UTILIZING PROTEOMICS TO IDENTIFY EXTRACELLULAR MATRIX CHANGES DURING BREAST CANCER METASTASIS

by

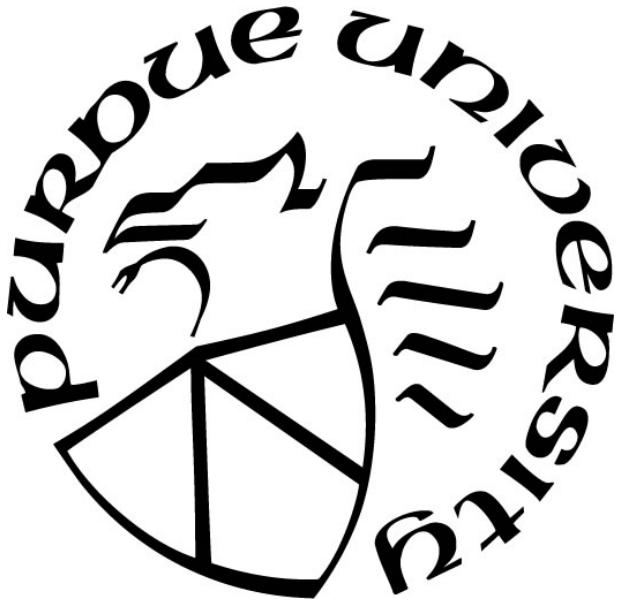
Elly Y Lambert

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Biomedical Engineering



Weldon School of Biomedical Engineering

West Lafayette, Indiana

August 2020

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Luis Solorio, Chair

Weldon School of Biomedical Engineering

Dr. Craig Goergen

Weldon School of Biomedical Engineering

Dr. Michael K Wendt

Department of Medicinal Chemistry and Molecular Pharmacology

Approved by:

Dr. George Wodicka

ACKNOWLEDGMENTS

I want to acknowledge all members of the Tumor MicroEnvironment and Therapeutics (TMET) Lab for all of their constant support during my time at Purdue. I have learned countless lessons professionally and personally being a member of this lab. I would like to thank Dr. Luis Solorio for all of his guidance and mentorship. The opportunity he provided me to pursue a higher education is invaluable and I could not have asked for a better mentor.

Many people have helped with this project, and their expertise and mentorship are greatly valued. I would like to acknowledge Kathryn Jacobson, whom this project could not have been completed without. Her assistance with mass spectrometry preparation and analysis is greatly appreciated. I would like to thank Dr. Sarah Calve for her expertise and advice about mass spectrometry. I, also, want to thank Dr. Kevin Buno for his mentorship and help with immunohistochemical staining. We appreciate the Purdue Proteomics Core for assistance with tandem mass spectrometry analysis.

Lastly, I want to thank my family and friends. Their love and support have been immeasurable during both my undergraduate and graduate career at Purdue.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

HL – Healthy Lungs; healthy lungs, no inoculation to mice

EL – Extracellular Vesicle Lungs; lungs primed with 4T1 extracellular vesicles (inoculation of EVs every other day for 21 days)

DL – Diseased Lungs; lungs from mice inoculated with 4T1 cells with metastatic lesions removed

ML – Metastasized Lungs; lungs from mice inoculated with 4T1 cells with metastatic lesions intact

MT – Mets; metastasized tumors in the lungs

PT – Primary Tumor; primary tumor at sight of inoculation

RT – Relapse Tumor; recurrent tumor at site of excised primary tumor

MG – Mammary Glands; mammary glands from healthy mice

FC – Fold Change

ABSTRACT

Breast cancer is one of the most commonly diagnosed cancers in women, with 1 in 8 women diagnosed during her lifetime. Distant metastatic breast cancer accounts for a majority of deaths in breast cancer patients. Changes in both the architecture and the biochemical composition of the extracellular matrix (ECM) occur during metastatic dissemination at both the primary tumor and the early metastatic niche. These changes play a significant role in the cell fate, and can alter proliferation, migration, and quiescence of cancer cells. This study utilizes tandem mass spectrometry to study ECM protein changes, specifically in the lungs, using an immune competent murine model of metastatic breast cancer. Liquid chromatography-tandem mass spectrometry was used to identify and quantify key ECM proteins in the primary tumors, lungs, and metastatic tumors during cancer progression. Fibronectin (FN) was upregulated in the primary tumor, suggestive of a more invasive mesenchymal-like cell. However, FN was decreased in abundance in metastatic tumors, which is favorable for a more epithelial phenotype, prompting tumor growth. The diseased lungs appear to have highly collagenous proteins, suggesting an increased stiffness in the matrix. This increase in stiffness would reduce physiologically induced strains, and potentially facilitate growth of metastatic lesions in the lungs. Characterization of the changes in the ECM during cancer progression will aid in development of future therapies as well as guide the design of relevant *in vitro* models, ultimately enhancing the knowledge of this phenomenon.

1. INTRODUCTION

Breast cancer remains one of the most commonly diagnosed cancers among women, with metastasis accounting for the majority of patient mortality [1–3]. During tumor progression, the extracellular matrix undergoes compositional and structural changes at both the primary tumor and during the establishment of the metastatic niche [3][4]. The ECM contributes to metastasis by providing a source of long-term stable signaling that facilitates processes such as invasion and tumor growth [5]. The dynamic relationship between the ECM and cancer cells may facilitate disease progression [6]. However, due to the poor solubility of ECM proteins and the extensive matrix remodeling that occurs during progression, knowledge about the matrix dynamics during this process has been limited.

The tumor microenvironment consists of the ECM, proteins and noncancerous cells, favorable to tumor growth [7]. Cells from the primary tumor disseminate to distant metastatic sites that also have favorable conditions to metastatic growth. This niche formation occurs and allows for cell seeding to occur [8]. Previous research suggests that extracellular vesicles (EVs) prime the metastatic site to facilitate niche formation [9][10]. EVs contain proteins and DNA packages that can modify the ECM and surrounding tissues into an environment more favorable for tumor growth. These changes have been demonstrated to recruit myeloid cells, which are hypothesized to alter the ECM and promote tumor growth [11]. A full description of the changes that occur in the ECM composition during niche formation and metastatic progression have not been investigated.

Recently, tandem mass spectrometry (LC-MS/MS) has been used as a method to characterize the ECM of different tissues, including cancer models [12]. Through ECM enrichment and digestion processes, LC- MS/MS becomes an applicable tool for characterizing protein abundance in whole organs. Homogenized tissue samples can be prepared for mass spectrometry analysis and then the ECM proteins found in the processed tissues can be identified. Using these tools the entire ECM proteome, referred to as the matrisome, can be identified [13][14]. This method of identifying the matrisome of whole organs has been translated to analyze and identify changes in the ECM composition that occur between healthy and diseased tissues [12].

Utilization of LC-MS/MS analysis to study metastatic breast cancer has revealed that the matrisome of metastatic tissues is unique to the organ that has been invaded [15]. Specifically, the

identification of COL4A4 and lamnin-121 as a markers of lung metastasis. The upregulation of SERPINB1 was identified in lung metastasis compared to healthy tissues, however knockdown of SERPINB1 only slightly, but not significantly, reduced lung metastasis [15]. These recent advances were performed using an immunocompromised murine model with human derived cancer cells, which cannot evaluate the effect that the immune system has on the ECM dynamics. However, it has been demonstrated that myeloid cells play a significant role in the ECM changes that occur in metastatic tissues during the formation of the early metastatic niche [10][16]. Furthermore, it is still unclear what changes occur in the ECM at both the primary tumor and the metastatic niche during the early stages of metastasis and metastatic niche formation [17].

Herein, we used an immunocompetent murine model to identify temporal changes in the matrisome that occur during tumor progression. Tissues were harvested from 4T1 tumor bearing Balb/c mice. The tissues were then analyzed using LC-MS/MS. Utilization of the 4T1 cell line in Balb/c mice produce tumors that consistently progress and disseminate to the lungs, making it an ideal model to study metastasis [18][19]. To evaluate the dynamic changes in the matrisome, healthy mammary glands were compared with tumors formed in the primary site as well as relapsed tumors that occur in the same location after surgical resection. Additionally, healthy lungs were compared with pre-metastatic (EV Primed) and late metastatic lungs. Interestingly, S100A8/A9 were identified in all diseased tissues, but were more abundant in tissues at later disease progression. This is indicative of an immune mediated cancer cell invasion of tissues.

2. MATERIALS AND METHODS

2.1 In Vivo Assays

All *in vivo* studies were conducted following an IACUC approved protocol from Purdue University. Luciferase expressing 4T1 cells were resuspended in 100 µL phosphate buffered saline (PBS) and injected ($5x10^4$ cells/mouse) into the second mammary fat pad of female 4-6 week old Balb/c mice (Jackson Labs, Bar Harbor, ME). Excision of primary tumors was performed after the tumor reached a volume of approximately 300-500 mm^3 to facilitate metastatic outgrowth. Bioluminescent imaging using the Advanced Molecular Imager (AMI) (Spectral Instruments, Tucson, AZ) was used to track primary tumor dissemination into the lungs. Upon metastatic growth into the lungs, mice were euthanized and whole lungs along with any relapse tumors were removed. Lungs and mammary glands were harvested from healthy mice as controls. To determine the effect the EVs have on the matrisome of the early metastatic niche, 50 µg of EVs were injected intraperitoneally BALB/c mice. This EV priming occurred every other day for 21 d. At day 21, mice were euthanized, and lung tissues were removed.

2.2 Extracellular Vesicle Enrichment

Extracellular vesicles (EV) were enriched via centrifugation using our previously described protocols [20]. Briefly, EVs were isolated from 4T1 cell serum starved for 72 hours in serum free conditioned DMEM media. The conditioned media was centrifuged at 300 g for 10 min to ensure removal of any living cells in the media. The supernatant was collected, and then centrifuged at 2000 g for an additional 10 min to remove cell debris from the media. The supernatant was conserved and filtered using a 0.22 µM pore size vacuum filter (Millipore) and then transferred to 3KD MWCO Amicon ultra-15 centrifugal filter tubes and centrifuged at 3220 g for 1 h. The filtered media was centrifuged at 3220 g for an additional hour and condensed the sample to approximately 1mL. The condensed sample was then ultra-centrifuged at 100,000 g for 2 h. The supernatant was discarded, and the pellet was resuspended and washed in 1X phosphate buffered saline (PBS) using ultracentrifugation at 100,000 g for 2 h. The supernatant was discarded again, and the EV enriched pellet was resuspended in 1X PBS for storage in a -80° C freezer until needed.

2.3 LC-MS/MS

Tissues were prepared as previously described [21][22]. Briefly, harvested tissues samples were homogenized using the TissueRuptor (Qiagen). Then, the homogenized tissues were fractionated using sequential extractions performed with buffers of increasing stringency to remove cellular fractions which left behind an ECM-rich pellet (Milipore ECM Extraction Kit).

The pellets were then resuspended with 8M Urea and 100mM ammonium bicarbonate. The suspension was then reduced with 10 mM dithiothreitol (DTT) for 2 h at 37°C under constant agitation. Samples were then brought to room temperature and alkylated using 25mM iodoacetamide for 30 min in the dark. Next, samples were diluted 1:4 with ammonium bicarbonate. Proteolytic digestion was completed using LysC, and followed by 2X trypsin (twice)[21]. The enzymes were then inactivated with 0.1% trifluoracetic acid (TFA).

Peptides were then washed using Pierce Detergent Removal Spin Columns (ThermoFisher Scientific). Next, samples were desalted using C18 MicroSpin columns (The Nest Group) conditioned with 100µL 100% acetonitrile (ACN) and equilibrated with 100µL 0.1% TFA. Samples were then added to the C18 columns and washed with 100µL of 0.1% TFA and eluted with 100µL 80% CAN and 25mM formic acid. The peptides were transferred to 0.5mL tubes and dried using vacuum centrifugation (CentriVap Benchtop Vacuum Concentrator) at 45°C. Dried peptides were resuspended in 10µL 3% ACN and 0.1% formic acid (FA). The peptide concentration was measured with the Pierce Quantitative Colorimetric Peptide Assay (ThermoFisher Scientific). The most concentrated sample was brought to 0.2µg/mL with 3% ACN and 1% FA. All other sample concentrations were normalized with 3% ACN and 1% FA. Peptides were analyzed using the Dionex UltiMate 2000 RSLC Nano System coupled with the Q exactive HF hybrid quadrupole-orbitrap mass spectrometer (QE HF; ThermoFisher Schientific) as previously reported [21] [22].

2.4 LC-MS/MS Statistical Analysis

Raw data files were analyzed using MaxQuant (v.1.6.1.0) as previously described [21]. ECM insoluble samples were analyzed with label free quantification (LFQ). Row z-score was calculated and utilized to compare abundance of proteins between tissue types. Data was analyzed with Microsoft Excel and Prism. LFQ intensities were used to complete statistical analysis and

visualized with volcano plots. Volcano plot comparisons were considered differentially expressed if the p-value was statistically significant (95% significance) and had a fold change of 2 or more.

3. RESULTS

3.1 Proteomics helps to identify ECM composition of diseased tissues

In order to identify ECM changes during disease progression, the highly metastatic 4T1 triple-negative breast cancer (TNBC) transplant murine model was utilized. To determine the effect of disease progression on the metastatic matrisome, tissues were obtained from healthy, pre-metastatic, and metastatic diseased states. Control tissues collected were healthy lungs and mammary glands. Primary tumors were then excised to help facilitate metastatic outgrowth. Pre-metastatic lungs were collected after being primed with EVs isolated from 4T1 cells to promote metastatic niche formation [20]. Metastasized lungs were then collected once bioluminescent imaging confirmed metastatic growth. Additionally, any relapse tumors that formed at the primary site were also collected. Sample names are outlined in **Table 1,** and are representative of tissue, tumor, and inoculation type. Label-free analysis identified 136 proteins across all tissue types evaluated. Each tissue type had approximately 80% of the proteins identified as matrisome or matrisome associated proteins (**Figure 2A**). Tumors and mammary glands had at least 80% of the matrisome classified as collagens, while lung tissues consisted of roughly 50% collagen (**Figure 2B**). High abundance of collagens is indicative of stiffer matrix properties as we have previously reported [23]. Within the lungs, the remainder of the identified proteins consisted of ECM glycoproteins. The full break down of matrisome protein distribution can be found in **Table 2**. Variations in protein classification identified show that ECM composition changes between healthy and diseased tissues during cancer progression.

Z-scores were calculated from normalized intensities to compare protein abundance across tissue types (**Figure 3**). This analysis allows for comparison of relative abundance during disease progression. The heat map (**Figure 3**) allows distinction of proteins unique to each tissue type, as well as identification of proteins common among tissues. Additionally, **Table 3** shows unique proteins to each tissue type.

Evaluating tumor progression at the primary location (mammary glands, primary tumors, and relapse tumors), many collagens that were identified had an increase in abundance in the relapse tumors, specifically COL1A1 and COL1A2. Some ECM Glycoproteins, FN1, EMILIN2, POSTN, and TNC, increased in abundance at the primary tumors and relapse tumor. Proteoglycans

did not readily increase in abundance in tissues at the primary location. However, the proteoglycans SERPINA3K, PLG, ITIH2, and F13A1 were uniquely identified only in the primary and relapse tumors, but not the mammary gland.

Evaluating the healthy lungs, EV lungs, and metastasized lungs revealed increased and decreased abundance of certain proteins. In particular the ECM glycoproteins were more highly expressed in lung tissues compared to tumors or mammary glands (**Figure 3**). Basement membrane proteins, NID1 and laminin isoforms are highly expressed in all three lung tissues compared to the tumors or mammary glands. Interestingly, these proteins also appear to be increased in abundance in premetastatic and metastatic tissues. COL4A1 and COL4A2 were highly expressed in metastasized lung tissues, as expected [15]. S100A8 was not identified in healthy lungs but was identified in EV lungs and metastasized lungs. S100A9 was increased in abundance in the EV (row-z increase of .26) and Metastasized lungs (row-z increase of 1.79) compared to healthy lungs. FN1 levels decreased in EV (row-z decrease of 1.03) and Metastasized lungs (row-z decrease of 1.28).

Significant differences were observed in the protein levels between the different tumor tissues to include primary, relapse and mets. The primary tumors had increased abundance of FN1. While both the relapse and mets had elevated levels of collagens. However, many collagens were unique to the mets, including COL2A1, COL6A5, COL9A1, COL9A2, COL9A3, and COL10A1 (**Table 3**). Additionally, TNC, a protein thought to be essential for metastatic outgrowth, was identified only in tumors (mets, primary tumors, and relapse tumors) [24]. FN1 was identified in all tissues at varying abundance. It appeared to be increased in expression in the primary and relapse tumors. However, FN1 appeared to be downregulated in metastasized lung tissue. Interestingly, S100A9 increased in row z-score intensity as a function of disease progression, being expressed higher in EV primed and metastasized lungs. Interestingly, S100A8 was only identified in diseased tissues, increasing row z-score at later metastasized time points. Additionally, several proteins involved in cartilage formation were identified or more abundant in metastatic lesions and tissues. Specifically, within the mets, a unique signature of proteins involved in cartilage formation were identified: COL2A1, COL6A5, COL10A1, COL9A1, COL9A2, COL9A3, COMP, MATN1, MGP, LTBP3, ACAN, CHAD, HAPLN1, PRELP, TGM1, TGM3 (**Table 3**).

Because an immunocompetent mouse was used, it was evident that the immune system would play an important role in ECM composition. This is consistent with the identification of

S100 proteins which are prevalent in body immune responses [25][26]. S100A8 was found only in diseased tissues including the EV Lungs, and S100A9 was found in all diseased tissues and low abundance in the healthy lungs (**Figure 3**).

3.2 ECM protein content changes at the primary tumor

Identifying distinctions in ECM composition of tumors at different stages of disease progression is important in understanding the factors that promote tumor outgrowth. There were 35 common proteins identified between native tissue and tumors (both primary and relapse) (**Figure 4A**). The mammary gland had three additional (38 total) identified proteins (ANXA5, HSPG2, and S100A10) in common with the primary tumors. Additionally, there were three identified proteins in common with the relapse tumors (COL14A1, COL6A6, and TNXB; **Figure 4A**). The primary tumor had 30 proteins that were distinct from the mammary gland and 38 proteins that were common between the two tissue types. Among the 38 common proteins, the primary tumor had significantly elevated levels of FN1 ($\log_2$(FC) of 3.02 and -$\log_{10}$(p-value) of 4.83) and TGM2 ($\log_2$(FC) of 1.29 and -$\log_{10}$(p-value) of 3.28) compared to the mammary gland (**Figure 4B**). These ECM proteins have been shown in our previous research to play significant roles in cancer metastasis [23] [16]. Comparing the mammary gland with the relapse tumor, there were 29 proteins unique to the relapse tumor and 38 common proteins between the two tissues. The relapsed tumors also had significantly elevated abundance of FN1 ($\log_2$(FC) of 1.99 and -$\log_{10}$(p-value) of 1.79) and slight increased abundance of TGM2 ($\log_2$(FC) of .79 and -$\log_{10}$(p-value) of 2.47) relative to the mammary gland. (**Figure 4C**). The mammary glands had significantly elevated abundance ($\log_2$(FC) greater than 1 and -$\log_{10}$(p-value) greater than 1.3) of COL4A1 and COL4A2 than both the primary and relapse tumors (**Figure 4B and 4C**).

The primary tumors and relapse tumors had an additional 24 (59 total) common proteins identified. Both tumors had similar abundance of TNC, S100A8 and S100A9 identified, which were not identified in the mammary glands. Even with all these similarities, the tumors still had unique proteins and protein abundances that distinguished the tumor types from each other. The primary tumor had statistically more COL4A1 and COL4A2, $\log_{10}$p-values of 1.81 and 1.63 respectively with a $\log_2$ FC of greater than 1, than the relapse tumor (**Figure 4D**). Relapse tumors contained A2M, COL14A1, COL6A6, and TNXB, which were not identified in the primary tumor. Proteins identified specifically to primary tumors (in this comparison) included ANXA5,

EMILIN1, HSPG2, LAMC2, LAMC2, S100A10. These differences were in addition to unique proteins found only in the primary or relapse tumor included in **table 3**.

3.3 The ECM content in lungs changes during metastasis

The ECM composition of the lungs was analyzed using healthy, EV primed lungs, and metastatic lungs. These samples allow comparison of ECM composition of healthy, pre-metastatic (EV lungs), and fully metastasized (Metastasized lungs) lung tissues. Similarities between healthy, EV, and metastasized lungs revealed 62 proteins in common (**Figure 5A**). In addition to common proteins in all lung tissues, the healthy and EV lungs had an additional 4 proteins (DCN, RPTN, SERPINA3K, and SFTPA1) common between them. The EV lungs and metastasized lungs had an additional 4 proteins (ADAMTS17, SERPINF2, and THBS1) common that were not found in the healthy lungs. One clear distinction is that S100A8 was found in both EV and metastasized lungs, but not in the healthy lungs. Additionally, the amount of S100A9 was significantly more abundant in the EV lungs by a $\log_2$(FC) of 1.93 and $-\log_{10}$(p-value) of 1.63 (**Figure 5C**) and in the metastasized lungs $\log_2$(FC) of 4.14 and $-\log_{10}$(p-value) of 2.94 (**Figure 5D**) than in the healthy lungs. S100A9 was more abundant in the metastasized lungs than the EV lungs by $\log_2$(FC) of 2.22 and $-\log_{10}$(p-value) of 2.23 (**Figure 5E**). This increasing abundance of S100A8 and S100A9 in further metastasized tissues is indicative of an ECM temporal change as disease progresses. This comparison correlates with previous statements that S100A9 was significantly abundant in diseased tissues only.

3.4 Metastatic tissues have distinct ECM proteins

There were some proteins found in all tissues, but with unique trends, including the identification of TGM2, and FN1. TGM2 and FN1 are known to contribute to breast cancer metastasis [20], [23]. However, FN1 appears to decrease in pre-metastatic and metastatic lung tissues and TGM2 appeared most abundant in the EV lungs and diseased lungs (**Figure 3**). Additionally, S100A8/A9 were identified primarily in diseased tissues (all tissue samples except controls), with seemingly increased abundance in more progressed diseased tissues. The unique protein signatures suggest that the ECM of pre-metastatic and metastatic tissues are distinct, contributing to outgrowth of metastatic lesions.

Similarities in tumor composition at the primary and distant metastatic site suggest certain proteins are necessary for cell survival in both primary and metastasized tumors (**Figure 6A**). Although 54 proteins were common between the mets and primary tumors, the relative abundance in these tissues were different. TNC and FN1 were found in the primary tumor and mets, but more highly abundant in the primary tumor (**Figure 6C**). This suggests that these protein changes are necessary to facilitate tumor outgrowth and then invasion and growth in metastatic tissues. Some proteins found in the primary tumor, but not the mets, were also found in the mammary gland (COL5A3, LGALS1, and S100A10). Also, some proteins in the mets were also found in the healthy lungs (AGRN, COL4A3, COL4A4, Col4A5, Col6A6, EFEMP1, FBLN5, LAMA2, LAMA3, LAMA4, LAMB2, LOX, LOXL1, RPTN, SERPINB12, SFTPA1, TNXB, and VWF). This suggests some of the differences in the primary tumors and mets are related to the surrounding mammary gland or lung tissue. There are proteins unique to both the mets and primary tumors (**Table 3**) as well as different abundance of common proteins. These differences in ECM content between primary tumors and metastatic lesions indicate that compositional changes are necessary to facilitate metastasis.

Comparing the metastatic tissues – metastasized lungs, diseased lungs, and mets – allows investigation of proteins unique to either the tumor or stroma. To reiterate, the metastasized lungs (ML) had metastatic lesions intact, the diseased lungs (DL) had them removed, and the mets are the lesions that grew in the lungs that were removed. There were 59 common proteins among all three of these tissues (**Figure 6B**). Additionally, 5 proteins (BMPER, EGFL7, MEGF6, TINAGL1, and WISP2) identified were common between just the diseased lungs and metastasized lungs, suggesting these are unique to either the stroma or lung tissue. Comparing common proteins between the diseased and metastasized lungs, they were not statistically different by a FC of 2 or more in abundance (**Figure 6D**). There were 5 additional proteins (ANXA5, COL12A1, S100A11, SERPINA3K, and SFTPA1) common between the diseased lungs and mets but not identified in the metastasized lungs. Although most common proteins between the diseased lungs and mets were similar in abundance or were not different by a large FC, TGM2 was more abundant in the diseased lungs ($\log_2$(FC) of 1.03 and $\log_{10}$(p-value) of 1.89), and PA1EFE was more abundant in the mets ($\log_2$(FC) of 1.41 and $\log_{10}$(p-value) of 1.54) (**Figure 6E**). Looking at the metastasized lungs and mets, there were no additional common proteins between these two tissues that were not also common with the diseased lungs. LAMB1 ($\log_2$(FC) of 2.24 and $\log_{10}$(p-value) of 1.38) and

COL45 (log_2(FC) of 2.71 and log_{10}(p-value) of 1.38) were more highly expressed in the metastasized lungs than the mets, indicative of these proteins being more highly expressed in tumor stroma to facilitate metastatic outgrowth (**Figure 6F**). ANXA1 (log_2(FC) of 3.70 and log_{10}(p-value) of 1.50) was more highly expressed in the mets than metastasized lungs. These distinctions indicate that tumor and stroma have different ECM composition, however this model was not adequate in completely separating tumor from stroma.

4. DISCUSSION

This study allowed quantitative analysis of ECM proteins during different stages of breast cancer progression using the 4T1 murine breast cancer allograft model. LC-MS/MS analysis was used to identify and quantify ECM protein composition involved in the progression of metastatic breast cancer. We were able to identify changes in the matrisome of tissues and tumors during disease progression. The study is consistent with previous research that TGM2 and FN1 play important roles throughout breast cancer metastasis [20] [23]. However, S100 proteins were clearly abundant in this model, specifically these proteins were significantly elevated in the pre-metastatic and metastatic tissues, prompting investigation into function of these proteins during metastasis.

The heatmap encompassing all data demonstrates that there are distinctions between tissue groups, such as high expression of ECM glycoproteins in the lung tissues compared to others, or COL1A1 and COL1A2 being higher in abundance in tumors than other tissues (**Figure S1**). Lung tissues had increased abundance of collagen expression, especially in metastasized lungs. This may be indicative of a stiffer matrix, favorable for metastatic growth [27]. Additionally, many of the proteins that were abundant in mets or relapse tumors were the collagens or basement membrane proteins, which may facilitate metastatic outgrowth in the lungs. These distinctions prompt investigation of protein effect on metastatic growth. Interestingly, cartilage forming proteins (including COL10A1, ACAN, CHAD) were found primarily in the mets, again suggesting a stiffer environment is needed for metastatic growth.

An immunocompetent model was used to gain insight into the role of the immune response on ECM composition. S100A8/A9 is known to be involved in the innate immune response [28]. S100A8 was identified only in diseased tissues and all tumor types in this study. Additionally, S100A8/A9 abundance increased as the disease progressed such that they were more abundant in EV primed lungs than healthy lungs, and the most abundant in fully metastasized lungs. Previous metastatic breast cancer work using a xenograft model also showed an increased abundance of COL4 as a marker of lung metastasis. However, that previous work lacked the effect of an immune system on ECM dynamics. Further, these studies did not identify S100A8/A9 proteins in the insoluble fraction, suggesting that these proteins occur as a result of the immune response to cancer cell invasion [15].

Previous work suggests that S100A8/A9 may also play a role in metastasis[28][29]. However, different concentrations of S100A8/A9 may contribute to different functions in disease progression. Low concentration levels of S100A8/A9 act as a tumor promoter, but high concentrations act as a tumor suppressor [28][26]. S100A9 may play a role in creation of a pre-metastatic niche for disseminated cells to invade and grow in [30]. These findings warrant further investigation into the role of the immune system on ECM protein expression of S100A8/A9 in cancer specifically.

Comparing the overall ECM compositions during breast cancer metastasis allows a better understanding of changes that facilitate cancer progression. By identifying the ECM composition at different stages of metastasis, we can identify proteins that may facilitate and promote metastatic growth. Future work identifying the origin of ECM proteins and their function in cancer progression would help identify biomarkers to aid in future diagnostics and therapeutics.

5. CONCLUSIONS AND FUTURE WORK

We utilized mass spectrometry to compare the ECM of healthy, pre-metastatic and metastatic tissues. We observed distinctions in content and relative abundance between tissue types, implying that the ECM of tissues changes as metastasis progresses to facilitate cell invasion and growth. There was a more collagenous ECM observed in diseased lung tissues suggesting a stiffening of matrix components to facilitate metastatic growth. S100A8/A9 was more highly expressed in diseased tissues, potentially as an immune response to invading cancer cells, prompting investigation into these protein interactions. These observations imply a dynamically changing ECM favorable for cancer progression.

Future studies, to solidify findings, include immunohistochemical staining and investigating the ECM composition in the lungs at early metastasis. Immunohistochemical staining will allow confirmation of mass spectrometry observations and localization of cells. Previous research suggests an increase in FN1 in the lungs occurs during early metastasis [16]. Looking at an earlier metastatic timepoint would allow further insight into the dynamic changes of breast cancer. One limitation of this study is the inability to determine if the protein came from the cancer cells or resident cells. Further investigation into the protein origin and effects on the ECM would allow identification of biomarkers of cancer progression.

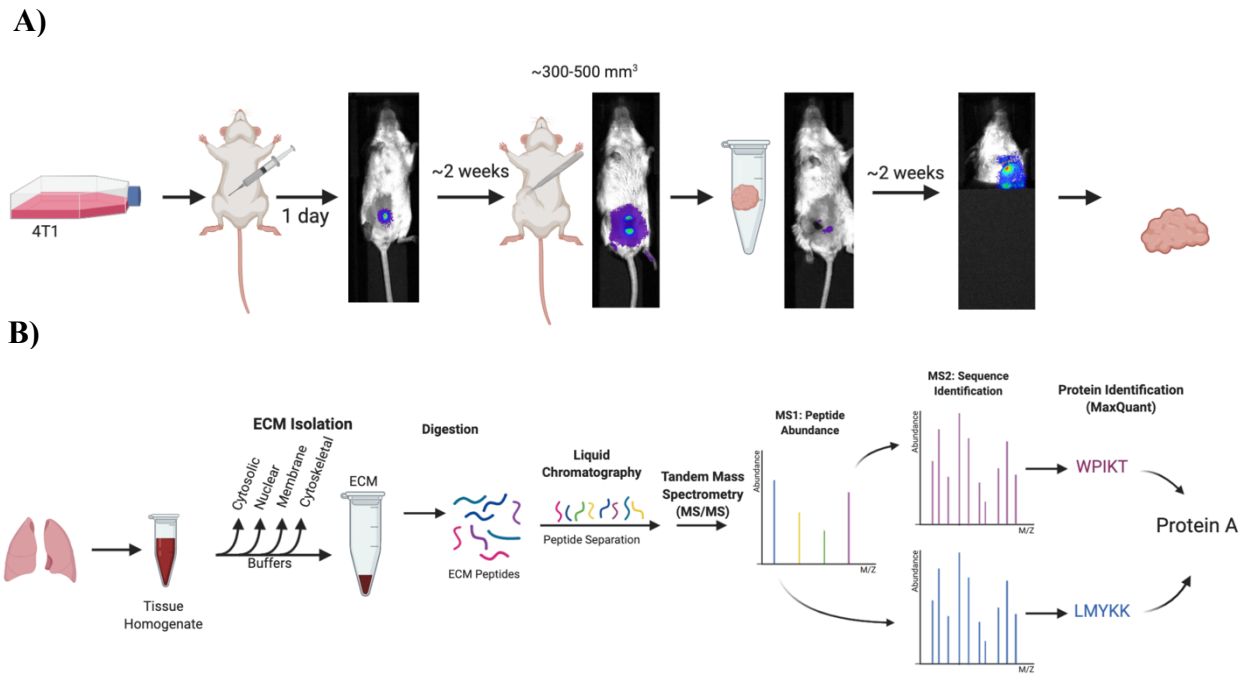


Figure 1. Experimental flow work of **A)** tissue extraction and **B)** mass spectrometry preparation and analysis. **A)** Luciferase positive 4T1 mammary carcinoma cells were inoculated into the fat pad of Balb/C mice. After primary tumors reached 300-500 mm^3 in volume, tumors were excised. After tumors metastasized to the lungs, confirmed with bioluminescent imaging, mice were euthanized, and lungs and any relapse rumors were dissected. **B)** Tissue samples were processed via an ECM isolation protocol. Tissues were fractioned using buffers leaving behind an ECM-rich pellet. The proteins in this insoluble fraction were then digested and prepared for LC-MS/MS analysis. Figure created using BioRender.

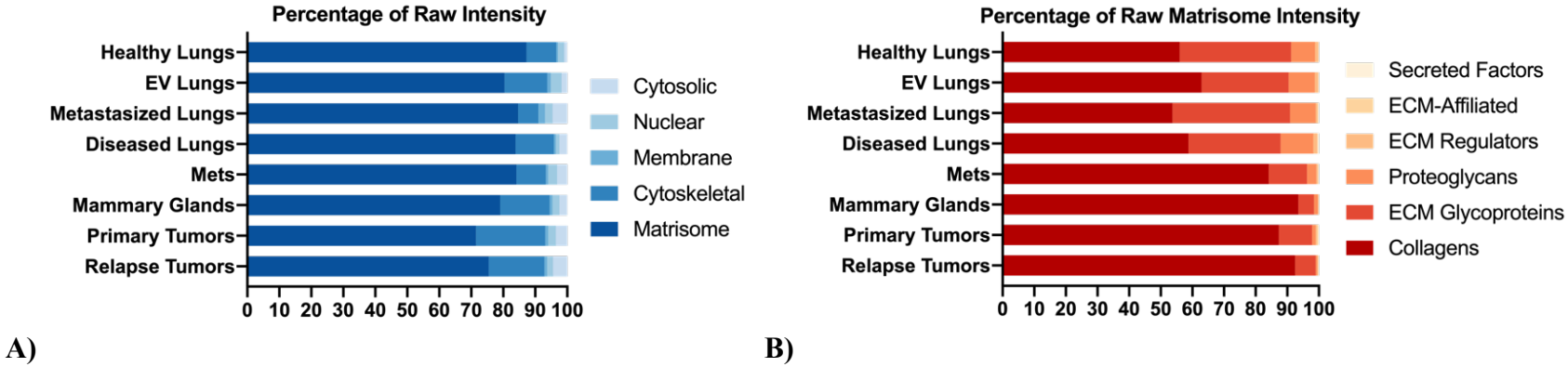


Figure 2. A) Fraction distribution of identified proteins. **B)** Matrisome classification of identified proteins for all tissue types.

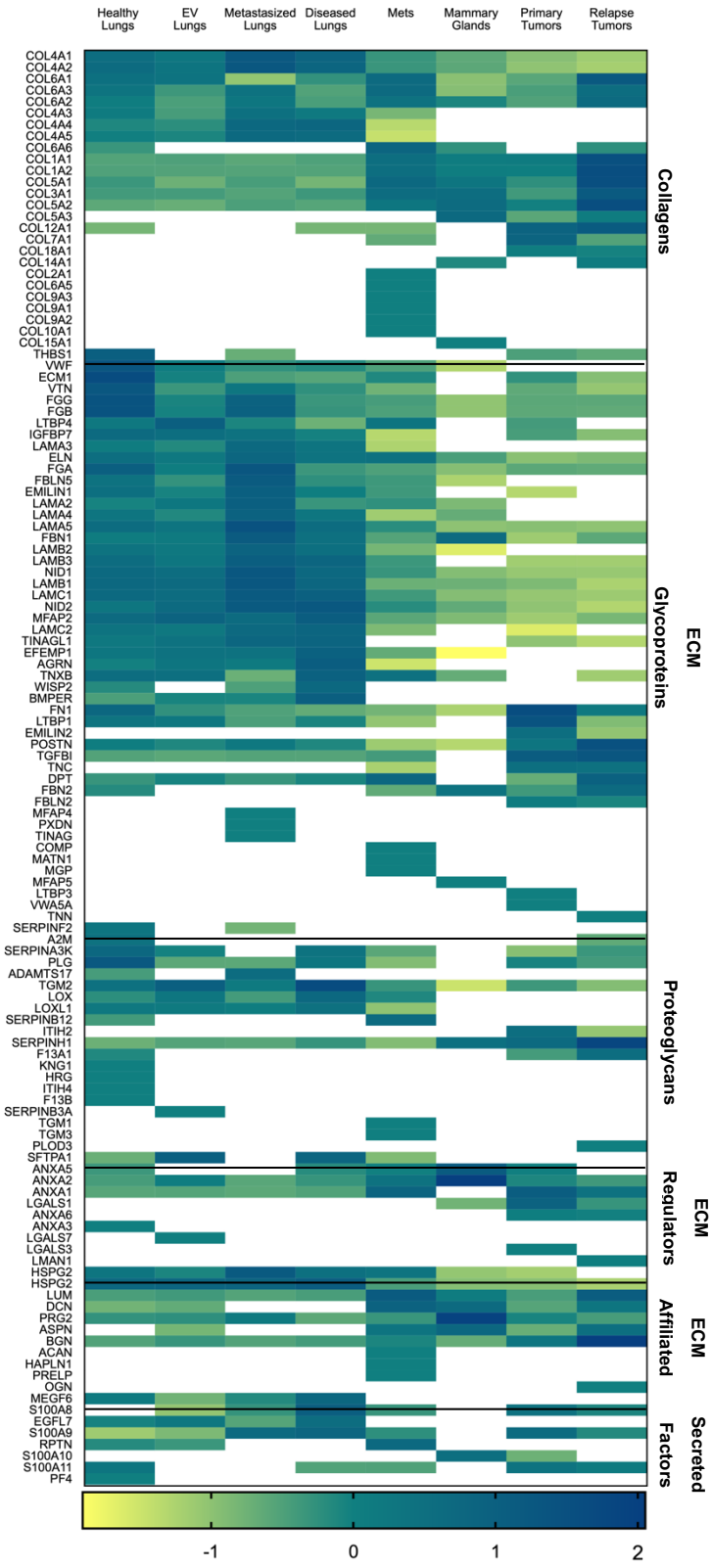


Figure 3. Heatmap from z-score distribution of all tissue types.

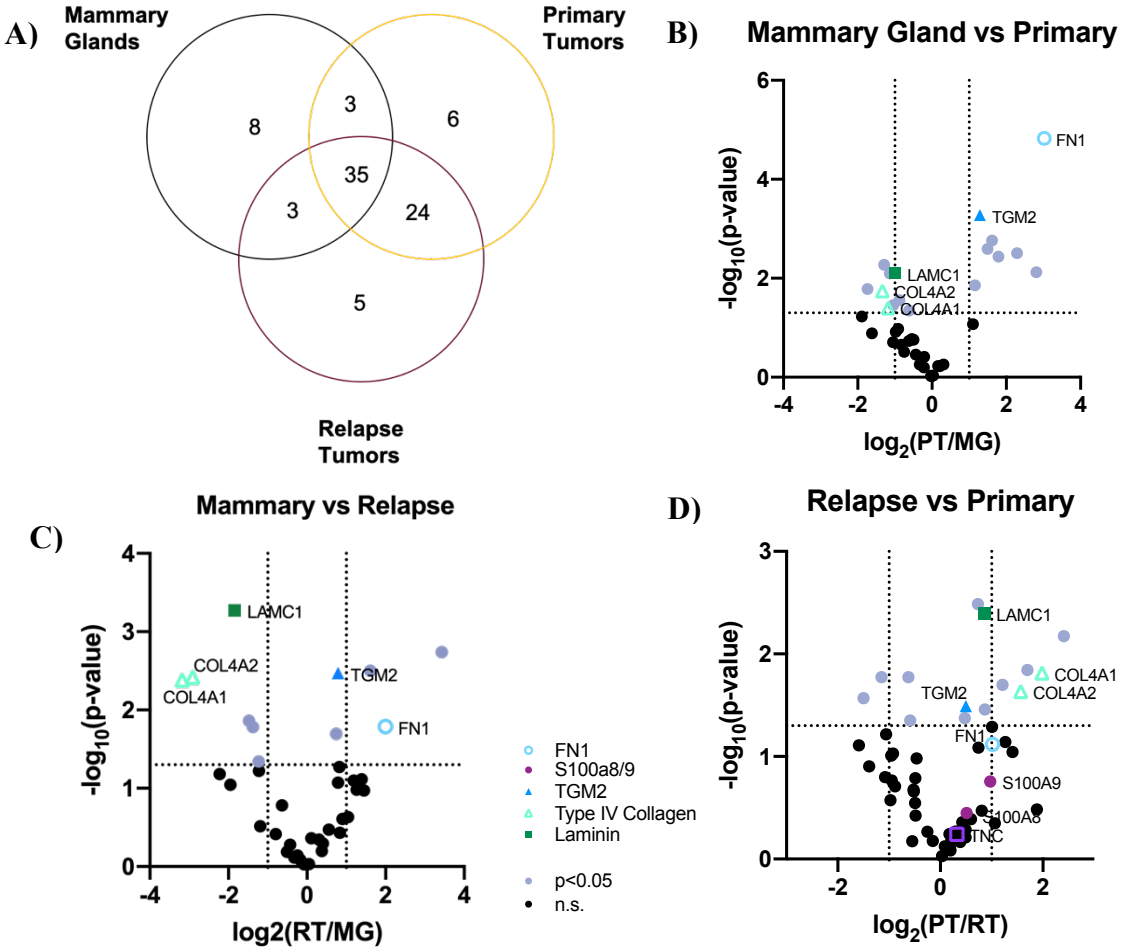


Figure 4. Matrisome graphs associated with tumor progression. **A)** Venn diagram comparison similar proteins identified in Mammary Glands, Primary Tumors, and Relapse Tumors. Volcano Plots of common identified proteins between the average of **B)** Mammary Glands and Primary Tumors, **C)** Mammary Glands and Relapse Tumors and **D)** Relapse Tumors and Primary Tumors.

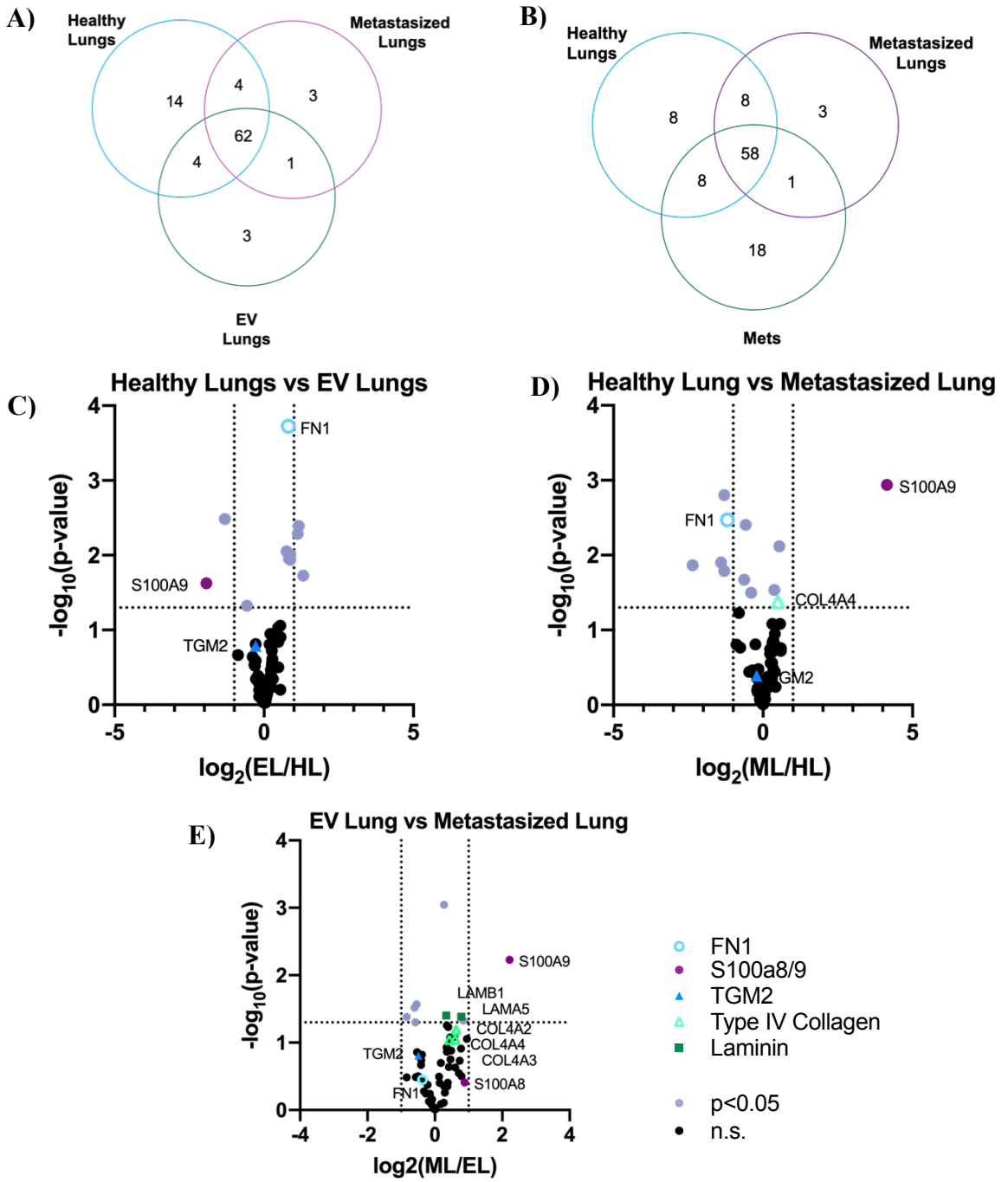


Figure 5. Comparison of ECM proteins identified during lung progression. **A)** Venn diagram comparison of proteins identified in the Healthy Lungs, EV Lungs, and Metastasized Lungs. **B)** Venn diagram comparison of proteins identified in the Healthy Lungs (HL), Metastasized Lungs (ML), and Mets. Volcano Plots of common identified proteins between the average of **C)** Healthy Lungs (HL) and EV Lungs (EL), **D)** Healthy Lungs (HL) and Metastasized Lungs (ML), and **E)** EV Lungs (EL) and Metastasized Lungs (ML).

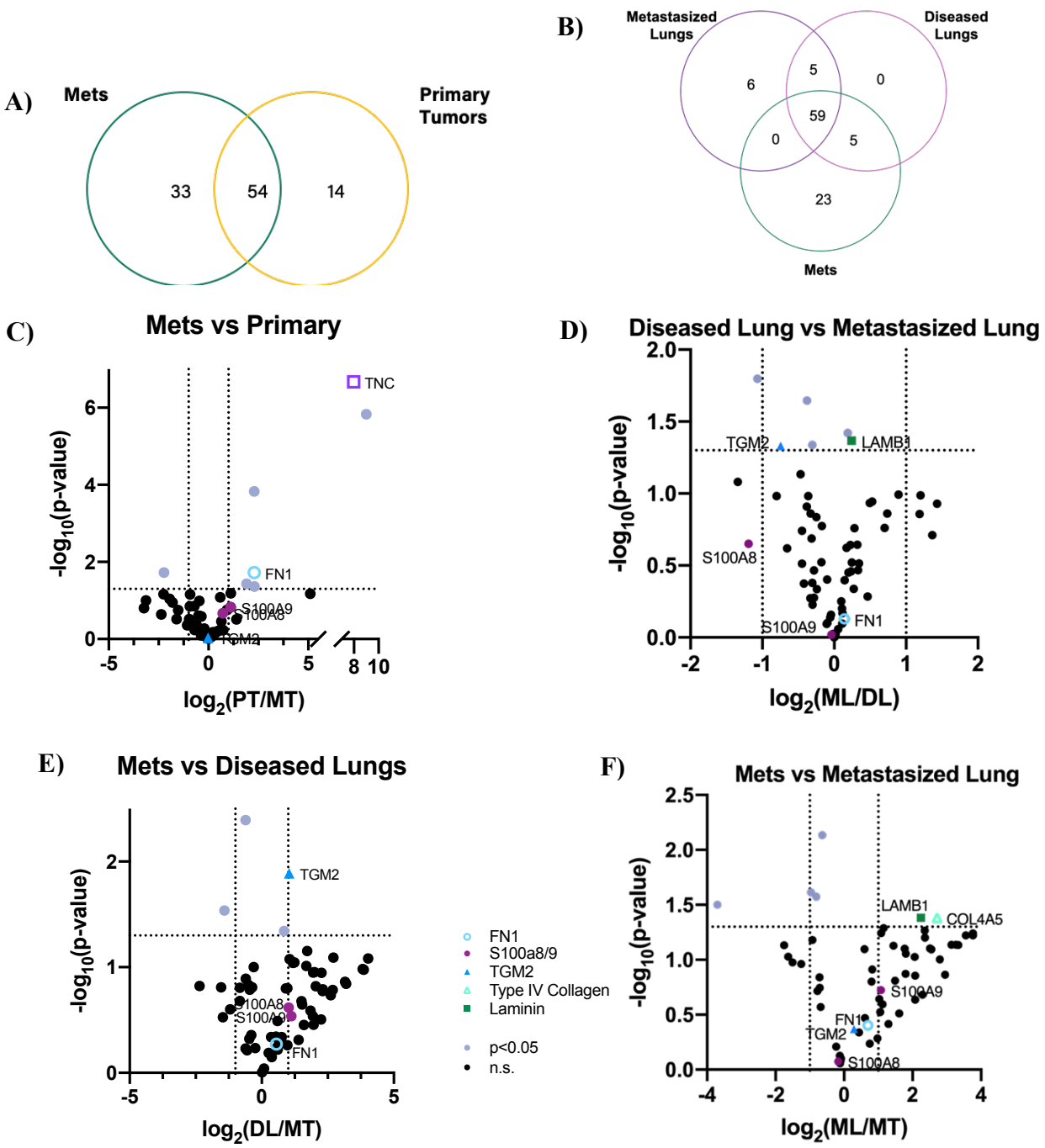


Figure 6. Comparison of Diseased Lungs and Tumors. **A)** Venn diagram comparison showing differences in proteins identified in fully metastasized Metastasized Lungs, Diseased Lungs, and Mets. **B)** Venn diagram comparison of identified proteins in the Mets and Primary Tumors. Volcano Plots of common identified proteins between the average of **C)** Mets (MT) and Primary Tumors (PT), **D)** Diseased Lungs (DL) and Healthy Lungs (HL), **E)** Mets (MT) and Diseased Lungs (DL), and F**)** Mets (MT) and Metastasized Lungs (ML).

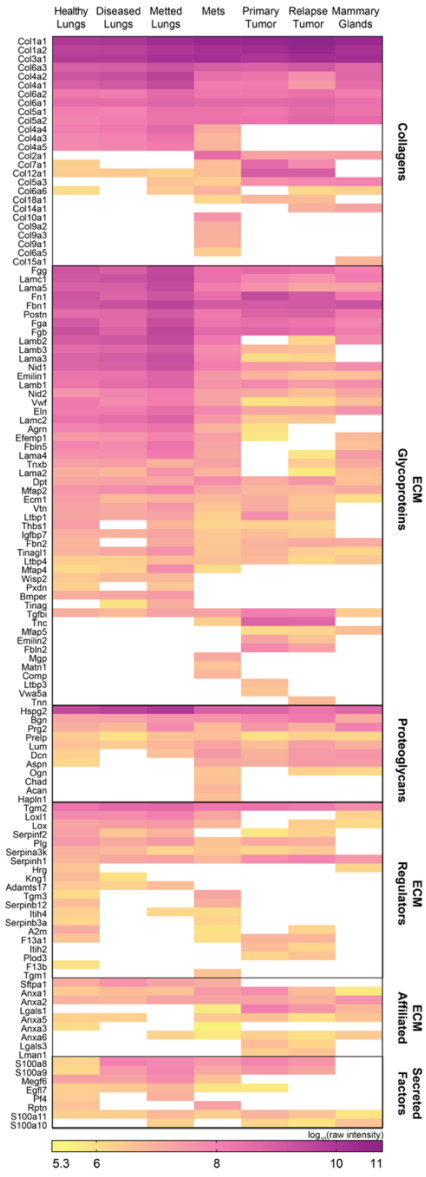


Figure S1. Raw Intensity Heat map

Table 1. Name, abbreviation, and description of tissue types

Name	Abbreviation	Description
Healthy Lungs	HL	Lungs removed from healthy Balb/C mice
EV Lungs	EL	Lungs removed from Balb/C mice inoculated with extracellular vesicles (EV) into the peritoneum
Metastasized Lungs	ML	Lungs from Balb/C mice inoculated with 4T1 cells with metastatic lesions intact
Diseased Lungs	DL	Lungs from Balb/C mice inoculated with 4T1 cells with metastatic lesions removed
Mets	MT	Metastatic tumors removed from lungs in Balb/C mice inoculated with 4T1 cells
Mammary Glands	MG	Mammary glands removed from healthy Balb/C mice
Primary Tumors	PT	Initial tumor excised from Balb/C mice inoculated with 4T1 cells into second fat pad
Relapse Tumors	RT	Any recurrent tumor removed from Balb/C mice after excision of primary tumor

Table 2. Percentage of Matrisome classification of raw intensity of proteins identified in all tissue types.

	Collagen	ECM Glycoproteins	Proteoglycans	ECM Regulators	ECM Affiliated	Secreted Factors
Healthy Lungs	56.00	35.22	7.55	1.07	0.07	0.09
Metastasized Lungs	53.71	37.17	8.07	0.73	0.05	0.27
Diseased Lungs	58.79	29.09	10.36	1.25	0.13	0.39
Mets	84.15	12.05	2.98	0.50	0.16	0.17
Mammary Glands	96.16	2.73	0.69	0.23	0.18	0.01
Primary Tumors	87.36	10.48	1.15	0.46	0.36	0.19
Relapse Tumors	92.52	6.43	0.57	0.27	0.14	0.09

Table 3. Unique proteins to each tissue type

Healthy Lungs	EV Lungs	Metastasized Lungs	Diseased Lungs	Mets	Mammary Glands	Primary Tumors	Relapse Tumors
ANXA3	LGALS7	MFAP4		ACAN	COL15A1	LGALS3	LMAN1
F13B	SERPINB3A	PXDN		CHAD	MFAP5	LTBP3	OGN
HRG		TINAG		COL10A1		VWA5A	PLOD3
ITIH4				COL2A1			TNN
KNG1				COL6A5			
PF4				COL9A1			
				COL9A2			
				COL9A3			
				COMP			
				HAPLN1			
				MATN1			
				MGP			
				PRELP			
				TGM1			
				TGM3			

APPENDIX A. PROTOCOLS

In Vivo Assays

All *in vivo* studies were conducted following an IACUC approved protocol from Purdue University.

1. Luciferase expressing 4T1 cells were resuspended in phosphate buffered saline (PBS) (50,000 cells/100 µL for each mouse).
2. Injected ($5x10^4$ cells/mouse) into the second mammary fat pad of female 4-6 week old Balb/c mice (Jackson Labs, Bar Harbor, ME).
3. After the tumor reached a volume of approximately 300-500 mm^3 excise primary tumor
4. Track tumor dissemination using bioluminescent imaging Advanced Molecular Imager (AMI) (Spectral Instruments, Tucson, AZ).
    a. Inject 150 uL luciferase into the peritoneum of the mice.
    b. After 5 minutes anesthetize the mouse and turn on the imager.
    c. Place mice into the imager ventral side up and close the door.
    d. Use the Spectral Instrument on the computer.
        i. 30,60, or 90 second exposure and save to new file.
        ii. Download images to flash drive and save.
5. Upon metastatic growth into the lungs, mice were euthanized using CO2 and whole lungs along with any relapse tumors were removed.
    a. Lungs and mammary glands were harvested from healthy mice as controls.
6. To determine the effect the EVs have on the matrisome of the early metastatic niche, 50 µg of EVs were injected intraperitoneally into Balb/c mice.
    a. This EV priming occurred every other day for 21 d.
    b. At day 21, mice were euthanized, and lung tissues were removed.

Extracellular Vesicle Enrichment

Extracellular vesicles (EV) were enriched via centrifugation using our previously described protocols [20].

1. EVs were isolated from 4T1 cell serum starved for 72 hours in serum free conditioned DMEM media.
2. The conditioned media is centrifuged at 300 g for 10 min.
3. The supernatant is collected, and then centrifuge at 2000 g for 10 min.
4. Conserve the supernatant and filter using a 0.22 µM pore size vacuum filter (Millipore).
5. Transferred to 3KD MWCO Amicon ultra-15 centrifugal filter tubes and centrifuge at 3220 g for 1 h.

6. Centrifuge filtered media at 3220 g for an additional hour condensing the sample to approximately 1mL.
7. Ultracentrifuge the condense sample at 100,000 g for 2 h.
8. Discard supernatant resuspend pellet and wash in 1X phosphate buffered saline (PBS) by ultracentrifugation at 100,000 g for 2 h.
9. Discard the supernatant and resuspend the EV enriched pellet in 1X PBS for storage in a -80° C freezer until needed.

ECM Enrichment

Enrichment followed a previous protocol [13], [14], [21], [22].Buffers used are part of the Compartment Protein Extraction Kit supplied by Millipore Sigma

1. Homogenize tissue samples in 200 µL C buffer (per 100 mg of starting tissue weight) using the TissueRuptor (Qiagen) until the tissue is disrupted and a homogenous suspension is obtained.
2. Extract cytosolic fraction (C) by incubating the homogenate on a shaker for 30 min at 4°C
3. Centrifuge homogenate at 16,000xg for 20 min at 4°C. Collect supernatant in a clean tube, snap freeze and store at -80°C.
4. Wash the pellet with 400 µL of Buffer W (per 10 mg starting weight) using a pipettor to fully resuspend the pellet. Place on shaker at 4°C for 15 minutes. Centrifuge at 16,000xg for 20 min at 4°C. Collect supernatant in a clean tube (W), snap freeze and store at -80°C.
5. Resuspend the pellet in 100 µL of Buffer N (per 10 mg starting weight) to extract first nuclear fraction (N). Place on shaker at 4°C for 20 minutes. Centrifuge at 16,000xg for 20 min at 4°C. Collect supernatant in a clean tube, snap freeze and store at -80°C. **Repeat** for a second nuclear extraction (N2).
6. Resuspend the pellet in 100 µL of Buffer M (per 10 mg starting weight) to extract the membrane fraction (M). Place on shaker at 4°C for 20 minutes. Centrifuge at 16,000xg for 20 min at 4°C. Collect supernatant in a clean tube, snap freeze and store at -80°C.
7. Resuspend the pellet in 50 µL of Buffer CS (per 10 mg starting weight) to extract the cytoskeletal fraction (CS). Place on shaker at room temperature for 20 minutes. Centrifuge at 16,000xg for 20 min at 4°C. Collect supernatant in a clean tube, snap freeze and store at -80°C.
8. Snap freeze insoluble pellet and store at -80°C.

ECM Digestion and MS Preparation

This is to digest and clean the Insoluble (IN) fraction from the ECM isolation protocol to prepare for Mass Spectrometry [13], [14], [21], [22].

ECM digestion

1. Suspend pellet with 8M Urea and 100mM ammonium bicarbonate. Reduce with 10 mM dithiothreitol (DTT) for 2 h at 37°C under constant agitation.
2. Remove samples from heat and cool to room temperature for 15 min. Alkylate samples with 10 mM iodoacetamide for 30 min in the dark.
3. Dilute samples 1:4 with ammonium bicarbonate ([urea]<2M). Add 5 µL of deglycosylation enzyme Chondroitinase ABC and incubate for 2 h at 37°C with constant agitation.
4. Add 1 µg of protease Lys-C and digest for 2h at 37°C with constant agitation.
5. Add 3µg of protease trypsin and digest overnight at 37°C with constant agitation.
6. Add an additional 1.5 3µg of protease trypsin and digest for 2 h at 37°C with constant agitation.
7. Inactivate trypsin by adding 0.1% trifluoroacetic acid (TFA).

Detergent Removal

1. Remove caps from Pierce detergent removal spin columns and place into 2 mL tube.
2. Remove storage solution by centrifugation for 1 min at 15000xg. Wash resin three times with 400µL of clean PBS
3. Transfer spin column to clean tube
4. Add 200 µL of clean sample to resin, SLOWLY, and incubate on resin for 2 min at room temperature
5. Centrifuge cleaned samples for 2 min at 1500xg

Desalting Clean Up

1. Condition Nest Group C18 desalting columns with 100 µL of 100% HPLC-grade acetonitrile (ACN) and spin for 1 min at 800 xg
2. Equilibrate with 100µL HPLC-grade H2O + 0.1% TFA and spin for 1 min at 800xg. Repeat.
3. Place desalting column in new tube. Add sample to desalting column and spin for 1 min at 800xg. Peptides will be retained as contaminants flow through to tube.
4. Wash peptides in desalting column with 100µL of HPLC grade H2O + 0.1% TFA and spin for 1 min at 800xg.
5. Transfer desalting column to clean tube and elute peptides with 100µL of 80% CAN _ 25 mM formic acid in HPLC-grade H2O. Spin for 2 min at 800xg.
6. Transfer solution to a 0.5mL tube and dry peptides in CentriVap Benchtop Vacuum Concentrator (45°C).
7. Resuspend dried peptides in 10µL of 3% CAN + 0.1% formic acid in HPLC-grade H2O

Peptide Concentration Normalization

1. The peptide concentration was measured with the Pierce Quantitative Colorimetric Peptide Assay (ThermoFisher Scientific).
    a. Prepare peptide standards starting with 1mg/mL stock supplied by kit and serial dilute to 0.0625mg/mL with sample buffer. Include a 0 mg/mL in sample buffer.
    b. Use a 384- well plate and carefully pipette 1.5µL of each standard and sample (in triplicate).
    c. Combine the correct amount of working reagent (13.5µL x 3 replicates x (# samples + 1)):
        i. 50% Reagent A
        ii. 48% Reagent B
        iii. 2% Reagent C
        iv. Vortex (solution should be green)
    d. Add 13.5µL of working reagent to each well containing standards and peptides.
    e. Incubate at 37°C for 15 min in the dark. Samples/standards with higher peptide concentration will change from green to orange.
    f. Read absorbance at 480 nm.
    g. Calculate sample peptide concentration using standard curve.
2. Normalize the concentration, using sample buffer, of each sample to e0.2 mg/mL
    a. The most concentrated sample was brought to 0.2µg/mL with 3% ACN and 1% FA.
    b. All other sample concentrations were normalized with 3% ACN and 1% FA.
3. Peptides were analyzed using the Dionex UltiMate 2000 RSLC Nano System coupled with the Q exactive HF hybrid quadrupole-orbitrap mass spectrometer (QE HF; ThermoFisher Schientific) as previously reported [21]

REFERENCES

[1] L. Jin, B. Han, E. Siegel, Y. Cui, A. Giuliano, and X. Cui, "Breast cancer lung metastasis: Molecular biology and therapeutic implications," *Cancer Biology and Therapy*, vol. 19, no. 10. Taylor and Francis Inc., pp. 858–868, 03-Oct-2018.

[2] H. Dillekås, M. S. Rogers, and O. Straume, "Are 90% of deaths from cancer caused by metastases?," *Cancer Med.*, vol. 8, no. 12, pp. 5574–5576, Sep. 2019.

[3] X. Guan, "Cancer metastases: Challenges and opportunities," *Acta Pharmaceutica Sinica B*, vol. 5, no. 5. Chinese Academy of Medical Sciences, pp. 402–418, 01-Sep-2015.

[4] B. Medeiros *et al.*, "Triple-negative primary breast tumors induce supportive premetastatic changes in the extracellular matrix and soluble components of the lung microenvironment," *Cancers (Basel).*, vol. 12, no. 1, Jan. 2020.

[5] S. Kaushik, M. W. Pickup, and V. M. Weaver, "From transformation to metastasis: deconstructing the extracellular matrix in breast cancer," *Cancer Metastasis Rev.*, vol. 35, no. 4, pp. 655–667, Dec. 2016.

[6] F. Jallow, K. A. O'Leary, D. E. Rugowski, J. F. Guerrero, S. M. Ponik, and L. A. Schuler, "Dynamic interactions between the extracellular matrix and estrogen activity in progression of ER+ breast cancer," *Oncogene*, vol. 38, no. 43, pp. 6913–6925, Oct. 2019.

[7] A. E. Place, S. Jin Huh, and K. Polyak, "The microenvironment in breast cancer progression: Biology and implications for treatment," *Breast Cancer Research*, vol. 13, no. 6. Breast Cancer Res, 01-Nov-2011.

[8] R. N. Kaplan, S. Rafii, and D. Lyden, "Preparing the 'soil': The premetastatic niche," *Cancer Research*, vol. 66, no. 23. Cancer Res, pp. 11089–11093, 01-Dec-2006.

[9] A. G. Clark and D. M. Vignjevic, "Modes of cancer cell invasion and the role of the microenvironment," *Current Opinion in Cell Biology*, vol. 36. Elsevier Ltd, pp. 13–22, 01-Oct-2015.

[10] J. Insua-Rodríguez and T. Oskarsson, "The extracellular matrix in breast cancer," *Advanced Drug Delivery Reviews*, vol. 97. Elsevier B.V., pp. 41–55, 01-Feb-2016.

[11] L. Haas and A. C. Obenauf, "Allies or Enemies—The Multifaceted Role of Myeloid Cells in the Tumor Microenvironment," *Frontiers in Immunology*, vol. 10. Frontiers Media S.A., 28-Nov-2019.

[12] A. Naba *et al.*, "Characterization of the Extracellular Matrix of Normal and Diseased Tissues Using Proteomics," *J. Proteome Res.*, vol. 16, no. 8, pp. 3083–3091, Aug. 2017.

[13] A. Naba, K. R. Clauser, S. Hoersch, H. Liu, S. A. Carr, and R. O. Hynes, "The matrisome: In silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices," *Mol. Cell. Proteomics*, vol. 11, no. 4, Apr. 2012.

[14] A. Naba, K. R. Clauser, and R. O. Hynes, "Enrichment of extracellular matrix proteins from tissues and digestion into peptides for mass spectrometry analysis," *J. Vis. Exp.*, vol. 2015, no. 101, Jul. 2015.

[15] J. D. Hebert *et al.*, "Proteomic profiling of the ECM of xenograft breast cancer metastases in different organs reveals distinct metastatic niches," *Cancer Res.*, vol. 80, no. 7, pp. 1475–1485, Apr. 2020.

[16] S. Libring *et al.*, "The Dynamic Relationship of Breast Cancer Cells and Fibroblasts in Fibronectin Accumulation at Primary and Metastatic Tumor Sites," *Cancers (Basel).*, vol. 12, no. 5, p. 1270, May 2020.

[17] R. R. Langley and I. J. Fidler, "The seed and soil hypothesis revisited-The role of tumor-stroma interactions in metastasis to different organs," *Int. J. Cancer*, vol. 128, no. 11, pp. 2527–2535, Jun. 2011.

[18] G. H. Heppner, F. R. Miller, and P. V. M. Shekhar, "Nontransgenic models of breast cancer," *Breast Cancer Research*, vol. 2, no. 5. BioMed Central, pp. 331–334, 2000.

[19] B. A. Pulaski and S. Ostrand-Rosenberg, "Mouse 4T1 Breast Tumor Model," *Curr. Protoc. Immunol.*, vol. 39, no. 1, Oct. 2000.

[20] A. Shinde, J. S. Paez, S. Libring, K. Hopkins, L. Solorio, and M. K. Wendt, "Transglutaminase-2 facilitates extracellular vesicle-mediated establishment of the metastatic niche," *Oncogenesis*, vol. 9, no. 2, pp. 1–12, Feb. 2020.

[21] A. M. Saleh, K. R. Jacobson, T. L. Kinzer-Ursem, and S. Calve, "Dynamics of Non-Canonical Amino Acid-Labeled Intra- and Extracellular Proteins in the Developing Mouse," *Cell. Mol. Bioeng.*, vol. 12, no. 5, pp. 495–509, Oct. 2019.

[22] K. R. Jacobson, A. M. Saleh, S. N. Lipp, A. R. Ocken, T. L. Kinzer-Ursem, and S. Calve, "Extracellular matrix protein composition dynamically changes during murine forelimb development," *Under Rev.*, 2020.

[23] A. Shinde *et al.*, "Autocrine fibronectin inhibits breast cancer metastasis," *Mol. Cancer Res.*, vol. 16, no. 10, pp. 1579–1589, Oct. 2018.

[24] T. Oskarsson *et al.*, "Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs," *Nat. Med.*, vol. 17, no. 7, pp. 867–874, Jul. 2011.

[25] S. Wang, R. Song, Z. Wang, Z. Jing, S. Wang, and J. Ma, "S100A8/A9 in inflammation," *Frontiers in Immunology*, vol. 9, no. JUN. Frontiers Media S.A., p. 1298, 11-Jun-2018.

[26] S. Ghavami *et al.*, "S100A8/A9 at low concentration promotes tumor cell growth via RAGE ligation and MAP-kinase dependent pathway."

[27] M. G. Ondeck *et al.*, "Dynamically stiffened matrix promotes malignant transformation of mammary epithelial cells via collective mechanical signaling," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 116, no. 9, pp. 3502–3507, Feb. 2019.

[28] C. Bergenfelz *et al.*, "S100A9 expressed in ER-PgR-breast cancers induces inflammatory cytokines and is associated with an impaired overall survival," *Br. J. Cancer*, vol. 113, no. 8, pp. 1234–1243, Oct. 2015.

[29] E. C. Bellavance *et al.*, "Clinical Significance of Elevated S100A8 Expression in Breast Cancer Patients," *Front. Oncol.* | *www.frontiersin.org*, vol. 8, p. 496, 2018.

[30] S. Rafii and D. Lyden, "S100 chemokines mediate bookmarking of premetastatic niches."