Quantitative Proteomic and Mutational Landscape of Metaplastic Breast Carcinoma and Generation of a 3D Organoid Model of Neoplastic Progression

by

Sabra I. Djomehri

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Molecular and Cellular Pathology)
in the University of Michigan
2020

Doctoral Committee:

Professor Celina G. Kleer, Chair
Professor Jennifer Linderman
Professor Sofia D. Merajver
Professor Alexey I. Nesvizhskii
Professor Shuichi Takayama, Georgia Institute of Technology

Sabra Djomehri

djomehri@umich.edu

ORCID iD: 0000-0003-3790-5949

Dedication

To my mother, Linda Ann (A'alia) Djomehri, whose life and untimely death gave the inspiration for a career in cancer research. Linda passed away in July 2004 from a recurrent breast cancer and her memory gives me strength every single day to strive and persevere. Linda was an incredible role model and mother who led by generosity, compassion, and the pursuit of truth. I love you and miss you every day.

And to my father, whose constant support and guidance through the difficult years have allowed me the confidence to pursue a career in science and who gave his all to see me succeed. I appreciate all your passion, patience and hard work in maintaining kindness and equity in the face of so much hardship. Your willpower inspires me to never give up hope.

Acknowledgements

The presented dissertation is the result of a unique synergy of collaborative efforts from multiple departments and individuals. I am endlessly grateful for the inspirational mentors who allowed me the ability to grow in a field, cancer research, which I never thought possible prior to coming to Michigan. I first extend full gratitude to Drs. Celina Kleer and Shuichi Takayama. When entering the PhD program in Molecular and Cellular Pathology (MCP), I came with a different background in physics and bioengineering and did not know what to expect in the vast realm of biology. Dr. Takayama was the first to allow me to discover a future career in breast cancer research. Having a mother who passed away from breast cancer, I was never inclined to pursue this field academically, as the memory of that painful incident cast a shadow on my mind. But Dr. Takayama enabled me to expand my mind away from that inner fear and embrace a new change. From there, I channeled into a new path under the mentorship of Dr. Kleer, who has taught me an enormous amount about balancing one's higher intellect while exuding the utmost humility and patience. These characteristics make Dr. Kleer an amazing mentor, and I am deeply humbled that she took me on at a difficult transition period. Dr. Kleer also embodies a mentorship style that fosters creativity through independence, collaboration, and positivity, and is committed to a highly equitable work environment that historically encourages women and people of various backgrounds to succeed. I recognize and I am deeply honored by this.

I am also extremely thankful to have such inspiring people on my committee. First, my deepest gratitude to Dr. Alexey Nesvizhskii, who enabled me to dream bigger. I was also not

thinking to pursue bioinformatics research, however after my first year under his direction in the proteome informatics of cancer training program (PICTP), I was hooked. Thanks to Dr. Nesvizhskii, my academic trajectory and goals have shifted significantly and I will be pursuing a future in using and developing new bioinformatics tools. Dr. Nesvizhskii is highly energetic and constantly pushed me to rise to new heights. Also, I am extremely humbled to have Drs. Sofia Merajver and Jennifer Linderman on my committee, who provided an incredible amount of intellectual poignancy and several invaluable and thought-provoking atmospheres. Thank you for the support you have given me and belief in me as a scientist.

I would next like to thank the members of Kleer lab, Takayama lab, and Nesvizhskii lab, both past and present, who have all made incredible contributions to this work and who have shaped my development as a scientist, and made this time period a stimulating and enjoyable place to work. Maria, Boris, Shilpa, Talha, Mai (Kleer lab); Felipe, Hui-Yin, Fengchao, Venky, Danny (Nesvizhskii lab); and Cameron, Ryan, Joyce, Suny, Ge-Ah, Eric (Takayama lab). I'm also extremely thankful to Dr. Zaneta Nikolovska-Coleska and Laura Labut of MCP who run an amazing and well-organized program, and to all the extended staff, students and faculty members in Pathology. Zaneta has greatly improved the quality of my training in a number of ways, and is a strong leader who has been very supportive and insightful. Laura is an incredible program administrator who is selfless, humble, and inspiring in her work ethic.

Most importantly, I would like to conclude by thanking my family. I wouldn't be here without the endless support and love of my beautiful wife Nada. I am blessed to be a part of your life and our sons' lives. What drives me most is knowing we will come together again with our beautiful sons, Moey and Ibby, very soon. And finally, to my mom and dad, who sacrificed everything in support of our thriving.

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

Triple-negative breast cancer (TNBC) is considered to be the most aggressive and has worse prognosis compared to other breast cancers and accounts for roughly 18% of all epithelial cancers of the breast, or carcinomas. TNBC exhibits complex molecular heterogeneity both inter- and intratumorally and likely consists of several distinct molecular subgroups that are currently unknown. Metaplastic breast carcinoma (MBC) is even more aggressive than triple-negative breast cancer (TNBC) but also typically presents as triple-negative histologically, and is defined by the admixture of both invasive glandular and non-glandular "metaplastic" heterologous elements of spindle, squamous or sarcomatoid subtypes.

The protein profiles underpinning the phenotypic diversity and metastatic behavior of MBC are unknown. We present a quantitative multi-subtype proteomic landscape of MBC, non-metaplastic TNBC, and normal breast from small yet well-annotated cohort of 27 patients, and also present the somatic mutational landscape on the same cohort. We used multiplex isobaric tandem mass tag (TMT) labeling for proteomics and quantified 5,798 proteins, and from whole-exome sequencing for genomics analysis we found 980 total somatic mutational variants. MBCs displayed increased epithelial-to-mesenchymal transition (EMT) and extracellular matrix (ECM) signaling, and reduced metabolic pathways compared to TNBC.

We discovered subtype-specific profiles among MBCs including distinct upregulated profiles; translation and ribosomal events in spindle, inflammation and apical junctions in squamous, and extracellular matrix in sarcomatoid. Comparison of the proteomes of spindle MBC with MMTV-cre;*Ccn6*$^{fl/fl}$ spindle MBC mouse tumors revealed a shared spindle-specific signature of 17 upregulated proteins involved in translation (e.g. RPL4,6,18, P3H1, PYCR1). The somatic mutational landscape also revealed MBCs share common TP53 mutations, and in PLEC, MUC17,

CRYBG2, and ZNF681. We identified that spindle and squamous MBC exhibit overlapping mutational profiles of genes involved in transcription, RNA metabolic processes and actin filament binding, while sarcomatoid tumors harbor distinct mutations in MAPK, WNT, protocadherin cluster genes, calcium binding and ECM organization. These data identify subtype-specific MBC protein profiles and mutational signatures that identified novel biomarkers for therapy.

Three-dimensional (3D) cell culture has been widely used in recent decades, compared with monolayer (2D) culture, because they better mimic the in vivo state. 3D systems utilize different types of gels critical for their success, such as collagen or the reconstituted basement membrane, Matrigel, which has enabled recapitulation of tissue architecture and function that is more physiologic compared to 2D. However, conventional 3D models using gel-embedded platforms have large variability and slow transport of biomolecules to the matrix-encapsulated cells. Here, we developed a highly reproducible, 3D scaffold-free hanging drop method amenable for primary tissues including mouse and human tumors, and our analyses describe a one drop-one organoid format using MCF10A cells, a non-tumorigenic breast cell line. We attained high-yield production of uniform organoids that resemble normal human breast acini, express both mammary gland-specific and progenitor markers, and we developed treatment assays for EMT induction and neoplastic progression delivering rapid quantification of phenotypic and morphological changes. Integration of 3D methods with omics analyses is envisioned to enhance the study of neoplastic progression and generate novel targets of both MBC and TNBC tumors.

Chapter 1 – Introduction

1-1 The human mammary gland and triple-negative breast cancers

The mammary gland is a complex structure that consists of an epithelial parenchyma situated in a bed of stromal cells that maintain and regulate both normal development and tumor progression [1], [2]. Breast cancer is the most common form of cancer in women and approximately 1 in 8 women in the U.S. (roughly 12%) will develop breast cancer in their lifetime [3]. Triple-negative breast cancer (TNBC) represents an estimated 15-18% of all breast cancer cases and are negative for estrogen receptor (ER), progesterone receptor (PR) or HER2 receptor and are among the most aggressive and high grade compared to hormone receptor positive forms of the disease [4], [5]. Therefore, it is of paramount importance to understanding the underlying biology that defines the complex molecular heterogeneity within TNBCs in order to develop new therapies to combat it.

The human mammary gland

The human mammary gland is a remarkably adaptive organ that consists of a system of branching ducts and lobules [1]. The mammary gland has a dynamic ability to undergo cyclic expansions and dramatic changes in both structure and function throughout lifespan of a woman, including stages of development, puberty, pregnancy, lactation, and involution [1], [6]. The functional unit of the breast is called the terminal duct lobular unit, or TDLU, which consists of a duct branched into lobules, or acini [7]. Acini are the functional secretory units in the mammary gland containing a central lumen. The TDLU is a highly organized tree-like structure with two

cell layers, the epithelium and basal cell/myoepithelium surrounded by a basement membrane [7]. The two major cell layers composing the ducts and acinar structures are basal and luminal cells, however the lobular units are also composed of secretory, milk-producing cells (alveolar cells) [8]. Acini are lined by an inner layer of luminal cells and an outer layer of myoepithelial (i.e. basal) cells [9]. With respect to the entire TDLU, the normal breast tissue microenvironment is also embedded in an adipose-rich stroma, which contains a mixture of cell types including mesenchymal cells, fibroblasts, myofibroblasts, macrophages and other immune cells [2], [10]. The distinguishing phenotype of breast acinar architecture is maintenance of a polarized structure, specialized cell-cell contacts, and attachment to the basement membrane [10]–[12]. In addition to providing structural support, the basement membrane is known to be a crucial regulator of cell proliferation, differentiation and organization of normal breast tissue function, and is a specialized type of extracellular matrix (ECM) that contains mainly type IV collagen and laminin [13].

The pathogenesis of epithelial tumors, called carcinomas, is characterized by the disruption of the intact, well-ordered acinar architecture. Recent studies of oncogenic transformation in the mammary gland and several cell-of-origin investigations suggest that human breast cancers arise from cells within the TDLU [7], [13]–[15]. A hallmark feature of the oncogenic transformation is the loss of epithelial polarity, tissue organization, cell-cell contacts, and rupturing of the basement membrane. Tumor malignancy has occurred once neoplastic cells escape the basement membrane and begin to invade through the adjacent stroma. Histologically, invasive breast cancers exhibit a wide range of molecular profiles with highly heterogeneous morphologies, and due to this, several aspects of tumorigenesis and mammary biology are not fully understood [16], [17].

Triple-negative breast cancer (TNBC) and current treatments

There are four main categories of breast cancer, these include luminal A, luminal B, human epidermal growth factor receptor 2 (HER2), and basal-like, as identified by studies on gene expression profiling [5], [6]. Triple-negative breast cancer (TNBC) represents an estimated 15-20% of all breast cancers and they do not express estrogen receptor (ER), progesterone receptor (PR) or HER2 [4], [18], [19]. TNBCs are typically of high pathologic grade, exhibit poor long-term prognostic outcomes, and are particularly more aggressive than HER2-amplified and hormone-positive breast cancers. They also share molecular features such as p53 and prevalence of germline BRCA mutations, are highly proliferative and express a phenotypically basal-like and claudin-low expression pattern [20]. In addition, this basal-like phenotype has a hallmark expression of basal cell markers such as cytokeratin 5/6 and 17. Overall, TNBC exhibits a complex heterogeneity at both inter- and intra-tumoral levels and likely consists of several distinct molecular subgroups that have not yet been clearly elucidated.

Emerging data highlights that TNBCs are not only very heterogeneous, but have variable prognosis across pathologic, genetic and clinical outcomes [21]. Central to understanding the prognostic significance of these outcomes, it is increasingly important to integrate pathologic, histologic, and other data sets together (e.g. genomics, proteomics) in order to pinpoint the underlying biological features, major drivers, and stage of disease. The treatment for TNBC has been less successful compared to other breast cancers since it lacks hormone receptor positive status and therefore, the use of current therapies have not been well targeted [22]. Cytotoxic chemotherapy has thus far been the backbone treatment option for TNBC. However, there have recently been encouraging results from molecularly targeted treatment options for TNBC, including PARP inhibitors (olaparib, talazoparib for germline BRCA mutation associated forms),

nab-paclitaxel for programmed death-ligand 1 (PD-L1+), and also checkpoint inhibitor atezolizumab [23]. As the molecular characteristics of TNBC is further revealed, a standard of care approach that is guided by tumor biology can be actualized as we enter a new era of TNBC therapeutics and management.

1-2 Epithelial-mesenchymal transition (EMT) and cancer stem cells (CSCs)

In the last decade, there has been increasing interest in the role of epithelial-mesenchymal transition (EMT) and cancer stem cells (CSCs) in tumorigenesis and cancer progression. These two concepts are now thought to be two sides of the same coin of a broader cancer cell plasticity program that exists in the tumor and its associated microenvironment. EMT involves the transdifferentiation of cells via a change in phenotype from epithelial to the acquisition of a mesenchymal-like state [24]–[26]. The process of EMT plays a well-established role in developmental programs and embryogenesis, and occurs when the mesoderm develops multiple tissue types and the generation of epithelial organs such as the breast and kidney [24]. It is also thought that EMT-derived cells have stem cell-like properties and are also capable of undergoing the reverse process, mesenchymal-epithelial transition (MET) once colonization of tumor cells have metastasized and invaded distant sites [27]–[29]. EMT is also facilitated by key regulators and transcription factors such as SNAIL, SLUG, TWIST, and ZEB1 and is characterized by the loss of epithelial markers (E-cadherin and important cell adhesion molecules) and gain of mesenchymal markers (N-cadherin, vimentin, fibronectin, and cytokines like TGFβ) [29]–[31].

Cancer cells are able to exploit this developmental EMT process during tumorigenesis and metastasis. However, since the cancer EMT program interacts very closely with the tumor microenvironment, which contains a wide range of cell types whose phenotypes are largely unexplored and difficult to distinguish, there has been controversy over actual molecular events

and clinical evidence behind EMT in cancer [32]. In addition, the presence of self-renewing cells in tumors, or CSCs, further complicates the situation. Cancer cell-of-origin models suggest two theories to explain the maintenance of malignant traits of neoplastic cells (**Figure 1-1**) [33], [34]. The first is the Stochastic Theory, which states that each cell in the tumor has the potential to become cancerous given the appropriate microenvironmental context. The second is the Hierarchy Theory, which postulates that CSCs are more likely to seed new tumors since they have a greater ability to self-renew compared to non-CSCs and can survive long-term as opposed to more mature cells, which have shorter lifespans.

In the last 15 years, a large body of research has described a small population of cells (~ 1%) in the tumor with evidence of self-renewal and stem-like features and are defined by their ability to "seed" or generate new tumors [33], [35]. Broadly, they are called "tumor-initiating cells" and were first discovered in the hematopoietic system and subsequently in solid tumors (e.g. breast, brain, colon). In fact, the discovery of hematopoietic stem cells (HSCs) 60 years ago has been an instrumental thrust towards elucidating a supposed CSC hierarchy [36]. In breast cancer, a small subset of CSCs expressing a $CD44^{+}/CD24^{-}$/ALDH+ phenotype is suggested to be a putative breast cancer stem cell (bCSC) population of notable importance linked with metastasis [37], [38]. This has been evidenced in transplantation studies that used immunocompromised mice, in which a few cells were sufficient to give rise to new tumors and cells that were both tumorigenic and non-tumorigenic [35], [39]. Genetic lineage-tracing approaches also emerged, enabling the identification of stem cells in tissues in situ, but when interrogating bCSCs, the multipotent mammary basal cell population was found to become unipotent [40]–[42]. Thus, the behavior of stem cells within transplantation and lineage-tracing platforms was noted to differ significantly, highlighting that the variable nature of stem cell

marker usage depends on the context of the particular platform used, in addition to the microenvironmental context the stem cell resides. Such discrepancies were also observed when studying HSCs. Nonetheless, these findings hinted that transplantation-based methods may be promising for revealing potential cancer stem cells, but may not ultimately uncover the fate of these cells.

To this end, recent investigations have aimed at correlating gene/protein level expression of EMT and/or CSC status in breast carcinomas with clinicopathological characteristics, molecular subtyping, and chemotherapy response to better define these phenomena in cancer. Because many current therapeutic options are targeted at cancer cells with conventional anti-proliferative chemotherapies, the benefits have been limited. In the future, the full scope of the differentiated CSC progeny that includes identifying those quiescent or "slow-cycling" CSC phenotypes should be elucidated, leading to novel therapies targeting these self-renewing cells that are therapy-resistant.

1-3 Metaplastic breast carcinoma

Metaplastic breast carcinomas (MBCs) are a rare and histologically diverse form of breast cancer that typically presents as triple-negative breast cancer and is thought to be enriched in EMT and CSC properties [43]–[45]. The word "metaplastic" itself is derived from Greek and means "change in form." This is due to a transformation of part or all of the tumor glandular carcinomatous component into a non-glandular, or metaplastic, component [46]. Although it accounts for roughly 1% of all breast tumors, it is present in up to 14% of breast carcinomas in women of African descent [47]. Metaplastic breast cancer was first documented in 1973 but only became recognized as a distinct histologic entity in breast cancer in 2000 [48], [49]. The heterologous, non-glandular elements within the carcinomatous region of MBC tumors are

exhibited histologically in three main forms: spindle, squamous, or sarcomatoid, in which the sarcomatoid also includes both osseous and chondroid differentiation [46], [50], [51]. All these forms are all extremely rare to arise in the breast, and compared with TNBC, metaplastic breast tumors are known to be more aggressive and are poorly responsive to chemotherapy. They also confer worse prognosis, are of higher pathological grade, and are known to have higher propensity for distant metastases than TNBC. Typically, these non-glandular metaplastic regions look like "solid nests" of tissue that are often observed next to a glandular carcinomatous region, such as invasive ductal carcinoma (IDC), and sometimes even ductal carcinoma in situ (DCIS) [43]. These findings suggest heterologous elements had initially progressed from associated ductal carcinoma. In addition, compared with conventional IDC patients, those diagnosed with MBC were shown to have less axillary lymph node involvement, and the 5-year survival rate for MBC is 65% compared with 89% with IDC [52].

Due to such histologically diverse forms of MBC neoplasms as spindle, squamous or mesenchymal elements, there is currently a lack of available therapies that specifically target these different subgroups. Previously, studies have looked at the role of immunohistochemical markers as predictors of survival, using p53, proliferation marker Ki-67, along with ER/PR and HER2 status [53]–[55]. By correlating immunohistochemical expression with clinicopathologic parameters, several recent investigations, including from our lab, have revealed an enriched combinatorial EMT and CSC phenotype present across MBCs ($CD44^+$ / $CD24^-$ / $ALDH1A1^+$, and $Ecad^-/Twist^+$) [56], [57]. Upon evaluation of these prognostic indicators, it has been observed that the combination of epithelial-mesenchymal transition markers and cancer stem cell overexpression are not only powerful predictors of disease-free and overall survival, but may

contribute to the reason why metaplastic carcinomas are particularly more aggressive than other TNBCs.

Due to the imminent need to find clinical markers of MBC, the last 5 years especially has seen a greater push towards genomic and transcriptomic profiling studies in the hopes of determining the underlying molecular alterations that define each histologic subtype. At the genomic level to date, metaplastic carcinomas have been found to be enriched in PI3K, TP53, CDKN2A deletions, EGFR amplifications, and WNT, PTEN, and CTNNB1-activating mutations [51], [58], [59]. Notably, an increasing number of studies observed that chondroid and general sarcomatoid-matrix producing MBCs seem to lack PI3K aberrations [58]. Most of these are also commonly found in TNBC, however it is suggested that the PI3K/AKT/mTOR pathway could be more frequently altered in MBCs compared to TNBCs. Also, MBCs were not found to be driven by recurrent fusion genes, and they seem to harbor similar patterns of gene copy number alterations. Taken together, it has been challenging to pinpoint specific somatic mutational profiles that define each MBC subtype, nor do MBCs seem to be underpinned by a pathognomonic copy number alteration profile, and that possibly other genetic and/or epigenetic aberrations affecting noncoding regulatory elements share an important role and need to be further examined [60].

However, a recent transcriptomic study has reported that based on RNA-based gene expression profiling, distinct expression patterns according to MBC histologic subtype were actually determined [61]. Specifically, MBCs with spindle differentiation showed distinct transcriptomic profiles consistent with an epithelial-mesenchymal transition profile, whereas chondroid and squamous differentiation were variably classified on the basis of basal-like or claudin-low profiles. Given this heterogeneity, and due to its rarity, it seems that future studies

with larger sample sizes are required to refine the boundaries of metaplastic versus non-metaplastic carcinoma components on the basis of genomic, epigenomic, transcriptomic and proteomic levels.

CCN6 (WISP3) and development of an MBC mouse model

To date, there is a lack of physiologically relevant animal models of metaplastic carcinomas. Recently, our lab has been one of the first to generate an MBC mouse model generated by Ccn6 deletion and a floxed Ccn6 mouse bred with an MMTV-Cre mouse [62]. CCN6, otherwise known as WISP3, is a secreted matrix protein of the CCN family that includes six members, all of which have profound regulatory roles in embryonic development, cell attachment and growth [63]. CCN proteins are tissue-specific and bind to cell surface receptors, are involved in integrin signaling, and modulate the effect of extracellular growth factors on epithelial cells. In addition, CCN6, located at 6q21-22, was first identified to be downregulated in aggressive inflammatory breast cancers, and later our lab showed that CCN6 is secreted by ductal epithelial cells in the breast [62], [64]. Several studies have revealed that the downregulation of CCN6 is linked with EMT, and thus, CCN6 was found to strongly regulate epithelial-mesenchymal transition and in general, any phenotypic changes between epithelial and mesenchymal states, and is known to modulate the TGFβ/BMP signaling axis [56], [64]. The $Ccn6^{fl/fl}$;MMTV-Cre mice formed invasive high grade mammary carcinomas with EMT features, and that CCN6 expression by immunostaining was reduced specifically in the metaplastic component compared to both normal breast and IDC tumors. Interestingly, this mouse model specifically recapitulates MBCs of spindle or of mixed spindle/squamous metaplastic elements. Overall, $Ccn6^{fl/fl}$;MMTV-Cre mouse tumors were found to have a protein expression profile analogous to human metaplastic carcinoma, and that Ccn6 deletion in the mammary gland is

sufficient to induce spindle MBCs on the basis of histological, immunophenotypic, and transcriptomic features. These are promising findings that could allow new targets to emerge in the diagnosis and management of MBC towards a subtype-specific treatement strategy.

1-4 3D culture systems in modeling breast cancer progression

Cell culture models are crucial for understanding biological and pathological processes for the evaluation of novel therapies. The literature suggests that three-dimensional (3D) culture systems are more physiologically relevant than conventional two-dimensional (2D) culture, as cells grown in a 2D environment are not capable of self-organizing and hence, tend to lose much of their in vivo phenotype [65]. However, when grown in 3D, cells are able to replicate physiological cell-ECM interactions, having been grown on an appropriate scaffold material or other extracellular matrix substrate that allows for a more faithful recapitulation of in vivo phenomena [66], [67]. Cancer cells grown in 3D also display signatures of drug resistance that are strikingly similar to that of in vivo tumors [68]–[70]. The use of 3D cell culture models has increased the understanding of breast cancer progression, enabled testing of new treatments, and may offer a unique platform for high throughput analyses that monitor molecular alterations during breast tumorigenesis, progression, and metastasis. Therefore, novel model systems that better recapitulate the mammary gland architecture and morphogenesis while also delivering reproducible quantitative parameters are greatly needed.

3D culture of mammary cells

Three-dimensional culture of mammary epithelial cells on a reconstituted basement membrane allows for the formation of polarized acini-like spheroids that resemble key aspects of glandular in vivo architecture. Early studies growing mammary epithelial cells in 3D found the main histological features of these 3D structures included a polarized morphology, cell-cell

contacts, and attachment to and presence of a basement membrane [10], [11], [71]. Glandular epithelial cells, such as in the mammary gland, require the development and maintenance of this polarized unit, which is critical for its function. On the other hand, the pathogenesis of mammary tumors requires the disruption of this highly organized architecture. Initially, studies began using a laminin-rich matrix, commonly known as Matrigel, or a collagen matrix, as a reconstituted basement membrane [69], [72]. To form acinar structures, mammary cells are typically cultured in a reconstituted basement membrane, Matrigel, which are derived from Engelbreth-Holm-Swarm (EHS) tumor and is mainly composed of laminin, collagen IV and entactin and is known for driving morphogenetic differentiation in 3D cultures. Studies subsequently demonstrated in vitro mammary gland morphogenesis, mammary gland branching, and discernment of normal and neoplastic mammary phenotypes, including seminal works from Brugge and Bissell groups and others, which laid the framework for standard 3D culture procedures of breast models that exist today [10], [14], [73]. Several early methods implementing 3D epithelial cultures used certain tumor-derived cell lines like LIM 1863 (colon carcinoma cells), DU4475 (mammary carcinoma cells), Madin-Darby canine kidney (MDCK) cells, S2 cells, MCF-10A cells (human origin), primary human mammary epithelial cells (HMECs) and also mouse mammary epithelial cell lines like TAC-2 and EpH4 [12].

MCF10A cells

MCF10A, a spontaneously immortalized non-transformed human breast epithelial cell line, is a well-established in vitro model of the benign breast. MCF10A cells are near-diploid cells with stable karyotype and are derived from the breast tissue of a 36-year-old patient with fibrocystic changes [71], [74]. The non-tumorigenic MCF10A cells form acini with hollow lumen in 3D culture, which are generated either by embedding cells inside or on top of Matrigel

after 1-2 weeks [10]. These scaffold-based systems using MCF10A retain important features found in glandular epithelium in vivo and have particular advantages studying tumor initiation, cell-ECM biophysical interactions, migration and invasion [75]. These cells also exhibit low proliferative capacity, making it easy to interrogate effects of oncogenic transformations as a result of treatment assays, despite being cytogenetically abnormal (they harbor genetic abnormalities including deletion of p16 and p14ARF, and amplifications of MYC, but express wild-type p53). Another limitation of MCF10A as a model of normal breast function are that these are immortalized cells and with long-term culture, harbor loss of architectural cues of normal mammary cells, and despite exhibiting a central hollow lumen, MCF10A lacks multiple cell types such as surrounding myoepithelial cells [75]. MCF10A cells are ER^- and associated with a basal epithelial phenotype, which may in turn make this cell line most relevant for studying triple-negative and/or basal breast cancers. Overall, mammary spheroid and organoid assays have provided a unique opportunity for drug development and are a promising pre-clinical cancer method that make it possible to interrogate the heterogenous composition of primary tumors.

Hanging droplet technique

Currently, drug development faces increasing challenges with issues of accuracy, sensitivity, and throughput of available in vitro testing platforms. The most common in vitro screening assays are still based on a monolayer, or 2D culture, which are not only the gold standard approach but are favored due to easy handling and better suitability for automation in high-throughput systems [69], [70], [76]. However, since it is known that 2D cultures lack the necessary physiologic and microenvironment interactions analogous to in vivo, this drastically limits the prediction of drug effects on humans. Therefore, there has been increasing support for

the use of 3D cultures and particularly spheroid and organoid systems in developing high-throughput screening assays. Consistent 3D structures are a top challenge of 3D culture methods, and if proven to be a scalable technique, must overcome limitations including cells that form slowly or have inefficient reaggregation capacity. Among these methods are the hanging drop technique, which allows for media exchanges and administration of chemical components with flexible timing [77], [78]. Others have achieved spheroid reproducibility using 96-well hanging drop systems, culture plate lids, spinner flask cultures, rotary cell systems, and various microfluidic chip devices to better control spheroid sizes, and increase spheroid formation efficiency [79], [80]. However, many of these platforms are tedious and produce highly variable spheroid sizes, are lower throughput, difficulty in handling. Microfluidics devices offer better consistency and simplified handling, however still suffers from compatibility with existing liquid handling robots, compatibility with drug screening, and long-term culturing.

An automated, modified high-throughput 384-well hanging drop platform developed recently by the Takayama group have demonstrated effective and compact formation of both cancer cells and normal cells using the macromolecular additive methylcellulose (MethoCel), along with lower concentration of Matrigel to the culture media only upon cell seeding (**Figure 1-2**) [81], [82]. This was found to contribute to highly uniform, compact spheroid morphologies and has been tested on spheroids generated from several breast cancer cell lines and MCF10A cells, and more recently, on primary tissues. This hanging droplet platform consists of a 384-well custom-made hanging drop array plate that is sandwiched between a standard Corning 96-well plate (**Figure 1-2a-b**), which is filled with distilled water and a standard plate lid. The distilled water from the bottom 96-well plate and the water reservoir on the two sides of the 384-well plate prevent evaporation of the hanging droplets, acting as a type of humidification chamber

(**Figure 1-2c**). The access holes on the top surface of the plate allow a liquid droplet containing live cells to be pipetted directly, resulting in consistent geometry of droplets and stable culturing conditions not possible using conventional flat hanging droplet substrates. Long-term culture is also possible, due to small volumes of droplets (~ 25 µl), allowing easier media exchanges using an automated liquid handling robot (**Figure 1-2d**) and enabling user-friendly high-throughput 3D culture applicable for multiple cell types and co-cultures.

1-5 Translational application of omics technologies

In order to extract the benefit of high-throughput studies from both in vitro (2D and 3D cultures) and in vivo studies (e.g. mouse models) to improve clinical outcome prediction in the era of precision medicine, cancer studies are now focused on omics data sets that have provided information regarding the etiology of oncogenesis. There is now an abundance of omics-based data currently available in biomedical research that has inspired the development of several robust statistical tools and software for interpreting and visualizing complex data sets [83], [84]. In addition, the complex and dynamic networks of molecular layers (e.g. genetics, epigenetics, mRNA, RNA, proteins, and metabolites) involved in cancer progression can be interrogated by omics technologies. A single omics layer can provide limited insights into the underlying mechanisms of disease, however, by combining these approaches obtained at different omics levels and assessing patterns in the cross talk between multiple molecular layers is not trivial, but is ultimately the goal. Therefore, the goal towards integrative multi-omics strategies may enhance functional analyses, discovery of key drivers contributing to specific cancers, and ultimately allow drug discovery, early detection, prevention and treatment at the individual patient level.

Genomics in the realm of medical research focuses on identifying genetic variants based on pathologic tissues, and in response to treatment. GWAS, genome wide association studies, made it possible from the early 2000s to map expression quantitative train loci (eQTL) for modeling biologic networks [85], [86]. Since then, a plethora of other omics-based technologies have been developed and are capable of capturing bulk information from transcripts, proteins and metabolites. GWAS studies generated information on multiple human populations consisting of thousands of individuals who have been genotyped for over a million genetic markers, finding statistically significant changes in minor allele frequencies and understanding complex phenotypes [86]. Associated techniques were also developed including genotype arrays, next-generation systems (NGS), and whole-genome and exome sequencing [87].

Epigenomics focuses on DNA-associated proteins such as DNA methylation and histone acetylation events, covalent modifications of DNA and histones, as these are considered to be major regulators of cell fate and gene transcription [86], [88]. Such modifications are impacted by genetic, heritable, and environmental factors. Several epigenome-wide association studies have revealed that the role of epigenetic modifications is important in disease development and biological processes, such as methylated regions of DNA, and certain epigenetic signatures known to be tissue-specific [89]. Recently established comprehensive epigenomic maps in multiple human tissues have emerged, such as the International Human Epigenome Consortium, giving insights into epigenetic modifications that correlate with disease [90].

Transcriptomics, the molecular intermediate between DNA and protein levels, looks at the RNA level and measure the expression of transcripts, and allows for the identification of novel splice sites, and RNA editing sites [61], [91]. RNA function such as those involving ribosomal complexes, are also important in pathogenesis, since up to 80% of the genome is

transcribed but only ~ 2% actually encodes proteins [84], [91]. RNA-seq investigations have therefore uncovered thousands of novel isoforms that enabled the discovery of an entirely new field of non-coding RNA biology that has been previously overlooked [92]. For example, it is now clear that many noncoding RNAs, both long non-coding (lncRNAs) and microRNAs play critical regulatory roles in both physiologic and pathologic processes in cancer.

At the protein level, proteomics has revolutionized the quantification of protein expression based on peptide abundances using mass spectrometry (MS) based methods [93]. These have recently been adapted for high-throughput analysis of in-depth proteomes which are capable of capturing thousands of proteins in human tissues and fluids. Fundamentally, MS methods measure the mass-to-charge ratio (m/z) of gas-phase ions, where the ion source converts analyte molecules into these gas-phase ions and a mass analyzer separates the ionized analytes based on m/z ratios [93]–[95]. The electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) are soft ionization techniques that can ionize peptides. Four types of mass analyzers are commonly used in proteomics, the quadrupole (Q), ion trap (QIT), time -of-flight (TOF) and Fourier-transform ion cyclotron resonance (FTICR) [93]. Over the past decade, developments in shotgun proteomics (bottom-up methods in which peptide detection is used to infer proteins) and chemical labeling using isobaric tandem mass tags (TMT) and isobaric tags for relative and absolute quantification reagents (iTRAQ) have been used in a variety of clinical samples for biomarker discovery and are highly robust for hyperplexing [96]–[98]. In addition, affinity purification which uses an antibody or genetic tag coupled with MS to identify associated proteins, has been widely used with chemical crosslinking and also previously adapted to measure interactions between proteins and nucleic acids (i.e. ChIP-seq). Proteomics also enables interrogation of proteins involved in post-translational modifications

such as phosphorylation, ubiquitination, proteolysis, glycosylation and nitrosylation, all of which play major roles in intracellular signaling cascades. Lastly, the field of metabolomics, which also uses liquid-chromatography-mass spectrometry (LC-MS) methods, has been developed to quantify metabolite levels from small molecules, amino acids, lipids, polar metabolites, carbohydrates and other products [99], [100]. Relative metabolite ratios are thought to reflect broad metabolic function, and some studies have even been able to use metabolomics to link genotype to phenotype information, and are proximal reporters of disease well-suited to serve as efficacy markers for drug development.

1-6 Summary

Here we have reviewed a broad body of literature on triple-negative breast cancer progression, specifically metaplastic breast cancers and how best to model them in vitro and in vivo. Understanding the key drivers and signaling axes of neoplastic and oncogenic transformation is likely to be enabled by integrative approaches that combine information on multiple scales and platforms, including human patient tumors and animal model data, 3D organoid screening and biobanking efforts, and multi-omics data sets which analyze these samples on multiple molecular levels. Both TNBC and MBC are highly complex, aggressive and heterogeneous tumors, with a lack of understanding on the underlying pathogenesis, limited models of disease progression and a lack of available therapies. Especially in the case of metaplastic breast carcinoma, which is more aggressive and rare owing to the presence of non-glandular components which display patterns of differentiation not typical to be found in the breast (spindle, squamous and sarcomatoid forms), we have introduced evidence that mouse models and 3D organoid platforms are capable of recapitulating key molecular drivers of such histological features. This makes it possible to understand the pathogenesis of this disease, how

these differentiation patterns arise, and what set of markers can be tested as novel therapies. However, in the future, studies will need to seamlessly integrate the information generated from animal models and 3D culture systems directly with omics technologies, which are useful for revealing multi-level information of thousands of genomic variants, transcripts, proteins, and other critical information. Pathogenesis occurs as a complex orchestration among all these molecular and cellular levels, resulting in the range of heterogeneity we see within these tumors, and it is of prime importance to collect these data sets both from large-scale and small-scale cohorts to be used for annotating future databases based on patient stratification approaches.

1-7 Figures

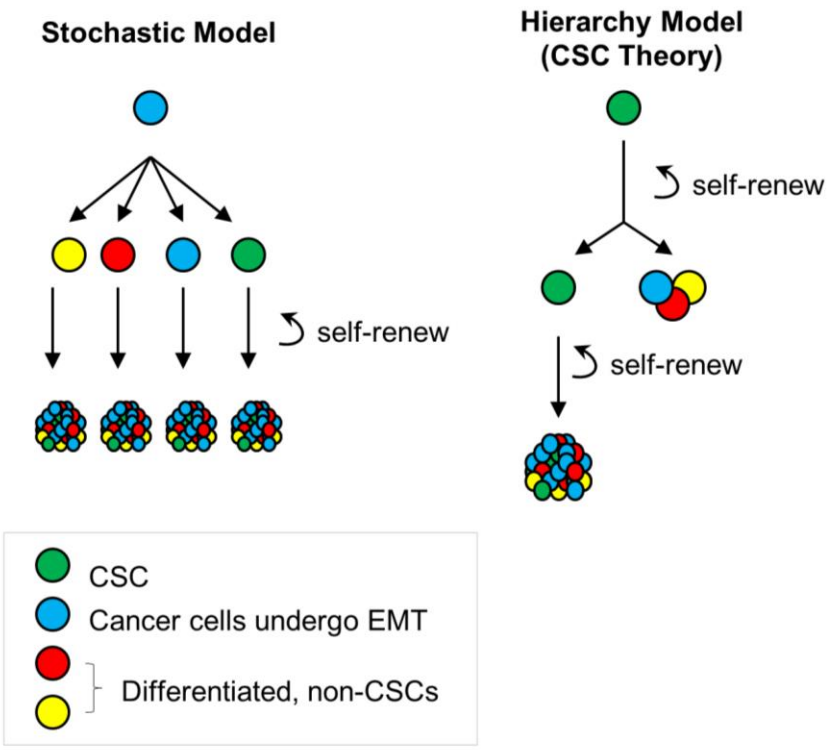


Figure 1-1. Two models proposed to explain the heterogenous nature of cancer cells, the Stochastic Model and the Hierarchy Model (Cancer Stem Cell Theory). In the Stochastic Model, any cell in the primary tumor can acquire the right set of somatic mutations to develop and seed new tumors with self-renewal and metastatic capacity. In the Hierarchy Model, tumors are only initiated by a small subset of rare self-renewing cells called cancer stem cells (CSCs) which are capable of self-renewal and seeding new tumors.

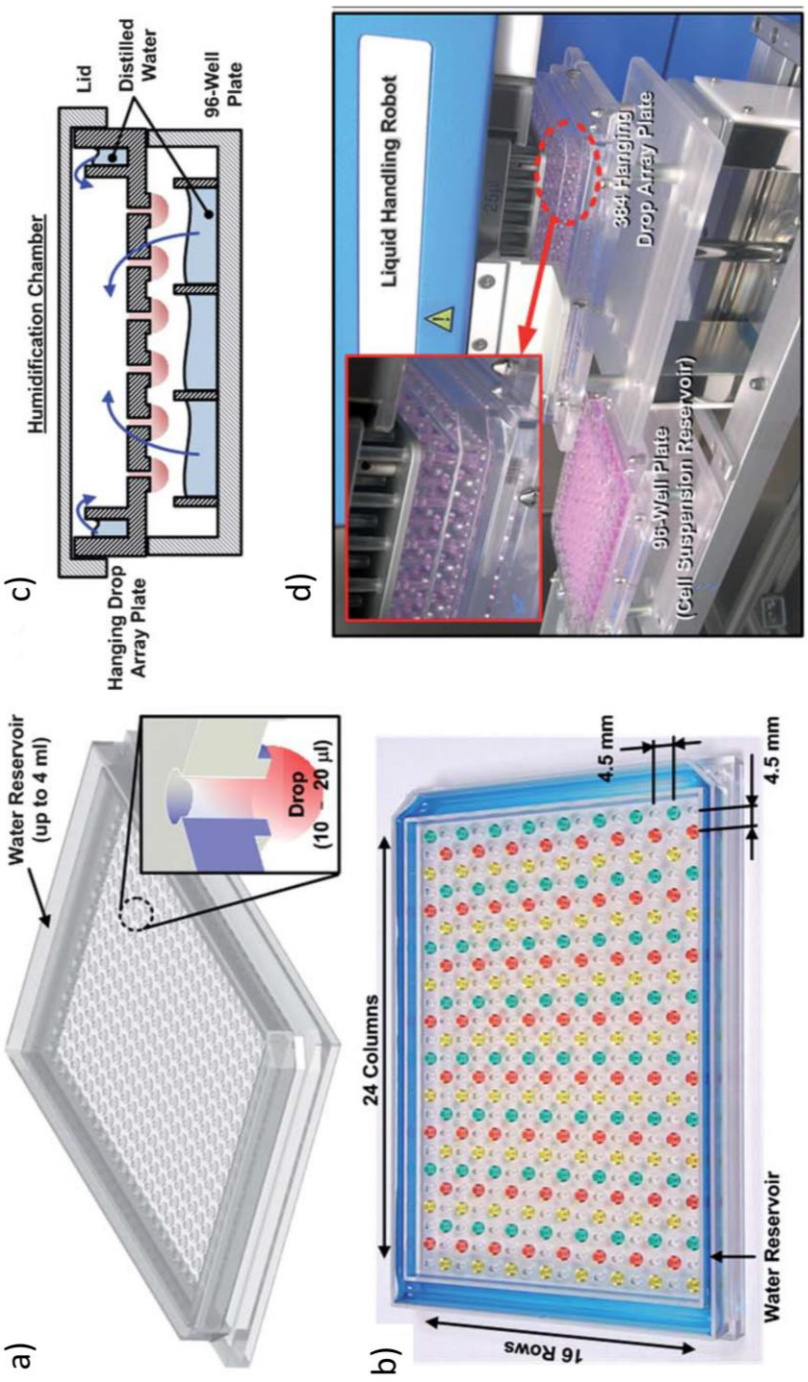


Figure 1-2. 3D culture platform with 384-well hanging droplet technique. a) Diagram of 384 hanging drop culture array plate, with cross-sectional view. b) Key dimensions of the array plate. c) Cartoon of the final humidification chamber used to culture 3D spheroids in the hanging drop array plate. The 384 hanging drop array plate is sandwiched between a 96-well plate filled with distilled water and a standard-sized plate lid. Distilled water from the bottom 96-well plate and the peripheral water reservoir prevent serious evaporation of the small volume hanging drops. d) The 384 hanging drop array plate operated with liquid handling robot capable of simultaneously pipetting 96 cell culture sites [81]. (Modified from Tung et al. 2011).

Chapter 2 – The Quantitative Proteomic and Mutational Landscape of Metaplastic Breast Carcinoma Pathological Subtypes and Their Relationship to Triple-Negative Tumors

The contents of the following chapter are under review in Nat Comm (NCOMMS-19-19350) and was written by: Sabra I. Djomehri, Maria E. Gonzalez, Felipe V. Leprevost, Shilpa Tekula, Hui-Yin Chang, Marissa J. White, Ashley Cimino-Mathews, Boris Burman, Venkatesha Basrur, Pedram Argani, Alexey I. Nesvizhskii, Celina G. Kleer

2-1 Abstract

Metaplastic breast carcinoma (MBC) is the most aggressive form of triple-negative cancer (TNBC), defined by the presence of "metaplastic" components of spindle, squamous, or sarcomatoid histology. The protein profiles underpinning the pathological subtypes and metastatic behavior of MBC are unknown. Using multiplex quantitative tandem mass tag-based proteomics we quantified 5,798 proteins in MBC, TNBC, and normal breast from 27 patients. MBC showed increased epithelial-to-mesenchymal transition and extracellular matrix, and reduced metabolic pathways compared to TNBC. MBC subtypes exhibited distinct upregulated profiles; translation and ribosomal events in spindle, inflammation and apical junctions in squamous, and extracellular matrix in sarcomatoid. Comparison of the proteomes of spindle MBC with MMTV-cre;*Ccn6*$^{fl/fl}$ spindle MBC tumors revealed a shared spindle-specific signature of 17 upregulated proteins involved in translation (e.g. RPL4,6,18, P3H1, PYCR1) and 19 downregulated proteins with roles in cell metabolism (e.g. ADH1B, ADH1C, LIPE, SOD1, FABP4). These data identify subtype specific MBC protein profiles providing biomarkers and therapeutic targets.

2-2 Introduction

Triple-negative breast cancer (TNBC) comprise a heterogeneous group of tumors with varying histologies and prognosis[101]. The most lethal subtype of TNBC is termed metaplastic breast carcinomas (MBC), a unique and heterogeneous group that account for 0.2 – 5% of all breast cancers[48], [49]. MBCs are characterized by the presence of a glandular epithelial component and a hallmark non-glandular "metaplastic" component that may consist of cells with spindle, squamous, and/or sarcomatoid (i.e. chondroid or osseous) features[50], [51]. Clinically, MBCs are more metastatic and chemoresistant than non-metaplastic TNBC, with the spindle subtype portending the worst prognosis[45]. This significant clinical challenge highlights the need to distinguish MBC tumors for diagnostic and precision treatment purposes.

The molecular alterations that distinguish MBC from TNBC, and the protein profiles that determine MBC histological subtypes are poorly understood[46], [102], [103]. To date, few studies have investigated the genomic and transcriptomic features of MBC[59], [61], [104]. Genetically, MBCs have a high level of genomic instability, display a complex copy number variation pattern, and tend to harbor significantly more mutations in PIK3CA, WNT, and TP53 compared to TNBCs, as well as presenting loss of CDKN2A and overexpression and amplification of EGFR[58], [105], [106]. Studies have been unable to show a genetic basis among MBC histologic subtypes, but have recently demonstrated distinct profiles at the transcriptomic level[61]. However, little is known about the mechanisms of MBC metastasis, and there are no reliable biomarkers or targets of therapy, underscoring the need to identify protein profiles specific to MBC and subtypes.

Our lab has generated MMTV-cre;Ccn6$^{fl/fl}$ knockout mice which form mammary tumors that recapitulate high-grade human MBC with predominant spindle components both morphologically and transcriptionally[107]. Using this model, we have identified an 87-gene

signature common between mouse and human MBCs including the transcription and translation regulatory proteins HMGA2 and IGF2BP2[108]. In addition, histological and immunophenotypical evidence suggests that MBCs are enriched in epithelial-to-mesenchymal (EMT) and different cellular compartment express stemness markers[37], [44], [57]. However, the protein profiles of human and mouse MBC are unknown.

Few proteomics studies to date explore invasive breast carcinoma signatures, with very limited information on MBC and normal tissue counterparts[18], [19], [109]. Here, we used human tissue samples of MBC, TNBC and normal breast and a quantitative custom-built platform to test the hypothesis that the histological subtypes of MBC may have distinct protein profiles that may result in the pathological phenotypic diversity and aggressive clinical behavior. We employed tandem mass tag (TMT) based proteomics technology and a novel Philosopher/TMT-Integrator computational pipeline that our group has recently developed and optimized for the analysis of large patient cohorts, to process the acquired mass spectrometry data. By unravelling the protein signatures within MBC and their relationship to TNBC and normal breast tissue, our study advances the understanding of the biology of MBC and provides potential diagnostic and prognostic markers, as well as testable targets of therapy specific to MBC pathological subtypes.

2-3 Methods

Tissue samples and histology and pathological evaluation

We employed frozen tissue samples from 14 women with MBC and adjacent normal tissues from the surgical pathology files at Johns Hopkins University, 6 TNBC and corresponding adjacent normal tissues from the surgical pathology files at the University of Michigan, with IRB approval. Each specimen was collected 30 minutes after operation and immediately transferred to

sterilized vials, snap frozen in liquid nitrogen and stored at –80 °C. Tumors were diagnosed according to the World Health Organization (WHO) classification into the following groups: spindle, squamous, and sarcomatoid (chondroid and/or osseous) MBC. Other pathological and clinical features such as estrogen and progesterone receptor analysis, Her2/Neu expression, Ki-67 proliferation index, tumor grade, tumor size, lymph node and distant metastasis were recorded in surgical pathology reports. At the time of sample preparation, we cut a representative 0.5 cm piece of each of MBC, TNBC and normal sample, which was embedded in paraffin, sectioned to 5 μm and stained with hematoxylin and eosin stained (H&E) for diagnosis confirmation.

Sample Preparation

Frozen specimens were kept on dry ice and cut to approximately 50 mg of tissue from each sample to be used for proteomic analyses. Specimens were diced and mechanically dissociated with a scalpel and placed in labeled 1.5ml Eppendorf tubes containing three stainless steel microbeads. Tubes were then submerged in liquid nitrogen for 60 seconds and immediately homogenized using a mixer mill (Retsch MM400) for 2-3 cycles at maximum speed (30 Hz vibrational frequency) at 60 seconds per cycle. If any tissue had not been homogenized, another cycle was repeated. Samples were then placed on ice and 495 μl of RIPA buffer and 5 μl protease inhibitor were added to each tube, resuspended, and placed on a rocker for 30 min on ice at 4°C. Beads were removed from tubes with a magnet and samples were ultrasonicated and centrifuged at 14,000 rpm for 30 minutes at 4°C. The supernatant was collected to retrieve a desired 2 mg/ml of protein, followed by standard protein quantification methods and storage at –80 °C.

Protein extraction and TMT labeling procedure

Tandem mass tag (TMT) labeling was performed for mass spectroscopy (MS) using three consecutive TMT-10plex isobaric labeling kits (ThermoFisher, Cat #90111) according to the

manufacturer's protocol. A master mix containing equal amount of protein from each of the twenty-seven samples was generated. 60 µg of protein from each sample and the master mix were reduced with DTT (5 mM) at 45 °C for 1 hr followed by alkylation with 2-chloroacetamide (15 mM) at room temperature (RT) for 30 min. Proteins were precipitated by adding six volumes of cold acetone and incubating overnight at –20 °C and pelleted by centrifuging at 8000 x g for 10 min at 4 °C. Supernatants were discarded and pellets resuspended in 100 µl of 100 mM TEAB, digested overnight at 37 °C by adding 1.1 µg of sequencing grade modified porcine trypsin (Promega, V5113). TMT reagents were reconstituted in 40 µl of anhydrous acetonitrile and digested peptides transferred to the TMT reagent vial, and incubated at RT for 1 hr. TMT channels for each of sample are given in Table 2. The reaction was quenched by adding 8 µl of 5% hydroxylamine and incubating for 15 min. For each of the three TMT experiments, 9 samples and 1 master mix (Table 2) were combined, dried, followed by 2D separation, with the first dimension containing an aliquot from each sample mix (100 µg) underwent fractionation using high pH reverse phase fractionation kit (following the manufacturer's protocol, Pierce). Fractions were then dried and reconstituted in 10 µl of loading buffer, 0.1% formic acid and 2% acetonitrile.

LC-MS/MS analysis

For data acquisition, an Orbitrap Fusion (ThermoFisher) and RSLC Ultimate 3000 nano-UPLC (Dionex) was used to obtain raw data. To increase accuracy and confidence in protein abundance measurements, a multinotch-MS3 method was employed for MS data analysis. 2 µl from each fraction were resolved in 2D on a nanocapillary reverse phase column (Acclaim PepMap C18, 2 micron, 75 µm i.d. x 50 cm, ThermoFisher) using a 0.1% formic/acetonitrile gradient at 300 nl/m (2-22% acetonitrile in 150m, 22-32% acetonitrile in 40m, 20 min wash at 90% followed by 50 min reequilibration) and directly sprayed onto Orbitrap Fusion with EasySpray

(ThermoFisher; Spray voltage (positive ion) = 1900V, Spray voltage (negative ion)= 600V, method duration= 180min, ion source type =NSI). The mass spectrometer was set to collect the MS1 scan (Orbitrap; 120 K resolution; AGC target $2x10^5$; max IT 100 ms), and then data-dependent "Top Speed" (3 s) MS2 scans (collision induced dissociation; ion trap; NCD 35; AGC $5x10^3$; max IT 100 ms). For multinotch-MS3, the top 10 precursor ions from each MS2 scan were fragmented by HCD followed by Orbitrap analysis (NCE 55; 60 K resolution; AGC $5x10^4$; max IT 120 ms; 100-500 m/z scan range).

Data analysis and protein quantification

Raw MS data were converted using msconvert in Proteowizard software suite[110] to mzML format. MS/MS spectra were searched using the MSFragger (v20181128) database search tool[111] against UniProt human protein database (UP000005640, last modified 2018-12-13; 73,928 proteins), appended with an equal number of decoy sequences and common contaminants. MS/MS spectra were searched using the following criteria: precursor-ion mass tolerance of 20 ppm, fragment mass tolerance of 0.6 Da (C12/C13 isotope errors (-1/0/1/2/3)), where cysteine carbamylation (+57.0215) and lysine TMT labeling (+229.1629) were specified as fixed modifications, and methionine oxidation (+15.9949), N-terminal protein acetylation (+42.0106), and TMT labeling of peptide N-terminus and serine residues were specified as variable modifications. The search was restricted to fully tryptic peptides, allowing up to two missed cleavage sites. MSFragger output files were processed using Philosopher toolkit (v20181128, github.com/Nesvilab/philosopher) as follows: The search results were first processed with PeptideProphet[112] (high-mass accuracy binning, semi-parametric mixture modeling options); ProteinProphet[113] was used to assemble peptides into proteins (protein inference) to create a combined file of high confidence proteins. Protein groups were filtered to 1% false discovery rate

(FDR) using the target-decoy strategy and the best peptide approach[98], [114], picked FDR adjustment[115]. The individual PSM lists for each TMT 10-plex we assembled, with peptides assigned either as a unique peptide to a particular protein group or assigned as a razor peptide[116] to a single protein group that had the most peptide evidence, and additionally filtered to 1% PSM-level FDR. For PSMs passing these filters, MS1 intensity of the corresponding precursor-ion was extracted using the Philosopher label-free quantification module based on the moFF method (10 ppm mass tolerance and 0.4 min retention time for extracted ion chromatogram peak tracing)[117]. For all PSMs corresponding to a TMT-labeled peptide, 10 TMT reporter ion intensities were extracted from the corresponding MS3 scans (using 0.002 Da window) and precursor ion purity scores were calculated using the intensity of the sequenced precursor ion and that of other interfering ions observed in MS1 data (within 0.7 Da isolation window).

Normalization and quality control

Three PSM tables from each TMT-10plex experiment generated as Philosopher output files were used as input to TMTIntegrator (v1.0.0; github.com/huiyinc/TMT-Integrator) for normalization and generation of integrative reports at the gene and protein level. For best quantitation quality, we only included PSMs (unique and razor peptides) that passed the following criteria: TMT label and quantification in the reference sample exist, minimum peptide probability 0.9, precursor ion purity >= 50%, minimum MS1 intensity 0.05%, summed MS2 intensity >= 5%. PSMs mapping to common contaminants were excluded, and for redundant PSMs, a single PSM of highest summed TMT intensity was kept. Then, intensities in each TMT channel were log2 transformed, and the reference channel intensity (pooled reference sample) was subtracted from that for the other nine channels (samples), thus converting the data into log2-based ratio to the reference scale. Here, the master mix TMT channel (131) was used as the reference channel. PSMs

were grouped by the corresponding gene/protein, and outlier removal using the interquartile range (1.5 IQR) algorithm was applied.

The gene-level median was calculated from the remaining PSM ratios, and then normalized using the median absolute deviation (MAD), given the p by n table of ratios for entry j in sample i, R_{ij}, the median ratio M_i = median(R_{ij}, j=1,…,p), and the global median across all n samples, M_0 = median(M_i, i=1,…,n). The ratios in each sample were median centered, $R^C_{ij} = R_{ij} - M_i$. The median absolute deviation of centered values in each sample was calculated, MAD_i = median(abs(R^C_{ij}), j=1…p), along with the global absolute deviation, MAD_0=median(MAD_i, i=1,…,n), and scaled to derive the final normalized ratios: $R^N_{ij} = (R^C_{ij}/ MAD_i) \times MAD_0$. Finally, normalized ratios were converted back to the absolute intensity scale using the estimated intensity of each entry in the reference sample. The reference intensity was estimated using the weighted sum of MS1 intensities of the top 3 most intense peptide ions [118], and in computing REF_i = mean(REF_{ik}, k = 1,…,q), missing intensity values were imputed with a global minimum intensity value. The final abundance (intensity) of entry i in sample j (log2 transformed) was computed as $A_{ij} = R^N_{ij} + \log_2(REF_i)$.

Missing data imputation

Upon inspection, experiments 1, 2, and 3 contained 18%, 14%, and 12% missing values, respectively. For missing data imputation, we used the multivariate imputation by chained equations algorithm using the statistical software 'mice' package (v3.4.0) in R is a gold standard method of choice to handle missing data and has been reported to produce lower estimate error rates compared to other methods (e.g. K nearest neighbors or singular value decomposition). We used the following input parameters: m=5 (number of imputed data sets), method='pmm' (predictive mean matching), and maxit=50 (number of iterations). Rubin's Rules (Rubin 1976 and

1987) were applied to pool estimates using logistic regression modeling. For statistical performance, we used significance testing of the categorical variables (i.e. patient expression values) to generate a covariance matrix containing pooled regression coefficient estimates, standard errors of parameter estimates, and p-values. We derived a pooled p-value < 1.6E-03 using the median p-value from the significance tests. The computational details are described in the original Rubin-Barnard approach[119].

Statistical analysis and subtyping of MBC

The batch corrected, imputed, combined data matrix was loaded and unsupervised PCA clustering was performed in R (v3.4.0) using standard k-means algorithms (e.g. Bayesian, Silhouette, Elbow) for finding optimal clusters, Cluster 3.0 was used for hierarchical clustering (median centering, uncentered correlation, and complete linkage), and visualization of results using Java Tree View (1.1.6r4). For differential expression analysis between tumors (MBC, TNBC) and normal samples, where data for each patient subgroup (e.g. spindle, triple-negative, normal) were pooled and results averaged, fold change was calculated by subtraction of averaged tumor subgroups to averaged control, and a Student's t-test was used to estimate $-\log_{10}$p-value. Proteins within a statistical region of FC>1 and p<0.05 were considered for analysis, including all proteins with FC>2 and p<0.01. For enrichment analyses, we performed gene ontology (GO) over-representation tests (GO annotations: biological process, molecular function, cellular compartment, protein domain) in PANTHER (v14.1), and STRING (v11.0) database for confirming enrichment results with topological features from protein-protein interaction networks among subgroups. For patient stratification tests, one-way ANOVA statistical analysis was performed between multiple groups and p<0.05 was considered significant.

Gene set enrichment analysis (GSEA)

GSEA is an aggregate score and running-sum statistic approach that enables molecular signature based statistical significance testing that considers the entire gene set containing a ranked list of all expression values in a data set without requiring a cutoff of differentially expressed values for functional analysis (Mootha et al., 2003; Subramanian et al., 2005). We supplied a pre-ranked list of two classes, up- and downregulated fold change values for each subgroup to be analyzed (MBC vs. TNBC, and across MBC subgroups: Spindle vs Squamous, Spindle vs Sarcomatoid, and Squamous vs. Sarcomatoid). To understand functional enrichment profiles of the proteomics results, we used annotated gene collections downloaded from the Molecular Signatures Database (MSigDB v7.0 for "H" (hallmark gene sets), "C2" (curated gene sets), and "C5" (GO gene sets). We determined normalized enrichment scores (NES) with the total protein list as background, and the following parameters: n=1000 permutations, where p-adjust < 0.05, and FDR < 0.05 were considered significant. The GSEA analysis was performed using the clusterProfiler and fgsea package in R and loading gene set collections from available gmt files from the BROAD Institute according to GSEA documentation (Subramanian, Tamayo 2005, Liberzon 2011).

Whole-exome sequencing (WES)

DNA samples with matched tumor and healthy counterparts from 10 pairs of samples in our patient cohort was subjected to whole exome sequencing using the NovaSeq 6000 Illumina system at the University of Michigan Advanced Genomics Core Facility in 150 bp paired-end format. Libraries were prepared used the NEBNext Ultra 2 FS DNA library prep kit for Illumina (NEB #E7805S; New England BioLabs) with 100ng DNA input, 15 min fragment, 275-475 bp size selection and 6 PCR cycles. Samples were captured with the IDT xGen hybridization capture

kit using 174ng of each library pooled for capture and a final PCR of 7 cycles. Fastq generation was performed using Illumina's bcl2fastq software version 2.20.

WES analysis and pathway enrichment

The exome sequencing data was analyzed by the variant calling pipeline developed by the University of Michigan Bioinformatics Core. FastQC v0.11.7 was used to assess the quality of raw reads, which were trimmed to remove Illumina adapters and low quality ends using Trimmomatic v0.39, aligned to the hg38 reference genome using BWA v0.7.17, followed by removal of sequence duplicates, SAM tag fixing, local realignment around INDELs, base quality score recalibration and target coverage summarization using GATK v4.1.4.0 or v3.8 (for indel realignment only). Normal-Tumor paired alignment files were submitted to GATK's MuTect2, Varscan v2.4.4, and Strelka v2.9.10 for the detection and filtration of somatic variants. Only the somatic variants on the canonical chromosomes that passed each caller's quality filter were kept. For Mutect2, variants were first called from the normal samples and a panel of normals (PoN) was created and used for somatic variant calling. Cross-sample contamination and orientation bias were estimated and used for variant filtration. For VarScan, an alignment coverage file was first created using SAMtools v1.5 from the pair of normal and tumor samples together, and then used for somatic variant calling. The selected high-confidence somatic variants were filtered using VarScan's fpfilter after allelic read counts were calculated using bam-readcount v0.8. For Strelka, structural variant/INDEL candidates were first called using Manta, as recommended by Strelka manual as a best practice.

Candidate variant calls across all samples and patients were merged using Jacquard v1.1.2 and included all variant loci whose filter field passed in MuTect2 or Strelka or VarScan (VarScan calls were limited to high confidence somatic variants confirmed in false-positive filter). Allele

frequency (AF) parameter from Strelka was calculated by Jacquard based on the allelic read depth reported by Strelka since Strelka does not report AF directly. Jacquard inferred consensus genotype (GT), and calculated average alternate allele frequency (AF), from individual caller's calling results. Variants were annotated using VarSeq v1.4.3. Variants excluding intergenic variants and common SNPs (SNPs whose allele frequency is higher than 5% in 1000 Genomes Phase 3 dataset) were integrated to create a gene-level variant and effect summary table using GeneRollup v0.3.2 (see additional files). Before filtering was applied, there were 11,652 total variants including those that fell in low complexity genomic regions, common SNPs, and duplicated regions. After filtering only out low complexity genomic regions, there were 980 total variants.

For enrichment analyses, gene lists of somatically mutated genes were generated for each MBC subtype and subjected to GO and pathway (KEGG, Panther, or Reactome) enrichment analysis using the WEB-based GEne SeT AnaLysis Toolkit (WebGestalt 2019, Liao, Zhang 2019). We applied BH correction, at least three genes per pathway, p-adjusted < 0.05 as significantly enriched, and used the human genome as the reference set.

Data Availability

All mass spectrometry proteomics data were deposited to the ProteomeXchange Consortium through the PRIDE partner repository.

2-4 Results

Human samples and clinical data

To elucidate the proteomic profile of MBC and understand the differences in protein expression with TNBC and normal breast tissues, we assembled a clinical cohort of 15 frozen MBC which were classified clinically according to their predominant metaplastic component into

the following subtypes: spindle (n=6), squamous (n=4), and sarcomatoid (n=5). In addition, we included 6 non metaplastic TNBCs and 6 normal adjacent breast tissues (**Table 1 and Fig. 2-1a**). All patients were women with a mean age of 55 years old (range 33 to 89 years old). The majority (14 of 15, 93.33%) of MBC and all TNBC were of histological grade 3, and all were negative for estrogen and progesterone receptor, and for HER2/neu overexpression. Of the 15 MBC, 11 (78.6%) were stage I/II and 3 (21.4%) stage III/IV at the time of diagnosis. Of the 6 TNBC, 5 (83.3%) were stage I/II and 1 (16.6%) stage III at diagnosis. At follow up, 4 of 14 (28.6%) MBC, and 1 of 6 (16.66%) TNBC developed distant metastasis to the lungs, liver, skin, and bone.

The proteome of metaplastic breast carcinoma

We leveraged the increased throughput of multiplexed TMT 10-plex proteomics and our automated and robust computational data analysis pipeline to generate a quantitative proteome profile of 27 human tissue samples (**Fig. 2-1b**). We arranged the samples into three experimental groups appropriate for the 10-plex TMT isobaric labeling strategy (n=10 samples per experiment; 9 tissue samples and 1 reference sample consisting of a pool of all 27 tissues) (**Table 2**). The three proteomic TMT 10-plex experiments identified 82,251, 84,667 and 84,386 peptides, respectively, to a depth of 5,798 unique proteins across all samples (1% protein and 1% PSM false discovery rate). We used the MSFragger and Philosopher tools (v20181128, github.com/Nesvilab/philosopher) for peptide identification, protein inference, FDR filtering, and extraction of quantification information form raw data, and TMT-Integrator for additional quality assessment and filtering, PSM selection, outlier removal, peptide-to-protein quantification roll-up, and normalization (**Fig. 2-1b**). We performed a gold standard data imputation method using the multivariate imputation by chained equations ("mice")[119] package in R, to preserve statistical power and sample size by producing unbiased estimates of the missing values. Next, we performed

standard batch correction in R, and the resulting expression matrix was used for all downstream analyses (**Fig. 2-2**).

Among the 5,798 proteins found in the MBC proteome, 5,635 unique proteins passed through quality filters in the TMT-Integrator algorithm, consistent with each sample. The distribution of all patient samples and principal component analysis (PCA) shows a clear distinction between normal breast and tumor proteomes (**Fig. 2-3a**). Unsupervised k-means clustering methods between MBC and TNBC proteomes and hierarchical clustering demonstrates modest distinction of MBC subtypes and TNBC, the latter overlapping most with spindle and squamous MBC (**Fig. 2-3b, Fig. 2-4**). In addition, 1 of the 15 MBC samples was excluded from downstream analyses since it was confirmed by histology and proteomics that the piece cut for analysis contained only normal tissue.

General features of the MBC proteome relative to normal tissues demonstrate a global downregulation program involving major tumor suppressors, extracellular matrix activities, and wound healing responses according to the clustering analysis (**Fig. 2-3b**), while enrichment analysis revealed the top GO molecular function for upregulated MBC proteins was procollagen-proline dioxygenase activity, and aldehyde dehydrogenase activity for downregulated proteins (**Fig. 2-5**). The MBC proteome establishes with functional relevance that MBC tumors and normal tissues have distinct protein profiles and exhibits deregulation of novel tumorigenic pathways.

Differential expression analysis of MBC proteome relative to TNBC and within MBC subtypes

Next, we sought to test the hypothesis that MBC histopathological subtypes are associated with specific proteomic signatures. We grouped patient samples according to histological MBC subtypes**,** where samples with mixed features were grouped in the subtype belonging to their

dominant feature (i.e. squamous with partial spindle features was grouped to squamous subgroup), and performed differential expression analysis of MBC relative to TNBC, and across MBC subtypes (**Fig. 2-6**). Compared to TNBC, MBC shows deregulation of the immune system (humoral immune responses; p<1-e04) and extracellular structure organization. Our analyses also revealed distinct functional processes among MBC subtypes, such as enriched keratinization (epidermal and keratinocyte differentiation) in squamous, regulation of proteolysis and protein activation cascade in spindle, and leukocyte activation and exocytosis in sarcomatoid. These results reveal common as well as distinct cellular differentiation profiles within the MBC proteome.

To gain a better understanding of the functional differences within MBC pathological subtypes and elucidate potentially unique protein signatures, we applied gene set enrichment (GSEA) analysis from the molecular signature database (MSigDB) using hallmark (**Fig. 2-7**), canonical pathways (**Fig. 2-8**) and GO gene sets (**Fig. 2-9**). Compared to TNBC, the top upregulated hallmark across MBCs is epithelial-to-mesenchymal transition (EMT) while oxidative phosphorylation is the top downregulated hallmark pathway (**Fig. 2-7**).

GSEA hallmark analyses revealed distinct up- and downregulated protein profiles among MBC subtypes (**Fig. 2-7**), and GSEA enrichment plots helped delineate specific differences in top pathways (**Fig. 2-10**). Comparison within pathological subtypes showed that spindle MBC has high MYC and E2F targets, and ribosome pathway proteins (KEGG pathways; **Fig. 2-8**), squamous MBC has high interferon gamma (and broad inflammatory responses), P53 and PI3K signaling, apical junction, and low OXPHOS, MYC and E2F targets, while sarcomatoid MBC has high EMT and OXPHOS, and low interferon gamma, MTORC1, and PI3K signaling (**Fig. 2-7 and 2-10**). Visualization of top enriched up- and downregulated terms and their associated proteins are

shown by protein networks (**Fig. 2-11**) which can be investigated as potential therapeutic candidates. These up- and downregulated protein signatures were further validated by additional differential expression analyses of each MBC relative to TNBC (**Fig. 2-12**). Together, these analyses pinpoint specific pathways operating in MBC compared to TNBC, and highlight MBC subtype specific pathways.

Patient stratification analysis: personalized signatures in MBC and TNBC tumors

To extend our proteomics approach to the individual patient level, we carried out differential expression analysis on hallmark pathways and makers of molecular subtypes in each tumor. Stratification of patient protein profiles demonstrate upregulation of EMT and cancer stem cell (CSC) proteins across MBC subtypes and TNBC, coupled with downregulation of stromal/stem cell proteins and endothelial-to-mesenchymal transition (EndMT; CDH5, IGFBP4, CD34) (**Fig. 2-13**). Stratification of patient protein profiles according to the expression of claudins, basal, and luminal proteins shows that spindle MBC is claudin-low, basal-low, and luminal-low, squamous MBC is claudin-low, basal-high, and luminal-low, sarcomatoid is claudin-low, basal-variable, and luminal-low, and TNBC is claudin-variable, basal-low, and luminal-high (**Fig. 2-13b**). These results show that a combination of claudins, basal, and luminal markers are able to stratify individual patient's tumors and may be useful for personalized strategies.

Patient-stratified differential expression analysis according to expression of our recently identified MBC markers, HMGA2 and IGF2BP2[108] shows that they are concordantly upregulated in spindle MBC (**Fig. 2-13c**). These data provide evidence that patient-stratified analyses distinguish MBC subtypes, and provide independent validation for using HMGA2 and IGF2BP2 diagnostically and as a therapeutic opportunity for spindle MBC.

Whole exome sequencing within human MBC and subtype-specific mutational signatures

Based on our discovery of subtype-specific hallmark protein pathways in MBC, we set out to elucidate the hypothesis that each pathway may exhibit distinct DNA mutational profiles. Out of the 15 MBC tumors, 10 pairs of tumors with their normal tissue counterparts passed initial quality control for whole exome sequencing (**Fig. 2-14**). Most of the variants observed were intronic (Fig. 6a), and when filtering only missense and loss-of-function or "LoF" (Fig. 6b), we found that MBCs shared frequent somatic mutations in TP53 (70%), MUC17 (60%), PLEC (30%), CRYBG2 (40%), and ZNF681 (30%) (**Fig. 2-14c**). Spindle and squamous shared AHNAK mutations (AHNAK, AHNAK2 (80%) in spindle; AHNAK (33%), AHNAK2 (67%) in squamous) and in PI3K genes (PIK3C2A (20%) in spindle; PIK3CA (33%) in squamous). Squamous MBC also harbored mutations MTOR, NOTCH3, and PTEN (33%). Sarcomatoid MBC shows genetic alterations in cadherin, calcium ion and WNT signaling, with frequent mutations in the protocadherin gene cluster (PCDH family), and CDH7, MAP3K2, and FAT1 (50%) (**Fig. 2-14b and e**). In addition, allele frequencies are similar for variants across MBC subtypes (**Fig. 2-14d**). Taken together, these data show that MBCs share mutations in five genes, with only TP53 having been reported previously [54], [55]. These data also reveal that sarcomatoid MBC has a distinct mutational profile from spindle and squamous MBC, highlighting the importance of the proteomics landscape in distinguishing MBC pathological subtypes, which paves the way for mechanistic and functional studies elucidating the biology of these tumors.

Comparison of human MBC with MMTV-Cre;Ccn6$^{fl/fl}$ MBC mouse model defines a protein signature of spindle MBC

To further refine the protein landscape of spindle MBC and determine common deregulated proteins between human and mouse tumors, we performed quantitative proteomics on MMTV-Cre;Ccn6$^{fl/fl}$ spindle MBC tumors followed by comparison with human MBC proteomes (**Fig. 2-15a-b, Fig. 2-16**). Compared to normal mouse mammary glands, MMTV-Cre;Ccn6$^{fl/fl}$ spindle MBCs showed increased MYC, E2F, unfolded protein response and ribosomal pathway proteins, which are similar to human spindle MBC (**Fig. 2-7**).

Based on shared pathology, metastatic ability, and hallmark pathways, we hypothesized that human and mouse spindle MBC may have an overlapping protein signature. Enrichment analyses in GO annotation and STRING databases uncovered a set of 36 proteins, 17 upregulated and 19 downregulated that overlapped between human and mouse spindle MBC proteome (**Fig. 2-15c**). The shared downregulated proteins have roles in metabolic processes including oxidation-reduction, carboxylic acid and ethanol metabolism (e.g. ALDH1L1, ADH1B, ADH1C, SOD1, LIPE, FABP4). Among the upregulated proteins are those involved in ribosomal function, translation, and RNA metabolism (e.g. RPL18A, RPL18, RPL6). In addition, we found a highly enriched protein-protein interaction network among the upregulated proteins ($p < 1e\text{-}16$). Together, these data delineate significantly deregulated pathways in spindle MBC in human and mouse tumors and nominate a subset of proteins that may be useful markers and targets of therapy.

2-5 Discussion

Here we present a quantitative proteomic landscape of human MBC, a subtype of TNBC with a defining histology, frequent chemoresistance, and distant metastases[120]. Using tissue resources from patients and from an MBC spindle mouse model, robust proteomics and novel bioinformatics tools we demonstrate shared and subtype-specific altered proteomes present in

spindle, squamous and sarcomatoid MBC, providing insights into the biology of this aggressive form of breast cancer and offering opportunities for precision medicine.

Despite the impact of protein expression on tumor phenotypes and clinical behavior, our knowledge on the protein landscapes of human breast cancer, and how differential profiles contribute to breast cancer phenotypes is very limited. Recent advances in proteomic technology and bioinformatics have enabled detailed characterization of breast cancer[19], [95]. However, studies to date have focused on frequent subtypes of breast cancer, with few to no MBC cases and normal breast tissues.

Upon in depth cross-analyses of MBC to both TNBC and normal tissues, and between each MBC subtype using GO and MSigDB databases, we uncovered common and subtype-specific differences among MBCs and their relationship with TNBC. Compared to TNBC, the proteome of MBC has a highly enriched EMT phenotype, with increased inflammatory responses (mainly in spindle and squamous), an active ECM, and reduced oxidative phosphorylation. These data are consistent with our previous studies showing that MBC express proteins involved in the EMT process, which may contribute to a more stem-like and aggressive phenotype than TNBC, and with transcriptome studies showing that MBCs cluster as basal-like and claudin-low.[57], [121] Subtype-specific proteins and pathways emerged when we compared proteomic profiles of spindle, squamous and sarcomatoid tumors. A salient feature of spindle MBCs, which are highly proliferative and exhibit pathological evidence of EMT with elongated cancer cells, is upregulated expression of E2F and MYC pathway proteins, ribosomal proteins and transcriptional and translational events. These findings are intriguing based on data showing that E2F and MYC are key drivers of cancer cell proliferation and EMT processes, where ribosome biogenesis and increased translation play an essential role.[122], [123] These data also suggest that spindle MBC

have a deregulated balance between translation and metabolic pathways which needs to be further investigated

Our data show that squamous MBC has upregulated inflammatory responses (e.g. IFN-γ, TNFα, PI3K/MTOR), keratinization, and widespread cell adhesion marker expression (e.g. apical junction, adherens, CAMs and cytoskeleton proteins), and decreased oxidative phosphorylation, MYC and E2F pathways relative to other MBC subtypes. On the other hand, sarcomatoid MBC, which includes chondroid and osseous differentiation, exhibit a predominance in extracellular matrix signaling cascades and an amplified EMT program, increased oxidative phosphorylation, and decreased inflammatory responses compared to both spindle and squamous subtypes, likely owing to its differentiation along mesenchymal lineages[124]. Also, we observed that in general, there is no significant difference in P53 and PI3K pathway between MBCs and TNBC. Collectively, our proteomic analysis suggests that while spindle, squamous and sarcomatoid MBCs and TNBC may share initial neoplastic events, each MBC subtype appears to have unique and active differentiation programs.

MBC are chemoresistant and metastatic, but the underlying molecular events and driver pathways are unclear. Further, there are no effective treatments against these tumors.[48], [52] The heterogeneity of MBC has been investigated at the genetic level[58], [60], [125], however the relationship between genomic alterations with proteomic profiles is unknown. While MBC were reported to harbor somatic mutations in TP53, PI3K/MTOR and Wnt signaling pathway genes,[60] no subtype-specific mutational profiles in MBCs have been shown to date. Our whole exome sequencing analyses of 10 paired tissue samples of MBC and normal breast tissue from the same patients identified somatic mutations common to all MBCs in five genes; TP53 in 70%, MUC17 in 60%, CRYBG2 in 40%, PLEC in 30%, and ZNF681 in 30%. Of these, mutations in MUC17,

CRYBG2, PLEC and ZNF681 have not previously been reported in MBC. We found that while spindle and squamous MBC exhibit overlapping mutational profiles of genes involved in transcription, RNA metabolic processes and actin filament binding, sarcomatoid tumors harbor distinct mutations, especially in MAPK, WNT, protocadherin cluster genes, calcium binding and ECM organization. Taken together, these results demonstrate novel somatic mutations in MBC tumors, distinct mutational profiles of sarcomatoid MBC compared to the overlapping landscape of spindle and squamous tumors. These data underscore the relevance of elucidating the proteomic landscape to nominate subtype-specific proteins and pathways, especially between tumors with spindle and squamous differentiation, that may lead to treatment targets.

Quantitative proteomics of our model of spindle MBC MMTV-Cre;Ccn6$^{fl/fl}$ tumors demonstrated a significant overlap with human spindle MBCs, with shared enrichment in E2F, MYC, EMT, and ribosomal proteins. Our analyses further pinpoint a set of 36 proteins commonly deregulated in mouse and human spindle MBCs which have not been previously considered in these tumors. The 19 downregulated proteins have functions in metabolic processes, while of the upregulated the majority are ribosomal proteins, and those involved in transcription and translation. Together these data validate the MMTV-Cre;Ccn6$^{fl/fl}$ model as a useful tool to investigate and test new therapeutic targets for spindle MBC, and nominate a set of significantly deregulated proteins and pathways as therapeutically operable targets. Our studies may pave the way towards mechanistic and functional investigations of these proteins in the biology of spindle MBC.

In addition to subtype-specific tumor signatures, our patient stratification analysis uncovered a stromal cell expression pattern in MBCs, with the most prominent being downregulated expression of mesenchymal stem cell proteins (CD59, CD248), macrophages

(CD209), immune cells (CD8A), hematopoietic stem / endothelial progenitors / macrophages (CD34, CD36, CDH5), a recently discovered mammary stem cell marker (CDH5)[126], and we detected evidence of endothelial-to-mesenchymal transition (EndMT; $CDH5^{lo}IGFBP4^{lo}CD34^{lo}$). The stromal profile is intriguing and has not been previously considered in MBC. Furthermore the identification of downregulated EndMT proteins is intriguing, as this process[127] has not been previously considered in MBC. These findings lead to the hypothesis that downregulation of these protein pathways, which include inflammatory mediators, may contribute to the therapeutic resistance characteristic of MBC or the promotion of aberrant differentiation programs. We also pinpointed previously unconsidered differences in claudin, basal and luminal marker expression in MBC and TNBC, highlighting the relevance of individual patient stratification when considering heterogeneous malignancies. Further, we found concordant upregulation of the CCN6 pathway proteins HMGA2 and IGF2BP2 in spindle MBC[108], which underscores the role of this pathway in the biology of this MBC subtype, and the advantage of personalized proteomics.

In summary, this patient cohort identifies a common proteomic landscape of MBC in relation to TNBC, and highlights the existence of specific protein profiles underlying the different histopathological subtypes of human MBC. We show that quantitative proteomics refines the mutational landscape of MBCs and allows for further distinction of the tumor subtypes. We provide evidence for a significant overlap between MMTV-Cre;$Ccn6^{fl/fl}$ spindle MBC mouse model and human disease, which together with a common histopathology and transcriptome, validate this mouse model to test new treatments and mechanistic investigations on the development of spindle MBC. The subtype-specific proteomes of MBC tumors emphasize a unique opportunity to impact precision therapies to improve the survival of women with this aggressive form of breast cancer.

2-6 Figures

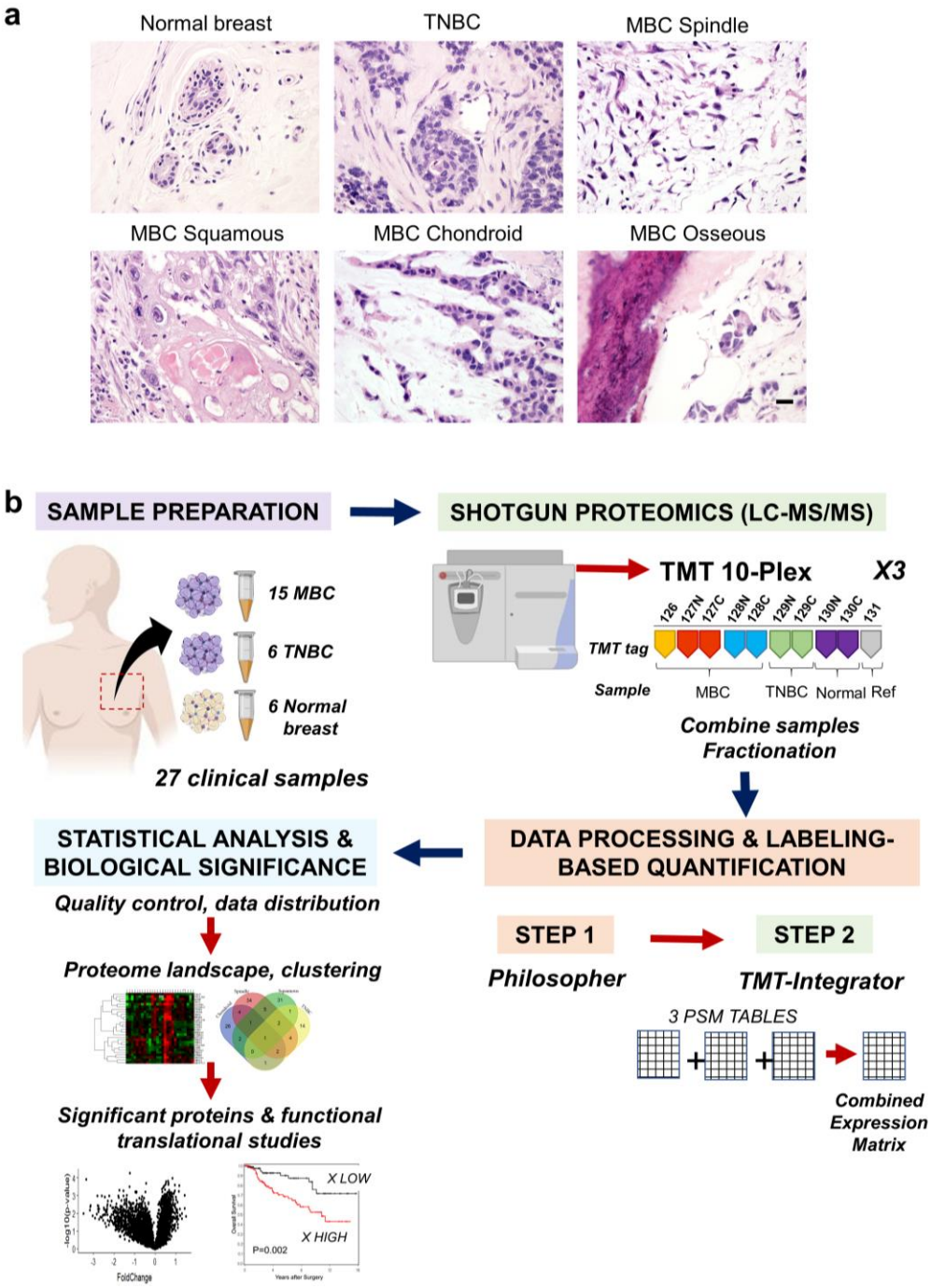
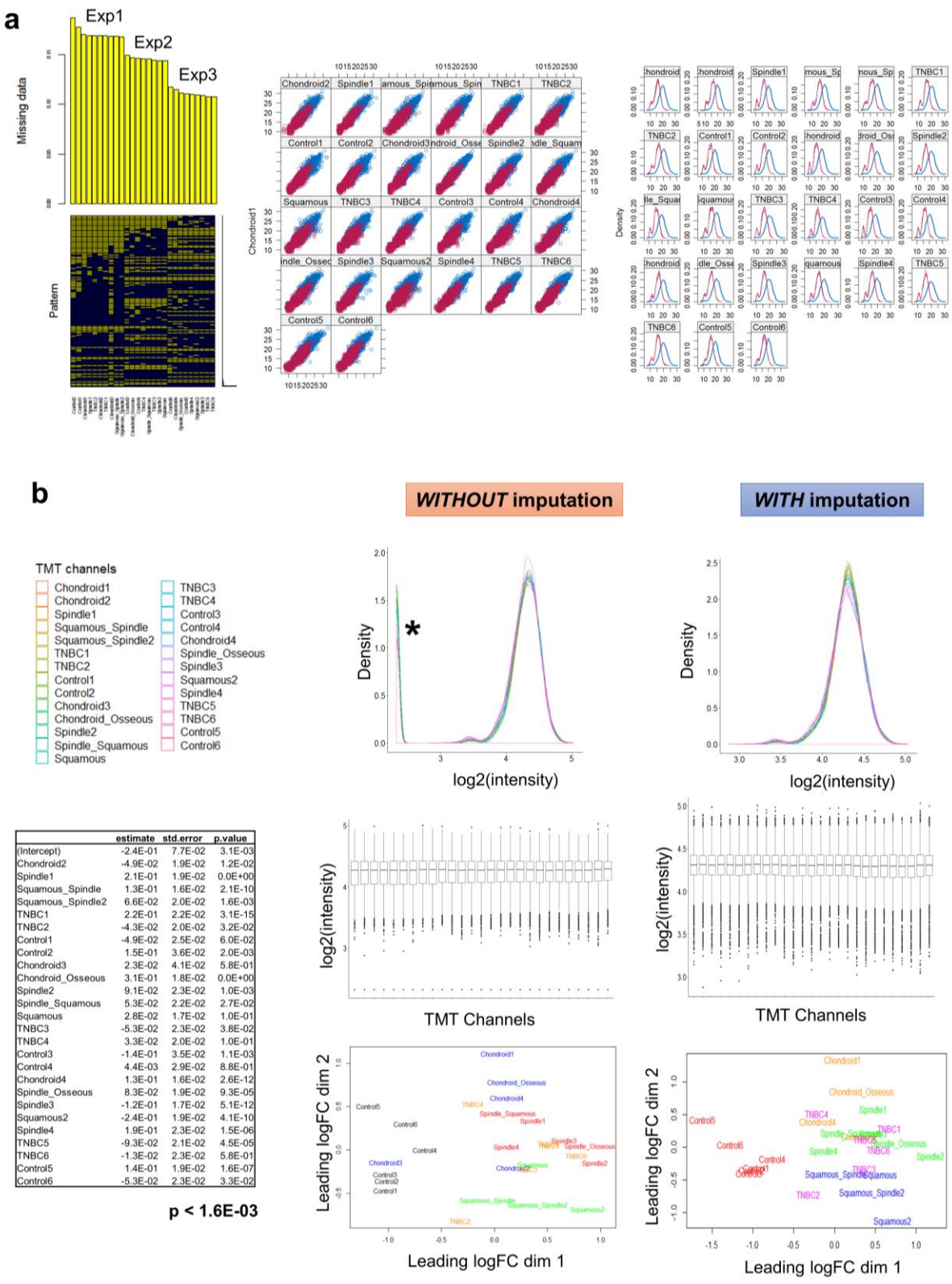


Figure 2-1. Human clinical samples and quantitative proteomics workflow.
a. Representative images of hematoxylin and eosin stained tissues from our patient cohort, including 15 metaplastic carcinomas (MBC, 6 spindle, 4 squamous, and 4 sarcomatoid), 6 triple-negative (TNBC), and 6 normal breast. Scale bar = 50 μm.
b. Workflow of quantitative mass spectrometry profiling. For data acquisition we assembled the 27 samples into 3 experimental groups for 10-plex LC-MS/MS tandem mass tag (TMT) isobaric

labeling. For data processing and quantification, we used two novel pipelines, Philosopher/TMT-Integrator, and generated a combined protein expression matrix for the 3 experiments used for downstream analyses, including hierarchical clustering, differential expression tests, statistical analysis and biological inference.



	estimate	std.error	p.value
(Intercept)	-2.4E-01	7.7E-02	3.1E-03
Chondroid2	-4.9E-02	1.9E-02	1.2E-02
Spindle1	2.1E-01	1.9E-02	0.0E+00
Squamous_Spindle	1.3E-01	1.6E-02	2.1E-10
Squamous_Spindle2	6.6E-02	2.0E-02	1.6E-03
TNBC1	2.2E-01	2.2E-02	3.1E-15
TNBC2	-4.3E-02	2.0E-02	3.2E-02
Control1	-4.9E-02	2.5E-02	6.0E-02
Control2	1.5E-01	3.6E-02	2.0E-03
Chondroid3	2.3E-02	4.1E-02	5.8E-01
Chondroid_Osseous	3.1E-01	1.8E-02	0.0E+00
Spindle2	9.1E-02	2.3E-02	1.0E-03
Spindle_Squamous	5.3E-02	2.2E-02	2.7E-02
Squamous	2.8E-02	1.7E-02	1.0E-01
TNBC3	-5.3E-02	2.3E-02	3.8E-02
TNBC4	3.3E-02	2.0E-02	1.0E-01
Control3	-1.4E-01	3.5E-02	1.1E-03
Control4	4.4E-03	2.9E-02	8.8E-01
Chondroid4	1.3E-01	1.6E-02	2.6E-12
Spindle_Osseous	8.3E-02	1.9E-02	9.3E-05
Spindle3	-1.2E-01	1.7E-02	5.1E-12
Squamous2	-2.4E-01	1.9E-02	4.1E-10
Spindle4	1.9E-01	2.3E-02	1.5E-06
TNBC5	-9.3E-02	2.1E-02	4.5E-05
TNBC6	-1.3E-02	2.3E-02	5.8E-01
Control5	1.4E-01	1.9E-02	1.6E-07
Control6	-5.3E-02	2.3E-02	3.3E-02

Figure 2-2. Pre-processing of LC-MS/MS TMT 10-plex proteomics data before and after data imputation and batch correction. a. Missing data percentages and patterns in each of the

three experiments 1, 2, and 3 (Exp = experiment) contained 18%, 14%, and 12% missing values, respectively. Scatter plots (middle) show one variable against all others (i.e. 26 comparisons between patients) and right plots are densities of imputed data against experimental data. The 'mice' package was used in R (v3.4.0) with the parameters: m=5 (number of imputed data sets), method=pmm (predictive mean matching), and maxit=50 (number of iterations). The multivariate imputation by chained equations method is a gold standard data imputation algorithm that accounts for uncertainty in the data when values are missing at random following Rubin's Rules (Rubin 1976 and 1987).
b. Comparison of the raw data distribution of 27 patient samples with and without data imputation. Density distribution plots (top and middle) showing log2-transformed protein abundances for all TMT channels of 3 experiments, and an MDS scatter plot (bottom) of computed distances between protein expression profiles of all samples. The asterisk (*) indicates the peak of missing values without data imputation. The table shows statistical performance (pooled estimates, p-values) of imputed data sets after fitting to a linear model and pooling the data to attain a median p-value<1.6E-03. The MS/MS spectra were first searched using MSFragger (v20181128) database search tool against UniProt human protein database (UP000005640), followed by processing by PeptideProphet with high-mass accuracy binning and semi-parametric mixture modeling to compute correct identification for each peptide-to-spectrum (PSM) searching. PSM lists were created for each TMT 10-plex experiment, and protein groups were assembled in ProteinProphet filtered to 1% false discovery rate (FDR). PSMs passing through filters in the Philosopher label-free quantification module corresponded to TMT reporter ion intensities, and used as input to TMT-Integrator for normalization to the reference channel and log2 transformation.

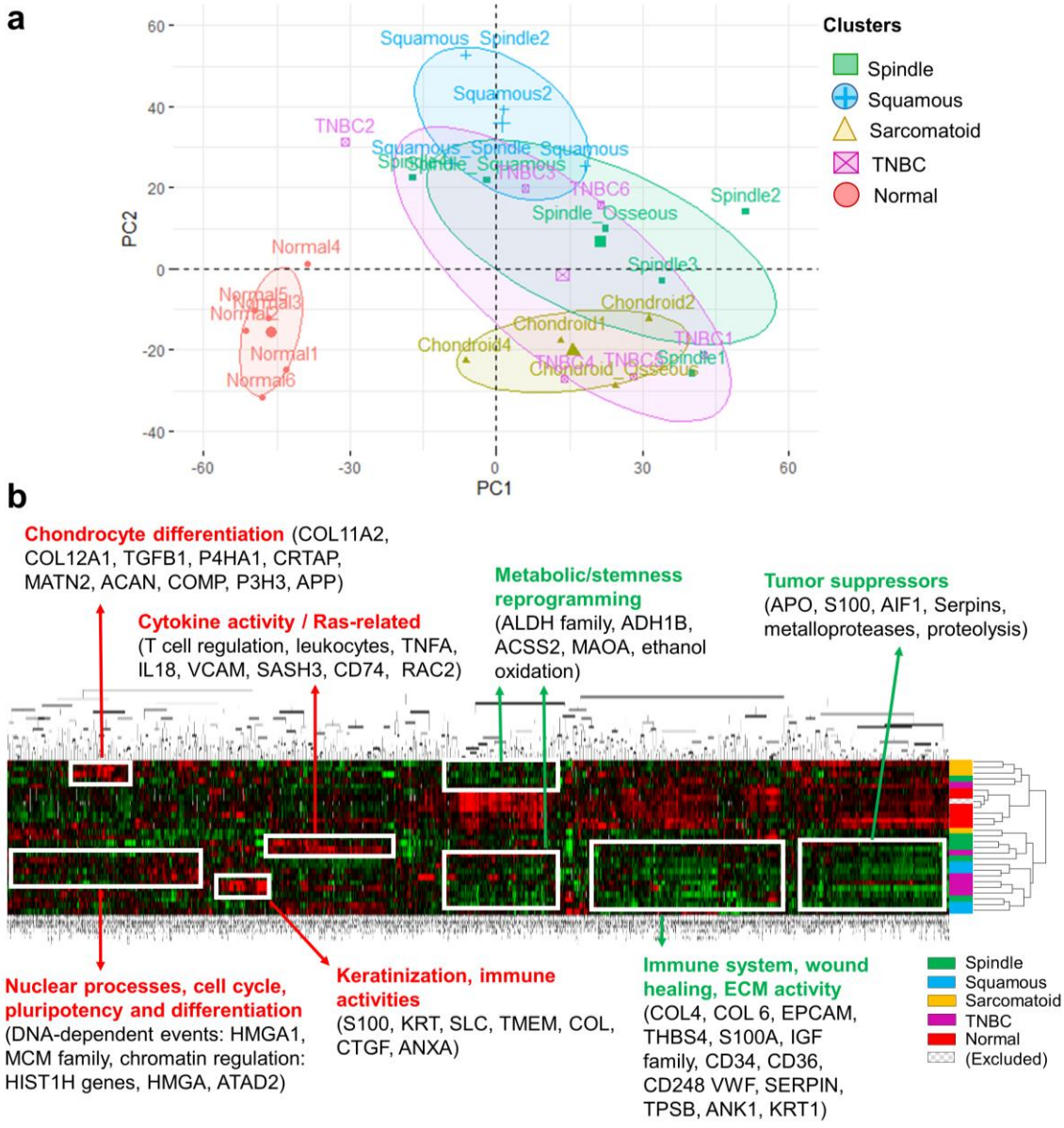
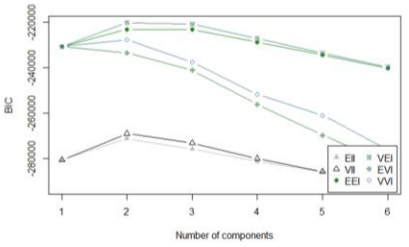
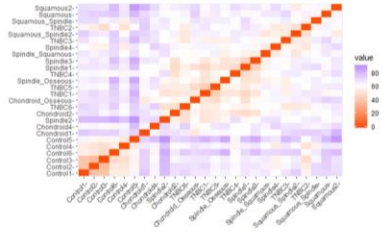
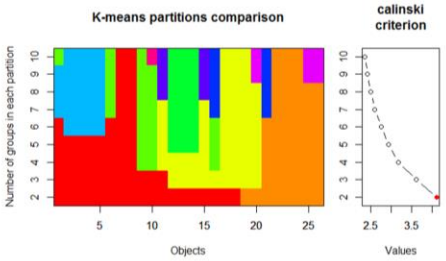


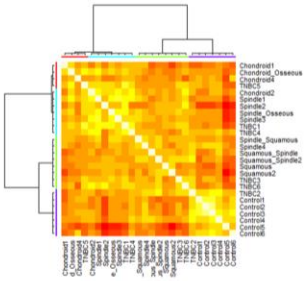
Figure 2-3. Quantitative proteomics of MBC, TNBC, and normal breast tissues.
a. Principal component analysis (PCA) plot shows unsupervised clustering among the 27 samples, demonstrating a clear distinction between normal and all tumors (MBC subtypes and TNBC) and between MBC squamous and sarcomatoid, while there was an overlap between MBC subtypes and TNBC.
b. Heat map of the 5,670 proteins that passed through quality filters in TMT-Integrator across all patients. Significantly enriched downregulated (green) and upregulated (red) GO biological processes ($p<0.05$). Dendogram from hierarchical clustering analysis in Cluster 3.0 using median centering, uncentered correlation, and complete linkage, and visualized in Java TreeView 1.1.6r4.

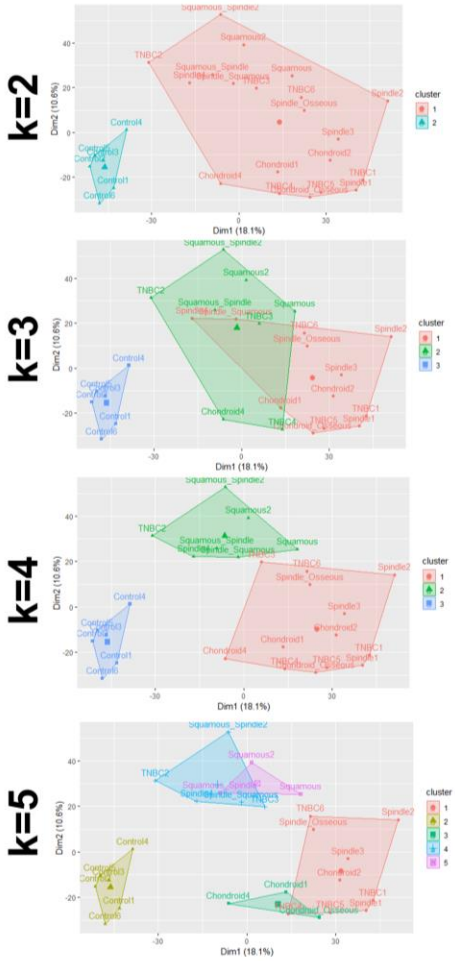
Method	Optimal # Clusters
Bayesian	2
WSS	2
Vegan	2
Elbow	3
Gap Statistic	3
Silhouette	4
Apcluster	4

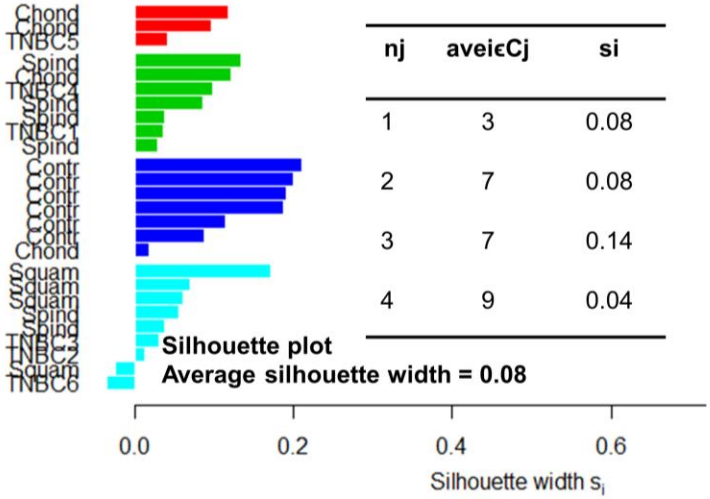












nj	aveiϵCj	si
1	3	0.08
2	7	0.08
3	7	0.14
4	9	0.04

k=2	k=3	k=4	k=5
Normal Tumor	Normal MBC+TN MBC+TN	Normal Spindle + TN Squamous + TN Chondroid	Normal Spindle + TN Squamous Chondroid Mixed

Figure 2-4. Summary of unsupervised k-means methods for determining optimal clusters. Table shows summary of results from seven different algorithms used in R (Bayesian, within-cluster sum of squares (WSS), Vegan, Elbow, Gap statistic, Silhouette, and Apcluster) for

determining the optimal number of clusters among our cohort of patient tumor samples. Representative plots (top right) show hierarchical clustering heatmaps and a range of 2-4 optimal clusters found. Representative silhouette plot with 4 clusters shows an average silhouette width of 0.08, where silhouette ranges from -1 to 1. PCA plots (left) and table of corresponding histological subtypes (bottom right) demonstrate a distinction between normal and tumor (MBC and TNBC) with 2-3 clusters present, and a distinction between both normal and MBC subtypes (spindle, squamous, sarcomatoid) with 4-5 clusters, where triple-negative (TNBC) clustered with spindle or squamous MBC.

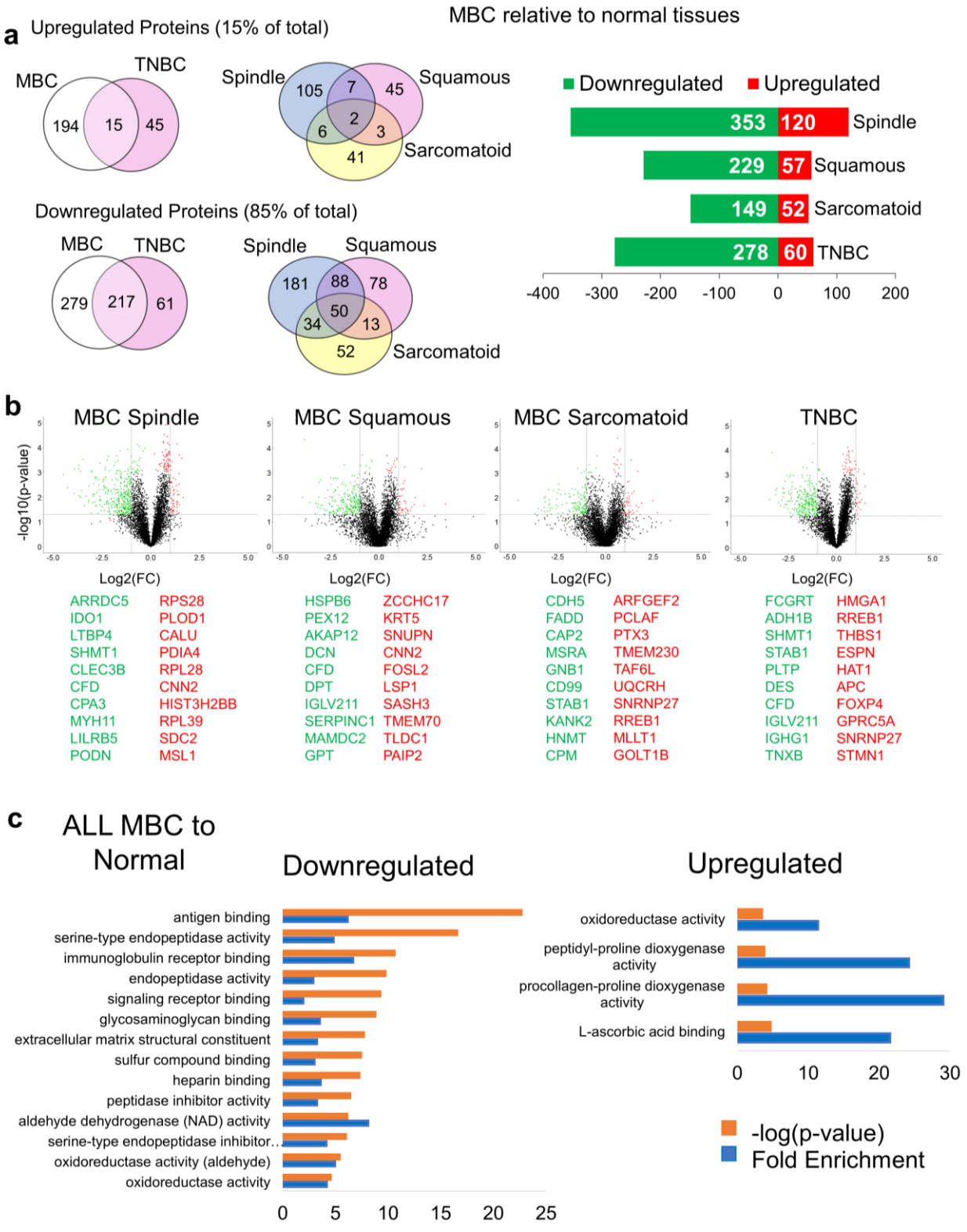


Figure 2-5. Supervised differential expression analysis reveals unique MBC protein signatures by histological subtype relative to normal breast. a. Total differentially expressed proteins in MBC and TNBC relative to normal tissues. Patient samples were grouped according

to their predominant histological MBC subtype (spindle, squamous, and sarcomatoid), and considered proteins of $p<0.05$ and $FC>1$, and highly significant outliers of $p<0.001$.
b. Statistical analysis shows volcano plots for MBC spindle, squamous and sarcomatoid and triple-negative (TNBC) subtypes as log2-Fold change versus –log10(p-value), where proteins indicated in green are downregulated and red indicates upregulated proteins. Horizontal boundary represents $p=0.05$ and two vertical boundaries indicate $FC = 1$ and -1. The list of the top 10 most significant up- and downregulated proteins are shown.
c. Enrichment analyses of all MBC relative to normal breast tissues demonstrating up- and downregulated enriched terms from gene ontology (GO) using PANTHER (v14.1) database. Fold enrichment scores show the GO annotation for biological process and the –log(p-value).

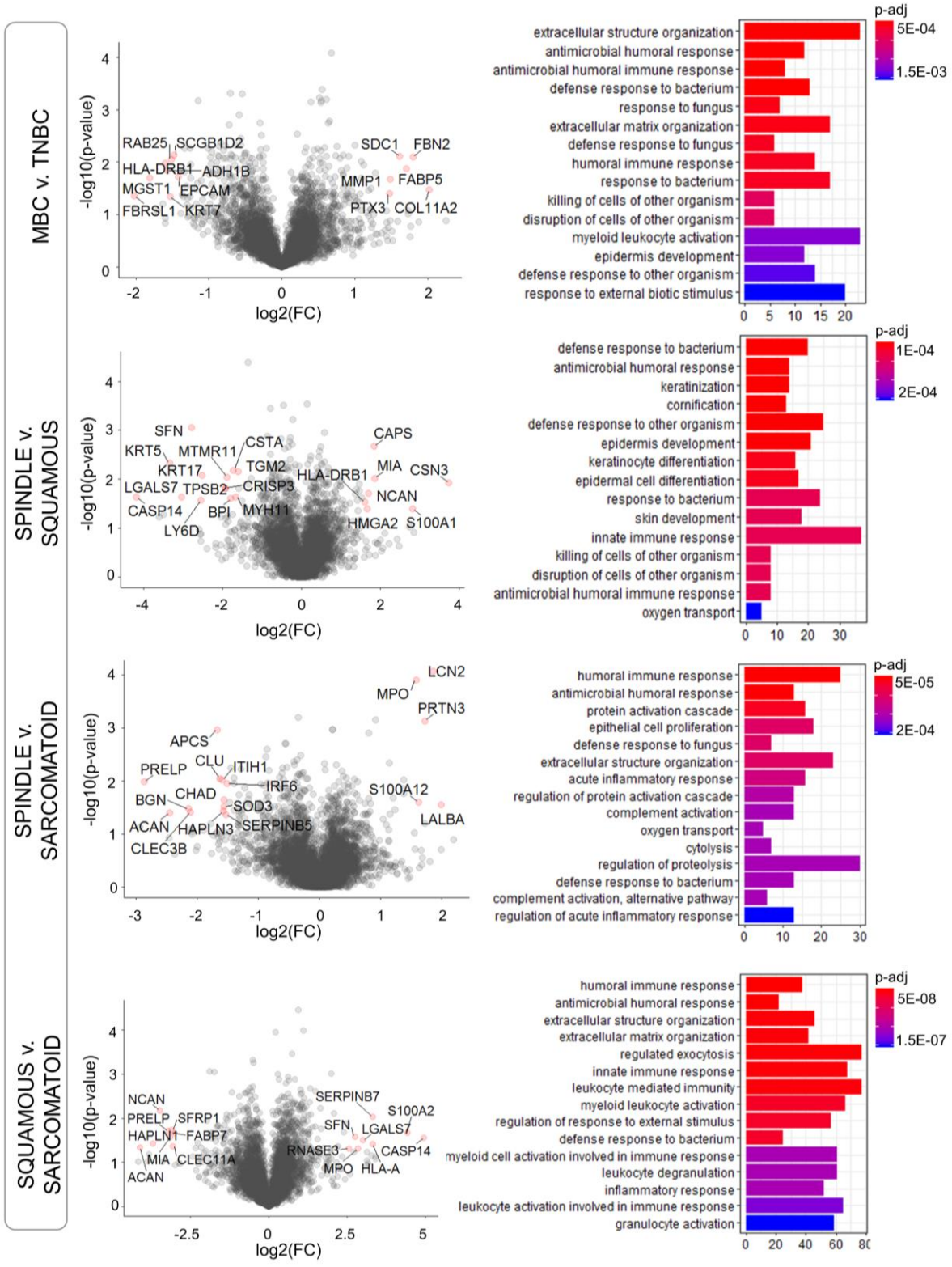


Figure 2-6. Differential expression analysis shows deregulated proteins and pathways within histological subtypes of MBC and in relationship to TNBC.

Left, Volcano plots comparing MBC with TNBC and within MBC subtypes, as indicated. Significantly differentially expressed proteins are highlighted in red. log2-Fold change versus – log10(p-value), where cutoff values on plots show FC > 1 or. FC < -1 (vertical lines) and p < 0.01 (horizontal line), n = 20 patients.
Right, Enrichment analysis using gene ontology (GO) annotations showing the top GO terms based on biological process, molecular function, or cellular compartment. Significant proteins were considered using p-value<0.05, q-value<0.1, fold change (log2-FC) greater than 1, Benjamini-Hochberg correction ("BH"), gene set size 5-500, and the total protein list (5670 proteins) as the background set. Barplot shows significant terms by the gradient legend as padjust < 0.001, the x-axis is gene ratio and the number of proteins belonging to given enriched terms.

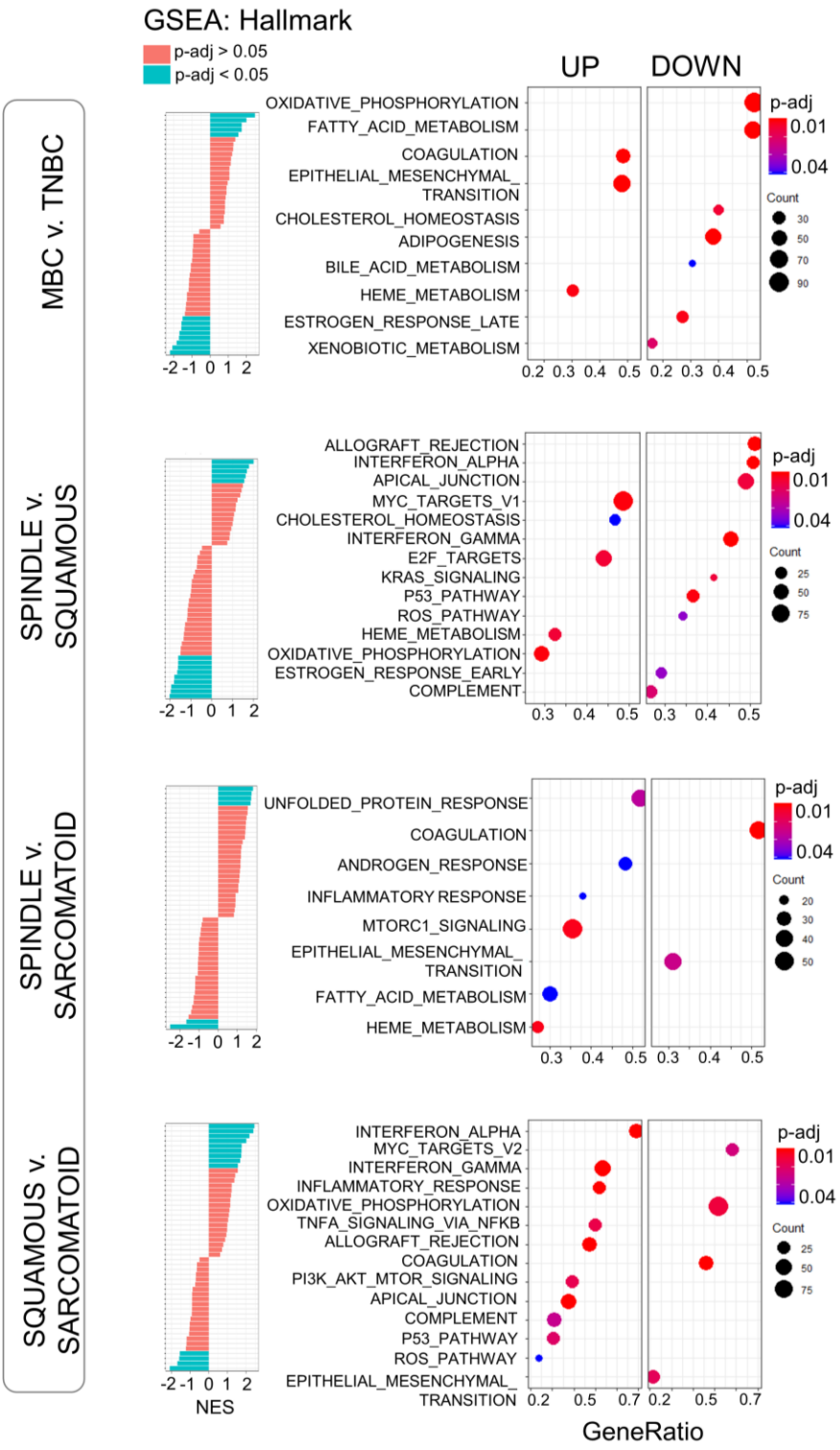


Figure 2-7. Gene set enrichment analysis (GSEA) reveals distinct hallmark pathways within MBC subtypes and relative to TNBC.
GSEA analysis show differentially expressed protein pathways in MBC compared to TNBC, and within MBC subtypes. Left, normalized enrichment scores (NES) versus the total list of GSEA

hallmark categories of up- and downregulated (marked “UP” and “DOWN”, respectively) significantly enriched terms (p-adjust <0.05) among non-enriched (p-adjust > 0.05). Right, top hallmark pathways highlighting unique significantly expressed categories. The GSEA analysis was performed using the clusterProfiler and fgsea package in R for the hallmark collection (“H”) (Broad Institute), with n=1000 permutations, where p-adjust < 0.05 and FDR < 0.05 were considered significant.

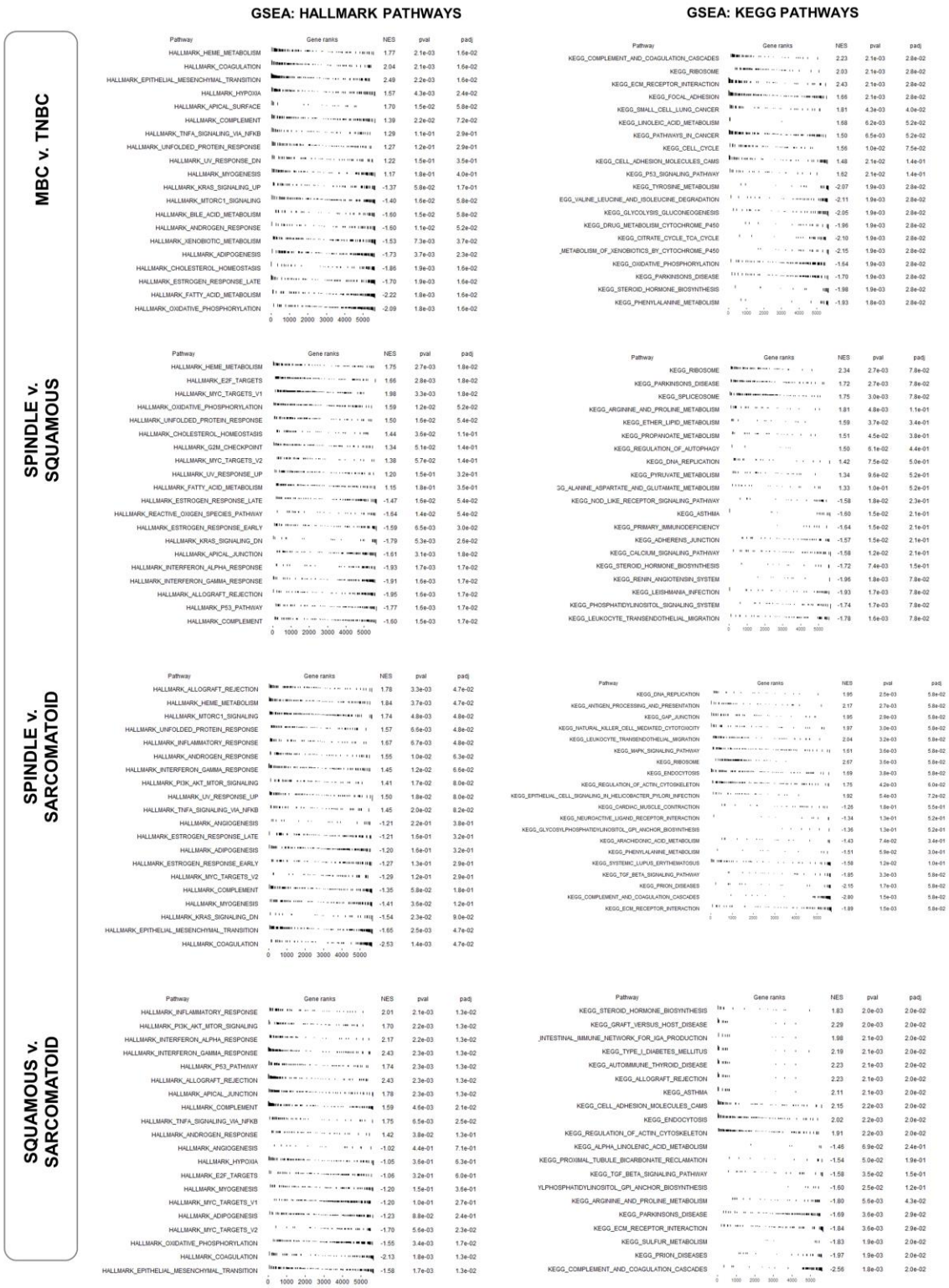
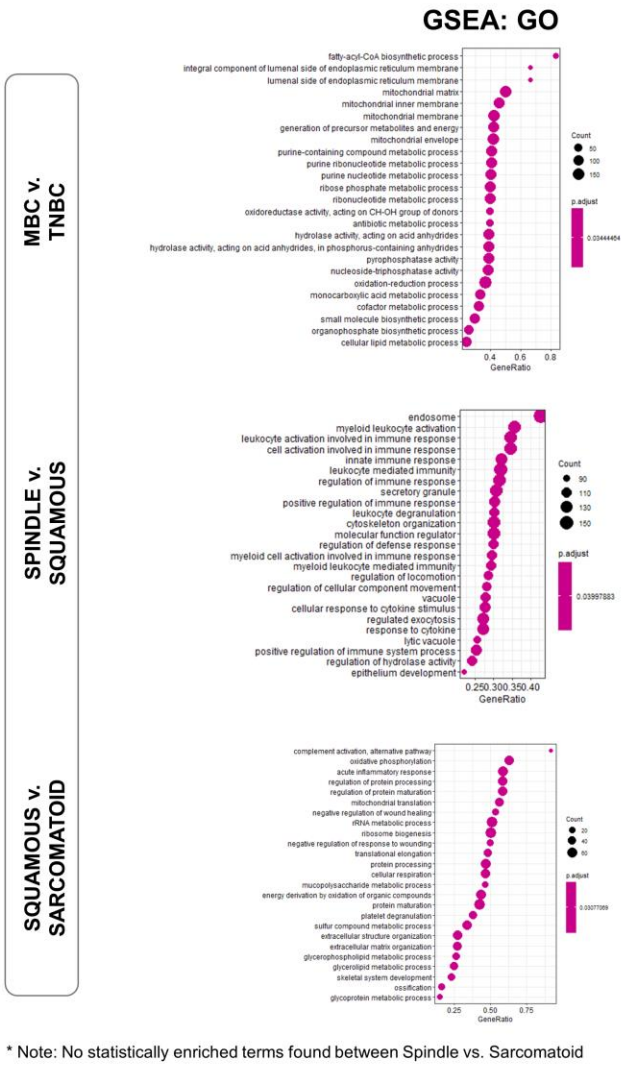


Figure 2-8. Top pathways of gene set enrichment analysis (GSEA) in Hallmark and KEGG gene sets within MBC and relative to TNBC. Ranked lists of gene sets were previously attained from differential expression analysis for each disease condition (MBC vs. TNBC, and across MBC subgroups (Spindle vs Squamous, Spindle vs Sarcomatoid, and Squamous vs.

Sarcomatoid) to run the GSEA analysis using MSigDB against hallmark gene sets and KEGG (Kyoto Encyclopedia of Genes and Genomes). Tables show top pathways, gene ranks, normalized enrichment score (NES), pvalue, and padjust value, where $p<0.05$ was considered significant.



* Note: No statistically enriched terms found between Spindle vs. Sarcomatoid

Figure 2-9. GSEA analysis of curated GO gene sets ("C5") from MSigDB for MBC relative to TNBC, spindle vs squamous, and squamous vs. sarcomatoid. Bar plot shows the total number of differentially expressed up- and downregulated proteins in each category. Networks display most significant enriched terms from the GSEA: GO analysis including biological process, molecular function, or cellular component for each carcinoma condition. Node size depicts protein count (i.e. "gene count"), and all terms are significant and filtered to padjust<0.05. No significant terms were found between spindle and sarcomatoid MBC.

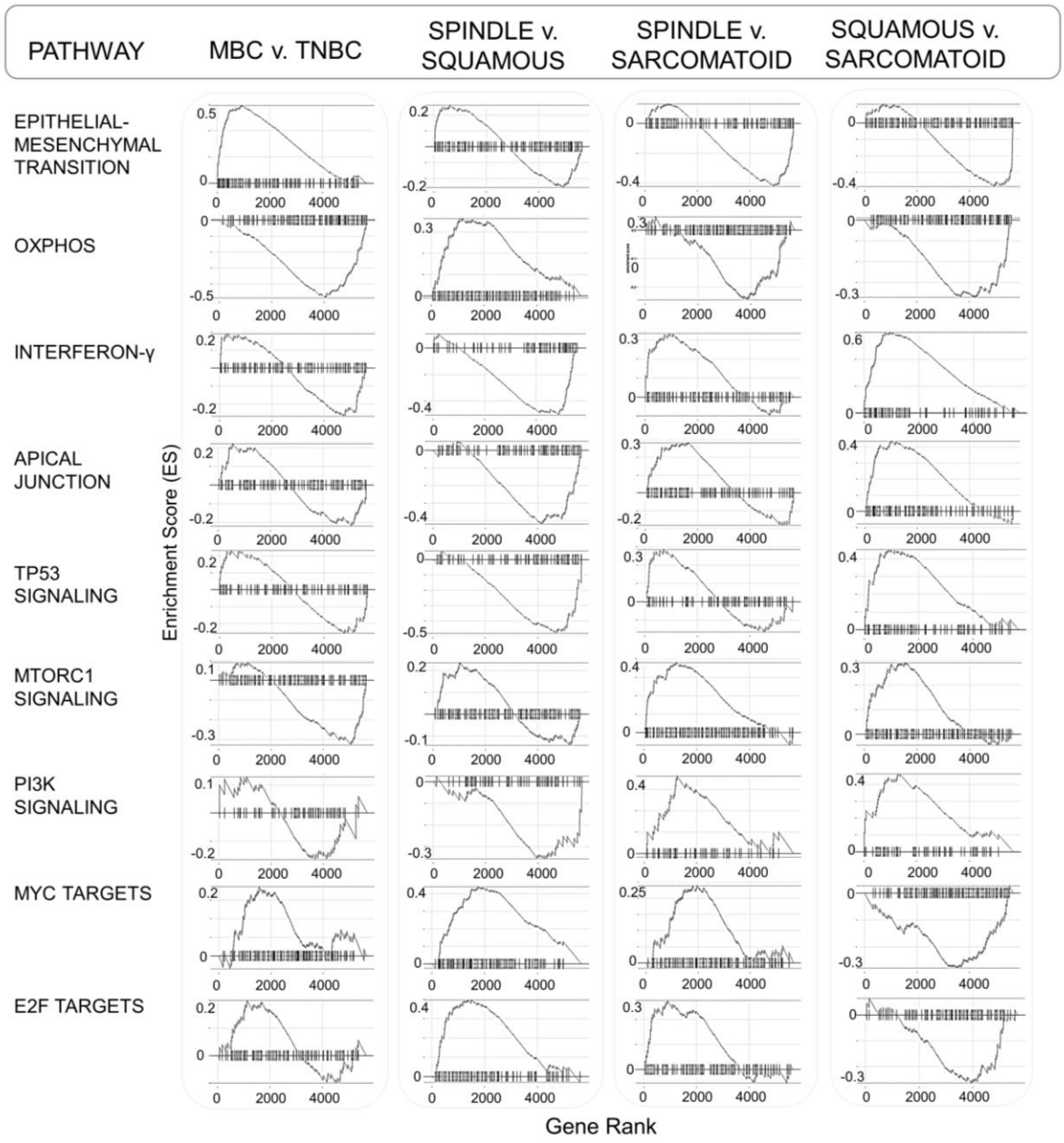


Figure 2-10. Top enriched pathway profiles distinguish among MBC subtypes and TNBC. GSEA enrichment plots of the highest up- or downregulated pathways by carcinoma type, including epithelial-mesenchymal transition (EMT), oxidative phosphorylation, interferon-γ, apical junction, TP53, MTORC1, and PI3K signaling, and MYC and E2F target pathways. The GSEA analysis was performed using the fgsea package in R for the hallmark collection ("H") (Broad Institute), with n=1000 permutations, where p-adjust < 0.05 and FDR < 0.05 were considered significant.

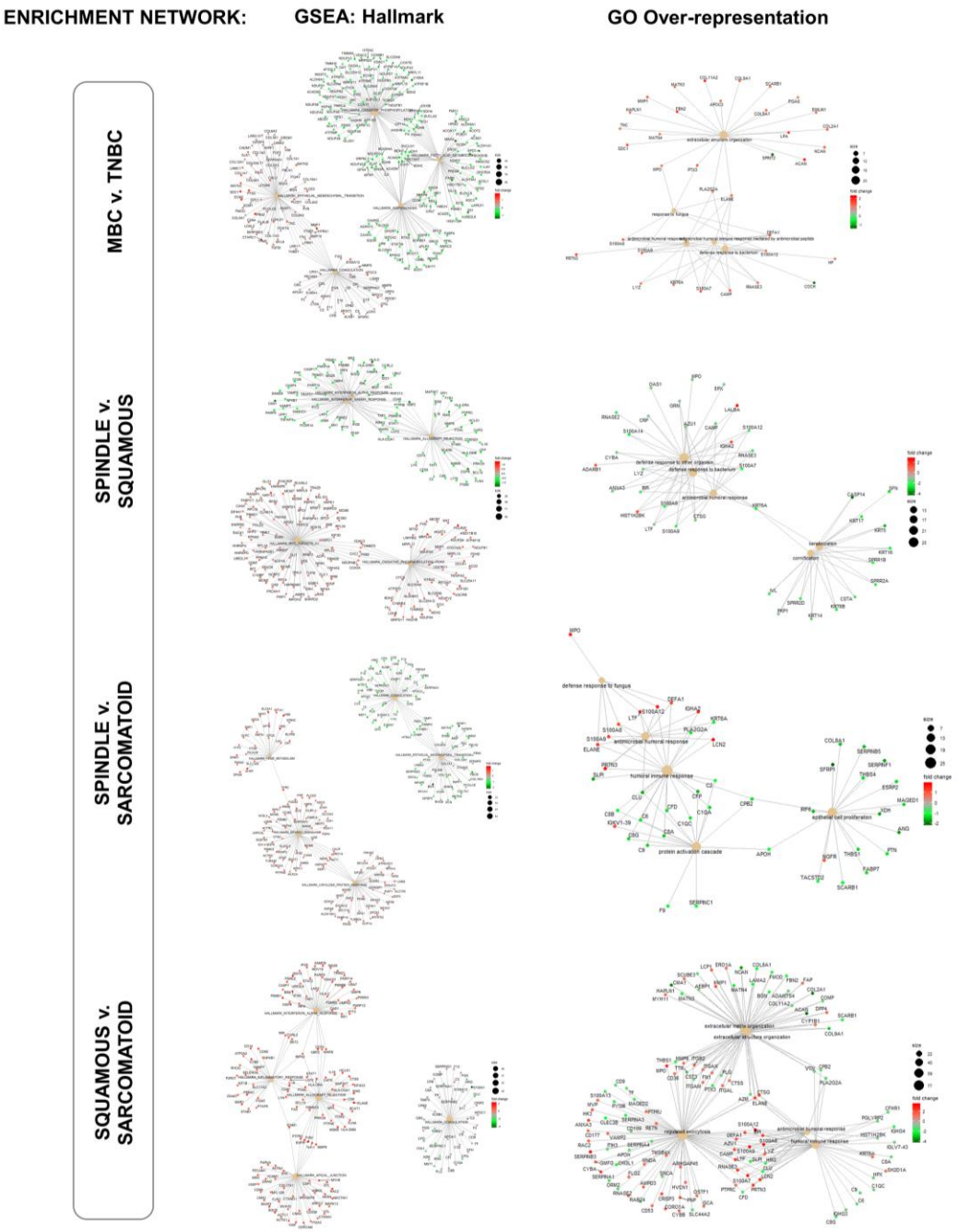


Figure 2-11. Enriched protein networks of the top GSEA hallmarks and GO terms within MBC and relative to TNBC. Networks display most significant enriched terms and associated proteins for each carcinoma condition. Node size depicts protein count, and the color scheme are fold change values (log2FC), where red=upregulated and green=downregulated proteins. All terms were filtered at pval<0.05 and padj<0.05.

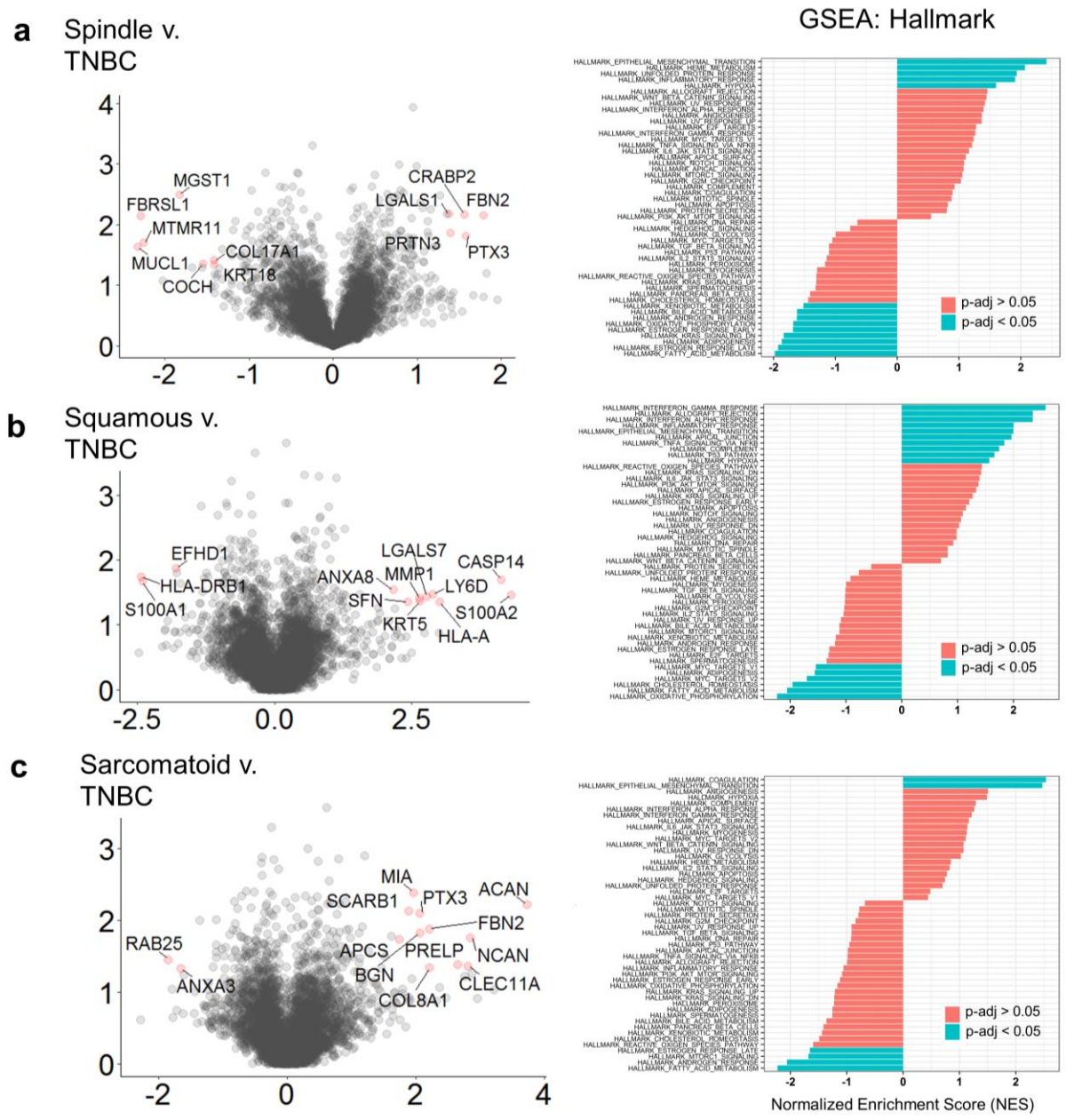


Figure 2-12. Differential expression analysis between TNBC and all MBC subtypes. Volcano plots for spindle MBC vs. TNBC, squamous MBC vs. TNBC, and sarcomatoid MBC vs TNBC, as log2-Fold change versus –log10(p-value) and GSEA hallmark analysis (MSigDB v7.0), where top pathways are shown as p-adj <0.05 (blue). We followed the clusterProfiler package in R with n=1000 permutations and FDR <0.05.

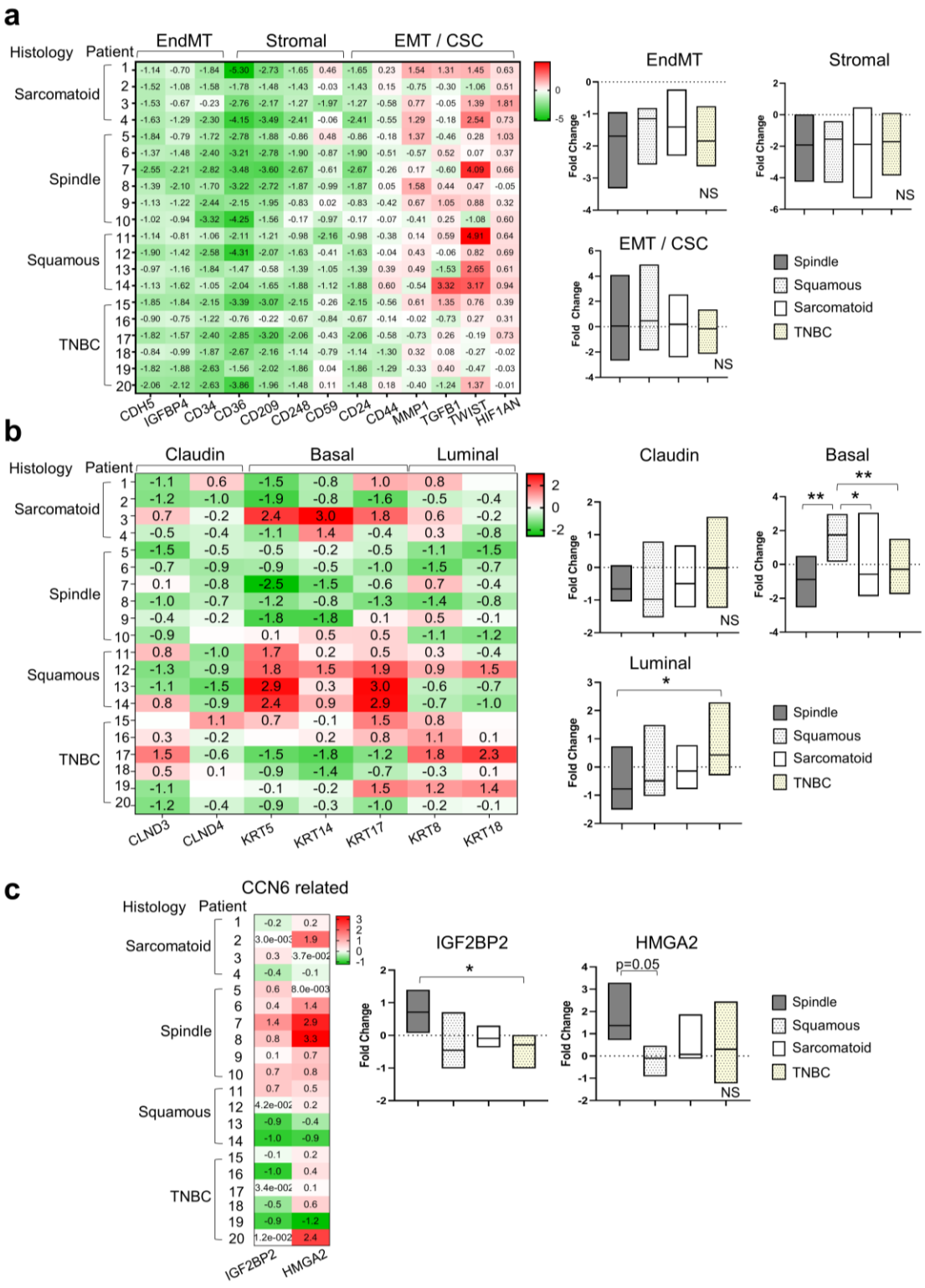


Figure 2-13. Patient-stratified unsupervised differential expression analysis provides personalized signatures across MBC and TNBC tumors. a. Visualization of patient protein profiles of hallmark cancer phenotypes/phenomena including endothelial-mesenchymal

transition (EndMT), stromal cells, and epithelial-mesenchymal transition (EMT) / cancer stem cell (CSC) markers. Heatmap shows fold change values of each patient relative to the average of normal breast tissues.
b. Stratification of patient profiles by well-known mammary gland proteins indicated by expression of claudin (CLDN3 and CLDN4), basal (KRT5, KRT14, KRT17) and luminal (KRT8, KRT18) markers.
c. Patient stratification of two recently identified and novel MBC markers, IGF2BP2 and HMGA2, specific to spindle/squamous in mice compared with human proteome confirms their upregulated feature and reveals a significant outcome in MBC spindle and spindle/squamous compared with a downregulation in MBC squamous. One-way ANOVA statistical analyses were performed between subtypes, where $p<0.05$ was considered significant (*), and $p<0.001$ was very significant (**).

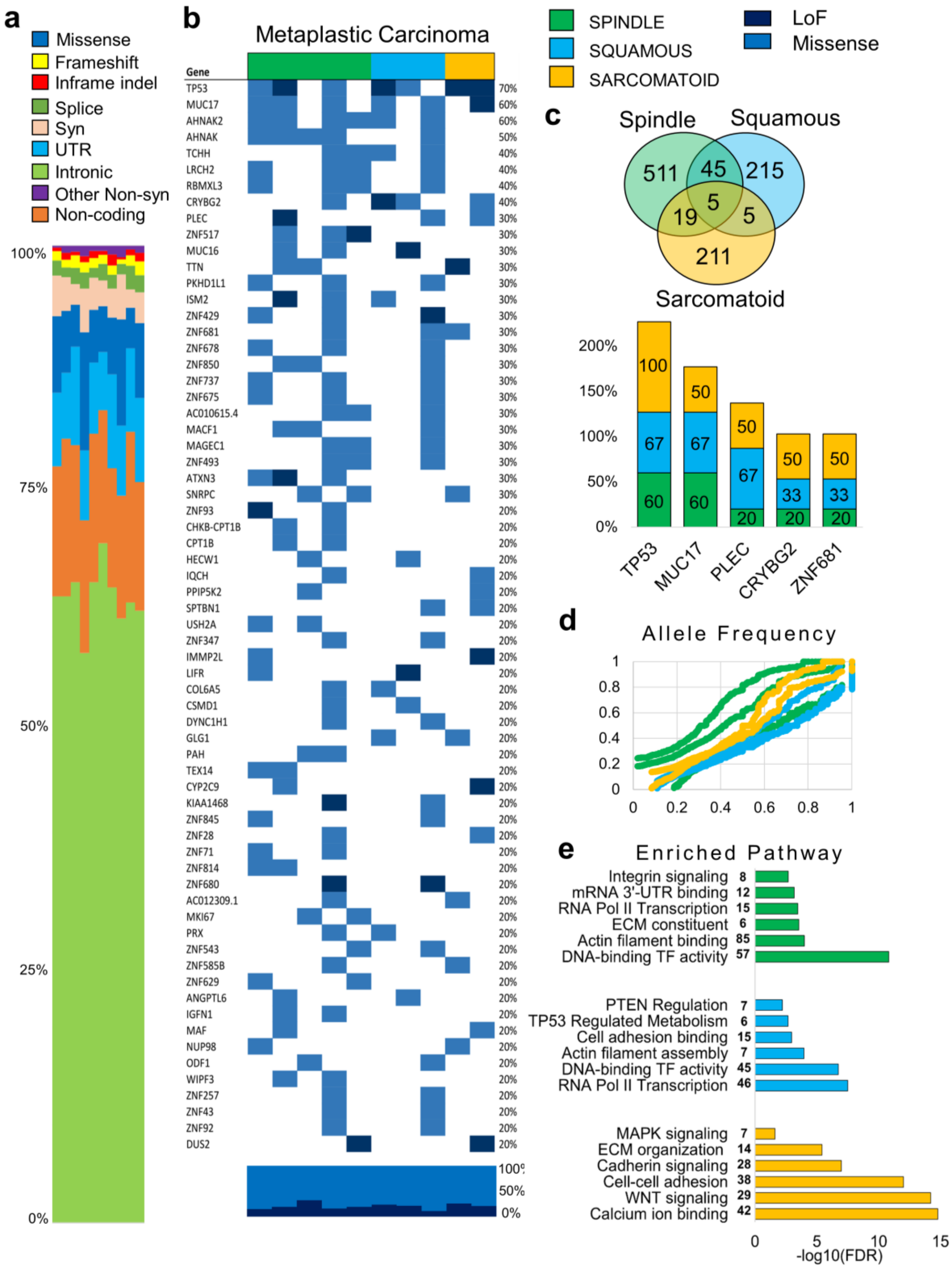


Figure 2-14. Whole exome sequencing (WES) analysis shows repertoire of somatic mutations of MBC and TNBC.

a. The landscape of somatic mutations common to MBC and within subtype-specific MBC histopathologies of 10 patients with MBC and matching normal breast tissues including spindle

(N=5), squamous (N=3) and sarcomatoid (N=2). The type of mutation is color-coded as indicated in the legend. Pathogenic mutational variants in MBC were defined as those of high complexity (1,011 of 11,652 total) were filtered in this analysis. Syn: synonymous, INDEL: inframe insertions and deletions.
b. Heat map shows the top mutated genes with 20% of greater frequency of missense (blue) or loss-of-function mutations, "LoF" (dark blue).
c. Venn diagram highlights the number common and distinct mutated genes in MBC subtypes. Bars show the frequency of the five commonly mutated genes (TP53, MUC17, PLEC, CRYBG2, ZNF681) in each subtype.
d. Scatterplot is average allele frequency (AF) for each tumor sample, and colors represent metaplastic subtype. Enrichment analysis was performed using gene lists extracted for each MBC according to the gene-level variant and effect summary tables (GeneRollupv0.3.2).
e. Top enriched GO and pathways (KEGG, Panther, or Reactome) of mutated genes in MBC subtypes. For this analysis, we excluded the variants with population allele frequency greater than 5% based on 1000 Genome Project data, and variants that fall in low complexity genomic regions, genomic super DUPS, and repeat masker regions.

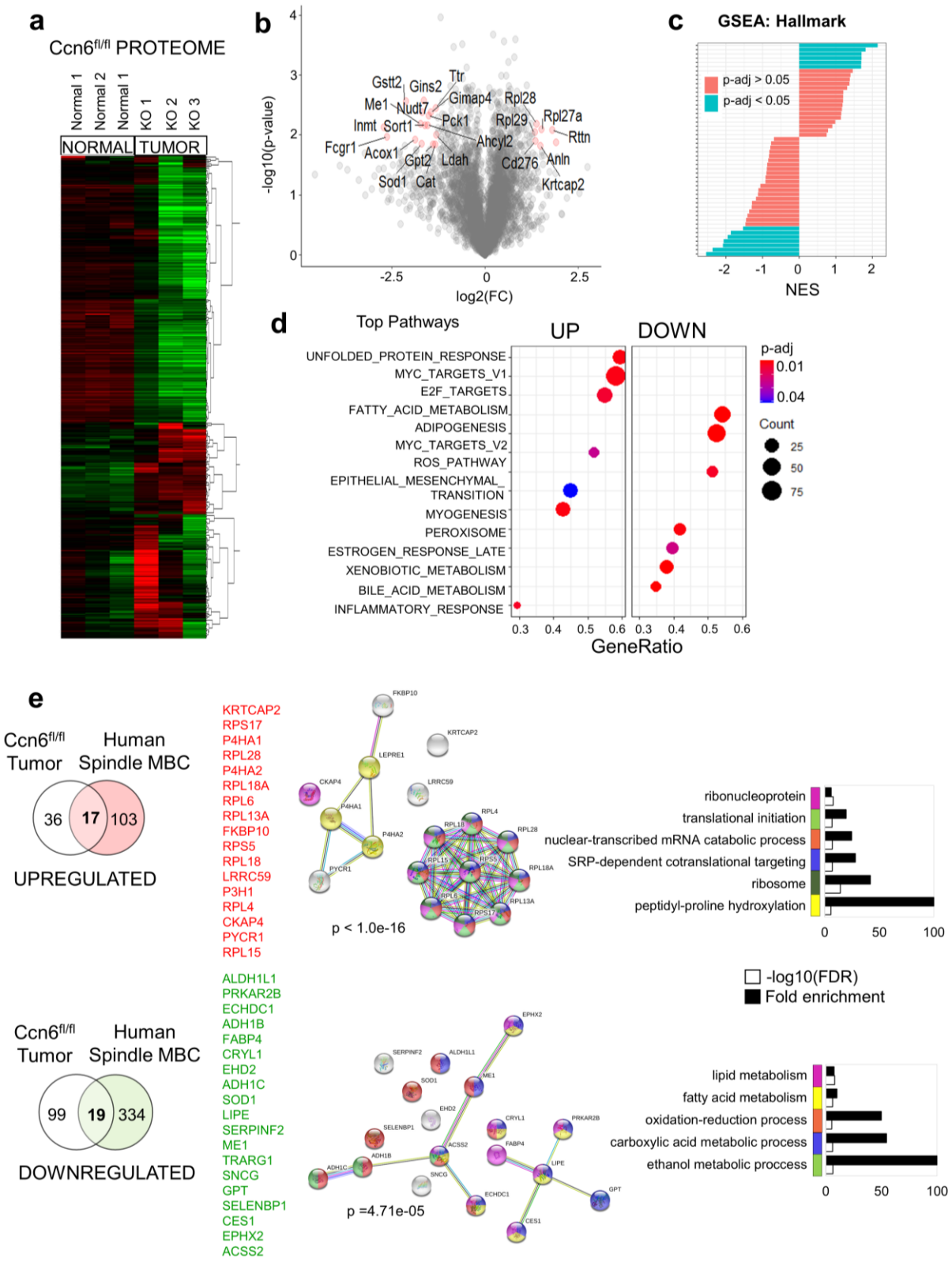


Figure 2-15. Quantitative proteomics analysis of MMTV-cre;Ccn6fl/fl spindle MBC tumors identify a common signature with human spindle MBCs. a. Heat map of the 4,609 proteins

that passed through quality filters in TMT-Integrator for all samples (3 MMTV-cre;Ccn6fl/fl mouse tumors (KO) and 3 normal mouse mammary glands).
b. Volcano plot comparing Ccn6fl/fl tumors with normal mammary gland. Significantly differentially expressed proteins are highlighted in red. $p<0.05$ and absolute value FC > 1 were considered significant.
c. Gene set enrichment analysis (GSEA) showing significant differentially expressed protein pathways in Ccn6fl/fl spindle MBCs compared with normal mammary glands. Normalized enrichment scores (NES) versus the total list of GSEA hallmark categories of up and down regulated hallmark pathways.
d. Top hallmark pathways highlighting significantly expressed up and downregulated protein pathways (marked "UP" and "DOWN", respectively). We used the molecular Signatures Database (MSigDB v7.0), hallmark gene sets with clusterProfiler and fgsea packages in R, with the biomaRT package in R to convert mouse gene IDs to human homolog associated gene symbols.
e. Venn diagrams demonstrate the overlap between the proteome of mouse MBC and human spindle MBC tumors, relative to their normal tissue counterparts, identifying a 17-protein upregulated and 19-protein downregulated protein set. Protein-protein interaction networks of up- and downregulated signatures highlight potential markers of interest, along with functional enrichment analysis using STRING v11.0. Color scheme of network proteins are matched by the legend of GO enrichment terms in barplots.

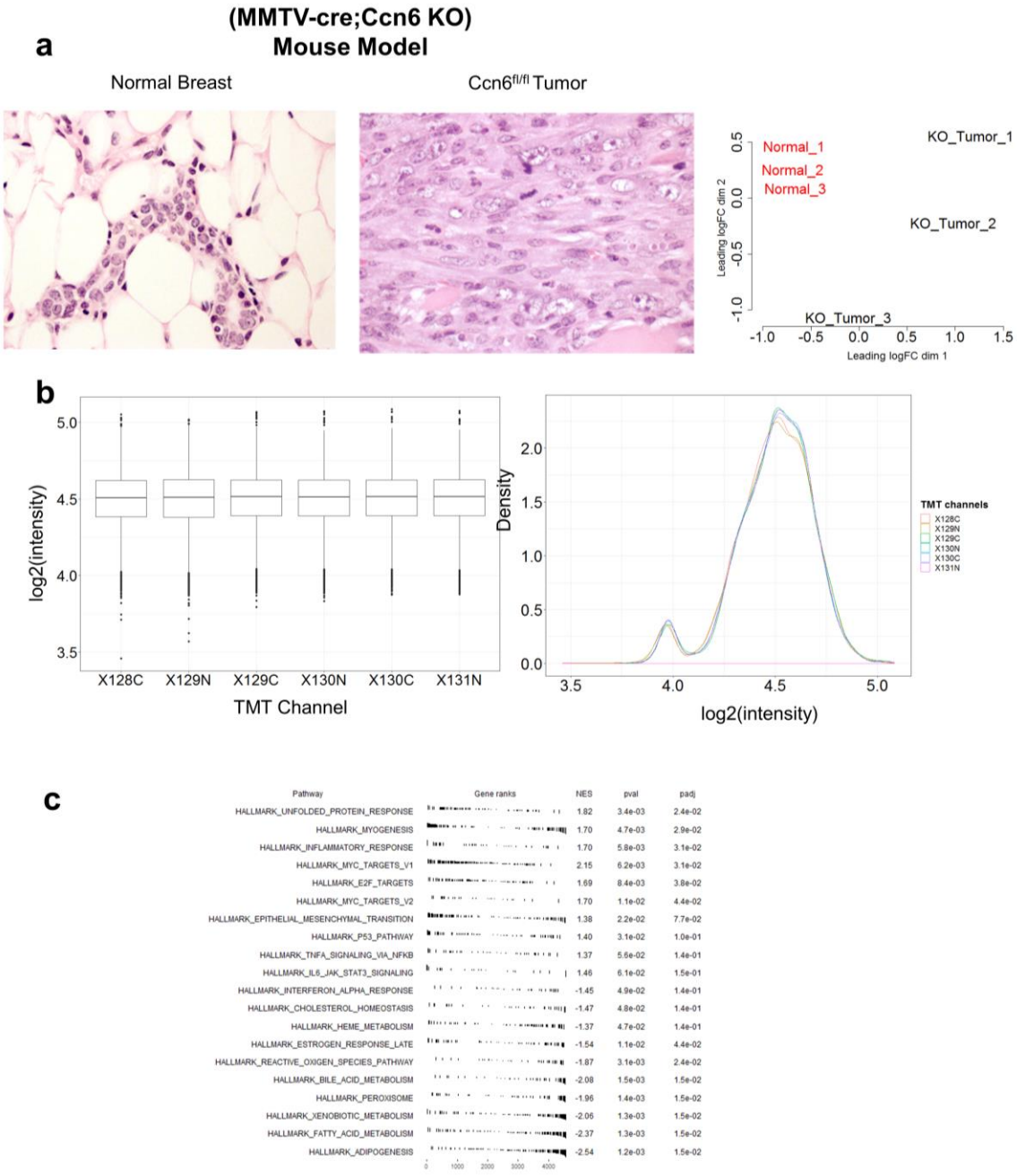


Figure 2-16. Pre-processing of LC-MS/MS TMT 10-plex proteomics of MBC mouse model (MMTV-cre;Ccn6 KO). a. Representative histology images stained with hematoxylin and eosin (H&E) of normal mouse breast tissues and MMTV-cre;Ccn6 mouse tumors. Scale bar = 50 μm. Raw data distribution of 6 mouse samples, 3 normal breast tissues and 3 CCN6 knockout tumors. b. TMT channel vs log2(intensity) for all samples and density distribution plot of log2-intensity vs. density of all expression values per sample.
c. Top hallmark pathways from GSEA analysis using MSigDB. Table shows gene ranks, normalized enrichment score (NES), p-value, and p-adjust value, where $p<0.05$ was considered significant.

Case	Age (years)	Tumor size (cm)	Diagnosis	Histologic subtype	Tumor grade	Stage at diagnosis	Distant Metastasis	Time to metastasize (months)
1	43	4	MBC	chondroid	3	T2N0	No	
2	89	2.8	MBC	chondroid	3	T2Nx	Liver, lung	12
3	39	1.8	MBC	chondroid	3	T1cN0	Bone	15
4	46	1.6	MBC	chondroid	3	T1cN0	No	
5	68	3	MBC	chondroid, osseous	3	T2N0	No	
6	50	29	MBC	spindle, osseous	3	T4N0	No	
7	52	2.7	MBC	spindle	3	T2Nx	Skin, abdomen	11
8	64	2	MBC	spindle	3	T1cNx	No	
9	68	5.5	MBC	spindle	3	T3N0	No	
10	47	4.2	MBC	spindle (partial squamous)	2	T2N0	No	
11	60	6	MBC	squamous (partial spindle)	3	T3N1	Liver, lung, bone	At diagnosis
12	53	4.5	MBC	squamous	3	T2Nx	Lung	6
13	81	10.1	MBC	squamous	3	T3N0	No	
14	55	3	MBC	squamous (partial spindle)	3	T2N0	No	
15	38	3	MBC	spindle	3	T2N2	No	At diagnosis
16	50	2.6	TNBC	IDC	3	T2N1mi	No	At diagnosis
17	33	2.1	TNBC	IDC	3	T2N0	No	
18	37	2	TNBC	IDC (apocrine features)	3	T1cN0	No	
19	57	5.9	TNBC	IDC (poorly differentiated with neuroendocrine features)	3	T3N1a	Liver	15
20	53	2.9	TNBC	IDC	3	T2N0	No	
21	73	2.8	TNBC	IDC (medullary features)	3	T2N0	No	
22-27	33-73 range	–	Normal					

Table 1. Clinical and histopathological features of the tumors in our patient cohort.

SAMPLE ID	HISTOLOGICAL SUBTYPE	SET	TMT CHANNEL
MBC (C1)	Sarcomatoid (Chondroid)	A	126
MBC (C2)	Sarcomatoid (Chondroid)	A	127N
MBC (Sp1)	Spindle	A	127C
MBC (Sq1)	Squamous	A	128N
MBC (Sq2)	Squamous	A	128C
TNBC1	Triple-negative	A	129N
TNBC2	Triple-negative	A	129C
Control1	Normal breast	A	130N
Control2	Normal breast	A	130C
Master Mix	Reference Sample	A	131
MBC (C3)	Sarcomatoid (Chondroid)	B	126
MBC (C4)	Sarcomatoid (Chondroid/Osseous)	B	127N
MBC (Sp2)	Spindle	B	127C
MBC (Sp3)	Spindle/Squamous	B	128N
MBC (Sq3)	Squamous	B	128C
TNBC3	Triple-negative	B	129N
TNBC4	Triple-negative	B	129C
Control3	Normal breast	B	130N
Control4	Normal breast	B	130C
Master Mix	Reference Sample	B	131
MBC (C5)	Sarcomatoid (Chondroid)	C	126
MBC (Sp4)	Spindle/Osseous	C	127N
MBC (Sp5)	Spindle	C	127C
MBC (Sq4)	Squamous	C	128N
MBC (Sp6)	Spindle	C	128C
TNBC5	Triple-negative	C	129N
TNBC6	Triple-negative	C	129C
Control5	Normal breast	C	130N
Control6	Normal breast	C	130C
Master Mix	Reference Sample	C	131

Table 2. LC-MS/MS study design and tandem mass tag (TMT) designations for patient samples

Chapter 3 – A Reproducible Scaffold-free 3D Organoid Model to Study Neoplastic Progression in Breast Cancer

The contents of the following chapter have been published in the Journal of Cell Communication and Signaling (2019) 13:129-143 (doi.org/10.1007/s12079-018-0498-7). Sabra I. Djomehri, Boris Burman, Maria E. Gonzalez, Shuichi Takayama, Celina G. Kleer

3-1 Abstract

While 3D cellular models are useful, gel-embedded organoids have large variability and slow transport of molecules to the matrix-encapsulated cells. This paper describes high-yield production of large (~1 mm diameter), scaffold-free, highly-spherical organoids in a one drop-one organoid format using MCF10A cells, a non-tumorigenic breast cell line. These organoids display a hollow lumen and secondary acini, and express mammary gland-specific and progenitor markers, resembling normal human breast tissue. Upon exposure to neoplastic conditions using TGF-β, a hypoxia-mimetic reagent, or co-culture with mesenchymal stem/stromal cells, the organoids increase collagen I production and undergo large phenotypic and morphological changes. Advantages of this scaffold-free, 3D breast organoid model include high consistency, ability to measure cellular collagen I production without noise from exogenous collagen, and capacity to subject the organoid to different stimuli from the microenvironment with precise timing without concern of matrix binding or slow diffusion. The morphological changes that accompany exposure to neoplastic conditions enable quantitative morphological analyses of the neoplastic progression. The method also allows for growth of organoids from cancer cells and primary tumors. The

platform is envisioned to be useful as a standardized 3D cellular model to study biomolecules from different neoplastic milieu as well as test potential therapeutics.

3-2 Introduction

The use of three-dimensional (3D) cell culture models has increased the understanding of breast cancer progression, enabled testing of new treatments, and may offer a unique platform for high throughput analyses of molecular changes during breast initiation, progression, and metastasis. However, novel models that better recapitulate the mammary gland architecture while also offering reproducible quantitative parameters are needed.

Being more physiologically relevant than conventional two-dimensional (2D) culture, cells grown in 3D are capable of organizing into distinct phenotypes resembling miniature functional units of the breast. Early investigations using a laminin-rich matrix, commonly known as Matrigel [11],[13], or a collagen matrix [128], as a reconstituted basement membrane demonstrated that mammary morphogenesis, mammary gland branching, and discernment of normal and neoplastic mammary phenotypes are assayable domains. Notable follow-up studies laid the framework for standard 3D culture procedures for breast models existing today [10],[65]. In this system, non-tumorigenic MCF10A cells form 3D microstructures called acini, which are generated either by embedding cells inside or on top of Matrigel. These scaffold-based systems have particular advantages when studying early events of tumor initiation, cell-ECM biophysical interactions [66],[74] and tumor migration and invasion [76],[129],[64]. However, although these systems recapitulate facets of the in vivo microenvironment with Matrigel substituting as a rudimentary extracellular matrix (ECM), the ECM scaffold can adsorb and trap growth factors, cytokines, and

chemokines as well as slow diffusion of therapeutic molecules to the cells complicating studies of the effect of different factors administered at different times.

Liquid-based systems, on the other hand, like the hanging drop technique, allow for media exchanges and administration of chemical components with flexible timing. Others have achieved spheroid reproducibility using 96-well hanging drop systems [76],[77] and even automated hanging drop culture with digital microfluidics [80] to generate a single spheroid per well. We have also developed a high-throughput 384-well hanging drop platform [81] that simplified the conventional hanging drop method, including optimized media additives [82]. The macromolecular additive methylcellulose (MethoCel) and a low concentration of Matrigel to the culture media only upon cell seeding contributed to highly uniform, compact spheroid morphologies. This system has allowed the development of spheroids from breast cancer cells [82], but until now has not been well explored using non-tumorigenic mammary epithelial cells to achieve consistent 3D structures, which demonstrate the challenges of slow or inefficient reaggregation capacity in a liquid-based setting [130].

Recent advances in cell culture are steering research towards organotypic culture (**Fig 3-1A**). Organotypic culture is characterized by the physiologic interaction of multiple cellular phenotypes to recapitulate the fundamental unit of the tissue of origin, the resulting cellular constructs often referred to as "organoids," whereas "spheroids" loosely describe 3D structures or cellular aggregates that may self-organize but need not recapitulate a tissue-like behavior (diagram, **Fig 3-1B**). Here, we define an "organoid," similar to that described in a recent review [68], as a 3D structure that self-organizes in culture to reproduce the basic functional unit of the native tissue.

In this study we use a scaffold-free 384-well hanging drop system to culture non-tumorigenic mammary MCF10A cells and show that they are able to expand into organoids that

resemble the architecture of normal human breast including presence of multiple cellular phenotypes [10]. We demonstrate 1) that laminin-rich components and not a laminin-rich matrix, together with a simple crowding agent and FBS in 3D suspension culture is sufficient to support formation of mammary acinar structures of high reproducibility, and 2) the formation of a large, self-organized mammary gland (MG)-like organoid structure using scaffold-free culture capable of exhibiting multiple lineage phenotypes. Importantly, we provide a new physiologically relevant 3D organoid model, able to recapitulate normal breast and neoplastic breast cancer progression.

3-3 Methods

Cell lines and cell culture. MCF10A and MDA-MBA-231 cell lines were obtained from the American Type Culture Collection (Manassas, VA). MCF10A, an immortalized non-transformed human breast epithelial cell line, is a well-established in vitro model of the benign breast. MCF10A are near-diploid cells with stable karyotype. MCF10A cells were cultured in growth medium containing DMEM/F12 (Invitrogen #11965-118), Horse serum (Invitrogen #16050-122) and supplements as previously reported by Debnath and colleagues [10] and MDA-MB-231 cells were maintained as per the manufacturer's instructions with culture medium composed of DMEM (Invitrogen) supplemented with 10% Fetal Bovine Serum (FBS). DsRed-labeled mesenchymal stem cells (MSCs) were generated by the Kleer Lab, where MSCs were isolated from fresh human breast cancer metastasis tissue (The Tissue Procurement Service at the University of Michigan, IRB#HUM00050330). MSCs were maintained as reported by Gonzalez et al. [131]. All cells were cultured at 37 °C and 5% CO_2 in 100 mm culture dishes, supplemented with 1% antibiotic-antimycotic, passaged before ~75% confluence was reached, and monitored with brightfield images collected at 10x magnification.

3D culture

A hanging drop array system was used to generate high-throughput 3D organoid culture in a 384-well format [81]. Hanging drop plates were coated in 0.1% Pluronic F108 (BASF) and UV sterilized. Hanging drops were formed using cell suspension solution with optimized media additives [82], 0.24% (w/v) methylcellulose (A4M MethoCel, Dow Chemical, MI) and 1.5-2% (v/v) Matrigel (growth factor reduced, phenol red-free, BD Biosciences, #356231), where 25 µl droplets were maintained. MCF10A cells were seeded at a density of 3,000 cells per droplet, MDA-MB-231 cells were seeded at 2,000 cells per droplet, and co-cultures with MCF10A and MSC cells were seeded at 1:1 with 2,000 total cells per droplet. To avoid evaporation, the array plate was sandwiched between a 96-well plate (Corning Costar 3596, Corning Inc., Lowell, MA) and plate lid to create a humidified environment. Fresh media was replenished every 2-3 days using a liquid handler (CyBi-Well, CyBio, Inc., Jena, Germany). Approximately 8-9 µl solution was removed from a drop and 9-10 µl fresh media added into a drop, repeated twice. For MCF10A organoids, we also added 10% FBS to the seeding solution to enhance organoid growth, and after spheres formed at day 3, we removed Matrigel, FBS, and MethoCel from the droplets with ~3 washes and replaced with complete media [10],[82] for the duration of culture. All organoids were maintained for 16 days and monitored by phase-contrast imaging (Nikon Eclipse TE300 inverted microscope). For comparison, MCF10A cells were seeded and maintained in a 96-well 3D format using the exact same parameters as in 3D hanging drop cultures except using 100 µl volumes and for media changes, 32 µl was removed from wells and replaced with 34 µl fresh media during each wash.

3D epithelial-to-mesenchymal (EMT) induction assays

Epithelial-to-mesenchymal induction assays were performed in the hanging drop system using MCF10A cells treated exogenously with either human recombinant TGF-β1 (Sigma, GF346)

or a hypoxia mimetic, $CoCl_2$ (Cobalt(II) Chloride 97%, Sigma, 232696). TGF-β1 and $CoCl_2$ solutions were prepared following manufacturer's instructions. For TGF-β1 induction assays, MCF10A organoids were grown for 5 days in hanging drop plates, followed by serum starvation of the droplets for 16 hours using serum-free DMEM/F12 medium (Invitrogen #11330-032). TGF-β1 was added to the serum-free medium, and a final concentration of 10 ng/ml TGF-β1 was added to droplets upon media exchanges and replenished after 2 days. Organoids were exposed to TGF-β1 for 4 days and removed from droplets with 7-8 washes of complete medium. For another 6 days, organoids were grown with complete medium and harvested from the plates. For $CoCl_2$ assays, hypoxia induction was performed as previously reported [132]. MCF10A organoids were grown as usual for 5-6 days. $CoCl_2$ was added directly to regular complete medium to induce a hypoxia-like response at a final concentration of 100 μM $CoCl_2$ for 24 hours. $CoCl_2$ was removed with 7-8 washes of complete medium, and organoids were allowed to grow for another week before harvesting.

Cryoblock embedding and sectioning

Organoid samples were collected from the array plate at a desired time, rinsed once with 1x PBS (GIBCO #10082), fixed in 4% PFA solution for 2 hours at room temperature, and washed again three times in 1x PBS. Prior to embedding, fixed organoids were stained with 0.1% (v/v) methylene blue for 10 min and rinsed with PBS. Next, a thin layer of optimal cutting temperature (OCT) compound (Tissue-Tek, Sakura Finetek, CA, #4583) was placed into a 5x5 mm cryomold, and the desired number of organoids were pipetted into the OCT layer. Additional OCT was filled to the top of the cryomold and snap-frozen on dry-ice, in block form, for 10 minutes in a Styrofoam container. The frozen blocks were stored at -80 °C, and subsequently sectioned into 5 μm sections on a Lecia 3050S Cryostat (Leica Biosystems Inc., IL) and mounted onto slides for staining.

H&E and Immunohistochemistry (IHC)

After thawing for 5 minutes, slides with sectioned organoids were rehydrated with 1X TBS buffer for 10 minutes. Heat induced epitope retrieval was performed with FLEX TRS Low pH Retrieval buffer (6.1, Dako, Carpinteria, CA) for 20 minutes. After peroxidase blocking, the antibody Collagen 1 rabbit polyclonal (Abcam, #ab34710) was applied at a dilution of 1:500 at room temperature for 60 minutes. The EnVision + Rabbit HRP System was used for detection. DAB chromagen was then applied for 10 minutes. Slides were counterstained with Harris' Hematoxylin and Eosin (H&E) for 5 seconds and then dehydrated and coverslipped. For mouse-derived organoids, primary antibodies used included Ki67 and Cleaved Caspase 3. Ki-67 rabbit monoclonal (Cell Marque, SP6 #475 R-16, Rocklin, CA) was applied at a dilution of 1:250 at room temperature (R.T.) for 30 minutes, and Cleaved Caspase-3 rabbit polyclonal (Cell Signaling #9661, Danvers, MA) at a dilution of 1:300 at R.T. for 60 minutes. To compare with organoid sections, we prepared H&E sections on human breast tissues samples obtained from University of Michigan (IRB approval HUM00050330).

Immunofluorescence and confocal imaging

Sectioned organoids were immediately fixed in 70% (v/v) methanol and stored at -20°C. After fixation, blocking of nonspecific binding of the antibodies was accomplished by incubation with Background Sniper (BioCare Medical, Pacheco, CA) for 30 min, followed by incubation overnight at 4°C with pairs of primary antibodies (mouse and rabbit). Primary antibodies included anti-E-cadherin (1:100, mouse monoclonal antibody, clone HECD-1, Thermo Fisher Scientific, Waltham MA, 13-1700), anti-CK5/6 (1:100, Thermo Fisher Scientific, Mouse monoclonal antibody, clone D5/16 B4), anti-CK18 (1:100, Rabbit monoclonal antibody clone ER431-1, AbCam, Cambridge, Ma, Ab32118), and anti-Vimentin (1:300, Rabbit monoclonal antibody,

clone EPR868(2), AbCam, Ab133260). After washing with Tris buffered saline (10 mM Tris HCl, pH7.4/ 0.154 M NaCl) containing 0.1% Tween 20 (TBST), the slides were incubated with fluorescent-conjugated secondary antibodies Cy3 and Cy5 for 30 minutes (1:200, #A10520-21, #A10523-24, Thermo Fisher Scientific) diluted in background sniper. After additional washes with TBST, sections were mounted using Prolong Gold (Thermo Fisher Scientific) containing 4′,6-diamidino-2-phenylindole (DAPI) and confocal images were captured with a Nikon A-1 Spectral Confocal microscope using a 20X and 40X oil objective.

Western Blot

Cells were collected and subsequently lysed in RIPA lysis buffer with protease and phosphatase inhibitors (Thermo Scientific #78410) at 100x dilution. Western blots were performed using 40µg of total protein, separating the protein samples by SDS-PAGE gel and transferring onto PVDF membranes. The membranes were blocked in TBS-T (Tris-buffered saline, 0.1% Tween 20) and 5% milk, and then incubated with primary antibodies in TBS-T at 4°C overnight. Abcam antibodies: CD49f (#AB112181), Vimentin (#AB16700), CD44 (#AB51037). Cell Signaling antibody: E-cadherin (#3195). Thermo Fisher antibodies: ALDH1A1 (#PA5-11537), EpCam (#710524). B-Actin-HRP (Santa Cruz, #47778) was used as a loading control.

Mouse-derived organoids (CCN6 knockout (KO) tumors)

CCN6 KO tumors from genetically engineered mice developed in the Kleer lab [62] were dissociated into tissue fragments and single cells prior to organoid generation. Tumor samples were washed for 5 minutes in 1x PBS, pH 7.4, then manually cut into small fragments (>1mm diameter) with a scalpel, tissue fragments were mechanically and enzymatically dissociated into a single-cell suspension using the MACS Tumor Dissociation Kit- mouse (Cat# 130-096-730, Miltenyi Biotec). After centrifugation and media aspiration, the remaining tumor cells together

with the cancer-associated stroma were resuspended in suspension media containing DMEM, 1x B27, 1x insulin transferrin selenite ethanolamine, 1x non-essential amino acids, EGF (10ng/ml), bFGF (10ng/ml), and 1x antibiotics. The cell suspension was then used to perform 3D hanging-drop culture as previously described. After 20 days, mouse-derived organoids were harvested, embedded, and sectioned as described above.

Morphological metrics and phenotypic assessment

Image processing for all morphological parameters was done in ImageJ software using the Watershed algorithm and binarization techniques as previously discussed [133]. The morphological parameters average organoid diameter, sphericity, and hollowness were measured by standard image segmentation techniques in ImageJ software using adaptive thresholding with the Watershed algorithm and binarization to create a mask of organoid boundaries. Morphological features were measured in triplicate with N=10 organoids for each group of cells and conditions, as indicated. Best fit ellipsoids of irregular organoid shapes were determined for average diameter and sphericity measurements using Feret's statistical diameter as an approximated equivalent diameter. Sphericity (named "circularity" in ImageJ) was measured as a ratio between 0 to 1, 1 being a perfect circle. The % hollowness was determined as a ratio of total area of cell and matrix regions to total organoid bounding area. For phenotype quantification, we measured the percentage of positive cells per organoid section by manual counting. This was calculated using threshold adjusted immunofluorescence signals from a Nikon E-800 microscope, with N=75-150 cells per organoid section and N=5 organoids per subgroup, and the results were triplicated. Similarly, we counted spindle-like cells in the same manner using H&E sections and confocal DAPI images.

Statistical Analysis

Results were expressed as means ± standard deviations and statistical analysis was performed using one-way ANOVA tests for morphologic and phenotypic data at days 8 and 16. Statistical significances of $P<0.05$ and highly significant $P<0.001$ results were considered. At each time point, ANOVA tests were carried out between non-tumorigenic (MCF10A), tumorigenic (MDA-MB-231) and neoplastic (co-culture, $CoCl_2$, TGFβ1) organoid subgroups.

3-4 Results & Discussion

Establishment and expansion of a scaffold-free 3D MCF10A breast organoid model

Modeling neoplastic progression in mammary organoids with multiple phenotypes and mixed cell populations remains a challenge [134]. A classic work by Wellings and colleagues [7] further supported by recent studies investigating cell-of-origin of breast cancers, suggests that human breast cancers originate from cells within the terminal duct lobular unit (TDLU, **Fig. 3-1A**), a composite of a duct and several lobules. In this study, we have generated a 3D mammary gland-like structure that resembles some of the signature and structural aspects of the normal breast.

In a previous report from our group [82], we found improved spheroid formation in a 384-well hanging drop system on spheroids derived from cancer cell lines using both Matrigel, collagen and other ECM hydrogel scaffolds. However, MCF10A cells could not be made to aggregate into tight single structures under these conditions (**Fig. 3-2A** no FBS conditions). Here we show conditions to form tight aggregates and further show that this scaffold-free environment allow 3D MCF10A organoids with hollow lumens to expand up to diameters of 1.2 mm as well as display multiple lineage markers (**Fig. 3-1A**, top panel). The large 3D organoids may also contain one or

more acini (**Fig. 3-1A**, bottom panel, acini indicated by white arrows). Because these features are distinct from previous reports of MCF10A acini [10],[64],[129],[14], and produce multiple cell lineages as should be observed in organotypic cultures [135], we refer to these 3D structures as MCF10A organoids.

Typically, MCF10A acini in the literature including studies from our lab [65],[64],[129], have matched physiologic acinar sizes (50-200 μm), but the MCF10A organoids formed here in hanging drop ranged from 900-1200 μm and 1070 ± 105 μm average diameter after 2 weeks in culture and typically exhibited multi-acinar structures resembling aspects of normal breast TDLU. These differences are depicted in **Fig. 3-1B**, which shows conventional MCF10A scaffold-based culture with polarized acini embedded in Matrigel (leftmost) or conventional spheroids grown in U-bottom plates (middle), MCF10A organoids developed in this study (rightmost), and legend showing typical sizes of the different biological structures formed in 3D. The size of the organoids may reflect not only the tightness of cellular interactions within the organoid but may also be a critical determinant of cell biology [136]. As seen in recent organotypic culture studies, larger sphere sizes may correlate with the ability of organoids to generate and display multiple phenotypes in 3D, as with multilayered optic cup structures from hESCs of ~550 μm diameters [130], to human brain organoids of ~2 mm diameters that produced regions of multiple cell types [137].

More critically, the use of MCF10A as a suitable cell line to model normal breast cell function has been established in the field [10],[13],[71]. More recently, Qu and others [75] found a mixed luminal and basal expression was observed. Our results confirm mixed luminal and basal expression of MCF10A organoids. As shown in **Figure 3-1**, MCF10A organoids expressed the basal marker cytokeratin (CK) 5/6 on the basolateral side, and luminal markers CK 18 and E-

cadherin on the apical side facing the lumen. Cytokeratin status has been intensely studied in clinical pathology on breast lesions and benign tissue, with distinct phenotypes described as stem/progenitor (CK 5/6+), glandular progenitor (CK 5/6+ and CK 18+), and committed glandular (CK 18+) [9],[138]. Interestingly, a population of CK 5/6+ cells on the basolateral side of MCF10A organoids appeared to co-express CK 18, suggesting glandular progenitor status. Supporting these data, recent studies generating TDLU-like organoids from primary human mammary epithelial cells found that the TDLU signature consisted of a co-expression of multiple lineage makers at similar positions as we observed in MCF10A organoids and in normal human TLDU [134],[15]. Together, these results support that the MCF10A model, when grown into 3D organoids, recapitulates some of the cellular heterogeneity of the normal breast TDLU.

Growth and remodeling of normal mammary organoids

The ability to form physiologic acinar structures in 3D is generally thought to require a scaffold-based setting because the support matrix provides the necessary pseudo-basement membrane. However, based on our studies, we postulate that the scaffold material itself might limit organoid expansion, by mechanical constraints and/or limiting the supply of critical nutrients [139], thus, MCF10A 3D structures only develop into smaller acinar structures of singular phenotype.

Since non-tumorigenic MCF10A cells are known to form acini in 3D without stemness capacity [75], we predicted low expansion ability in hanging drop. However, we found that a combination of low Matrigel concentration (~1-2%) and higher serum (10% FBS) at seeding strikingly promoted expansion of organoids (**Fig. 3-2A**), which was similar to a recipe of additives noted in organotypic culture, where self-forming 3D retina from hESCs used 1% Matrigel and 10% FBS over 18 days [130]. We also compared MCF10A cells grown in 96 well U-bottom plates

(H&E sections, **Fig. 3-2A**) and found greater hyperplastic regions within the lumens of the organoids in U-bottom plates compared to hanging drop organoids.

To further optimize the system, we explored organoid formation at varying concentrations of media-dissolved Matrigel over 3 days (**Fig. 3-2B**, **Fig. 3-3**). In the absence of Matrigel, MCF10A forms highly compact spheres but were unable to differentiate beyond a cellular aggregate (**Fig. 3-2C**), 1% (v/v) Matrigel was the minimum concentration needed to observe differentiation, while increasing to 2% Matrigel rendered lower aggregate-forming ability. Thus, we achieved one large organoid per droplet at 95% forming efficiency using 1.5% Matrigel and 10% FBS over 3 days, an optimal recipe to establish uniform spheres. This is consistent with literature, in which FBS was shown to augment and enhance cellular differentiation [130], while increasing Matrigel concentrations can decrease the number and function of acini formed [140].

We found by keeping cell seeding number and all other culture parameters constant, MCF10A organoid sizes increased by ~75% every 4 days for 16 days, co-culture and cancer organoids had slightly slower growth, while 10A MG^- spheres and EMT transformed organoids stopped proliferating after days 4 and 8, respectively (**Fig. 3-2C**). This suggests our cocktail of additives is critical for generating reproducible, uniform spheroids that can expand long-term, as shown by their range of sizes across cell subgroups (**Fig. 3-2C**), which has been a major challenge with conventional 3D platforms.

Available data suggest that hanging drop culture might provide advantages over other 3D spheroid assays for achieving organoid expansion. For example, conventional 3D gel embedded or on-top Matrigel culture of benign MCF10A acini [14], and 3D microfluidic and PDMS microwell systems for cancer spheroids [141],[78] show significant restriction in size (<300 μm diameter) and proliferation potential, possibly due to mechanical confinement. Also, a recent study

using microsensors found oxygen levels of multi-spheroids grown in conventional U-bottom 96 well plates dropped significantly after 18 hours, while hanging drops containing one spheroid had stabilized oxygen values [139]. Our observation of irregular morphologies, lumen filling, and no acini formation in 96 well U-bottom plates (**Fig. 3-2A**), suggests that hanging drops are organoid-supportive by providing more oxygen, by enhancing autocrine effects relative to the 96 well plates by using smaller media volumes, or by other unknown mechanisms.

We next investigated whether the expansion observed in MCF10A organoids reflected a stem cell population since organoids described in literature are either propagated by stem cells (e.g. mammary progenitors, hESCs, iPSCs) or maintained by a known stem cell population. Western blot analysis of MCF10A organoids demonstrated the presence of a stem cell population with an EpCAM+/CD49f^{+}/ALDH1^{+}/CD44^{+} phenotype (**Fig. 3-4**). This is an important finding since previous studies of MCF10A cells in 3D Matrigel systems were not able to observe stem cell populations or reported a loss of stem/progenitor markers when moving from monolayer to 3D [75]. This suggests the MCF10A cell line, in the absence of ECM scaffold, could be maintained by a self-renewing stem/progenitor cell population in 3D culture allowing substantial expansion. This may be a potential advantage of growing cells in a free floatation context, whereby the investigation of stem cell populations over time can be readily monitored despite having less physiologically relevant ECM mimicry than typical 3D scaffold systems.

Organoids as models of neoplastic progression

We next compared the phenotypic features of normal MCF10A cells grown with and without Matrigel, and MDA-MB-231 breast cancer organoids in our model, noting similarities and differences with other 3D breast models [65],[142]–[129]. By day 8, MCF10A organoids develop a hollow lumen similar to acini [10] but unlike typical acini embedded in an ECM gel, the

organoids are able to grow (approximately 700 μm in diameter), and form a complex network of acini, resembling a normal human TDLU in tissues (**Fig. 3-5**, in vivo TDLU image, leftmost column; arrows indicate acinar structures). This contrasts with the limited 3D growth of MCF10A formed under suboptimal conditions (MG^{-}) which remain as cellular aggregates and do not exhibit tissue-specific markers (**Fig. 3-5**, middle column), and with MDA-MB-231 breast cancer organoids that display an invasive phenotype similar to that of invasive breast carcinoma in clinical tissue samples (**Fig. 3-5**, rightmost panel) [129]. Over time, from day 4 through 16, MCF10A organoids undergo significant remodeling with lumen clearance, reduction of ECM matrix, and a tendency toward strict circular morphologies with increasingly cord-like basolateral peripheries (**Fig. 3-6A**). Together, these results demonstrate that the hanging drop organoid system allows expansion of the cell populations with normal and neoplastic phenotypes, which can be applied to the study of breast tumorigenesis.

To directly test whether the MCF10A organoid platform in hanging drop is useful to study changes that occur during neoplastic progression, we subjected our model to different treatments shown to induce an epithelial-to-mesenchymal transition (EMT), an important process during tumorigenesis [25]. MCF10A and other normal breast epithelial cells cultured in Matrigel were shown to undergo EMT and malignant transformation upon exogenous treatments [10],[142],[143], but whether these changes occur in scaffold-free 3D culture is unknown. We found that treatment of MCF10A organoids with TGFβ1 and the hypoxia mimetic agent, $CoCl_2$, two well established inducers of EMT and neoplastic transformation, led to a widespread loss of E-cadherin and increase in vimentin expression, which are hallmark responses of EMT (**Fig. 3-7**) [144].

We selected treatment with TGFβ and the hypoxia-mimetic agent given their reported effect in induction of EMT and malignant transformation in breast cancer. Tests using TGFβ-induced transformation have been used more commonly in 2D MCF10A models [31]–[145] than in 3D, with few reports utilizing a MCF10A gel-embedded system [70]. Studies of hypoxia and MCF10A acini have been even fewer, in which hypoxia chambers were used, but these were unsuccessful in detecting EMT in MCF10A 3D culture [146]. More recently, we found the literature points to certain chemical agents able to induce hypoxic effects rapidly, within minutes of administering, generating strong responses and higher controllability than hypoxia chambers [147]. Thus, we optimized a protocol using $CoCl_2$ induction from Wu and Yotnda [132] and successfully generated rapid EMT-driven responses after just 24 hours (**Fig. 3-8A**, H&E sections). For both TGFβ1 and $CoCl_2$ induction, we found significant deviation from normal MCF10A organoid morphology and signatures over days 8 through 16, with key differences highlighted in **Figure 3-6B**.

The ability to produce collagen and other ECM components has been ascribed to stromal cells [148] and to cancer cells [149]–[150]. Unexpectedly, we found that TGFβ and $CoCl_{2-}$ induced a strong fibrotic-like response (**Fig. 3-8A**, **Fig. 3-6B**, collagen I stains). Quantification of immunostaining results for TGFβ1 and $CoCl_2$ treated MCF10A organoids showed increased collagen I compared to untreated controls and were similar to levels observed in stromal cells and cancer cells (**Fig. 3-8A**, bottom plot). These data reveal the previously unknown ability of MCF10A cells to undergo de novo synthesis and deposition of collagen I upon EMT induction. Since conventional 3D gels often make use of collagen-containing matrices, the production of collagen from transformed normal epithelial cells would be an indiscernible or challenging feature to capture. The results indicate that benign breast cells such as MCF10A may have an inherent

plasticity that is revealed by using a gel-free culture platform. We also show that the de novo remodeling processes can easily be detectable by H&E or IHC analyses. From a clinical perspective, the 3D organoid system can be applied to study ways to inhibit stromal fibrosis in cancer progression using pharmacological approaches.

Organoids allow epithelial-stromal cross-talk

The microenvironment consisting of stromal, immune, and vascular cells as well as structural ECM components has been shown to exert a powerful influence on breast cancer initiation and progression [2]. Physiologically relevant in vitro models to study these interactions are needed. Towards this end, we generated a 3D heterotypic model with a simple test compartment of MCF10A and MSCs (**Fig. 3-8B**; MCF10A=blue, MSC=red).

Comparing to non-tumorigenic MCF10A organoids at day 8, co-culture organoids were significantly altered, with key features being disruption of CK5/6 expression, no acini formation, and lumen filling (**Fig. 3-8B**, H&E and confocal images). We also noted spindle-like MCF10A cell phenotypes and loss of cell-cell contact. These features reflect the profound influence of MSC-epithelial cell interactions, and suggests that the increased collagen output by MSCs may induce EMT on MCF10A cells (**Fig. 3-8A**, **Fig. 3-6B**). This is consistent with studies from our lab, which showed that direct contact with MSCs induced a spindle morphology and increased migration and invasion of MCF10A cells through the collagen receptor DDR2 in co-culture experiments [131]. Our 3D organoid platform may be useful to study the details and consequences epithelial-stromal cross-talk during neoplastic initiation and progression and may offer an ideal system to test new therapies to block the stromal pro-tumorigenic influences on epithelial cells.

To investigate whether our new model can be used to grow fresh tumors, we set out to generate organoids derived from $Ccn6^{fl/fl}$;MMTV-Cre mice, a breast cancer mouse model of

metaplastic carcinomas developed in our lab [62]. These tumor-derived organoids contain cancer cells, and cells from the stromal and immune compartments. Methods for generating mouse mammary organoids in 3D culture have been limited, and typically describe propagation of epithelial organoids [151]. In addition, we demonstrate detectable protein expression (e.g. proliferation marker Ki67 and apoptotic marker cleaved caspase 3). Because of growing interest in patient tumor derived organoids for purposes of biobanking [152] and single cell sequencing [153], these efforts have mainly focused on the primary epithelial compartment, but there is a great need to expand our understanding of patient intratumoral heterogeneity analyses. Our results of mouse-derived organoids show a step in that direction, forming in a highly reproducible format over shorter time periods.

Overall, the two major substrata used in 3D breast models, collagen I gels and Matrigel, are essential for providing the context whereby systemic cues can mimic the microenvironment, such as directing gene expression and cell responses [73]. But there are no specific guidelines for 3D systems on how best to mimic particular features of the in vivo situation [136]. Although use of a scaffold can preserve MCF10A acinar mechanical integrity by reproducing pressure forces like those found in vivo [71], spheroids grown under free suspension may not experience similar pressure cues, which could be a potential drawback. The lack of ECM biophysical forces could explain why MCF10A organoids undergo substantial expansion to approximately twenty-five times the size of normal breast acini. Even though as a structural unit, they are larger than the reported in vivo range, their ability to expand highlights a tight cellular commitment to the normal morphogenetic program.

Moreover, a major disadvantage of gel systems is their restricted ability to express tissue-specific markers [154]. Our data support the idea that a 3D free-floatation context might provide

a more permissive environment for differentiation as previously stipulated [68]. Thus, a general guideline for investigators using scaffold or suspension systems is to consider the relevant biophysical and spatiotemporal factors to facilitate a desired outcome.

Assessment of morphologic and phenotypic parameters of normal and neoplastic progression in organoids

Despite an increase in studies using 3D tumor models in recent years, there remains a lack of standardization in spheroid and organoid analyses. Some investigators advised monitoring a combination of morphological parameters [79],[155],[156] such as volume, sphericity, and area, however this information can be quite variable in the literature, stemming from inconsistencies in reproducibility and uniformity. Therefore, it is important to use a quantifiable metric to link our "observables" to biological relevance. With this outlook, we performed morphometric analyses on organoid parameters for average diameter, sphericity, hollowness, and phenotypic analyses on epithelial-like (E-cadherin+), mesenchymal-like (spindle +), basal / progenitor (CK5/6+), and fibrotic-like (Col1+) features for all tested organoid subgroups, i.e. non-tumorigenic (MCF10A), tumorigenic (MDA-MB-231), and neoplastic (co-culture, $CoCl_2$, TGFβ1) (**Fig. 3-9**).

ANOVA tests performed across all culture conditions for organoid-level morphologic and cellular-level phenotypic parameters at days 8 and 16 were statistically highly significant ($P<0.001$) (**Fig. 3-9A**). We found from days 8 to 16, sphericity and hollowness had diverged across conditions resulting in decreased P-values ($P<1E-06$). Since tumorigenic and non-tumorigenic morphologic parameters become well distinguished at 16 days, this suggests sphericity and hollowness could be suitable biological indicators for analyzing dynamic changes in organoids in short-term culture. Others have also noted overall shape is a highly variable parameter yet crucial indicator of tumorigenicity, as in a recent study that examined over 100 primary breast cancer

organoid lines [157]. The improved significance over time suggests when looking for distinct morphologic traits in organoids, they should be cultured > 2 weeks.

On the other hand, all cell phenotypic features were distinct by day 8 for all organoid subgroups (P<1E-08) (**Fig. 3-9A**). For neoplastic organoids, mesenchymal and fibrotic-like features converged to a tumorigenic phenotype just 2 days after treatment, with increased P-values at day 16 (P=0.005 and P=0.19, respectively), becoming phenotypically indistinguishable to MDA-MB-231 organoids. Thus, although organoid phenotypes are statistically unique, upon neoplastic progression they become harder to distinguish. As a result, we observed that epithelial and basal traits were harder to decipher since over time significances did not change as much as mesenchymal and fibrotic-like traits in response to a treatment. In relating phenotype to morphology, we found organoid sizes < 400 μm in general do not display organotypic expansion or differentiation (e.g. 10A MG– 3D structures), whereas after organoids reached > 800 μm in diameter, we detected unique multi-phenotypic and multi-morphologic states, summarized in the Figure 3-9B table. Here, sphericity/lumen status provided approximate regimes as follows: a normal epithelial, non-fibrotic state (> 0.75), a mesenchymal, fibrotic-like "invasive" state with largely filled lumen (< 0.5), and a mixed phenotype "irregular" regime (0.5-0.75) reflecting all neoplastic states (e.g. EMT, cancer, heterotypic).

Highly-reproducible 3D evaluations can form a basis for future experiments involving patient-derived organoids, which can in turn facilitate the study of large-scale drug screening and patient organoid biobanking [158] to guide personal therapies. We advise a quantitative analytical assessment based on a combination of morphological and phenotypic parameters for 3D tumor models, outlined in the Figure 3-9B table, which can be applied across cell types and conditions. In relating morphology with phenotype, filled lumen status and fibrotic-like features of neoplastic

organoids could reflect an ability to produce their own ECM which normal epithelial cells clearly lacked. However, those more difficult to discern traits such as basal / progenitor status, which might drive organoid regeneration, suggest an underlying mechanistic complexity to be followed up in future studies. Our model offers important advantages over previous organoid and 3D embedded culture methods, which are summarized in Table 3.

In summary, the present work describes a non-embedded, gel-free, hanging drop culture that enabled consistent production of highly-spherical MCF10A non-tumorigenic breast organoids with surprisingly large diameters in good yield. We further developed neoplastic breast organoids also with uniform, reproducible sizes and shapes, making it easy to standardize against other parameters. The obtained MCF10A organoids contain cells of multiple phenotypic lineages including tissue-specific and progenitor-like populations mimicking signatures of normal TDLUs. When exposed to conditions that mimic neoplastic progression, the organoids underwent cellular-level phenotypic changes and organoid-level morphologic changes that could be quantified to stratify organoid responses. The one droplet – one organoid simplicity, the highly regular shape and size of the obtained organoids, and dramatic morphologic changes upon exposure to neoplastic conditions open the door for assay standardization and high-throughput testing.

3-5 Figures

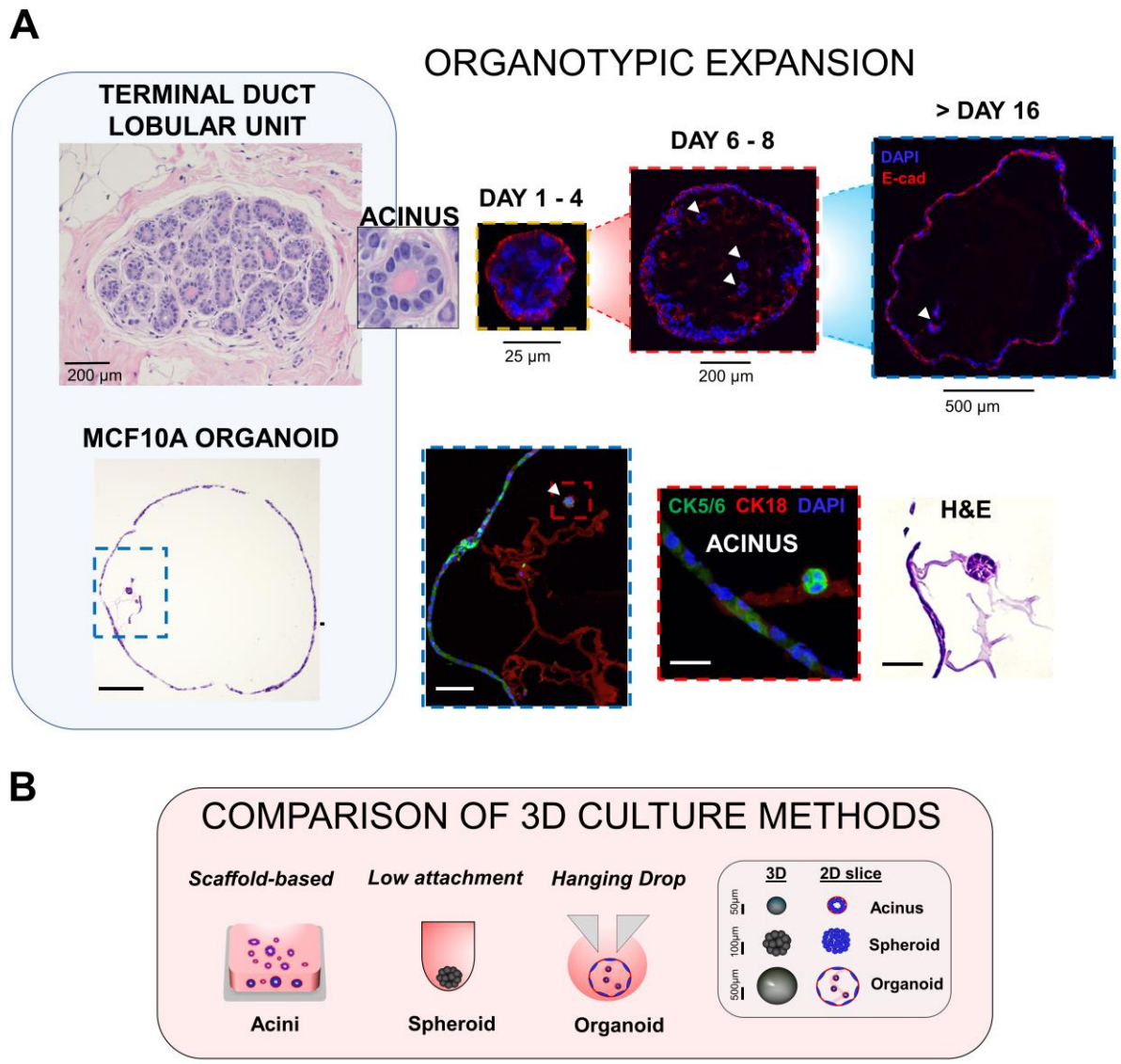


Figure 3-1. Organotypic expansion in MCF10A cells cultured in 3D hanging drop system. A) TOP PANEL: representative image of a normal terminal duct lobular unit (TDLU) in a human breast tissue sample (University of Michigan IRB HUM00050330) and inset showing an acinar structure comparing with organoid formation through 16 days, with expression of epithelial marker E-cadherin and hollow lumen formation. White arrows indicate 3D acinar structures within organoids. BOTTOM PANEL: 3D MCF10A organoid after 16 days by H&E showing a large hollow structure (scale bar = 200 µm), blue inset at 20X showing internal features (scale bar = 50 µm), red inset at 60X magnification of a developing acinar structure (scale bar = 15 µm) (confocal images: CK5/6=green, CK18=red, nuclei=DAPI) and corresponding H&E image (scale bar = 15 µm). B) diagram of conventional scaffold-based Matrigel culture or U-bottom spheroid formation compared with the hanging drop system. Legend shows characteristic sizes of acini, spheroid or organoid structures grown in 3D.

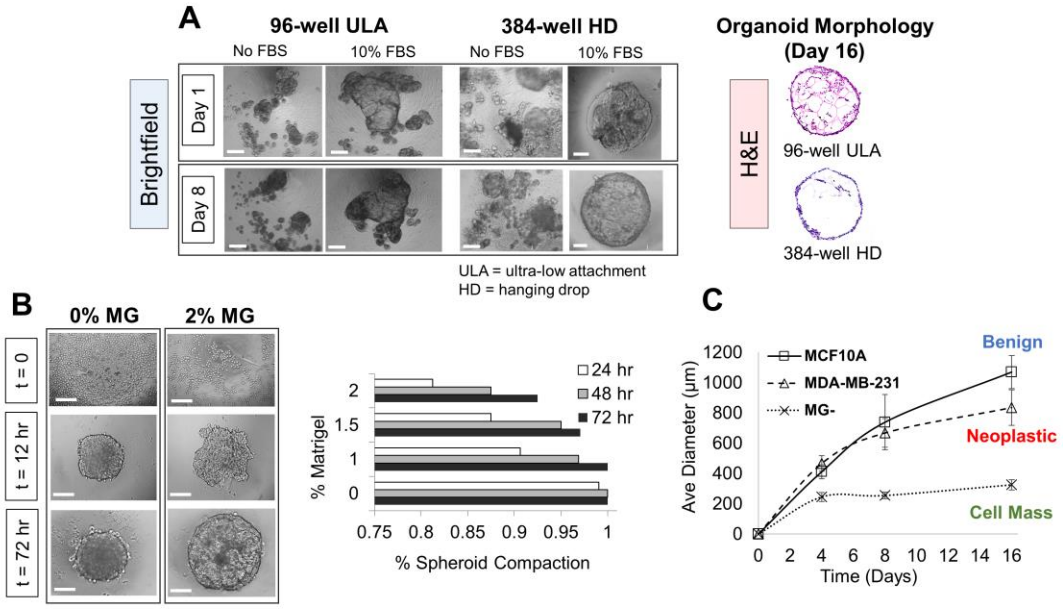


Figure 3-2. **Optimization of organoid formation technique in 384-well hanging drop (HD) system**. A) Representative brightfield images of 96-well ultra-low attachment (ULA) or 384-well HD 3D culture methods without FBS and with 10% FBS at days 1 and 8, and corresponding H&E sections at day 16. B) 0% and 2% Matrigel (v/v) at t=0, 12, and 72 hours (scalebar = 100 μm), and measurements of spheroid compaction efficiency rate at 0%, 1%, 1.5% and 2% Matrigel (v/v) at 24, 48 and 72 hours (N=25 per subgroup). C) average diameter of benign, neoplastic, and cell mass organoid subgroups (MCF10A, MDA-MB-231, and 0% Matrigel (MG^{-}), respectively) and time-course measurements at t=0 through day 16. Average diameter was measured by morphometric analyses in triplicate using N=10 organoids per subgroup, where best fit ellipsoids of round or irregular organoid morphologies were determined using Feret's statistical diameter in ImageJ. Scalebar = 500 μm for 96-well ULA and 150 μm for 384-well HD brightfield images.

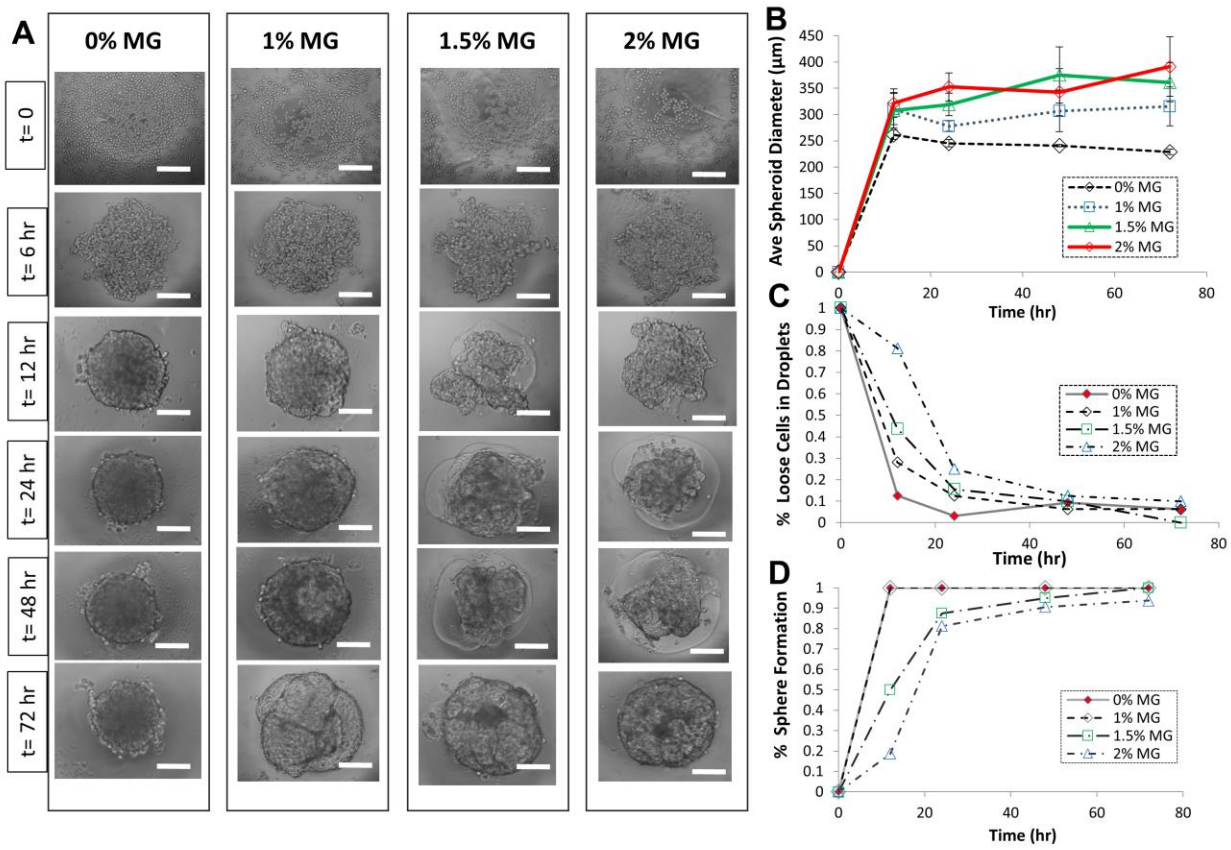


Figure 3-3. Spheroid optimization assay at varying Matrigel concentrations (0%, 1%, 1.5% and 2% v/v) from time=0 to 72 hours (N=25 per subgroup). A) representative brightfield images, B) average spheroid diameter, C) percentage of droplets containing multi-spheres per droplet, called "loose cell aggregates", and D) percentage of droplets that form one sphere per droplet. Scalebar = 100 µm.

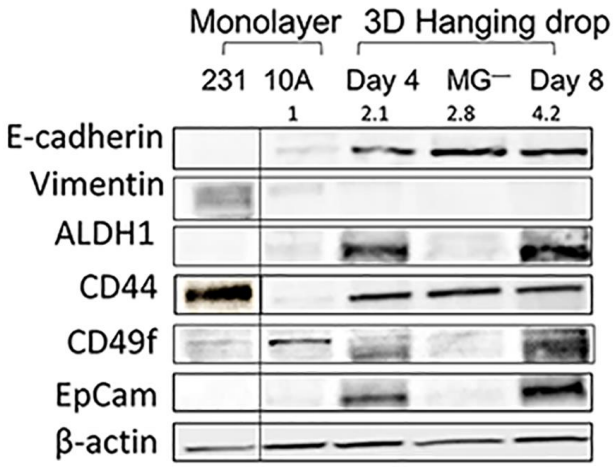


Figure 3-4. Comparison of stem cell vs differentiation marker expression for western blot analysis of MCF10A cells cultured in monolayer and 3D at days 4 and 8. "MG⁻" identifies MCF10A spheroids seeded without Matrigel. MDA-MB-231 cells cultured in 2D were shown as a control. Antibodies used: E-cadherin, Vimentin, ALDH1, CD44, CD49f, EpCam, and β-actin. Values above E-cadherin blot shows densitometric analysis of relative concentrations.

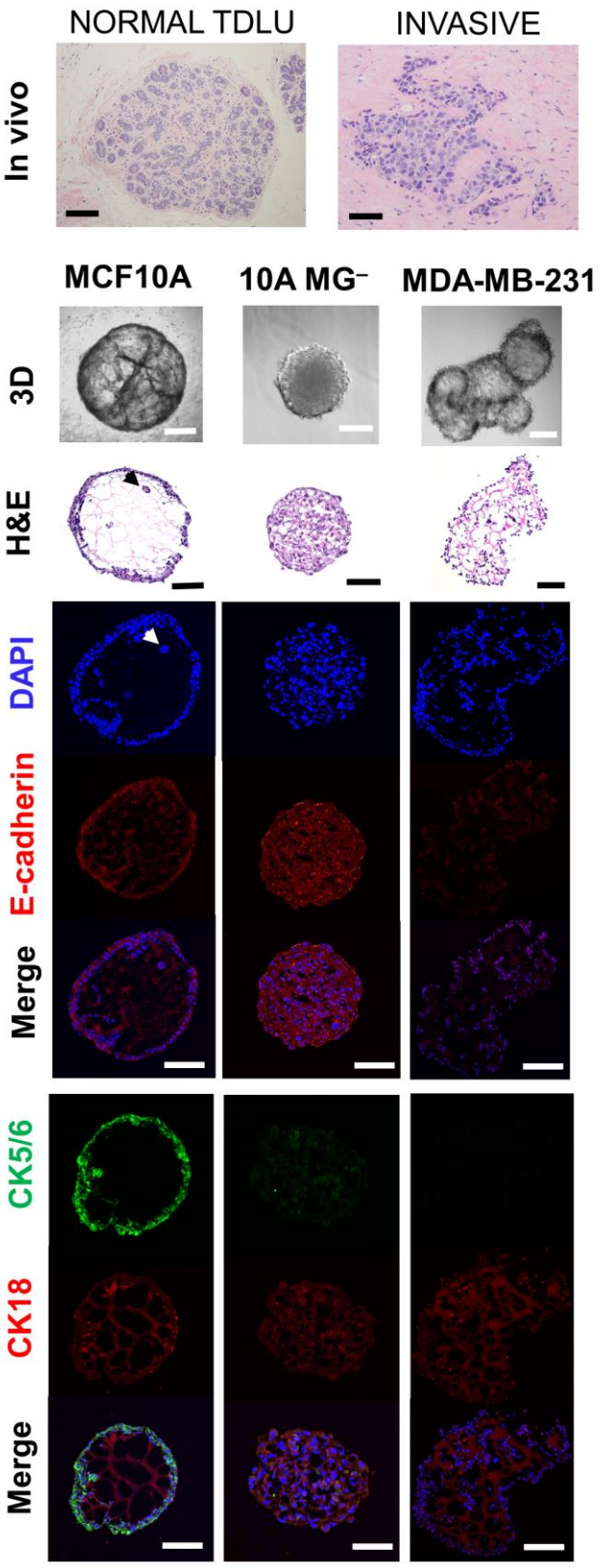

NORMAL TDLU
INVASIVE
In vivo
MCF10A
10A MG⁻
MDA-MB-231
3D
H&E
DAPI
E-cadherin
Merge
CK5/6
CK18
Merge

Figure 3-5. Organoid formation with normal MCF10A, MCF10A Matrigel-free (MG$^-$), and breast cancer MDA-MB-231 cells. Shown are representative images of the organoids at 8 days in 3D hanging drop with brightfield, H&E, and confocal images against Dapi, E-cadherin, CK5/6, and CK18. For comparison, images of a normal breast TDLU and an invasive ductal carcinoma from human tissues are shown with their characteristic morphology. Scalebar = 100 μm for 10A MG$^-$ and 200 μm for all others. Black or white arrows indicate 3D acinar structures within organoids. Distinct phenotypic features of normal or invasive TDLU (top panel, in vivo histologic images) share phenotypic similarities with in vitro organotypic structures in 3D hanging drop culture for benign (MCF10A), tumorigenic or neoplastic (MDA-MB-231), and non-tissue-like cell mass (10A MG$^-$) subgroups for the study of breast tumorigenesis and neoplastic progression.

(Figure on previous page).

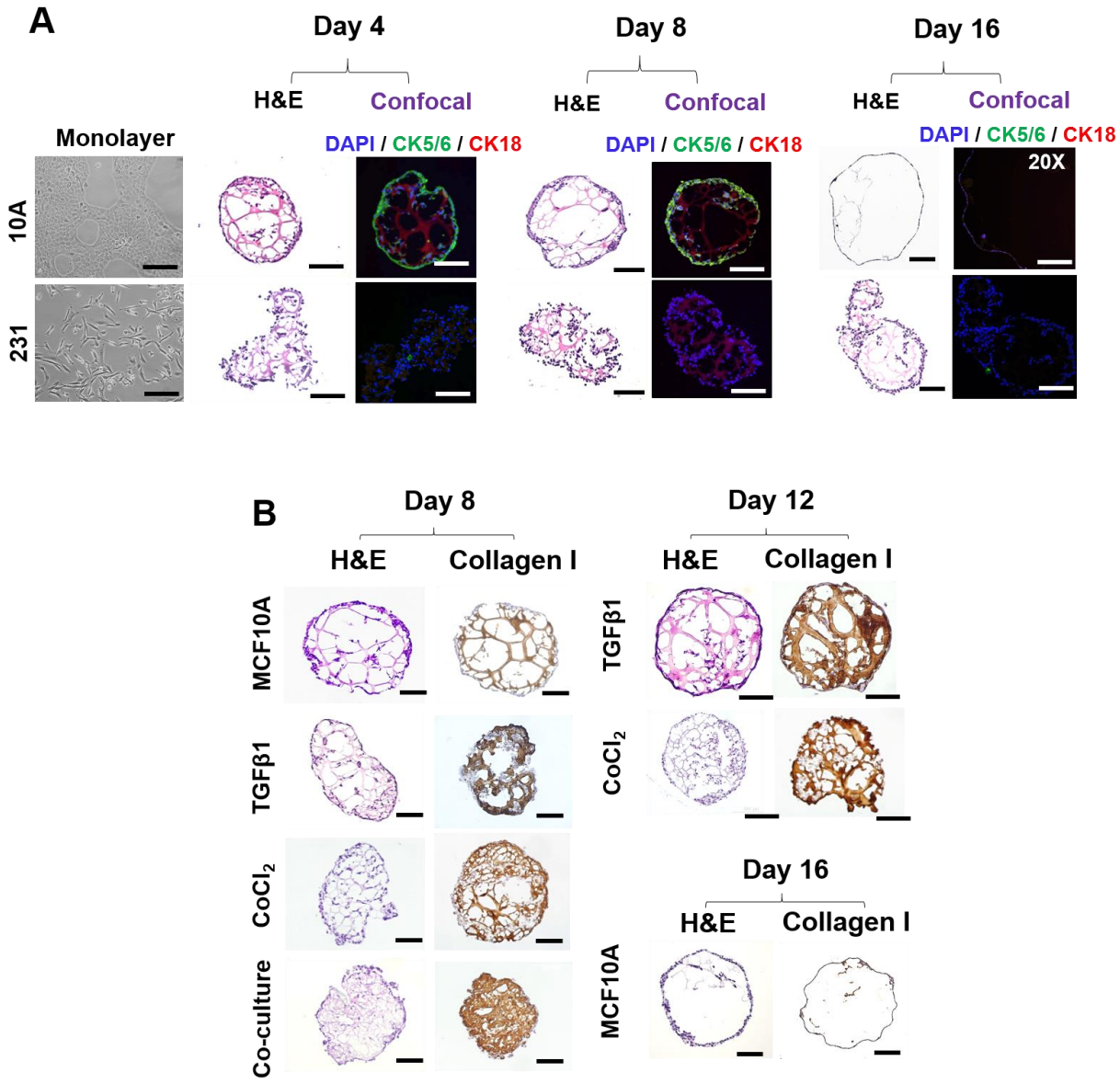


Figure 3-6. Summary of organoids developed in hanging drop. A) MCF10A vs MDA-MB-231 organoids at days 4, 8, and 16 with a comparison of monolayer phase contrast images, H&E staining, and confocal imaging with Dapi, CK5/6, and CK18 status (expression results merged), B) comparison of H&E and collagen I staining from MCF10A, TGFβ1, $CoCl_2$, and co-culture organoids at days 8, 12, and 16. Scale bar =200 μm.

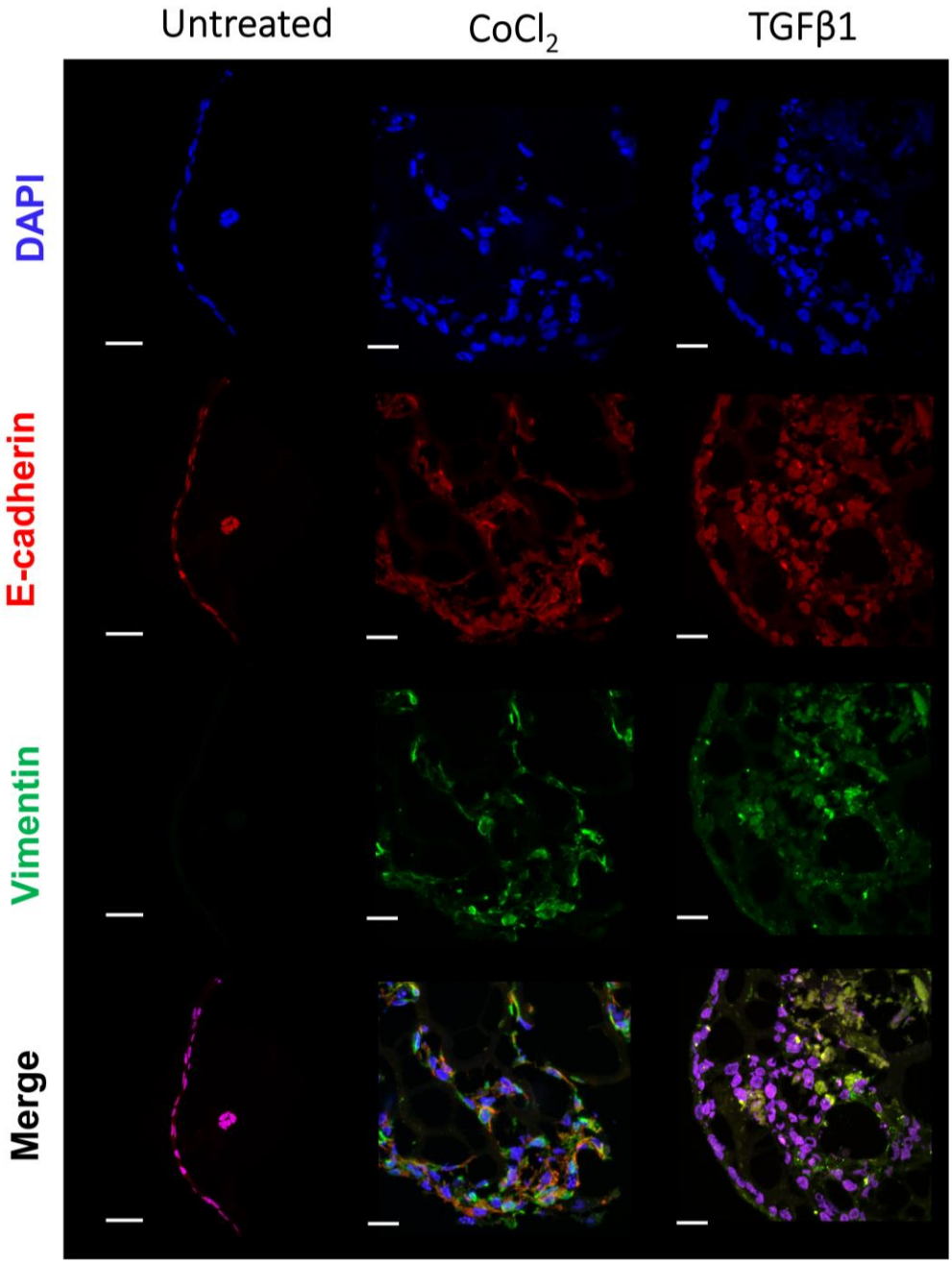


Figure 3-7. **Organoids as models of neoplastic progression.** Epithelial-to-mesenchymal (EMT) induction assays were developed for the study of neoplastic transformation in MCF10A organoids by TGFβ1 and $CoCl_2$ treatment at Day 16, two well-studied agents known to induce EMT and neoplastic progression of MCF10A cells. Representative confocal images show altered expression of E-cadherin (red) and Vimentin (green) at 20X magnification, with enlarged inset at 60X showing regions of co-expression. Scale bar = 100 μm. Neoplastic progression has not been explored in free floatation, scaffold-free 3D culture and allows organoids to be readily quantified using a 384-well HD format.

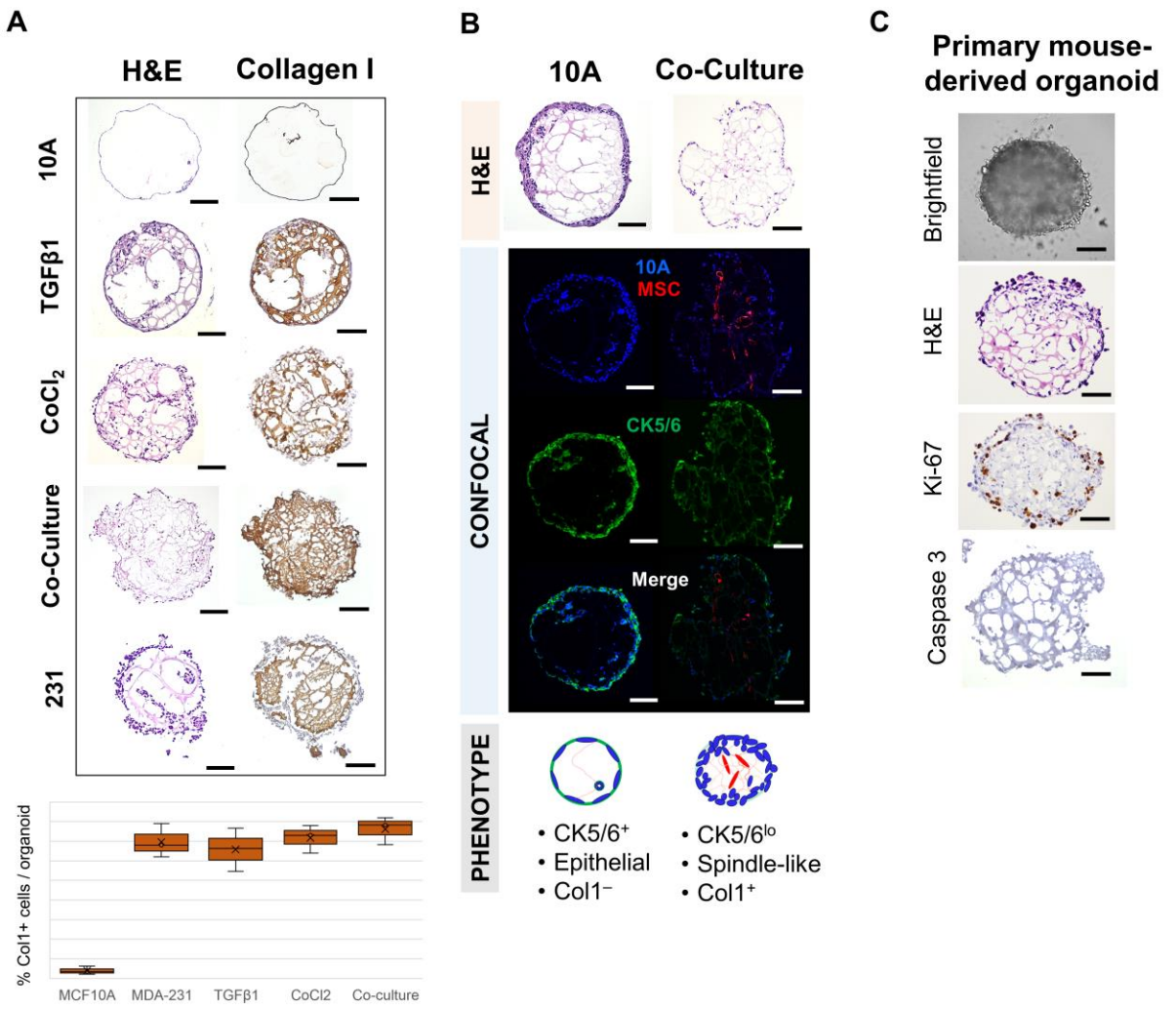
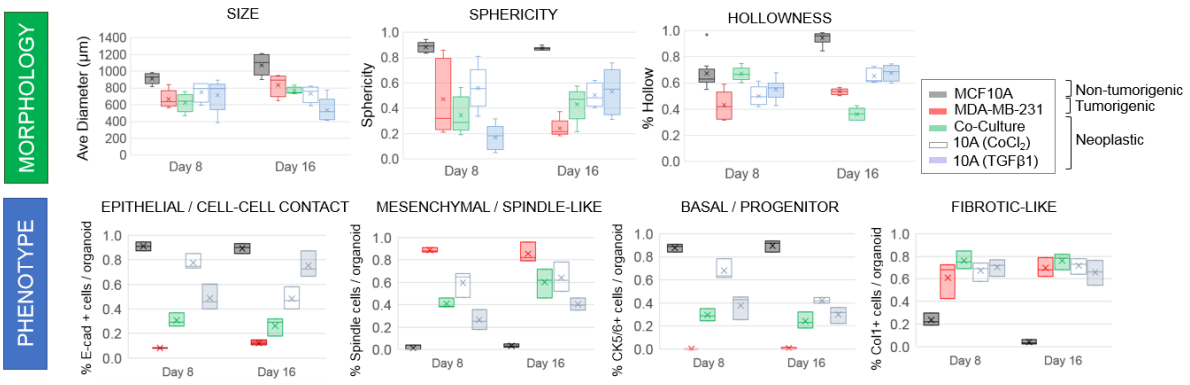


Figure 3-8. **Organoids developed in hanging drop exhibit phenotypic changes when subjected to different conditions and co-culture with MSCs.** Phenotypes in organoids developed in hanging drop with A) collagen I profiles for MCF10A, TGFβ1, $CoCl_2$, co-culture, and MDA-MB-231 at day 16 (N=5 per subgroup) compared with H&E and their associated fibrotic-like status via phenotypic quantitative analysis, B) H&E sections and confocal images at day 8 of normal MCF10A and co-culture organoids (MCF10A and MSCs seeded at a 1:1 ratio) showing spatial locations of MCF10A (blue) and MSC (red) cells within organoids, along with basal CK5/6 (green) status, and a schematic of typical morphology and phenotype based on epithelial/spindle-like features, or CK5/6 or Col1 expression, and C) CCN6 KO mouse mammary tumor-derived organoids with a live brightfield image at day 20, H&E and IHC sections for Ki67, and Cleaved Caspase 3. Scale bar = 200 μm.

A



B

Cell Type	MORPHOLOGIC PARAMETERS			PHENOTYPIC PARAMETERS			
	Size (Ave Diameter, µm)	Shape* (Sphericity)	Lumen† (Hollowness)	Epithelial / cell-cell contact	Mesenchymal / spindle-like	Basal or progenitor	Fibrotic-like
MCF10A	1070±105	spherical (0.87-0.90)	hollow (0.91-0.98)	++	–	++	–
MDA-MB-231	835±115	irregular (0.19-0.30)	filled (0.51-0.56)	–	++	–	++
Co-culture	765±40	irregular (0.22-0.58)	filled (0.32-43)	+	++	+	++
$CoCl_2$ treated	755±90	semi-spherical (0.41-0.61)	partial filled (0.61-0.74)	++	+	+	++
TGFβ1 treated	540±130	semi-spherical (0.35-0.71)	partial filled (0.61-0.72)	++	+	+	++

Figure 3-9. Evaluation of organoid morphologic and phenotypic parameters. A) Morphology parameters: average organoid diameter, sphericity, and hollowness, and phenotype parameters: epithelial/cell-cell contact (E-cadherin+), mesenchymal-like (spindle-like+), basal/progenitor (CK5/6+), and fibrotic-like (Collagen I+). Phenotypic parameters were based on the percentage of positive stained cells per organoid (N=5 organoids per subgroup, triplicated, 75-150 cells analyzed per section). Organoids were categorized as non-tumorigenic (MCF10A), tumorigenic (MDA-MB-231) and neoplastic (co-culture, TGFβ1 and $CoCl_2$ treated). B) table of the combined morphologies and phenotypes represented in a table showing results of quantitative measurements of all organoid subgroups at day 16. For morphologic parameters (N=10 per group): * = Sphericity > 0.75 on scale of [0 – 1] is "spherical," < 0.5 is "irregular," and ~ 0.5-0.75 is "semi-spherical". † = % Hollowness > 0.75 is "hollow," 0.5-0.75 is "partial filled," and < 0.5 is "filled," where "–" indicates no observed expression, "++" indicates a percentage of positive cells per organoid > 50%, and "+" indicates percent positive cells < 50%.

Table 3. Advantages of scaffold-free organoids to study neoplastic progression over other 3D methods

Cell culture system	Advantages and Applications	Disadvantages	References
Conventional scaffold-based 3D culture	• In vivo-like 3D organization compared to 2D cultures • Physiologic microenvironments • Direct cell-ECM interactions, biophysical responses • Luminal and basal marker expression • Long-term culture, easily maintained	• Highly variable 3D structures and variable differentiation capacity • Low throughput • Difficult to access aggregates • Limited modeling tumorigenic signaling • Mass transfer barriers • Potential loss of stem/progenitor populations	Barcellos-Hoff et al. 1989 Petersen et al. 1992 Debnath et al. 2003 Lee et al. 2007 Pal and Kleer 2014
Scaffold-free hanging droplet	• Single organoid per droplet, high-throughput, high reproducibility, compact aggregates • Direct cell-cell interaction • Ideal for modeling tumorigenic signaling over time • Maintains stem/progenitor expression • Easily quantifiable tumorigenicity status • Multipotent phenotypic differentiation of breast cells • Well suited for primary tumor organoids • Rapid testing of pharmacological interventions • In vivo-like microenvironment and cell-ECM interaction when derived from primary cultures	• Lack of in vivo microenvironment and cell-ECM cues when derived from cell lines • Difficulty to maintain long-term • Risk of droplet dehydration or mechanical disruption • Time consuming and/or labor intensive	This study
Hanging droplet (our previous work)	• High-throughput, compact spheres, simple usage • Scaffold-free, easy manipulation of droplets • Single tumor spheroids per droplet by gravity • Rapid testing with various therapeutics, easy media changes and access	• Unable to generate spheres from normal breast cells • Limited differentiation, variable 3D aggregates • Lack of standardization approach	Tung et al. 2011 Leung et al. 2015

Chapter 4 – Conclusion

4-1 Discussion and Future Directions

High-throughput translational studies are paving the way for a new era of improved clinical diagnosis and treatment options by combining patient data with available in vitro and in vivo models. Understanding the key signaling drivers of neoplastic transformation will enable a diverse range of integrative approaches that combine information on manifold scales and platforms, including human patient tumors and animal model data, 3D organoid screening and biobanking, and multi-omics technologies to analyze these on multiple molecular levels. Aggressive breast cancers such as triple-negative and metaplastic carcinomas are highly complex and heterogeneous tumors, but there is currently a drastic lack of knowledge on their underlying pathogenesis. This has resulted in a severe deficiency of models of oncogenic progression and thus, a lack of available therapies, especially in the case of metaplastic breast carcinoma, which is more aggressive and rare compared to TNBC due to the presence of heterologous non-glandular components displaying uncommon patterns of differentiation that are very abnormal to reside in the breast (spindle, squamous and sarcomatoid). Here we have introduced evidence that both human and mouse models, together with 3D organoid and multi-omics data are capable of recapitulating key molecular drivers of known histological features, and that it is possible to correlate phenotypic information about subtype-specific histologies on multiple molecular levels, such as comparing genomic and proteomic levels to tissue presentation via imaging data (histologic and

morphologic). This enables investigators to understand the pathogenesis of this complex disease, how these differentiation patterns arise, and what set of markers can be tested as novel therapies. In the future, studies will need to undertake the difficult task of seamlessly integrating the information generated from these multi-modalities directly with omics technologies, which are critical for elucidating multi-level information (e.g. thousands of data points on the expression of genomic variants, proteins, etc). Pathogenesis occurs as a complex orchestration among all these molecular and cellular levels, resulting in the range of heterogeneity we see within these tumors at the tissue level, and therefore it is of critical importance to assemble these data from large-scale and small-scale cohorts and create well-annotated databases in a patient stratified format to easily allow identification of novel biomarkers in the selection of therapeutically operable targets.

Here we presented a quantitative proteomic landscape of human MBC, a subtype of TNBC with a defining histology, frequent therapy resistance, and distant metastases [120]. Using tissue resources from patients and from an MBC spindle mouse model, robust proteomics and novel bioinformatics tools we demonstrated shared and subtype-specific altered proteomes present in spindle, squamous and sarcomatoid MBC, providing insights into the biology of this aggressive form of breast cancer and offering unique opportunities for precision medicine.

Despite the impact of protein expression on tumor phenotypes and clinical behavior, our knowledge on the protein landscapes of human breast cancer, and how differential profiles contribute to breast cancer phenotypes is very limited. Advances in proteomic methodology and bioinformatics technologies have enabled detailed characterization of breast cancer [19], [95]. However, studies to date have focused on frequent subtypes of breast cancer, with few to no MBC cases and normal breast tissues.

Upon comprehensive analyses we revealed common and subtype-specific differences among MBCs and their relationship with TNBC. Relative to TNBC, the proteome of MBC is enriched in EMT, and has a highly activated ECM, and reduced oxidative phosphorylation. These data are consistent with our previous studies showing that MBC express proteins involved in EMT, which might contribute to a more stem-like and aggressive phenotype compared to TNBC. Also, transcriptome studies have shown that MBCs cluster as basal-like and claudin-low [57], [121] and subtype-specific proteins and pathways emerged when we compared proteomic profiles of spindle, squamous and sarcomatoid tumors. A distinct feature of spindle MBCs was upregulated E2F and MYC signaling, ribosomal proteins, nuclear processing, and translational events. These are intriguing findings since recent reports have shown a possible synergy exists at the E2F/MYC axis, which is known to be involved in the cell cycle, chromatin remodeling and EMT, where ribosome biogenesis and increased translation play an essential role [122], [123]. These findings also suggest that spindle MBC has a deregulated balance between translation and metabolic pathways. Squamous MBC on the other hand, was found to have predominantly upregulated inflammatory responses (e.g. IFN-γ, TNFα, PI3K/MTOR), keratinization, and widespread cell adhesion marker expression and decreased oxidative phosphorylation, E2F/MYC signaling, and sarcomatoid MBC exhibited prominent profiles in extracellular matrix signaling and increased EMT program compared to other MBCs, possibly owing to it being analogous to a differentiation pattern along mesenchymal lineages [124]. Also, we observed that in general, there is no significant difference in P53 and PI3K pathway between MBCs and TNBC, and that while our proteomic analysis suggests that spindle, squamous and sarcomatoid MBCs and TNBC may share initial neoplastic events, each MBC subtype appears to have unique and active differentiation programs.

MBC are chemoresistant and metastatic, but the underlying molecular events and driver pathways are unclear and there are no effective treatments currently available [48], [52]. The heterogeneity of MBC has been investigated at the genetic level [58], [60], [125], however the relationship between genomic and proteomic alterations is unknown. Recent reports at the genetic level show that MBC harbor somatic mutations in TP53, PI3K/MTOR and Wnt signaling pathway genes [60], but no underlying subtype-specific mutational profiles in MBCs have been uncovered. Our whole exome sequencing analyses of 10 paired tissue samples of MBC and normal breast tissue from the same patient cohort identified somatic mutations common to all MBCs in TP53, MUC17, CRYBG2, PLEC, and ZNF681, where only TP53 has been previously reported in MBC. While spindle and squamous MBC exhibit overlapping mutational profiles of genes involved in RNA metabolic processes and actin filament binding, sarcomatoid tumors harbor distinct mutations in MAPK, WNT, protocadherin cluster genes, calcium binding and ECM organization. Together, these findings demonstrate novel somatic mutations in MBC tumors, distinct mutational profiles of sarcomatoid MBC compared to the overlapping landscape of spindle and squamous MBCs, which underscore the importance of elucidating the proteomic landscape to nominate subtype-specific proteins and pathways, especially between tumors with spindle and squamous differentiation, that may lead to novel treatment targets.

Quantitative proteomics results from our mouse model of spindle MBC using MMTV-Cre;Ccn6$^{fl/fl}$ mice demonstrated a significant overlap with human spindle MBCs, with shared enrichment in E2F, MYC, EMT, and ribosomal proteins. Our analyses further pinpoint a set of 17 upregulated and 19 downregulated proteins commonly deregulated in mouse and human spindle MBCs which have not been previously considered in these tumors. The 19 downregulated proteins have functions in metabolic processes, while of the 17 upregulated proteins, the majority are

ribosomal proteins. These data validate the MMTV-Cre;Ccn6$^{fl/fl}$ mouse model as a relevant tool to investigate and test new targets for spindle MBC, and we suggest this overlapping set of proteins between mouse and human MBCs represents a set of therapeutically operable targets. Our studies could pave the way towards mechanistic and functional investigations of these proteins in elucidating the underlying biology and pathogenesis of spindle MBC.

In addition to subtype specific tumor signatures, our study uncovered a stromal cell expression pattern in MBCs in the form of a significant number of mesenchymal and stromal proteins being downregulated (e.g. CD34, CD36, CD59, CD209 CD248). The stromal profile is intriguing and has not been previously considered in relation to metaplastic breast carcinoma, and hints the interplay of cellular transitions are critical to either stemness, invasion, or both, and that processes like EMT, angiogenesis and other unexplored events like EndMT (endothelial-mesenchymal transition) might also be simultaneously involved in oncogenic transformation [127].

In summary, we presented a patient cohort that describes a common proteomic landscape of MBC relative to TNBC, and highlights the existence of specific protein signature that underly the different histopathological subtypes of human MBC. We also show that quantitative proteomics assisted in refining the somatic mutational landscape of MBCs and allowed for the distinction of the MBC tumor subtypes. We provide evidence for a significant overlap between MMTV-Cre;Ccn6$^{fl/fl}$ spindle MBC mouse model and human disease, which consist of a common histopathology on the protein level, and validate that this mouse model can be utilized for testing new treatments and mechanistic investigations on the development of spindle MBC. The subtype-specific proteomes of MBC tumors highlight an ability to impact precision therapies to improve the survival of women with this aggressive form of breast cancer.

4-2 Translational applications with 3D modeling

Modeling neoplastic progression in mammary organoids with multiple phenotypes and mixed cell populations remains a challenge [134]. Here, we have developed a 3D mammary gland-like structure that resembles morphological and phenotypic aspects of the normal breast and allows highly reproducible quantification in vitro.

Early work in the field discovered that human breast cancers originate from cells within the terminal duct lobular unit (TDLU), a composite of a duct and several lobules. Here, we selected MCF10A to model normal breast cell function. Although MCF10A has been previously established in 2D and scaffold-based 3D cultures using Matrigel [10], [13], [71]), they typically are limited in tissue-specific marker expression, physiologic sizes (50-200 μm), and are known to dramatically influence stem cell populations. Our MCF10A organoids in the hanging drop system potentially mitigate these issues, and offer increased flexibility in modeling neoplastic progression while keeping certain populations intact, such as progenitors. We developed MCF10A organoids that were strikingly larger compared to other 3D models of MCF10A, and they ranged from 900-1200 μm with an average diameter of 1070 ± 105 μm after 2 weeks in culture. They also exhibited multi-acinar structures resembling aspects of normal breast TDLU. Recently, Qu and others [75] have investigated the use of MCF10A as an accurate model of normal breast, and found mixed luminal and basal expression of MCF10A acini. Our findings also confirmed a mixed luminal and basal expression of MCF10A organoids. Cytokeratin status has been intensely studied in clinical pathology on breast lesions and benign tissue, with distinct phenotypes described as stem/progenitor (CK 5/6+), glandular progenitor (CK 5/6+ and CK 18+), and committed glandular (CK 18+) [9],[138]. Interestingly, a population of CK 5/6+ cells on the basolateral side of MCF10A organoids appeared to co-express CK 18, suggesting glandular progenitor status. These

findings seem to suggest that the MCF10A model, when grown into 3D organoids, recapitulates some of the cellular heterogeneity present in normal breast TDLU. The size of the organoids may reflect not only the tightness of cellular interactions within the organoid but may also be a crucial determinant of cell biology [136]. As seen in recent organotypic culture studies, larger sphere sizes may correlate with the ability of organoids to generate and display multiple phenotypes in 3D, seen in other cell types such as the development of multilayered optic cup structures with hESCs growing to ~550 μm diameters [130], and human brain organoids of ~2 mm diameters that produced regions of multiple cell types [137]. Overall, this ability to form physiologic acinar structures in 3D is generally thought to require a scaffold setting because the support matrix provides the necessary pseudo-basement membrane. However, based on our studies, we postulate that the scaffold material itself might limit organoid expansion, by mechanical constraints and/or limiting the supply of critical nutrients [139], and that the actual concentration of Matrigel or other scaffold substrates used in 3D culture models should be lowered dramatically, and only used upon cell seeding to enable differentiation and self-organization activities.

Available data suggest that hanging drop culture might provide advantages over other 3D spheroid assays for achieving organoid expansion, such as conventional 3D gel embedded or on-top Matrigel culture, or microfluidic systems for cancer progression modeling [141],[78] which show significant restriction in size (<300 μm diameter) and proliferation potential, possibly due to mechanical confinement. Also, when using microsensors, the oxygen levels of multi-spheroids grown in conventional U-bottom 96 well plates is greatly reduced after 18 hours, while hanging droplets with only one spheroid had stabilized oxygen values [139]. Our observation of irregular morphologies, lumen filling, and no acinar formation in 96 well U-bottom plates, suggests that hanging droplets are organoid-supportive by providing more oxygen, enhancing autocrine effects

considering the use of small media volumes. In addition, MCF10A cells were not able to aggregate into tight single structures without FBS at seeding, consistent with literature, in which FBS was shown to augment cellular differentiation [130].

When looking more closely at the enhanced differentiation pattern of MCF10A organoids, we demonstrated the presence of a stem cell population where previous studies of MCF10A cells in 3D Matrigel systems were not able to observe stem cell populations and actually reported a loss of stem/progenitor markers when moving from 2D to 3D systems [75]. This suggests the MCF10A cell line, when using drastically lower concentrations of Matrigel in 3D hanging drop, could maintain a self-renewing progenitor population. This is a potential advantage of growing cells in a free-floatation context, whereby the investigation of stem cell populations in real-time can be readily monitored despite having less physiologically relevant ECM mimicry compared to conventional 3D scaffold-based platforms.

Overall, the two major substrata used in 3D breast models, collagen I gels and Matrigel, are essential for providing the biophysical context and cues for mimicking the microenvironment, although it remains challenging for 3D systems to mimic specific features of the in vivo case [136]. Organoids grown under free suspension may experience less pressure cues than scaffold-embedded culturing, which might be a potential drawback. The lack of ECM biophysical forces could explain why MCF10A organoids undergo substantial expansion to approximately twenty-five times the size of normal breast acini, but this ability to expand highlights a tight cellular commitment to the normal morphogenetic program. Moreover, a major disadvantage of gel systems is their restricted ability to express tissue-specific markers [154]. Our data support the idea that a 3D free-floatation context might provide a more permissive environment for differentiation.

We also showed that MCF10A could undergo malignant neoplastic progression by exogenous treatments rapidly, within days of treatment, and that it is relatively easy to quantify both morphology and phenotypic properties of such organoid states. We selected treatment with TGFβ and the hypoxia-mimetic agent, $CoCl_2$, given their reported effect in induction of EMT and malignant transformation in breast cancer, and demonstrated that the hanging drop organoid system allows expansion of cell populations with normal and neoplastic phenotypes, which can be applied to the study of breast tumorigenesis. Unexpectedly, our results revealed the previously unknown ability of MCF10A cells to undergo de novo synthesis and deposition of collagen I upon EMT induction. Since conventional 3D gels often make use of collagen-containing matrices, the production of collagen from transformed normal epithelial cells would be an indiscernible or challenging feature to both measure and capture.

Due to growing interest in patient tumor derived organoids for purposes of biobanking [152] and single cell sequencing [153], which have focused on the primary epithelial cell compartment, there is a great need to unravel patient tumor heterogeneity with high-throughput quantitative approaches. Here, we have provided a proof-of-concept that our 3D organoid platform is useful to study the microenvironment, particularly epithelial-stromal cross-talk during neoplastic initiation and progression. This might offer an ideal system to test new therapies to block stromal pro-tumorigenic influences on epithelial cells. In summary, our results indicate that benign breast cells such as MCF10A may have an intrinsic plasticity that is revealed by using a gel-free culture platform. We also show that the de novo remodeling processes can easily be detectable by histological methods. From a clinical perspective, the 3D organoid system can be applied to study rare and aggressive carcinomas such as TNBC and MBC, using patient primary tissues and performing a panel of treatments.

Despite an increase in studies using 3D tumor models in recent years, there remains a lack of standardization in spheroid and organoid analyses. Although parameters such as volume, sphericity, and area are very useful, they can be quite variable in the literature, stemming from inconsistencies in reproducibility and uniformity. Others have also noted overall shape is a highly variable parameter yet is an important indicator of tumorigenicity, so finding a quantifiable metric that correlates well with biological relevance will be a critical step going forward. Thus, as our 3D platform is a highly-reproducible method, future investigations involving patient-derived organoids which facilitate the study of large-scale drug screening and patient organoid biobanking ([158]) will become necessary in order to guide precision therapies.

4-3 Future Directions

In chapter 2, we identified specific pathways and drivers unique to each MBC, and in chapter 3 we optimized a 3D model capable of testing these therapeutically operable candidates using patient tissues. Future investigations are necessary to validate the set of genes/proteins we identified for each MBC subtype building upon our proteomics and WES analysis. A general workflow for such future work could be to validate these results by a combination of IHC, integrative multi-level omics profiling, functional assays, and 3D patient-derived organoid treatments. However, it is important to note that each MBC should be analyzed as a separate disease and experiments should be tailored to their major phenotypes. Below, we outline a brief translational vignette going forward for the analysis of MBC tumors:

Spindle MBC: These tumors tend towards an undifferentiated, EMT-like mesenchymal phenotype. Studies can perform IHC on E2F, MYC, and ribosomal proteins, "RP" (e.g. RPS17, RPL28, RPL6, RPL15), followed by an integrative analysis of proteome and phosphoproteome to interrogate the E2F/MYC/RP axis to reveal the connectivity of key transcription factors ("TFs") and kinases that

control the protein translation machinery. This will provide a proof-of-principle experiment for the reconstruction of the connectivity between TFs and kinases using proteomic and phosphoproteomic data along this signaling cascade and reveal candidate TFs (e.g. protein abundance vs phosphorylation levels) and screen each TF by functional phosphosites using a known kinase-substrate database. Also, since we found mutations in AHNAK and PI3K/MTOR genes, these experiments could reveal links to these key pathways and ribosome biogenesis and provide functional responses of oxidative phosphorylation. In addition, the MMTVcre;Ccn6$^{fl/fl}$ KO mice can be further investigated to identify actionable targets of CCN6 treated mice compared to BSA controls and normal mammary mouse tissues with a follow-up proteomics analysis. Organoids will also be developed using human and mouse primary tissues and subjected to CCN6 treatment.

Squamous MBC: Tumors present with epithelioid status and are more similar to a squamous cell carcinoma phenotype, have a prominent immune profile, and a hypothesis is that these tumors could be potentially responsive to immunotherapies. Genomically, these tumors were observed with mutations in PI3K/MTOR and MUC17, and were upregulated at the protein level in several inflammation-associated pathways (IFNγ, TNFα, NFκB). We also suggest an integrative analysis using IHC of these immune-related markers, and proteomics and phosphoproteomics profiling analysis. However, for these tumors, one could instead interrogate T cell activation and interconnected functional modules of TF-kinase signaling and TF-TF interactions to identify important immune regulators. Single-cell sequencing can then be used to cluster immune cells from tumor cells and provide insight into tumor infiltrating T cell phenotypes. A database of 3D patient-derived organoids can be established using a biobanking method of patient tumors and a

screening of these organoids using panel of available immunotherapies (e.g. PD-L1 inhibitor, durvalumab).

Sarcomatoid MBC: This metaplastic component tends toward a more differentiated, EMT-like mesenchymal phenotype, and was the only MBC subtype to present a similar genomic and proteomic profile with molecular and protein alterations in extracellular matrix receptor, cadherin, and calcium ion signaling. For example, IHC analysis could specifically focus on a panel of protocadherin cluster, which had a novel mutational profile in these tumors. A proteomics and phosphoproteomics analysis could potentially resolve the cadherin interactome and identify cell-cell adhesion activity, and demonstrate which kinases are involved in regulating cadherin activity comparing sarcoma tumor samples as controls. We could also examine cadherin signaling by cadherin-dependent adhesion functional assays, for example by fluorescently labeled cadherin fusion proteins (Ca+2 dependent adhesion) and Ca+2 binding properties (e.g. immunoblotting on the levels of N-cad or VE-Cad vs. interactions with β-catenin). In addition, single-cell analysis using sarcomatoid vs. sarcoma samples can cluster mesenchymal cell phenotypes to elucidate whether sarcomatoid MBC tends more toward the molecular properties of epithelial or mesenchymal tumors. If the latter, a hypothesis is that sarcomatoid tumors should be treated like other sarcomas, and 3D patient-derived organoids might involve treatments using a combination of doxorubicin and ifosfamide.

Due to the rarity of the disease limited access to available patient tumor samples, a typical study might utilize a small patient cohort. A future strategy for rare tumor analysis is to build a growing database of metaplastic carcinoma samples that can assemble multi-level omics datasets (from both primary tissues and patient-derived organoids) to better consolidate information in a patient-stratified format.

4-4 Final Thoughts

Our proteomic and genomic data sets suggest that a clear snapshot of the microenvironment brings a molecular context to processes like EMT and metabolism, which could point to which features of cellular plasticity such as CSC phenotypes that might be dominant or enriched within a particular MBC subtype in seeking novel candidate therapies. In addition, the data indicates that there may be certain subtype-specific differences in MBC that may allude to the presence of metastable E/M states along an epithelial (E) to mesenchymal (M) spectrum [28]. EMT was the top activated process compared to other TNBCs on the protein level, and this feature is also the most widely reported by histological evidence. Within MBC subtypes, it might suggest that squamous tends toward a basal or epithelioid phenotype that is apparently driven by inflammation, whereas spindle tends toward a mesenchymal status but is more poorly differentiated and consists of a seemingly "primitive" ECM relative to sarcomatoid. Among the MBCs, sarcomatoid presented both histologically and proteomically with higher mesenchymal status (e.g. higher EMT and ECM signaling) than spindle and squamous forms, possibly indicating a more extensive and/or well-differentiated ECM. Thus, our results broaden the notion of the conventional EMT enriched feature in MBCs suggesting that each MBC might reside in much different stages of EMT/MET, and general stemness states that influence their invasiveness. We also showed a proteomic link to genomic features of MBCs, which could prompt future pathognomonic studies that explore how each MBC subtype are tightly regulated, and how their complex epithelial–mesenchymal "metastable" phenotypes might contribute to increased aggressiveness and poor survival outcomes. In addition, the knowledge of proteomic and genomic alterations in MBC could be further tested using our 3D neoplastic breast organoid platform that produces highly uniform, reproducible 3D structures that are easy to standardize against a variety of parameters. MCF10A

organoids contained cells of multiple phenotypic lineages, including those along the epithelial and mesenchymal spectrum, and those that include progenitor-like populations. Our 3D organoid platform also revealed the ability to quantify and define cellular-level phenotypic changes and organoid-level morphologic changes that could be stratified at the patient level if primary tissues were used in this system. The one droplet – one organoid simplicity and rapid treatment assay strategy opens the door for standardization in a high-throughput testing.

Appendix 1 – MMTVcre;Ccn6$^{fl/fl}$ metaplastic carcinomas express IGF2BP2, HMGA2 and EMT markers and are recapitulated in mouse derived organoids

The following are work performed by the Kleer lab, in which I was a co-author and contributed to development of mouse-derived organoids from primary tissues, sectioning, histology and immunohistochemistry. In this study by McMullen, E. R. et al. CCN6 regulates IGF2BP2 and HMGA2 signaling in metaplastic carcinomas of the breast. Breast Cancer Res. Treat. 172, 577–586 (2018), our lab generated a novel MMTVcre;Ccn6$^{fl/fl}$ knockout mouse using the Cre-lox system to specifically study the effect of Ccn6 deletion in the mammary epithelium. Using this model, we discovered that Ccn6 is important for normal mammary gland development and tumorigenesis. MMTVcre;Ccn6$^{fl/fl}$ mice mammary glands had fewer terminal buds, decreased breast complexity, and displayed ductal hypoplasia when compared to wild-type mice. MMTVcre;Ccn6$^{fl/fl}$ mice developed mammary tumors morphologically and transcriptionally similar to high-grade spindle and squamous human metaplastic carcinomas.

IGF2BP2 (insulin-like growth factor 2 mRNA binding protein 2/IMP2) is an oncofetal protein important for embryonic development and downregulated in normal adult tissues, which functions by binding and stabilizing mRNAs to extend their half-life. In breast cancer, IGF2BP2 overexpression is associated with decreased cell adhesion and migration. HMGA2 (high mobility group AT-Hook 2), another oncofetal protein, functions as an architectural transcription factor. Similar to IGF2BP2, HMGA2 contributes to embryonic development and tumorigenesis. HMGA2 levels are associated with migration, EMT, and worse outcome in breast cancer. IGF2BP2 and HMGA2 have a reciprocal relationship as IGF2BP2 prevents Let7-mediated

HMGA2 inhibition and IGF2BP2 gene transcription and protein expression are highly dependent on HMGA2. MMTVcre;Ccn6$^{fl/fl}$ mice tumors recapitulate human high-grade triple-negative spindle/squamous metaplastic carcinomas. However, whether they express IGF2BP2 and HMGA2 proteins is unknown. Here, we found that all MMTVcre;Ccn6$^{fl/fl}$ tumors showed increased expression of IGF2BP2 and HMGA2 by immunohistochemistry. IGF2BP2 protein was expressed in the cytoplasm of cancer cells, while HMGA2 was exclusively nuclear (Fig. A-1). IGF2BP2 and HMGA2 expression were low in adjacent normal mammary epithelium of MMTVcre;Ccn6$^{fl/fl}$ metaplastic carcinomas (Fig. A-1). Our laboratory has previously reported that CCN6 knockdown in non-tumorigenic breast cells induces a spindle and invasive phenotype with up-regulation of the EMT transcription factors SNAI1 and ZEB1. Consistent with their spindle morphology, MMTVcre;Ccn6$^{fl/fl}$ tumors also had upregulated expression ZEB1 and SNAI2 (Fig. A-1). Immunostaining of organoids derived from MMTVcre;Ccn6$^{fl/fl}$ tumors demonstrated high expression of IGF2BP2 and HMGA2 proteins in the spindle cancer cells, similar to the observed expression in whole tumor sections.

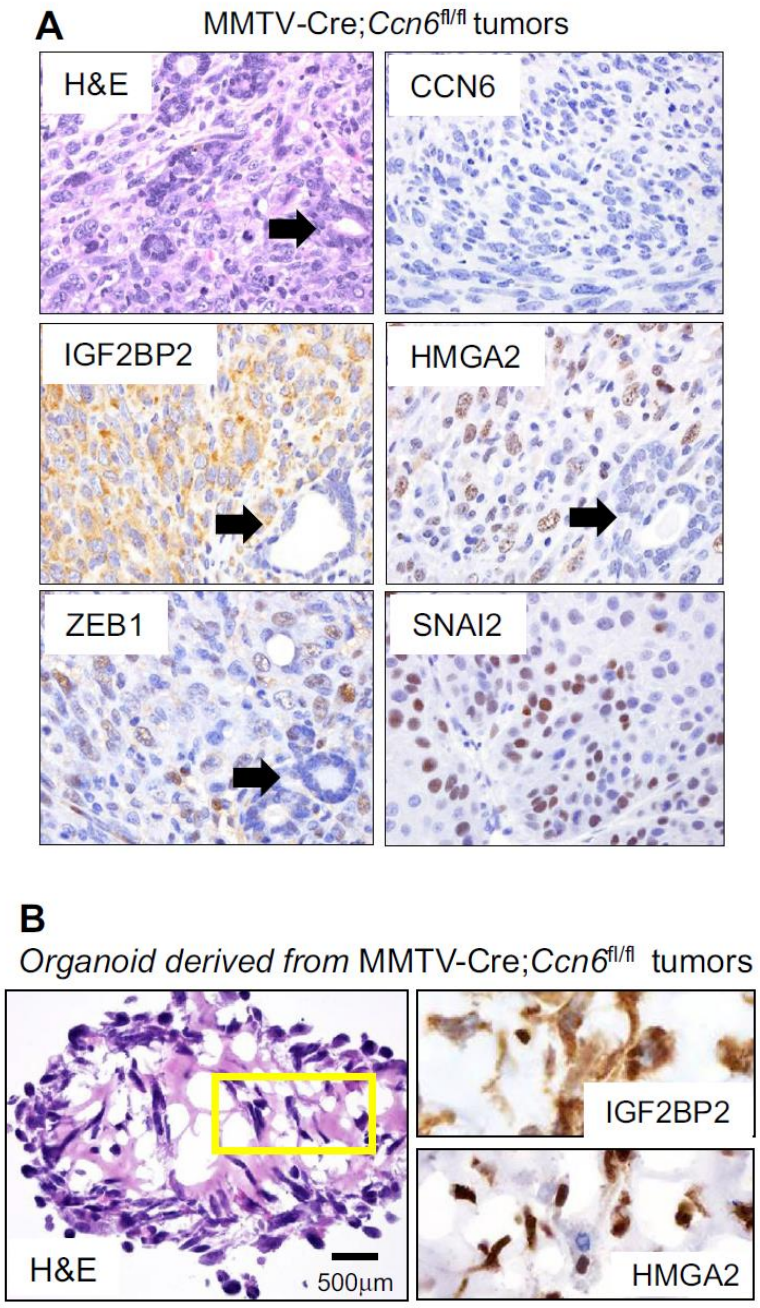


Figure A-1: MMTV-cre;Ccn6$^{fl/fl}$ metaplastic carcinomas express IGF2BP2, HMGA2, and EMT markers. A) Histological sections of MMTVcre;Ccn6$^{fl/fl}$ tumors showing low CCN6, high IGF2BP2 and HMGA2 proteins and upregulated expression of EMT transcription factors ZEB1 and SNAI2. Arrows indicate normal glandular structures in the murine mammary gland. B) Organoids developed from MMTVcre;Ccn6$^{fl/fl}$ tumors demonstrating a spindle morphology and high levels of IGF2BP2 and HMGA2 proteins.

Appendix 2 - Recombinant human CCN6 protein reduces IGF2BP2 an HMGA2 expression and regulates metaplastic tumor growth in vivo

The Kleer lab reported that recombinant human CCN6 protein regulates IGF2BP2 and HMGA2 expression, and is sufficient to reduce growth of MMTV‑cre;$Ccn6^{fl/fl}$ metaplastic carcinomas in vivo [108]. As CCN6 is a secreted protein, we used recombinant human CCN6 (rhCCN6) to elucidate its role on IGF2BP2 and HMGA2 regulation. We employed MDA-MB-231 and −468 cells, which are TNBC cells with a spindle morphology and gene expression profiles of mesenchymal-like and basal type-A breast carcinomas, respectively. Addition of rhCCN6 was sufficient to reduce IGF2BP2 and HMGA2 protein levels in both breast cancer cell lines (Figure A-2). Furthermore, ectopic overexpression of CCN6 in MDA-MB-231 cells reduced ZEB1 and SNAI2 protein levels compared to controls. The effect of rhCCN6 was investigated in vivo, using orthotopic syngeneic model by transplanting MMTVcre;$Ccn6^{fl/fl}$ tumors into the mammary fat pads of mice with the same genetic background (FVB). Mice were treated with rhCCN6 (1 ng/g, i.v., twice a week, initiated 2 weeks after transplantation) or BSA in the same regimen. rhCCN6 significantly reduced tumor growth, induced a morphological change from spindled towards epithelial, and reduced the expression of IGF2BP2 and HMGA2, compared to BSA treatment (Figure A-2). These findings demonstrated that CCN6 regulates IGF2BP2 and HMGA2 signaling in spindle carcinoma cells and that rhCCN6 is sufficient to reduce tumor growth in vivo.

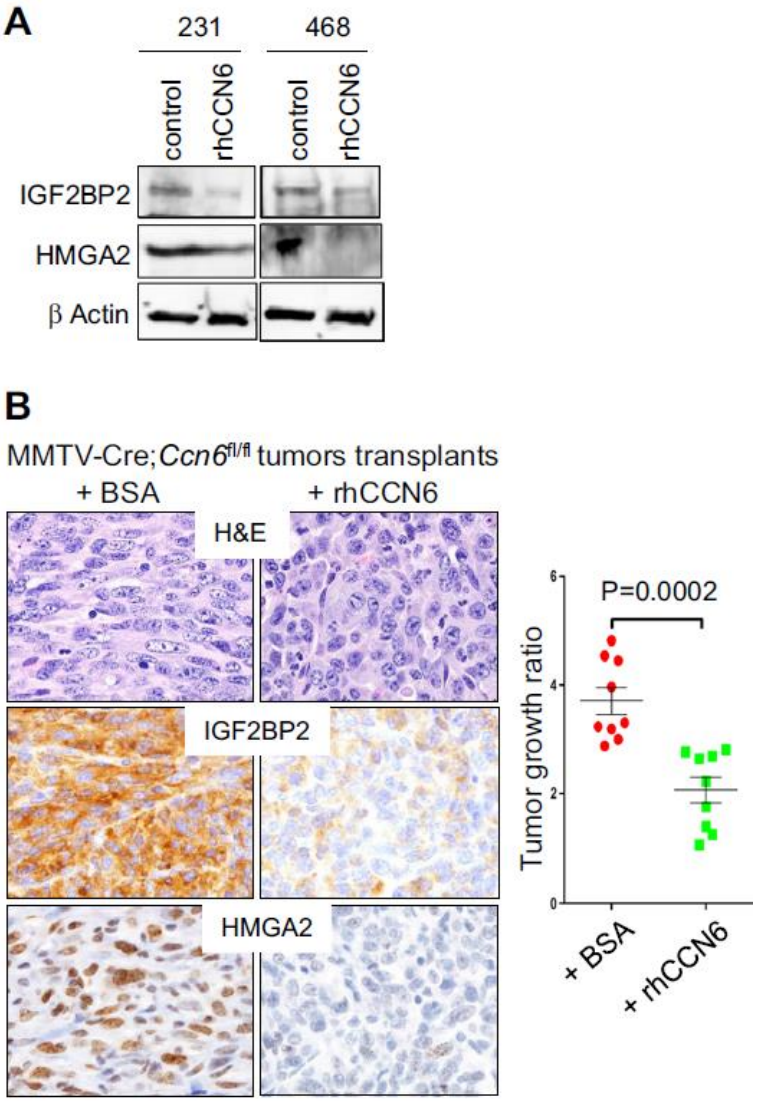


Figure A-2: Human recombinant CCN6 protein reduces expression of IGF2BP2 and HMGA2 and regulates metaplastic tumor growth in vivo. A) MDA-MB-231 and −468 breast cancer cells with spindle morphology were treated with rhCCN6 or BSA. B) MMTVcre;Ccn6$^{fl/fl}$ tumors were implanted in the mammary fat pads of 8–10 weeks old female MMTVcre;Ccn6$^{fl/fl}$ mice to avoid the possible contribution of microenvironment-derived CCN6 on tumor growth. Mice were treated with rhCCN6 (1 ng/g) or BSA (1 ng/g) i.v. twice a week starting 2 weeks after tumor implantation. Shown are histological sections of the tumors after 3 weeks of treatment. The graph shows differences in tumor growth at 15 days of treatment with rhCCN6 or BSA.

Bibliography

[1] H. M. and L. Hinck, "Mammary Gland Development," *Wiley Interdiscip. Rev. Dev. Biol.*, vol. 1, no. 4, pp. 533–557, 2012.

[2] M. Wang *et al.*, "Role of tumor microenvironment in tumorigenesis," *J. Cancer*, vol. 8, no. 5, pp. 761–773, 2017.

[3] American Cancer Society, "American Cancer Society: Breast Cancer Facts & Figures," *Am. Cancer Soc. Inc.*, 2012.

[4] J. Mehanna, F. G. H. Haddad, R. Eid, M. Lambertini, and H. R. Kourie, "Triple-negative breast cancer: Current perspective on the evolving therapeutic landscape," *Int. J. Womens. Health*, vol. 11, pp. 431–437, 2019.

[5] M. Hubalek, T. Czech, and H. Müller, "Biological Subtypes of Triple-Negative Breast Cancer," *Breast Care*, vol. 12, no. 1, pp. 8–14, 2017.

[6] J. E. Visvader, "Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis," *GENES Dev.*, vol. 23, pp. 2563–2577, 2009.

[7] S. R. Wellings, "A hypothesis of the origin of human breast cancer from the terminal ductal lobular unit.," *Pathol. Res. Pract.*, vol. 166, no. 4, pp. 515–35, Apr. 1980.

[8] B. Tiede and Y. Kang, "From milk to malignancy: The role of mammary stem cells in development, pregnancy and breast cancer," *Cell Res.*, vol. 21, no. 2, pp. 245–257, 2011.

[9] W. Boecker and H. Buerger, "Evidence of progenitor cells of glandular and myoepithelial cell lineages in the human adult female breast epithelium: a new progenitor (adult stem) cell concept.," *Cell Prolif.*, vol. 36 Suppl 1, pp. 73–84, 2003.

[10] J. Debnath, S. K. Muthuswamy, and J. S. Brugge, "Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures," *Methods*, vol. 30, no. 3, pp. 256–268, 2003.

[11] M. H. Barcellos-Hoff, J. Aggeler, T. G. Ram, and M. J. Bissell, "Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane.," *Development*, vol. 105, no. 2, pp. 223–235, 1989.

[12] A. T. Lo, H. Mori, J. Mott, and M. J. Bissell, "Constructing three-dimensional models to study mammary gland branching morphogenesis and functional differentiation," *J. Mammary Gland Biol. Neoplasia*, vol. 17, no. 2, pp. 103–110, 2012.

[13] O. W. Petersen, L. Ronnov-Jessen, A. R. Howlett, and M. J. Bissell, "Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells.," *Proc. Natl. Acad. Sci.*, vol. 89, no. 19, pp. 9064–9068, 1992.

[14] P. Vidi, M. J. Bissell, and S. A. Lelièvre, "Epithelial Cell Culture Protocols," vol. 945, pp. 193–219, 2013.

[15] J. R. Linnemann *et al.*, "Quantification of regenerative potential in primary human mammary epithelial cells," *Development*, vol. 142, no. 18, pp. 3239–3251, 2015.

[16] W. Tsuji and J. A. Plock, "Breast Cancer Metastasis," *Introd. to Cancer Metastasis*, vol. 320, pp. 13–31, 2017.

[17] S. Lal, A. E. McCart Reed, X. M. de Luca, and P. T. Simpson, “Molecular signatures in breast cancer,” *Methods*, vol. 131, pp. 135–146, 2017.
[18] M. Komatsu *et al.*, “Molecular features of triple negative breast cancer cells by genome-wide gene expression profiling analysis,” *Int. J. Oncol.*, vol. 42, no. 2, pp. 478–506, 2013.
[19] R. T. Lawrence *et al.*, “The Proteomic Landscape of Triple-Negative Breast Cancer,” *Cell Rep.*, vol. 11, no. 4, pp. 630–644, 2015.
[20] B. Na *et al.*, “Therapeutic targeting of BRCA1 and TP53 mutant breast cancer through mutant p53 reactivation,” *NPJ Breast Cancer*, vol. 5, no. 14, pp. 1–14, 2019.
[21] S. R. Wellings, H. M. Jensen, and R. G. Marcum, “An Atlas of Subgross Pathology of the Human Breast With Special Reference to Possible Precancerous Lesions,” *JNCI J. Natl. Cancer Inst.*, vol. 55, no. 2, pp. 231–273, 1975.
[22] P. Jézéquel *et al.*, “Identification of three subtypes of triple-negative breast cancer with potential therapeutic implications,” *Breast Cancer Res.*, vol. 21, no. 1, pp. 1–14, 2019.
[23] A. C. Garrido-Castro, N. U. Lin, and K. Polyak, “Insights into molecular classifications of triple-negative breast cancer: Improving patient selection for treatment,” *Cancer Discov.*, vol. 9, no. 2, pp. 176–198, 2019.
[24] M. A. Nieto, “Epithelial plasticity: A common theme in embryonic and cancer cells,” *Science (80-.).*, vol. 342, no. 6159, 2013.
[25] S. A. Mani *et al.*, “The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells,” *Cell*, vol. 133, no. 4, pp. 704–715, 2008.
[26] M. Takaishi, M. Tarutani, J. Takeda, and S. Sano, “Mesenchymal to epithelial transition induced by reprogramming factors attenuates the malignancy of cancer cells,” *PLoS One*, vol. 11, no. 6, pp. 1–15, 2016.
[27] U. D. Kahlert, J. V. Joseph, and F. A. E. Kruyt, “EMT- and MET-related processes in nonepithelial tumors: importance for disease progression, prognosis, and therapeutic opportunities,” *Mol. Oncol.*, vol. 11, no. 7, pp. 860–877, 2017.
[28] G. Sannino, A. Marchetto, T. Kirchner, and T. G. P. Grünewald, “Epithelial-to-mesenchymal and mesenchymal-to-epithelial transition in mesenchymal tumors: A paradox in sarcomas?,” *Cancer Res.*, vol. 77, no. 17, pp. 4556–4561, 2017.
[29] P. Mallini, T. Lennard, J. Kirby, and A. Meeson, “Epithelial-to-mesenchymal transition: What is the impact on breast cancer stem cells and drug resistance,” *Cancer Treat. Rev.*, vol. 40, no. 3, pp. 341–348, 2014.
[30] J. H. Taube *et al.*, “Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 35, pp. 15449–15454, 2010.
[31] D. D. Tran, C. A. S. Corsa, H. Biswas, R. L. Aft, and G. D. Longmore, “Temporal and Spatial Cooperation of Snail1 and Twist1 during Epithelial-Mesenchymal Transition Predicts for Human Breast Cancer Recurrence,” *Mol. Cancer Res.*, vol. 9, no. 12, pp. 1644–1657, 2011.
[32] A. P. Morel *et al.*, “EMT inducers catalyze malignant transformation of mammary epithelial cells and drive tumorigenesis towards claudin-low tumors in transgenic mice,” *PLoS Genet.*, vol. 8, no. 5, 2012.
[33] H. Liu *et al.*, “Cancer stem cells from human breast tumors are involved in spontaneous metastases in orthotopic mouse models,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 42, pp. 18115–18120, 2010.
[34] A. Kreso and J. E. Dick, “Evolution of the cancer stem cell model,” *Cell Stem Cell*, vol.

14, no. 3, pp. 275–291, 2014.
[35] E. C. Kordon and G. H. Smith, “An entire functional mammary gland may comprise the progeny from a single cell,” *Development*, vol. 125, no. 10, pp. 1921–1930, 1998.
[36] M. K. Juric *et al.*, “Milestones of hematopoietic stem cell transplantation - From first human studies to current developments,” *Front. Immunol.*, vol. 7, pp. 1–16, 2016.
[37] M. L. Oon, A. A. Thike, S. Y. Tan, and P. H. Tan, “Cancer stem cell and epithelial–mesenchymal transition markers predict worse outcome in metaplastic carcinoma of the breast,” *Breast Cancer Res. Treat.*, vol. 150, no. 1, pp. 31–41, 2015.
[38] S. P. McDermott and M. S. Wicha, “Targeting breast cancer stem cells,” *Mol. Oncol.*, vol. 4, no. 5, pp. 404–419, 2010.
[39] K. B. Deome, L. J. Faulkin, H. A. Bern, and P. B. Blair, “Development of Mammary Tumors from Hyperplastic Alveolar Nodules Transplanted into Gland-free Mammary Fat Pads of Female C3H Mice,” *Cancer Res.*, vol. 19, no. 5, p. 515, 1959.
[40] J. Zhou, Q. Chen, Y. Zou, S. Zheng, and Y. Chen, “Stem Cells and Cellular Origins of Mammary Gland: Updates in Rationale, Controversies, and Cancer Relevance,” *Stem Cells Int.*, vol. 2019, pp. 1–12, 2019.
[41] R. Bjerkvig, B. B. Tysnes, K. S. Aboody, and J. Najbauer, “Current Controversies and New Insights,” *Nat. Rev.*, vol. 5, pp. 899–904, 2005.
[42] X. Yang, H. Wang, and B. Jiao, “Mammary gland stem cells and their application in breast cancer,” *Oncotarget*, vol. 8, no. 6, pp. 10675–10691, 2017.
[43] Y. Kang, S. Kang, Q. Li, and X. Zheng, “Mixed epithelial and mesenchymal metaplastic carcinoma (carcinosarcoma) of the breast: A case report,” *Eur. J. Med. Res.*, vol. 19, no. 1, pp. 1–6, 2014.
[44] Y. Zhang, K. Toy, and C. Kleer, “Metaplastic breast carcinomas are enriched in markers of tumor-initiating cells and epithelial to mesenchymal transition,” *Mod. Pathol.*, vol. 25, no. 2, pp. 178–184, 2012.
[45] D. El Zein *et al.*, “Metaplastic Carcinoma of the Breast Is More Aggressive Than Triple-negative Breast Cancer: A Study From a Single Institution and Review of Literature,” *Clin. Breast Cancer*, vol. 17, no. 5, pp. 382–391, 2017.
[46] C. L. Cooper, R. Z. Karim, C. Selinger, H. Carmalt, C. S. Lee, and S. A. O’Toole, “Molecular alterations in metaplastic breast carcinoma,” *J. Clin. Pathol.*, vol. 66, no. 6, pp. 522–528, 2013.
[47] S. Abouharb and S. Moulder, “Metaplastic Breast Cancer: Clinical Overview and Molecular Aberrations for Potential Targeted Therapy,” *Curr. Oncol. Rep.*, vol. 17, no. 3, pp. 1–7, 2015.
[48] J. Huvos AG, Lucas JC, and Foote, FW, “Metaplastic breast carcinoma. Rare form of mammary cancer.,” *N Y State J Med*, vol. 73, no. 9, pp. 1078–82, 1973.
[49] H. A. Oberman, “Metaplastic carcinoma of the breast. A clinicopathologic study of 29 patients,” *American Journal of Surgical Pathology*, vol. 11, no. 12. pp. 918–929, 1987.
[50] N. S. Salemis, “Metaplastic carcinoma of the breast with mesenchymal differentiation (carcinosarcoma). A unique presentation of an aggressive malignancy and literature review,” *Breast Dis.*, vol. 37, no. 3, pp. 169–175, 2018.
[51] A. Cimino-Mathews *et al.*, “A clinicopathologic analysis of 45 patients with metaplastic breast carcinoma,” *Am. J. Clin. Pathol.*, vol. 145, no. 3, pp. 365–372, 2016.
[52] M. P. Budzik, J. Patera, M. Sobol, A. I. Czerw, A. Deptała, and A. M. Badowska-Kozakiewicz, “Clinicopathological characteristics of metaplastic breast cancer – analysis

of the basic immunohistochemical profile and comparison with other invasive breast cancer types," *Breast*, vol. 43, pp. 135–141, 2019.
[53] B. B. Choi and K. S. Shu, "Metaplastic carcinoma of the breast: Multimodality imaging and histopathologic assessment," *Acta radiol.*, vol. 53, no. 1, pp. 5–11, 2012.
[54] H. Lien, C. Lin, T. Mao, S. Kuo, and C. Hsiao, "p53 overexpression and mutation in metaplastic carcinoma of the breast : genetic evidence for a monoclonal origin of both the carcinomatous and the heterogeneous sarcomatous components," *J. Pathol.*, vol. 204, pp. 131–139, 2004.
[55] X. Wang *et al.*, "Metaplastic Carcinoma of the Breast : p53 Analysis Identified the Same Point Mutation in the Three Histologic Components," *Mod Pathol*, vol. 14, no. 11, pp. 1183–1186, 2001.
[56] Y. Zhang, K. A. Toy, and C. G. Kleer, "Metaplastic breast carcinomas are enriched in markers of tumor-initiating cells and epithelial to mesenchymal transition," *Mod. Pathol.*, vol. 25, no. 2, pp. 178–184, 2012.
[57] B. T. Hennessy *et al.*, "Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics," *Cancer Res.*, vol. 69, no. 10, pp. 4116–4124, 2009.
[58] C. Ng *et al.*, "The landscape of somatic genetic alterations in metaplastic breast carcinomas," *Clin. Cancer Res.*, vol. 23, no. 14, pp. 3859–3870, 2017.
[59] B. E. Avigdor *et al.*, "Whole-exome sequencing of metaplastic breast carcinoma indicates monoclonality with associated ductal carcinoma component," *Clin. Cancer Res.*, vol. 23, no. 16, pp. 4875–4884, 2017.
[60] S. Piscuoglio *et al.*, "Genomic and transcriptomic heterogeneity in metaplastic carcinomas of the breast," *npj Breast Cancer*, vol. 3, no. 1, 2017.
[61] S. Piscuoglio *et al.*, "Genomic and transcriptomic heterogeneity in metaplastic carcinomas of the breast," *npj Breast Cancer*, vol. 3, no. 1, 2017.
[62] and C. G. K. E.E. Martin, W. Huang, T. Anwar, C. Arellano-Garcia, B. Burman, J-L. Guan, M.E. Gonzalez, "MMTV-cre;Ccn6 knockout mice develop tumors recapitulating human metaplastic breast carcinomas," *Oncogene*, vol. 36, no. 16, pp. 2275–2285, 2017.
[63] M. N. Tran and C. G. Kleer, "Matricellular CCN6 (WISP3) protein: a tumor suppressor for mammary metaplastic carcinomas," *J. Cell Commun. Signal.*, vol. 12, no. 1, pp. 13–19, 2018.
[64] A. Pal, W. Huang, K. A. Toy, and C. G. Kleer, "CCN6 Knockdown Disrupts Acinar Organization of Breast Cells in Three-dimensional Cultures through Up-regulation of Type III TGF-β Receptor," *Neoplasia*, vol. 14, no. 11, pp. 1067–74, 2012.
[65] G. Y. Lee, P. a Kenny, E. H. Lee, and M. J. Bissell, "Three-dimensional culture models of normal and malignant breast epithelial cells," *Nat. Methods*, vol. 4, no. 4, pp. 359–365, 2007.
[66] E. M. Chandler *et al.*, "Stiffness of photocrosslinked RGD-alginate gels regulates adipose progenitor cell behavior," *Biotechnol. Bioeng.*, vol. 108, no. 7, pp. 1683–1692, 2011.
[67] R. C. Dutta and A. K. Dutta, *3D Cell Culture*. 2018.
[68] S. Marina and M. J. Bissell, "Organoids: a historical perspective of thinkin in three dimensions," *J. Cell Biol*, pp. 1–10, 2017.
[69] A. Pal and C. G. Kleer, "Three dimensional cultures: A tool to study normal acinar architecture vs. malignant transformation of breast cells," *J. Vis. Exp.*, no. 86, 2014.
[70] H. P. H. Naber, E. Wiercinska, P. ten Dijke, and T. van Laar, "Spheroid Assay to Measure

TGF-β-induced Invasion," *J. Vis. Exp.*, no. 57, p. e3337, 2011.
[71] A. Gaiko-Shcherbak, G. Fabris, G. Dreissen, R. Merkel, B. Hoffmann, and E. Noetzel, "The acinar cage: Basement membranes determine molecule exchange and mechanical stability of human breast cell acini," *PLoS One*, vol. 10, no. 12, pp. 1–20, 2015.
[72] X. Wang, L. Sun, M. V. Maffini, A. Soto, C. Sonnenschein, and D. L. Kaplan, "A complex 3D human tissue culture system based on mammary stromal cells and silk scaffolds for modeling breast morphogenesis and function," *Biomaterials*, vol. 31, no. 14, pp. 3920–3929, 2010.
[73] M. J. Bissell, H. G. Hall, and G. Parry, "How does the extracellular matrix direct gene expression?," *J. Theor. Biol.*, vol. 99, no. 1, pp. 31–68, 1982.
[74] G. Venugopalan *et al.*, "Multicellular architecture of malignant breast epithelia influences mechanics," *PLoS One*, vol. 9, no. 8, 2014.
[75] Y. Qu *et al.*, "Evaluation of MCF10A as a reliable model for normal human mammary epithelial cells," *PLoS One*, vol. 10, no. 7, pp. 1–16, 2015.
[76] M. Vinci *et al.*, "Advances in establishment and analysis of three-dimensional tumor spheroid-based functional assays for target validation and drug evaluation," *BMC Biol.*, vol. 10, no. March, 2012.
[77] C. R. Thoma, S. Stroebel, N. Rösch, B. Calpe, W. Krek, and J. M. Kelm, "A high-throughput-compatible 3D microtissue co-culture system for phenotypic RNAi screening applications," *J. Biomol. Screen.*, vol. 18, no. 10, pp. 1330–1337, 2013.
[78] Y. Zhou, T. Arai, Y. Horiguchi, K. Ino, T. Matsue, and H. Shiku, "Multiparameter analyses of three-dimensionally cultured tumor spheroids based on respiratory activity and comprehensive gene expression profiles," *Anal. Biochem.*, vol. 439, no. 2, pp. 187–193, 2013.
[79] J. M. Kelm, N. E. Timmins, C. J. Brown, M. Fussenegger, and L. K. Nielsen, "Method for generation of homogeneous multicellular tumor spheroids applicable to a wide variety of cell types," *Biotechnol. Bioeng.*, vol. 83, no. 2, pp. 173–180, 2003.
[80] A. P. Aijian and R. L. Garrell, "Digital Microfluidics for Automated Hanging Drop Cell Spheroid Culture," *J. Lab. Autom.*, vol. 20, no. 3, pp. 283–295, 2015.
[81] Y.-C. Tung, A. Y. Hsiao, S. G. Allen, Y. Torisawa, M. Ho, and S. Takayama, "High-throughput 3D spheroid culture and drug testing using a 384 hanging drop array," *Analyst*, vol. 136, no. 3, pp. 473–478, 2011.
[82] B. M. Leung, S. C. Lesher-Perez, T. Matsuoka, C. Moraes, and S. Takayama, "Media additives to promote spheroid circularity and compactness in hanging drop platform," *Biomater. Sci.*, vol. 3, no. 2, pp. 336–344, 2015.
[83] Y. Hasin, M. Seldin, and A. Lusis, "Multi-omics approaches to disease," *Genome Biol.*, vol. 18, no. 1, pp. 1–15, 2017.
[84] D. L. Gibbs, L. Gralinski, R. S. Baric, and S. K. McWeeney, "Multi-omic network signatures of disease," *Front. Genet.*, vol. 4, no. JAN, pp. 1–11, 2013.
[85] Z. Zhu *et al.*, "Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets," *Nat. Genet.*, vol. 48, no. 5, pp. 481–487, 2016.
[86] Y. V. Sun and Y. J. Hu, *Integrative Analysis of Multi-omics Data for Discovery and Functional Studies of Complex Human Diseases*, vol. 93. Elsevier Ltd, 2016.
[87] E. T. Cirulli and D. B. Goldstein, "Uncovering the roles of rare variants in common disease through whole-genome sequencing," *Nat. Rev. Genet.*, vol. 11, no. 6, pp. 415–425, 2010.

[88] J. Zhu, Z. Shi, J. Wang, and B. Zhang, “Empowering biologists with multi-omics data: Colorectal cancer as a paradigm,” *Bioinformatics*, vol. 31, no. 9, pp. 1436–1443, 2015.
[89] A. Piunti and A. Shilatifard, “Epigenetic balance of gene expression by polycomb and compass families,” *Science (80-.).*, vol. 352, no. 6290, 2016.
[90] Y. Kanai and E. Arai, “Multilayer-omics analyses of human cancers: Exploration of biomarkers and drug targets based on the activities of the International Human Epigenome Consortium,” *Front. Genet.*, vol. 5, pp. 1–7, 2014.
[91] C. Trapnell *et al.*, “Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation,” *Nat. Biotechnol.*, vol. 28, no. 5, pp. 511–515, 2010.
[92] R. A. Gupta *et al.*, “Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis,” *Nature*, vol. 464, no. 7291, pp. 1071–1076, 2010.
[93] X. Han, A. Aslanian, and J. R. Yates III, “Mass Spectrometry for Proteomics,” *Curr Opin Chem Biol*, vol. 12, no. 5, pp. 483–490, 2008.
[94] B. Domon and R. Aebersold, “Mass spectrometry and protein analysis,” *Science (80-.).*, vol. 312, no. 5771, pp. 212–217, 2006.
[95] A. I. Nesvizhskii, O. Vitek, and R. Aebersold, “Analysis and validation of proteomic data generated by tandem mass spectrometry.,” *Nat. Methods*, vol. 4, no. 10, pp. 787–97, 2007.
[96] A. Thompson *et al.*, “Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.,” *Anal. Chem.*, vol. 75, no. 8, pp. 1895–904, 2003.
[97] R. Moulder, S. D. Bhosale, D. R. Goodlett, and R. Lahesmaa, “Analysis of the plasma proteome using iTRAQ and TMT-based isobaric labelling,” *Mass Spectrom. Rev.*, vol. 37, no. 5, pp. 583–606, 2018.
[98] A. I. Nesvizhskii, “A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics.,” *J. Proteomics*, vol. 73, no. 11, pp. 2092–123, 2010.
[99] C. B. Clish, “Metabolomics: an emerging but powerful tool for precision medicine,” *Mol. Case Stud.*, vol. 1, no. 1, p. a000588, 2015.
[100] C. Gieger *et al.*, “Genetics meets metabolomics: A genome-wide association study of metabolite profiles in human serum,” *PLoS Genet.*, vol. 4, no. 11, 2008.
[101] E. Montagna *et al.*, “Heterogeneity of triple-negative breast cancer: Histologic subtyping to inform the outcome,” *Clin. Breast Cancer*, vol. 13, no. 1, pp. 31–39, 2013.
[102] F. Geyer *et al.*, “Molecular analysis reveals a genetic basis for the phenotypic diversity of metaplastic breast carcinomas,” *J. Pathol.*, vol. 216, no. July, pp. 399–407, 2010.
[103] J. S. Ross *et al.*, “Genomic profiling of advanced-stage, metaplastic breast carcinoma by next-generation sequencing reveals frequent, targetable genomic abnormalities and potential new treatment options,” *Arch. Pathol. Lab. Med.*, vol. 139, no. 5, pp. 642–649, 2015.
[104] R. R. Johnson *et al.*, “Metaplastic breast carcinoma: a clinical-pathologic study of 97 cases with subset analysis of response to neoadjuvant chemotherapy,” *Mod. Pathol.*, 2019.
[105] S. Bartels *et al.*, “CDKN2A loss and PIK3CA mutation in myoepithelial-like metaplastic breast cancer,” *J. Pathol.*, vol. 245, no. 3, pp. 373–383, 2018.
[106] G. Bataillon *et al.*, “High rate of PIK3CA mutations but no TP53 mutations in low-grade adenosquamous carcinoma of the breast,” *Histopathology*, vol. 73, no. 2, pp. 273–283, 2018.

[107] E. E. Martin *et al.*, "MMTV-cre;Ccn6 knockout mice develop tumors recapitulating human metaplastic breast carcinomas," *Oncogene*, vol. 36, no. 16, pp. 2275–2285, 2017.
[108] E. R. McMullen *et al.*, "CCN6 regulates IGF2BP2 and HMGA2 signaling in metaplastic carcinomas of the breast," *Breast Cancer Res. Treat.*, vol. 172, no. 3, pp. 577–586, 2018.
[109] H. J. Johansson, "Breast cancer quantitative proteome and proteogenomic landscape," *Nat. Commun.*, vol. 10, pp. 1–14, 2019.
[110] D. Kessner, M. Chambers, R. Burke, D. Agus, and P. Mallick, "ProteoWizard: Open source software for rapid proteomics tools development," *Bioinformatics*, vol. 24, no. 21, pp. 2534–2536, 2008.
[111] A. T. Kong, F. V. Leprevost, D. M. Avtonomov, D. Mellacheruvu, and A. I. Nesvizhskii, "MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics," *Nat. Methods*, vol. 14, no. 5, pp. 513–520, 2017.
[112] A. Keller, A. I. Nesvizhskii, E. Kolker, and R. Aebersold, "Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search," *Anal. Chem.*, vol. 74, no. 20, pp. 5383–5392, 2002.
[113] A. I. Nesvizhskii, A. Keller, E. Kolker, and R. Aebersold, "A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry abilities that proteins are present in a sample on the basis," *Anal. Chem.*, vol. 75, no. 17, pp. 4646–4658, 2003.
[114] A. K. Shanmugam, A. K. Yocum, and A. I. Nesvizhskii, "Utility of RNA-seq and GPMDB protein observation frequency for improving the sensitivity of protein identification by tandem MS," *J. Proteome Res.*, vol. 13, no. 9, pp. 4113–4119, 2014.
[115] M. M. Savitski, M. Wilhelm, H. Hahne, B. Kuster, and M. Bantscheff, "A Scalable Approach for Protein False Discovery Rate Estimation in Large Proteomic Data Sets," *Mol. Cell. Proteomics*, vol. 14, no. 9, pp. 2394–2404, 2015.
[116] F. Ghali *et al.*, "Tools (Viewer, Library and Validator) that Facilitate Use of the Peptide and Protein Identification Standard Format, Termed mzIdentML," *Mol. Cell. Proteomics*, vol. 12, no. 11, pp. 3026–3035, 2013.
[117] A. Argentini *et al.*, "moFF: a robust and automated approach to extract peptide ion intensities," *Nat. Methods*, vol. 13, no. 12, pp. 964–966, 2016.
[118] K. Ning, D. Fermin, and A. I. Nesvizhskii, "Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data," *J. Proteome Res.*, vol. 11, no. 4, pp. 2261–2271, 2012.
[119] R. J. A. Little and D. B. Rubin, *Statistical Analysis with Missing Data*, vol. 26, no. 3. 2002.
[120] I. C. Chen *et al.*, "Lack of efficacy to systemic chemotherapy for treatment of metaplastic carcinoma of the breast in the modern era," *Breast Cancer Res. Treat.*, vol. 130, no. 1, pp. 345–351, 2011.
[121] C. Perou, "Molecular Stratification of Triple-Negative Breast Cancers," *Oncologist*, vol. 15, no. suppl 5, pp. 744–749, 2010.
[122] N. B. La Thangue, "Chromatin Control--a Place for E2F and Myc to Meet," *Science (80-.).*, vol. 296, no. May, pp. 1034–1036, 2002.
[123] V. Prakash *et al.*, "Ribosome biogenesis during cell cycle arrest fuels EMT in development and disease," *Nat. Commun.*, vol. 10, pp. 1–16, 2019.
[124] A. I. Alford, K. M. Kozloff, and K. D. Hankenson, "The International Journal of Biochemistry Extracellular matrix networks in bone remodeling," *Int. J. Biochem. Cell Biol.*, vol. 65, pp. 20–31, 2015.

[125] G. Krings and Y. Y. Chen, “Genomic profiling of metaplastic breast carcinomas reveals genetic heterogeneity and relationship to ductal carcinoma,” *Mod. Pathol.*, vol. 31, no. 11, pp. 1661–1674, 2018.
[126] H. Sun *et al.*, “Single-cell RNA-Seq reveals cell heterogeneity and hierarchy within mouse mammary epithelia,” *J. Biol. Chem.*, vol. 293, no. 22, pp. 8315–8329, 2018.
[127] K. M. Welch-Reardon, N. Wu, and C. C. W. Hughes, “A role for partial endothelial-mesenchymal transitions in angiogenesis?,” *Arterioscler. Thromb. Vasc. Biol.*, vol. 35, no. 2, pp. 303–308, 2015.
[128] M. Simian, Y. Hirai, M. Navre, Z. Werb, a Lochter, and M. J. Bissell, “The interplay of matrix metalloproteinases, morphogens and growth factors is necessary for branching of mammary epithelial cells.,” *Development*, vol. 128, no. 16, pp. 3117–3131, 2001.
[129] C. G. Pal, A. and Kleer, “Three Dimensional Cultures: A Tool to Study Normal Acinar Architecture vs. Malignant Transformation of Breast Cells,” *J Vis Exp*, no. 86, p. e51311, 2014.
[130] T. Nakano *et al.*, “Self-formation of optic cups and storable stratified neural retina from human ESCs,” *Cell Stem Cell*, vol. 10, no. 6, pp. 771–785, 2012.
[131] M. E. Gonzalez *et al.*, “Mesenchymal Stem Cell-Induced DDR2 Mediates Stromal-Breast Cancer Interactions and Metastasis Growth,” *Cell Rep.*, vol. 18, no. 5, pp. 1215–1228, 2017.
[132] D. Wu and P. Yotnda, “Induction and Testing of Hypoxia in Cell Culture,” *J. Vis. Exp.*, no. 54, pp. 4–7, 2011.
[133] J. B. T. M. Roerdink and A. Meijster, “The Watershed Transform : Definitions , Algorithms and Parallelization Strategies,” vol. 41, pp. 1–40, 2001.
[134] B. A. Gusterson, D. T. Ross, V. J. Heath, and T. Stein, “Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer,” *Breast Cancer Res.*, vol. 7, no. 4, pp. 143–148, 2005.
[135] X. Yin, B. E. Mead, H. Safaee, R. Langer, J. M. Karp, and O. Levy, “Cell Stem Cell Engineering Stem Cell Organoids,” *Stem Cell*, vol. 18, pp. 25–38, 2016.
[136] C. R. Thoma, M. Zimmermann, I. Agarkova, J. M. Kelm, and W. Krek, “3D cell culture systems modeling tumor growth determinants in cancer target discovery,” *Adv. Drug Deliv. Rev.*, vol. 69–70, pp. 29–41, 2014.
[137] G. Quadrato *et al.*, “Cell diversity and network dynamics in photosensitive human brain organoids,” *Nature*, vol. 545, no. 7652, pp. 48–53, 2017.
[138] D. M. Abd El-Rehim *et al.*, “Expression of luminal and basal cytokeratins in human breast carcinoma,” *J. Pathol.*, vol. 203, no. 2, pp. 661–671, 2004.
[139] S. C. Lesher-Pérez *et al.*, “Dispersible oxygen microsensors map oxygen gradients in three-dimensional cell cultures,” *Biomater. Sci.*, vol. 5, no. 10, pp. 2106–2113, 2017.
[140] A. et al. Lance, “Increased extracellular matrix density decreases MCF10A breast cell acinus formation in 3D culture conditions,” *J. Tissue Eng. Regen. Med.*, vol. 10, pp. 71–80, 2016.
[141] K. Ziółkowska, R. Kwapiszewski, A. Stelmachowska, M. Chudy, A. Dybko, and Z. Brzózka, “Development of a three-dimensional microfluidic system for long-term tumor spheroid culture,” *Sensors Actuators, B Chem.*, vol. 173, pp. 908–913, 2012.
[142] K. S. Klos *et al.*, “Building bridges toward invasion: Tumor promoter treatment induces a novel protein kinase C-dependent phenotype in MCF10A mammary cell acini,” *PLoS One*, vol. 9, no. 3, pp. 1–11, 2014.

[143] J. E. Fata *et al.*, “The MAPKERK-1,2pathway integrates distinct and antagonistic signals from TGFα and FGF7 in morphogenesis of mouse mammary epithelium,” *Dev. Biol.*, vol. 306, no. 1, pp. 193–207, 2007.
[144] T. E. Anwar and C. G. Kleer, “Tissue-based identification of stem cells and epithelial-to-mesenchymal transition in breast cancer,” *Hum. Pathol.*, vol. 44, no. 8, pp. 1457–1464, 2013.
[145] J. Zhang, X. Tian, H. Zhang, Y. Teng, R. Li, and F. Bai, “TGF- b – induced epithelial-to-mesenchymal transition proceeds through stepwise activation of multiple feedback loops,” *Sci. Signal.*, vol. 7, no. 345, pp. 1–12, 2015.
[146] M. Vaapil *et al.*, “Hypoxic Conditions Induce a Cancer-Like Phenotype in Human Breast Epithelial Cells,” *PLoS One*, vol. 7, no. 9, 2012.
[147] W. G. Mueller S, Millonig G, “The GOX/CAT system: a novel enzymatic method to independently control hydrogen peroxide and hypoxia in cell culture.,” *Adv Med Sci*, vol. 54, no. 2, pp. 121–135, 2009.
[148] J. Zhu, G. Xiong, C. Trinkle, and R. Xu, “Integrated extracellular matrix signaling in mammary gland development and breast cancer progression,” *Histol Histopathol*, vol. 29, no. 9, pp. 1083–1092, 2014.
[149] A. Naba, S. Hoersch, and R. O. Hynes, “Towards definition of an ECM parts list: An advance on GO categories,” *Matrix Biol.*, vol. 31, no. 7–8, pp. 371–372, 2012.
[150] A. Rizwan *et al.*, “Metastatic breast cancer cells in lymph nodes increase nodal collagen density,” *Sci. Rep.*, vol. 5, pp. 1–6, 2015.
[151] T. Sato *et al.*, “Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett’s epithelium,” *Gastroenterology*, vol. 141, no. 5, pp. 1762–1772, 2011.
[152] M. Van De Wetering *et al.*, “Prospective Derivation of a Living Organoid Biobank of Colorectal Cancer Patients Resource Prospective Derivation of a Living Organoid Biobank of Colorectal Cancer Patients,” *Cell*, vol. 161, no. 4, pp. 933–945, 2015.
[153] W. Chung *et al.*, “Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer,” *Nat. Commun.*, vol. 8, no. May, 2017.
[154] J. & P. D. R. Emerman, “Maintenance and induction of morphological differentiation in dissociated mammary epithelium on floating collagen membranes.,” *In Vitro*, vol. 13, no. 5, pp. 316–328, 1977.
[155] J. P. Celli *et al.*, “An imaging-based platform for high-content, quantitative evaluation of therapeutic response in 3D tumour models,” *Sci. Rep.*, vol. 4, pp. 1–10, 2014.
[156] M. Zanoni *et al.*, “3D tumor spheroid models for in vitro therapeutic screening: A systematic approach to enhance the biological relevance of data obtained,” *Sci. Rep.*, vol. 6, no. August 2015, pp. 1–11, 2016.
[157] N. Sachs *et al.*, “A Living Biobank of Breast Cancer Organoids Captures Disease Heterogeneity,” *Cell*, vol. 172, no. 1–2, pp. 373-386.e10, 2018.
[158] F. Weeber, S. N. Ooft, K. K. Dijkstra, and E. E. Voest, “Tumor Organoids as a Pre-clinical Cancer Model for Drug Discovery,” *Cell Chem. Biol.*, vol. 24, no. 9, pp. 1092–1100, 2017.