Altered Expression of Secreted Mediator Genes That Mediate Aggressive Breast Cancer Metastasis to Distant Organs

Apama Maiti  
Roswell Park Cancer Institute

Ichiro Okano  
Roswell Park Cancer Institute

Masanori Oshi  
Roswell Park Cancer Institute

Maiko Okano  
Roswell Park Cancer Institute

Wanqing Tian  
Roswell Park Cancer Institute

Tsutomu Kawaguchi  
Roswell Park Cancer Institute

Eriko Katsuta  
Roswell Park Cancer Institute

Kazuaki Takabe  
Roswell Park Cancer Institute

Yan Li  
Roswell Park Cancer Institute

Santosh Patnaik  
Roswell Park Cancer Institute

Nitai Hait  (nitai.hait@roswellpark.org)  
Roswell Park Cancer Institute

Research Article

Keywords: Breast cancer, metastases, mediator genes, aggressive subtypes, patients survival

Posted Date: December 7th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-109330/v1
Altered expression of secreted mediator genes that mediate aggressive breast cancer metastasis to distant organs

Aparna Maiti\textsuperscript{1,2}, Ichiro Okano\textsuperscript{1}, Masanori Oshi\textsuperscript{1}, Maiko Okano\textsuperscript{1}, Wanqing Tian\textsuperscript{3}, Tsutomu Kawaguchi\textsuperscript{1}, Eriko Katsuta\textsuperscript{1}, Kazuaki Takabe\textsuperscript{1}, Yan Li\textsuperscript{3}, Santosh Patnaik\textsuperscript{4}, Nitai C. Hait\textsuperscript{1,2}

\textsuperscript{1}Division of Breast Surgery and Department of Surgical Oncology, \textsuperscript{2}Department of Molecular & Cellular Biology, \textsuperscript{3}Department of Biostatistics & Bioinformatics, \textsuperscript{4}Department of Thoracic Surgery, Roswell Park Comprehensive Cancer Center, Buffalo, New York, 14263, USA.

Running title
Breast cancer distant metastases genes

Keywords
Breast cancer, metastases, mediator genes, aggressive subtypes, patients survival

Competing Interests
The authors have declared that no competing interests exist

Correspondence: Nitai C. Hait, Ph.D., Department of Surgical Oncology, Department of Molecular & Cellular Biology, Roswell Park Comprehensive Cancer Center, Medical Research Complex (MRC), Elm and Carlton Streets, Buffalo, NY 14263 (USA). Tel: (716) 845-8527; Fax: (716) 845-1668; E-mail: Nitai.Hait@roswellpark.org; or Aparna Maiti, Ph.D., Department of Surgical Oncology, Department of Molecular & Cellular Biology, Roswell Park Comprehensive Cancer Center, Medical Research Complex (MRC), Elm and Carlton Streets, Buffalo, NY 14263 (USA). Tel: (716) 845-3505; E-mail: Aparna.Maiti@roswellpark.org
ABSTRACT Breast cancer most often recurs and metastasizes to the distal organs with their primary tumors surgically excised. Due to the heterogeneous nature of breast cancer, metastasis organotropism has poorly understood. This study assessed the specific cancer-related gene expression changes occurring with metastatic breast cancer recurrence to distant organs compared with non-metastatic breast cancer. We found that secreted mediators encoding genes (ANGPTL7, MMP3, LCN2, S100A8, and ESM1) are upregulated in primary breast tumors of 4T1.2, a highly metastatic variant of mouse triple-negative breast cancer (TNBC) 4T1 cells. Those genes are variably expressed at distant metastatic sites such as the spine, bone, and the lung in the syngeneic implantation/tumor-resection metastasis mouse model. The altered expression score of an individual gene was strongly associated with the survival of breast cancer patients. Notably, LCN2 and S100A8 overexpressed at the distant metastatic site spine (LCN2, 5 fold; S100A8, 6 fold) and bone (LCN2, 5 fold; S100A8, 3 fold) vs. primary tumors. In contrast, the ESM-1 encoding gene is overexpressed in the primary tumors and markedly downregulated at distant metastatic sites such as the spine, bone, and the lung. LCN2, S100A8, and ESM-1 mediators encoding individual gene expression scores were strongly associated with disease-specific survival (DSS) in the METABRIC cohort (hazard ratio [HR]>1, p < 0.0004). The gene expression scores predicted worse clinically aggressive tumors, such as high Nottingham histological grade and advanced cancer staging. High gene expression score of ESM-1 gene was strongly associated with worse overall survival (OS) in the triple-negative breast cancer (TNBC) and hormonal receptor (HR)-positive/Her2-negative subtype in METABRIC cohort, HER2+ subtype in TCGA-BRCA and SCAN-B breast cancer cohorts.

Our data suggested that mediators encoding genes with both prognostic and predictive values may, therefore, be clinically useful for breast cancer spine, bone, and lung metastasis in particularly in more aggressive subtypes such as TNBC or HER2+ breast cancer.
Introduction

Cancer shows organ specificity during metastasis, known as organotropism, an unanswered question in cancer research. Many theories have been proposed to explain the mechanism of metastasis; however, it is still a challenge that is not fully understood. It has been suggested that metastatic dispersion occurs when the primary tumor is very large (1). Another model postulates that metastatic dissemination occurs in the very early stage of disease progression. Based on all these observations and postulation, it is evident that disseminating cancer cells evolve independent of the primary tumor, that tumor clones can be seeded in parallel or independently to distant sites (2, 3). Thus it implies that cancer requires systemic treatment at the very stage for efficient elimination (3, 4). Research on metastatic mechanisms is mostly hindered because the lack of a proper experimental model mimics the complex metastatic process. The recent development of patient-derived models and mega-analysis of circulating tumor cells (CTCs) brought insights into metastasis molecular mechanisms and suggested that CTCs take advantage of distant organ supportive niches to establish metastasis. The immune system excludes most CTCs, and a minority of escaping cancer cells form metastasis lesions (5). Since the mid-1980s, it has been evident that the gene expression portrayal within the tumor cells dictates cancer metastasis (6, 7). A limited number of tumor cells acquire metastatic potential through the aberrant expression of several intertwining genes.

Further, besides the set of genes of a positive regulator of dissemination, the genes for metastasis suppression also need to be inactivated for tumor cells to disseminate (8, 9).

Heterogeneic clonal evolutions, interacting host factors, and consistent genetic adjustment are required to complete metastasis.

Several decades of research have identified several genes are associated with distant organ-specific metastasis (8-11); still, it is not enough to explain the molecular mechanisms of organ-specific metastasis. We compared the gene expression profiles of poorly metastatic mouse breast cancer 4T1 cells with a highly-metastatic variant 4T1.2 cells. 4T1.2 cells were identified from the 4T1 breast cancer bone metastasis by several rounds of breast implantation and resection of primary tumors using immune-competent mouse models (12, 13). We have performed
transcriptomic analyses to compare gene expression profiles between TNBC metastatic variant 4T1.2 and parental 4T1 cells. Our goal was to identify and validate the set of genes of 4T1.2 tumors are associated with distant metastasis in the mouse model.

We use syngeneic implantation/resection distant metastasis models to validate the gene expression data and compare the gene expression at the primary site vs. distant metastatic sites. Using concurrent clinical and molecular information from a large cohort of breast cancer patients, we developed an individual highly overexpressing gene score associated with distant survival metastasis in multiple cohorts.

We hypothesized that altered expression of genes between primary tumors vs. distant sites might reflect breast cancer's overall aggressiveness, and those genes could be potential targets for distant organ metastasis.

Results

**Identifying potential gene signatures in TNBC metastatic variants of breast cancer cells.**

To comprehensively investigate genes playing a role in breast cancer metastatic phenotypes, we have performed RNA-seq analysis of 4T1 and derivative cell line 4T1.2 with varied metastatic abilities to lung, bone, and spine. The cells are grown in vitro, and RNA was isolated to perform RNA-seq analysis to understand how these two cell lines are different in terms of their gene expression, explaining the metastatic nature of 4T1.2 cells. We found a cluster of top 50 genes differentially expressed between 4T1 and 4T1.2. (Fig. 1A, B). As shown in Fig. 1A, several pro-angiogenic and pro-inflammatory secretory genes are highly upregulated in 4T1.2 cells compared to 4T1 parental cells. Interestingly, genes namely Angiopoietin related protein 7 (ANGPTL7) [Log2fold change (10.042)], Serpine2 [Log2fold change (8.99)], Teraspanin11 [Log2fold change (8.9)], Endothelial cell-specific molecule 1 (ESM1) [Log2fold change (8.88)], Cadherin 5 (CDH5) [Log2fold change (8.59)], Matrix metalloproteinase protein 3 (MMP3) [Log2fold change (7.8)], S100A8 [Log2fold change (6.9]), and Lipocalin2 (LCN2) [Log2fold change (6.6)] are the most important genes that are downregulated in 4T1 while upregulated in 4T1.2. However, Claudin4 [Log2fold change (-10.6)], Epithelial splicing regulated protein1 [Log2fold change (-10.4)], Fermitin family homolog1 Met transcriptional regulator [Log2fold change (-10.2)] are downregulated in
4T1.2 compared to 4T1 cell lines (Fig. 1B). This heatmap data analysis gives us an allusion that although 4T1.2 is a metastatic derivative of 4T1, they have differential gene expression levels. We wanted to validate some of our gene expression RNA-seq data of 4T1.2 cells using the metastatic syngeneic mouse tumor model. As expected, 4T1.2-Luc+ cells show metastatic lesions in the lymph node (LN), lung, spine, and hind bones after primary tumors were resected, identified by the IVIS imaging signal on both ventral and dorsal sides (Fig. 1C-D). Further, MRI analysis of those mice (Fig. 1E, representative MRI images, upper and lower panels) showed cancer cell colonization in the spinal cord (Fig. 1E, bottom panels), explaining recurrence and metastatic spread of the 4T1.2-Luc+ syngeneic tumor model.

**ANGPTL7 secreted angiogenesis-related protein-encoding gene alteration in breast cancer distant metastasis.**

Profiling of human tissue can distinguish the gene expression between normal and tumor tissue. However, there could be an interference of therapies on the alteration of gene expressions. Gene signatures derived from treatment-naive mouse metastatic models could predict the impact of gene expression on cancer prognosis. Our goals are to 1) identify genes that are responsible for breast cancer organotrophic distant metastases, 2) validate our metastatic 4T1.2 cell line RNA-seq data using breast cancer metastatic recurrence mouse models; finally, 3) link those altered genes with the breast cancer patients survival using publicly available breast cancer metastasis cohorts. Based on our RNA-seq analysis, ANGPTL7 is one of the most upregulated genes in 4T1.2 cells. Angiopoietin-like (ANGPTL) proteins belong to a family composed of eight members (from ANGPTL1 to ANGPTL8) (14). ANGPTL proteins are secreted proteins showing structural similarity to members of the Angiopoietin (ANG) family, with a coiled-coil domain at the N-terminus and a fibrinogen-like domain at the C-terminus. Several ANGPTL proteins potently modulate angiogenesis (15-18). ANGPTL7 protein was initially found in the human corneal stroma, trabecular meshwork, and sclera. Protein levels of ANGPTL7 are over-expressed in glaucomatous aqueous humor (19-21). ANGPTL7 has been emerging as an important factor in tumor progression and metastasis (21-23).
Therefore, to validate our RNA-seq data of ANGPTL7, a qPCR analysis was performed using primary breast tumors of 4T1 and 4T1.2 compared with the expression level at the metastatic sites, e.g., spine, bone, and lung using 4T1.2 syngeneic recurrence/metastasis mouse model. As shown in Fig. 2A, ANGPTL7 expression was significantly higher in 4T1.2 primary breast tumors than 4T1 primary breast tumors (p=0.0001), suggesting that cells maintained the increased expression of ANGPTL7 when they formed breast tumors in mice. In order to understand that ANGPTL7 contributed to establishing metastatic lesions in distant organs, we have performed qPCR analyses to examine and compare the expression level of ANGPTL7 in the metastatic lesions isolated from the spine, bone, and lung. Our qPCR data showed that ANGPTL7 expression was reduced in metastatic lesions in the spine, bone, and lung. However, reduced expression of ANGPTL7 at the distant sites is significantly higher (Fig. 2A) than the normal spine, bone, and lung endogenous tissue-specific expression (Fig 2B), suggesting truncated expression ANGPTL7 are linked to breast cancer metastasis. Although ANGPTL7 is overexpressed in cancer, as shown in Fig. 2C, its expression levels were significantly reduced in the distant metastasis group of patients compared to patients with primary tumors in the TCGA-BRCA cohort (24). In contrast, the ANGPTL7 expression score for the TCGA-BRCA cohort using a paired comparison of absolute log2 fold change (FC) gene expression values with Welch’s t-test suggested that ANGPTL7 was significantly higher in adjacent normal tissue compared to the tumor (p<0.00001).

We also examined the ANGPL7 expression levels in primary comparing with bone, lymph node (LN), and lung-metastatic tumors of the breast cancer in the GSE110590 cohort (25). As shown in Fig. 2D, expression levels of ANGPTL7 were significantly higher in bone, LN than the primary tumor, with a tendency of reduced expression in lung metastasis. To comprehend the prognostic value of the ANGPTL7 expression using Kaplan–Meier analysis of OS in SCAN-B cohort having data of 3,273 breast cancer patients (26), we chose to divide the patients into ANGPTL7 high- and low-expression groups using the top one-third of the score as the cutoff. Data showed in Fig. 2E (right panel), the hazard ratio (HR) for progression-free survival (PFI) or disease specific survival (DSS) for METABRIC and TCGA-BRCA with the high expression of ANGPTL7 was not significant.
However, the HR for OS with the ANGPTL7 expression was 0.657 (95% CI = 0.529–0.816; p < 0.001) for the SACN-B cohort, suggesting breast cancer patients with a high ANGPTL7 expression score have better survival (Fig. 2E, left panel). ANGPTL7 is able to promote vascularization by inducing angiogenesis and thereby overexpressed in several cancers. Henceforth, more vascularization might help better drug delivery, which might be the reason for better survival.

**MMP3 matrix metalloproteinase secreted endopeptidase protein-encoding gene inversely linked in breast cancer distant metastasis.**

Matrix metalloproteinases (MMPs) belong to a large proenzymes family of secreted proteins. These proteins play a significant role in the degradation and remodeling of the extracellular matrix (ECM). MMPs are found overexpressed in a variety of diseases, including cancers. We found MMP3 and MMP13 were upregulated in metastatic 4T1.2 cell lines in our RNA seq data; we pursue determining if MMP3 expression is elevated at the distant metastasis sites in mouse models and could be a useful prognostic marker for patient survival. One of the earlier studies conducted by Lochter et al. discovered the role of MMP3 in cancer metastasis (27). However, several studies reporting that MMP3 exhibits both tumor-promoting and tumor-inhibiting effects based on its actions on the targeted substrates.

For example, the interaction between MMP3 and connective tissue growth factor results in the release of angiogenesis-promoting factors (28). MMP3-mediated cleavage of other growth factors such as heparin-bound epidermal growth factor and transforming growth factor β promotes cancer cell proliferation and epithelial-mesenchymal transition (EMT). In these cases, MMP3 exhibits its tumor-promoting effects. As per our animal experiment data (Fig. 3A), the expression of MMP3 was higher in primary tumors of 4T1.2 than primary tumors of 4T1 in our metastatic mouse model (p=0.0001). However, MMP3 gene expression levels were reduced in metastatic lesions of the spine, bone, and lung (Fig. 3A). MMP3 expression at the lung's metastatic lesion is higher (Fig. 3A) than the expression of control mouse lung tissues (Fig. B). In contrast, TCGA-BRCA cohort (24) gene expression analysis revealed that MMP3 expression levels were higher in adjacent normal tissues than patients' tumors (data not included). Further analysis using breast cancer metastasis GSE110590 cohort (25) showed that MMP3 was significantly altered in the bone, LN,
and lung metastasis tissues compared to the primary tumor, with a propensity of reduced expression in LN and lung metastasis (Fig. 3C). The prognostic value of the MMP3 was also tested by Kaplan–Meier analysis of OS using the SCAN-B cohort of 3,273 breast cancer patients (26) and validated using the METABRIC cohort of 1904 breast cancer patients (29) or TCGA-BRCA cohort of 1091 patients (24). Fig. 3D (upper panel) demonstrated that the HR values for OS with the MMP3 expression of high and low-group were less than 1 for both the SCAN-B and METABRIC cohorts (p=0.0001) but not for TCGA-BRCA cohort (p=0.88). These data suggest that breast cancer patients with higher MMP3 expression scores are associated with a better prognosis of breast cancer patients' survival. However, higher MMP3 expression scores are not associated with or DSS for both the METABRIC and TCGA-BRCA cohorts (Fig. 3D, lower panel).

Matrix metalloprotease-13 (MMP-13) also belongs to the MMP superfamily and is also called collagenase 3. MMP-13 in mice plays a critical role in musculoskeletal development. Mouse MMP-13 is markedly upregulated in the stroma during tumor growth and the progression of breast cancer.

In humans, MMP-13 is highly overexpressed in patients with rheumatoid arthritis (30), osteoarthritis (31), and lymphangiogenesis of multiple myeloma (32). A similar pattern of elevated MMP13 expression levels was observed in 4T1.2 primary tumors and spine metastasis (data not shown). However, MMP13 expression scores are not linked with the prognosis of breast cancer patients' survival using TCGA-BRA, SCAN-B, or METABRIC cohorts. So, we decided not to include the MMP13 data.

Positive association of Lipocalin-2 secreted glycoprotein-encoding gene and breast cancer distant metastasis

Lipocalin 2 (LCN2) is a secreted glycoprotein belonging to a lipocalin protein family. LCN2 expression levels are elevated in various human diseases (33, 34) and several cancers (35, 36), including breast cancer (37, 38). We hypothesized that LCN2 is elevated in breast cancer metastasis and is linked with breast cancer aggressiveness, which could be a prognostic factor for the patient's survival.
In compliance with our RNA-seq data of cell lines, the LCN2 transcript is elevated significantly (p=0.0001) in the 4T1.2 primary tumor compared to the 4T1 primary tumor in the metastatic mouse model (Fig. 4A). It is noteworthy that Bone and Lung have a significant level of endogenous expression of LCN2 (Fig. 4B). Furthermore, our data indicated that LCN2 is markedly elevated (>4-6 folds vs. primary tumor) in the distant metastatic sites of the spine (p=0.0001) and bone (0.0001) compared to the primary tumors (Fig. 4A) in addition to the endogenous expression level of control spine and bone tissues of mice (Fig. 4B). We found a reduced expression level of LCN2 in the primary tumor of breast cancer patients vs. adjacent normal tissues by analyzing gene expression data of the TACGA-BRCA cohort (data not included). Further, we found that LCN2 expression levels are altered in distant sites compared to primary tumors of breast cancer patients' as revealed from metastatic cohort (GSE110590) data analysis (Fig. 4C). We performed survival analysis on the METABRIC data set and used Hazard regression analysis models to evaluate the association of high LCN2 gene expression score with OS, DSS, and PFI (Fig 4D). In supportive of our hypothesis, higher LCN2 expression was associated with worse DSS (HR 1.36, 95% CI 1.09-1.57, p=0.003), PFI (HR=1.31, (95% CI=1.134-1.65, p=0.001), but not for OS (p=0.43) among breast cancer patients of the METABRIC cohort (Fig. 4D). Comprehensive analysis of higher LCN2 expression and survival prognosis did not show any prognostic correlation for SCAN-B and TCGA cohort data sets (data not included). Of our interest, we explore further to understand if a higher expression of LCN2 is linked to aggressiveness and advance stage of breast cancer using the METABRIC cohort. Interestingly we found that the LCN2 expression score was higher in TNBC and HER2+ subtypes, which are known to be more clinically aggressive than hormone receptor (HR+) and HER2- subtype (Fig. 4E, p=0.0001). Similar data were also obtained in the SCAN-B cohort (Fig. 4F, p=0.0001), despite no correlation with survival in this cohort, as mentioned earlier, suggesting LCN2 expression is elevated in more aggressive types of breast cancers. We also hypothesized that the LCN2 gene expression score is associated with aggressive clinical parameters such as pathological grade and cancer stage. Further digging detail into the clinic-pathological variable in the METABRIC cohort, our data explained that the LCN2 gene expression was higher for advanced cancer staging in the METABRIC cohort (Fig. 4D).
4G), and the trend was observed in the TCGA cohort (data not included). LCN2 gene expression score was significantly associated with Nottingham pathological grade (grade 3) in the METABRIC cohort (Fig. 4H), and a similar trend was observed in the SCAN-B cohort (p=0.06, grade 3 vs. grade 1 or 2; data not included).

Thus, all these results suggested that LCN2 expression is elevated at the distant metastatic sites and may add prognostic value, particularly to Nottingham pathological grade and advanced cancer staging.

**S100A8 secreted pro-inflammatory mediator encoding gene expression correlates with breast cancer progression and metastasis.**

Among the top 50 genes, those were upregulated in 4T1.2 cell line RNA-seq analysis, S100A8 is another secreted molecule. S100 groups of proteins are cytosolic calcium-binding families of protein play important roles in inflammation and cancer (39-41). S100A8 expression levels are found elevated in several cancers (42-44), including breast cancer (45), could be a biomarker in solid tumors. As S100A8 functions to maintain immune homeostasis (40, 46), we first look at the endogenous expression of S100A8 in the control bone, spine, and lung of the mice. It is not our surprise, qPCR data demonstrated that control bone and lung have a high expression (200 fold normalized value with GAPDH) than the spine (Fig. 5B).

Interestingly, in conformity with our cell line RNA-seq data, 4T1.2 primary breast tumors showed increased expression of S100A8 compared to 4T1 primary breast tumors (Fig. 5A). Surprisingly, S100A8 expression levels are significantly elevated in the spine and metastatic bone lesions (p values are 0.0005 and 0.005, respectively). Lung metastatic lesion showed no further increased expression of S100A8 compared to 4T1.2 primary breast tumors (Fig. 5A). After analyzing TCGA-BRCA and breast cancer metastatic cohort (GSE110590) gene expression data, we found that S100A8 transcripts are altered in metastatic sites with a higher expression tendency in metastasis (Fig. 5C). Given the strong association between high expression of S100A8 in the primary and metastatic lesion in animal models, we hypothesized that breast cancer with a high S100A8 expression score is also associated with worse prognosis of survival and associated with clinical parameters with aggressive breast cancers. Our extensive analysis divulged that a high
expression score of S100A8 was associated with worse survival (PFI, OS, and DSS; p<0.05) of breast cancer patients in METABRIC cohorts (Fig. 5D and the OS data were validated in the SCAN-B cohort of 3,273 patients (HR=1.799, CI 95%; 1.451-2.231, log-rank p= 0.0000001)(Fig. not included)).

Further exploration using TCAGA-BRCA, SCAN-B, and METABRIC cohorts data analysis, revealed that a high S100A8 expression score was associated with the aggressive subtype TNBC or HER2+ breast cancers compared to the HR+ or HER2- breast cancer subtypes (Fig. 5E-G). As shown in Fig. 5H-J, using multiple cohorts data analyses, we have also found that a high S100A8 expression score was associated with higher stages (II/IV) and higher pathological grade (grade 3) of breast cancer patients.

Together, these results suggest that S100A8 is one of the important mediators associated with aggressive breast cancer metastasis and a prognostic factor for worse survival of breast cancer patients.

**ESM-1 secreted endothelial proteoglycan encoding gene expression is associated with poor prognosis of aggressive subtypes of breast cancers.**

Endothelial cell-specific molecule-1 (ESM-1) is a 50-kDa proteoglycan secreted by diverse endothelial cells (ECs) (47). The angiogenic growth factor VEGF or the pro-inflammatory cytokine TNFα upregulates the expression of ESM-1 (48-50). A variety of studies have reported that a high level of ESM-1 secretion is found in several cancers (51-54), including TNBC (55), and ESM-1 levels have been implicated to play a role in tumor metastasis, migration, and vascular invasion in human cancers by regulating the expression of MMPs (53, 56, 57). However, the expression levels of ESM-1 at the metastatic sites, its role in aggressive breast cancers, and whether it could be a prognostic factor of aggressive types of breast cancer were not investigated before.

Consistent with our RNA-seq data, metastatic 4T1.2 cell primary tumors have almost six-fold elevated ESM-1 gene expression compared to the 4T1 primary tumors (Fig 6A).

Further, ECM-1 expression levels were reduced at the metastatic sites such as the spine, bone, and lung; however, its levels at the distant metastatic sites were higher (Fig. 6A) than those found at corresponding control tissues of mice (Fig. 6B). The ECM-1 expression levels were found
elevated in the primary tumors compared to the adjacent control tissues of breast cancer patients of TCAGA-BRCA cohort (P<0.000001, n=1091) and showed diverse distribution at the metastatic sites compared to primary tumors of breast cancer patients of the metastatic cohort (Fig. 6C). To establish the prognostic value of ECM-1 expression, we analyzed survival data in several breast cancer cohorts and represented our analysis in table format (Table1). Deep analysis of survival in different cohorts implies, indeed, ECM-1 expression was linked with poor survival (OS, and DSS, and PFI) of breast cancer patients (Table 1). We hypothesized that ECM-1 is a prognostic factor associated with the aggressive breast cancer progression, higher stage, and pathological grade of breast cancer. As shown in Fig. 6E-F, a higher expression score of ECM-1 was associated with more metastatic (stage III/IV)(Fig. 6F) and higher pathological grade (grade 2/3)(Fig. 6E) of breast cancers. We have also determined the clinical parameters associated with the higher expression score of the ECM-1 gene. Data analysis using multiple cohorts has suggested that a higher expression score of ECM-1 was associated with the more aggressive subtype TNBC or HER2+ compare to the HR+ or HER2- subtypes of breast cancers (Fig. 6D). We further extend our analysis comparing the higher expression of ECM-1 associated with TNBC patients and survival in three breast cancer cohorts (Table 2). Data from explained that the TNBC subtype with high expression of ESM-1 had worse survival (DSS, HR 1.67, p=0.02) and also worse survival for HR+/HER2- patients (OS, DSS, or RFS) compared to HER2+ subtype in the METABRIC cohort (Table 2). ESM-1 high expression was also linked with worse survival (DSS or PFI) of HER2+ vs. other subtype patients in the TCGA-BRCA cohort, associated with the poor OS survival of HER2+ patients of SCAN-B cohort (Table 2). Together, data suggested that the ECM-1 gene is overexpressed in metastatic aggressive subtypes of breast cancers and could be an independent prognostic factor.

**DISCUSSION**

Breast cancer metastasis is a complex process that requires alteration of gene expression allowing tumor cells to escape from the primary tumor site (58). Among several molecular mechanisms that enhance the metastasis process, the secretion of mediator molecules by tumor
cells may facilitate evasion from the immune system’s detection. Primary tumor-derived secretory factors impact the cancer cells in the tumor microenvironment (59). Metastatic cells are challenged to acquire a specific characteristic for the aggressive phenotype (60). Indeed, these secretory molecules may have roles to impact primary tumor cells and influence tumor microenvironment to create favorable conditions for metastases to occur. Therefore, there is a selection of specific phenotypes in cancer cells due to interaction with the tumor microenvironment that evolves with the primary tumor during tumor progression.

So far, to our knowledge, there is no such study where there is a comparison of gene expression between primary tumors and how each particular gene expression changes at the metastatic sites such as the spine, bone, or the lung. To identify genes for metastatic breast cancer, we compared the gene expression of a poorly metastatic 4T1 cell line and its variant, 4T1.2, which was selected in vivo after multiple implantation/resection, a higher tendency for spontaneous metastatic spread using syngeneic mouse models (12, 13). Our study aimed to validate the overexpressed mediator encoding genes of 4T1.2 in implantation/resection metastasis mouse model and develop a prognostic biomarker for the occurrence of distant metastasis in breast cancer based on gene expression profiles of bulk tumors.

Interestingly, primary tumor expressing ANGPTL7, MMP3, LCN2, S100A8, and ESM-1 levels that were upregulated in 4T1.2 cells was strongly associated with breast cancer patients' survival outcomes. ANGPTL7, a member of the angiopoietin-like (ANGPTL) protein family, is a novel pro-angiogenetic inflammatory mediator emerging as an important regulator of metastasis development (22). Although ANGPTL7 is overexpressed in breast cancer (22), the role of ANGPTL7 in breast cancer progression and metastasis is still unclear. The expression of ANGPTL7 is downregulated in the distant metastatic sites such as the spine, bone, and lung compared to the 4T1.2 primary tumors. Furthermore, these data were persistent with the cohort data, suggesting that levels of ANGPTL7 are altered or even downregulated in metastatic breast cancers. Overall survival (OS) data analysis using the SCAN-B cohort of 3,273 breast cancer patients suggested that a high expression score of ANGPTL7 was associated with a better OS. The role of ANGPTL7 in the tumor microenvironment (TME) may be complex associated with
tumor progression and metastasis. For example, CD11b+myeloid cells contribute to hepatic metastasis through the downregulation of ANGPTL7. Further, depletion of CD11b+myeloid cells or overexpression of ANGPTL7 significantly inhibits hepatic metastasis formation and angiogenesis (61). These experimental data suggest that cancer cells inhibit angiopoietin-like protein gene expression to progress metastasis, consistent with our observation in our metastatic animal model.

Matrix metalloproteinases (MMPs), the inflammatory mediators, are important for tumor cell invasion by degrading ECM and regulating cancer progression and metastasis (28). We found that MMP3 expression is high in the primary tumor that was overexpressed in the 4T1.2 cells. MMP3 levels were downregulated in distant metastatic sites such as the spine, bone, and lung in the implantation/resection mouse model. MMP3 exhibits both tumor-promoting and tumor-inhibiting effects based on the substrates that it degrades. Although, by using web-based KM Plotter (62), it has been shown that patients expressing higher levels of MMP3 had a significantly poorer outcome for distant metastasis-free survival (DMFS) (63). However, its role in metastatic sites and as a prognostic factor for breast cancer patients using larger cohorts (TCGA, METABRIC, or SCAN-B) was unknown. Out data suggested that MMP3 levels are downregulated at distant organs such as the lung in the metastatic cohort of breast cancer patients.

Further data analysis using SCAN-B and METABRIC cohorts of breast cancer patients suggested that a high MMP3 expression score was associated with the better OS of patients (HR<1, P=0.0001), suggesting MMP3 is not a poor prognostic factor of breast cancer. Altogether, our data explained that ANGPTL7 and MMPs 3 and 13 were regulated inversely in metastatic sites among several genes that were upregulated in primary tumors. Cancer researchers are paying attention to those upregulated genes and finding a strategy to downregulate these genes and associated pathways to get better prognostic clinical output. It is proven that antiangiogenesis drug treatment increased resistance and metastasis. Interestingly, in our animal model, we observed that ANGPTL7, MMP3, and MMP13 might be required to be downregulated in cancer cells to enhance metastasis distantly without any drug treatment.

LCN2 is a secreted inflammatory mediator, is elevated in varieties of cancers, and is associated with breast cancer progression (36-38). However, its role at the metastatic sites such as spine,
bone, and the lung compared to the primary tumors and whether it could be a prognostic factor for aggressive breast cancer using larger cohorts was unknown. LCN2 expression that was high in 4T1.2 cells and 4t1.2 primary tumors was further markedly elevated in the metastatic sites such as spine and bone, but not the lung in the resection/metastasis mouse model. Although web-based KM plotter database data analysis suggested that LCN2 expression predicts poor clinical outcome in TNBC (64), our results using larger cohorts SCAN-B and METABRIC of breast cancer patients data analysis suggested that LCN2 expression score is high in aggressive breast cancer HER2+ or TNBC, but not linked with any survival outcome of these subtypes. Interestingly, we found that a high LCN2 expression score was associated with the higher stages and higher pathological grade of breast cancer patients. We found that a high LCN2 expression score was associated with poorer DSS and PFI survival outcomes of the METABRIC cohort of breast cancer patients, suggesting LCN2 is an important target for breast cancer distant metastasis.

S100A8 is a secreted inflammatory mediator, is increasingly recognized as a biomarker in many solid tumors, including breast cancers (45, 65). However, a few studies have uncovered that individual S100A8 plays a vital role as a prognostic factor for breast cancer patients. Infiltrating S100A8+ myeloid cells coordinated by macrophage inhibitory factor (MIF) result in poor overall survival (OS) and shorter metastasis-free survival in breast cancer patients. Studies strongly suggest that S100A8 is expressed by cancer cells as well as by infiltrating immune and myeloid cells (66). Since S100A8 could be upregulated by different conditions such as oxidative stress, cytokines, and growth factors in many types of cells, in agreement, we found that control bone and lung have a high expression compared to the spine. LCN2 is markedly elevated in distant sites such as spine and bone compared to the primary 4T1.2 tumors, suggesting it could be an important target for breast cancer spine and bone metastases. In agreement with the idea, a high expression score of LCN2 was strongly associated with patient survival of the METABRIC cohort (OS, PFI, and DSS) and was associated with aggressive subtypes (TNBC and HER2+) with higher stages (II/IV) and higher grade (grade 3) of breast cancer patients.

ESM-1 endothelial secreted mediator is elevated in a variety of cancers, including breast cancers (55), and it has been implicated in playing a role in tumor metastasis in other cancers. However,
its role in breast cancer metastasis and if it could be a prognostic factor for more aggressive subtypes of breast cancer was not known. ESM1 expression that was elevated in the 4T1.2 cells and primary tumors was further downregulated in distant sites such as the spine, bone, and the lung in our implantation/resection metastasis mouse model. Using multiple larger cohorts of breast cancer patients (METABRIC, SCAN-B, and TCGA-BRCA), we have demonstrated that the high expression score of ESM-1 instead exhibits as an independent prognostic factor for worse survival of breast cancer patients. Based on our hypothesis that ESM-1 is an important prognostic factor for aggressive subtypes of breast cancers, we found that a high expression score of ESM-1 was associated with a higher grade and higher stages of breast cancer patients. High expression of ESM-1 was also associated with TNBC, HER2+, in some cases, HR+/HER2- breast cancer patients’ survival outcomes. Our data suggested that mediators encoding genes with both prognostic and predictive values may, therefore, be clinically useful for breast cancer spine, bone, and lung metastasis in particularly in more aggressive subtypes such as TNBC or HER2+ breast cancer.

Materials and Methods

Chemicals and reagents. Cell culture medium, α-MEM (Minimum Essential Medium α), trypsin, penicillin-streptomycin, Phosphate Buffered Saline (PBS), sodium pyruvate and D-luciferin were purchased from Thermo-Fisher Scientific (Waltham, MA, USA). Fetal bovine serum (FBS) was purchased from Peak Serum (Peak Serum, Wellington, CO, USA). Cell culture.

The 4T1-Luc+ mouse breast cancer cell line and the 4T1.2-Luc+ metastatic variant of 4T1 parental cell line (12) were kindly provided by Prof. Cheryl L. Jorcyk of Boise State University, Department of Biological Sciences, Biomolecular Sciences Program, 1910 University Drive, Boise, ID, 83725, USA. Both the cell lines were cultured and maintained in α-MEM supplemented with 10% fetal bovine serum, 1 mM Penicillin/Streptomycin, and 1 mM sodium pyruvate at 37°C in 5% CO2 and 95% humidity, as mentioned before (67-69).

RNA preparation and RNA-seq.
4T1-Luc+ and 4T1-Luc+ cell lines were used in triplicates for RNA-seq analysis. The cell lines were cultured in a full-serum medium, as mentioned above, and harvested at sub-confluence (60% confluence) for RNA isolation by using a total RNA purification kit with a DNAse treatment step (Qiagen, Valencia, CA, USA). The total RNA quality was evaluated using the Agilent 2100 bioanalyzer (Agilent, Palo Alto, CA, USA) with the RNA 6000 Nano LabChip kit. RNA-seq libraries were prepared using TruSeq Stranded Total RNA Library Prep Gold kit (Illumina, San Diego, CA, USA). The quality of the libraries was validated by assaying using TapeStation D1000 ScreenTape (Agilent) and Library Quantification Kit (Kapa Biosystems, Wilmington, MA, USA). All the top quality libraries were sequenced together on an Illumina HiSeq 2500 instrument using HiSeq Rapid Cluster Kit v2—Paired-End and Rapid SBS Kit v2 reagents to obtain paired reads of 100 bases. Casava software (version 1.8.2, Illumina Inc, San Diego, CA, USA) was used to demultiplex the sequencing data. An average of 100 million sequence read-pairs was obtained for each sample. Raw read data were filtered and mapped in a splicing-aware manner with the data processing using TopHat2 software (70). Correct read alignment across splice junctions of the raw data was performed by Gencode v25 gtf (71). HTSeq framework (72) was used for gene-level mapped read count values of the RNA-seq data. Finally, the read count data were normalized, and log2 fold changes were estimated for further analyses with DESeq2 (73).

**Gene expression analyses.**

The DESeq2 Bioconductor package for R was used for differential gene expression analyses. Gene levels were considered as differentially expressed based on the expression levels changes of absolute log2 fold-change (FC) >1.2 and false discovery rate (FDR) <0.05 after adjustment for multi-testing by Benjamini–Hochberg method. The Benjamini–Hochberg method was used to calculate the FDR (<0.05) as a cutoff to identify the candidate genes for multiple testing adjustments.

**Quantitative-real-time PCR (qRT-PCR).**

According to the manufacturer's instructions, as described before (67-69, 74), cDNA was synthesized from DNase pre-treated 1 μg RNA using the SuperScript cDNA Synthesis kit (Life Technologies, Carlsbad, CA, USA). The PCR reaction was performed by the thermal cycler (Bio-
Rad, Hercules, CA, USA) using SYBR Green qPCR SuperMixes with PCR primers for the mouse (ANGPTL7: F- TGACTGTTCTTCCCTGTACCA, R- CAAGGCCACTCTTACGTCTCT; MMP3: F- ACATGGAGACTTTGTCCCTTTTG, R- TTGGCTGAGTGTTAGAGTCCC; LCN2: F- TGGCCCTGAGTGTCATGTG, R- CTCTTTGTAGCTCATAGATGGTC; S100A8: F- AAATCACCATGCCCTCTACAAG, R- CCCACTTTTATCACCATCGCAA; ESM-1: F- CTGGAGCGCCAAATATGGCG, R- TGAGACTGTACGGTAGCAGGT; GAPDH: F- TGATTTTGGACGCATTGGTC, R- TTTGCACTGGTACGTGTTGAT). All the primer sequences were used from the publicly available mouse primer bank (https://pga.mgh.harvard.edu/primerbank/index.html). Each qPCR sample was run at least in triplicate. The relative level of target genes from each sample was calculated by the 2^(-delta Cq) method (75, 76) and normalizing to the house-keeping gene GAPDH.

Ethical statement

All animal methods were approved by the Institutional Animal Care and Use Committee of Roswell Park Comprehensive Cancer Center and were performed in accordance with the relevant guidelines and regulations for the American Association of Laboratory Animal Care.

Animals and tumor cell implantations

Female Balb/c mice, 12 weeks of age, and approximately 20g/mice were obtained from The Jackson Laboratories (Bar Harbor, ME). 4T1-Luc+ and 4T1.2-Luc+ cells suspended in the culture medium at a concentration of 1×10^7 cells/ml, and 10 μl of this solution were then implanted as described below (77, 78). All cancer cell implantations were performed under isoflurane anaesthesia using sterile technique. A 5 mm incision was made medial to the nipple, and a cotton swab was used to expose the mammary gland. The cells were implanted directly into the mammary gland of mice (n=5) under direct vision, using ×10 microscopic magnifications, and the wound was closed with a nylon suture. Xenogen In Vivo Imaging Systems (IVIS®) 200 and Living Image® software (Caliper Life Sciences, Hopkinton, MA) was used to quantify the photon/sec emitted by 4T1-Luc+/4T1.2-Luc+ cells after intraperitoneal injection of 200 μl (150 mg/kg) of D-luciferin (Fisher Scientific, Inc. USA) for the determination of in vivo tumor burden. After day 3 of
the inoculation of cancer cells, mice were randomized based on equal initial low levels of photon counts.

On day 7, IVIS live imaging was performed for the tumor-bearing animals, and no distal metastatic spread of cancer cells was observed. Seven days after implantation, the entire primary tumors were resected from the animals, and the incision was closed. IVIS live imaging confirmed no residual Luc+ cells at the primary sites or any metastatic spread. 4T1.2-Luc+-tumor resected mice were kept for another 10 days for tumor recurrence at distant sites. Distant metastasis of 4T1.2-Luc+ tumors were regularly monitored by IVIS live imaging. On day 17 of inoculation, IVIS live imaging was performed, and mice images were shown, suggesting 4T1.2-Luc+ cells recurred to the distant sites, including the lung and the bone. On day 17, the magnetic resonance imaging (MRI) data confirmed the 4T1.2-Luc+ cancer cell colonization in the spinal cord.

Following instructions, MRI was performed using a laboratory animal MRI scanner (Bruker Medical Inc.) with a magnetic field of 4.7T (Translational Imaging Shared Resource, Roswell Park Cancer Institute). 4T1.2-Luc+ metastatic cancer-bearing mice (n=5) were scanned in multiple stages using transverse and sagittal projections. T2-weighted spin-echo images were acquired for the mice. In T2-weighted imaging, the field of view was 3.2x3.2 cm, and the thickness of the slice was 1 mm. The total duration of the MRI scan for each mouse was 20 min. A representative image of the MRI results was shown, suggesting cancer cell recurrence at spine bone. Mice were sacrificed, and metastatic lesions of the spine, bone, and lung were collected based on ex vivo IVIS confirmation. Primary tumors and the metastatic lesions of the distant organs were used for molecular analyses.

**Clinical and gene expression data analyses of breast cancer patient cohorts.**

Publicly available clinical parameters and tumor gene expression data for 1091 patients of the TCGA breast cancer (BRCA) Project (24) and 1094 patients for the METABRIC cohort (29) were obtained for analyses from cBioPortal (79). Tumor gene expression and clinical data of 3273 breast cancer patients were also obtained for the SCAN-B cohort (26). In terms of survival data, disease-specific survival (DSS), overall survival (OS), and progression-free survival (PFI) were available in TCGA and METABRIC cohorts, and only OS data were available for the SCAN-B cohort. Normalized microarray-based gene expression data (log2-transformed data) for primary
tumors and metastatic sites, including the bone lymph node (LN) and the lung, were obtained for analyses from the GEO repository (GSE110590 cohort of 16 patients (25)). Univariate Cox regression analysis was performed using the cohorts to determine which gene's expression may be an independent prognostic marker for patient survival. Among the top 50 differentially upregulated genes in 4T1.2 cells compared to parental 4T1 cells, ANGPTL7, MMP3, LCN2, S100A8, and ESM-1 were used for analyses for the cohorts. The high/low cutoff for any gene was defined as the top-third vs. the bottom two-thirds within any cohort.

Statistical Analyses

For qPCR data, an unpaired two-tailed Student's t-test was used to compare two groups (using GraphPad Prism version 8.0). The levels of mRNAs are expressed as the means ± SEM. We had used a one-way ANOVA test for datasets containing multiple group comparisons and Tukey's posthoc test for the family-wise error rate comparison. Data plotting were performed using GraphPad Prism 8 or Microsoft Excel (version 16 for Windows, Redmond, WA, USA). Kaplan–Meier method with log-rank test was used for survival analysis. For all analyses, P≤0.05 was considered statistically significant.

Author Contributions: A.M., I.O., M.O., M.O., W.T., T.K., E.K., Y.L., and S.P., were involved in analyzing data and study design. K.T. provided reagents. A.M., I.O., M.O., M.O., K.T, and N.C.H., were involved in animal experiments and data analyses. A.M. and N.C.H. were involved in molecular biology experiments and data analyses. W.T., S.P., E.K., and L.Y. analyzed data. A.M. and N.C.H. wrote the original draft of the manuscript. I.O., M.O., W.T., T.K., E.K., K.T., Y.L., S.P., reviewed, and edited the manuscript. All authors contributed to the editing of the manuscript. N.C.H. and A.M. developed the concept of the study, supervised the study, editing the manuscript, and prepared the final figures and tables. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Roswell Park Health Research Incorporated (HRI) Start Up Funds #714084-01 (NCH). This work was also supported by NCI cancer center support Grant P30CA016056, involving the use of Roswell Park Comprehensive Cancer Center's Bioinformatics and Biostatistics Shared Resources.
Acknowledgments: We thank Prof. Cheryl L. Jorcyk of Boise State University, ID, USA for providing the 4T1.2-Luc+ and 4T1-Luc+ cell lines. We also thank the Genomics Shared Resource of RPCCC for performing the sequencing experiment.

Conflicts of Interest: The authors declare no conflict of interest.

References:

1. Chambers, A. F., Groom, A. C., and MacDonald, I. C. (2002) Dissemination and growth of cancer cells in metastatic sites. *Nat Rev Cancer* 2, 563-572
2. Klein, C. A. (2009) Parallel progression of primary tumours and metastases. *Nat Rev Cancer* 9, 302-312
3. Husemann, Y., Geigl, J. B., Schubert, F., Musiani, P., Meyer, M., Burghart, E., Forni, G., Eils, R., Fehm, T., Riethmüller, G., and Klein, C. A. (2008) Systemic spread is an early step in breast cancer. *Cancer Cell* 13, 58-68
4. Fisher, B., and Redmond, C. (1992) Systemic therapy in node-negative patients: updated findings from NSABP clinical trials. *National Surgical Adjuvant Breast and Bowel Project. J Natl Cancer Inst Monogr*, 105-116
5. Nagrath, S., Sequist, L. V., Maheswaran, S., Bell, D. W., Irimia, D., Ulkus, L., Smith, M. R., Kwak, E. L., Digumarthy, S., Muzikansky, A., Ryan, P., Balis, U. J., Tompkins, R. G., Haber, D. A., and Toner, M. (2007) Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* 450, 1235-1239
6. Matrisian, L. M., Bowden, G. T., Krieg, P., Furstenberger, G., Briand, J. P., Leroy, P., and Breathnach, R. (1986) The mRNA coding for the secreted protease transin is expressed more abundantly in malignant than in benign tumors. *Proc Natl Acad Sci U S A* 83, 9413-9417
7. Gunthert, U., Hofmann, M., Rudy, W., Reber, S., Zoller, M., Haussmann, I., Matzku, S., Wenzel, A., Ponta, H., and Herrlich, P. (1991) A new variant of glycoprotein CD44 confers metastatic potential to rat carcinoma cells. *Cell* 65, 13-24
8. Kim, S. H., Choi, S. J., Park, J. S., Lee, J., Cho, Y. B., Kang, M. W., Lee, W. Y., Choi, Y. S., Kim, H. K., Han, J., Chun, H. K., and Kim, J. (2012) Tropism between hepatic and pulmonary metastases in colorectal cancers. *Oncol Rep* 28, 459-464
9. Minn, A. J., Kang, Y., Serganova, I., Gupta, G. P., Giri, D. D., Dubrovin, M., Ponomarev, V., Gerald, W. L., Blasberg, R., and Massague, J. (2005) Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. *J Clin Invest* 115, 44-55
10. Santulli, G. (2014) Angiopoietin-like proteins: a comprehensive look. *Front Endocrinol (Lausanne)* 5, 4
11. Morisada, T., Kubota, Y., Urano, T., Suda, T., and Oike, Y. (2006) Angiopoietins and angiopoietin-like proteins in angiogenesis. *Endothelium* 13, 71-79
12. Oike, Y., Ito, Y., Maekawa, H., Morisada, T., Kubota, Y., Akao, M., Urano, T., Yasunaga, K., and Suda, T. (2004) Angiopoietin-related growth factor (AGF) promotes angiogenesis. *Blood* 103, 3760-3765
13. Oike, Y., Yasunaga, K., and Suda, T. (2004) Angiopoietin-related angiopoietin-like proteins regulate angiogenesis. *Int J Hematol* 80, 21-28
14. Hato, T., Tabata, M., and Oike, Y. (2008) The role of angiopoietin-like proteins in angiogenesis and metabolism. *Trends Cardiovasc Med* 18, 6-14
19. Peek, R., Verbraak, F., Coevoet, H. M., and Kijlstra, A. (1998) Muller cell-specific autoantibodies in a patient with progressive loss of vision. *Invest Ophthalmol Vis Sci* **39**, 1976-1979

20. Peek, R., van Gelderen, B. E., Bruinenberg, M., and Kijlstra, A. (1998) Molecular cloning of a new angiopoietin-like factor from the human cornea. *Invest Ophthalmol Vis Sci* **39**, 1782-1788

21. Kuchtey, J., Kallberg, M. E., Gelatt, K. N., Rinkoski, T., Komaromy, A. M., and Kuchtey, R. W. (2008) Angiopoietin-like 7 secretion is induced by glaucoma stimuli and its concentration is elevated in glaucomatous aqueous humor. *Invest Ophthalmol Vis Sci* **49**, 3438-3448

22. Parri, M., Pietrovito, L., Grandi, A., Campagnoli, S., De Camilli, E., Bianchini, F., Marchio, S., Bussolino, F., Jin, B., Sarmientos, P., Grandi, G., Viale, G., Pileri, P., Chiarugi, P., and Grifantini, R. (2014) Angiopoietin-like 7, a novel pro-angiogenic factor over-expressed in cancer. *Angiogenesis* **17**, 881-896

23. Carbone, C., Piro, G., Merz, V., Simionato, F., Santoro, R., Zecchetto, C., Tortora, G., and Melisi, D. (2018) Angiopoietin-Like Proteins in Angiogenesis, Inflammation and Cancer. *Int J Mol Sci* **19**

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

25. Siegel, M. B., He, X., Hoadley, K. A., Hoyle, A., Pearson, J. B., Garrett, A. L., Kumar, S., Moylan, V. J., Brady, C. M., Van Swearingen, A. E., Marron, D., Gupta, G. P., Thorne, L. B., Kieran, N., Livasy, C., Mardis, E. R., Parker, J. S., Chen, M., Anders, C. K., Carey, L. A., and Perou, C. M. (2018) Integrated RNA and DNA sequencing reveals early drivers of metastatic breast cancer. *J Clin Invest* **128**, 1371-1383

26. Brueffer, C., Vallon-Christersson, J., Grabau, D., Ehinger, A., Hakkinen, J., Hegardt, C., Malina, J., Chen, Y., Bendahl, P. O., Manjer, J., Malberg, M., Larsson, C., Loman, N., Ryden, L., Borg, A., and Saal, L. H. (2018) Clinical Value of RNA Sequencing-Based Classifiers for Prediction of the Five Conventional Breast Cancer Biomarkers: A Report From the Population-Based Multicenter Sweden Cancerome Analysis Network-Breast Initiative. *JCO Precis Oncol* **2**

27. Lochter, A., Galosy, S., Muschler, J., Freedman, N., Werb, Z., and Bissell, M. J. (1997) Matrix metalloproteinase stromelysin-1 triggers a cascade of molecular alterations that leads to stable epithelial-to-mesenchymal conversion and a premalignant phenotype in mammary epithelial cells. *J Cell Biol* **139**, 1861-1872

28. Duffy, M. J., Maguire, T. M., Hill, A., McDermott, E., and O’Higgins, N. (2000) Metalloproteinases: role in breast carcinogenesis, invasion and metastasis. *Breast Cancer Res* **2**, 252-257

29. Curtis, C., Shah, S. P., Chin, S. F., Tarashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A. G., Samarajiwa, S., Yuan, Y., Graf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Group, M., Langerod, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Borresen-Dale, A. L., Brenton, J. D., Tavare, S., Caldas, C., and Aparicio, S. (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346-352

30. Lindy, O., Konttinen, Y. T., Sorsa, T., Ding, Y., Santavirta, S., Ceponis, A., and Lopez-Otin, C. (1997) Matrix metalloproteinase 13 (collagenase 3) in human rheumatoid synovium. *Arthritis Rheum* **40**, 1391-1399

31. Neuhold, L. A., Killar, L., Zhao, W., Sung, M. L., Warner, L., Kulik, J., Turner, J., Wu, W.,Billinghurst, C., Meijers, T., Poole, A. R., Babij, P., and DeGennaro, L. J. (2001) Postnatal expression in hyaline cartilage of constitutively active human collagenase-3 (MMP-13) induces osteoarthritis in mice. *J Clin Invest* **107**, 35-44

32. Wang, X., and Cao, X. (2014) Regulation of metastasis of pediatric multiple myeloma by MMP13. *Tumour Biol* **35**, 8715-8720

33. Nickolas, T. L., Barash, J., and Devarajan, P. (2008) Biomarkers in acute and chronic kidney disease. *Curr Opin Nephrol Hypertens* **17**, 127-132

34. Yan, Q. W., Yang, Q., Mody, N., Graham, T. E., Hsu, C. H., Xu, Z., Houstis, N. E., Kahn, B. B., and Rosen, E. D. (2007) The adipokine lipocalin 2 is regulated by obesity and promotes insulin resistance. *Diabetes* **56**, 2533-2540

35. Devarajan, P. (2007) Neutrophil gelatinase-associated lipocalin: new paths for an old shuttle. *Cancer Ther* **5**, 463-470

36. Yang, J., and Moses, M. A. (2009) Lipocalin 2: a multifaceted modulator of human cancer. *Cell Cycle* **8**, 2347-2352

37. Shi, H., Gu, Y., Yang, J., Xu, L., Mi, W., and Yu, W. (2008) Lipocalin 2 promotes lung metastasis of murine breast cancer cells. *Journal of experimental & clinical cancer research : CR* **27**, 83-9966-9927-9983

38. Yang, J., Bielenberg, D. R., Rodig, S. J., Doiron, R., Clifton, M. C., Kung, A. L., Strong, R. K., Zurakowski, D., and Moses, M. A. (2009) Lipocalin 2 promotes breast cancer progression. *Proc Natl Acad Sci U S A* **106**, 3913-3918
39. Leclerc, E., Fritz, G., Vetter, S. W., and Heizmann, C. W. (2009) Binding of S100 proteins to RAGE: an update. *Biochim Biophys Acta* 1793, 993-1007
40. Donato, R., Cannon, B. R., Sorci, G., Ruiuzzi, F., Hsu, K., Weber, D. J., and Geczy, C. L. (2013) Functions of S100 proteins. *Curr Mol Med* 13, 24-57
41. Gross, S. R., Sin, C. G., Barraclough, R., and Rudland, P. S. (2014) Joining S100 proteins and migration: for better or for worse, in sickness and in health. *Cell Mol Life Sci* 71, 1551-1579
42. Funk, S., Mark, R., Bayo, P., Flechtenmacher, C., Grabe, N., Angel, P., Plinkert, P. K., and Hess, J. (2015) High S100A8 and S100A12 protein expression is a favorable prognostic factor for survival of oropharyngeal squamous cell carcinoma. *Int J Cancer* 136, 2037-2046
43. Fujita, Y., Khateb, A., Li, Y., Tinoco, R., Zhang, T., Bar-Yoseph, H., Tam, M. A., Chowers, Y., Sabo, E., Gerassy-Vainberg, S., Starosvetsky, E., James, B., Brown, K., Shen-Orr, S. S., Bradley, L. M., Tessier, P. A., and Ronai, Z. A. (2018) Regulation of S100A8 Stability by RNF5 in Intestinal Epithelial Cells Determines Intestinal Inflammation and Severity of Colitis. *Cell Rep* 24, 3296-3311 e3296
44. Tajmul, M., Parween, F., Singh, L., Mathur, S. R., Sharma, J. B., Kumar, S., Sharma, D. N., and Yadav, S. (2018) Identification and validation of salivary proteomic signatures for non-invasive detection of ovarian cancer. *Int J Biol Macromol* 108, 503-514
45. Miller, P., Kidwell, K. M., Thomas, D., Sabel, M., Rae, J. M., Hayes, D. F., Hudson, B. I., El-Ashry, D., and Lippman, M. E. (2017) Elevated S100A8 protein expression in breast cancer cells and breast tumor stroma is prognostic of poor disease outcome. *Breast Cancer Res Treat* 166, 85-94
46. Donato, R. (2003) Intracellular and extracellular roles of S100 proteins. *Microsc Res Tech* 60, 540-551
47. Bechard, D., Gentina, T., Delehedde, M., Scherpereel, A., Lyon, M., Aumercier, M., Vazeux, R., Richet, C., Degand, P., Jude, B., Janin, A., Fernig, D. G., Tonnel, A. B., and Lassalle, P. (2001) Endocan is a novel chondroitin sulfate/dermatan sulfate proteoglycan that promotes hepatocyte growth factor/scatter factor mitogenic activity. *J Biol Chem* 276, 48341-48349
48. Bechard, D., Meignin, V., Scherpereel, A., Oudin, S., Kervoaze, G., Bertheau, P., Janin, A., Tonnel, A., and Lassalle, P. (2000) Characterization of the secreted form of endothelial-cell-specific molecule 1 by specific monoclonal antibodies. *J Vasc Res* 37, 417-425
49. Rennel, E., Mellberg, S., Dimberg, A., Petersson, L., Botling, J., Ameer, A., Westholm, J. O., Komorowski, J., Lassalle, P., Cross, M. J., and Gerwins, P. (2007) Endocan is a VEGF-A and PI3K regulated gene with increased expression in human renal cancer. *Exp Cell Res* 313, 1285-1294
50. Shin, J. W., Huggenberger, R., and Detmar, M. (2008) Transcriptional profiling of VEGF-A and VEGF-C target genes in lymphatic endothelium reveals endothelial-specific molecule-1 as a novel mediator of lymphangiogenesis. *Blood* 112, 2318-2326
51. Gamarra, F., Noel, J. L., Brunelli, A., Dingemans, A. C., Felip, E., Gaga, M., Grigoriu, B. D., Hardavella, G., Huber, R. M., Janes, S., Massard, G., Putora, P. M., Sculier, J. P., Schnabel, P. A., Ramella, S., Van Raemdonck, D., and Meert, A. P. (2016) Thoracic oncology HERMES: European curriculum recommendations for training in thoracic oncology. *Breathe (Sheff)* 12, 249-255
52. Leroy, X., Aubert, S., Zini, L., Franquet, H., Kervoaze, G., Villers, A., Delehedde, M., Copin, M. C., and Lassalle, P. (2010) Vascular endocan (ESM-1) is markedly overexpressed in clear cell renal cell carcinoma. *Histopathology* 56, 180-187
53. Kang, Y. H., Ji, N. Y., Han, S. R., Lee, C. I., Kim, J. W., Yeom, Y. I., Kim, Y. H., Chun, H. K., Kim, J. W., Chung, J. W., Ahn, D. K., Lee, H. G., and Song, E. Y. (2012) ESM-1 regulates cell growth and metastatic process through activation of NF-kappaB in colorectal cancer. *Cell Signal* 24, 1940-1949
54. Maurage, C. A., Adam, E., Mineo, J. F., Sarrazin, S., Debuinne, M., Siminski, R. M., Baroncini, M., Lassalle, P., Blond, S., and Delehedde, M. (2009) Endocan expression and localization in human glioblastomas. *J Neuropathol Exp Neurol* 68, 633-641
55. Sagara, A., Igarashi, K., Otsuka, M., Kodama, A., Yamashita, M., Sugisura, R., Karasawa, T., Arakawa, K., Narita, M., Kuzumaki, N., Narita, M., and Kato, Y. (2017) Endocan as a prognostic biomarker of triple-negative breast cancer. *Breast Cancer Res Treat* 161, 269-278
56. Liu, N., Zhang, L. H., Du, H., Hu, Y., Zhang, G. G., Wang, X. H., Li, J. Y., and Ji, J. F. (2010) Overexpression of endothelial cell specific molecule-1 (ESM-1) in gastric cancer. *Ann Surg Oncol* 17, 2628-2639
57. Kang, Y. H., Ji, N. Y., Lee, C. I., Lee, H. G., Kim, J. W., Yeom, Y. I., Kim, D. G., Yoon, S. K., Kim, J. W., Park, P. J., and Song, E. Y. (2011) ESM-1 silencing decreased cell survival, migration, and invasion and modulated cell cycle progression in hepatocellular carcinoma. *Amino Acids* 40, 1003-1013
58. Jin, X., and Mu, P. (2015) Targeting Breast Cancer Metastasis. *Breast Cancer (Auckl)* 9, 23-34
59. Klemm, F., and Joyce, J. A. (2015) Microenvironmental regulation of therapeutic response in cancer. *Trends in cell biology* 25, 198-213
60. Nguyen, D. X., Bos, P. D., and Massague, J. (2009) Metastasis: from dissemination to organ-specific colonization. Nat Rev Cancer 9, 274-284

61. Lim, S. Y., Gordon-Weeks, A., Allen, D., Kersemans, V., Beech, J., Smart, S., and Muschel, R. J. (2015) Cd11b(+) myeloid cells support hepatic metastasis through down-regulation of angiopoietin-like 7 in cancer cells. Hepatology 62, 521-533

62. Gyorffy, B., Lanczky, A., Eklund, A. C., Denkert, C., Budczies, J., Li, Q., and Szallasi, Z. (2010) An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. Breast Cancer Res Treat 123, 725-731

63. Mehner, C., Miller, E., Nassar, A., Bamlet, W. R., Radisky, E. S., and Radisky, D. C. (2015) Tumor cell expression of MMP3 as a prognostic factor for poor survival in pancreatic, pulmonary, and mammary carcinoma. Genes Cancer 6, 480-489

64. Cheng, G., Sun, X., Wang, J., Xiao, G., Wang, X., Fan, X., Zu, L., Hao, M., Qu, Q., Mao, Y., Xue, Y., and Wang, J. (2014) HIC1 silencing in triple-negative breast cancer drives progression through misregulation of LCN2. Cancer Res 74, 862-872

65. Acharyya, S., Oskarsson, T., Vanharanta, S., Malladi, S., Kim, J., Morris, P. G., Manova-Todorova, K., Leversha, M., Hogg, N., Seshan, V. E., Norton, L., Brogi, E., and Massague, J. (2012) A CXCL1 paracrine network links cancer chemo resistance and metastasis. Cell 150, 165-178

66. Gebhardt, C., Nemeth, J., Angel, P., and Hess, J. (2006) S100A8 and S100A9 in inflammation and cancer. Biochem Pharmacol 72, 1622-1631

67. Hait, N. C., Allegood, J., Maceyka, M., Strub, G. M., Harikumar, K. B., Singh, S. K., Luo, C., Marmorstein, R., Kordula, T., Milstien, S., and Spiegel, S. (2009) Regulation of histone acetylation in the nucleus by sphingosine-1-phosphate. Science (New York, N.Y.) 325, 1254-1257

68. Hait, N. C., Maiti, A., Xu, P., Qi, Q., Kawaguchi, T., Okano, M., Takabe, K., Yan, L., and Luo, C. (2020) Regulation of hypoxia-inducible factor functions in the nucleus by sphingosine-1-phosphate. FASeB J 34, 4293-4310

69. Maiti, A., Qi, Q., Peng, X., Yan, L., Takabe, K., and Hait, N. C. (2019) Class I histone deacetylase inhibitor suppresses vasculogenic mimicry by enhancing the expression of tumor suppressor and anti-angiogenesis genes in aggressive human TNBC cells. International journal of oncology

70. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14, R36

71. Mudge, J. M., Jungrvais, I., Hunt, T., Gonzalez, J. M., Wright, J. C., Kay, M., Davidson, C., Fitzgerald, S., Seal, R., Tweedie, S., He, L., Waterhouse, R. M., Li, Y., Bruford, E., Choudhary, J. S., Frankish, A., and Kellis, M. (2019) Discovery of high-confidence human protein-coding genes and exons by whole-genome PhyloCSF helps elucidate 118 GWAS loci. Genome Res 29, 2073-2087

72. Anders, S., Pyl, P. T., and Huber, W. (2015) HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166-169

73. Love, M. I., Huber, W., and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550

74. Maiti, A., Takabe, K., and Hait, N. C. (2017) Metastatic triple-negative breast cancer is dependent on SphKs/S1P signaling for growth and survival. Cellular signalling 32, 85-92

75. Schmittgen, T. D., and Livak, K. J. (2008) Analyzing real-time PCR data by the comparative C(T) method. Nature protocols 3, 1101-1108

76. Livak, K. J., and Schmittgen, T. D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif.) 25, 402-408

77. Rashid, O. M., Nagahashi, M., Ramachandran, S., Graham, L., Yamada, A., Spiegel, S., Bear, H. D., and Takabe, K. (2013) Resection of the primary tumor improves survival in metastatic breast cancer by reducing overall tumor burden. Surgery 153, 771-778

78. Hait, N. C., Avni, D., Yamada, A., Nagahashi, M., Aoyagi, T., Aoki, H., Dumur, C. I., Zelenko, Z., Gallagher, E. J., Leroith, D., Milstien, S., Takabe, K., and Spiegel, S. (2015) The phosphorylated prodrug FTY720 is a histone deacetylase inhibitor that reactivates ERalpha expression and enhances hormonal therapy for breast cancer. Oncogenes 4, e156

79. Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., and Schultz, N. (2013) Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci Signal 6, pl1
Figure 1. Proangiogenic and pro-inflammatory secretory protein-encoding genes are upregulated in metastatic variant 4T1.2 vs. parental 4T1 cells. (A-B) RNA-Seq heat map for 50 upregulated and 50 downregulated genes in 4T1.2 (red) vs. 4T1 cells (blue). Genes indicated in bold red arrow revealed a fold change (>6 fold vs. 4T1) is the mediator protein-encoding genes are elevated in the 4T1.2 cells. (C-D) On day 17 of the inoculation of cancer cells in the implantation/resection model, IVIS live images of Balb/c mice bearing metastatic 4T1.2-Luc+ breast cancer cells at primary and metastatic sites, supine (C), and prone (D) positions are shown. Arrow indicates hind bone 4T1.2-Luc+ metastasis. Representative mice images are shown (n=5). (E), MRI of 4T1.2-Luc+ spine bone metastasis in vivo. On the day of sacrifice (day 17) of 4T1.2-Luc+ metastatic cancer-bearing Balb/c mice, an MRI scan was performed for the mice, and representative images were shown. E, top left and right, two representative T2-weighted transverse slice images of Balb/c mice with 4T1.2-Luc+ metastatic cancer. The spinal cord was indicated with the arrow in both images. E, bottom left and right, two representative spinal cord images visualized by the T2-weighted MRI, sagittal projection. Arrows indicate spine bone metastasis. Data are representative of n=5 mice.

Figure 2. ANGPTL7 is downregulated at the metastatic sites vs. primary tumor and positively impacts breast cancer patients' survival. (A) ANGPTL7 mRNA levels, determined by qPCR analysis, were compared between primary tumors of 4T1 vs. 4T1.2 cells (n=3/group). Metastatic lesions of 4T1.2 cells at distant sites such as the spine, bone, and the lung, were used for ANGPTL7 mRNA levels using qPCR analysis (n=3). (B) ANGPTL7 expression levels were also estimated in the control mice (n=3) tissues of the spine, bone, and lung. ANGPTL7 mRNA levels were normalized with GAPDH. All the qPCR experiments were repeated at least twice to obtain consistent results. Data are mean ± SEM, oneway ANOVA P=0.0001, Tukey’s posthoc test, P<0.05, n=3. (C) A boxplot shows a high expression score of the ANGPTL7 gene in breast cancer patients of TCGA-BRCA cohort with primary tumors (n=797) vs. patients with metastases (n=64). Student's t-test, P=0.05. (D) Boxplots show a high ANGPTL7 expression score of primary tumors vs. metastasis to bone, LN, or the lung in GSE110590 cohort of 16 patients (25), boxplots analyses, Kruskal-Wallis test (non-parametric method), P<0.05; Tukey’s t-test with medians and
inter-quartile ranges, $P<0.05$, as a significant difference. (E) Patients with a high ANGPTL7 expression score and association of survival in three breast cancer cohorts. Overall survival (OS) in METABRIC, TCGA-BRCA, and SCAN-B cohorts (left) or progression-free survival (PFI) and disease-specific (DSS) survival (right) in METABRIC and TCGA-BRCA cohorts along with hazard ratios (HR) and their 95% confidence intervals (CI) and $P$-values are shown. Note: Only SCAN-B OS data for a high expression score of ANGPTL7 is significant ($HR=0.65$, $P=0.0001$).

**Figure 3. MMP3 is downregulated at the metastatic sites vs. primary tumor and positively impacts breast cancer patients' survival.** MMP3 mRNA levels were determined by qPCR analysis in the primary tumors, metastatic lesions (A), and corresponding control tissues (B) of mice, as mentioned in Fig. 2A. MMP3 mRNA levels were normalized with GAPDH. Data are mean ± SEM, oneway ANOVA $P=0.0001$, Tukey's posthoc test, $P<0.05$, $n=3$. (C) Boxplots show a high MMP3 expression score of primary tumors vs. metastasis to bone, LN, or the lung. Boxplots analyses, Kruskal-Wallis test (non-parametric method), $P<0.05$; Tukey’s t-test with medians and inter-quartile ranges, $P<0.05$, as a significant difference. (D) Patients with a high MMP3 expression score and association of survival in three breast cancer cohorts. Overall survival (OS) in METABRIC, TCGA-BRCA, and SCAN-B cohorts (upper) or progression-free survival (PFI) and disease-specific (DSS) survival in METABRIC and TCGA-BRCA cohorts along with HR, CI, and $P$ values are shown (lower). Note: METABRIC and SCAN-B OS data for a high expression score of MMP3 is significant ($HR<1$, $p=0.0001$).

**Figure 4. LCN2 is upregulated at the metastatic sites (spine and bone) vs. primary tumor and an independent prognostic factor for breast cancer patients.** (A, B) qPCR analyses were performed to validate RNA-seq data between 4T1 vs. 4T1.2 cell lines using primary tumors. LCN2 mRNA normalized levels were determined in 4T1.2 primary tumors vs. metastatic lesions in the spine, bone, and lung (A). LCN2 mRNA normalized levels were also determined in the control tissues isolated from the spine, bone, and lung (B). mRNA levels were normalized with GAPDH, as described before. Data are mean ± SEM, oneway ANOVA $P<0.0001$, Tukey’s posthoc test, $P<0.05$, $n=3$. (C) Boxplots show a high LCN2 expression score of primary tumors vs. metastasis to bone, LN, or the lung. Boxplots analyses, the Kruskal-Wallis test (non-parametric method),
P<0.05, suggested an altered expression of LCN2. (D) Patients with a high LCN2 expression score and association of survival of breast cancer patients for the METABRIC cohort. LCN2 high vs. low expression scores survival rates (PFI, OS, or DSS) for the patients of METABRIC cohort and HR, CI, and P values are shown (D). Boxplots of the LCN2 high expression score by immunohistochemistry (IHC) determined subtype in the METABRIC (E) and SCAN-B cohorts (F). Similar data were obtained from the TCGA-BRCA cohort (data not shown). All boxplots are of Tukey type, and boxes depict medians and inter-quartile ranges. One-way ANOVA was used to calculate the p-value. Boxplots of high expression of LCN2 score of tumors of different American Joint Committee on Cancer (AJCC) stages (G) and Nottingham pathological grades (H) are shown for the METABRIC cohort. Similar data obtained from the TCGA-BRCA cohort (data not shown). One-way ANOVA and Tukey's tests were used to calculate P values.

**Figure 5. S100A8 is upregulated at the metastatic sites (spine and bone) vs. primary tumor and an independent prognostic factor for breast cancer patients.** (A, B) qPCR analyses for S100A8 gene. (A) Elevated levels of S100A8 mRNA in the primary tumors of 4T1.2 vs. 4T1 tumors were confirmed using qPCR analysis. S100A8 mRNA levels were also measures at the distant metastatic lesions of the spine, bone, and lung (A). S100A8 mRNA levels were also measured in the corresponding normal tissues of the control mice (B). mRNA levels were normalized with GAPDH. Data are mean ± SEM, oneway ANOVA P<0.0006, Tukey's posthoc test, P<0.05, n=3. (C) Altered expression of the S100A8 mRNA was also observed in patients with primary tumors vs. metastatic tumors, as shown by boxplots. (D) Patients with a high S100A8 expression score and association of survival (PFI, OS, and DSS) in the breast cancer cohort. Patient survival (PFI, OS, and DSS) data based on the high expression of S100A8 in the METABRIC cohort of breast cancer patients and HR, CI, and P values are shown (D). Boxplots of the S100A8 high expression score by immunohistochemistry (IHC) determined subtype in the TCGA-BRCA (E), SCAN-B (F), and METABRIC cohorts (G). All boxplots are of Tukey type, and boxes depict medians and inter-quartile ranges. One-way ANOVA was used to calculate the p-value. Boxplots of high expression of S100A8 score of tumors of different AJCC stages (H) for
METABRIC cohort and Nottingham pathological grades are shown for the METABRIC (I) and SCAN-B cohorts (J). One-way ANOVA and Tukey's tests were used to calculate P values.

**Figure 6. Expression of ESM-1 is downregulated in the 4T1.2 metastatic lesions vs. primary tumors and an independent prognostic factor for aggressive breast cancer subtypes.** (A-B) ESM-1 gene expression levels were analyzed by qPCR in the primary tumors, metastatic lesions (A), and control mice tissues (B), as labeled. ESM-1 mRNA levels were normalized with the GAPGH. Data are mean ± SEM, one-way ANOVA P<0.0006, Tukey's posthoc test, P<0.05, n=3. (C) Boxplots data showed that ESM-1 mRNA levels were altered in primary tumors vs. metastatic tumors of breast cancer patients, Kruskal-Wallis test (non-parametric method), P<0.05. (D) Boxplots of the ESM-1 high expression score by immunohistochemistry (IHC) determined subtype in the SCAN-B (left), METABRIC (middle), and TCGA-BRCA cohorts (right). Boxplots of high expression of ESM-1 score of tumors of different AJCC stages (F) for METABRIC (left panel) and TCGA-BRCA cohorts (right panel) and (E) Nottingham pathological grades in METABRIC (upper panel) and SCAN-B cohorts (lower panel). One-way ANOVA and Tukey's tests were used to calculate P values.
| Cohorts          | Survival | P-Value   | HR       | Lower 95% CI of HR | Upper 95% CI of HR |
|------------------|----------|-----------|----------|--------------------|--------------------|
| TCGA-BRCA (1,091) | OS       | 0.2651561 | 1.241098 | 0.847944           | 1.81654            |
|                  | DSS      | 0.0260737 | 1.802839 | 1.081905           | 3.00417            |
|                  | PFI      | 0.0404124 | 1.507879 | 1.018059           | 2.23336            |
| SCAN-B (3,273)   | OS       | 0.0027642 | 1.393542 | 1.122641           | 1.72981            |
| METABRIC (1,904) | OS       | 0.0002073 | 1.313813 | 1.135676           | 1.51989            |
|                  | DSS      | 0.0000432 | 1.488734 | 1.229534           | 1.80257            |
|                  | PFI      | 0.0001991 | 1.421232 | 1.180513           | 1.71103            |
| Breast cancer Cohorts | Subtypes | Survival | P-Value   | HR       | Lower 95% CI of HR | Upper 95% CI of HR |
|----------------------|----------|----------|-----------|----------|--------------------|--------------------|
| METABRIC             | TNBC     | OS       | 0.0678599 | 1.405336 | 0.9650234          | 2.046551           |
|                      |          | DSS      | 0.0271319 | 1.671391 | 1.05759            | 2.64142            |
|                      |          | RFS      | 0.06816303| 1.531018 | 0.9659142          | 2.426733           |
| HER2+                | OS       | 0.2023138| 1.304144  | 0.8662699 | 1.963351           |                    |
|                      | DSS      | 0.4882853| 1.170697  | 0.7486584 | 1.830649           |                    |
|                      | RFS      | 0.4749246| 1.177973  | 0.7512301 | 1.847131           |                    |
| HR+/HER2-            | OS       | 0.0084293| 1.255485  | 1.058645 | 1.488925           |                    |
|                      | DSS      | 0.0060594| 1.392947  | 1.098825 | 1.765796           |                    |
|                      | RFS      | 0.0075160| 1.357008  | 1.084659 | 1.697742           |                    |
| TCGA-BRCA            | TNBC     | OS       | 0.5326945 | 1.329982 | 0.5391484          | 3.280826           |
|                      | DSS      | 0.3297474| 1.753202  | 0.5614073 | 5.47502            |                    |
|                      | PFI      | 0.690275 | 1.205813  | 0.476718 | 3.049989           |                    |
| HER2+                | OS       | 0.2602478| 1.738325  | 0.6522829 | 4.632614           |                    |
|                      | DSS      | 0.0343557| 6.291787  | 1.570462 | 25.20697           |                    |
|                      | PFI      | 0.0204855| 4.075747  | 1.366649 | 12.15506           |                    |
| HR+/HER2-            | OS       | 0.4214821| 1.284964  | 0.6960809 | 2.37204            |                    |
|                      | DSS      | 0.2860177| 1.617747  | 0.6729301 | 3.889119           |                    |
|                      | PFI      | 0.304155 | 1.384029  | 0.7437917 | 2.575366           |                    |
| SCAN-B               | TNBC     | OS       | 0.4688024 | 1.321067 | 0.6211149          | 2.809815           |
|                      | HER2+    | OS       | 0.0171431 | 2.245766 | 1.188923           | 4.242043           |
|                      | HR+/HER2-| OS       | 0.2057267 | 1.190957 | 0.9086451          | 1.560983           |
Figure 1
**Figure 2**

**Panel A**: ANGPTL7 mRNA expression levels across different tissues, with a significant ANOVA result (P=0.0001).

**Panel B**: ANGPTL7 mRNA expression levels across different tissues, showing a trend with a specific focus on Cnt spine, Cnt bone, and Cnt lung comparisons.

**Panel C**: Expression of ANGPTL7 and TCGA BRCA, with a P-value of 0.02.

**Panel D**: Comparison of ANGPTL7 expression across different tumor types (PRIMARY, Bone, LN, LUNG) with a P-value less than 0.05.

**Panel E**: Hazard ratios (HR) and 95% confidence intervals (CI) for ANGPTL7 OS, PFI, and DSS across different cohorts and survival periods, with significant results indicated (P<0.05).
Figure 3
Figure 5
Figure 6
Figures

Figure 1

Proangiogenic and pro-inflammatory secretory protein-encoding genes are upregulated in metastatic variant 4T1.2 vs. parental 4T1 cells. (A-B) RNA-Seq heat map for 50 upregulated and 50 downregulated genes in 4T1.2 (red) vs. 4T1 cells (blue). Genes indicated in bold red arrow revealed a fold change (>6 fold vs. 4T1) is the mediator protein-encoding genes are elevated in the 4T1.2 cells. (C-D) On day 17 of the inoculation of cancer cells in the implantation/resection model, IVIS live images of Balb/c mice bearing metastatic 4T1.2-Luc+ breast cancer cells at primary and metastatic sites, supine (C), and prone (D) positions are shown. Arrow indicates hind bone 4T1.2-Luc+ metastasis. Representative mouse images are shown (n=5). (E), MRI of 4T1.2-Luc+ spine bone metastasis in vivo. On the day of sacrifice (day 17) of 4T1.2-Luc+ metastatic cancer-bearing Balb/c mice, an MRI scan was performed for the mice, and representative images were shown. E, top left and right, two representative T2-weighted transverse slice images of Balb/c mice with 4T1.2-Luc+ metastatic cancer. The spinal cord was indicated with the arrow in both images. E, bottom left and right, two representative spinal cord images visualized by the T2-weighted MRI, sagittal projection. Arrows indicate spine bone metastasis. Data are representative of n=5 mice.
Figure 2

ANGPTL7 is downregulated at the metastatic sites vs. primary tumor and positively impacts breast cancer patients' survival. (A) ANGPTL7 mRNA levels, determined by qPCR analysis, were compared between primary tumors of 4T1 vs. 4T1.2 cells (n=3/group). Metastatic lesions of 4T1.2 cells at distant sites such as the spine, bone, and the lung, were used for ANGPTL7 mRNA levels using qPCR analysis (n=3). (B) ANGPTL7 expression levels were also estimated in the control mice (n=3) tissues of the spine, bone, and lung. ANGPTL7 mRNA levels were normalized with GAPDH. All the qPCR experiments were repeated at least twice to obtain consistent results. Data are mean ± SEM, oneway ANOVA P=0.0001, Tukey's posthoc test, P<0.05, n=3. (C) A boxplot shows a high expression score of the ANGPTL7 gene in breast cancer patients of TCGA-BRCA cohort with primary tumors (n=797) vs. patients with metastases (n=64). Student's t-test, P=0.05. (D) Boxplots show a high ANGPTL7 expression score of primary tumors vs. metastasis to bone, LN, or the lung in GSE110590 cohort of 16 patients (25), boxplots analyses, Kruskal-Wallis test (non-parametric method), P<0.05; Tukey’s t-test with medians and inter-quartile ranges, P<0.05, as a significant difference. (E) Patients with a high ANGPTL7 expression score and association of survival in three breast cancer cohorts. Overall survival (OS) in METABRIC, TCGA-BRCA, and SCAN-B cohorts (left) or progression-free survival (PFI) and disease-specific (DSS) survival (right) in METABRIC and TCGA-BRCA cohorts along with hazard ratios (HR) and their 95% confidence intervals (CI) and P-values are shown. Note: Only SCAN-B OS data for a high expression score of ANGPTL7 is significant (HR=0.65, P=0.0001).
MMP3 is downregulated at the metastatic sites vs. primary tumor and positively impacts breast cancer patients' survival. MMP3 mRNA levels were determined by qPCR analysis in the primary tumors, metastatic lesions (A), and corresponding control tissues (B) of mice, as mentioned in Fig. 2A. MMP3 mRNA levels were normalized with GAPDH. Data are mean ± SEM, one-way ANOVA P=0.0001, Tukey's posthoc test, P<0.05, n=3. (C) Boxplots show a high MMP3 expression score of primary tumors vs. metastasis to bone, LN, or the lung. Boxplots analyses, Kruskal-Wallis test (non-parametric method), P<0.05; Tukey's t-test with medians and inter-quartile ranges, P<0.05, as a significant difference. (D) Patients with a high MMP3 expression score and association of survival in three breast cancer cohorts. Overall survival (OS) in METABRIC, TCGA-BRCA, and SCAN-B cohorts (upper) or progression-free survival (PFI) and disease-specific (DSS) survival in METABRIC and TCGA-BRCA cohorts along with HR, CI, and P values are shown (lower). Note: METABRIC and SCAN-B OS data for a high expression score of MMP3 is significant (HR<1, p=0.0001).
LCN2 is upregulated at the metastatic sites (spine and bone) vs. primary tumor and an independent prognostic factor for breast cancer patients. (A, B) qPCR analyses were performed to validate RNA-seq data between 4T1 vs. 4T1.2 cell lines using primary tumors. LCN2 mRNA normalized levels were determined in 4T1.2 primary tumors vs. metastatic lesions in the spine, bone, and lung (A). LCN2 mRNA normalized levels were also determined in the control tissues isolated from the spine, bone, and lung (B). mRNA levels were normalized with GAPDH, as described before. Data are mean ± SEM, oneway ANOVA P<0.0001, Tukey's posthoc test, P<0.05, n=3. (C) Boxplots show a high LCN2 expression score of primary tumors vs. metastasis to bone, LN, or the lung. Boxplots analyses, the Kruskal-Wallis test (non-parametric method), P<0.05, suggested an altered expression of LCN2. (D) Patients with a high LCN2 expression score and association of survival of breast cancer patients for the METABRIC cohort. LCN2 high vs. low expression scores survival rates (PFI, OS, or DSS) for the patients of METABRIC cohort and HR, CI, and P values are shown (D). Boxplots of the LCN2 high expression score by immunohistochemistry (IHC) determined subtype in the METABRIC (E) and SCAN-B cohorts (F). Similar data were obtained from the TCGA-BRCA cohort (data not shown). All boxplots are of Tukey type, and boxes depict medians and interquartile ranges. One-way ANOVA was used to calculate the p-value. Boxplots of high expression of LCN2 score of tumors of different American Joint Committee on Cancer (AJCC) stages (G) and Nottingham pathological grades (H) are shown for the METABRIC cohort. Similar data obtained from the TCGA-BRCA cohort (data not shown). One-way ANOVA and Tukey's tests were used to calculate P values.
S100A8 is upregulated at the metastatic sites (spine and bone) vs. primary tumor and an independent prognostic factor for breast cancer patients. (A, B) qPCR analyses for S100A8 gene. (A) Elevated levels of S100A8 mRNA in the primary tumors of 4T1.2 vs. 4T1 tumors were confirmed using qPCR analysis. S100A8 mRNA levels were also measured at the distant metastatic lesions of the spine, bone, and lung (A). S100A8 mRNA levels were also measured in the corresponding normal tissues of the control mice (B). mRNA levels were normalized with GAPDH. Data are mean ± SEM, oneway ANOVA P<0.0006, Tukey's posthoc test, P<0.05, n=3. (C) Altered expression of the S100A8 mRNA was also observed in patients with primary tumors vs. metastatic tumors, as shown by boxplots. (D) Patients with a high S100A8 expression score and association of survival (PFI, OS, and DSS) in the breast cancer cohort. Patient survival (PFI, OS, and DSS) data based on the high expression of S100A8 in the METABRIC cohort of breast cancer patients and HR, CI, and P values are shown (D). Boxplots of the S100A8 high expression score by immunohistochemistry (IHC) determined subtype in the TCGA-BRCA (E), SCAN-B (F), and METABRIC cohorts (G). All boxplots are of Tukey type, and boxes depict medians and inter-quartile ranges. One-way ANOVA was used to calculate the pvalue. Boxplots of high expression of S100A8 score of tumors of different AJCC stages (H) for METABRIC cohort and Nottingham pathological grades are shown for the METABRIC (I) and SCAN-B cohorts (J). One-way ANOVA and Tukey's tests were used to calculate P values.
Expression of ESM-1 is downregulated in the 4T1.2 metastatic lesions vs. primary tumors and an independent prognostic factor for aggressive breast cancer subtypes. (A-B) ESM-1 gene expression levels were analyzed by qPCR in the primary tumors, metastatic lesions (A), and control mice tissues (B), as labeled. ESM-1 mRNA levels were normalized with the GAPGH. Data are mean ± SEM, oneway ANOVA P<0.0006, Tukey's posthoc test, P<0.05, n=3. (C) Boxplots data showed that ESM-1 mRNA levels were altered in primary tumors vs. metastatic tumors of breast cancer patients, Kruskal-Wallis test (non-parametric method), P<0.05. (D) Boxplots of the ESM-1 high expression score by immunohistochemistry (IHC) determined subtype in the SCAN-B (left), METABRIC (middle), and TCGA-BRCA cohorts (right). Boxplots of high expression of ESM-1 score of tumors of different AJCC stages (F) for METABRIC (left panel) and TCGA-BRCA cohorts (right panel) and (E) Nottingham pathological grades in METABRIC (upper panel) and SCAN-B cohorts (lower panel). One-way ANOVA and Tukey's tests were used to calculate P values.