group (PcG) complexes (Figure 30b). Similar pattern of gene set enrichment was found in the biological processes GO meaning that results are robust and not restricted only to one GO term (Figure 30c). This analysis pinpoints that MSK1 may regulate gene expression by two mechanisms. In the first one epigenetic changes in chromatin modifications alter gene transcription. Whereas the second possible mechanism is related to splicing and regulation of gene translation.

The GSEA results combined with the literature review suggest that the loss of MSK1 expression or activity can promote bone homing by epigenetic regulation of developmental gene transcription. We developed this hypothesis based on two facts. First, chromatin modification such as serine phosphorylation in histone H3 tail is a transcription-activating epigenetic mark. This modification is in particular important for the transcription of PcG-regulated genes.

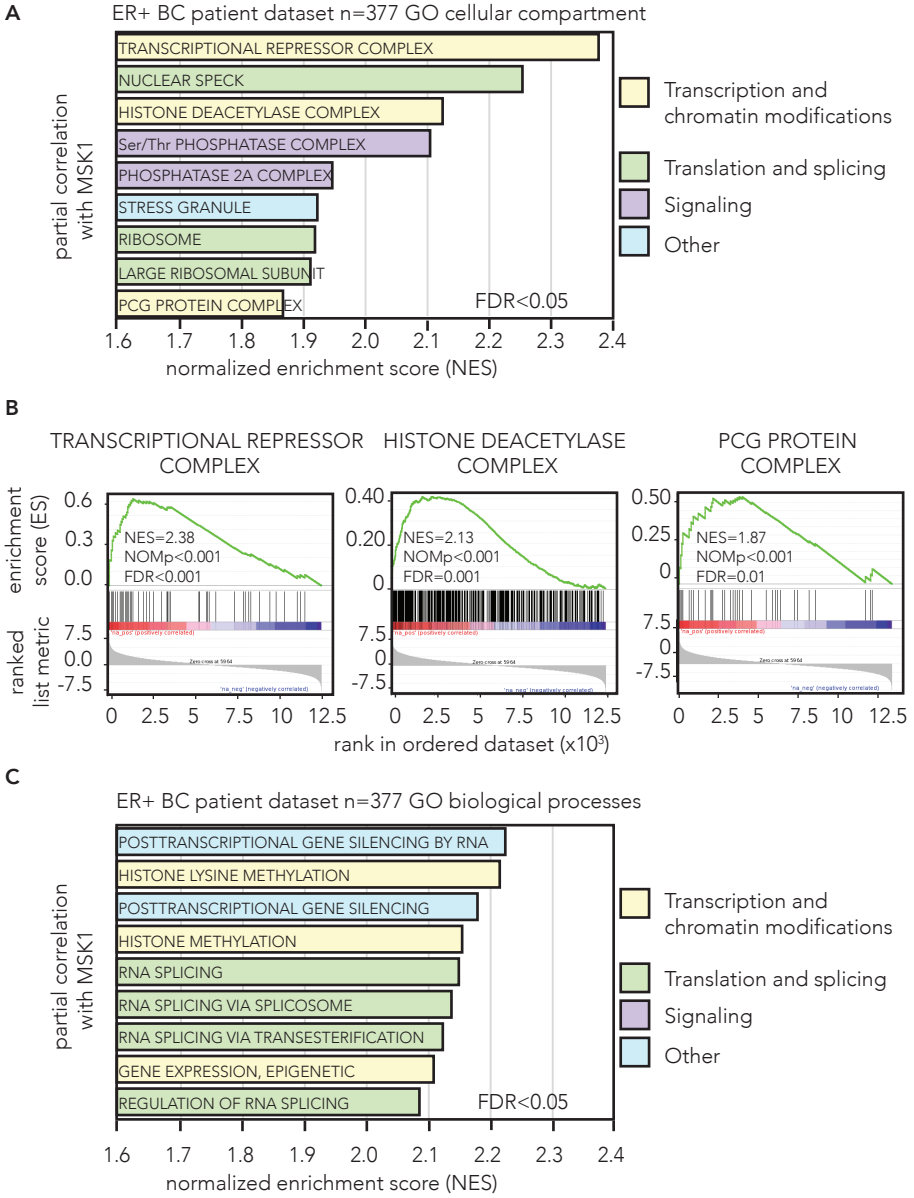

Figure 30. MSK1 positively correlates with genes implicated in transcription and chromatin modification. (**a**) Cellular component gene ontology enrichment analysis of genes correlated with MSK1 in ER+ breast cancer (BC) dataset. GO were assigned to 4 color-coded categories based on the function. (**b**) Gene set enrichment analysis histograms of 3 GOs from transcription and chromatin modification category. (**c**) Biological processes gene ontology enrichment analysis of genes correlated with MSK1 in ER+ breast cancer dataset. GO were assigned to 4 color-coded categories based on the function. FDR, false discovery rate; NES, normalized enrichment score; NOMp, nominal *p* value.

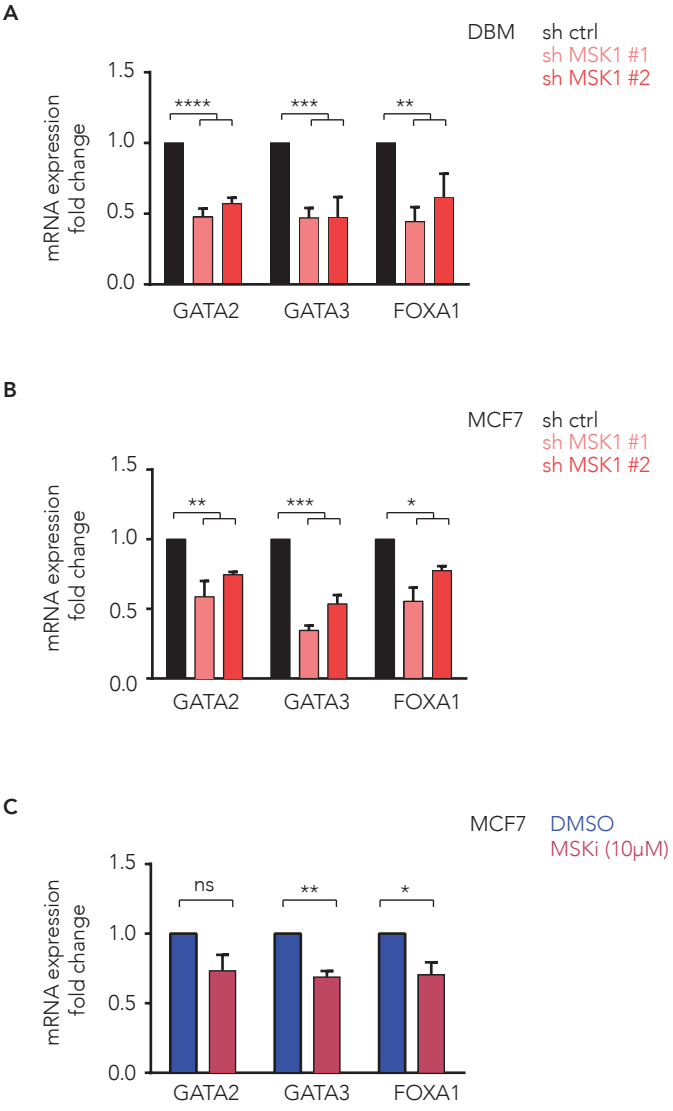

Figure 31. MSK1 controls GATA and FOXA transcription factor expression profile in ER+ breast cancer cells. *GATA2*, *GATA3*, *FOXA1* mRNA expression in (**a**) sh control and sh MSK1 infected DBM cells. (**b**) sh control and sh MSK1 infected MCF7 cells. (**c**) MCF7 cells treated with MSK1 inhibitor and DMSO. All panels show data as mean ± SEM from 3 biological replicates. Statistical significance: ns, not significant $p>0.05$; *, $p\leq0.05$; **, $p\leq0.01$; ***, $p\leq0.001$; ****, $p\leq0.0001$.

Two independent studies showed that, in fibroblasts, MSK mediated histone H3 serine 28 phosphorylation (H3S28ph) can profoundly disrupt PcG silencing and activate transcription (Gehani, Agrawal-Singh et al. 2010; Lau and Cheung 2011).

Second, the majority of PcG target genes plays a pivotal role in developmental regulation and cell differentiation. Genome-wide approaches identified genes encoding the transcription factor families: FOX, SOX, GATA, and TBX as targets of PcG complexes in human and mouse embryonic stem cells (Boyer, Plath et al. 2006; Lee, Jenner et al. 2006). Moreover, some of the transcription factors belonging to these families have been associated with breast cancer. In particular, high expression of PcG-regulated genes GATA binding protein 3 (*GATA3)* and forkhead box A1 (*FOXA1)* strongly correlates with differentiated ER+ luminal A breast cancer subtype and good prognosis for patients (Mehra, Varambally et al. 2005; Ross-Innes, Stark et al. 2012).

To test whether MSK1 regulates transcription of PcG target genes in ER+ breast cancer we analyzed expression of three transcription factors *GATA2*, *GATA3*, *FOXA1*. Analyzed PcG targets were significantly downregulated upon MSK1 silencing in DBM and MCF7 cell lines (Figure 31a-b). In line with this data, MSK inhibitor treatment also decreases expression of *GATA2*, *GATA3*, *FOXA1* (Figure 31 c). Next, we extended our findings to *in vivo* setting using DBM cells. Results of xenograft experiment indicate that MSK1 depletion is associated with reduced FOXA1 protein levels in bone metastatic lesions (Figure 32).

These results suggest that, in ER+ breast cancer, metastatic cells may benefit from the repression of differentiation genes belonging to GATA and FOX families. In consequence, reduced expression of differentiation-related and enhanced expression of

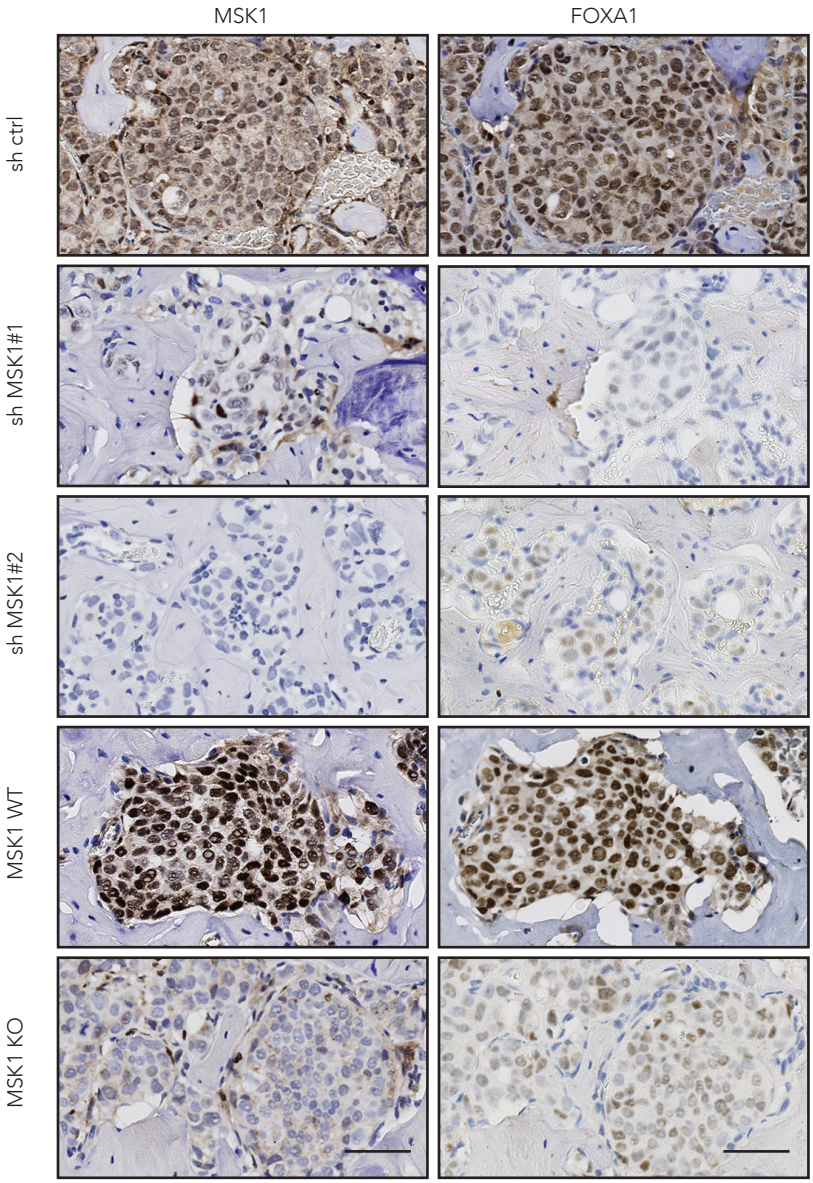

Figure 32. FOXA1 expression decreases upon MSK1 downregulation in bone metastatic lesions. (**left**) MSK1 staining of hind limbs bones from mice injected with DBM control (sh control or WT), MSK1- silenced (shMSK1) or MSK1- ablated (KO) cell lines. (**right**) FOXA1 staining of hind limbs bones. Scale bar: 50 μm

stem-cell factors enforces cancer cell plasticity. Upon MSK1 depletion metastatic cells shift their phenotype from differentiated luminal toward poorly differentiated basal.

DISCUSSION

1. Long-latency is maintained by dormant micrometastases composed of quiescent, slow-cycling, and rapidly proliferating cells and apoptotic cells

In a significant fraction of breast cancer patients, symptomatic metastases in the bone emerge after years or even decades of latency. How metastatic cells dissaminated in the bone or micrometastatic lesions keep dormant and undetecteble is a major question in metastasis research. Here, we report the identification and functional analysis of molecular mechanisms involved in the latency of ER+ breast cancer. We developed a model of breast cancer dormancy to identify genes relevant for long-latent relapse. Next, we used this model to perform *in vivo* loss-of-function shRNA screening. The screening revealed MSK1 as a metastatic dormancy regulator. The *in vivo* and *in vitro* validation results indicate that MSK1 plays a role in homing and differentiation of metastatic cells. Importantly, low MSK1 gene expression levels associate with early metastasis in ER + breast cancer.

Based on *in vivo* selection we isolated metastatic clones that form bone lesions after a notable and reproducible phase of latency. This technique has the capacity to enhance organ tropism of metastatic cells (Kang, Siegel et al. 2003; Minn, Gupta et al. 2005). In most studies, ER- cell lines have served as a model of bone metastasis, although studies based on ER+ cell lines, including MCF7, ZR-75.1, MCF7/Neu and T47D, have been reported (Yi, Williams et al. 2002; Yin, Mohammad et al. 2003; Lu, Mu et al. 2011). Nevertheless, these models failed to recapitulate long-latent disease in the ER+ setting. In our model, a derivative population of ER+ DBM cells homes the bone, manifests tumor mass dormancy, and forms osteolytic lesions, thus making it a novel and unique tool in the field. Strikingly, in our model of ER+ breast cancer bone metastasis

the latent phase lasts up to 5 weeks. In a series of experiments, we characterized the proliferation status, division turnover, and apoptosis rates of cells in the latent phase and during outgrowth into overt metastases. The results show that solitary cells proliferate after reaching the bone, while dormancy is established in the micrometastatic lesion, where fewer cycling but more label-retaining cells and apoptotic cells in comparison to the metastatic lesions were observed. Therefore, tumor mass dormancy mechanisms mainly govern latency in our model. Despite we can not exclude that mechanisms of cellular dormancy manifested as quiescence are exclusive to solitary cells we suggest that they may contribute to tumor mass dormancy.

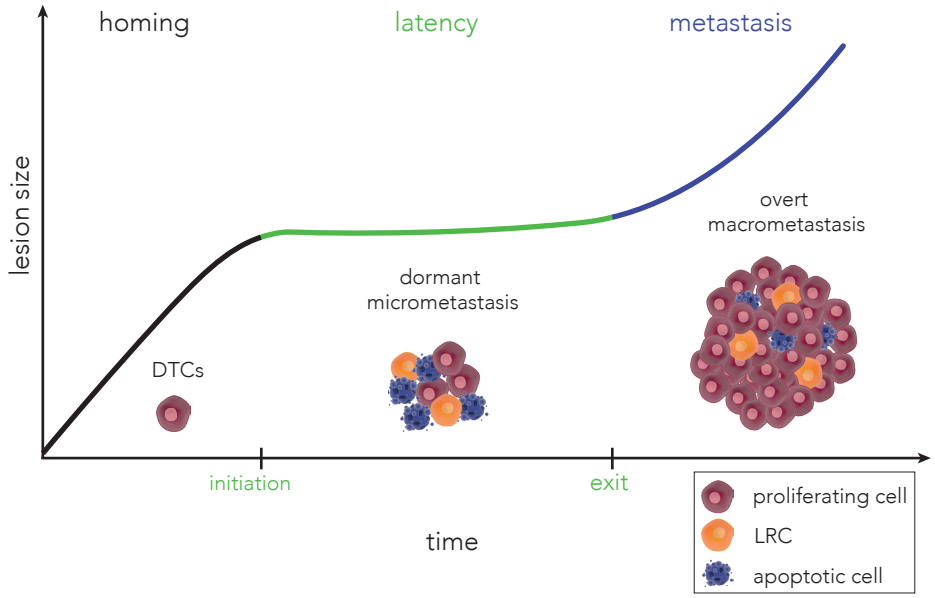

Figure 33. Metastatic latency is maintained by dormant micrometastases. In our model of ER+ breast cancer metastatic progression to bone divides into three steps: homing, latency, and metastasis. In homing (black) phase DTC proliferate to form micrometastases, which enter latency phase (green) by activation of dormancy. These dormant lesions are composed of arrested and slow cycling LRC, proliferating and apoptotic cells in balanced numbers that reflect on stable lesion size. Micrometastases eventually exit from dormancy, grow fast and form overt metastasis (blue), consist in majority of proliferating cells.

Recent studies revealed that the dormancy of micrometastatic lesions is an important contributor to long-latency. Using a human ER- breast cancer cell line (Lu, Mu et al. 2011), a single clone population was isolated that infrequently formed overt metastases from dormant micrometastases in the bone. Also, mathematical modelling showed that patients with long-latent breast cancer have between 1 and 5 micrometastases at 10 years post-resection, thereby indicating that small numbers of lesions maintain dormancy (Willis, Alarcon et al. 2010). In contrast, several lines of evidence indicate that cellular quiescence of solitary cells is a major contributor to long latency. DTCs in the bone marrow of breast cancer patients are largely non-proliferative and, in contrast to CTCs, can persist in the target organs for long periods (Klein 2011). A recent report showed that, upon orthotropic injection, human ER- breast cancer cells disseminate to various organs, including liver, lung, brain, and bone marrow, and undergo cellular dormancy before the formation of micrometastases (Ghajar, Peinado et al. 2013). Also, in a syngeneic mouse model of breast cancer, dormancy is governed by the quiescence of solitary cells. Single 4TO7 cells enter arrest immediately upon infiltrating the lung and are therefore unable to form micrometastatic lesions (Gao, Chakraborty et al. 2012).

Collectively, this set of observations derived from different experimental models suggests that the cellular dormancy is associated mainly with solitary cells, while dormant macrometastatic lesions are considered to consist of actively proliferating cells balanced with the same number of apoptotic cells. Moreover, these two forms of dormancy seem to be exclusive and sequential events (Wells, Griffith et al. 2013). In our model of latency, we demonstrate that dormant micrometastases are composed of proliferating and apoptotic cells; however, we also found arrested and slow cycling

cells. We believe that, G0 cell dormancy is responsible not only for the arrest of solitary cells, but also contributes to tumor mass dormancy, thereby suggesting that a variety of mechanisms can synergistically promote long latency (Figure 33).

2. Genome-wide screening as a tool to identify dormancy-associated genes

To identify genes that control dormancy in the long-latent bone metastasis we chose the RNAi-based whole-genome screen approach. This method has several features that secure the robustness and feasibility of the screening. For example, the mechanisms of RNA silencing by shRNA are well established. Moreover in comparison to genome-editing by CRISPR/Cas9, RNAi does not require a clonal selection, which is a drawback in studding heterogeneous population of cancer cells.

However, in addition to the identification of genes specific for metastatic progression, the screen strategy might also lead to uncover general controllers of cell proliferation or survival *in vitro* and *in vivo*. Therefore, mechanisms of micrometastatic dormancy did not exclusively select for the enrichment of shRNAs in the screen. Viral transduction and *in vitro* culture may have a direct effect on the results due to depletion or enrichment of proliferation-associated genes. We addressed this caveat by quantifying shRNAs from cultured cells, which served as control samples, followed by eliminating the under- and over-represented constructs. The top cut-off chosen corresponded to the quadrupled representation of each shRNA in the pooled library used for the transduction. On the other hand, under-represented shRNA constructs silence genes that are essential for cell proliferation. In the screen, 18% of shRNAs

were undetected in the control samples. In addition, because the candidate gene selection method was based on the fold change enrichment, we eliminated shRNA represented below the lower cut-off point of 10 in the control samples to reduce false-positive hits. As a result of the analysis, more than 28% of the initial shRNAs corresponding to 55% of genes in the library were detected in cultured cells after transduction, and these shRNAs were used as a reference for changes in shRNA representation post-injections. In contrast, after the metastatic process, 0.5% of shRNAs were enriched more than 1.5 fold, suggesting that a minority of genes drives metastatic traits in breast cancer cells. Other groups have reasoned and operated on similar premises in order to select target genes from *in vivo* shRNA screens. For example, the method of shRNA abundance comparison was used to identify genes that became significantly depleted during primary tumor formation in mammary fat pad xenografts (Possemato, Marks et al. 2011). Methods for the exclusion of general proliferation-associated genes in a loss-of-function screen to uncover transcriptional regulators that govern the self-renewal capacity of hair follicle stem cells were also used previousely (Chen, Heller et al. 2012).

In our screen, the list of enriched shRNAs in the metastatic process consisted of 328 corresponding to 322 genes. We refined it by creating two short-lists of potential dormancy controllers containing genes that were down-regulated in 2 or more animals or were silenced by 2 or more independent shRNAs. These selection steps dramatically reduced the number of hits; yet only 2 genes fulfilled both criteria. Similar refinement principles have been used previously, and there is general consensus that selecting a gene consistently silenced in distinct biological and technical replicates increases the specificity of the screen (Chen, Heller et al. 2012).

Strikingly, the screen results indicated that metastatic properties were enriched in all of the library-transduced groups, independently of the pool. This observation can be partially explained by the elimination of genes whose silencing induces cell death or reduces proliferation *in vitro* and *in vivo*. Alternatively, the heterogeneity of cells transduced with different shRNAs can lead to greater manifestation of the growth phenotype over cell death. For instance, some oncogenes could be silenced insufficiently and still play an oncogenic role in metastasis. On the other hand, in each pool, downregulation of one of the metastasis suppressors can lead to the formation of lesions. This effect can be synergized by the activity of oncogenes in the pooled cell population.

The use of RNAi carries a risk of potential off-target effects caused by the convergence of the shRNA processing machinery with the endogenous microRNA biosynthesis pathway. The shRNA sequence expressed from the lentiviral vector requires export from the nucleus to the cytoplasm and processing into functional small interfering RNA (siRNA) (Pan, van der Laan et al. 2012). For this reason, its impact on the microRNA machinery is greater than that of the raw synthetic siRNA. In addition to the cleavage of the target mRNA sequence based on base-pairing with the homologous siRNA, off-targets can be introduced in a sequence-dependent manner. For instance, particular shRNA may suppress the expression or translation of a subset of unspecific target genes via partial base pairing. Therefore, optimal design of shRNA constructs in the pooled library using bioinformatics tools helps to reduce sequence-dependent off-targets. In addition, libraries often contain an average of 4-5 distinct shRNAs targeting the same gene. This strategy is based on the premise that each shRNA in a pool has different off-targets but the same on-target effects, thus the shRNA

pools can dilute off-target effects (Jackson, Burchard et al. 2006). In line with this, top hit candidate genes should be downregulated by at least two shRNA constructs. Given the broad activities of micro-RNAs in regulating gene expression and other biological processes such as virus infection (Pan, van der Laan et al. 2012), off-targets can be introduced in a sequence-independent manner. Cell transduction of the viral vector also potentially induces non-specificity. Therefore, including the appropriate experimental controls is crucial to minimize risks of misinterpretation of the RNAi screening caused by non-specific effects. For our shRNA screening, we used a control sequence as negative control for viral transduction and metastatic progression.

3. Relevance of MSK1 in metastatic dormancy of long-latent ER+ breast cancer

As a first approach, we aimed to functionally validate the role of MSK1 in the dormancy. To this end, we performed a loss-of-function experiment and selectively downregulated the protein level in DBM cells. MSK1 downregulation by two shRNAs enriched the tumor-initiation capacity of DBM cells, resulting in a greater number of bone metastatic lesions. This assay recapitulates the *in vivo* conditions used in the screening and excludes potential false-positive or passanger hits. Despite the significant increase in fold change and the effective silencing of a gene, passenger hits would have no effect or a minimal effect on promoting metastasis *in vivo*. In contrast, downregulation of drivers reflects on the number of metastases or the kinetics of the metastatic process since their downregulation leads to a selective advantage for metastatic cells.

As mentioned above, RNAi technology is prone to off-target effects that can influence screening results. In order to exclude pos-

sible non-specific effects on the phenotype, we used an alternative technique of gene-editing. The CRISPR/Cas 9 method allowed us to ablate MSK1 protein by disrupting its gene. Therefore we were able to by-pass the sequence-dependent and -independent off-targets induced by RNAi. Moreover, we chose a plasmid delivery method other than the lentiviral system, thus eliminating another potential bias. Importantly, low MSK1 levels increased the metastatic potential of DBM cells independently of the MSK1 depletion method used.

Low or medium levels of MSK1 expression in patients with ER+ breast cancer significantly decreases bone metastasis-free survival. This finding suggests the clinical relevance of our result in the long-latent type of brest cancer. Furthermore, bioinformatics analysis revealed that MSK1 downregulation correlates with early recurrence, thereby linking MSK1 not only to general bone metastatic progression but specifically to early onset of bone relapse.

4. MSK1 mediates luminal differentiation by regulating the expression of GATA-3 and FOXA1 transcription factors

Using an RNAi genome-wide screening approach, we have identified MSK1 as a dormancy-enforcing gene in ER+ breast cancer to bone metastasis. Functionally, reduced MSK1 levels increase the tumor-initiating capacities of metastatic cells and the number of poorly differentiated cells. Mechanistically, we showed that MSK regulates the expression of GATA-3 and FOXA1 transcription factors, which are involved in luminal cell differentiation. Clinically, low MSK levels are associated with early relapse in ER+ patients (Figure 34).

Our data indicate that MSK1 can be an upstream controller of luminal differentiation through the transcriptional regulation of

FOXA1 and GATA-3. FOXA1 is a 'pioneer' forkhead transcription factor that can directly bind condensed chromatin, displace repressive linker histones, and recruit other transcription factors to promote transcription (Sekiya, Muthurajan et al. 2009). Importantly, in ER+ breast cancer, FOXA1 is a prognostic marker and it may also prevent metastatic progression of luminal subtype breast cancers by controlling differentiation (Mehta, Jain et al. 2012). GATA-3 is another transcription factor that acts as a master regulator of luminal

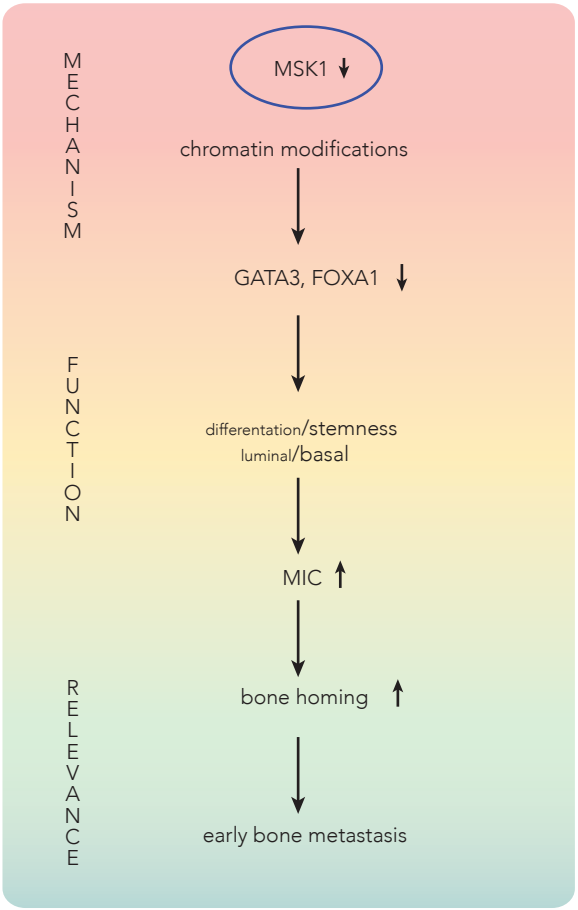

Figure 34. MSK1 loss in breast cancer cells increases metastatic capacities. Metastatic cells with low MSK1 levels have downregulated expression of luminal transcription factors, and in consequence, less differentiated phenotype. This leads to an increase in metastatic-initiation capacities and altered bone homing, which in patients corresponds to early metastasis onset.

differentiation. Experiments using the MMTV-PyMT mouse model of breast cancer showed that GATA-3 maintains tumor differentiation and suppresses dissemination and metastasis (Kouros-Mehr, Bechis et al. 2008). In breast cancer, GATA-3 has emerged as a strong and independent predictor of tumor differentiation and clinical outcome since low GATA-3 expression was found to be strongly predictive of high tumor grade, positive lymph node status, and large tumor size (Jenssen, Kuo et al. 2002).

The mechanistic insights into the involvement of FOXA1 and GATA-3 in maintaining luminal differentiation and plasticity of human breast cancer cells have been reported in two studies. The first revealed an ER-independent, luminal-specific function for FOXA1 in maintaining the differentiated characteristics of luminal breast cancer cells through transcriptional regulation of both luminal and basal genes (Bernardo, Bebek et al. 2013). In particular, FOXA1 regulates plasticity between basal and luminal breast cancer cells, not only by inducing luminal genes, but also by repressing the basal phenotype, and thus aggressiveness. Those authors concluded that loss of FOXA1 may lead to growth arrest of a subpopulation of differentiated cells, while the remaining cells may have greater plasticity to de-differentiate towards the basal phenotype. Using an integrative network approach of transcription factors and regulons in breast cancer, the second study identified 2 clusters of transcription factors that regulate gene expression in ER+ and ER- breast tumors (Castro, de Santiago et al. 2016). FOXA1 and GATA-3 belong to a cluster that controls gene expression in ER+ (luminal) breast cancer (cluster 1). Importantly, these results also highlight the possibility that repression of the luminal cluster of transcription factors leads to a shift in cancer cell state toward more basal-like,

which is associated with a more aggressive tumor phenotype and resistance to therapy.

Therefore, we hypothesize that the loss of MSK1 expression may lead metastatic cells to a partial shift from a differentiated luminal to a basal phenotype - the latter mediated by FOXA1 and GATA-3. Crucially, long periods of metastatic latency are predominantly observed in the patients with ER+ breast cancer, in particular luminal A subtype. In this subtype luminal genes, including FOXA1 and GATA-3, are expressed at high levels, while basal at low. Whereas, metastasis in basal type subtypes, with low expression of luminal genes, emerges after a short latency.

5. MSK1 may regulate the expression of luminal and basal genes by epigenetic mechanisms

The mechanism by which MSK1 regulates the expression of basal and luminal genes in metastatic cells remains unknown. Nevertheless, based on a preliminary data and the available literature, we propose a working model in which MSK1 contributes to histone-crosstalk to regulate the expression of luminal differentiation and basal genes. A bioinformatic analysis of MSK1 correlation with transcription of other genes in cohort of primary breast tumors suggests that MSK1 controls gene transcription through epigenetic modifications. Alternatively, we hypothesize that MSK1 is involved in the translational regulation. Currently, our results support the first hypothesis because MSK1 depletion was found to modify the expression of genes belonging to the FOX and GATA and. The expression of these transcription factors is controlled by a Polycomb group complex in embryonic development and during the differentiation of adult tissues (Boyer, Plath et al. 2006; Lee, Jenner et al. 2006). PcG-dependent regulation of gene transcription is

tightly related to chromatin modifications, including methylation, acetylation, and phosphorylation. Studies performed on fibroblasts demonstrate that stress-induced activation of MAPK signaling activates MSK1 to phosphorylate histone 3 serine 28 (see introduction section 3.2). Furthermore, H3S28ph was demonstrated to counteract PcG-mediated silencing by facilitating the dissociation of Polycomb repressive complexes (PRC) in response to external signaling. In details, stress-induced MSK1 mediates a functional interplay between methylated (H3K27me3) or acetylated (H3K27ac) lysine 27 and phosphorylated serine 28 (H3S28ph) on histone H3 function on immediate-early and developmental gene promoters. Another study reported a MSK1-dependent transcription of PcG target genes in terminally differentiated neurons (Sodersten, Feyder et al. 2014). While H3S28ph mark increases, repressive mark H3K27me3 decreases with a concomitant release of the transcription repressor PcG, which specifically binds histones at H3K27me. Importantly, this mechanism of PcG-mediated repression by H3K27me is not found for immediate-early genes and appears to be restricted to developmental genes (Gehani, Agrawal-Singh et al. 2010). Also, the activating mark H3K27ac increases with H3S28ph at developmental gene promoters, thereby suggesting a codependency of histone phosphorylation and acetylation (Lau and Cheung 2011). Additionally, a genome-wide analysis of H3S28 phosphorylation in stress-induced fibroblasts also revealed a high enrichment of H3S28ph mark for genes involved in development and morphogenesis (Sawicka, Hartl et al. 2014). Finally, a recent study on *Drosophila* demonstrated that the H3S28 residue is essential for efficient PcG-mediated gene repression during development. Impairment in S28 phosphorylation in the serine to alanine (H3S28A)

mutant reduces H3K27 methylation and compromises PcG-mediated silencing (Yung, Stuetzer et al. 2015).

The above observations and results led us to conclusion that MSK1 regulates the expression of basal and luminal differentiation genes in metastatic cells through epigenetic mechanisms. However, to date our data support two distinct effects of MSK1 depletion on PcG-regulated gene transcription. First, the loss of MSK1 activity enhances PcG-mediated silencing of luminal differentiation genes GATA-3 and FOXA1.This might be a result of a decrease in H3S28ph and subsequent increase in a repressive mark H3K27me3 and reduction of an activating H3K27ac mark (Figure 35a). The mechanism of epigenetic silencing of FOXA1 gene expression in breast cancer has been demonstrated in BRCA-1 (breast cancer 1) deficient and wild-type cell lines (Gong, Fujino et al. 2015). Accord-

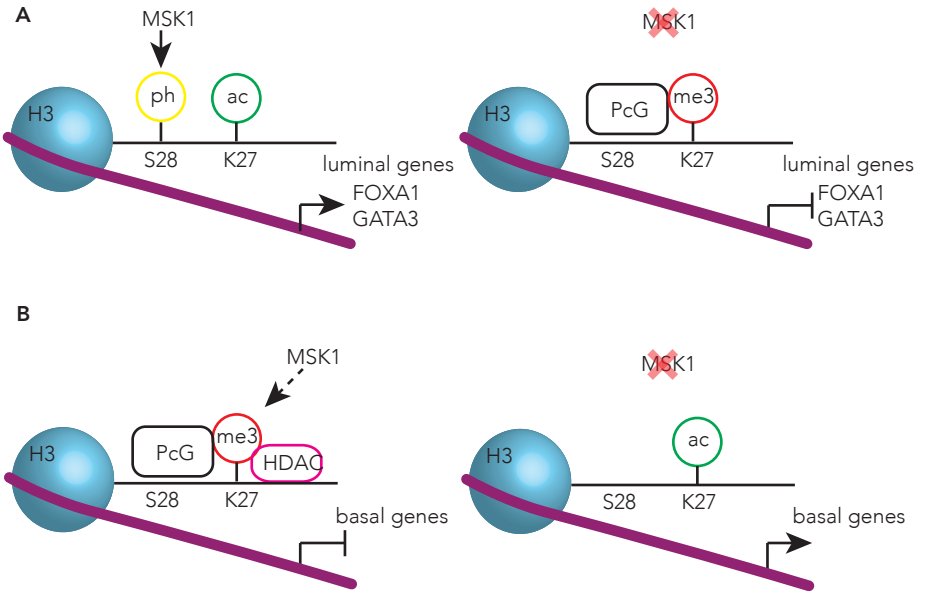

Figure 35. A working model of MSK1 role in epigenetic regulation of gene expression. (**a**) MSK1 phosphorylates H3S28 facilitating a transcription of luminal differentiation genes (left). Upon MSK1 depletion and lack of H3S28ph mark PcG complex represses transcription of luminal differentiation genes (right). (**b**) High level of MSK1 indirectly enhances transcriptional repression of basal genes (left). Loss of MSK1 leads to expression of basal genes (right).

ing to these studies, BRCA-1 activates FOXA1 expression in BRCA-1 wild-type breast cancer cell lines. In particular, FOXA1 is negatively regulated by promoter methylation since the H3K27me3 repressive histone mark is highly enriched on the FOXA1 promoter in BRCA-1-deficient cells compared with the wild-type BRCA-1-expressing cells.

Second, co-expression correlation data suggest that loss of MSK1 activates the transcription of basal genes, potentially from SOX family, since histone methylation, transcriptional repression, and histone deacetylase complexes are co-expressed with MSK1. Histone deacetylases remove the acetyl groups from histones and, consequently, deactivate transcription (Figure 35b). ChIP-sequencing analysis could be used to unreveal how MSK1contribute to the transcriptional differences in basal and luminal gene expression and subsequently to the latency in breast cancer. ChIP focused on the immunoprecipitation of the DNA bound to histone H3 activating marks H3K9ac, H3K27ac, repressive marks H3K9me3, H3K27me3, and phosphorylated histone H3 H3S10ph and H3S28ph would provide valuable insights into this point. Sequencing of DNA fragments would show which marks are present on the promoters of basal and luminal genes after MSK1 depletion. This would help to study the changes in the chromatin landscape that lead to differences in the gene expression pattern.

Collectively, we provide the clinical and mechanistical data indicating a role of MSK1 in dormancy and the establishment of the latent bone metastases of ER+ breast cancer. Dormant tumor cells and micrometastases have significant clinical importance during the course of long-latent metastasis,. The eventuall exit from dormancy leads to metastatic relapse as dormant cells may serve as the source of therapy-resistant cells.

CONCLUSIONS

We have esteblished a model of dormancy that mimics metastatic patterns of long-latent ER positive breast cancer metastasis.

Tumor mass dormancy characterized as a balance of cell proliferation and apoptosis mainly controls latency in our model. However, solitary cell dormancy also contributes dormancy of micrometastases in latent phase.

Genome-wide loss of function screening identifies MSK1 as a dormancy-enforcing gene in ER positive breast cancer.

MSK1 depletion supports bone homing of metastatic cells and is associated with early bone metastasis in ER positive breast cancer patients.

MSK1 depletion in dormant metastatic cells increases tumor-initiating capacities.

MSK1 supports luminal differentiation regulating the expression of luminal transcription factors FOXA1 and GATA-3.

MATERIALS AND METHODS

1. Methods

Cell culture

Human ductal adenocarcinoma cell line T47D and human embryonic kidney 293T cells were purchased from ATCC. Dormant bone metastatic subline T47D-5516 (DBM) was derived from parental T47D cell line following *in vivo* selection procedure similar to described in (Kang et al. 2003) and BoM2 bone metastatic subline was derived from MCF7. All cell lines were cultured in standard conditions (37°C, 5% CO_2) in DMEM medium supplemented with 10% FBS, 50 U/ml penicillin, 50 µg/ml streptomycin and 5% glutamine. For genetically modified derivatives puromycin (2µg/ml) was added to select the expression of transgenes. All cell lines with exception of HEK293T were stably transduced with TK-GFP-Luc construct and sorted for GFP expression.

Animal studies

All animal work was approved by the institutional animal care and use committee of IRB Barcelona. Female BALB/c nude mice of 11-13 weeks of age were used for all studies. Prior to all surgical procedures mice were anesthetized using a mixture of ketamine (100 mg/kg) and xylazine (10 mg/kg). For experiments with T47D, DBM or BoM2 cell lines 60 or 90-day release estrogen (ß-estradiol 0.18 mg/pellet) pellets were subcutaneously provided unless otherwise indicated. Mice were monitored weekly using IVIS imaging system, unless otherwise indicated.

For intra-cardiac injections, $5x10^5$ cells were injected into the left cardiac ventricle of the mice, using a 26G needle as previously described (Yin et al. 1999). Immediately after injection mice were imaged for luciferase activity to confirm successful xenograft.

For intra-tibiae injections, the injection site was prepped with betadine scrub followed by a 70% alcohol wipe. A 1-cm skin incision was made on the antero-medial part of the leg and the muscle was moved using blunt forceps. The bone was drilled using a 26G syringe-needle. 6 µl of cell suspensions in PBS were injected into the upper half of the tibia medullary cavity. The skin was sutured back and inoculations were confirmed by BLI. The number of injected cells depends on experimental design but does not exceed $3x10^4$.

EdU or BrdU incorporation experiments were done by single or multiple intraperitoneal injection of indicated compound (50 mg/kg).

Bioluminescent imaging

For bioluminescent imaging, mice were anesthetized and injected retro-orbitally with D-luciferin (75mg/kg) at the indicated times after xenografting. Animals were imaged in an IVIS 100 chamber within 1 min after D-luciferin injection, and data were recorded using Living Image software (Xenogen). To measure bone colonization, photon flux was calculated for each mouse by using 2 circular regions of interest encompassing the hind limb of the mouse. After subtracting a background value obtained at indicated day of the experiment, photon flux was normalized to the value obtained at the day of xenografting. Metastatic colonization was detected when the value of photon flux of the bone lesion was greater that BLI signal at day 0.

IVIS SpectrumCT instrument was used to obtain BLI images integrated with low dose microCT. Anesthetized animal, previously injected with D-luciferin, was placed on stable revolving platform that rotates 360° to acquire 3D images. Quantitative biolumines-

cence data and CT images were processed, reconstructed and co-registrated using DLIT Reconstruction option in the Living Image software.

Microarray assay

RNA was extracted from cells using the PureLink mini kit following manufacturer's instructions. Labelling and hybridization of the samples to the HG1.0ST gene expression chip (Affymetrix) were performed by the Functional Genomics Core Facility of IRB Barcelona using standard methodology. Data analysis was performed using R (Bioconductor) by the Biostatistics/Bioinformatics unit at IRB Barcelona. The posterior expected false discovery rate was set at 0.05.

Copy number alteration analysis

High-molecular genomic DNA was isolated from cells using GeneElute Mammalian Genomic DNA Miniprep Kit following manufacturer's instructions. DNA quantity and quality was determined by NanoDrop spectrophotometer and electrophoresis in 1% agarose gel. Genetic aberrations were detected using NimbleGen Human CGH 3x720K Whole-Genome Tiling v3.0 Array consisting of 72,000 probes. Samples were independently labeled with Cy3 and Cy5 fluorochromes, and co-hybridized. The copy number analysis was performed using Bioconductor. Briefly, log2 fold changes were normalized the by a mode normalization and outlier smoothing (Venkatraman and Olshen 2007).

X-ray scanning

Development of bone metastasis was monitored by X-ray imaging CT-Scan (SkyScan). Images were acquired at 50 kV with a 0.5 aluminium filter using a detection pixel size of 5 mm. Visible meta-

static lesions were measured using Image J software and osteolytic area and calculated in arbitrary units.

Histopathology and immunohistochemistry

Hind limb bones were excised, fixed in 10% neutral-buffered formalin O/N at RT. Then, they were washed twice with PBS and decalcified with Osteosoft bufferfor 15 days at RT before embedding in paraffin. Paraffin sections were stained with hematoxylin and eosin (H&E) or subjected to immunostaining. For staining with antibodies, paraffin sections were deparaffinized and rehydrated through a series of decreasing alcohol dilutions. Antigen retrieval was performed with citrate buffer (pH 6) or TRIS-EDTA buffer (pH 9) depending on the antibody (see Table 8) for 30 min at 95°C. Next, sections were treated with peroxidase-blocking solution for 15 minutes and washed twice with PBS. Incubation with primary antibody diluted in the blocking buffer was done from for 1h at RT or O/N at 4°C. After washing 3 times with PBS secondary HRP-conjugated antibody was applied for 45 min at RT. Slides were washed 3 times with PBS and incubated with DAB for 10 seconds to 3 minutes. Hematoxylin was used as counterstaining dye. Stained sections were dehydrated and mounted.

Images of the stained bone sections were taken at 40x magnifications (2-5 fields per bone lesion, number of bone lesions depends on the animal) and analyzed using ImageJ software. Percentage of positive cells relative to total cell number was quantified. Number of analyzed samples (animal or hind limbs) is indicated below each experiment.

Label retention assay

Cells were harvested and labeled with Vybrant DiD dye diluted 1:500 in PBS for 25 min at 30°C. Next, $2.5x10^5$ cells were seeded in p60 plate and passaged twice a week for 3 weeks. With every passage half of cell number was fixed in 70% ethanol and stored at -20°C for FACS analysis. DiD fluorescence was measured using APC channel and normalized to GFP fluorescence.

Flow cytometry and FACS sorting

Cell from culture were collected by trypsinization while cells from bone metastatic lesions were purified using physical and enzymatic method protocol. Briefly, hind limbs bones were excised and femur was separated from the tibia. Bones were placed in morter filled with 3 ml of ice-cold PBS supplemented with 2% FBS and 1 mM EDTA and crushed. Suspended bone marrow cells were filtrated through cell strainer (70 μm) and collected. After multiple repetition of this procedure, bone fragments were incubated for 45 min at 37°C with digesting medium composed of 0,25% collagenase type 1 and 20% FBS in PBS. After incubation, digestion media were passed through cell strained and added to previously obtained cells. Cells were concentrated by centrifugation for 7 minutes at 1200 rpm and suspended in single-cell condition after passing through a mesh (40μm). Depending on experimental design cells were analyzed by flow cytometer Gallios or sorted by FACS Aria 2.0. For EdU detection Click-It Plus kit for FACS with Alexa 647 was used following manufacturer's instructions.

Cell surface markers were labeled on trypsinized cells with primary antibodies conjugated with fluorophors for 30-45 min at RT (see Table 8).

For analysis cells were gated as following. First, cells were selected in the FSC/SSC dot plot to remove debris. The cells were then gated to exclude cellular aggregates in the FSC/FSC dot plot. Gates of GFP, RFP, APC or Alexa 647 positive cells were set comparing to a control sample that has no detectable fluorochrome expression. For cell surface markers isotype control antibody was used to establish the gates.

Apoptosis *in vivo* assay

In vivo activity of Caspase 3/7 was measured by administration to the mice of VivoGlo Z-DEVD-aminoluciferin (166 mg/kg). BLI imaging was done immediately after substrate injection. 6 hours later animals were imaged for remaining luciferase activity. In case of non-detectable signal, standard D-luciferin was administrated to animals and BLI signals were measured. Apoptotic cell content was assessed by normalization of Z-DEVD BLI signals to BLI signals from the lesion.

Genome-scale loss-of-function screening with a lentivral RNAi library

To identify genes that facilitates exit from dormancy genome-scale loss-of-function screening was performed. The MISSION LentiPlex human pooled shRNA library TRC1.0 used for the screening was purchased from Sigma-Aldrich and all experimental design and optimization was done with support of manufacturer´s instruction. The library consists of over 75,000 shRNA constructs from the TRC collection targeting 16,019 human genes and was pre-divided into ten subpools of approximately 8,000 shRNA constructs each. Between 5 and $8x10^6$ of DBM cells were infected with each pool separately at a multiplicity of infection of 0.4 to ensure that most cells receive only 1 viral construct with high probability

and cultured with puromycin (2 μg/ml) selection for 10 days. After puromycin selection, confirmation of insert integration by PCR and GFP expression cells were injected into mice resulting that each pool was inoculated intracadiacly in 10 animals ($5x10^5$ cells per mouse).Population of DBM cells infected with control shRNA that does not affect exit from dormancy were also injected to animals. Animals were monitored weekly after cell inoculation by BLI. Xenografted cells that had formed metastatic lesions in hind limbs were flushed and GFP-sorted. gDNA was extracted from the preinoculation DBM cells and bone metastasis using GeneElute kit and tested for quality and quantity by NanoDrop spectrophotometer and Qubit DNA assay. The shRNA sequences were amplified using provided primers complementary to the shRNA containing regions.

High-throughput sequencing and data analysis

gDNA samples were submitted to Sigma Deconvolution platform where custom-design procedure was performed in order to assess the abundances of each shRNA clone in samples. Briefly, procedure consisted of amplification of shRNA region and sample barcoding which was followed by next generation sequencing at coverage 1000 and alignment of short reads to the reference. After bioinformatics data clean-up data were obtained as number of shRNA sequences per clone per sample.

Screening hit selection

First, number of shRNA sequences per clone was analyzed in preinoculation samples in order to select the clones within the 10 and 300 range representation. Next, the fold change of the number of shRNA sequences from the metastasis to preinoculation was calculated and genes with FC ≥ 1.5 were chosen. Hits were defined

as genes that had increased in representation by two different shRNA´s or (2 shRNA for gene) or in two or more of the replicate metastatic samples (2 samples for gene).

Patient gene expression data sets

The patients information is publically available and has been downloaded from Gene Expression Omnibus (GEO) (Barrett, Troup et al. 2007). MSKCC/EMC cohort was used for analysis of gene expression in primary tumors and its association to the time and site of metastasis (survival anlysis) and for co-expression correlation analysis. This union cohort is pooled of GSE2603, GSE2034, and GSE12276 cohorts and consist of 560 patients with annotated clinical records including time and site of metastasis. In order to remove systematic biases, the expression measurements were converted to z-scores for all genes prior to merging. ER+ patients were selected based on the bimodality of gene *ESR1* encoding estrogen receptor.

GSE14020 cohort was used for analysis of gene expression in primary tumors and metastatic sites. It is composed of microarray data form breast tumors and bone, lung, liver, and brain metastases collected from 55 patients with annotated clinical records. ER+ patients were selected based on the bimodality of gene *ESR1*.

Survival analysis

Cox model adjusted by ER status (*ESR1* expression based) was applied for survival analysis using coxme R package for fitting. Eklund metrics within each dataset (interaction) scan batch as random effect and a time dependent covariate were used for hazard ratio (HR) estimation before and after 3 years. Next, Kaplan-Meier analysis standardized to the overall population was performed us-

ing reweighting on ER. For specific site metastasis other metastasis sites were excluded from analysis.

Co-expression correlation and GSEA

Association of each gene expression with *RPS6KA5* expression using a mixed-effect model was computed, adjusting by ER status (based on *ESR1* expression), dataset, Ecklund metrics effects within each dataset and scanning batch as random effect. Partial correlations were computed using expression values after correction performed by the mixed-effect model. Multiple contrasts adjustment was done by Benjamini-Yekutieli FDR. Pre-ranked GSEA (Gene Set Enrichment analysis) was used to analyze co-expression correlation analysis (Subramanian, Tamayo et al. 2005) by Bioconductor R package to annotate genes according to gene ontology (GO) terms labels and Kegg pathways.

Viral production and transduction

pLKO lentiviral vectors containing human shRNA sequences were obtained from MISSION TRC1 library (see Table 6). After expansion in *E.coli* bacteria and plasmid isolation 3 µg of pLKO plasmid was mixed in ratio 1:1 with packaging vectors V-SVG, RRE and RSV in the NaCl (150 mM) supplemented with PEI (5.8 µg/ml). For retroviral production 6µg of vector was mixed with V-SVGR and GAG-POL in ratio 1:0.1:1 in transfection media (see Table 5). HEK293T cell seeded in 70% confluency were transfected with plasmid mix and left for 24h incubation. Following day media was changed and cells were transferred to 33°C incubator for additional 48h in order for viral particles production. Medium containing virus was collected, filtrated (0.45 µM), concentrated using 100K centrifugal filters (3000 x g for 1h in 4°C), and supplemented with poly-

brene (8 μg/ml). Recipient cells in 50% confluency were transduced with the viral media using centrifugation at 2250 rpm for 45 min in RT followed by O/N incubation. Viral particles were removed with media and fresh medium was added for 24h to let cells express transgenes. Finally, infected cells were trypsinized and seeded in the media containing selection marker puromycin (2 μg/ml) in order to obtain stable transgene containing population.

Protein extraction and Western blotting

Cells were lysed with a RIPA buffet composed of 25 mM Tris-HCl (pH 7.6), 150 mM NaCl, 1% NP-40, 1% sodium deoxycholate and 0.1% SDS supplemented with protease and phosphatase inhibitors cocktail and sonicated for 5 minutes at medium intensity. After sonication, extracts were centrifuged at 13200 rpm for 15 minutes at 4°C and supernatant was collected for storage in -80°C. Protein concentration was quantified using assay based in the Bradford method. Equal amount of protein in each sample was mixed with the sample buffer (45 mM Tris pH 6.8, 10% glycerol, 1% SDS, 52 mM DTT and 1% bromophenol blue) and heated at 99°C for 5 min. Proteins were separated by standard SDS-PAGE technique and transferred to PVDF membranes. In order to avoid unspecific binding of antibodies membranes were incubated with TBS buffer containing 0.1% of tween and 5% of milk for 1h at RT. Primary antibodies were incubated for 1h at RT or O/N at 4°C (see Table 8). After washing with TBS-T 0.1% membranes were incubated for 1h at RT with HRP-conjugated secondary antibodies diluted 1/1000. Immobilized antigen-antibody complexes were incubated with ECL substrates for 1 min and visualized trough exposure on X-ray films. Densitometry of bands was calculated using ImageJ software.

mRNA extraction quantitative real-time PCR

Total RNA was isolated using commercially available kit following manufacturer's instructions and its quantity and quality was checked by NanoDrop spectrophotometer. cDNA was obtained by reverse-transcription reaction. Quantitative PCR was performed using TaqMan gene expression assay in which designed probes for gene of interest are run in parallel to *B2M* probe serving as a control (see Table 7 and Table 9).The gene expression results were normalized using comparative CT method.

Cell proliferation assay

Cell proliferation was assessed using two methods. One based on cell courting and other based on cell metabolic activity that reflect on number of viable cells (MTT assay). $1x10^5$ cells were seeded in triplicated in 6 well plate. At day 1, 3, 5, and 7 cells were harvested and counted in automatic cell counter. MTT assay was performed in 96 well plate following manufacturer's instructions. $1x10^3$ cells per well were seeded in triplicates. Assay was performed at day 1, 3 and 6 by incubation of cells with 10 µl of MTT reagent for 4 hours followed by solubilizing of precipitates overnight. The absorbance of solution was quantified by a spectrophotometer at wavelength range 595 nm to 570 nm.

CRISPR gene editing

The CRISPR/Cas 9 based gene editing methods have been previously described (Cong et al. 2013; Ran et al. 2013) and used with small modifications. The backbone plasmid vector pX330 was obtained from Addgene. All gRNAs sequences were designed using ChopChop software in default setting (https://chopchop.rc.fas.harvard.edu/index.php) (Montague et al. 2014). Briefly, top 5

target sites in *RPS6KA5* (NCBI sequence ID# NM_004755) gene ranked by software were chosen and none of them had reported off-target sites (see Table 6 for details). Ligation adapter sequences were added to 5' end of 20 nucleotide gRNA sequences for each target site. For sgRNA assembly, a pair of synthesized oligos was annealed, phosphorylated, and ligated to *BbsI* linearized vector. Constructs have been sequenced using hU6 promoter primer. Liposome-based chemical transfection method was used for all plasmids and cell lines. $2 \cdot 10^5$ cells were transfected using 2.5µg of plasmid diluted in reagent and booster (ratio 2:3) and serum free medium up to final volume of 100 µl. 24h after transfection, medium was changed. Cells were harvested for sorting 48h post transfection and single cells were sorted into 96 well plates. MSK1-pX330 transfected cells were sorted based on GFP expression in comparison to non-transfected cells. Clones derived from single cell were expanded and gDNA, RNA and protein extracts were isolated from each clone. MSK1 knockout clones were was identified by mRNA expression and protein level. Knockout clones were pooled together and the same procedure was applied to control clones (non-edited or expressing MSK1) (see Table 4).

Hypoxia assay

In hypoxia assay cells were cultured as described in paragraph Cell culture, but placed in a hypoxic chamber instead of a standard incubator. The gas in the chamber was composed of 94.5% N_2, 5.0% CO_2 and 0.5% O_2. In the assay $5x10^4$ cells were seeded for 24, 48 or 72 hours, harvested and stained with Annexin V APC kit and PI. Percentage of living, early-apoptotic, late apoptotic and dead cells was determined by FACS analysis using APC and PI channel.

Cell adhesion

$1x10^5$ cells were seeded in triplicates on fibronectin, collagen (10mg/ml) or matrigel coated 24 well plates. After 2 hours later unattached cells were washed and attached were fixed with formalin for 10 min. Cells were visualized by crystal violet staining for 15 minutes and the dye was dissolved in 2% SDS. Absorbance was measured at 570nm.

Cell migration and invasion assay

Cell invasion was assayed using human fibronectin treated 3 μm pore transwell inserts and cell invasion assay was done using Martigel coated 8 μm pore transwell inserts. 24hours prior to experiment cells were starved by culture in DMEM without phenol red supplemented with 50 U/ml penicillin, 50 μg/ml streptomycin and 5% glutamine (serum-free medium). After starvation cells were labeled with CellTracker green fluorescent dye for 30 minutes. $5x10^4$ cells were suspended in a serum-free medium and seeded in duplicates on a top of transwell insert. Full medium (containing 10% serum) was used as a chemoattractant in a lower compartment and assay was performed for 24 hours at 37°C. Next, cells were fixed with 10% buffered formalin and top membrane of the insert was swab in order to remove cells. Migrated or invaded cells were counted from the bottom part of a membrane and from the bottom the lower compartment.

Oncosphere formation assay

To assess tumor initiation capacity *in vitro*, cells were counted and plated into low-attachment 96-well plates at dilution of 1 cell per well for 15 days. For i*n vitro* culture mammary epithelial basal medium (MEBM), supplemented with MEGM Single- Quots (which

contain Insulin, EGF, Hydrocortisone and GA-1000), 1X B27 without retinoic acid and 20 ng/ml of recombinant fibroblast growth factor was used. Oncospheres were counted under the microscope.

Statistical analysis

All statistical analyses were performed using GraphPad Prism 6 software. Fisher exact test was used for binomial variables. For continuous variables Student t-test was used if data were distributed normally or Mann-Whitney test for non-Gaussian population. Kaplan–Meier estimates and log-rank test were used for metastasis-free survival data plotting and comparison. Unless indicated otherwise, two-sided and unpaired tests were used for data analysis. P values were calculated and $P < 0.05$ was considered to be statistically significant. For simplicity reason P values are presented on figures as ns, not significant $P>0.05$; *, $P\leq0.05$; **, $P\leq0.01$; ***, $P\leq0.001$; ****, $P\leq0.0001$.

2. Materials

Table 5. List of plasmids

name	type	description	selection	source
TGL	retroviral	luciferase, GFP	GFP	Lab resources
pLKO	lentiviral	shRNA	puromycin	Sigma-Aldrich
pBABE puro GFP	retroviral	empty	puromycin	A.T.J Wierenga UMCG
pBABE MSK1 GFP	retroviral	MSK1 OE	puromycin	A.T.J Wierenga UMCG
pX330-EGFP	transient	CRISPR/Cas9 backbone	EGFP	E. Battle lab

Table 6. List of shRNAs and gRNAs

name	target	sequence 5' - 3'; PAM	ID
MSK1 shRNA1	*RSP6KA5*	CCGGGCTGAGAAGGTGGGAATAGAACTC-GAGTTCTATTCCCACCTTCTCAGCTTTTT	TRCN0000001497
MSK1 shRNA2	*RSP6KA5*	CCGGCGCGGTGGAAATCATGAAGAACTC-GAGTTCTTCATGATTTCCACCGCGTTTTT	TRCN0000001498
MSK2 shRNA1	*RSP6KA4*	CCGGCCGAAATCATCCGTAGCAAGACTC-GAGTCTTGCTACGGATGATTTCGGTTTTT	TRCN0000021515
MSK2 shRNA2	*RSP6KA4*	CCGGCGAAATCATCCGTAGCAAGACCTC-GAGGTCTTGCTACGGATGATTTCGTTTTT	TRCN0000021518
sh ctrl	N/A	N/A	SHC016
MSK1 gRNA1	*RSP6KA5*	TGGTCTGAGTAAGGAGTTTGTGG	N/A
MSK1 gRNA2	*RSP6KA5*	GGCACCAGATATTGTCAGAGGGG	N/A
MSK1 gRNA3	*RSP6KA5*	GCTAAAGCACTCATTTCTTGGGG	N/A
MSk1 gRNA4	*RSP6KA5*	TGGTGCAGGCACTTTTTTGGCGG	N/A
MSK1 gRNA5	*RSP6KA5*	GTGATGGAACTTCTGAATGGAGG	N/A

Abbreviations: N/A, non applicable

Table 7. List of primers

target	ID
B2M	Mm00437762_m1
RPS6KA5	HS01046596_m1
Cfos	
EGR1	HS00152928_m1

COX2	HS00153133_m1
KRT7	HS00559840_m1
KRT14	HS00265033_m1
KRT18	HS02827483_g1
GATA2	HS00231119_m1
GATA3	HS00231122_m1
FOXA1	HS04187555_m1

Table 8. List of antibodies

antigen	dillution	source	ID	vendor	other details
Western blotting					
MSK1*	1:500	rabbit	3489	Cell Signalling	5% BSA/TBST
MSK2	1:500	rabbit	42101	Abcam	5% BSA/TBST
tubulin	1:5000	mouse		Sigma Aldrich	1% BSA/TBST
anti-mouse	1:3000	rabbit	31452	Sigma Aldrich	5% BSA/TBST, HRP conjugated
anti-rabbit	1:3000	donkey	NA934V	Amersham	5% BSA/TBST, HRP conjugated
Immunohistochemistry					
BrdU	1:100	mouse	347580	BD	antigen retrieval TRIS-EDTA pH=9
Ki67	1:100	rabbit			antigen retrieval citrate pH=6
ERα	1:100	rabbit	16660	Abcam	antigen retrieval TRIS-EDTA pH=9
GFP	1:100	rabbit	11122	Invitrogen	antigen retrieval TRIS-EDTA pH=9
FOXA1	1:100	mouse	05-1466	Millipore	antigen retrieval TRIS-EDTA pH=9
anti-mouse IF	1:500	goat goat	O11033 A11004	Molecular Probes	alexa fluor 488 or 568
anti-rabbit IF	1:500	goat goat	A11008 A11011	Molecular Probes	alexa fluor 488 or 568
anti mouse IHC	ready to use	goat	DPVMHRP	Dako	HRP conjugated
anti-rabbit IHC	ready to use	goat	DPVRHRP	Dako	HRP conjugated
FACS					
CD24_PerCP-Cy5.5	1:100		561647	BD Pharmigen	PBS/0.2% BSA

CD44_PE	1:50	555479	BD Pharmigen	PBS/0.2% BSA
CD49f_PE7Cy7	1:100	313621	Biolegend	PBS/0.2% BSA

* this antibody served for Western blotting and immunohistochemical analyses

Table 9. List of comercial kits

application	name	vendor
RNA extraction	PureLink RNA Mini kit	Ambion
gDNA extraction	GeneElute mammalian gemonic DNA miniprep kit	Sigma Aldrich
plasmid isolation	PureLink HiPure Maxiprep kit	Invitrogen
PCR	KOD Hot Start DNA Polymerase	Novagen
RT-PCR	High-Capacity cDNA Reverse Transcription kit	Applied Biosystems
qPCR	TaqMan Universal Master Mix	Applied Biosystems
apoptosis detection	Annexin V Apoptosis Detection APC kit	eBioscience
EdU detection	Click-iTPlus EdU for FACS kit	Molecular Probes
proliferation	Cell Proliferation Kit MTT	Roche
transfection	NanoJuice Transfecion kit	Millipore

BIBLIOGRAPHY

Abd El-Rehim, D. M., S. E. Pinder, et al. (2004). "Expression of luminal and basal cytokeratins in human breast carcinoma." J Pathol 203(2): 661-671.

Aguirre-Ghiso, J. A. (2007). "Models, mechanisms and clinical evidence for cancer dormancy." Nat Rev Cancer 7(11): 834-846.

Aguirre-Ghiso, J. A., Y. Estrada, et al. (2003). "ERK(MAPK) activity as a determinant of tumor growth and dormancy; regulation by p38(SAPK)." Cancer Res 63(7): 1684-1695.

Aguirre-Ghiso, J. A., D. Liu, et al. (2001). "Urokinase receptor and fibronectin regulate the ERK(MAPK) to p38(MAPK) activity ratios that determine carcinoma cell proliferation or dormancy in vivo." Mol Biol Cell 12(4): 863-879.

Al-Hajj, M., M. S. Wicha, et al. (2003). "Prospective identification of tumorigenic breast cancer cells." Proc Natl Acad Sci U S A 100(7): 3983-3988.

Anborgh, P. H., J. C. Mutrie, et al. (2010). "Role of the metastasis-promoting protein osteopontin in the tumour microenvironment." J Cell Mol Med 14(8): 2037-2044.

Anjum, R. and J. Blenis (2008). "The RSK family of kinases: emerging roles in cellular signalling." Nat Rev Mol Cell Biol 9(10): 747-758.

Balic, M., H. Lin, et al. (2006). "Most early disseminated cancer cells detected in bone marrow of breast cancer patients have a putative breast cancer stem cell phenotype." Clin Cancer Res 12(19): 5615-5621.

Barrett, T., D. B. Troup, et al. (2007). "NCBI GEO: mining tens of millions of expression profiles--database and tools update." Nucleic Acids Res 35(Database issue): D760-765.

Beavon, I. R. (2000). "The E-cadherin-catenin complex in tumour metastasis: structure, function and regulation." Eur J Cancer 36(13 Spec No): 1607-1620.

Bernardo, G. M., G. Bebek, et al. (2013). "FOXA1 represses the molecular phenotype of basal breast cancer cells." Oncogene 32(5): 554-563.

Bidwell, B. N., C. Y. Slaney, et al. (2012). "Silencing of Irf7 pathways in breast cancer cells promotes bone metastasis through immune escape." Nat Med 18(8): 1224-1231.

Boyer, L. A., K. Plath, et al. (2006). "Polycomb complexes repress developmental regulators in murine embryonic stem cells." Nature 441(7091): 349-353.

Brabletz, T. (2012). "To differentiate or not--routes towards metastasis." Nat Rev Cancer 12(6): 425-436.

Bragado, P., Y. Estrada, et al. (2013). "TGF-beta2 dictates disseminated tumour cell fate in target organs through TGF-beta-RIII and p38alpha/beta signalling." Nat Cell Biol 15(11): 1351-1361.

Bulavin, D. V., C. Phillips, et al. (2004). "Inactivation of the Wip1 phosphatase inhibits mammary tumorigenesis through p38 MAPK-mediated activation of the p16(Ink4a)-p19(Arf) pathway." Nat Genet 36(4): 343-350.

Cameron, M. D., E. E. Schmidt, et al. (2000). "Temporal progression of metastasis in lung: cell survival, dormancy, and location dependence of metastatic inefficiency." Cancer Res 60(9): 2541-2546.

Cancer Genome Atlas, N. (2012). "Comprehensive molecular portraits of human breast tumours." Nature 490(7418): 61-70.

Castro, M. A., I. de Santiago, et al. (2016). "Regulators of genetic risk of breast cancer identified by integrative network analysis." Nat Genet 48(1): 12-21.

Clark, C. J., D. M. McDade, et al. (2007). "Contrasting roles of neuronal Msk1 and Rsk2 in Bad phosphorylation and feedback regulation of Erk signalling." J Neurochem 102(4): 1024-1034.

Cong, L., F. A. Ran, et al. (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science 339(6121): 819-823.

Cox, T. R., R. M. Rumney, et al. (2015). "The hypoxic cancer secretome induces pre-metastatic bone lesions through lysyl oxidase." Nature 522(7554): 106-110.

Crosio, C., G. M. Fimia, et al. (2002). "Mitotic phosphorylation of histone H3: spatio-temporal regulation by mammalian Aurora kinases." Mol Cell Biol 22(3): 874-885.

Chambers, A. F., A. C. Groom, et al. (2002). "Dissemination and growth of cancer cells in metastatic sites." Nat Rev Cancer 2(8): 563-572.

Chen, T., E. Heller, et al. (2012). "An RNA interference screen uncovers a new molecule in stem cell self-renewal and long-term regeneration." Nature 485(7396): 104-108.

De Cesare, D., S. Jacquot, et al. (1998). "Rsk-2 activity is necessary for epidermal growth factor-induced phosphorylation of CREB protein and transcription of c-fos gene." Proc Natl Acad Sci U S A 95(21): 12202-12207.

Dyson, M. H., S. Thomson, et al. (2005). "MAP kinase-mediated phosphorylation of distinct pools of histone H3 at S10 or S28 via mitogen- and stress-activated kinase 1/2." J Cell Sci 118(Pt 10): 2247-2259.

Early Breast Cancer Trialists' Collaborative, G. (2005). "Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials." Lancet 365(9472): 1687-1717.

Eckhardt, B. L., P. A. Francis, et al. (2012). "Strategies for the discovery and development of therapies for metastatic breast cancer." Nat Rev Drug Discov 11(6): 479-497.

Erler, J. T., K. L. Bennewith, et al. (2009). "Hypoxia-induced lysyl oxidase is a critical mediator of bone marrow cell recruitment to form the premetastatic niche." Cancer Cell 15(1): 35-44.

Ewton, D. Z., J. Hu, et al. (2011). "Inactivation of mirk/dyrk1b kinase targets quiescent pancreatic cancer cells." Mol Cancer Ther 10(11): 2104-2114.

Ezkurdia, I., D. Juan, et al. (2014). "Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes." Hum Mol Genet 23(22): 5866-5878.

Feuerer, M., M. Rocha, et al. (2001). "Enrichment of memory T cells and other profound immunological changes in the bone marrow from untreated breast cancer patients." Int J Cancer 92(1): 96-105.

Friedrichs, K., P. Ruiz, et al. (1995). "High expression level of alpha 6 integrin in human breast carcinoma is correlated with reduced survival." Cancer Res 55(4): 901-906.

Gao, H., G. Chakraborty, et al. (2014). "Forward genetic screens in mice uncover mediators and suppressors of metastatic reactivation." Proc Natl Acad Sci U S A 111(46): 16532-16537.

Gao, H., G. Chakraborty, et al. (2012). "The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites." Cell 150(4): 764-779.

Gehani, S. S., S. Agrawal-Singh, et al. (2010). "Polycomb group protein displacement and gene activation through MSK-dependent H3K27me3S28 phosphorylation." Mol Cell 39(6): 886-900.

Ghajar, C. M., H. Peinado, et al. (2013). "The perivascular niche regulates breast tumour dormancy." Nat Cell Biol 15(7): 807-817.

Giampieri, S., C. Manning, et al. (2009). "Localized and reversible TGFbeta signalling switches breast cancer cells from cohesive to single cell motility." Nat Cell Biol 11(11): 1287-1296.

Giancotti, F. G. (2013). "Mechanisms governing metastatic dormancy and reactivation." Cell 155(4): 750-764.

Gong, C., K. Fujino, et al. (2015). "FOXA1 repression is associated with loss of BRCA1 and increased promoter methylation and chromatin silencing in breast cancer." Oncogene 34(39): 5012-5024.

Goss, P. E. and A. F. Chambers (2010). "Does tumour dormancy offer a therapeutic target?" Nat Rev Cancer 10(12): 871-877.

Goto, H., Y. Yasui, et al. (2002). "Aurora-B phosphorylates Histone H3 at serine28 with regard to the mitotic chromosome condensation." Genes Cells 7(1): 11-17.

Gutierrez, M. C., S. Detre, et al. (2005). "Molecular changes in tamoxifen-resistant breast cancer: relationship between estrogen receptor, HER-2, and p38 mitogen-activated protein kinase." J Clin Oncol 23(11): 2469-2476.

Ha, S. A., Y. S. Lee, et al. (2011). "The prognostic potential of keratin 18 in breast cancer associated with tumor dedifferentiation, and the loss of estrogen and progesterone receptors." Cancer Biomark 10(5): 219-231.

Hess, K. R., L. Pusztai, et al. (2003). "Estrogen receptors and distinct patterns of breast cancer relapse." Breast Cancer Res Treat 78(1): 105-118.

Howlader N, N. A., Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (2016). "SEER Cancer Statistics Review, 1975-2013, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2013/, based on November 2015 SEER data submission." posted to the SEER web site.

Hsu, P. D., D. A. Scott, et al. (2013). "DNA targeting specificity of RNA-guided Cas9 nucleases." Nat Biotechnol 31(9): 827-832.

Hu, J., H. Deng, et al. (2013). "Ovarian cancer cells, not normal cells, are damaged by Mirk/Dyrk1B kinase inhibition." Int J Cancer 132(10): 2258-2269.

Jackson, A. L., J. Burchard, et al. (2006). "Position-specific chemical modification of siRNAs reduces "off-target" transcript silencing." RNA 12(7): 1197-1205.

Jenssen, T. K., W. P. Kuo, et al. (2002). "Associations between gene expressions in breast cancer and patient survival." Hum Genet 111(4-5): 411-420.

Jones, C., E. Ford, et al. (2004). "Molecular cytogenetic identification of subgroups of grade III invasive ductal breast carcinomas with different clinical outcomes." Clin Cancer Res 10(18 Pt 1): 5988-5997.

Joyce, J. A., A. Baruch, et al. (2004). "Cathepsin cysteine proteases are effectors of invasive growth and angiogenesis during multistage tumorigenesis." Cancer Cell 5(5): 443-453.

Joyce, J. A. and J. W. Pollard (2009). "Microenvironmental regulation of metastasis." Nat Rev Cancer 9(4): 239-252.

Kang, Y., P. M. Siegel, et al. (2003). "A multigenic program mediating breast cancer metastasis to bone." Cancer Cell 3(6): 537-549.

Kelly, T., L. J. Suva, et al. (2005). "Expression of heparanase by primary breast tumors promotes bone resorption in the absence of detectable bone metastases." Cancer Res 65(13): 5778-5784.

Kennecke, H., R. Yerushalmi, et al. (2010). "Metastatic behavior of breast cancer subtypes." J Clin Oncol 28(20): 3271-3277.

Kessenbrock, K., V. Plaks, et al. (2010). "Matrix metalloproteinases: regulators of the tumor microenvironment." Cell 141(1): 52-67.

Klein, C. A. (2011). "Framework models of tumor dormancy from patient-derived observations." Curr Opin Genet Dev 21(1): 42-49.

Kobayashi, A., H. Okuda, et al. (2011). "Bone morphogenetic protein 7 in dormancy and metastasis of prostate cancer stem-like cells in bone." J Exp Med 208(13): 2641-2655.

Kouros-Mehr, H., S. K. Bechis, et al. (2008). "GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model." Cancer Cell 13(2): 141-152.

Kozlow, W. and T. A. Guise (2005). "Breast cancer metastasis to bone: mechanisms of osteolysis and implications for therapy." J Mammary Gland Biol Neoplasia 10(2): 169-180.

Labelle, M. and R. O. Hynes (2012). "The initial hours of metastasis: the importance of cooperative host-tumor cell interactions during hematogenous dissemination." Cancer Discov 2(12): 1091-1099.

Lau, P. N. and P. Cheung (2011). "Histone code pathway involving H3 S28 phosphorylation and K27 acetylation activates transcription and antagonizes polycomb silencing." Proc Natl Acad Sci U S A 108(7): 2801-2806.

Lawler, J. (2002). "Thrombospondin-1 as an endogenous inhibitor of angiogenesis and tumor growth." J Cell Mol Med 6(1): 1-12.

Lee, T. I., R. G. Jenner, et al. (2006). "Control of developmental regulators by Polycomb in human embryonic stem cells." Cell 125(2): 301-313.

Liu, H., M. R. Patel, et al. (2010). "Cancer stem cells from human breast tumors are involved in spontaneous metastases in orthotopic mouse models." Proc Natl Acad Sci U S A 107(42): 18115-18120.

Liu, R., X. Wang, et al. (2007). "The prognostic role of a gene signature from tumorigenic breast-cancer cells." N Engl J Med 356(3): 217-226.

Louie, D. F., K. K. Gloor, et al. (2000). "Phosphorylation and subcellular redistribution of high mobility group proteins 14 and 17, analyzed by mass spectrometry." Protein Sci 9(1): 170-179.

Lu, X., E. Mu, et al. (2011). "VCAM-1 promotes osteolytic expansion of indolent bone micrometastasis of breast cancer by engaging alpha4beta1-positive osteoclast progenitors." Cancer Cell 20(6): 701-714.

Luo, M., S. G. Clouthier, et al. (2015). "Breast cancer stem cells: current advances and clinical implications." Methods Mol Biol 1293: 1-49.

Luzzi, K. J., I. C. MacDonald, et al. (1998). "Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases." Am J Pathol 153(3): 865-873.

Malanchi, I., A. Santamaria-Martinez, et al. (2012). "Interactions between cancer stem cells and their niche govern metastatic colonization." Nature 481(7379): 85-89.

Malladi, S., D. G. Macalinao, et al. (2016). "Metastatic Latency and Immune Evasion through Autocrine Inhibition of WNT." Cell 165(1): 45-60.

Massague, J. and A. C. Obenauf (2016). "Metastatic colonization by circulating tumour cells." Nature 529(7586): 298-306.

Mehra, R., S. Varambally, et al. (2005). "Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis." Cancer Res 65(24): 11259-11264.

Mehta, R. J., R. K. Jain, et al. (2012). "FOXA1 is an independent prognostic marker for ER-positive breast cancer." Breast Cancer Res Treat 131(3): 881-890.

Meyer, M. J., J. M. Fleming, et al. (2010). "CD44posCD49fhiCD133/2hi defines xenograft-initiating cells in estrogen receptor-negative breast cancer." Cancer Res 70(11): 4624-4633.

Minn, A. J., G. P. Gupta, et al. (2005). "Genes that mediate breast cancer metastasis to lung." Nature 436(7050): 518-524.

Mitra, A., L. Mishra, et al. (2015). "EMT, CTCs and CSCs in tumor relapse and drug-resistance." Oncotarget 6(13): 10697-10711.

Moffat, J., D. A. Grueneberg, et al. (2006). "A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen." Cell 124(6): 1283-1298.

Montague, T. G., J. M. Cruz, et al. (2014). "CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing." Nucleic Acids Res 42(Web Server issue): W401-407.

Mueller, H., N. Flury, et al. (2000). "Potential prognostic value of mitogen-activated protein kinase activity for disease-free survival of primary breast cancer patients." Int J Cancer 89(4): 384-388.

Murugaesu, N., M. Iravani, et al. (2014). "An in vivo functional screen identifies ST6GalNAc2 sialyltransferase as a breast cancer metastasis suppressor." Cancer Discov 4(3): 304-317.

Naqvi, S., A. Macdonald, et al. (2012). "Characterization of the cellular action of the MSK inhibitor SB-747651A." Biochem J 441(1): 347-357.

Nesbitt, J. C., J. B. Putnam, Jr., et al. (1995). "Survival in early-stage non-small cell lung cancer." Ann Thorac Surg 60(2): 466-472.

Nguyen, D. X., P. D. Bos, et al. (2009). "Metastasis: from dissemination to organ-specific colonization." Nat Rev Cancer 9(4): 274-284.

Nieswandt, B., M. Hafner, et al. (1999). "Lysis of tumor cells by natural killer cells in mice is impeded by platelets." Cancer Res 59(6): 1295-1300.

Obenauf, A. C. and J. Massague (2015). "Surviving at a distance: organ specific metastasis." Trends Cancer 1(1): 76-91.

Osisami, M. and E. T. Keller (2013). "Mechanisms of Metastatic Tumor Dormancy." J Clin Med 2(3): 136-150.

Oskarsson, T., S. Acharyya, et al. (2011). "Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs." Nat Med 17(7): 867-874.

Padua, D., X. H. Zhang, et al. (2008). "TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4." Cell 133(1): 66-77.

Pan, Q., L. J. van der Laan, et al. (2012). "A dynamic perspective of RNAi library development." Trends Biotechnol 30(4): 206-215.

Pavlovic, M., A. Arnal-Estape, et al. (2015). "Enhanced MAF Oncogene Expression and Breast Cancer Bone Metastasis." J Natl Cancer Inst 107(12).

Pearce, L. R., D. Komander, et al. (2010). "The nuts and bolts of AGC protein kinases." Nat Rev Mol Cell Biol 11(1): 9-22.

Peinado, H., M. Aleckovic, et al. (2012). "Melanoma exosomes educate bone marrow progenitor cells toward a pro-metastatic phenotype through MET." Nat Med 18(6): 883-891.

Perou, C. M., T. Sorlie, et al. (2000). "Molecular portraits of human breast tumours." Nature 406(6797): 747-752.

Possemato, R., K. M. Marks, et al. (2011). "Functional genomics reveal that the serine synthesis pathway is essential in breast cancer." Nature 476(7360): 346-350.

Prat, A., J. S. Parker, et al. (2010). "Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer." Breast Cancer Res 12(5): R68.

Qian, B. Z. and J. W. Pollard (2010). "Macrophage diversity enhances tumor progression and metastasis." Cell 141(1): 39-51.

Ran, F. A., P. D. Hsu, et al. (2013). "Genome engineering using the CRISPR-Cas9 system." Nat Protoc 8(11): 2281-2308.

Redig, A. J. and S. S. McAllister (2013). "Breast cancer as a systemic disease: a view of metastasis." J Intern Med 274(2): 113-126.

Romero, I., F. Garrido, et al. (2014). "Metastases in immune-mediated dormancy: a new opportunity for targeting cancer." Cancer Res 74(23): 6750-6757.

Root, D. E., N. Hacohen, et al. (2006). "Genome-scale loss-of-function screening with a lentiviral RNAi library." Nat Methods 3(9): 715-719.

Ross-Innes, C. S., R. Stark, et al. (2012). "Differential oestrogen receptor binding is associated with clinical outcome in breast cancer." Nature 481(7381): 389-393.

Sarkar, A. and K. Hochedlinger (2013). "The sox family of transcription factors: versatile regulators of stem and progenitor cell fate." Cell Stem Cell 12(1): 15-30.

Sawicka, A., D. Hartl, et al. (2014). "H3S28 phosphorylation is a hallmark of the transcriptional response to cellular stress." Genome Res 24(11): 1808-1820.

Sawicka, A. and C. Seiser (2012). "Histone H3 phosphorylation - a versatile chromatin modification for different occasions." Biochimie 94(11): 2193-2201.

Sekiya, T., U. M. Muthurajan, et al. (2009). "Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA." Genes Dev 23(7): 804-809.

Sethi, N., X. Dai, et al. (2011). "Tumor-derived JAGGED1 promotes osteolytic bone metastasis of breast cancer by engaging notch signaling in bone cells." Cancer Cell 19(2): 192-205.

Shankaran, V., H. Ikeda, et al. (2001). "IFNgamma and lymphocytes prevent primary tumour development and shape tumour immunogenicity." Nature 410(6832): 1107-1111.

Sheridan, C., H. Kishimoto, et al. (2006). "CD44+/CD24- breast cancer cells exhibit enhanced invasive properties: an early step necessary for metastasis." Breast Cancer Res 8(5): R59.

Shimamura, A., B. A. Ballif, et al. (2000). "Rsk1 mediates a MEK-MAP kinase cell survival signal." Curr Biol 10(3): 127-135.

Shiozawa, Y., E. A. Pedersen, et al. (2010). "GAS6/AXL axis regulates prostate cancer invasion, proliferation, and survival in the bone marrow niche." Neoplasia 12(2): 116-127.

Simon, J. A. and R. E. Kingston (2009). "Mechanisms of polycomb gene silencing: knowns and unknowns." Nat Rev Mol Cell Biol 10(10): 697-708.

Sodersten, E., M. Feyder, et al. (2014). "Dopamine signaling leads to loss of Polycomb repression and aberrant gene activation in experimental parkinsonism." PLoS Genet 10(9): e1004574.

Soloaga, A., S. Thomson, et al. (2003). "MSK2 and MSK1 mediate the mitogen- and stress-induced phosphorylation of histone H3 and HMG-14." EMBO J 22(11): 2788-2797.

Stephens, J. K., G. T. Everson, et al. (2000). "Fatal transfer of malignant melanoma from multiorgan donor to four allograft recipients." Transplantation 70(1): 232-236.

Straume, O., T. Shimamura, et al. (2012). "Suppression of heat shock protein 27 induces long-term dormancy in human breast cancer." Proc Natl Acad Sci U S A 109(22): 8699-8704.

Strauss, D. C. and J. M. Thomas (2010). "Transmission of donor melanoma by organ transplantation." Lancet Oncol 11(8): 790-796.

Subramanian, A., P. Tamayo, et al. (2005). "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." Proc Natl Acad Sci U S A 102(43): 15545-15550.

Urosevic, J., X. Garcia-Albeniz, et al. (2014). "Colon cancer cells colonize the lung from established liver metastases through p38 MAPK signalling and PTHLH." Nat Cell Biol 16(7): 685-694.

Valastyan, S. and R. A. Weinberg (2011). "Tumor metastasis: molecular insights and evolving paradigms." Cell 147(2): 275-292.

Venkatraman, E. S. and A. B. Olshen (2007). "A faster circular binary segmentation algorithm for the analysis of array CGH data." Bioinformatics 23(6): 657-663.

Vermeulen, L., G. De Wilde, et al. (2003). "Transcriptional activation of the NF-kappaB p65 subunit by mitogen- and stress-activated protein kinase-1 (MSK1)." EMBO J 22(6): 1313-1324.

Vermeulen, L., W. Vanden Berghe, et al. (2009). "The versatile role of MSKs in transcriptional regulation." Trends Biochem Sci 34(6): 311-318.

Vicent, G. P., C. Ballare, et al. (2006). "Chromatin remodeling and control of cell proliferation by progestins via cross talk of progesterone receptor with the estrogen receptors and kinase signaling pathways." Ann N Y Acad Sci 1089: 59-72.

Visvader, J. E. and G. J. Lindeman (2012). "Cancer stem cells: current status and evolving complexities." Cell Stem Cell 10(6): 717-728.

Wang, H., C. Yu, et al. (2015). "The osteogenic niche promotes early-stage bone colonization of disseminated breast cancer cells." Cancer Cell 27(2): 193-210.

Weilbaecher, K. N., T. A. Guise, et al. (2011). "Cancer to bone: a fatal attraction." Nat Rev Cancer 11(6): 411-425.

Wells, A., L. Griffith, et al. (2013). "The dormancy dilemma: quiescence versus balanced proliferation." Cancer Res 73(13): 3811-3816.

Wiggin, G. R., A. Soloaga, et al. (2002). "MSK1 and MSK2 are required for the mitogen- and stress-induced phosphorylation of CREB and ATF1 in fibroblasts." Mol Cell Biol 22(8): 2871-2881.

Willis, L., T. Alarcon, et al. (2010). "Breast cancer dormancy can be maintained by small numbers of micrometastases." Cancer Res 70(11): 4310-4317.

Wyckoff, J. B., Y. Wang, et al. (2007). "Direct visualization of macrophage-assisted tumor cell intravasation in mammary tumors." Cancer Res 67(6): 2649-2656.

Yang, J. Y., C. S. Zong, et al. (2008). "ERK promotes tumorigenesis by inhibiting FOXO3a via MDM2-mediated degradation." Nat Cell Biol 10(2): 138-148.

Yi, B., P. J. Williams, et al. (2002). "Tumor-derived platelet-derived growth factor-BB plays a critical role in osteosclerotic bone metastasis in an animal model of human breast cancer." Cancer Res 62(3): 917-923.

Yin, J. J., K. S. Mohammad, et al. (2003). "A causal role for endothelin-1 in the pathogenesis of osteoblastic bone metastases." Proc Natl Acad Sci U S A 100(19): 10954-10959.

Yin, J. J., K. Selander, et al. (1999). "TGF-beta signaling blockade inhibits PTHrP secretion by breast cancer cells and bone metastases development." J Clin Invest 103(2): 197-206.

Yoneda, T. (2000). "Cellular and molecular basis of preferential metastasis of breast cancer to bone." J Orthop Sci 5(1): 75-81.

Yung, P. Y., A. Stuetzer, et al. (2015). "Histone H3 Serine 28 Is Essential for Efficient Polycomb-Mediated Gene Repression in Drosophila." Cell Rep 11(9): 1437-1445.

Zender, L., W. Xue, et al. (2008). "An oncogenomics-based in vivo RNAi screen identifies tumor suppressors in liver cancer." Cell 135(5): 852-864.

Zhang, X. H., M. Giuliano, et al. (2013). "Metastasis dormancy in estrogen receptor-positive breast cancer." Clin Cancer Res 19(23): 6389-6397.

Zhang, X. H., Q. Wang, et al. (2009). "Latent bone metastasis in breast cancer tied to Src-dependent survival signals." Cancer Cell 16(1): 67-78.

Zhang, Y., G. Liu, et al. (2001). "MSK1 and JNKs mediate phosphorylation of STAT3 in UVA-irradiated mouse epidermal JB6 cells." J Biol Chem 276(45): 42534-42542.

Zheng, D., Y. Y. Cho, et al. (2008). "Cyclin-dependent kinase 3-mediated activating transcription factor 1 phosphorylation enhances cell transformation." Cancer Res 68(18): 7650-7660.

Zhong, H., H. SuYang, et al. (1997). "The transcriptional activity of NF-kappaB is regulated by the IkappaB-associated PKAc subunit through a cyclic AMP-independent mechanism." Cell 89(3): 413-424.