Molecular determinants of mammary differentiation and breast cancer progression

by
Dexter X. Jin

B.S., Biology (2010)
University of Wisconsin – Madison, WI

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2018

Signature redacted

Signature of Author..... .....
Dexter X. Jin
Department of Biology
February 6, 2018

Signature redacted

Certified By..... .....
Piyush B. Gupta
Professor of Biology
Thesis Supervisor

Signature redacted

Accepted by ....
Stephen P. Bell
Professor of Biology
Co-Chair, Biology Graduate Committee

Molecular determinants of mammary differentiation and breast cancer progression

by

Dexter X. Jin

Submitted to the Department of Biology on February 6, 2018

in partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Abstract

The mammary epithelium is an architecturally complex tissue comprising of multiple cell lineages. Development and maintenance of this tissue are carefully orchestrated by balancing stem and progenitor cell self-renewal and differentiation. The mammary epithelium must also endure the successive regenerative cycles of pregnancy and lactation. Therefore, it is not surprising that the fidelity of these processes is of the utmost importance to ensure proper homeostasis of this tissue. In fact, dysregulation of these processes frequently results in progression toward cancer, and later, potentially metastatic disease. The clinical relevance of metastasis is hard to overstate, as it is responsible for over 90% of cancer-related deaths. In this thesis, I have identified a number of determinants involved in breast cancer progression and mammary differentiation. First, I describe SMARCE, a SWI-SNF component, as a prognostic factor of carcinoma progression. We show that SMARCE1 cooperates with ILF3 to regulate a basement membrane module and that it is functionally required to degrade basement membrane. Afterwards, I describe CREB3L1 as a key mediator of PERK-driven metastasis. We also showed that the unique mode of action of CREB3L1 provides a therapeutic opportunity to drug invasive breast cancers. Finally, I describe a 3D differentiation screen which identified the collagen receptor tyrosine kinase, DDR1, as a regulator of mammary stem cell differentiation. Mechanistically, we coupled *ex vivo* functional assays with single cell transcriptomic sequencing to show that DDR1 is required for basal fate commitment to activate JAG1 expression, which indirectly stimulates luminal NOTCH1 signaling to drive lobulogenesis. Collectively, these data provide insight into key molecular regulators of breast cancer progression and mammary differentiation.

Thesis Supervisor: Piyush B. Gupta

Title: Professor of Biology

I dedicate this thesis to

my wife, Carolyn Marie Jin,

and my daughter, Evelyn Song Jin

for their love, patience, and inspiration.

Acknowledgements

This thesis is a showcase of not only my own effort, but also the efforts of many other individuals whom I was fortunate enough to cross paths with. I would like to start by expressing my gratitude toward my thesis advisor, Piyush Gupta. In addition to stewarding a warm and scholastic environment, Piyush frequently encouraged me to explore exciting projects, regardless of the risk. Never fearful of failure, he always focused on the positives – a trait that I hope to always keep with me. I would also like acknowledging the entire Gupta lab, past and present, for creating an environment that encouraged exciting discussions that kept my enthusiasm at my all-time high, for fostering a strong camaraderie that I continue to return to, and for scientific guidance. In particular, I would like to mention Ethan Sokol, Daniel Miller, and Yuxiong Feng. Ethan introduced me into the Gupta lab and closely mentored me. But more than that, Ethan's passion toward his work, his dedication to details, and his thoughtfulness have truly inspired me. I also want to thank Daniel Miller. His insightful advice and his ability to contextualize have taught me and helped me to grow scientifically and personally. And I would like to acknowledge Yuxiong Feng, who will soon be opening his own lab. My time with Yuxiong has shown me that the people who join his lab will be very fortunate to have a leader such as him at the helm.

I would like to acknowledge my committee members, Richard Hynes and Jacqueline Lees, for their scientific input and their career advice over the years, as well as Christopher Love for serving as my external faculty advisor. I would also like to thank the facilities at the Whitehead and MIT that make this work possible, particularly Wendy Salmon of the Keck Imaging Facility.

I would like to express my gratitude my advisor during my time at the University of Wisconsin – Madison, Karen Downs. When I first entered college, I was rather lost in my search for a career. But my time in her lab inspired me to take the next step to MIT. I would also like to thank my former lab mates from the Downs lab. In particular, I would like to mention Maria Mikedis and Jake Daane for their wonderful advice and friendship over the years.

I am also extremely grateful to my family and friends. My parents, Kim and Eva Jin, for always pushing me to be my very best. In doing so, they shaped me into the person I am today. Steve and Linda Lucas, for their encouragement and advice. They always helped me to see the grand scheme of things.

Most importantly, I would like to acknowledge my wife, Carolyn, for her guidance, encouragement, and wisdom. It is hard for me to overstate how critical she was to my successes. Throughout our time together, Carolyn has been everything to me. As my confidante, she supported me through all manner of challenges. As my mentor, she taught me how to balance all the aspects of my life. As my partner, she celebrated my successes. For everything she was and is, I will always cherish her.

Last but not least, I would like to thank my daughter, Evelyn, for her company through the many late nights.

Table of Contents

Abstract 2
Dedication 3
Acknowledgements 4
Chapter 1: Introduction 6
1.1 : Introduction 7
1.2 : References 21
1.3 : Figures 28
Chapter 2: SMARCE1 is required for the invasive progression of in situ cancers 30
2.1 : Abstract 31
2.2 : Introduction 32
2.3 : Results 33
2.4 : Discussion 38
2.5 : Methods 40
2.6 : Acknowledgements 41
2.7 : References 42
2.8 : Figures and Tables 45
2.9 : Supplemental Data 51
Chapter 3: Cancer-specific PERK signaling drives invasion and metastasis through CREB3L1 67
3.1 : Abstract 68
3.2 : Introduction 69
3.3 : Results 70
3.4 : Discussion 75
3.5 : Methods 77
3.6 : Acknowledgements 83
3.7 : References 84
3.8 : Figures and Tables 88
3.9 : Supplemental Data 95
Chapter 4: Cell-matrix interaction drives multi-lineage differentiation of mammary stem cells through the discoidin domain receptor tyrosine kinase, DDR1 121
4.1 : Abstract 122
4.2 : Introduction 123
4.3 : Results and Discussion 124
4.4 : Methods 129
4.5 : Acknowledgements 133
4.6 : References 134
4.7 : Figures and Tables 136
4.8 : Supplemental Data 144
Chapter 5: Conclusions 147
5.1 : Summary 148
5.2 : Discussions and Future Directions 149
5.3 : Concluding Remarks 152
5.4 : References 153

Chapter 1

Introduction

Chapter 1.1: Introduction

Introduction

Breast cancer is one of the most prevalent and deadliest forms of cancer in women (1). The most common cause of breast cancer death is metastasis, which is responsible for the vast majority of solid cancer mortalities in general (2). Therefore, it is critical to understand breast cancer and its progression towards metastatic disease. However, breast cancer is also a very complex disease with numerous subtypes resulting from aberrant differentiation (3, 4). Despite dedifferentiating from normal cells within the mammary gland, parallels have often been drawn between the subtypes of breast cancer and the cell types within the mammary gland, with some cells of origin even having been identified (5, 6). In drawing these parallels, it has been noted that some of the most recurrently mutated genes in breast cancer play vital roles in mammary differentiation (7-9). Moreover, these genes have provided mechanistic insights into how cancers arise and progress (10, 11). Thus, understanding differentiation in the mammary gland has proven to be useful to understanding breast cancer development and progression.

Mammary gland and its development

The mammary gland is a distinctive organ which separates mammals from other animals (12). This exocrine gland is specialized for milk production following childbirth (12). The mammary gland exists among a variety of species, ranging from the egg-laying platypus to placental mammal primates. Within this variety of species is mice, which is fortuitously a great model system to understand development. This is owing to the fact that the mice are amenable to sophisticated *in vivo* manipulations.

The mammary gland can broadly be subdivided between the epithelial compartment and the stromal compartment (12). The stroma of the mammary gland is comprised of several distinctly functioning components including fibroblasts, endothelial cells, adipocytes, and immune cells (12). The mammary stroma also includes several non-cellular components which make up the extracellular matrix (ECM), such as collagens (12).

Development of the mammary gland can be generally subdivided into embryonic development and postnatal development (13). Embryonic mammary development initiates during the first trimester (in the mouse this happens at approximately embryonic day 10.5) from the ectoderm, one of the three primary cell types that form during embryogenesis (12, 13). In humans, this layer of cells migrates to the mammary bud sites to give rise to distinct multilayered ectodermal structures called mammary placodes (13, 14). This site is surrounded by a layer of mesenchymal cells, which arose from mesoderm – another one of the three primary cell types that form during embryogenesis (13). The mesenchymal component gives rise to multiple cell types of the mammary gland, including the adipose cells and the fibroblast cells (12). The mesenchyme and epithelium of the gland interact with each other to promote the development of the gland. Then the

epithelial cells, which developed from the prior ectodermal cells, proliferate to form multiple rudimentary tree-like structures that invade into the mesenchymally-derived components in a process called branching morphogenesis (12, 13). Meanwhile, the epidermal cells, which also developed from ectodermal cells, that are adjacent to an end of the tree-like epithelial structure gives rise to the nipple. Eventually, the aforementioned trees connect at the nipple (15).

Afterwards, the mammary gland allometrically grows until puberty, which initiates post-natal development. This aspect of organ development that is unique to the mammary gland (15). During puberty, branching morphogenesis reinitiates and the ducts of the mammary gland rapidly proliferate and invade through the mammary stroma. However, this time the ducts of the gland also fill in the gland by sprouting secondary ductal branches which are lateral to the primary ducts (13). During this stage, composite structures, known as terminal ductal lobular units (TDLUs) also arise (13). These composite structures are comprised of two major anatomical mammary units – ducts and lobules (Figure 1A). Adolescent branching morphogenesis requires hormone signals and growth factors (13, 16). For instance, one of the most well-known hormones, estrogen, has been shown to contribute heavily to ductal growth since studies of mice without the estrogen receptor noted deficiencies in ductal outgrowth and branching morphogenesis (16).

The next stage of mammary development is pregnancy. During this period, the mammary gland undergoes another drastic expansion. First, the gland undergoes another round of increased branching, which results in tertiary ductal branches, to allow for more lobules to develop throughout the gland (13). By the end of pregnancy, the lobular component represents the majority of the mammary gland and these lobules have gained the capacity to secrete (13). During this stage, hormones such as estrogen were shown to play a key role again. Mice that had the estrogen receptor conditionally knocked out after ductal elongation lacked lobular development (13). Unlike during pubertal development, progesterone, another hormone, was also found to be important for development during pregnancy (13, 17). Studies of mice that lacked the progesterone receptor reported deficiencies in tertiary branching of ducts and in lobular development (13, 17). Following pregnancy, the mammary gland continues to expand and secretes milk in a process called lactation. Additionally, a third major hormone, oxytocin, is released during nurturing to activate contraction to push out milk (13).

After pregnancy and lactation, the mammary gland retracts to a virgin-like state in a process termed involution. Here, the milk-producing cells are apoptotically lost and the ductal architecture is pruned to reduce branching (18). Moreover, proteases break down the ECM of the mammary gland, which in turn promotes additional apoptosis throughout the gland resulting in further retraction of the mammary tree (18). Interestingly, while the gland appears morphologically similar to the virgin-like gland, studies that performed gene expression profiling of the mammary gland pre- and post-pregnancy have found that the post-involution gland is actually quite distinct from the virgin gland (18, 19).

Much of what is understood about the development of the mammary gland was revealed by mouse models. However, a number of differences hinder direct translation of studies from mouse mammary gland development to human mammary gland development. For instance, during gestation, the human gland forms several trees which coalesce at the nipple, whereas the mouse gland forms a single tree (15). Additionally, during fetal and adolescent development, the mouse mammary gland has a terminal end bud (TEB) that is not present in the human mammary gland (15). Moreover, following puberty, the human gland has lobules as a part of the TDLUs, whereas the mouse mammary gland, which has lost its TEB, is primarily a ductal tree. Only after the first round of pregnancy will the mouse mammary gland have lobules that have evaded involution. Furthermore, mouse and human mammary glands also differ in that the human lobules tend to be surrounded by fibroblast- and ECM-rich stroma (18). In contrast, the stroma of the mouse mammary gland tends to be much more adipocyte-rich (18).

The mammary epithelium

Epithelial cells are defined, in part, by their ability to form tightly packed sheets of cells. To do so, epithelial cells use specialized cell adhesion molecules such the epithelial cadherin (E-cadherin; CDH1), which function to bind cells to each other (20, 21). These sheets of cells can be organized in a variety of shapes such as simple single-layered epithelia or as complex as multilayered epithelia (21). Epithelial cells can also function in a diverse and distinct manner of ways ranging from secreting to absorbing to sensing (21).

In the mammary gland, the epithelium is a remarkably complex and dynamic tree-like structure (5, 12). In adults, the branched ductal component of the epithelia networks the nipple to the TDLUs. This network of ducts acts as a route for milk to travel; starting from production centers found in the lobular component of the TDLUs.

Cross-sectional inspection of the mammary epithelium reveals that it is organized as an intricate bilayered structure with two major epithelial cell lineages. In the inner, lumen-facing portion of the epithelium are the luminal cells (Figure 1B). These cells form a circular perimeter around the lumen. Luminal cells characteristically express cytokeratins such as KRT8, KRT18, and KRT19 (5, 22). From a cell surface marker perspective, luminal also express high amounts of another epithelial cell adhesion marker (EpCAM) and a variable amount of CD49f, otherwise known as ITGA6 (5). Immediately adjacent to these luminal cells are the basal cells, which sequester the luminal cells from the basement membrane (Figure 1B) (5, 12). Basal cells express high amounts of cytokeratins such as KRT5 and KRT14 (5). These cells typically also have high expression of CD49f, but low expression of EpCAM (5). This distinction is in cell surface markers (EpCAM and CD49f) is often used to isolate each of the two epithelial compartments

in humans. The mouse ductal organization is similarly subdivided among the two cell lineages, except at the TEBs which contain multilayer outgrowths (12).

The mammary epithelial hierarchy

In a proposed mammary epithelial hierarchy, putative MaSCs would give rise to distinct unipotent progenitors, which would then contribute to the luminal and basal compartments (Figure 2). Lineage tracing studies, which function by following the individually labeled cells and their matching labeled progeny, had elucidated the existence of lineage-committed mammary cells. In 2011, Van Keymeulen et al had identified long-lived lineage-committed clones that contribute during various stages of mammary development (23). Based on the trait that these cells survive multiple rounds of pregnancy, lactation, and involution, they suggest that these long-lived unipotent clones drive adult development and homeostasis. However, these labeled luminal cells were not sufficient to recreate a mammary epithelium without a basal counterpart (23). In contrast, a small subset of the basal compartment was shown to be sufficient to repopulating the entire mammary epithelium with high efficiency (24).

In the luminal compartment of the hierarchy, there are multiple subtypes of luminal cells. First, a putative common luminal progenitor is thought to give rise to the ductal luminal lineage and the alveolar luminal lineage through ductal luminal progenitors and alveolar luminal progenitors, respectively. The ductal luminal cells are found in the inner layer of the ducts, while the alveolar luminal cells are found in the inner layer of the lobules. In mouse models, CD61, otherwise known as ITGB3, was shown to mark luminal progenitors (25). Later studies revealed that ductal committed luminal progenitors were a distinct subset of luminal progenitors and they are typically positive for CD133 (PROM1) but have lost CD61. Consistent with its role as a ductal progenitor, Anderson et al reported that branching morphogenesis was severely dysregulated when analyzing mice that lacked CD133 (26). Moreover, Sleeman et al indicated that CD133 specifically marks the estrogen-responsive cells in the mammary gland (27). Most recently, Wang et al confirmed that CD133 marks all the estrogen-responsive cells (28). They also reported that these cells robustly contributed to the ductal tree (28). Interestingly, lineage tracing also showed that estrogen receptor negative cells could also contribute to ducts (28). Taken together, these studies highlighted the existence of a ductal luminal progenitor and suggests that multiple ductal luminal progenitors may even exist. Meanwhile, alveolar luminal progenitors were also found to be a unique subset of luminal progenitors. In 2002, Wagner et al had tracked cells that expressed the whey acidic protein (WAP) and found that these cells could contribute to the secretory cells of the luminal layer, but not to the hormone-responsive cells (29). During pregnancy, these alveolar luminal progenitors later go on to give rise to functional mature alveolar cells (5). These mature alveolar cells are responsible for secreting the components that make up milk (30). They are localized in the lobules.

In the basal portion of the hierarchy, the distinction between various basal cell types has been challenging to make. The current model posits a putative myoepithelial progenitor which gives rise to terminally differentiated myoepithelial cells, which provide the muscular force necessary to guide milk throughout the gland. However, the basal compartment is much less stratified than the luminal compartments. Even in recent single-cell transcriptomic sequencing studies, distinctions between basal cells has been challenging to demarcate (31).

Given the existence of a basal and a luminal lineage, many had hypothesized the existence of a bipotent stem cell that is capable of giving rise to cell types in both lineages. In 1998, Stingl et al had reported the isolation of a putative human multipotent stem cell based on two metrics: the ability of a cell to form a single colony that spans both lineages when seeded onto two-dimensional surfaces and the ability of a cell to form a structure that also spanned both lineages when embedded into a three-dimensional (3D) matrix. By 2006, Shackleton et al. and Stingl et al. showed that a single mammary stem cell could generate a functional mammary gland (32, 33). More recently, by using genetic mouse models, which combine 3D imaging with a multi-color based fate mapping system, Rios et al followed the progeny of individually labeled cells (34). In this mouse model, a unique combination of fluorescent proteins is expressed in KRT5- or KRT14-positive cells. This specific fluorescent combination is passed to daughter cells allowing for the progeny to be distinguishable from the progeny of other labeled cells. Each uniquely labeled set of cells is called a clone. Rios et al found that a subset of these uniquely labeled clones was capable of contributing to both basal and luminal lineages *in vivo* (34). These clones could also contribute to morphogenesis during pregnancy, lactation, and could even be followed after involution (34). Taken together, these data indicated the existence of bipotent MaSCs *in vivo*. Following this observation, Wang et al identified a subset of basal cells which express Procr (24). Lineage tracing of Procr-positive epithelial cells revealed that they contributed to all lineages of the mammary epithelium, indicating that Procr marks a MaSC subset of basal cells. With the identification of the bipotent stem cell, the apex of the potential hierarchy was unveiled.

Mammary epithelial differentiation

Alongside the understanding of the numerous cell types along the mammary hierarchy, much effort has gone into understanding the regulators of mammary differentiation. This body of work has revealed several functional and mechanistic insights into how mammary epithelial cells differentiate and the cell types they give rise to.

For instance, SOX9 and SLUG were found to be key determinants of the mammary stem cell state (35). Functionally, the combination of both SOX9 and SLUG were sufficient to transition cells to a mammary stem cell state which was capable of repopulating a mammary gland (35). These transcription factors were shown to activate autoregulatory networks to maintain a stem cell state (35). Consistent with

these ideas, recent lineage tracing studies by Malhotra et al, and more recently by Wang et al, found that a subset of SOX9 positive clones contributed to multiple lineages, indicating that they were bipotent stem cells (28, 36). Guo et al had also noted that expression of SOX9 in differentiated luminal cells also conferred differentiated luminal cells with luminal progenitor activity, suggesting an independent role for SOX9 in luminal differentiation (35). In keeping with these data, Malhotra et al and Wang et al also found that SOX9 positive clones often specifically contributed significantly to luminal differentiation (28, 36). Moreover, Malhotra et al showed that conditional loss of SOX9 after puberty results in a loss of the luminal compartment (36).

GATA3, which is expressed in the luminal layer of the mammary epithelium, was found to be involved in differentiation and maintenance of luminal cells. Asselin-Labat et al showed that overexpression of GATA3 in a MaSC-enriched population promotes luminal fate commitment (25). A prior study by Kouros-Mehr et al showed that loss of GATA3 in mice led to a disruption in ductal and lobular formation (11). Furthermore, Kouros-Mehr had found that short-term, conditional loss of GATA3 in adult mammary glands pushed luminal cells toward a proliferative and less differentiated stated (11). Corroborating these data, Asselin-Labat et al found that GATA3 loss led to an expansion of luminal progenitors (25). Mechanistically, GATA3 has been shown to promote ductal differentiation by activating expression of FOXA1 (11).

Bernardo et al had reported FOXA1 to be a determinant of ductal development but not lobular development (37). By using a FOXA1 knock-out mouse model, they showed that ductal expansion was defective, while milk-producing alveoli had formed normally during pregnancy (37). FOXA1 was also found to be necessary for hormone-responsive cells which are found in the ducts. They showed that this was, in part, due to transcriptional regulation of ESR1 expression by FOXA1 (37). This potentially establishes a feedback loop to stabilize the differentiation state, since ESR1 has been shown to regulate expression of GATA3 (38).

ELF5 was shown to contribute to alveologenesis during pregnancy. Mice that lacked ELF5 have also been found to be deficient in their ability to lactate (39-41). Loss of ELF5 also resulted in an increase in the progenitor population of the luminal lineage, suggesting it may be particularly important for luminal maturation. ELF5 was shown to regulate STAT5A, and reciprocally STAT5A was also shown to regulate ELF5 (42, 43). Interestingly, GATA3 has also been indirectly implicated in regulating STAT5A (44). Taken together, this implicates a second potential loop downstream of GATA3, which regulates lobular development independent of regulating ductal development.

In addition to the aforementioned phenotypes, loss of ELF5 was also shown to hyperactivate Notch signaling (39). While it has not yet been shown in mammary development, Notch signaling has been found to upregulate ELF5 expression in other systems for development (45). In mouse models,

NOTCH1 was also shown to be important for luminal differentiation down the alveolar lineage (46). NOTCH1 is turned off by the time cells have terminally differentiated (46). Furthermore, Notch signaling was shown to play an important role in luminal lineage determination in an independent manner. Notch signaling was reported to suppress p63 to promote luminal fate commitment (47). Conversely, p63 has also been shown to antagonize Notch signaling (47).

RUNX1 was shown to be critical for exit from the bipotent state. By screening in multipotent cells, loss of RUNX1 was found to inhibit terminal differentiation (48). Moreover, reactivation of RUNX1 allowed cells to differentiate as measured by structure maturation in 3D and by following lineage commitment in colony assays (48). RUNX1 was also found to be expressed in all epithelial cells, except alveolar cells, suggesting that it represses alveolar differentiation (49). Consistent with this idea, RUNX1 was found to repress ELF5 (49).

During mammary development, epithelial cells must often invade through the stroma to expand the gland. However, epithelial cells are not well equipped to push through the gland due to their propensity to stay tightly packed. However, mesenchymal cells, unlike their epithelial counterpart, are loosely organized and motile cells (50). The epithelial-to-mesenchymal transition (EMT) is a process which confers epithelial cells with mesenchymal properties. These cells, and by extension the process that creates these cells, have been well characterized in the context of developmental processes and during wound healing. Important transcription factors implicated in the EMT program include SNAI1, SNAI2 (otherwise known as SLUG) and TWIST1 (51, 52). These transcription factors act as key repressors of E-cadherin, which is important for stabilizing the epithelial cell state. While activation of these transcription factors promotes EMT, loss of E-cadherin has also been shown to be sufficient to induce an EMT (53). EMT has been shown to play an important role during branching morphogenesis, whereby cells in the mouse TEB undergo an EMT to gain the capacity to invade through the stroma (12, 54).

Another process important for the development of the mammary gland revolves around the endoplasmic reticulum (ER), which is a major organelle responsible for protein folding and maturation of a significant fraction of the proteome (55). Since homeostasis of this organelle is critical to the survival of the cell, the ER is protected by three major pathways which make up the unfolded protein response (UPR). The UPR has often been linked to specialized secretory cell types such as the pancreatic ß cells (56). In these cells, an immense load is placed on the ER to create the products important for the function of these specialized cell types (56). However, the UPR also plays a key role during mammary morphogenesis. Hasegawa et al had found that loss of XBP1, a key mediator of one of the three major pathways of the UPR, resulted in deficient branching morphogenesis in mice (57). These mice also had trouble lactating (57). Independently, Zhu et al reported that mammary epithelial cells lacking another UPR regulator, BiP, were incapable of repopulating mammary glands (58).

The cellular microenvironment and its role in mammary development

The stromal cells of the mammary gland are also quite important for the development of the mammary gland. In fact, it has been reported that stromal cells of the mammary gland can reprogram other cell types to adopt the fate of the mammary epithelia. For instance, Sakakura et al had shown that the mammary epithelium could adopt salivary gland fate when mixed with the salivary mesenchyme (59). Conversely, Cunha et al demonstrated that the salivary epithelium could adopt a mammary epithelial fate when mixed with mammary mesenchyme (60). Furthermore, Gilbert Smith and colleagues showed that many cell types could be inoculated with dispersed normal mouse mammary epithelial cells into cleared mammary fat-pads and incorporate into mammary outgrowths (61-63). Taken together, these results highlight the importance of the stroma in defining the mammary epithelium. Several of the major stromal cell types have been implicated in additional specific roles in mammary development (12).

For instance, the majority of the mammary gland by area is filled with adipocytes. Adipocyte precursors appear as early as during embryonic development of the mammary gland. When developed, adipocytes are lipid-filled cells. In the mammary gland, adipocytes have been reported to regulate angiogenesis by secreting VEGF, a key angiogenic factor (64). It was also noted that adipocytes may regulate milk production with lipid reservoirs. This hypothesis arose since it was noted that during pregnancy and lactation, lipid content was reduced in adipocytes (12, 65, 66).

Fibroblasts are another cell type that appears as early as during embryonic development of the mammary gland. Fibroblasts were found to support mammary epithelial survival and morphogenesis. A major function of the fibroblasts is to provide the components that make up the ECM (12). They also secrete growth factors and cytokines which regulate cellular function in the gland. Consistent with these characteristics, fibroblasts have been implicated in branching morphogenesis by providing growth factors and by breaking down the environment with proteases (67).

Outside of the cellular component, the stroma also plays critical roles in mammary gland development by way of its non-cellular component, the ECM. The ECM of the mammary gland surrounds the epithelium. This positioning allows the ECM to provide signals in the form of direct interactions and growth factors which impact numerous stages of mammary development (68). While the most well-known secreted regulators of mammary gland development are steroid hormones such as estrogen and progesterone, there are many other secreted regulators of the mammary gland. Many ECM components contain domains that can attract and bind ligands such as growth factors which, upon contact with the local epithelia, bind to the receptors found on the epithelial cells. This ligand-receptor interaction then stimulates signaling pathways in the epithelial cells to promote a variety of activities. For instance, in mice, growth factors, such as those in the FGF family, drive ductal expansion (69). These FGFs are

frequently deposited by the stromal cells into the ECM. FGFs are embedded into the ECM by binding Perlecan, a heparin chain containing component of the ECM. Upon binding to their receptors, FGFs were reported to activate mammary epithelial proliferation (68). Consistent with this, it was also found that overexpression of heparanases, which free FGFs from Perlecan, in the mammary gland results in excessive branching, which mimics excessive FGF signaling (68, 70, 71).

Outside of growth factor regulation, the composition of the ECM was also shown to be important for mammary development in many ways, including by establishing epithelial cell polarity. Through defining cell polarity, ECM composition was shown to be important to establishing the spatial orientation of luminal and myoepithelial cells. It was found that it was necessary for ECM to have Laminin-111 to establish the proper luminal/myoepithelial spatial orientation (68). Cell polarity has also been shown to be important for end bud formation and ductal development (72). ECM density has also been shown to be important for directing ductal morphogenesis (68).

Another hint that the ECM composition is important to the development of the mammary gland comes in the form of studies on remodeling. In particular, proteases such as MMP-3 have been shown to be important for secondary and tertiary ductal branching (73). MMP-2 was reported to be important for TEBs to penetrate the stroma. MMP-2 has also been shown to be important for epithelial cell survival (73).

In addition to the ECM, cellular receptors of the ECM have been implicated in mammary development. One of the most famous sets of ECM sensors, the integrin family has been implicated in roles during a variety of stages of mammary development. Integrins are a family of transmembrane receptors which bind the cells to the ECM. Upon activation, integrins can transduce signals to activate many cellular functions such as proliferation, motility, and differentiation. While aforementioned discussions of integrins indicated that are used as markers to distinguish lineages in the mammary gland, studies have shown that integrins also play functional roles in regulating mammary development (74). For instance, ß1-integrin was implicated in ductal branching, since blocking engagement with ß1-integrins resulted in a deficiency in ductal branching (75). Similarly, α2-integrin loss also resulted in deficient ductal branching, albeit with a milder phenotype (76). Furthermore, genetic ablation of ß1-integrins in the basal lineage of mouse mammary glands was shown to perturb stem cell self-renewal (77). Integrin-mediated complexes are also thought to act as mechanical sensors of the physical properties of the ECM. Fitting this idea, filamin, an integrin-associated scaffold protein, has been found to transduce signals defined by the mechanical cues of the ECM to control morphogenesis (78).

The Discoidin Domain Receptor (DDR) family contains another group of transmembrane receptors that are activated by collagens. These DDRs are receptor tyrosine kinases (RTKs). Characterization of DDRs revealed that DDR1 recognizes all collagens, while DDR2 specifically

recognizes fibrillary collagens. Studies during mouse mammary development by Vogel et al found that loss of DDR1 in mice disrupts ductal morphogenesis and lactation (79). DDR1 was also shown to maintain the epithelial differentiation state of cells by stabilizing E-cadherin (80). Interestingly, induction of EMT in systems has been shown to correlate with a switch from DDR1-expression to DDR2-expression (53).

Breast cancer progression

The vast majority of breast cancers are carcinomas, a type of cancer that initiates from epithelial cells. Broadly speaking, breast carcinomas can be histopathologically subdivided into *in situ* and invasive variants of lobular and ductal carcinomas, with ductal carcinomas *in situ* being the most prevalent form of breast cancer (3). While the exact initiating steps of breast cancer are poorly understood, progression has been histologically characterized. First, cells become clonally hyper-proliferative and begin to exhibit cytological abnormalities to form atypical ductal or lobular hyperplasias, two types of precancerous lesions. After measuring beyond a size threshold, fulfilling cytological requisites, or being found in more than a single duct, atypical ductal hyperplasias are classified as ductal carcinomas *in situ* (DCIS) (81). Similarly, once atypical lobular hyperplasias have expanded beyond individual lobes and have fulfilled cytological criteria, they are considered to be lobular carcinomas *in situ (82)*. As carcinomas *in situ* grow, they distend the mammary gland but they continue to remain encapsulated in the site or *in situ*. Molecular, epidemiological, and pathological studies have established that *in situ* carcinomas are, indeed, predecessors to invasive carcinomas (83, 84).

Progression of breast cancer can be classified by clinical stage, which ranges from 0 to 4 (85). Pathologic staging of breast cancers focuses on three different characteristics: Tumor size (T), the amount of carcinoma spread to lymph nodes (N), and metastatic spread (M). In this TNM system, carcinomas *in situ* are considered stage 0, with no spread to lymph nodes or metastatic dissemination (85). Despite this staging, carcinomas *in situ* are quite heterogeneous, spanning from less aggressive forms that are difficult to distinguish from atypical hyperplasias to more aggressive forms which are molecularly very similar to invasive carcinomas (83, 84).

As carcinomas progress, they can lose their epithelial properties, resulting in dedifferentiation. A distinct but similarly prognostic manner of categorizing breast cancers is focused on the degree of dedifferentiation via histological grade. Dedifferentiation is associated with poor prognosis. To categorize the amount of dedifferentiation in breast cancer, one often uses the Nottingham Grading System, which focuses on three morphological features (86). The first characteristic revolves around the architectural pattern of the tumor. Normally, breast tissues have tubular structures, but as breast cancers become dedifferentiated, they lose this architectural tendency. The second attribute focuses on the size and the

shape of the nucleus. The third attribute focuses on the mitotic rate, which is used to measure the amount of growth occurring in the tumor. Taken together, these features measure the abnormality of a tumor relative to a normal mammary gland. As tumors become increasingly dissimilar from the normal gland, their grade increases (86).

Studies have attempted to identify differential genetic lesions between carcinomas *in situ* and invasive carcinomas. But they revealed that invasive cancers are surprisingly similar to their carcinomas *in situ* counterparts (84). Alongside these clinical stages, breast cancers are also typically immunohistochemically profiled for expression of the estrogen receptor, progesterone receptor, as well as for HER2 overexpression as measured by protein or RNA. However, carcinomas *in situ* and invasive carcinomas have also been shown to be similar in these clinical aspects as well (84). Part of the challenge in identifying differences between *in situ* and invasive carcinomas lies in the fact that carcinomas *in situ* are variable. In fact, high-grade DCIS has been surprisingly associated with the disruption of the myoepithelium and the basement membrane (87).

Complicating these classifications of breast cancer, gene expression profiling studies have reported that breast cancers can also be further distinguished into a series of intrinsic subtypes including luminal A, luminal B, Her2-amplified, claudin-low, and basal-like (7). Based on the gene expression profiles of these subtypes, parallels have been drawn to normal mammary cell types. For instance, the claudin-low subtype of breast cancer has been most closely aligned with the MaSC/basal signature (88). In BRCA1 mutated breast cancer patients, luminal progenitor cells were even found to be the cell of origin for the basal-like subtype of breast cancer (6). The remaining subtypes are thought to fall along the luminal hierarchy, however, the exact origins of the remaining intrinsic subtypes are an active area of scholarship (5). This is particularly important because these cells of origin are thought to influence the metastatic capabilities of carcinoma cells (89).

Dysregulation of developmental pathways in breast cancer

Given these observations, many have pursued the idea that dysregulation of genes involved in differentiation may be important to tumor initiation and progression. Indeed, profiling of breast cancers by several groups including the cancer genome atlas (TCGA) have revealed recurrent mutations of a number of factors involved in differentiation (7-9). In fact, some of the most recurrently mutated genes in breast cancer include the aforementioned CDH1, GATA3, FOXA1, and RUNX1.

As outlined earlier, lobules and ducts in the mammary gland are quite distinct, both morphologically and transcriptomically. As their names would suggest, lobular carcinomas arise from lobules and ductal carcinomas arise from ducts. Unsurprisingly, different genes drive the pathology of these different variants of breast cancer. For instance, RUNX1 and FOXA1 had been indicated to play

more of a role in driving lobular carcinomas. Large-scale mutation profile studies of lobular carcinomas showed enrichment for mutations in RUNX1 and FOXA1 (7). Functionally, Kas et al reported that mutagenic screening in a mouse model for invasive lobular carcinomas identified FOXA1 and RUNX1 as candidate drivers, indicating functional roles for these genes in lobular carcinoma progression (90).

GATA3 expression was found to be prognostic for breast cancer survival, with higher expression correlating with better survival. Moreover, forced expression of GATA3 in breast cancers differentiated mammary tumors and reduced metastatic capacity (11). More recently, it was also shown that GATA3 suppresses metastatic dissemination by upregulating a microRNA which consequently suppresses angiogenesis and ECM remodeling factors that promote invasion and promotes dedifferentiation of carcinomas (10). Taken together, these data indicate a tumor suppressive role for GATA3.

While initially described as a transdifferentiation process to convert epithelial cells to mesenchymal cells, EMT is also often associated with dedifferentiation (91, 92). Even though EMT is essential for proper development, the traits acquired during EMT are dangerous in cancers (50, 91). Activation of EMT confers carcinoma cells with stem cell properties in the form of tumor-initiating capacity (50). EMT also endows cells with the ability to resist apoptosis, senescence, and suppress the immune system (50). Therefore, it is not surprising that one of the most mutated genes in breast cancer is CDH1 (7-9). As mentioned earlier, loss of CDH1 is used as a model to induce an EMT in cells (53). EMT has also been shown to be critically necessary for carcinoma cells to disseminate from the primary tumors to distant sites (91).

A growing body of scholarship has the elucidated the increasing importance of the UPR in carcinomas (93). Within these studies, Feng et al reported that while generally more resistant to therapeutic agents, tumor cells that have undergone an EMT are especially sensitized toward small-molecules that stress the ER (94). This was due to the baseline activation of the UPR, which was later found to provide cells with the capacity to efflux cytotoxic drugs (95). Additionally, PERK and XBP1, two key components of the UPR, were shown to promote metastatic capabilities in mouse models for human breast cancer (94, 96). Interestingly, prolonged activation of the UPR has been implicated with activation of apoptotic pathways, yet these tumor cells seem to evade such a response (97).

Metastasis

Metastasis is a multi-step process which begins with encapsulated primary tumors and ends with disseminated macroscopic lesions at sites away from the initial location (98). During this process, tumors first locally disrupt their environment and obtain access to the vasculature. Subsequently, tumor cells break into the lumen of blood or lymphatic vessels and travel through the vasculature until they arrive at a distant organ site. At this point, it becomes important for these cells to extravasate through the blood

vessels into the distant tissues. To complete the cascade, cells must survive the foreign environment and proliferate to generate macroscopic growths.

During the early stages of cancer, the primary tumor is confined by and to its environment. In particular, the basement membrane (BM) – a subset of ECM – plays a key role in delimiting the epithelial cells from stromal cells (68). As carcinomas progress, they secrete proteases which break down components of the basement membrane. This loss in signals from the basement membrane alters several characteristics in tumor cells, such as cell polarity (99). Tumor cells must also overcome the tumor suppressive effects of myoepithelial cells which can secrete protease inhibitors to maintain the integrity of the basement membrane (100).

After invading through the adjacent stroma, carcinoma cells eventually reach lymphatic sites. During intravasation, invasive carcinoma cells gain access to the lumen of the vessels by breaking through the endothelial architecture. In doing so, tumor cells enter circulation, where they must endure many challenges such as withstanding cell death in the form of anoikis, a programmed cell death that is triggered by the loss of anchorage, or in the form of shear stress due to the high flow rates in the vessels (98). Circulating Tumor Cells (CTCs) are also challenged with evading the immune system (98). However, CTCs now have the potential to access many distant sites by traversing the endothelial lumen. CTCs are often thought to persist in circulation until they are trapped in smaller vessels.

To access the parenchyma of a distant site, tumor cells must escape the vasculature. One manner of exiting the vasculature is through a process called extravasation (98). During this process, cells permeabilize the walls of the vasculature (101). Once they have entered these new distant sites, tumor cells must survive the new environment which are usually hostile to these invasive tumor cells. Most of the time, tumor cells will become dormant or die (102). However, sometimes cells are capable of exiting dormancy and colonizing the metastatic site, which ends the cascade (98, 102).

Unlike with the primary tumor, the metastatic spread is often incurable due to the systemic nature of dissemination. Currently, treatment of metastasis is more limiting than those of primary tumors since metastases are inaccessible for surgical resection. Moreover, metastatic cells are also frequently resistant to therapeutic agents (98, 103). Finally, cytotoxic therapies typically target proliferative cells. This approach precludes treatment of dormant tumor cells, which can later exit dormancy and cause relapse (98).

Breast cancer treatment

Understanding breast cancer and its progression would greatly benefit treatment of breast cancers. Currently, invasive carcinomas are treated in a similar manner to a subset of *in situ* tumors. The treatment of carcinomas *in situ* involves surgically removing tumors. To improve treatment, radiation has been used

in conjunction with surgery. This combination was shown to significantly reduce the rate of local recurrence (84). Afterwards, clinical features such as expression of the estrogen receptor or HER2 overexpression are used to decide on the additional therapies such as estrogen antagonists or HER2 targeting antibodies that will be used. Alongside those treatments, more aggressive carcinomas *in situ* are also treated with chemotherapy. However, chemotherapy has been shown to have marginal benefits, while still being a cytotoxic therapy with several side effects (104).

By understanding breast cancers and their progression, it would be possible to design more rationally targeted therapeutics. In fact, examples of these sorts of success include the aforementioned anti-HER2 treatments for tumors which have been clinically categorized as HER2 high (105). Another example of a targeted therapy is anti-estrogen therapy, which is used to treat tumors which depend on estrogen signaling (105). However, for breast cancer, these remain the major standard lines of targeted treatment.

In this thesis, I sought to improve breast cancer treatment by understanding how breast cancers progress toward invasiveness, identifying novel targeted approaches to treat invasive breast cancers, and understanding mammary differentiation to open up future avenues of study into breast cancer development.

Chapter 1.2: References

1. Bernard W. Stewart PK (2014) World Cancer Report 2014 (World Health Organization).
2. Sporn MB (1996) The war on cancer. *The Lancet* 347(9012):1377-1381.
3. Lakhani SR (2012) WHO Classification of Tumours of the Breast. *Fourth Edition. International Agency for Research on Cancer.*
4. Prat A & Perou CM (2011) Deconstructing the molecular portraits of breast cancer. *Molecular Oncology* 5(1):5-23.
5. Visvader JE & Stingl J (2014) Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes & development* 28(11):1143-1158.
6. Lim E, *et al.* (2009) Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nature medicine* 15(8):907-913.
7. Ciriello G, *et al.* (2015) Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* 163(2):506-519.
8. Lawrence MS, *et al.* (2014) Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505(7484):495-501.
9. Pereira B, *et al.* (2016) The somatic mutation profiles of 2,433 breast cancers refines their genomic and transcriptomic landscapes. *Nature communications* 7:11479.
10. Chou J, *et al.* (2013) GATA3 suppresses metastasis and modulates the tumour microenvironment by regulating microRNA-29b expression. *Nature cell biology* 15(2):201-213.
11. Kouros-Mehr H, Slorach EM, Sternlicht MD, & Werb Z (2006) GATA-3 Maintains the Differentiation of the Luminal Cell Fate in the Mammary Gland. *Cell* 127(5):1041-1055.
12. Inman JL, Robertson C, Mott JD, & Bissell MJ (2015) Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development* 142(6):1028-1042.
13. Sternlicht MD (2006) Key stages in mammary gland development: The cues that regulate ductal branching morphogenesis. *Breast Cancer Research* 8(1):201-201.
14. Howard BA & Gusterson BA (2000) Human Breast Development. *Journal of Mammary Gland Biology and Neoplasia* 5(2):119-137.
15. Sternlicht MD, Kouros-Mehr H, Lu P, & Werb Z (2006) Hormonal and local control of mammary branching morphogenesis. *Differentiation; research in biological diversity* 74(7):365-381.
16. Korach K, *et al.* (1996) *Estrogen receptor gene disruption: Molecular characterization and experimental and clinical phenotypes* pp 159-186; discussion 186.
17. Soyal S, *et al.* (2002) Progesterone receptors - animal models and cell signaling in breast cancer: Progesterone's role in mammary gland development and tumorigenesis as disclosed by experimental mouse genetics. *Breast Cancer Research* 4(5):191-196.

18. Macias H & Hinck L (2012) Mammary Gland Development. *Wiley interdisciplinary reviews. Developmental biology* 1(4):533-557.
19. Balogh GA, *et al.* (2006) Genomic signature induced by pregnancy in the human breast. *International journal of oncology* 28(2):399-410.
20. Hagios C, Lochter A, & Bissell MJ (1998) Tissue architecture: the ultimate regulator of epithelial function? *Philosophical Transactions of the Royal Society B: Biological Sciences* 353(1370):857-870.
21. Lowe JS & Anderson PG (2015) Chapter 3 - Epithelial Cells. *Stevens & Lowe's Human Histology (Fourth Edition) (Fourth Edition)*, (Mosby, Philadelphia), pp 37-54.
22. Bartek J, Bartkova J, & Taylor-Papadimitriou J (1990) Keratin 19 expression in the adult and developing human mammary gland. *The Histochemical journal* 22(10):537-544.
23. Van Keymeulen A, *et al.* (2011) Distinct stem cells contribute to mammary gland development and maintenance. *Nature* 479(7372):189-193.
24. Wang D, *et al.* (2015) Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* 517(7532):81-84.
25. Asselin-Labat ML, *et al.* (2007) Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nature cell biology* 9(2):201-209.
26. Anderson LH, Boulanger CA, Smith GH, Carmeliet P, & Watson CJ (2011) Stem cell marker Prominin-1 regulates branching morphogenesis, but not regenerative capacity, in the mammary gland. *Developmental dynamics : an official publication of the American Association of Anatomists* 240(3):674-681.
27. Sleeman KE, *et al.* (2007) Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland. *The Journal of cell biology* 176(1):19-26.
28. Wang C, Christin JR, Oktay MH, & Guo W (2017) Lineage-Biased Stem Cells Maintain Estrogen-Receptor-Positive and -Negative Mouse Mammary Luminal Lineages. *Cell Reports* 18(12):2825-2835.
29. Wagner K-U, *et al.* (2002) An adjunct mammary epithelial cell population in parous females: its role in functional adaptation and tissue renewal. *Development* 129(6):1377.
30. Capuco AV & Akers RM (2009) The origin and evolution of lactation. *Journal of Biology* 8(4):37-37.
31. Pal B, *et al.* (2017) Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nature communications* 8(1):1627.
32. Shackleton M, *et al.* (2006) Generation of a functional mammary gland from a single stem cell. *Nature* 439(7072):84-88.

33. Stingl J, *et al.* (2006) Purification and unique properties of mammary epithelial stem cells. *Nature* 439(7079):993-997.

34. Rios AC, Fu NY, Lindeman GJ, & Visvader JE (2014) In situ identification of bipotent stem cells in the mammary gland. *Nature* 506(7488):322-327.

35. Guo W, *et al.* (Slug and Sox9 Cooperatively Determine the Mammary Stem Cell State. *Cell* 148(5):1015-1028.

36. Malhotra GK, *et al.* (2014) The role of Sox9 in mouse mammary gland development and maintenance of mammary stem and luminal progenitor cells. *BMC Developmental Biology* 14(1):47.

37. Bernardo GM, *et al.* (2010) FOXA1 is an essential determinant of ERα expression and mammary ductal morphogenesis. *Development (Cambridge, England)* 137(12):2045-2054.

38. Eeckhoute J, *et al.* (2007) Positive Cross-Regulatory Loop Ties GATA-3 to Estrogen Receptor α Expression in Breast Cancer. *Cancer Research* 67(13):6477.

39. Chakrabarti R, *et al.* (2012) Elf5 Regulates Mammary Gland Stem/Progenitor Cell Fate by Influencing Notch Signaling. *Stem cells (Dayton, Ohio)* 30(7):1496-1508.

40. Lee HJ, *et al.* (2013) Progesterone drives mammary secretory differentiation via RankL-mediated induction of Elf5 in luminal progenitor cells. *Development* 140(7):1397.

41. Oakes SR, *et al.* (2008) The Ets transcription factor Elf5 specifies mammary alveolar cell fate. *Genes & development* 22(5):581-586.

42. Choi YS, Chakrabarti R, Escamilla-Hernandez R, & Sinha S (2009) Elf5 conditional knockout mice reveal its role as a master regulator in mammary alveolar development: Failure of Stat5 activation and functional differentiation in the absence of Elf5. *Developmental Biology* 329(2):227-241.

43. Yamaji D, *et al.* (2009) Development of mammary luminal progenitor cells is controlled by the transcription factor STAT5A. *Genes & development* 23(20):2382-2387.

44. Haricharan S & Li Y (2014) STAT signaling in mammary gland differentiation, cell survival and tumorigenesis. *Molecular and cellular endocrinology* 382(1):10.1016/j.mce.2013.1003.1014.

45. Grassmeyer J, *et al.* (2017) Elf5 is a principal cell lineage specific transcription factor in the kidney that contributes to Aqp2 and Avpr2 gene expression. *Developmental Biology* 424(1):77-89.

46. Rodilla V, *et al.* (2015) Luminal progenitors restrict their lineage potential during mammary gland development. *PLoS biology* 13(2):e1002069.

47. Yalcin-Ozuysal O, *et al.* (2010) Antagonistic roles of Notch and p63 in controlling mammary epithelial cell fates. *Cell death and differentiation* 17(10):1600-1612.

48. Sokol ES, *et al.* (2015) Perturbation-expression analysis identifies RUNX1 as a regulator of human mammary stem cell differentiation. *PLoS computational biology* 11(4):e1004161.
49. van Bragt MPA, Hu X, Xie Y, & Li Z (2014) RUNX1, a transcription factor mutated in breast cancer, controls the fate of ER-positive mammary luminal cells. *eLife* 3:e03881.
50. Kalluri R & Weinberg RA (2009) The basics of epithelial-mesenchymal transition. *The Journal of Clinical Investigation* 119(6):1420-1428.
51. Nieto MA (2002) The snail superfamily of zinc-finger transcription factors. *Nature reviews. Molecular cell biology* 3(3):155-166.
52. Qin Q, Xu Y, He T, Qin C, & Xu J (2011) *Normal and disease-related biological functions of Twist1 and underlying molecular mechanisms* pp 90-106.
53. Onder TT, *et al.* (2008) Loss of E-Cadherin Promotes Metastasis via Multiple Downstream Transcriptional Pathways. *Cancer Research* 68(10):3645.
54. Nelson CM, VanDuijn MM, Inman JL, Fletcher DA, & Bissell MJ (2006) Tissue Geometry Determines Sites of Mammary Branching Morphogenesis in Organotypic Cultures. *Science* 314(5797):298.
55. Wang M & Kaufman RJ (2014) The impact of the endoplasmic reticulum protein-folding environment on cancer development. *Nature reviews. Cancer* 14(9):581-597.
56. Eizirik DL & Cnop M (2010) ER Stress in Pancreatic β Cells: The Thin Red Line Between Adaptation and Failure. *Science Signaling* 3(110):pe7.
57. Hasegawa D, *et al.* (2015) Epithelial Xbp1 is required for cellular proliferation and differentiation during mammary gland development. *Mol Cell Biol* 35(9):1543-1556.
58. Zhu G, *et al.* (2014) Differential requirement of GRP94 and GRP78 in mammary gland development. *Scientific Reports* 4:5390.
59. Sakakura T, Nishizuka Y, & Dawe CJ (1976) Mesenchyme-dependent morphogenesis and epithelium-specific cytodifferentiation in mouse mammary gland. *Science* 194(4272):1439.
60. Cunha G, *et al.* (1995) Mammary phenotypic expression induced in epidermal cells by embryonic mammary mesenchyme. *Cells Tissues Organs* 152(3):195-204.
61. Booth BW, *et al.* (2008) The mammary microenvironment alters the differentiation repertoire of neural stem cells. *Proceedings of the National Academy of Sciences* 105(39):14891-14896.
62. Boulanger CA, Mack DL, Booth BW, & Smith GH (2007) Interaction with the mammary microenvironment redirects spermatogenic cell fate in vivo. *Proceedings of the National Academy of Sciences* 104(10):3871-3876.
63. Bruno RD, *et al.* (2017) Mammary extracellular matrix directs differentiation of testicular and embryonic stem cells to form functional mammary glands in vivo. *Sci Rep* 7:40196.

64. Hovey RC, Goldhar AS, Baffi J, & Vonderhaar BK (2001) Transcriptional Regulation of Vascular Endothelial Growth Factor Expression in Epithelial and Stromal Cells during Mouse Mammary Gland Development. *Molecular Endocrinology* 15(5):819-831.

65. Gregor MF, *et al.* (2013) The role of adipocyte XBP1 in metabolic regulation during lactation. *Cell reports* 3(5):1430-1439.

66. Hovey RC & Aimo L (2010) Diverse and Active Roles for Adipocytes During Mammary Gland Growth and Function. *Journal of Mammary Gland Biology and Neoplasia* 15(3):279-290.

67. Howard BA & Lu P (2014) Stromal regulation of embryonic and postnatal mammary epithelial development and differentiation. *Seminars in Cell & Developmental Biology* 25-26:43-51.

68. Muschler J & Streuli CH (2010) Cell–Matrix Interactions in Mammary Gland Development and Breast Cancer. *Cold Spring Harbor Perspectives in Biology* 2(10):a003202.

69. Zhang X, *et al.* (2014) FGF ligands of the postnatal mammary stroma regulate distinct aspects of epithelial morphogenesis. *Development (Cambridge, England)* 141(17):3352-3362.

70. Patel VN, *et al.* (2007) Heparanase cleavage of perlecan heparan sulfate modulates FGF10 activity during ex vivo submandibular gland branching morphogenesis. *Development* 134(23):4177.

71. Zcharia E, *et al.* (2004) Transgenic expression of mammalian heparanase uncovers physiological functions of heparan sulfate in tissue morphogenesis, vascularization, and feeding behavior. *The FASEB Journal* 18(2):252-263.

72. McCaffrey LM & Macara IG (2009) The Par3/aPKC interaction is essential for end bud remodeling and progenitor differentiation during mammary gland morphogenesis. *Genes & development* 23(12):1450-1460.

73. Wiseman BS, *et al.* (2003) Site-specific inductive and inhibitory activities of MMP-2 and MMP-3 orchestrate mammary gland branching morphogenesis. *The Journal of cell biology* 162(6):1123-1133.

74. Raymond K, Faraldo MM, Deugnier M-A, & Glukhova MA (2012) Integrins in mammary development. *Seminars in Cell & Developmental Biology* 23(5):599-605.

75. Klinowska TCM, *et al.* (1999) Laminin and β1 Integrins Are Crucial for Normal Mammary Gland Development in the Mouse. *Developmental Biology* 215(1):13-32.

76. Taddei I, *et al.* (2003) Integrins in Mammary Gland Development and Differentiation of Mammary Epithelium. *Journal of Mammary Gland Biology and Neoplasia* 8(4):383-394.

77. Taddei I, *et al.* (2008) Beta1 integrin deletion from the basal compartment of the mammary epithelium affects stem cells. *Nature cell biology* 10(6):716-722.

78. Schedin P & Keely PJ (2011) Mammary Gland ECM Remodeling, Stiffness, and Mechanosignaling in Normal Development and Tumor Progression. *Cold Spring Harbor Perspectives in Biology* 3(1):a003228.

79. Vogel WF, Aszodi A, Alves F, & Pawson T (2001) Discoidin Domain Receptor 1 Tyrosine Kinase Has an Essential Role in Mammary Gland Development. *Molecular and Cellular Biology* 21(8):2906-2917.

80. Yeh Y-C, Wu C-C, Wang Y-K, & Tang M-J (2011) DDR1 triggers epithelial cell differentiation by promoting cell adhesion through stabilization of E-cadherin. *Molecular Biology of the Cell* 22(7):940-953.

81. Tavassoli FA & Norris HJ (1990) A comparison of the results of long-term follow-up for atypical intraductal hyperplasia and intraductal hyperplasia of the breast. *Cancer* 65(3):518-529.

82. Simpson PT, Gale T, Fulford LG, Reis-Filho JS, & Lakhani SR (2003) The diagnosis and management of pre-invasive breast disease: Pathology of atypical lobular hyperplasia and lobular carcinoma in situ. *Breast Cancer Research : BCR* 5(5):258-262.

83. Reed AEM, Kutasovic JR, Lakhani SR, & Simpson PT (2015) Invasive lobular carcinoma of the breast: morphology, biomarkers and 'omics. *Breast Cancer Research : BCR* 17(1):12.

84. Burstein HJ, Polyak K, Wong JS, Lester SC, & Kaelin CM (2004) Ductal Carcinoma in Situ of the Breast. *New England Journal of Medicine* 350(14):1430-1441.

85. Singletary SE & Connolly JL (2006) Breast Cancer Staging: Working With the Sixth Edition of the AJCC Cancer Staging Manual. *CA: A Cancer Journal for Clinicians* 56(1):37-47.

86. Rakha EA, *et al.* (2008) Prognostic Significance of Nottingham Histologic Grade in Invasive Breast Carcinoma. *Journal of Clinical Oncology* 26(19):3153-3158.

87. Mohan N, Black JO, Schwartz MR, & Zhai QJ (2016) Invasive ductal carcinoma with in situ pattern: how to avoid this diagnostic pitfall? *American Journal of Translational Research* 8(8):3337-3341.

88. Herschkowitz JI, *et al.* (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biology* 8(5):R76-R76.

89. Hanahan D & Weinberg Robert A (2011) Hallmarks of Cancer: The Next Generation. *Cell* 144(5):646-674.

90. Kas SM, *et al.* (2017) Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma. *Nature genetics* 49(8):1219-1230.

91. Chaffer CL, San Juan BP, Lim E, & Weinberg RA (2016) EMT, cell plasticity and metastasis. *Cancer and Metastasis Reviews* 35(4):645-654.

92. Chaffer CL, *et al.* (2013) Poised chromatin at the ZEB1 promoter enables cell plasticity and enhances tumorigenicity. *Cell* 154(1):61-74.

93. Urra H, Dufey E, Avril T, Chevet E, & Hetz C (Endoplasmic Reticulum Stress and the Hallmarks of Cancer. *Trends in Cancer* 2(5):252-262.

94. Feng Y, *et al.* (2014) Epithelial-to-mesenchymal transition activates PERK-eIF2a and sensitizes cells to endoplasmic reticulum stress. *Cancer Discovery.*

95. Del Vecchio CA, *et al.* (2014) De-Differentiation Confers Multidrug Resistance Via Noncanonical PERK-Nrf2 Signaling. *PLoS biology* 12(9):e1001945.

96. Chen X, *et al.* (2014) XBP1 Promotes Triple Negative Breast Cancer By Controlling the HIF1 α Pathway. *Nature* 508(7494):103-107.

97. Fribley A, Zhang K, & Kaufman RJ (2009) Regulation of Apoptosis by the Unfolded Protein Response. *Methods in molecular biology (Clifton, N.J.)* 559:191-204.

98. Lambert AW, Pattabiraman DR, & Weinberg RA (2016) Emerging Biological Principles of Metastasis. *Cell* 168(4):670-691.

99. Manninen A (2015) Epithelial polarity – Generating and integrating signals from the ECM with integrins. *Experimental Cell Research* 334(2):337-349.

100. Pandey PR, Saidou J, & Watabe K (2010) Role of myoepithelial cells in breast tumor progression. *Frontiers in bioscience : a journal and virtual library* 15:226-236.

101. Reymond N, d'Agua BB, & Ridley AJ (2013) Crossing the endothelial barrier during metastasis. *Nature reviews. Cancer* 13(12):858-870.

102. Luzzi KJ, *et al.* (1998) Multistep Nature of Metastatic Inefficiency: Dormancy of Solitary Cells after Successful Extravasation and Limited Survival of Early Micrometastases. *The American Journal of Pathology* 153(3):865-873.

103. Gupta PB, *et al.* (2009) Identification of Selective Inhibitors of Cancer Stem Cells by High-Throughput Screening. *Cell* 138(4):645-659.

104. Morgan G, Ward R, & Barton M (2004) The contribution of cytotoxic chemotherapy to 5-year survival in adult malignancies. *Clinical Oncology* 16(8):549-560.

105. den Hollander P, Savage MI, & Brown PH (2013) Targeted Therapy for Breast Cancer Prevention. *Frontiers in Oncology* 3:250.

Chapter 1.3: Figures

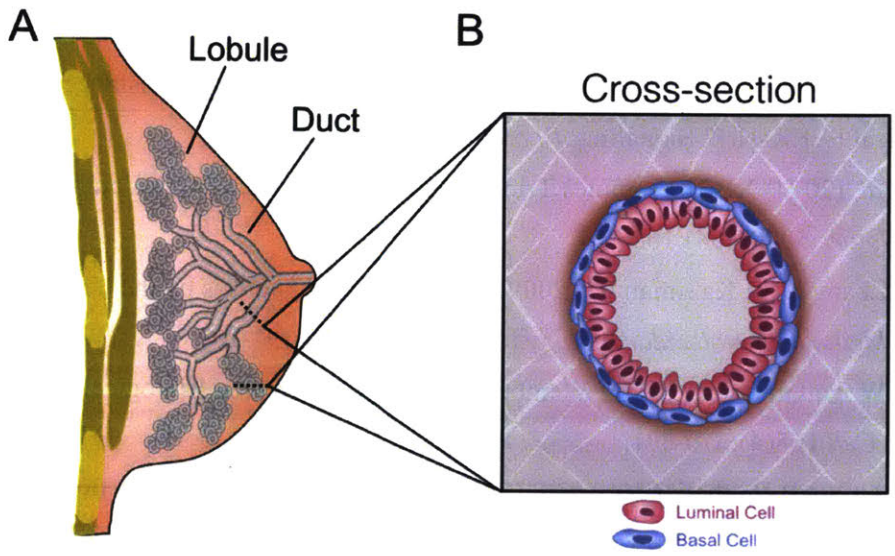


Figure 1. Overview of the mammary gland (A) Schematic of the mammary gland. The mammary epithelium has been colored gray. Black lines indicate sample ducts and lobules. (B) Cross-section of the mammary epithelium reveals a bilayered architecture. The inner layer contains luminal cells and the outer layer contains basal cells.

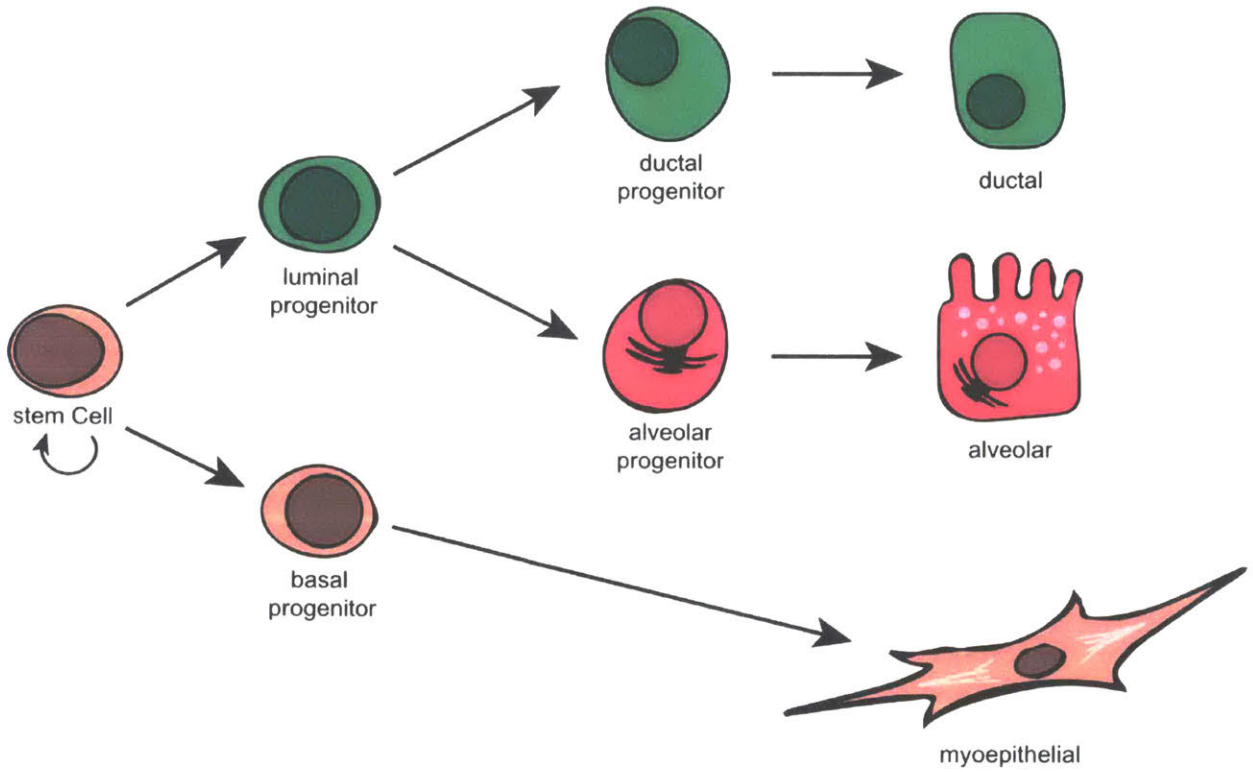


Figure 2. A proposed mammary hierarchy. On the leftmost side are stem cells which give rise to a common luminal progenitor and a basal progenitor. The common luminal progenitor can give rise to ductal and alveolar luminal progenitors, while basal progenitors differentiate into myoepithelial cells. The ductal luminal progenitor differentiates into mature ductal cells and the alveolar luminal progenitor differentiates into mature alveolar cells.

Chapter 2

SMARCE1 is required for the invasive progression of in situ cancers

Published as: Ethan S. Sokol*, Yu-Xiong Feng*, Dexter X. Jin*, Minu D. Tizabi, Daniel H. Miller, Malkiel A. Cohen, Sandhya Sanduja, Ferenc Reinhardt, Jai Pandey, Daphne A. Superville, Rudolf Jaenisch, and Piyush B. Gupta (2017). SMARCE1 is required for the invasive progression of in situ cancers. PNAS.

*These authors contributed equally to this work.

DXJ performed experiments shown in Figures 1A-B, Figures 4B-C, Figure 5B and 5D, Figure 6, Figure S1C-D, Figure S4, Figure S6, Figure S7B-C, Figure S8. All authors contributed to the design of experiments, analysis of data, or editing of the manuscript.

Chapter 2.1: Abstract

Advances in mammography have sparked an exponential increase in the detection of early-stage breast lesions, most commonly ductal carcinoma in situ (DCIS). More than 50% of DCIS lesions are benign and will remain indolent, never progressing to invasive cancers. However, the factors that promote DCIS invasion remain poorly understood. Here, we show that SMARCE1 is required for the invasive progression of DCIS and other early-stage tumors. We show that SMARCE1 drives invasion by regulating the expression of secreted proteases that degrade basement membrane, an ECM barrier surrounding all epithelial tissues. In functional studies, SMARCE1 promotes invasion of in situ cancers growing within primary human mammary tissues and is also required for metastasis in vivo. Mechanistically, SMARCE1 drives invasion by forming a SWI/SNF-independent complex with the transcription factor ILF3. In patients diagnosed with early-stage cancers, SMARCE1 expression is a strong predictor of eventual relapse and metastasis. Collectively, these findings establish SMARCE1 as a key driver of invasive progression in early-stage tumors.

Chapter 2.2: Introduction

The past two decades have brought an exponential increase in the diagnosis of early-stage breast lesions, most commonly ductal carcinoma in situ (DCIS). DCIS remains encapsulated within the ductal-lobular architecture of mammary epithelium; in contrast, invasive breast cancers have escaped this architecture by breaking through the basement membrane, a layer of ECM rich in collagen (IV) and laminins that separates epithelial tissues from the adjacent stromal microenvironment (Fig. S1*A*) (1). This distinction has a critical impact on patient prognosis: whereas women with DCIS show no reduction in survival 5 y after diagnosis, those with invasive cancers have a 15–74% reduction in 5-y survival rates depending on the extent of tumor invasion at diagnosis (2).

Given these observations, there is significant interest in finding genes that promote the invasive progression of early-stage tumors (3). Previous studies have sought molecular alterations present in invasive tumors but not DCIS, leading to the identification of hundreds of genomic and gene-expression alterations specifically associated with invasive cancers (4–6). However, it is unclear if genes that are amplified or up-regulated in invasive cancers also functionally drive DCIS invasion. In large part, the difficulty in addressing this question can be traced to a paucity of experimental systems that model cancer invasion within a microenvironment that faithfully replicates human breast tissue.

The treatment of early-stage cancers remains an unresolved issue. Women with early-stage breast cancers—which include DCIS and stage I tumors that have not entered the lymph nodes—are typically treated by lumpectomy followed by localized radiation. However, recurrence with metastasis occurs in a significant fraction of women with stage I cancers; if such tumors could be prospectively identified, it would be possible to preemptively adopt a more aggressive therapy. Conversely, even though the standard treatment is curative for DCIS, more than half of these lesions are indolent and would never become life-threatening if left untreated (7–9), indicating that there is systematic overtreatment of a significant fraction of patients with DCIS. Collectively, these considerations underscore the importance of defining the genetic drivers of DCIS progression.

In the present study, we identify SMARCE1 as a key driver of early-stage tumor invasion and show that its expression in patients is a strong predictor of whether early-stage tumors will ultimately progress and metastasize.

Chapter 2.3: Results

SMARCE1 Regulates an ECM Invasion Module That Is Up-Regulated upon DCIS Progression

Expression profiling studies have identified ~350 genes that are up-regulated as DCIS tumors progress to invasive cancers (4). We hypothesized that upstream regulators of these transcriptional changes might be master regulators of DCIS progression. Genes with shared upstream regulators form "transcriptional modules" that exhibit correlated fluctuations in their expression (10). We identified two transcriptional modules associated with DCIS progression, containing genes that were highly correlated in their expression across 158 breast cancers (average $\rho = 0.44$, $P < 10^{-15}$; Fig. 1*A*, Fig. S1*B*, and *SI Materials and Methods*).

The larger of these modules encoded for multiple secreted proteases that degrade collagen and laminin, the two main components of basement membrane. Among these proteases were three collagen-degrading matrix metalloproteinases (MMPs): MMP1, MMP2, and MMP13 (11, 12). This module also included the urokinase plasminogen activator (*PLAU*) and its membranous receptor (*PLAUR*), which together degrade laminin (13). In addition, this module included ECM components that stimulate cancer cell invasion, such as *COL1A2*, fibronectin (*FN1*), periostin (*POSTN*), and SPARC, among others (see complete list of genes above). The module also included three lysyl oxidases (*LOX*, *LOXL1*, and *LOXL2*) that remodel ECM by cross-linking collagen, previously implicated in invasion (14). Given these observations, we refer to this set of genes as an "ECM invasion module."

To identify potential regulators of this ECM invasion module, we applied perturbation gene signatures and the Apriori algorithm to estimate the contributions of transcription and chromatin-modifying factors to module gene expression. This analysis identified SMARCE1 as a candidate regulator of the ECM invasion module ($P < 10^{-42}$; Fig. 1*B*); SMARCE1 was not identified as a candidate regulator of identically sized random gene modules (Fig. S1 *C* and *D*) (15). To validate that SMARCE1 regulates this module, we used shRNAs to inhibit SMARCE1 expression in the invasive SUM159 and MDA.MB.231 breast cancer cell lines (Fig. S2*A*). Whereas SMARCE1 inhibition significantly reduced the expression of genes in the ECM invasion module, it had no effect on the expression of a random set of control genes (Fig. 1*C* and Fig. S2*B*).

SMARCE1 Promotes Invasion Through Basement Membrane

We next assessed if SMARCE1 is required for cancer cells to up-regulate proteases and invade through basement membrane. When seeded into 3D basement membrane, the SUM159 and MDA.MB.231 breast cancer lines form clonal spheroids that are initially noninvasive but, over time, invade into the surrounding matrix (16, 17). One week after seeding, the cultured spheroids can be classified by automated image analyses as noninvasive (T-I), partially invasive (T-II) or highly invasive

(T-III; Fig. 2*A* and Fig. S3 *A* and *B*). To quantify matrix protease activity, we supplemented the basement membrane cultures with a modified collagen (IV) substrate that fluoresces upon proteolytic cleavage. Although minimal in the noninvasive spheroids, MMP activity increased progressively upon invasion, with a fourfold increase in partially invasive spheroids and a 10-fold increase in highly invasive spheroids (Fig. S3*C*). Consistent with prior studies, this indicated that invasive progression is associated with up-regulated protease activity.

In the MDA.MB.231 and SUM159 lines, SMARCE1 inhibition almost completely blocked the formation of highly invasive spheroids (i.e., T-III) while also significantly reducing partially invasive spheroids (i.e., T-II; Fig. 2*B* and Fig. S3*D*); SMARCE1 inhibition had no effect on the number or size of the spheroids formed and overall cell numbers (Fig. 2*B* and Fig. S3 *D–F*). SMARCE1 inhibition also resulted in a 75% reduction in the activity of secreted matrix proteases that cleave collagen type IV (Fig. 2*B* and Fig. S3*D*). Consistent with these findings, SMARCE1 protein levels were elevated in partially and strongly invasive spheroids (Fig. S3*G*). To assess reversibility and sufficiency, we inhibited SMARCE1 expression with a doxycycline-inducible shRNA (Fig. S3*H*). Although doxycycline addition for 7 d led to a fivefold increase in noninvasive spheroids, reexpression of SMARCE1 by doxycycline removal was sufficient to trigger invasiveness within 30 h (Fig. 2 *C*and *D*). SMARCE1 overexpression also triggered invasiveness in the noninvasive HMLER breast cancer cell line (Fig. S3*I*). Collectively, these findings indicated that SMARCE1 was dispensable for proliferation, but was required for tumor spheroids to up-regulate protease activity and invade through basement membrane.

Because SMARCE1 is a component of the SWI/SNF complex, its inhibition could, in principle, disrupt the functions of the complex. If this were the case, disrupting the SWI/SNF complex would phenocopy SMARCE1 inhibition. However, inhibiting SMARCC1, a core component of the SWI/SNF complex, abolished cell growth (Fig. S4*A*). Because SMARCE1 is not required for proliferation (Fig. S4*B*), we conclude that its inhibition does not disrupt the core functions of the SWI/SNF complex.

SMARCE1 Is Required for Invasion and Metastasis in Vivo

We next assessed SMARCE1's in vivo function by using an orthotopic mouse model of human breast tumor formation and spontaneous metastasis. In this model, primary tumors are formed by introducing MDA.MB.231-LM2 cells stably expressing luciferase and GFP into the mammary glands of nonobese diabetic (NOD)/SCID mice. Inhibiting SMARCE1 had no effect on primary tumor growth (Fig. S5*A*); however, SMARCE1 inhibition significantly reduced local tumor invasion and entry into the circulation (Fig. 3 *A* and *B*). In contrast to control tumors, which had prominent invasive fronts with numerous cancers cells invading into the surrounding tissue, SMARCE1-inhibited tumors were well-

encapsulated with few cells invading into the adjacent tissue (Fig. 3*A*). Mice harboring SMARCE1-inhibited tumors also exhibited 30-fold lower levels of circulating tumor cells (Fig. 3*B*).

In addition, inhibition of SMARCE1 resulted in a 500-fold reduction in lung metastases (Fig. 3*C*). Staining with an anti-GFP antibody further confirmed that the lungs of animals bearing SMARCE1-inhibited tumors were nearly devoid of cancer cells (Fig. S5*B*). To further clarify the steps in the metastatic cascade in which SMARCE1 was required, we performed tail-vein injections and longitudinally monitored lung metastasis. Eighteen days after injection, metastatic tumor burden was ~10-fold lower in mice injected with SMARCE1-inhibited cancer cells compared with mice injected with cells expressing a control LacZ shRNA (Fig. S5*C*). This 10-fold difference in metastatic burden remained unchanged at later time points (Fig. S5*C*), suggesting that SMARCE1 was important specifically for extravasation or metastatic colony formation, but not for growth of the metastases within the lung parenchyma. Collectively, these observations indicated that SMARCE1 is required for the invasion and metastasis of breast cancers, but is dispensable for their growth.

SMARCE1 Predicts Prognosis in Patients with Early-Stage Tumors

We next investigated the clinical relevance of these findings by assessing if SMARCE1 expression could be used to prospectively identify breast tumors with a propensity to metastasize. Immunostaining for SMARCE1 in tissue microarrays indicated that its expression was lowest in early-stage breast cancers, increased during tumor progression, and was highest in tumors invading into adjacent lymph nodes ($P = 0.007$; Fig. 4*A*). Patients with early-stage breast tumors expressing high levels of SMARCE1 were significantly more likely to show relapse with metastases over a follow-up period of more than 15 y [hazard ratio (HR) = 4.13, $P < 0.0003$; Fig. 4*B*]. Importantly, this prognostic value was observed across multiple independent breast cancer datasets (Fig. S6*A*) and was independent of confounding factors such as grade and tumor size (Fig. S6 *B* and *C*). In contrast, SMARCE1 expression was not predictive of metastasis for patients diagnosed with later-stage tumors that had already invaded to adjacent lymph nodes (Fig. 4*C* and Fig. S6*D*). In addition, stratifying tumors based on expression of other members of the SWI/SNF complex was not predictive of metastasis (Fig. S6 *E-H*).

Stratifying by SMARCE1 expression had similar prognostic value for other types of epithelial tumors. In patients diagnosed with early-stage lung cancers, SMARCE1 expression was strongly predictive of future relapse and metastasis (HR = 7.30, $P < 0.0001$; Fig. S6*I*), but was not predictive for patients diagnosed with later-stage lung cancers (Fig. S6*J*). SMARCE1 expression was also predictive of relapse in early-stage ovarian cancers (HR = 3.35, $P = 0.0052$; Fig. S6*K*), but had no predictive value for later-stage cancers (Fig. S6*L*). These findings complemented the functional observations detailed earlier

and indicated that SMARCE1 expression is strongly predictive of relapse and metastasis for early-stage tumors.

SMARCE1 Is Required to Escape the Ductal-Lobular Architecture of Normal Mammary Tissues

Breast tumors are initially confined in situ within the architecture of normal mammary tissue. To assess SMARCE1 in this context, we used a recently reported 3D model that supports the outgrowth of mammary tissues from primary human breast cells (18). In this model, human mammary tissues are expanded in hydrogels that mimic the microenvironment of the human breast (Fig. S7*A*) (18). To model in situ cancer, we microinjected fluorescently labeled cancer cells into the expanded mammary tissues (Fig. 5*A*).

When inoculated in situ, SUM159-dsRed breast cancer cells proliferated and, over a span of 6 d, migrated to ducts and lobules adjacent to the initial site of injection. As early as 3 d after inoculation, the cells projected long filopodia (>100-μm average length) into the surrounding matrix (Fig. 5 *B–F*). By 11 d, a subset of the inoculated cancer cells had escaped into the surrounding ECM (Fig. 5*B* and Fig. S7*B*). In contrast, nonneoplastic MCF10A cells only spread internally within the mammary tissues and were unable to invade into the surrounding ECM (Fig. 5*B* and Fig. S7*B*).

When SMARCE1 expression was inhibited, the SUM159 cells still proliferated and spread to adjacent ducts and lobules within the normal breast tissues, indicating that SMARCE1 was dispensable for both of these processes. However, the SMARCE1-inhibited cells were unable to extend filopodial projections or invade into the surrounding ECM (Fig. 5 *B–F*). To control for differences in proliferation, we coinjected dsRed-labeled SUM159 cells expressing a control shRNA together with Venus-labeled SUM159 cells inhibited for SMARCE1 expression into shared tissues. The proliferation rates of these lines were indistinguishable, indicating that differences in proliferation were not responsible for the phenotypic differences observed (Fig. S7*C*). These observations indicated that SMARCE1 was dispensable for proliferation and migration of cancer cells growing in situ, but was essential for their escape from the normal tissue architecture and invasion into the surrounding matrix.

SMARCE1 Binds ILF3 and Is Recruited to ILF Motifs

To probe SMARCE1's function, we identified its binding partners in noninvasive and invasive cells by using immunoprecipitation (IP)-MS (Fig. S8*A*). To determine if SMARCE1 binds to its partners together with the SWI/SNF complex, we also performed IP-MS against a core component of the complex (SMARCC1). As expected, SMARCE1 and SMARCC1 were both associated with the SWI/SNF complex in the invasive and noninvasive cells (Fig. 6*A*). However, SMARCE1 was uniquely bound to one factor, ILF3, specifically in invasive cells. Unlike SMARCE1, SMARCC1 did not bind ILF3 in noninvasive or

invasive cells. These findings indicated that SMARCE1 specifically interacts with ILF3 in invasive cells independently of the core SWI/SNF complex (Fig. 6 *A* and *B*).

Because SMARCE1 lacks a sequence-specific DNA binding domain, these findings suggested a model in which ILF3 could be directing the genomic localization of SMARCE1. If this were the case, we would expect ILF3 to also be required for the expression of SMARCE1-regulated genes, in particular those in the ECM-invasion module. Consistent with this, mouse embryonic fibroblasts overexpressing ILF3 up-regulated nearly all of the genes in the SMARCE1-regulated ECM invasion module (63 of 81; $P < 7.1 \times 10^{-8}$; Fig. 6*C*) (19). Moreover, inhibition of ILF3 caused a twofold reduction in the formation of invasive spheroids in the 3D basement membranes (Fig. 6*D*). However, unlike SMARCE1, ILF3 expression was not correlated with or predictive of progression in patients with early-stage cancers ($P = 0.0685$; Fig. S8*B*), suggesting that its mRNA expression is not limiting in patient tumors.

We next performed ChIP and sequencing (ChIP-seq) to identify the binding sites of SMARCE1 in the genomes of noninvasive and invasive cells (Fig. 6*E*) and examine their proximity to ILF3 motifs. SMARCE1 was bound to ~550 genomic sites in noninvasive cells, and 58% of these sites were also bound by SMARCE1 in invasive cells (321 of 554; $P < 0.01$; Fig. S8*C*). However, SMARCE1 was bound to an additional 8,000 sites in invasive cells (Fig. S8*C*), the majority of which were localized to regulatory regions with acetylated H3K27 histones (63% in invasive vs. 7% in noninvasive; $P < 0.01$). The SMARCE1-bound sites in invasive cells were strongly enriched at the enhancers of genes associated with DCIS progression ($P < 0.01$; Fig. 1). In addition, the SMARCE1-bound sites in invasive cells were frequently associated with ILF3 motifs, which, when present, were invariably found at the center of the SMARCE1-bound sites (Fig. 6*F*). In contrast, ILF3 motifs were not enriched at SMARCE1-bound sites in noninvasive cells or at SMARCC1 binding sites (Fig. 6*F* and Fig. S8*D*). These findings were consistent with the observation that SMARCE1 binds to ILF3 only in invasive cells, and further supported a model in which ILF3 was directing the genomic localization of SMARCE1 in invasive cells.

Chapter 2.4: Discussion

These observations establish SMARCE1 as a clinically relevant driver of the invasive progression of early-stage breast cancers. We find that SMARCE1 drives invasion by regulating a module of genes encoding proinvasive ECM and secreted proteases that degrade basement membrane. In functional studies in 3D cultures and animal models, SMARCE1 is dispensable for tumor growth but is required for the invasive and metastatic progression of cancers. The clinical relevance of these findings is underscored by how we were first led to SMARCE1—namely through the analysis of heterogeneous breast tumors containing regions of DCIS and invasive cancer—and by our subsequent analyses of hundreds of patient tumors, indicating that its expression is strongly predictive of metastasis across a spectrum of cancer types.

Currently, most women with early-stage tumors are treated with breast-conserving surgery (i.e., lumpectomy) followed by localized radiation. Although effective for most early-stage tumors, this treatment leads to recurrence with metastasis in ~25% of women with lymph node-negative stage I or stage IIA cancers over a follow-up period of 10 y (20). Our analyses of patient tumors indicate that these poor-prognosis cancers express high levels of SMARCE1 at the time of diagnosis. As a consequence, stratifying patients on the basis of SMARCE1 expression would prospectively identify a subset of early-stage tumors that would be better treated with a more aggressive therapy regimen.

Conversely, epidemiological studies have suggested that more than half of DCIS lesions are indolent and would never become life-threatening if left untreated (7, 21). This has led to an increasing awareness that the current clinical paradigm may be overtreating a significant fraction of women diagnosed with indolent DCIS tumors. To rigorously assess this possibility, several clinical trials are currently under way to determine whether watchful surveillance would lead to equivalent outcomes for such women. Our findings indicate that SMARCE1 should be added to the pathological features used to identify indolent DCIS lesions that are candidates for watchful surveillance.

Although our functional studies focused exclusively on SMARCE1's role in breast cancer, our analyses of the expression profiles of patient tumors suggest that it is likely to also play an important role in regulating the invasiveness of other cancer types. High SMARCE1 expression is predictive of relapse and prognosis in early-stage lung and ovarian cancers, but provided little predictive value for later-stage cancers of these types. This suggests that these tumors use a common underlying mechanism to invade and, ultimately, metastasize. Consistent with this, a shared functional requirement for the invasive progression of all carcinomas is the degradation of the basement membrane that surrounds epithelial tissues. This is the very function conferred by SMARCE1 through the up-regulation of proteases that degrade collagen IV and laminins.

At a mechanistic level, prior studies of SMARCE1 function have focused on its ability to recruit the SWI/SNF complex to genes regulated by hormone receptors (22⇓–24). However, our ChIP-seq and MS observations indicate that SMARCE1 binds most enhancers in the absence of hormone receptors and the SWI/SNF complex, indicating that both are dispensable for SMARCE1's regulatory functions in invasive cells. Moreover, our findings suggest that ILF3 is a key cofactor that mediates SMARCE1's proinvasive regulatory functions. Further study will be needed to determine how SMARCE1 partitions between ILF3-bound and SWI/SNF-bound complexes, and why our findings contrast with those of a recently published study on SMARCE1 in breast cancer (25).

To date, experimental studies of early-stage breast cancer have been significantly limited by a lack of model systems that faithfully recapitulate the tissue microenvironment of early-stage lesions as they occur in human breast tissue. We have addressed this limitation by contributing a model of breast cancer progression in which human cancer cells are integrated in situ into the ductal-lobular architecture of primary human mammary tissues. The development of this model leverages a recently reported method for expanding primary human mammary tissues in culture by using 3D hydrogel scaffolds (18). Because comparable methods have recently been reported for other human tissue types (26, 27), we are hopeful that the strategy used here will prove broadly useful for modeling in situ tumors arising within the relevant human tissue microenvironment.

Chapter 2.5: Methods

In vivo and 3D hydrogel studies were performed as previously described (18, 28). All computational analyses, reagents, public datasets used, and other protocols are described in SI Materials and Methods.

Chapter 2.6: Acknowledgements

We thank Wendy Salmon for assistance with microscopy and Prof. Hazel Sive for sharing microinjection equipment. This research was supported by National Science Foundation Graduate Research Fellowship Grant 1122374 (to E.S.S.), a Ludwig Fund for Cancer Research fellowship (to Y.-X.F), and a grant from the Ludwig Fund for Cancer Research.

Chapter 2.7: References

1. Place AE, Jin Huh S, Polyak K (2011) The microenvironment in breast cancer progression: Biology and implications for treatment. Breast Cancer Res 13:227.
2. Howlader N, et al. (2016) SEER Cancer Statistics Review, 1975-2013 (National Cancer Institute, Bethesda, MD).
3. Kerlikowske K, et al. (2010) Biomarker expression and risk of subsequent tumors after initial ductal carcinoma in situ diagnosis. J Natl Cancer Inst 102:627–637.
4. Schuetz CS, et al. (2006) Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. Cancer Res 66:5278–5286.
5. Hernandez L, et al. (2012) Genomic and mutational profiling of ductal carcinomas in situ and matched adjacent invasive breast cancers reveals intra-tumour genetic heterogeneity and clonal selection. J Pathol 227:42–52.
6. Yao J, et al. (2006) Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. Cancer Res 66:4065–4078.
7. Sanders ME, Schuyler PA, Dupont WD, Page DL (2005) The natural history of lowgrade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. Cancer 103:2481–2484.
8. Collins LC, et al. (2005) Outcome of patients with ductal carcinoma in situ untreated after diagnostic biopsy: Results from the Nurses' Health Study. Cancer 103:1778–1784.
9. Bleyer A, Welch HG (2012) Effect of three decades of screening mammography on breast-cancer incidence. N Engl J Med 367:1998–2005.
10. Wang X, Dalkic E, Wu M, Chan C (2008) Gene module level analysis: Identification to networks and dynamics. Curr Opin Biotechnol 19:482–491.
11. Knäuper V, López-Otin C, Smith B, Knight G, Murphy G (1996) Biochemical characterization of human collagenase-3. J Biol Chem 271:1544–1550.
12. Matrisian LM (1990) Metalloproteinases and their inhibitors in matrix remodeling. Trends Genet 6:121–125.
13. Nakagami Y, Abe K, Nishiyama N, Matsuki N (2000) Laminin degradation by plasmin regulates long-term potentiation. J Neurosci 20:2003–2010.
14. Siegel RC, Fu JC, Chang Y (1976) Collagen cross-linking: The substrate specificity of lysyl oxidase. J Biol Chem 251(18):5779–5785.

15. Wilson BG, Roberts CWM (2011) SWI/SNF nucleosome remodellers and cancer. Nat Rev Cancer 11:481–492.
16. Petersen OW, Rønnov-Jessen L, Howlett AR, Bissell MJ (1992) Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells. Proc Natl Acad Sci USA 89: 9064–9068.
17. Weaver V, Petersen O (1997) Reversion of the malignant phenotype of human breast cells in three-dimensional culture and in vivo by integrin blocking antibodies. J Cell Biol 137(1):231–245.
18. Sokol ES, et al. (2016) Growth of human breast tissues from patient cells in 3D hydrogel scaffolds. Breast Cancer Res 18:19.
19. Todaka H, et al. (2015) Overexpression of NF90-NF45 represses myogenic MicroRNA biogenesis, resulting in development of skeletal muscle atrophy and centronuclear muscle fibers. Mol Cell Biol 35:2295–2308.
20. Schmidt M, et al. (2008) The humoral immune system has a key prognostic impact in node-negative breast cancer. Cancer Res 68:5405–5413.
21. Alvarado M, Ozanne E, Esserman L (2012) Overdiagnosis and overtreatment of breast cancer. Am Soc Clin Oncol Educ Book e40–e45.
22. Link KA, et al. (2005) BAF57 governs androgen receptor action and androgendependent proliferation through SWI/SNF. Mol Cell Biol 25:2200–2215.
23. García-Pedrero JM, Kiskinis E, Parker MG, Belandia B (2006) The SWI/SNF chromatin remodeling subunit BAF57 is a critical regulator of estrogen receptor function in breast cancer cells. J Biol Chem 281:22656–22664.
24. Balasubramaniam S, et al. (2013) Aberrant BAF57 signaling facilitates prometastatic phenotypes. Clin Cancer Res 19:2657–2667.
25. Sethuraman A, et al. (2016) SMARCE1 regulates metastatic potential of breast cancer cells through the HIF1A/PTK2 pathway. Breast Cancer Res 18:81.
26. Dye BR, et al. (2015) In vitro generation of human pluripotent stem cell derived lung organoids. eLife 4:e05098.
27. Sato T, et al. (2011) Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. Gastroenterology 141: 1762–1772.
28. Feng Y, et al. (2014) Epithelial-to-mesenchymal transition activates PERK-eIF2a and sensitizes cells to endoplasmic reticulum stress. Cancer Discov 4(6):702–715.
29. Minn AJ, et al. (2005) Genes that mediate breast cancer metastasis to lung. Nature 436:518–524.

30. Gyorffy B, Lánczky A, Szállási Z (2012) Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. Endocr Relat Cancer 19:197–208.
31. Marson A, et al. (2008) Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell 134(3):521–533.

Chapter 2.8: Figures and Tables

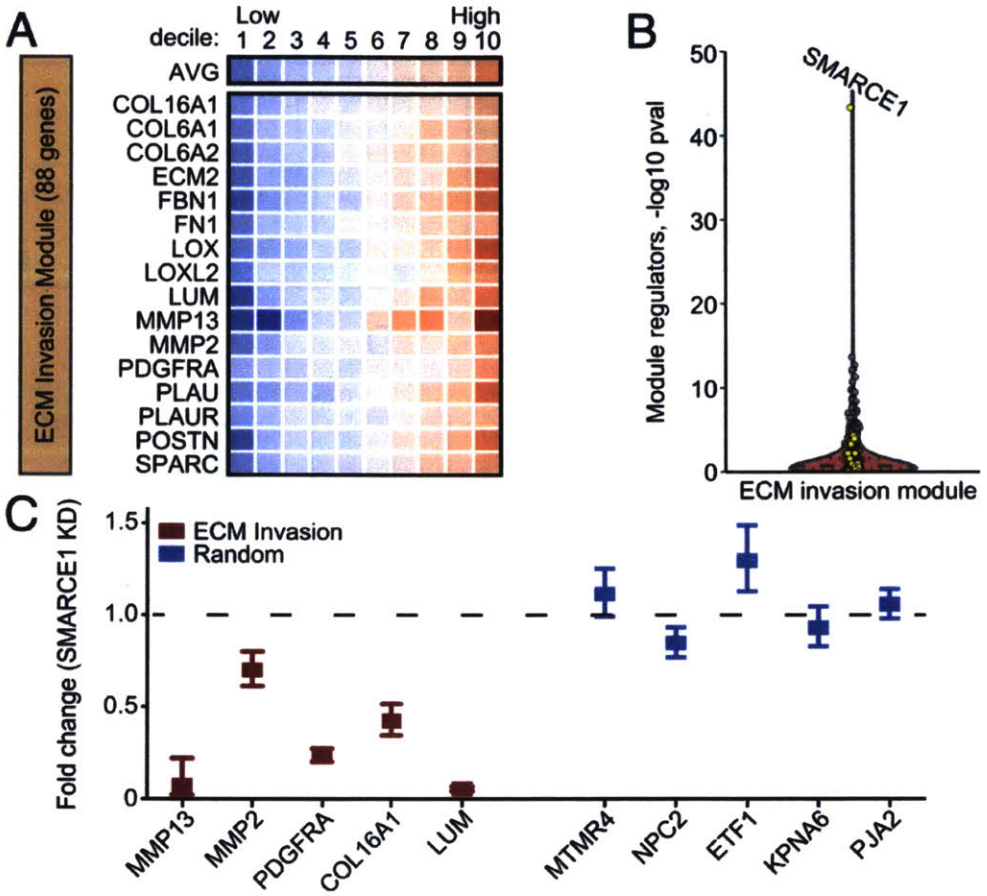


Fig. 1. SMARCE1 regulates an ECM invasion module that is up-regulated upon DCIS progression. (A) Expression of genes in the ECM invasion module across 158 primary human breast tumors. Tumors were sorted based on their average expression of the 88 genes in the module (AVG), and divided into ten groups (deciles). Heat map denotes the average expression in the corresponding decile. (B) Violin plot showing the contributions of 1,124 transcription and chromatin-modifying factors to expression of the ECM invasion module. Statistical significance was computed with the hypergeometric test. SWI/SNF complex genes are highlighted in yellow. (C) Quantitative PCR analysis of expression of ECM invasion module genes and random module genes in SUM159 cells transduced with control or SMARCE1 shRNAs. Gene expression is normalized to GAPDH and plotted as fold change relative to the control cell line (n = 4; mean ± SEM). 6

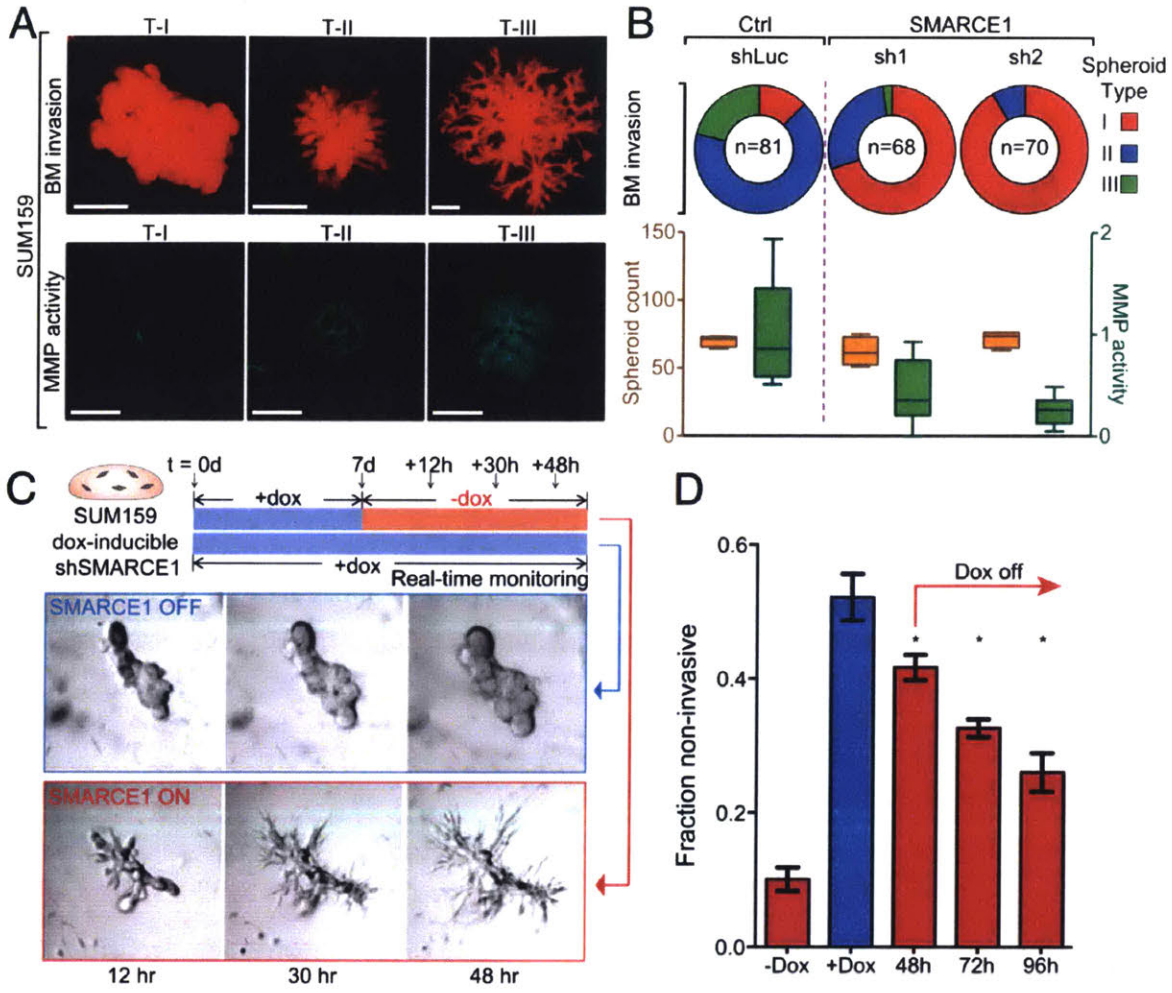


Fig. 2. SMARCE1 is required for cancer cell invasion through basement membrane. (A) Epifluorescence images of tumor spheroids formed by SUM159 cells in 3D basement membrane (BM). (*Top*) Noninvasive (T-I), partially invasive (T-II), or highly invasive (T-III) spheroids. (*Bottom*) Collagen IV hydrolysis (green) in DQ collagen IV-supplemented BM. (Scale bars: 100 μm.) (B) Quantification of tumor spheroid invasiveness, protease activity, and number in control and shSMARCE1 SUM159 cells. (*Top*) Quantification of T-I, T-II, and T-III spheroids. (*Bottom*) Spheroid MMP activity relative to shLuc controls and spheroid number per BM. (C) SMARCE1 reexpression rescues invasive progression in tumor spheroids. (*Top*) Experimental design to measure effect of SMARCE1 reexpression on spheroid invasiveness. (*Bottom*) Representative images of tumor spheroids 12, 30, and 48 h after doxorubicin (dox) withdrawal (SMARCE1 ON) or continued treatment with doxorubicin (SMARCE1 OFF). (D) Quantification of noninvasive spheroids after SMARCE1 inhibition for 7–9 d (+dox) and subsequent reexpression of SMARCE1 (dox off) for 48, 72, or 96 h. All spheroids were quantified at day 11 (n = 4; *$P < 0.05$).

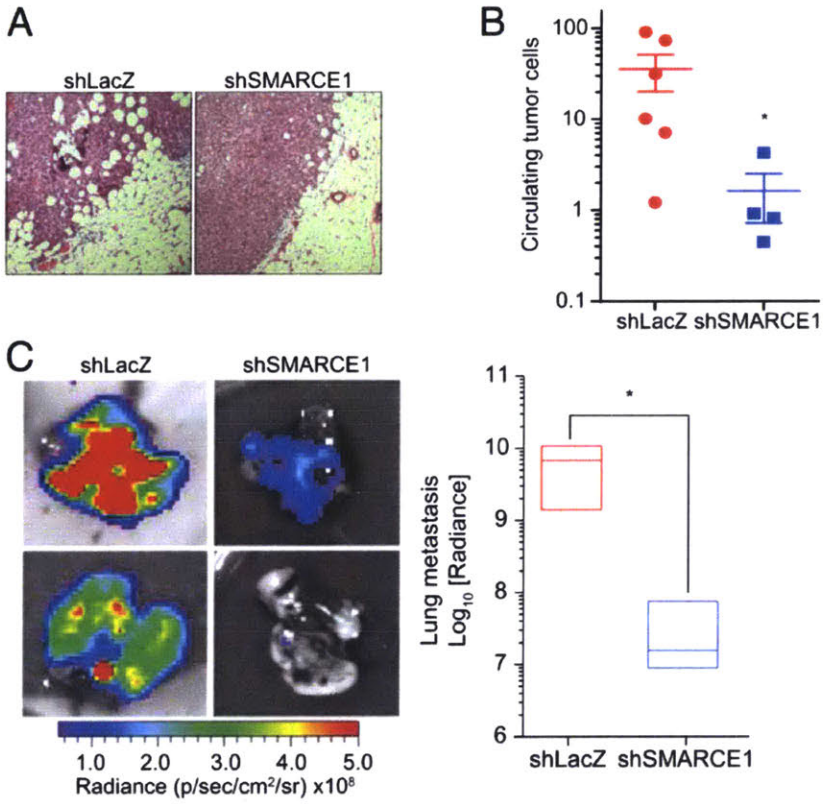


Fig. 3. SMARCE1 is essential for metastasis in vivo. (A) Representative H&E- stained sections of the tumor boundary from MDA.MB.231-LM2–injected control (shLacZ) and SMARCE1-inhibited (shSMARCE1) tumors 4 wk after injection. (B) Number of circulating tumor cells in mice bearing control (n = 6) or SMARCE1-inhibited (n = 4) tumors. (C) Representative luminescence images and quantification of metastatic burden in whole lungs of mice inoculated with shLacZ (n = 3) or shSMARCE1 (n = 4) MDA.MB.231-LM2 cells (*P < 0.05).

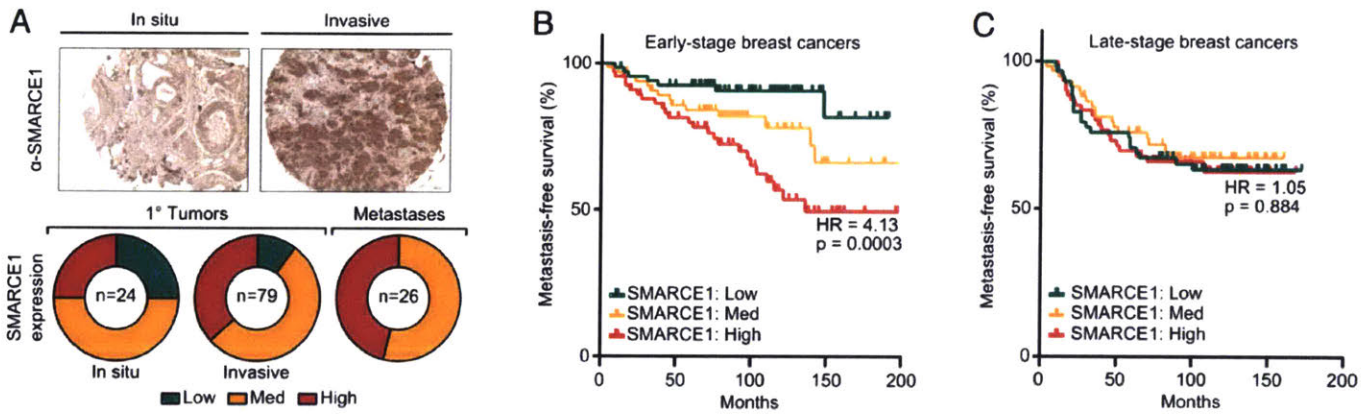


Fig. 4. SMARCE1 expression is prognostic in early-stage tumors. (A) Patient breast tissues from in situ and invasive breast cancers were immunohistochemically stained for SMARCE1 (*Top*). In situ breast cancer, invasive breast cancer, and metastasis staining intensities were quantified (*Bottom*). (B and C) SMARCE1 expression was examined in a cohort of patients with early-stage breast tumors [N stage 0 (lymph node-negative), GSE11121; n = 200] and late- stage breast tumors (N stage ≥ 1, GSE20685; n = 190). Metastasis-free survival curves in patients stratified into tertiles (high, medium, low) based on tumor SMARCE1 expression. HRs and P values were determined with the log-rank statistical test.

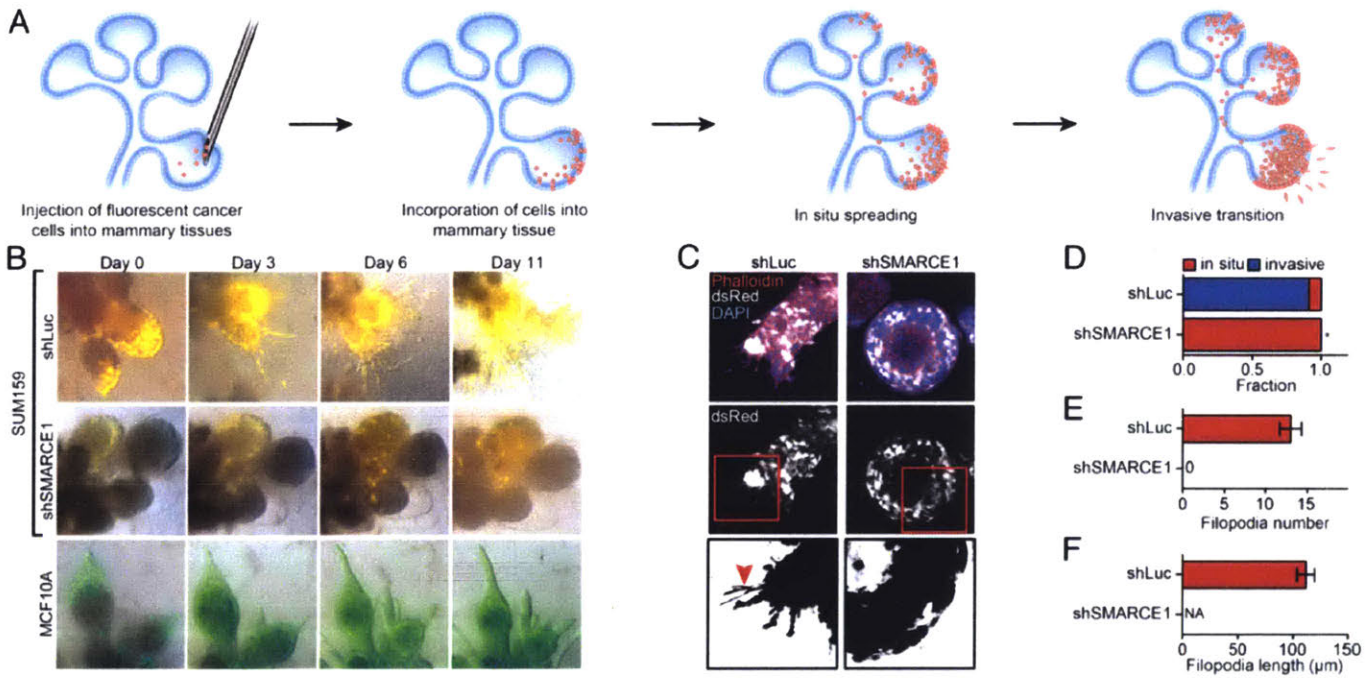


Fig. 5. SMARCE1 is required for cancers to escape the ductal-lobular architecture of normal mammary tissues. (A) Schematic of the tissue model of invasive progression of in situ cancer cells. (B) Representative bright-field images with a fluorescent overlay of dsRed-labeled SUM159 cells transduced with shLuc or shSMARCE1 (pseudocolored yellow) and Venus-labeled MCF10A cells injected into mammary tissues. (C) Confocal images of tissues 3 d postinjection. Red arrowhead indicates filopodia. (D) Fraction of SUM159 cells (shLuc or shSMARCE1) that remained encapsulated (in situ) or escaped the tissue architecture (invasive) 3 d postinjection. (E) Number of filopodia formed by SUM159 cells per tissue injected in D. (F) Length of the filopodia from E (*$P < 0.05$).

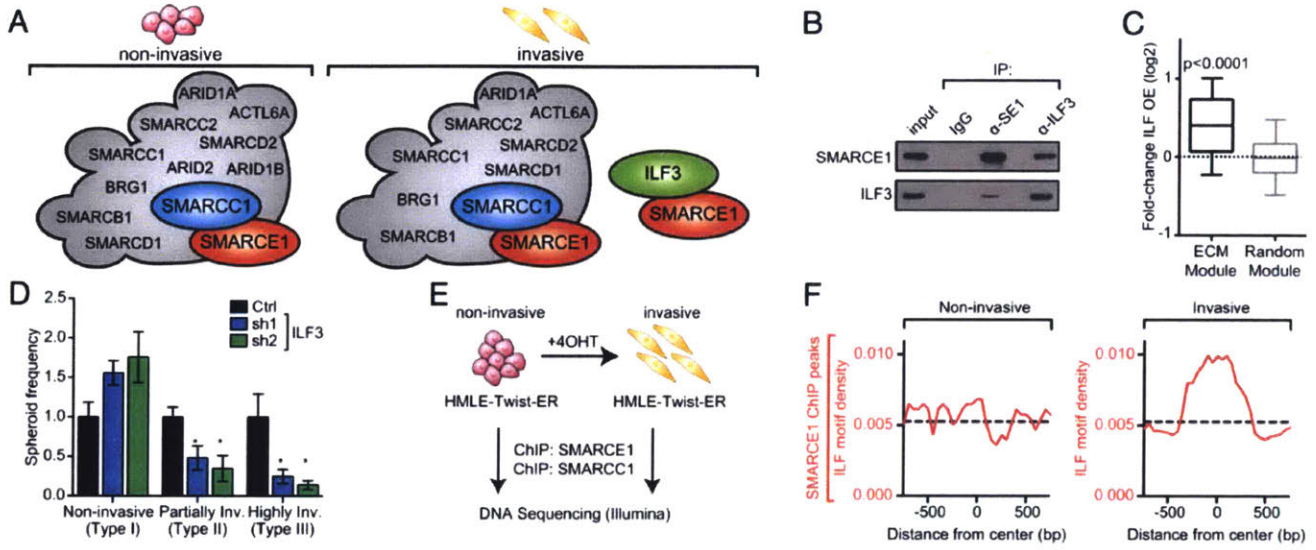


Fig. 6. SMARCE1 binds ILF3 and is recruited to ILF motifs. (A) Schematic representation of the list of binding partners of SMARCE1 or SMARCC1 in non- invasive or invasive HMLE-Twist-ER cell [untreated or treated with 125 nM 4-hydroxytamoxifen (4OHT), respectively] identified by co-IP and MS. SMARCE1 and SMARCC1 interacted with each other as well as other members of the SWI/SNF complex (shown in gray) in invasive and noninvasive cells. An additional unique interaction between SMARCE1 and ILF3 was detected in invasive HMLE-Twist-ER cells. (B) Co-IP of endogenous SMARCE1 (SE1) and ILF3 in nuclear lysates from invasive SUM159 cells. (C) Expression of the ECM invasion module or a random module of 88 genes in embryonic fibroblasts from transgenic mice over- expressing ILF2/3 (GSE67591) relative to expression in genetically matched mice without ILF2/3 overexpression. (D) Quantification of tumor spheroid invasiveness in control (shLuc) or shILF3 (sh1, sh2) SUM159 cells grown in basement membrane. (E) Experimental design for ChIP and sequencing (ChIP-seq) of SMARCC1 and SMARCE1 in noninvasive and invasive HMLE-Twist-ER cells (untreated or treated with 125 nM 4OHT, respectively). (F) SMARCE1 is localized to ILF motifs specifically in invasive HMLE-Twist-ER cells. Motif density was calculated by using a sliding window of 50 bp extending to 750 bp on either side of SMARCE1-bound peaks identified by ChIP-seq (*$P < 0.05$).

Chapter 2.9: Supplemental Data

SI Materials and Methods

Cell Culture and Reagents

SUM159 breast cancer cells (Asterand) were cultured in F12 with 5% FBS (Sigma Aldrich), insulin, and hydrocortisone. MDA.MB.231 cells and their highly metastatic derivative line (LM2) (29) were cultured in DMEM supplemented with 10% FBS. MCF10A cells (American Type Culture Collection) were cultured in MEGM (CC-3150; Lonza) with 100 ng/mL cholera toxin. All media was supplemented with L-glutamine and penicillin-streptomycin, and cells were grown on tissue culture grade plastic at 37 °C, 5% CO_2 in a humidified chamber.

Module Identification

Gene expression data from 158 breast tumors was downloaded from the Gene Expression Omnibus (GEO) database (accession no. GSE3143), and the list of genes up-regulated as DCIS tumors progress to invasiveness was obtained from Neubauer and coworkers (4). Expression data were log-transformed and median-normalized, and Spearman correlations were computed across the 158 tumors for each pair of genes in the DCIS-to-invasive progression signature. In cases in which multiple probes corresponded to a single gene, the maximum intensity probe was used. Hierarchical clustering of this pairwise correlation matrix identified two modules of sizes 88 and 45 (Fig. S1*B*).

Computational Identification of Candidate Upstream Regulators

A set of 1,124 chromatin modifiers and transcription factors was identified by selecting all genes containing the following Gene Ontology (GO) terms: GO:0016568, GO:0006338, or GO:0003700. With each gene, we associated a set of direct and indirect downstream targets by mining tumor expression data (accession no. GSE3143) for pairwise association rules with the Apriori algorithm and by extracting perturbation gene sets when available from the MolSigDB.

As the Apriori algorithm requires categorical data, genes were categorized into "high" or "not high" expression in each tumor. For any given gene and breast tumor, the threshold for "high expression" was set to be at least 1.8-fold higher than the median expression across all 158 tumors. Any genes that did not vary significantly across tumors were removed before running the algorithm. We implemented Apriori by identifying, for a given gene A, all genes B such that Pr(B high | A high) > Pr(B high), discovering all genes B whose expression was conditionally increased at least twofold when gene A was highly expressed. This associated a set of genes with each gene A. In cases in which both Pr(B high | A high) > Pr(B high) and Pr(A high | B high) > Pr(A high), we assigned directionality to this statistical

interaction if there was at least a twofold difference between the conditional probabilities. Activated downstream targets were also identified by extracting, when available, perturbation gene sets from MolSigDB for chromatin modifiers and transcription factors. Gene sets larger than 1,000 or smaller than 100 were excluded to improve statistical robustness.

We identified the overlap between each of the target sets derived and the gene modules of interest ("ECM invasion" or a "random" module of 88 genes), and determined statistical significance by using the hypergeometric test. The following is a list of genes in the ECM invasion module: *AEBP1*, *ANGPTL2*, *AXL*, *BGN*, *C1R*, *C1S*, *C5orf13*, *CALD1*, *CDH11*, *CFH*, *CLEC11A*, *CLEC2B*, *COL10A1*, *COL11A1*, *COL16A1*, *COL1A2*, *COL3A1*, *COL5A2*, *COL6A1*, *COL6A2*, *COL6A3*, *COMP*, *CTGF*, *CTSK*, *DPYSL3*, *ECM2*, *EDNRA*, *EFEMP2*, *EMP3*, *FAP*, *FBLN2*, *FBN1*, *FN1*, *GAS1*, *GPNMB*, *HSD17B6*, *IGFBP3*, *IGFBP7*, *ISLR*, *KIAA1199*, *KIAA1462*, *LAMA4*, *LAMB1*, *LAMC1*, *LGALS1*, *LOX*, *LOXL1*, *LOXL2*, *LRRC15*, *LUM*, *MAF*, *MFAP2*, *MICAL2*, *MMP1*, *MMP11*, *MMP13*, *MMP2*, *MRC2*, *MYL9*, *NBL1*, *NDN*, *NID2*, *NNMT*, *OLFML2B*, *PDGFRA*, *PDGFRB*, *PLAU*, *PLAUR*, *PLS3*, *POSTN*, *PTRF*, *RARRES2*, *RECK*, *SEC23A*, *SERPINF1*, *SERPING1*, *SERPINH1*, *SLC16A3*, *SPARC*, *SPON1*, *SRPX2*, *SULF1*, *TGFB1I1*, *TGFBI*, *THBS2*, *THY1*, *TWIST1*, and *WWTR1*.

Three-Dimensional Invasion Assays

Invasion through basement membranes was assessed by seeding cells in growth factor-stripped Matrigel (BD Biosciences). Briefly, Matrigel was thawed on ice, and 0.1 mL was added per chamber and allowed to solidify for 30 min at 37 °C, 5% CO_2, in a humidified incubator. When indicated, the Matrigel was supplemented with 1% DQ-collagen IV to assess protease activity (Life Technologies). Cells were seeded on top of the basement membrane in a single-cell suspension with 500 cells in 0.2 mL and allowed to settle for 20–30 min. The cultures were layered with additional basement membrane through the dropwise addition of 200 μL of media supplemented with 10% Matrigel. A total of 0.2 mL of media was exchanged after 4 d, and cultures were fixed and imaged on day 7. Invasion through polymerized ECM was assessed by seeding 1,000 cells into 200 μL of polymerized 1.3 mg/mL collagen I (Corning). Spheroids were scored on a three-point scale based on the degree of invasiveness. Noninvasive type I structures have smooth edges with no visible cellular protrusions. Type II structures exhibit partial invasiveness as evidenced by cells protruding from the colony mass and occasional scattering of the peripheral cells. Type III structures are fully scattered and have lost a central mass. Form factors for the structures were calculated in Cell Profiler.

For microinjection assays, patient-derived mammary tissues were grown in 3D hydrogel culture for 2 wk, as previously described (18). Briefly, reduction mammary tissue was mechanically and enzymatically dissociated with collagenase and hyaluronidase. Purified epithelial tissue fragments were

seeded into ECM hydrogels consisting of 1.7 mg/mL collagen I (Corning), 10 μg/mL hyaluronan (Millipore), 40 μg/mL laminin isolated from Engelbreth-Holm-Swarm sarcoma cells (Life Technologies), and 20 μg/mL fibronectin (Life Technologies), pH 7.3, in MEGM (Lonza).

Cells for injection were harvested by trypsinization and resuspended in their respective medium at concentrations of 20–40 million cells per milliliter. Immediately before injection, Trypan blue solution was added at a concentration of 0.03% for visualization and cell viability monitoring purposes. Injection pipets were made by using a pipet puller (P-97; Sutter Instrument) and a microforge (MF-900; Narishige). A typical injection pipette, made from a capillary glass (Sutter Instrument), had an outside diameter of approximately 50 μm and an opening with a diameter of approximately 25 μm. Cells were injected into the luminal space of mammary tissues growing in an ECM hydrogel by using a micromanipulator (MN-151; Narishige) and a microinjection device (IM-200; Narishige) under a stereomicroscope (Stemi SV6; Zeiss).

Live imaging was performed at indicated time points after injection by using a stereomicroscope (Discovery V8; Zeiss) and a camera (AxioCam MRc; Zeiss). Filopodial length was calculated from bright-field images in ImageJ. For comixing experiments, an equal number of Venus-labeled shSMARCE1 cells and dsRed-labeled shLacZ cells were injected into the tissues. The structures were imaged for 6 d, and the fraction of total fluorescent intensity within the imaged focal plane of a structure was determined for Venus and dsRed signal by using ImageJ.

Microscopic Imaging and Analysis

Confocal images of spheroids were captured by using a Zeiss LSM 700 series microscope, and processed by using the ZEN imaging software using a maximum intensity projection from a *Z*-stack.

Co-IP and MS Analysis

A total of 2×10^7 cells were seeded onto a 15-cm dish 24 h before nuclear protein extraction. Nuclear protein was collected by using the NE-PER kit (Thermo Fisher). A total of 1 mg nuclear protein extract was incubated overnight with 50 μL Protein A/G Dynabeads conjugated with 2 μg control IgG, anti-SMARCE1 (no. 70540; Abcam) or anti-SMARCC1 antibodies (no. 9746; Santa Cruz). Eluate from the aforementioned IP was separated by SDS/PAGE, the gel was silver-stained, and bands were cut out for analysis by LC-MS/MS.

Western Blotting

Western blotting was performed as previously described (28). Primary antibodies include anti-SMARCE1 (no. 70540; Abcam), anti-SMARCC1 (no. 9746; Santa Cruz), anti-FLAG (F3165; Sigma), anti-ILF3 (no. 133354; Abcam), and anti–β-Actin (no. 5125; Cell Signaling Technology).

Tumor Dataset Analysis

Cancer patient gene expression datasets were downloaded from GEO (www.ncbi.nlm.nih.gov/geo). Early-stage breast tumor data were obtained from accession nos. GSE11121 (Fig. 3) and GSE21653 (Fig. S6) and late-stage breast tumor data from accession nos. GSE20685 (Fig. 3) and GSE21653 (Fig. S6). Early-stage lung tumor data were obtained from accession no. GSE31210 (Fig. S6) and late-stage lung tumor data from accession no. GSE30219 (Fig. S6). Ovarian cancer data were obtained from KMplot Ovarian Cancer 2015 version (30) (Fig. S6). Early-stage breast tumors were defined as node-negative (i.e., N stage 0), whereas late-stage breast tumors were defined as those with N stage $\geq$ 1. Early-stage lung tumors were defined as stage IA or IB (N stage 0), whereas late-stage lung tumors were defined as those with N stage $\geq$ 1. Early-stage ovarian cancers were defined as stage I or II, whereas late-stage cancers were defined as stage III or IV. Tumors were binned into SMARCE1-high (top third of SMARCE1 expression) and SMARCE1-low (bottom third of SMARCE1 expression) based on the expression of probe 211989_at. ILF3 binning was performed with 208930_s_at. HRs and *P* values were calculated for these two groups by log-rank (Mantel–Cox) test.

In Vivo Tumorigenicity and Metastasis Assays

All mouse procedures were approved by the animal care and use committees of the Massachusetts Institute of Technology and conducted in accordance with institutional policies. For tumor-seeding studies, 1×10^6 of the highly metastatic MDA.MB.231 (LM2) cells were suspended in 50 μL of 1:1 mix of Matrigel and DMEM and injected into the mammary fat pad of NOD/SCID mice (Jackson Labs). At 3 wk, tumors were surgically resected and weighed. Mice were monitored for an additional 4 wk until they were killed and lungs were collected. To measure metastases, freshly collected lungs were soaked in D-luciferin (150 μg/mL; PerkinElmer) for 15 min and imaged with the IVIS system (Perkin-Elmer) at nonsaturating exposures. Radiance was quantified with LivingImage version 4.4 software.
For tail-vein injections, 1×10^6 LM2 cells were suspended in DMEM and injected into the tail vein of NOD/SCID mice. Living mice were visualized for tumor cell burden using the IVIS System (Perkin-Elmer) approximately weekly; mice were injected with 200 μL of luciferin and imaged 10 min later at nonsaturating exposures.

Measurement and Quantification of Circulating Tumor Cells

A total of 1×10^5 luciferase-labeled MDA.MB.231-LM2 cells were injected into the fat pad of NOD/SCID mice. Seven weeks after tumor implantation, 400 μL of peripheral blood was collected. Blood samples were centrifuged at 600 × g to precipitate the cell fractions. One milliliter of red blood cell lysis buffer (Stem Cell Technology) was applied to resuspend the pellet on ice for 5 min. The samples were then centrifuged again at 2,000 rpm. A total of 200 μL PBS solution containing luciferin was used to resuspend the pellet, and luciferase activity of all samples were immediately measured by using the IVIS imaging system. To quantify the number of luciferase-positive cells, a standard curve of MDA.231.LM2 cells, ranging from 1,000 to 12, were mixed with 400 μL of blood samples from tumor-free NOD/SCID mice and followed by the same purification and measurement processes described earlier. A linear correlation between luciferase activity and cell number was then generated by linear regression, and the number of luciferase-positive circulating tumor cells per milliliter from each tumor-bearing mouse was calculated by using the resulting formula.

Immunohistochemistry

Matrigel spheroids were fixed with 4% paraformaldehyde for 30 min, washed with PBS solution, and permeabilized by using 0.1% Triton X-100 in PBS. Spheroids were treated with H_2O_2-MeOH solution (10% H_2O_2 + 50% MeOH + 40% PBS solution) at 4 °C for 30 min and washed with PBS solution. Blocking, antibody incubations, and HRP detection was performed by using a DAKO EnVision Kit (K4010) following the manufacturer's instructions. Anti-SMARCE1 (Abcam 70540) was used at 1:250.

Resected tumors were fixed in 10% neutral-buffered formalin and embedded in paraffin. Serial sections at 5 μm were used for histological analyses. For SMARCE1 staining, breast tumor microarray was ordered from US Biomax (BR2082). Briefly, slides were deparaffinized by using xylene and a series of alcohols, followed by heating in 10 mM citrate buffer (pH 6.0) for antigen retrieval. Sections were blocked and incubated with primary antibodies (GFP; Cell Signaling no. 2956; or SMARCE1; Abcam no. 70540), washed, incubated with corresponding secondary antibodies, and stained with appropriate detection reagents. Nuclei were counterstained with hematoxylin. H&E staining was performed by using standard procedures.

To quantify SMARCE1 expression, each tissue section was scored on (*i*) the percentage of positive cells on a scale of 0–3—0 (0% positive cells), 1 (<25%), 2 (25–75%), 3 (>75%)—and (*ii*) the staining intensity on a scale of 0–3: 0, negative staining; 1, weak staining; 2, moderate staining; or 3, strong staining. The two scores were multiplied, resulting in a value ranging from 0 to 9, whereby low SMARCE1 expression ranged from 0 to 3, medium SMARCE1 expression ranged from 4 to 6, and high SMARCE1 expression ranged from 7 to 9.

ChIP

ChIP was performed with a SMARCE1 antibody (no. 70540; Abcam) or a SMARCC1 antibody (no. 9746; Santa Cruz) as described previously (27). ChIP-seq libraries were prepared at the Massachusetts Institute of Technology BioMicroCenter by using Beckman Coulter SPRIworks, and run on a HiSeq 2000. Reads were aligned to the human genome build hg19 with a maximum of two mismatches. ChIP-seq peaks were identified by using MACS with default settings. Peaks with a *P* value less than 10^{-9} were used in follow-up analyses.

ILF3 Gene Expression and Motif Generation

Log2-transformed gene expression data were downloaded from GEO accession no. GSE67591. Gene expression of the 88 ECM module genes or a module of 88 random genes was plotted as a ratio of expression in the overexpression vs. WT conditions. All genes with a greater than twofold change in gene expression in the ILF2/3-overexpressing cells vs. WT cells were used to run motif analysis with HOMER (Hypergeometric Optimization of Motif Enrichment). An enriched ILF motif was generated by analyzing sequences flanking the transcription start site (±2 kb) of the selected genes.

Overlap Analysis

ChIP-Seq peaks were mapped to the closest mapped gene by using HOMER. H3K27Ac peaks called from seven cell lines from the ENCODE (Encyclopedia of DNA Elements) project were downloaded. Active enhancers from all cell lines were merged and filtered for blacklist regions. SMARCE1 peaks that overlapped with merged H3K27Ac peaks were called by using HOMER. Significance of the overlap between the invasive ductal carcinoma (IDC) signature genes (4) and the SMARCE1-bound genes was performed by using a binomial test using the background of all genes.

SI Figures

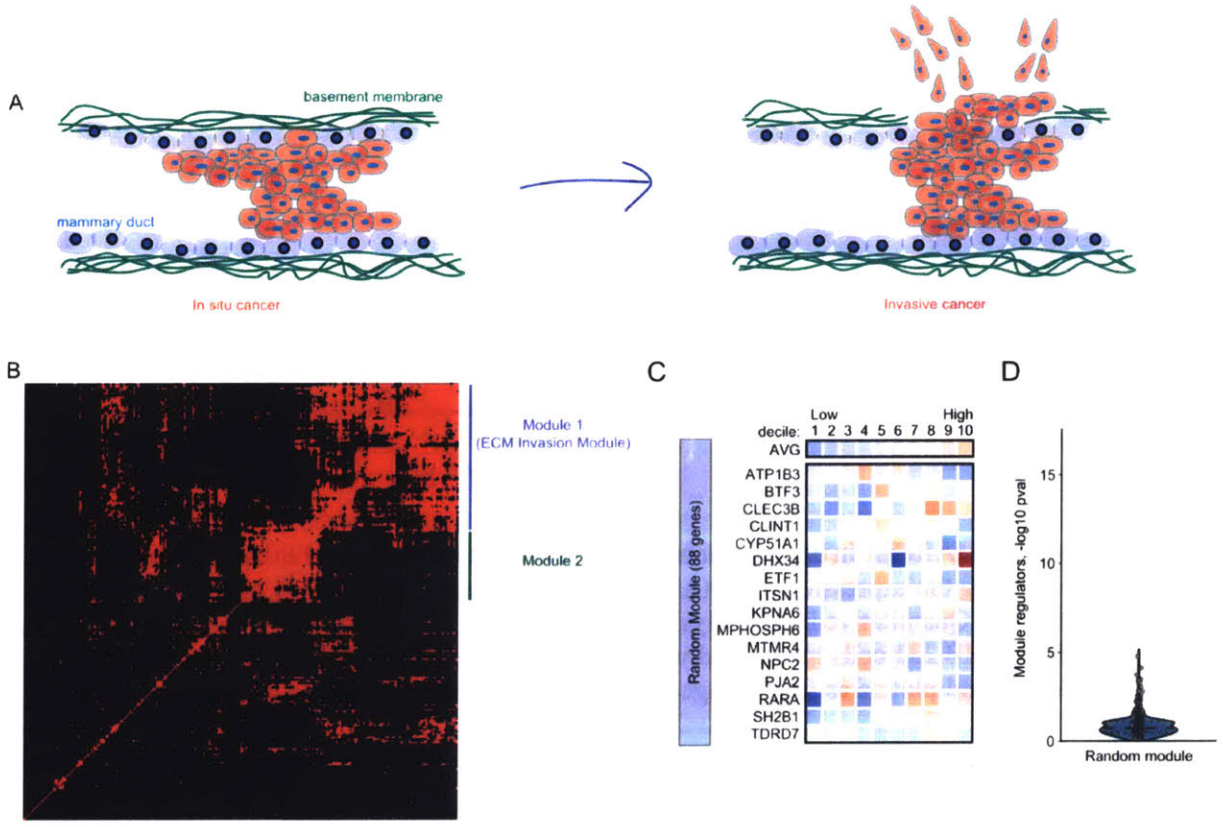


Fig. S1. Computational analysis of gene modules involved in DCIS-to-IDC progression. (A) Schematic representation of the transition from DCIS to invasive breast cancer. In DCIS (Left), cancer cells (red) remain encapsulated within the ducts (blue) and are surrounded by the basement membrane (green). Subsequently, as cells transition to invasive carcinoma (Right), they degrade the basement membrane and invade into surrounding tissues. (B) Spearman correlations were calculated from all pairwise comparisons of genes up-regulated in invasive tumors vs. DCIS tumors. Hierarchical clustering was performed on the pairwise correlation matrix. Plotted is a heat map of the clustered correlation matrix. Red indicates a high correlation, and black indicates low/no correlation. The blue bar marks genes present in module 1, and the green bar marks genes present in module 2. (C) Expression of genes in a random module across 158 primary human breast tumors. Tumors were sorted based on their average expression of the 88 genes in the module (AVG), and divided into ten groups (deciles). The heat map denotes the average expression in the corresponding decile. (D) A violin plot showing the contributions of 1,124 transcription and chromatin-modifying factors to expression of the random module. Statistical significance was computed by using the hypergeometric test. SWI/SNF complex genes are highlighted in yellow.

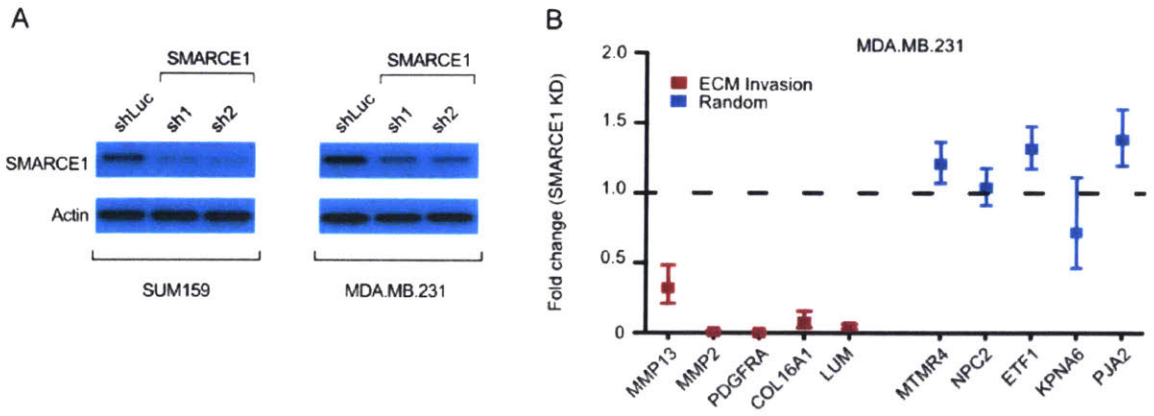


Fig. 2. Knockdown of SMARCE1 in breast cancer cells reduces the expression of ECM invasion module genes. (A) Western blot showing the expression of SMARCE1 in control and shSMARCE1 (sh1 and sh2) SUM159 and MDA.MB.231 cells. (B) Quantitative PCR analysis of expression of ECM invasion module genes and random module genes in MDA.MB.231 cells transduced with control or SMARCE1 shRNAs. Gene expression is normalized to GAPDH and plotted as fold change relative to the control cell line (n = 4 plotted as mean ± SEM).

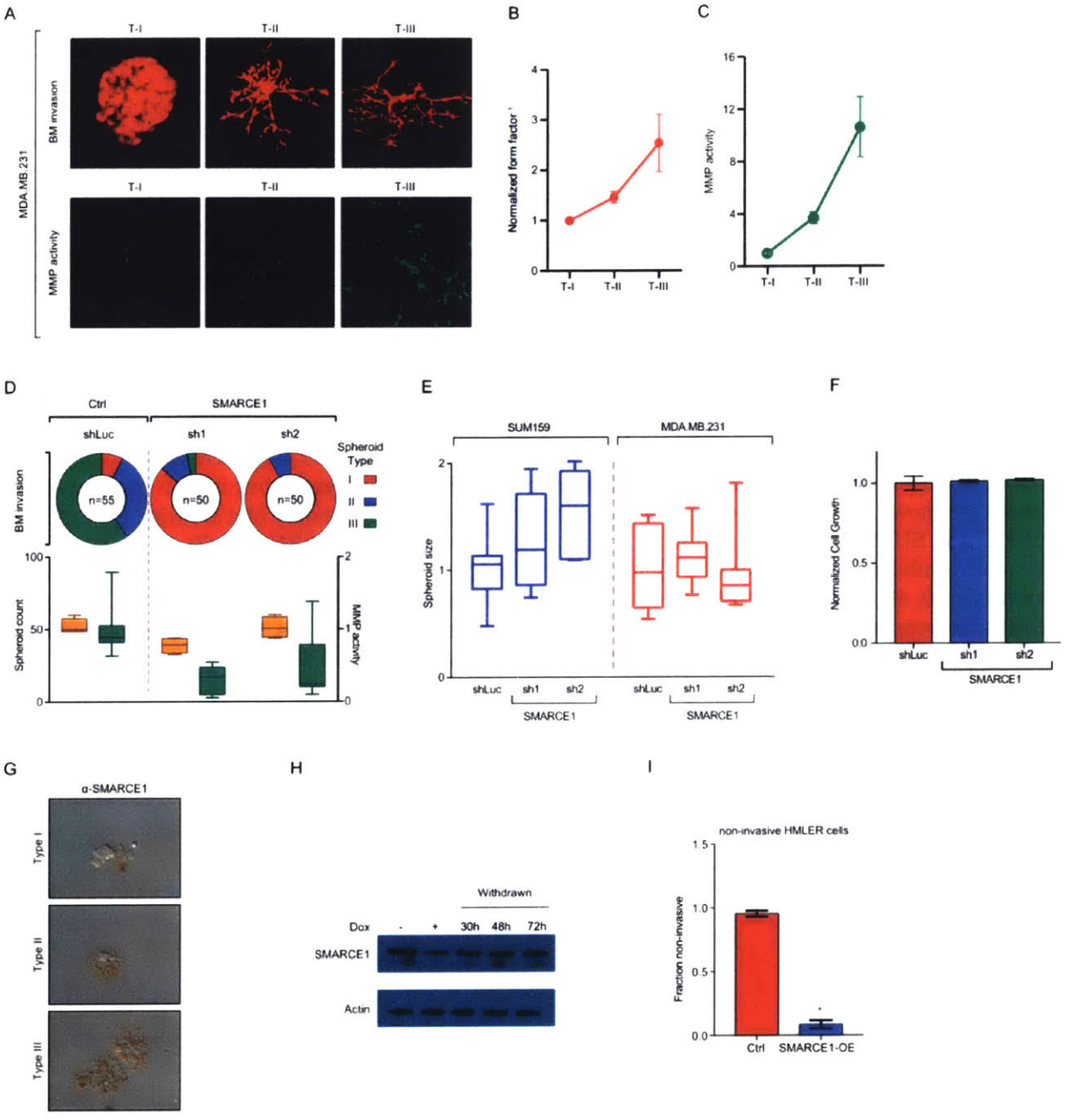

A
T-I
T-II
T-III
BM invasion
MMP activity
MDA.MB.231
B
Normalized form factor
C
MMP activity
D
Ctrl
SMARCE1
shLuc
sh1
sh2
Spheroid Type
I
II
III
n=55
n=50
n=50
BM invasion
Spheroid count
MMP activity
E
SUM159
MDA.MB.231
Spheroid size
SMARCE1
F
Normalized Cell Growth
G
α-SMARCE1
Type I
Type II
Type III
H
Withdrawn
Dox
30h
48h
72h
SMARCE1
Actin
I
non-invasive HMLER cells
Fraction non-invasive
Ctrl
SMARCE1-OE

Fig. S3. SMARCE1 inhibition blocks invasion through basement membrane. (A) Epifluorescence images of tumor spheroids formed by MDA.MB.231 cells in 3D basement membrane (BM) showing noninvasive (T-I), partially invasive (T-II), or highly invasive (T-III) characteristics. Below are epifluorescence images showing collagen hydrolysis (green signal) of type I, II, and III spheroids cultured in DQ collagen (IV)-supplemented BM cultures. (Scale bar: 100 μm.) (B) Basement membrane invasiveness of type I, II, and III structures in SUM159 cells was calculated by using image segmentation analysis and form factor calculation and normalized to T-I spheroids. (C) Quantification of MMP activity in each SUM159 spheroid type, normalized to T-I. (D) Quantification of spheroids formed by control and shSMARCE1 MDA.MB.231 cells (Upper) and quantification of spheroid counts and MMP activity of the same cells (Lower). (E) Quantification of the size of spheroids formed by control and shSMARCE1 (sh1 and sh2) SUM159 and MDA.MB.231 cells, respectively. (F) MTS proliferation assays (Celltiter Glo-based) were performed in basement membrane cultures 3 d after seeding for SUM159 control (shLuc) and SMARCE1 KD lines (sh1 and sh2). (G) SUM159 cells were cultured in basement membrane for 7 d and immunohistochemically stained for SMARCE1. Representative images of type I, II, and III colonies are shown. Note the stronger staining in invasive regions. (H) Western blot showing the expression of SMARCE1 in SUM159 cells infected with a doxorubicin (dox)- inducible shRNA. Cells were grown in the absence of doxorubicin (−), presence of 1 μg/mL doxorubicin (+), or with doxorubicin followed by doxorubicin withdrawal for 30, 48, or 72 h. (I) HMLER cells infected with control (Ctrl) or FLAG-SMARCE1 (SMARCE1-OE) constructs were seeded into polymerized collagen matrices and grown for 7 d. Shown is the quantification of the fraction of noninvasive structures from control ($n = 87$) and SMARCE1-OE ($n = 71$; $^*P < 0.05$).

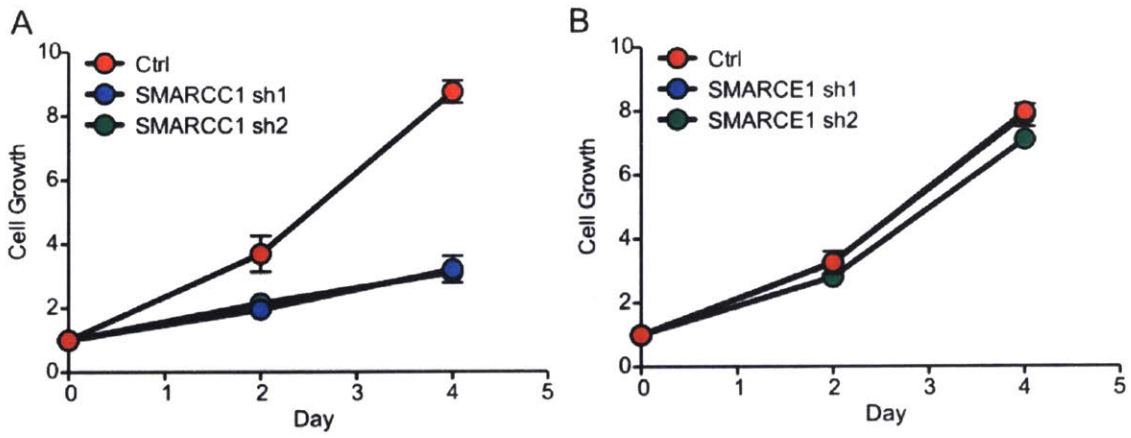


Fig. S4. SMARCE1 does not affect proliferation of breast cancer cells. SUM159 breast cancer cells expressing a control shRNA (Ctrl) or shRNAs targeting SMARCC1 (A) or SMARCE1(B) were cultured in 2D for 4 d. Cell viability was measured with CellTiter Glo and plotted for 0, 2, and 4 d.

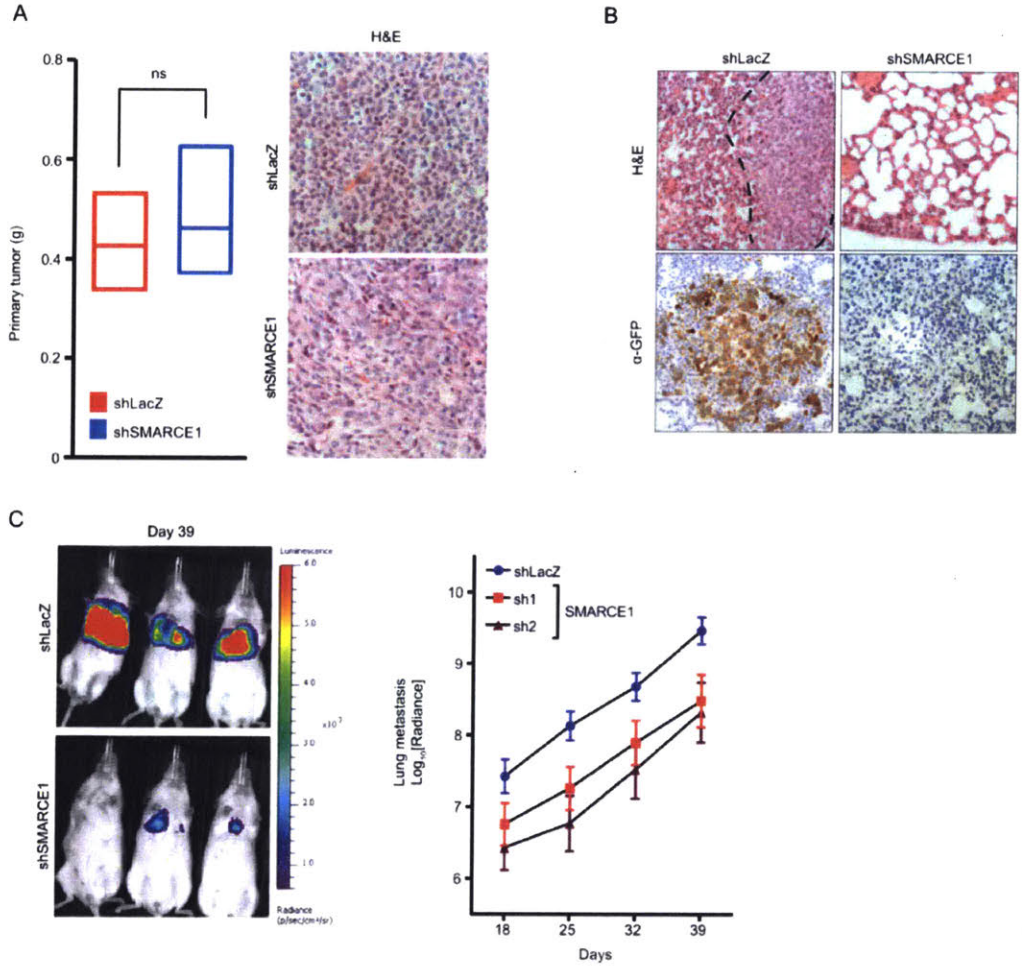
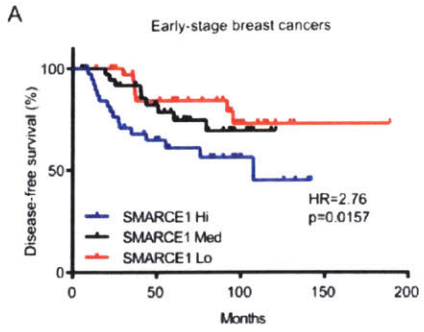
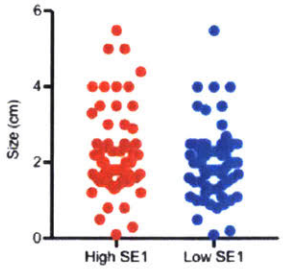


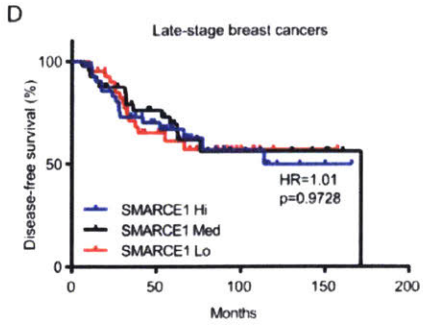
Fig. S5. The role of SMARCE1 in tumorigenesis and metastasis in vivo. (A) (Left) Quantification of primary tumor weight in mice inoculated with MDA.MB.231- LM2 cells expressing an shRNA targeting LacZ (shLacZ; n = 6) or SMARCE1 (shSMARCE1; n = 6). (Right) Representative images of tumor sections stained with H&E. (B) Lung sections from mice inoculated with shLacZ or shSMARCE1. MDA.MB.231-LM2 cells were stained with H&E and an α-GFP antibody. Dotted line indicates the border of a metastatic lesion. (C) Quantification of metastatic burden in lungs of mice injected in the tail vein with shLacZ or shSMARCE1 MDA. MB.231-LM2 cells at 18, 25, 32, and 39 d postinjection. Shown are representative luminescent images from tail vein-injected mice (shLacZ or shSMARCE1) 39 d following injection.

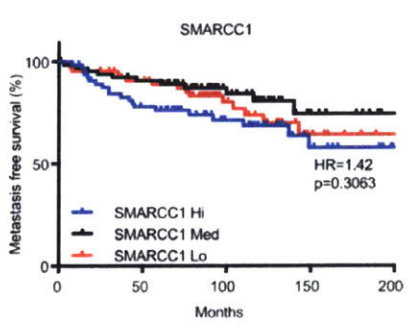


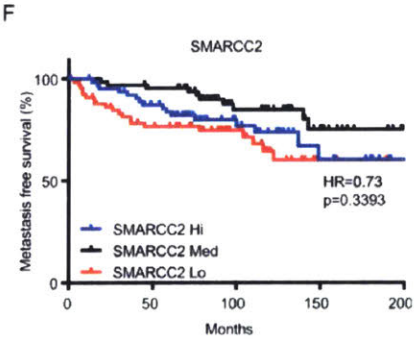


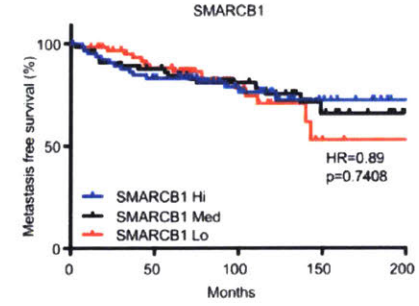
C

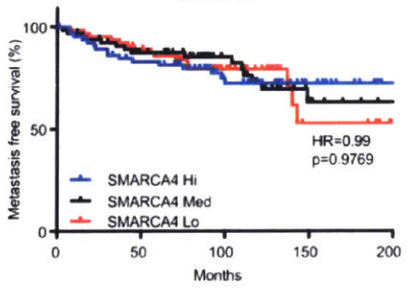
Grade	High SE1	Low SE1
1	6	9
2	50	43
3	11	15

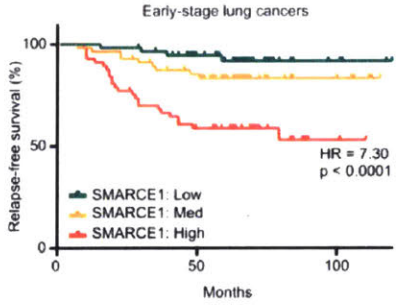


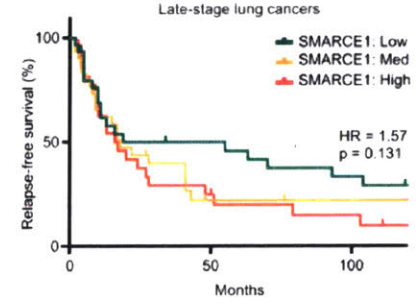


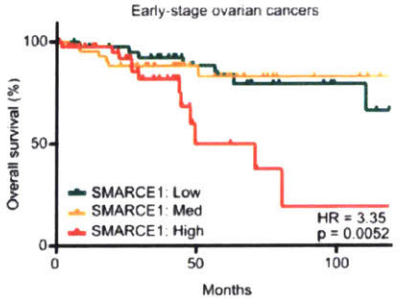












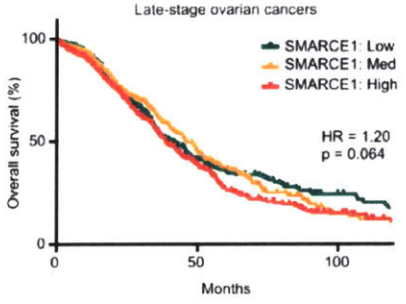


Fig. S6. SMARCE1 predicts outcome in patients with early-stage tumors. (A) A cohort of patients with early-stage breast tumors (GSE21653, node-negative patients; n = 116) were binned into SMARCE1 high (top tertile), SMARCE1 medium (middle tertile), or SMARCE1 low (bottom tertile) categories, and disease-free survival curves were plotted. The top and bottom tertiles were compared with the log-rank test to determine an HR and P value. (B and C) A cohort of patients with breast cancer (node-negative, GSE11121; n = 200) were binned into SMARCE1 high, medium, or low by tertile, and tumor size (B) and tumor grade (C) were plotted. (D) A cohort of patients with late-stage breast tumors (GSE21653, node-positive patients; n = 133) were binned into SMARCE1 high (top tertile), SMARCE1 medium (middle tertile), or SMARCE1 low (bottom tertile) categories, and disease-free survival curves were plotted. The top and bottom tertiles were compared with log-rank test to determine an HR and P value. (E–H) Early-stage breast tumors (GSE11121) were stratified based on the tertile expression of SMARCC1 (E), SMARCC2 (F), SMARCB1 (G), or SMARCA4 (H), and were plotted for metastasis-free survival. (I–L) Patients were binned into SMARCE1 high (top tertile), SMARCE1 medium (middle tertile), or SMARCE1 low (bottom tertile) categories, and survival curves were plotted. A cohort of patients with early-stage lung tumors (stage IA or IB, GSE31210; n = 168) (I) and late-stage lung tumors (stage II or above, GSE30219; n = 94) (J) were analyzed for relapse-free survival. A cohort of patients with early-stage (stage I or II; n = 133) (K) and late-stage (stage III or IV; n = 1,148) (L) ovarian cancers (KMplot ovarian cancer, 2015 version) were analyzed for overall survival. The top and bottom tertiles were compared with log-rank test to determine an HR and P value.

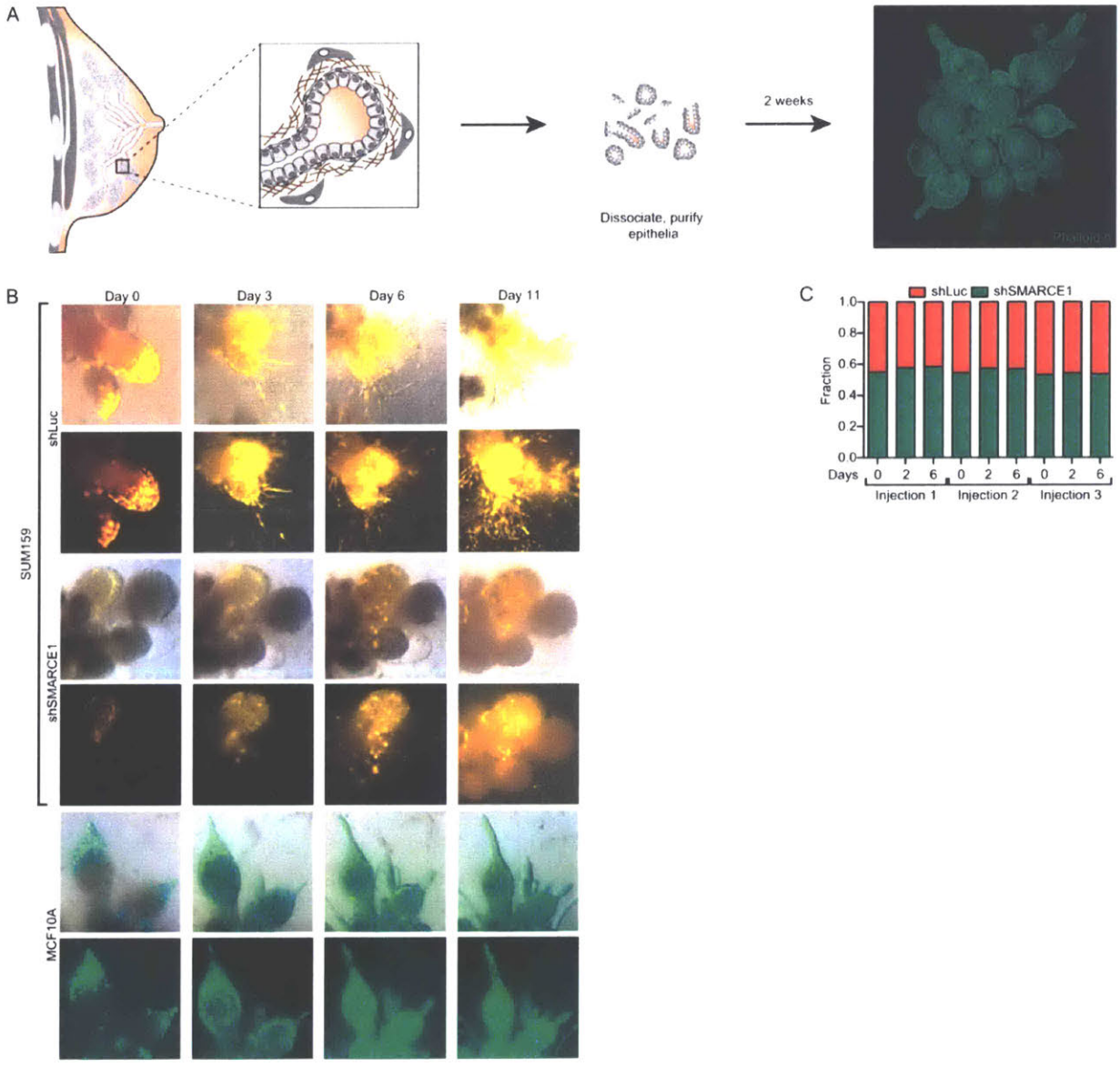


Fig. S7. SMARCE1 is required for cancer cells to escape the ductal-lobular architecture of normal mammary tissues. (A) Schematic of how primary mammary samples were processed to produce tissue outgrowths. Patient samples were mechanically and enzymatically dissociated then purified with differential pelleting and plating (Materials and Methods). Shown is a representative tissue grown in culture for 3 wk and stained for phalloidin (green, Right). (B) Images from Fig. 5B were split into a merge (bright-field and epifluorescence; Top) and epifluorescence-only image (Bottom) for each condition indicated. Representative images at 0, 3, 6, and 10/11 d after injection are shown. (C) DsRed-labeled SUM159 shLuc cells and Venus-labeled SUM159 shSMARCE1 cells were coinjected into tissues, and fluorescence intensities were measured over 6 d. Plotted is the fraction of fluorescent intensity contributed from dsRed (red) and Venus (green).

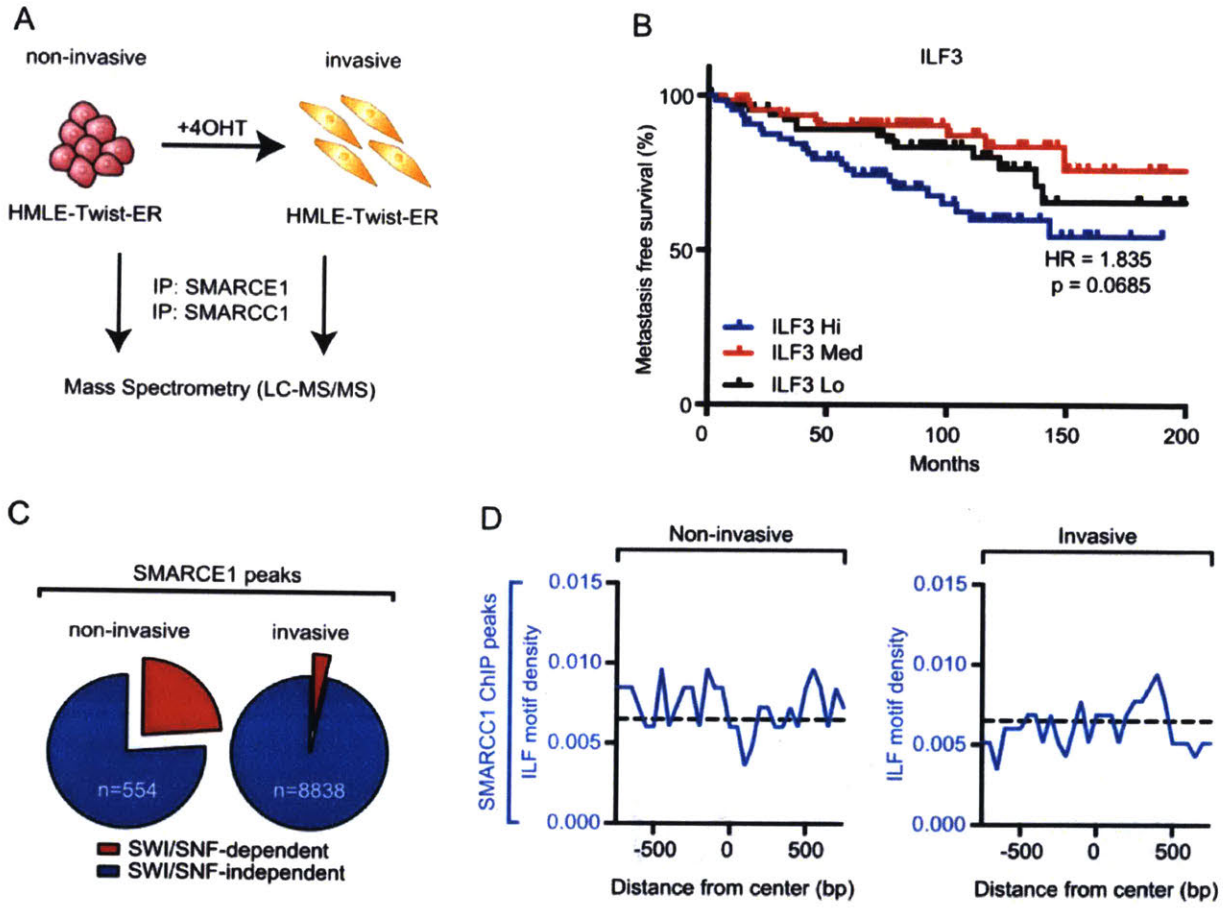


Fig. S8. SMARCE1 physically binds ILF3 and is recruited to ILF motifs in the genome. (A) Schematic representation of the MS design. The HMLE-Twist-ER cell line is noninvasive in the absence of 4OHT, but can be induced into an invasive state through the addition of 4OHT. SMARCE1 and SMARCC1 were immunoprecipitated in HMLE-Twist-ER cells grown in the presence (125 nM) or absence of 4OHT, and their coprecipitates were determined by using MS. (B) ILF3 expression was examined in a cohort of early-stage breast tumors (N stage 0, GSE11121; n = 200). Patients were binned into ILF3 high, ILF3 medium, or ILF3 low tertiles, and metastasis-free survival curves were plotted. (C) Binding of SMARCE1 and SMARCC1 in the genome of HMLE-Twist-ER cells was analyzed by ChIP-Seq. SMARCE1 and SMARCC1 peaks were determined by using MACS. Shown is the fraction of SMARCE1 peaks that overlap with SMARCC1 (red, SWI/SNF-dependent) in noninvasive (no 4OHT) cells and invasive (+4OHT) cells. Also shown is the total number of observed peaks. (D) A de novo motif for ILF was generated by using HOMER on genes that were more than twofold up-regulated in ILF2/3-overexpressing murine fibroblasts relative to genetically matched controls. Regions around SMARCC1-bound peak centers were subdivided into bins of size 50 bp. Within each bin, motif density per peak was calculated.

Chapter 3

Cancer-specific PERK signaling drives invasion and metastasis through CREB3L1

Published as: Yu-Xiong Feng*, Dexter X. Jin*, Ethan S. Sokol, Ferenc Reinhardt, Daniel H. Miller & Piyush B. Gupta (2017). Cancer-specific PERK signaling drives invasion and metastasis through CREB3L1. Nature Communications.

*These authors contributed equally to this work.

DXJ performed experiments shown in Figure 1-2, Figures 3B-E, Figures 4B and 4D-E, and Figures 5A-E and 5G, Figure S1B, Figure S1C-D, Figures S2A and S2C, Figures S3A-B, S3E and S3H, Figure S7B-C, Figures S4A-B, Figures S6A-C and S6F, and Figure S9. All authors contributed to the design of experiments, analysis of data, or editing of the manuscript.

Chapter 3.1: Abstract

PERK signaling is required for cancer invasion and there is interest in targeting this pathway for therapy. Unfortunately, chemical inhibitors of PERK's kinase activity cause on-target side effects that have precluded their further development. One strategy for resolving this difficulty would be to target downstream components of the pathway that specifically mediate PERK's pro-invasive and metastatic functions. Here we identify the transcription factor CREB3L1 as an essential mediator of PERK's pro-metastatic functions in breast cancer. CREB3L1 acts downstream of PERK, specifically in the mesenchymal subtype of triple-negative tumors, and its inhibition by genetic or pharmacological methods suppresses cancer cell invasion and metastasis. In patients with this tumor subtype, CREB3L1 expression is predictive of distant metastasis. These findings establish CREB3L1 as a key downstream mediator of PERK-driven metastasis and a druggable target for breast cancer therapy.

Chapter 3.2: Introduction

The PERK kinase plays a critical role in tumor invasion and metastasis. PERK signaling—which is activated downstream of the unfolded protein response (UPR) (1, 2) and the integrated stress response 3—enables cancer cells to survive the adverse conditions typically observed in tumor microenvironments (4,5,6). Upon its activation, PERK phosphorylates eIF2a, which reduces global protein synthesis while selectively increasing the translation of ATF4 (7). In addition to enabling cell survival, PERK–ATF4 signaling triggers multiple steps in the metastatic cascade (8), including angiogenesis (9), migration (10), survival11, and colonization at secondary organ sites12. PERK is also required for the metastatic dissemination of cancer cells that have undergone an epithelial-to-mesenchymal transition (EMT) (13).

Given its critical role in driving metastatic progression, PERK has been a focus of drug discovery programs for cancer, which have identified several small-molecule inhibitors of this kinase that reduce metastatic dissemination (13). While these molecules reduce metastatic spread, they also cause rapid onset of pancreatic atrophy, precluding their further consideration for clinical development (14). Since similar phenotypes are observed in PERK-knockout mice, PERK is likely to be essential for normal pancreatic function (15, 16). This has raised concerns about whether this key pathway is a viable target for therapy. One way to circumvent this difficulty would be to target downstream factors that specifically mediate PERK's pro-metastatic functions, but do not contribute to the pathway's functions in non-cancerous tissues. While this strategy is promising in principle, it is not currently known if the PERK pathway's pro-metastatic functions can be genetically separated from its role in normal homeostasis.

In this study, we show that this is indeed the case, and identify a key transcription factor (CREB3L1) downstream in the pathway, that specifically functions to promote metastasis in cancer cells that have activated PERK. Since CREB3L1 must be proteolytically cleaved in order to become active, this provides a unique opportunity to target this factor with small-molecule inhibitors for therapy (17). While CREB3L1's role in cancer is currently controversial—with some studies suggesting that this factor promotes metastasis, and others suggesting it suppresses metastasis (18, 19)—we resolve this discrepancy by showing that CREB3L1 specifically promotes metastasis in tumors that have activated both PERK signaling and the EMT program. Collectively, our findings establish CREB3L1 as a promising target downstream of the PERK pathway for therapeutic blockade in cancer.

Chapter 3.3: Results

Cancer-specific PERK signaling correlates with metastasis

Consistent with prior reports (13), we found that PERK is activated by phosphorylation in human breast cancers (Supplementary Fig. 1a). To identify factors downstream of PERK specifically upregulated in human breast cancers, we compared PERK pathway gene expression between a large cohort of breast cancers ($n = 1093$) and normal breast tissues ($n = 112$) (TCGA, breast cancer data set). Of the ~400 genes downstream of PERK (20) (Supplementary Data 1), only 23 showed at least a twofold increase in expression in cancers relative to normal tissues (Fig. 1a, b; CSPS, cancer-specific PERK set). Expression of these genes depended on PERK activity, since inhibition of PERK with a small-molecule inhibitor (GSK2656157 (14)) led to a significant decrease in the expression of 18 out of 23 CSPS genes (Supplementary Fig. 1b). Breast cancers that highly expressed these cancer-specific PERK genes were significantly more likely to develop distant metastases over 10 years (HR = 1.8, $p < 0.005$, log-rank test; Fig. 1c). Expression of these genes was also predictive of tumor progression in lung and gastric cancers (Fig. 1d, e).

CREB3L1 upregulates CSPS genes and is required for invasion

CREB3L1 was the only transcription factor among the cancer-specific PERK genes we identified (Fig. 1a, b). Analysis of ChIP-seq data indicated that CREB3L1 was enriched near CSPS gene loci (14/23 genes, $p < 0.01$, binomial test; Fig. 2a and Supplementary Fig. 2a); CREB3L1 was localized to active regulatory regions marked by H3K27-acetylated histones (86% of bound sites; Fig. 2b) (21). To functionally assess CREB3L1's role in regulating these genes, we used shRNAs to stably inhibit CREB3L1 expression in two invasive breast cancer cell lines, SUM159 and MDA.MB.231 (Supplementary Fig. 2b). Inhibiting CREB3L1 reduced the expression of 10 of these 14 CSPS genes (Fig. 2c and Supplementary Fig. 2c). Taken together, these data indicated that CREB3L1 directly promotes the transcription of CSPS genes.

We next used three approaches to assess if CREB3L1 was required for cancer cell invasion. First, we compared CREB3L1 expression between non-invasive and invasive breast cancer lines (22, 23), which revealed that this factor's expression was increased over 50-fold in the invasive lines (Supplementary Fig. 3a). Second, we inhibited CREB3L1 expression to test if it was required for the invasiveness of invasive cancer lines. In both the SUM159 and MDA.MB.231 lines, CREB3L1 inhibition caused a sixfold reduction in cellular invasion through basement membrane-coated transwell chambers (Fig. 2d and Supplementary Fig. 3b). In contrast, knockdown of CREB3L1 does not affect cell growth of SUM159 and MDA.MB.231 cells (Supplementary Fig. 3c, d). Lastly, we tested the requirement of CREB3L1 for invasion in 3D. When seeded as single cells into 3D cultures in basement membrane,

SUM159 and MDA.MB.231 breast cancer lines form clonal structures that can be classified as benign (non-invasive), intermediate (partially invasive), or invasive. Knockdown of CREB3L1 prevented the formation of invasive spheroids, while increasing the formation of non-invasive spheroids (Fig. 2e and Supplementary Fig. 3e).

CREB3L1 is localized to ER membranes and can migrate to the nucleus to activate transcription. Its nuclear migration is triggered via cleavage by the site 1 (S1P) and site 2 proteases (S2P), which produces an activated, truncated form of CREB3L1 (24); the truncated CREB3L1 is then free to migrate into the nucleus and activate transcription (Fig. 2f). Both the full length and truncated form of CREB3L1 were upregulated in invasive cancer cells when compared to non-invasive cells (Supplementary Fig. 3f). We found that the aforementioned decrease in invasive spheroids following knockdown of CREB3L1 is rescued by overexpression of a constitutively active form of CREB3L1 ($CREB3L1^{\Delta 375-519}$; Supplementary Fig. 3g, h). Next, we examined if CREB3L1 was sufficient to promote invasion in non-invasive cells. Overexpression of $CREB3L1^{\Delta 375-519}$ resulted in a fourfold increase in cellular invasion in a non-invasive HMLE cell line (Fig. 2g), indicating that CREB3L1's expression was sufficient to promote invasion.

CREB3L1-induced ECM deposition activates FAK

To further explore the mechanism through which CREB3L1 promotes invasion, we performed gene ontology analysis of CREB3L1-regulated CSPS genes. The analysis suggested that CREB3L1 might regulate ECM production and remodeling (Supplementary Fig. 4a; $p < 10^{-5}$, hypergeometric test), key processes associated with cancer cell invasion. Consistent with this, CREB3L1 inhibition strongly reduced activation of FAK—a kinase regulated by cell–ECM interactions and known to be important for cellular migration (Fig. 2h) (25, 26). To assess if CREB3L1 was important for migration, we seeded CREB3L1-inhibited or control cells into transwell chambers. We found that CREB3L1 inhibition caused a fivefold reduction in migratory potential (Supplementary Fig. 4b). Since CREB3L1 regulates the expression of ECM genes, e.g., *COL1A2* and *FN1* (Fig. 2c and Supplementary Fig. 2c), we hypothesized that its inhibition could be reducing FAK activity by decreasing ECM production. Consistent with this, addition of type I collagen (encoded by *COL1A1* and *COL1A2*) rescued FAK activity in the CREB3L1-inhibited cells (Supplementary Fig. 4c). If CREB3L1's pro-invasive effects were arising through ECM production, it should be possible to rescue the migration of CREB3L1-inhibited cells by supplementing them with exogenous pro-invasive ECM. In fact, cells inhibited for CREB3L1 expression were able to better heal wounds when treated with exogenous type I collagen or fibronectin (Fig. 2i). Collectively, these results suggested that CREB3L1 may promote migration and invasion by inducing ECM production.

CREB3L1 is required for breast cancer metastasis

We next assessed CREB3L1's role in vivo using the MDA.MB.231-LM2 orthotopic transplantation model of human breast cancer, which forms tumors that invade the vasculature and metastasize to the lungs. While CREB3L1 inhibition does not affect primary tumor formation or growth, as indicated by similar tumor weights and Ki67-positive proliferative indices (Supplementary Fig. 5), there was a >80-fold reduction in the number of circulating tumor cells (CTCs) (Fig. 3a). Moreover, CREB3L1 inhibition resulted in a >400-fold reduction in lung metastases, as measured both by luminescence emitted by metastatic cells in whole lungs, as well as by immunohistochemistry staining for GFP-positive cancer cells in lung sections (Fig. 3b, c). These findings indicated that CREB3L1 was required for cancer cells to enter into the circulation and form metastases.

To assess the clinical relevance of these findings, we examined CREB3L1 expression in primary breast cancers and metastatic breast cancers from patients. In comparison with primary breast cancer tissues, CREB3L1 gene expression was significantly upregulated in the metastatic growths of breast cancers that had disseminated to distant organ sites ($p < 0.01$, Student's *t*-test; Fig. 3d, GSE20565, GSE20685, GSE7904, and GSE3744). Consistent with these findings, nuclear and total CREB3L1 protein were expressed at significantly higher levels in the lymph node metastases of breast cancers, relative to primary breast tumors present on the same tissue microarray (Fig. 3e).

These findings indicated that CREB3L1 is required for metastatic dissemination in animal models of breast cancer, and is upregulated in the breast cancers of patients as they progress toward metastatic disease.

Chemical inhibition of CREB3L1 blocks metastasis

Although most transcription factors are currently not "druggable", CREB3L1's unique mechanism of activation provides an opportunity to target its activity using protease inhibitors. A chemical inhibitor of proteases, AEBSF, has been used previously to inhibit the protease (S1P) that cleaves CREB3L1 to its active form (24,27). In SUM159 cells, AEBSF treatment blocked cleavage of FLAG-CREB3L1$^{\text{full-length}}$ protein to its active form, but had no effect on the levels of FLAG-CREB3L1$^{\Delta 375-519}$protein (Fig. 4a). Consistent with this, AEBSF treatment decreased the expression of CREB3L1 target genes *COL1A1* (28) and *COL1A2* (Fig. 4b). However, AEBSF treatment did not decrease the expression of these genes in cells expressing the constitutively active FLAG-CREB3L1$^{\Delta 375-519}$(Fig. 4b). While this finding was expected, because this truncated protein does not need to be cleaved to become active, it nonetheless indicated that AEBSF's effects on *COL1A1* and *COL1A2* were mediated by CREB3L1. Inhibition of CREB3L1 activation by another S1P inhibitor, PF429242, also decreased the expression of *COL1A1* and *COL1A2*, which could be rescued by expression of FLAG-CREB3L1$^{\Delta 375-}$

^{519}(Supplementary Fig. 6a). In addition, AEBSF treatment strongly reduced FAK phosphorylation, which could be restored by expression of FLAG-CREB3L1$^{\Delta 375-519}$ (Fig. 4c). Furthermore, inhibition of CREB3L1 activation by AEBSF or PF429242 significantly negated the invasiveness of the SUM159 cells, an effect that can be rescued by overexpression of the truncated FLAG-CREB3L1$^{\Delta 375-519}$ (Fig. 4d and Supplementary Fig. 6b–d). In contrast, overexpression of the constitutively active form of two other targets of AEBSF, ATF6 and SREBP-1, was not able to rescue the decrease of cell invasion caused by AEBSF (Supplementary Fig. 6e, f). These results showed that AEBSF inhibits invasion through targeting CREB3L1.

We next assessed if treatment with AEBSF could inhibit metastasis in vivo in the orthotopic MDA.MB.231-LM2 model of human breast cancer. While AEBSF did not affect primary tumor growth (Supplementary Fig. 7a, b), it caused a 15-fold reduction in lung metastases, as measured by luminescence emitted from whole lungs (Fig. 4e). However, AEBSF was not able to inhibit the metastasis of LM2 tumors that overexpressed FLAG-CREB3L1$^{\Delta 375-519}$, indicating that its effects were mediated through CREB3L1. These observations suggest that inhibiting S1P activity to block CREB3L1 activation could be an effective therapeutic strategy for preventing metastasis.

PERK signaling requires an EMT to upregulate CREB3L1

Lastly, we investigated how this factor is regulated in invasive cancer cells. Consistent with the finding that CREB3L1 is downstream of PERK signaling (Supplementary Fig. 8), inhibition of ATF4 expression decreased CREB3L1 transcription in MDA.MB.231 and SUM159 lines (Fig. 5a). Surprisingly, overexpression of ATF4 in non-invasive lines (HMLE and T47D) is not sufficient to activate transcription of CREB3L1, suggesting that PERK signaling requires additional pathways to induce CREB3L1 (Fig. 5b). To explore these additional pathways, we segregated breast cancers that have active PERK signaling into two categories according to the expression level of CREB3L1 (CREB3L1-high and CREB3L1-low, respectively), and employed gene set enrichment analysis (GSEA) to compare these two cohorts (Fig. 5c). We found that the EMT hallmark set was the most enriched pathway in tumors that had high CREB3L1 expression relative to those with low CREB3L1 expression (Fig. 5c). This implicated EMT signaling as a potential regulator of CREB3L1 expression in the context of PERK activation. To functionally assess this, we assayed CREB3L1 expression with or without EMT induction in cells overexpressing ATF4. We found that HMLE cells induced into EMT had a greater than fivefold higher CREB3L1 expression than those that were not (Fig. 5d). Consistently, expression of CREB3L1 was highest in breast tumors that have both active PERK signaling and EMT signaling (Fig. 5e). Furthermore, expression of CREB3L1 is predictive of distant metastasis-free survival in patients with

triple-negative breast cancer of the mesenchymal subtype, which is usually enriched for pathways associated with EMT (Supplementary Fig. 9) (29).

We next further dissected which component of EMT signaling cooperates with PERK signaling to induce CREB3L1. Using the proteomic interaction database, Biogrid, we found that ATF4 might interact with Fra-1, a potent EMT transcription factor (30, 31). By performing co-immunoprecipitation, we validated that ATF4 interacts with Fra-1 in HMLE cells (Fig. 5f). Double overexpression of ATF4 and Fra-1 strongly induces the expression of CREB3L1 in non-EMT HMLE cells, while single overexpression of either gene is not sufficient to induce CREB3L1 (Fig. 5g and Supplementary Fig. 10). These results suggested that PERK signaling requires EMT to upregulate CREB3L1 through an ATF4–Fra-1 interaction.

Chapter 3.4: Discussion

The present findings establish CREB3L1 as a key downstream mediator of PERK's pro-metastatic function in cancers. CREB3L1's unique mode of activation makes it amenable—unlike the vast majority of transcription factors—to inhibition by small molecules. Thus, our findings identify a viable new avenue for therapeutically targeting the PERK pathway in cancer.

The phenotypes observed in prior studies with transgenic mice provide strong support for the notion that targeting CREB3L1 would circumvent the toxicities associated with PERK inhibitors. While CREB3L1-knockout mice only have bone defects (28), PERK-knockout mice exhibit not only these bone defects, but also reduced birth rates and develop diabetes (15, 16, 32, 33). Additionally, mice lacking ATF4—which is upstream of CREB3L1 but downstream of PERK—also have bone defects, lower birth rates, and develop anemia (34, 35, 36, 37). These additional deficiencies strongly suggest that both PERK and ATF4 have CREB3L1-independent functions. In fact, in addition to activating CREB3L1, PERK also activates NRF2—a transcriptional regulator of the cellular antioxidant response (38)—whereas ATF4 also activates the transcription of CHOP and BiP (39). These observations are in consonance with the general principle that upstream components of signaling pathways typically have more pleotropic effects when compared to downstream components.

While our results provide a proof-of-concept for inhibiting PERK-CREB3L1 signaling in vivo, the use of site 1 protease (S1P) inhibitors may raise concerns regarding drug specificity. Three additional ER-bound transcription factors, ATF6 and SREBP-1/2, are also affected by agents that inhibit S1P/S2P-mediated proteolysis (40). However, our data indicate that these factors are not required for cell invasion, precluding their role in mediating the anti-invasive effects of AEBSF in our experiments. Moreover, Nelfinavir, a S2P inhibitor used in treating HIV patients, has been well tolerated in clinical trials41, indicating that inhibition of ATF6 and SREBP-1 is not likely to be problematic in patients.

Defects in ECM remodeling are hallmarks of human diseases—including cancer and fibrosis (42, 43)—and are frequently associated with increased UPR signaling2,44. Our findings provide the first molecular explanation for how PERK might regulate ECM remodeling, namely through activation of CREBL1. Consistent with this model, S1P inhibition—which would inhibit CREB3L1—has also been shown to significantly decrease expression of ECM genes, including *COL1A1*, *COL1A2*, and *FN1* (27). This model also explains the bone defects observed in CREB3L1, ATF4, and PERK-knockout mice. The primary function of osteoblasts is to produce ECM, which spontaneously mineralizes to form bone (45). Our results would suggest that these bone defects arise because osteoblasts in CREB3L1, ATF4, and PERK-knockout mice produce insufficient ECM (36,37). In fact, a prior study has shown that expression of *COL1A1* and *COL1A2* is significantly reduced in CREB3L1-knockout mice (28)—consistent with our finding that CREB3L1 regulates these collagens in invasive cancer cells.

Our observations indicate that the signaling downstream of PERK activation is highly dependent on the differentiation state of cancer cells. Thus, while PERK induces CREB3L1 in cancer cells that have undergone an EMT, it fails to do so in cells that have not undergone an EMT. Consistent with this, although PERK is required for invasion in cancer cells that have undergone an EMT, its activation is insufficient to drive invasion in a non-EMT context (13). This suggests that CREB3L1 specifically promotes metastasis in tumors that have activated both PERK signaling and the EMT program. This observation may provide an explanation for the conflicting reports in the literature regarding CREB3L1's role in metastasis (18, 19). More generally, our findings suggest that the ability of stress signaling to promote metastasis will depend on the differentiation state of tumors.

Our observations provide a new avenue for treating the mesenchymal subtype of triple-negative breast cancers (TNBC). As a class, TNBCs afford a relatively poor prognosis, and are therefore the focus of molecular subtyping and drug development efforts (29, 46). We have found that CREB3L1 expression strongly correlates with the metastatic potential of this subtype of TNBCs. Since the mechanism through which CREB3L1 is activated can be effectively targeted by small-molecule inhibitors, our findings suggest a promising new therapeutic strategy for this clinically important subtype of breast cancer.

Chapter 3.5: Methods

Human cell lines and culture conditions

MCF7, T47D, BT474, BT549, ZR-75-30, Hs578T, MDA-MB-157, and MDA.MB.231 were obtained from the American Type Culture Collection (ATCC), and were cultured in DMEM + 10% FBS. SUM159 cells were obtained from Asterand, and were cultured in F12 + 5% FBS, insulin (10 μg/ml), and hydrocortisone (0.5 μg/ml). MDA.MB.231-luc-LM2 (LM2) cells were a kind gift from Dr Joan Massagué and cultured in DMEM + 10% FBS. HMLE cells expressing the coding sequence of Twist1 fused to the mutated estrogen receptor (HMLE-TwER) were obtained from Dr Robert A. Weinberg's lab, and maintained in a 1:1 mixture of DMEM + 10% FBS, insulin (10 μg/ml), hydrocortisone (0.5 μg/ml), EGF (10 ng/ml), and MEGM. To induce an EMT, the HMLE cells were treated with 25 nM of 4-hydroxy-tamoxifen (4-OHT) for a period of 12 days.

Lentiviral production and infection

Lentiviral particles were produced by co-transfection with 0.25 μg pCMV-VSV-G, 0.75 μg pCMV-dR8.2-dvpr, and 1 μg pLKO-shRNAs into 5×10^5 293T cells using 6 μl Fugene 6. Cell culture media of 293T cells were harvested 24, 48, and 72 h post transfection. After harvesting, lentiviral transduction was performed in the presence of protamine sulfate by spin infecting for 1 h, then incubating overnight. Following transduction, cells were selected using puromycin or blasticidin.

Plasmids

pCW-ATF4 was generated by digesting pCW57.1 (Addgene plasmid #41393) with NheI and AgeI. ATF4-3xFlag-V5 was amplified with NheI and AgeI overhangs and digested to insert into cut pCW57.1. pCW-CREB3L1$^{\Delta 375-519}$ was generated by cloning CREB3L1 without the amino acids between position 375 and the C-terminus with NheI and AgeI overhangs, digested, and inserted into pCW57.1 cut with NheI and AgeI. shATF4 and shCREB3L1 plasmids were from the Broad Institute TRC platform. pLX304-FOSL1-3xFlag-V5 was generated by digesting pLX304 (Addgene plasmid #25890) with BamHI and NheI. FOSL1-3xFlag-V5 was amplified with BamHI and NheI overhangs and digested to insert into cut pLX304. Cleaved N-terminus of ATF6 (1-373) and cleaved N-terminus of SREBP-1-c (1-490) were from Addgene (Ron Prywes lab, Addgene plasmid #27173, and Timothy Osborne lab, Addgene plasmid #26802). The pCW57.1 plasmid was a gift from David Root (Addgene plasmid #41393) and pLKO.1-blast plasmid was a gift from Keith Mostov (Addgene plasmid #26655).

Western blot and co-immunoprecipitation

Western blotting was performed according to a previous study with the following modifications (13). Cultured cells were washed twice with PBS and lysed in radioimmunoprecipitation assay buffer for 15 min on ice. Cell lysates were clarified by centrifugation at 12,000×*g* for 10 min, and protein concentration was determined by the BCA Reagent. Lysates were separated on NuPAGE 4–12% Bis-Tris gel electrophoresis, proteins were then transferred to nitrocellulose membrane and immunoblotted with indicated antibodies. All immunoblots were visualized by enhanced chemiluminescence. Antibodies used for immunoblotting were as follows: ATF4 (Cell Signaling, 11815, 1:1000), Flag (Sigma, M2 clone, 1:10,000), GAPDH (Cell Signaling, 3683, 1:2000), p-FAK (Cell Signaling, 8556 and 3281, 1:1000), Total FAK (Cell Signaling, 13009, 1:2000), and Fra-1 (Cell Signaling, 5281, 1:1000). In collagen rescue experiments, culture plates were coated with type I collagen (EMD, CC050) for 1 h at room temperature. Subsequently, 10^5 cells were seeded onto the collagen and grown for 24 h before protein harvest. In co-immunoprecipitation experiments, 5×10^7 cells were seeded 24 h prior to harvesting. Whole cell lysate was collected and incubated with 10 μl anti-Flag antibodies and 50 μl Dynabeads Protein G, or 10 μl anti-Fra-1 antibodies and 50 μl Dynabeads Protein A, respectively, for 24 h at 4 °C. Immunoprecipitants were eluted and analyzed by western blot. Raw data of all western blots were included in Supplementary Fig. 11.

In vitro transwell assay

To assay migratory and invasive potential, we used a 24-well 8-μm pore transwell plate coated without or with Matrigel, respectively (Corning Inc.). Briefly, we seeded 5×10^4 cells in the upper chamber, and incubated the cells for 8 h before removing cells on the upper surface of the chamber to count the number of cells that migrated onto the lower surface of the chamber. We use crystal violet staining (Sigma-Aldrich) to label cells, and imaged five random fields per well with a microscope at ×10. Images were then quantified using ImageJ.

In vitro wound healing migration assay

Total of 5×10^5 cells were seeded in six-well plate 12 h prior to wound cutting. A 200 μl pipette tip was used to make scratches on the single cell layer. Images of cells were taken immediately or 8 h after wound cutting. All images were analyzed using ImageJ.

3D invasion assay

Basement membrane remodeling was assessed by seeding cells in growth factor reduced Matrigel (BD Biosciences). Matrigel was thawed at 4 °C. Then 100 μl of Matrigel was added to each chamber and allowed to solidify at 37 °C with 5% CO_2 in a humidified incubator for 30 min. A total of 500 single cells

in a 200 µl suspension was seeded onto the initial layer of Matrigel and allowed to settle for 20 min. An aliquot of 200 µl of media supplemented with 10% Matrigel was gently layered on top. An aliquot of 200 µl of media was replaced every 4 days. Cultures were fixed and imaged 7 days after growth. Spheres were binned into three categories based on invasiveness: benign, intermediate, and invasive. Benign spheres were defined as structures with no visible cellular protrusions, intermediate spheres were defined as structures with protrusions extending from a central colony, and invasive structures were defined as fully scattered colonies.

Spontaneous metastasis assay

NOD/SCID mice were purchased from Jackson Labs. All mouse procedures were approved by the Animal Care and Use Committees of the Massachusetts Institute of Technology. For the spontaneous metastasis assay, 1×10^6 MDA.MD.231-LM2 cells were suspended in 50 µl 1:1 mix of Matrigel and DMEM, and injected into a mammary pad of each mouse. Animals were randomized by weight before treatment. After 7 weeks, mice were weighed. Tumors and lungs were harvested 7 weeks after injection. To measure metastases, freshly collected lungs were soaked in D-Luciferin (150 µg/ml; PerkinElmer) for 15 min and imaged with the IVIS system (PerkinElmer) at non-saturating exposures. Radiance was quantified with LivingImage v4.4 software. Tumors were weighed after collection.

Drug treatment

In the chemical treatment experiment, 2 weeks after tumor cells implantation, 200 µl pure PBS or PBS with 1 mg of AEBSF was administrated daily through i.p. injection into the animals for a period of 4 weeks.

Immunohistochemistry and human tissue microarray

All immunohistochemistry was performed at the Koch Institute Histology Core using the Thermo Scientific IHC Autostainer 360 (Thermo). Tumors obtained from the spontaneous metastasis assay were fixed in 10% neutral-buffered formalin, and then embedded in paraffin. Paraffin-embedded tumors were sectioned at 5 µm for histological analyses. Immunohistochemistry for Ki67 and GFP was performed. Stained sections were imaged and quantified. Briefly, two fields were scored per animal, resulting in a total of 10 fields per group. On each cross-section, ~5% of the total area was scored for GFP positivity using ImageJ. Scores were normalized to the mean of the control group to calculate relative lung metastasis. Tissue microarrays were ordered from Biomax (BR10010d). Immunohistochemistry for CREB3L1 and phosphorylated PERK (pPERK) was performed. CREB3L1 and pPERK staining was independently quantified in a blinded manner. Antibodies used in IHC were: GFP (Cell Signaling, 2956,

1:100), CREB3L1 (R&D systems, AF4080, 1:100), Ki67 (Cell Signaling, 9449, 1:250), and pPERK (PIERCE CHEMICAL MS, PA537773, 1:100).

Quantification of circulating tumor cells

A total of 1×10^6 luciferase-labeled MDA.MB.231-LM2 cells was inoculated into the fat pad of NOD/SCID mice. Seven weeks post tumor implantation, 400 μl of peripheral blood samples was collected. Blood samples were centrifuged at 1000×*g* to precipitate the cell fractions. An aliquot of 1 ml of red blood cell lysis buffer (Stemcell Technology) was applied to resuspend the pellet on ice for 5 min. The samples were then centrifuged again at 2000 rpm. An aliquot of 200 μl PBS containing luciferin was used to resuspend the pellet, and luciferase activity of all samples were immediately measured using the IVIS imaging system. To quantify the number of luciferase-positive cells, different numbers of MDA.MB.231-LM2 cells, 1000, 333, 111, 37, and 12, were mixed with 400 μl of blood samples from tumor-free NOD/SCID mouse, respectively, and followed by the same purification and measurement processes mentioned above. A linear relationship between luciferase activity and cell number was then generated by linear regression, and the number of luciferase-positive circulating tumor cells from tumor-bearing mice were then calculated.

GSEA and gene ontology analysis

We defined the PERK gene set as the top 400 genes downregulated in an ATF4 knockout MEFs relative to wild-type MEFs following thapsigargin treatment (GSE35681). Gene symbols were converted to human genes that overlapped with those in the TCGA data set, which left 388 remaining symbols. Gene ontology enrichment analysis for cellular components was performed using the molecular signature database. Gene set enrichment analysis was performed using the GSEA software developed by the Broad Institute. $PERK^{high}$ patients were defined as those whose scores for the PERK gene set ranked in the top half. Subsequently, these $PERK^{high}$ patients were ranked based on CREB3L1 expression. For each gene, expression from the top half ($CREB3L1^{high}$) of $PERK^{high}$ tumors were compared to those from the bottom half ($CREB3L1^{low}$) of $PERK^{high}$ tumors. Gene set enrichment analysis was performed using hallmark gene sets provided by the molecular signature database on the ranked list of genes produced by comparing $CREB3L1^{high}$ to $CREB3L1^{low}$.

Chromatin immunoprecipitation

Chromatin immunoprecipitation was previously performed by the ENCODE Project using a CREB3L1 antibody (Sigma-Aldrich, HPA024069) as described (21). Library was downloaded from ascension ENCLB300HOB. Reads were aligned to the human genome build GRCh38 with up to one

mismatch. Chip-seq peaks were identified using MACS2 with a q value threshold of 0.01. q value was calculated using the Benjamini-Hochberg procedure. Each peak was annotated with the closest associated gene using HOMER. CSPS bound genes were considered overlapping with H3K27ac if they directly overlapped an H3K27ac peak. Distribution, as a fold change of CREB3L1 signal relative to input signal, and peaks were displayed using UCSC genome browser for hg38.

Primary tumor and metastasis gene expression analyses

Normalized RNA-seq data were obtained from https://gdac.broadinstitute.org/. Data were log transformed after adding a pseudocount. Gene expression for PERK signature genes was binned into normal samples or tumor samples and was plotted as a heat map using GENE-E. For each gene, a p value and the difference in log transformed expression between cancer and normal was calculated.
For CREB3L1 expression across PERK and EMT bins, TCGA tumors were first binned into PERK-high or PERK-low bins based on if they were in the top 33 percentile or the bottom 33 percentile, respectively. For each PERK bin, EMT signaling was categorized as high or low if it was in the top 33 percentile of the bin or the bottom 33 percentile of the bin. For each tumor in each respective bin, CREB3L1 expression was plotted.

For primary tumor vs. metastasis comparisons, GSE20565, GSE20685, GSE7904, and GSE37444 were downloaded and quantile normalized. CREB3L1 expression across metastatic sites was compared to expression in primary tumors.

Survival analysis

Clinical survival analyses were performed through an online tool, KMplotter (kmplot.com) (47). The information of data sets used was included in Fig. 1. To calculate a score for the signatures, average gene expression for each signature was determined. Patients were binned into those in the top tertile and those in the bottom tertile for the signature score. Data were plotted as a Kaplan–Meier curve.

Statistics

All statistical analyses were performed using Graphpad Prism software. All data unless otherwise specified are presented as mean ± s.e.m. A two-tailed Student's t-test was performed for comparisons between two groups of data. Significance of gene ontological enrichment was calculated using the hypergeometric test. Statistical significance of Kaplan–Meier survival analysis was calculated using the log-rank test. Significance of CREB3L1 binding to the CSPS genes was calculated using the binomial test. For each gene in tumor vs. normal comparisons, $p$values were calculated using a Student's t-test, where only those that surpassed the Bonferroni-corrected threshold were considered significant.

Data availability

The data that support the findings of this study are available from the corresponding authors on reasonable request.

Chapter 3.6: Acknowledgements

We thank Dr David Pincus for the helpful discussions, and Dr Yuting Liu for the help on microarray data analyses. This research was supported by grants from the Richard and Susan Smith Family Foundation and the Breast Cancer Alliance (to P.B.G.), the Ludwig cancer research postdoctoral fellowship (to Y.-X.F.), and the National Science Foundation Graduate Research Fellowship Program (1122374; to E.S.S.).

Chapter 3.7: References

1. Walter, P. & Ron, D. The unfolded protein response: from stress pathway to homeostatic regulation. Science 334, 1081–1086 (2011).
2. Wang, M. & Kaufman, R. J. The impact of the endoplasmic reticulum proteinfolding environment on cancer development. Nat. Rev. Cancer 14, 581–597 (2014).
3. Wek, R. C., Jiang, H. Y. & Anthony, T. G. Coping with stress: eIF2 kinases and translational control. Biochem. Soc. Trans. 34, 7–11 (2006).
4. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. Nat. Med. 19, 1423–1437 (2013).
5. Caino, M. C. et al. Metabolic stress regulates cytoskeletal dynamics and metastasis of cancer cells. J. Clin. Invest. 123, 2907–2920 (2013).
6. Harding, H. P., Zhang, Y., Bertolotti, A., Zeng, H. & Ron, D. Perk is essential for translational regulation and cell survival during the unfolded protein response. Mol. Cell 5, 897–904 (2000).
7. Harding, H. P. et al. Regulated translation initiation controls stress-induced gene expression in mammalian cells. Mol. Cell 6, 1099–1108 (2000).
8. Bu, Y. & Diehl, J. A. PERK integrates oncogenic signaling and cell survival during cancer development. J. Cell. Physiol. 231, 2088–2096 (2016).
9. Blais, J. D. et al. Perk-dependent translational regulation promotes tumor cell adaptation and angiogenesis in response to hypoxic stress. Mol. Cell Biol. 26, 9517–9532 (2006).
10. Nagelkerke, A. et al. Hypoxia stimulates migration of breast cancer cells via the PERK/ATF4/LAMP3-arm of the unfolded protein response. Breast Cancer Res. 15, R2 (2013).
11. Dey, S. et al. ATF4-dependent induction of heme oxygenase 1 prevents anoikis and promotes metastasis. J. Clin. Invest. 125, 2592–2608 (2015).
12. Bobrovnikova-Marjon, E. et al. PERK promotes cancer cell proliferation and tumor growth by limiting oxidative DNA damage. Oncogene 29, 3881–3895 (2010).
13. Feng, Y. X. et al. Epithelial-to-mesenchymal transition activates PERKeIF2alpha and sensitizes cells to endoplasmic reticulum stress. Cancer Discov. 4, 702–715 (2014).
14. Atkins, C. et al. Characterization of a novel PERK kinase inhibitor with antitumor and antiangiogenic activity. Cancer Res. 73, 1993–2002 (2013).
15. Harding, H. P. et al. Diabetes mellitus and exocrine pancreatic dysfunction in perk−/− mice reveals a role for translational control in secretory cell survival. Mol. Cell 7, 1153–1163 (2001).
16. Zhang, W. et al. PERK EIF2AK3 control of pancreatic beta cell differentiation and proliferation is required for postnatal glucose homeostasis. Cell Metab. 4, 491–497 (2006).

17. Brown, M. S., Ye, J., Rawson, R. B. & Goldstein, J. L. Regulated intramembrane proteolysis: a control mechanism conserved from bacteria to humans. Cell 100, 391–398 (2000).
18. Vellanki, R. N., Zhang, L. & Volchuk, A. OASIS/CREB3L1 is induced by endoplasmic reticulum stress in human glioma cell lines and contributes to the unfolded protein response, extracellular matrix production and cell migration. PLoS ONE 8, e54060 (2013).
19. Mellor, P. et al. CREB3L1 is a metastasis suppressor that represses expression of genes regulating metastasis, invasion, and angiogenesis. Mol. Cell Biol. 33, 4985–4995 (2013).
20. Han, J. et al. ER-stress-induced transcriptional regulation increases protein synthesis leading to cell death. Nat. Cell Biol. 15, 481–490 (2013).
21. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57–74 (2012).
22. Keller, P. J. et al. Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. Breast Cancer Res. 12, R87 (2010).
23. Sarrio, D. et al. Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. Cancer Res. 68, 989–997 (2008).
24. Murakami, T. et al. Cleavage of the membrane-bound transcription factor OASIS in response to endoplasmic reticulum stress. J. Neurochem. 96, 1090–1100 (2006).
25. Michael, K. E., Dumbauld, D. W., Burns, K. L., Hanks, S. K. & Garcia, A. J. Focal adhesion kinase modulates cell adhesion strengthening via integrin activation. Mol. Biol. Cell 20, 2508–2519 (2009).
26. Schlaepfer, D. D. & Mitra, S. K. Multiple connections link FAK to cell motility and invasion. Curr. Opin. Genet. Dev. 14, 92–101 (2004).
27. Gorski, J. P. et al. Inhibition of proprotein convertase SKI-1 blocks transcription of key extracellular matrix genes regulating osteoblastic mineralization. J. Biol. Chem. 286, 1836–1849 (2011).
28. Murakami, T. et al. Signalling mediated by the endoplasmic reticulum stress transducer OASIS is involved in bone formation. Nat. Cell Biol. 11, 1205–1211 (2009).
29. Lehmann, B. D. et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. J. Clin. Invest. 121, 2750–2767 (2011).
30. Reinke, A. W., Baek, J., Ashenberg, O. & Keating, A. E. Networks of bZIP protein-protein interactions diversified over a billion years of evolution. Science 340, 730–734 (2013).
31. Tam, W. L. et al. Protein kinase C alpha is a central signaling node and therapeutic target for breast cancer stem cells. Cancer Cell. 24, 347–364 (2013).

32. Delepine, M. et al. EIF2AK3, encoding translation initiation factor 2-alpha kinase 3, is mutated in patients with Wolcott-Rallison syndrome. Nat. Genet. 25, 406–409 (2000).
33. Zhang, P. et al. The PERK eukaryotic initiation factor 2 alpha kinase is required for the development of the skeletal system, postnatal growth, and the function and viability of the pancreas. Mol. Cell Biol. 22, 3864–3874 (2002).
34. Masuoka, H. C. & Townes, T. M. Targeted disruption of the activating transcription factor 4 gene results in severe fetal anemia in mice. Blood 99, 736–745 (2002).
35. Yang, X. & Karsenty, G. ATF4, the osteoblast accumulation of which is determined post-translationally, can induce osteoblast-specific gene expression in non-osteoblastic cells. J. Biol. Chem. 279, 47109–47114 (2004).
36. Yang, X. et al. ATF4 is a substrate of RSK2 and an essential regulator of osteoblast biology; implication for Coffin-Lowry syndrome. Cell 117, 387–398 (2004).
37. Wei, J., Sheng, X., Feng, D., McGrath, B. & Cavener, D. R. PERK is essential for neonatal skeletal development to regulate osteoblast proliferation and differentiation. J. Cell. Physiol. 217, 693–707 (2008).
38. Cullinan, S. B. & Diehl, J. A. Coordination of ER and oxidative stress signaling: the PERK/Nrf2 signaling pathway. Int. J. Biochem. Cell Biol. 38, 317–332 (2006).
39. Luo, S., Baumeister, P., Yang, S., Abcouwer, S. F. & Lee, A. S. Induction of Grp78/BiP by translational block: activation of the Grp78 promoter by ATF4 through and upstream ATF/CRE site independent of the endoplasmic reticulum stress elements. J. Biol. Chem. 278, 37375–37385 (2003).
40. Okada, T. et al. A serine protease inhibitor prevents endoplasmic reticulum stress-induced cleavage but not transport of the membrane-bound transcription factor ATF6. J. Biol. Chem. 278, 31024–31032 (2003).
41. Rengan, R. et al. A phase I trial of the HIV protease inhibitor nelfinavir with concurrent chemoradiotherapy for unresectable stage IIIA/IIIB non-small cell lung cancer: a report of toxicities and clinical response. J. Thorac. Oncol. 7, 709–715 (2012).
42. Hynes, R. O. & Naba, A. Overview of the matrisome--an inventory of extracellular matrix constituents and functions. Cold Spring Harb. Perspect. Biol. 4, a004903 (2012).
43. Pickup, M. W., Mouw, J. K. & Weaver, V. M. The extracellular matrix modulates the hallmarks of cancer. EMBO Rep. 15, 1243–1253 (2014).
44. Wang, M. & Kaufman, R. J. Protein misfolding in the endoplasmic reticulum as a conduit to human disease. Nature 529, 326–335 (2016).

45. Shi, S., Kirk, M. & Kahn, A. J. The role of type I collagen in the regulation of the osteoblast phenotype. J. Bone Miner. Res. 11, 1139–1145 (1996).
46. Foulkes, W. D., Smith, I. E. & Reis-Filho, J. S. Triple-negative breast cancer. N. Engl. J. Med. 363, 1938–1948 (2010).
47. Szasz, A. M. et al. Cross-validation of survival associated biomarkers in gastric cancer using transcriptomic data of 1,065 patients. Oncotarget 7, 49322–49333 (2016).

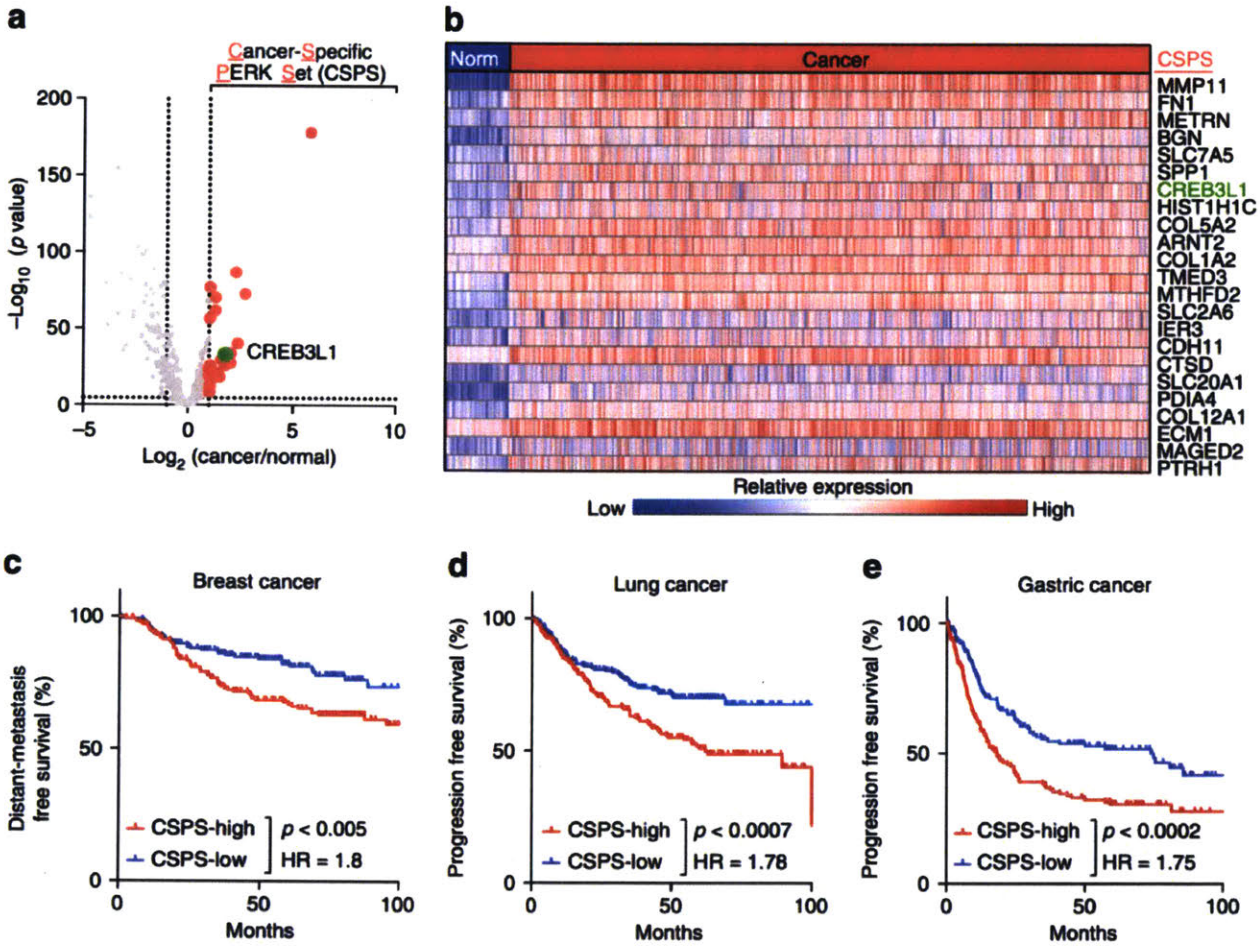


Fig. 1 Cancer-specific PERK signaling correlates with progression and metastasis. **a** A volcano plot comparing the expression of 388 PERK pathway genes between breast cancer tissues and normal breast tissues. Plotted for each gene are the negative $\log_{10}$ of the p value and the $\log_2$ of the fold change of gene expression of cancer samples relative to normal samples. A set of 23 genes that are twofold overexpressed in cancer and significant after Bonferroni correction were defined as the cancer-specific PERK set (CSPS; red dots). **b** A heat map showing the expression of the CSPS genes across 112 normal breast samples and 1093 breast tumors. **c–e** Average expression of CSPS genes was used to examine survival in breast, lung, and gastric tumor data sets. Patients were binned into CSPS high (top tertile) and CSPS low (bottom tertile), and survival curves were plotted. A cohort of **c** breast cancers was analyzed for metastasis-free survival (GSE17907, GSE9195, GSE20685, GSE16446, and GSE19615, $n = 182$ for each arm). A cohort of **d** lung cancers and **e** gastric cancers were analyzed for progression-free survival (lung cancer: GSE8894, GSE50081, GSE29013, and GSE31210, $n = 195$ for each arm; gastric cancer: GSE22377, GSE15459, and GSE62254, $n = 167$ for each arm). The indicated p-values were calculated using the log-rank test

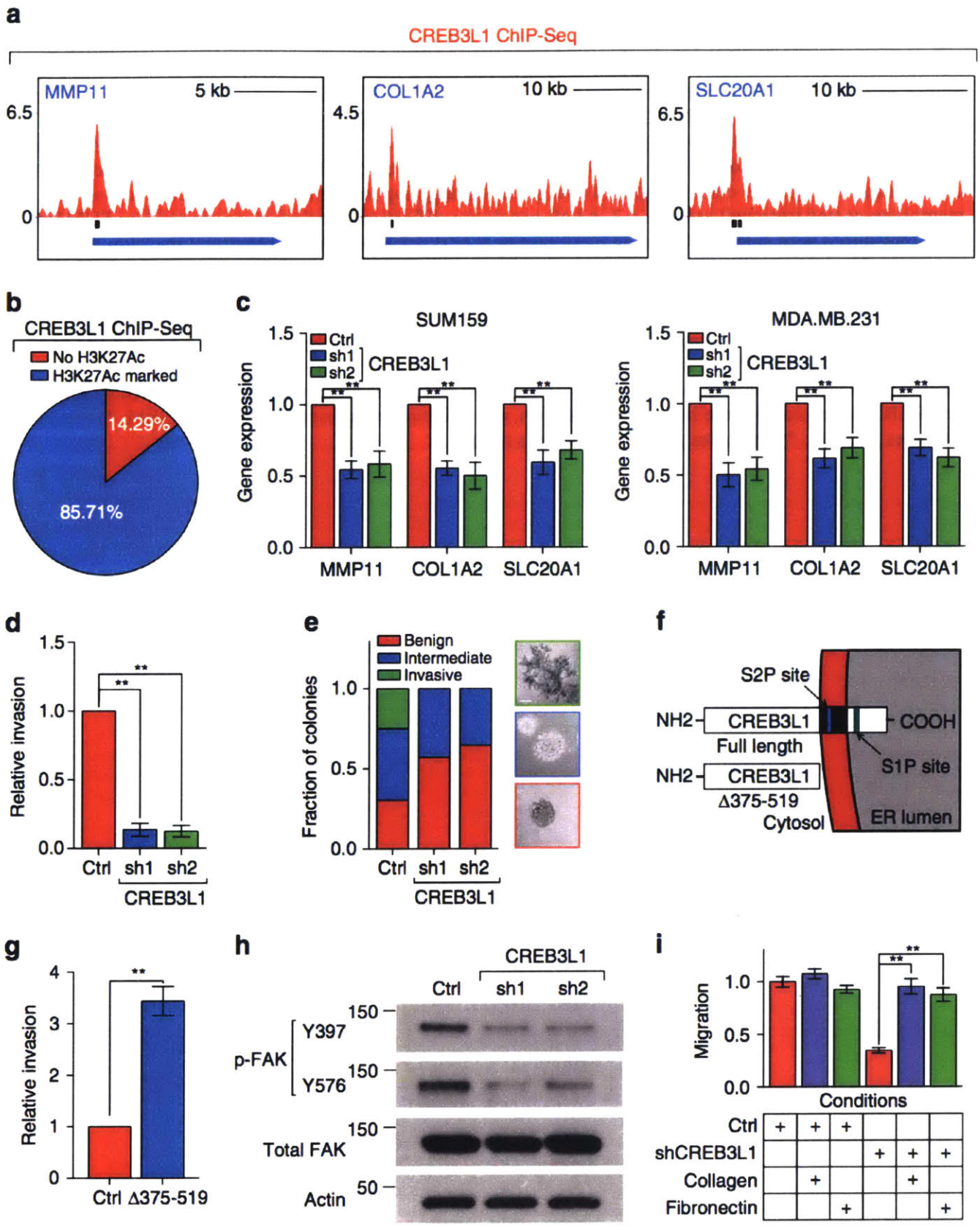

a
CREB3L1 ChIP-Seq
MMP11
5 kb
6.5
0
COL1A2
10 kb
4.5
0
SLC20A1
10 kb
6.5
0
b
CREB3L1 ChIP-Seq
No H3K27Ac
H3K27Ac marked
14.29%
85.71%
c
SUM159
MDA.MB.231
Ctrl
sh1
sh2
CREB3L1
Gene expression
1.5
1.0
0.5
0.0
**
MMP11
COL1A2
SLC20A1
d
Relative invasion
1.5
1.0
0.5
0.0
**
Ctrl
sh1
sh2
CREB3L1
e
Benign
Intermediate
Invasive
Fraction of colonies
1.0
0.5
0.0
Ctrl
sh1
sh2
CREB3L1
f
S2P site
NH2
CREB3L1
COOH
Full length
S1P site
NH2
CREB3L1
Δ375-519
Cytosol
ER lumen
g
Relative invasion
4
3
2
1
0
**
Ctrl
Δ375-519
h
CREB3L1
Ctrl
sh1
sh2
p-FAK
Y397
Y576
Total FAK
Actin
150
150
150
50
i
Migration
1.0
0.5
0.0
**
Conditions
Ctrl
shCREB3L1
Collagen
Fibronectin

Fig. 2 CREB3L1 upregulates cancer-specific PERK genes and is required for invasion. **a** Distribution of CREB3L1 ChIP-seq fold-enrichment over input control, along with called-peaks, at the locus of CSPS genes. Shown were the CREB3L1 peaks for three CSPS genes: MMP11, COL1A2, and SLC20A1. **b** A pie chart showing the fraction of CREB3L1 ChIP-Seq peaks from the CSPS genes that were co-localized with acetylated H3K27. **c** qPCR showing the relative expression of the CSPS genes shown in **a** in SUM159 and MDA.MB.231 cells transduced with shRNAs targeting luciferase or CREB3L1, $n = 4$. **d** Quantification of the cell invasion of SUM159 cells transduced with shRNAs targeting luciferase or CREB3L1 in a basement membrane-coated transwell assay, $n = 3$. **e** Left panel: quantification of colony types formed by SUM159 cells transduced with shRNAs targeting luciferase or CREB3L1 in a 7-day 3D Matrigel invasion assay, $n = 50$; right panel: representative images of benign, intermediate, and invasive colonies (scale bar: 80 μm). **f** Schematic of full-length CREB3L1 and CREB3L1$^{\Delta 375\text{-}519}$. **g**Quantification of cell invasion of HMLE cells transduced with control plasmid or a constitutive active form of CREB3L1 (CREB3L1$^{\Delta 375\text{–}519}$) in a basement membrane-coated transwell assay, $n = 3$. **h** Western blotting of phospho-FAK (pY397 and pY576), total FAK, and Actin in SUM159 cells transduced with shRNAs targeting luciferase or CREB3L1. **i**Quantification of cell migration of SUM159 cells transduced with shRNAs targeting luciferase or CREB3L1 treated with the indicated conditions, $n = 8$. Data are represented as mean ± SEM. * indicates $p < 0.05$; ** indicates $p < 0.01$ (Student's *t*-test)

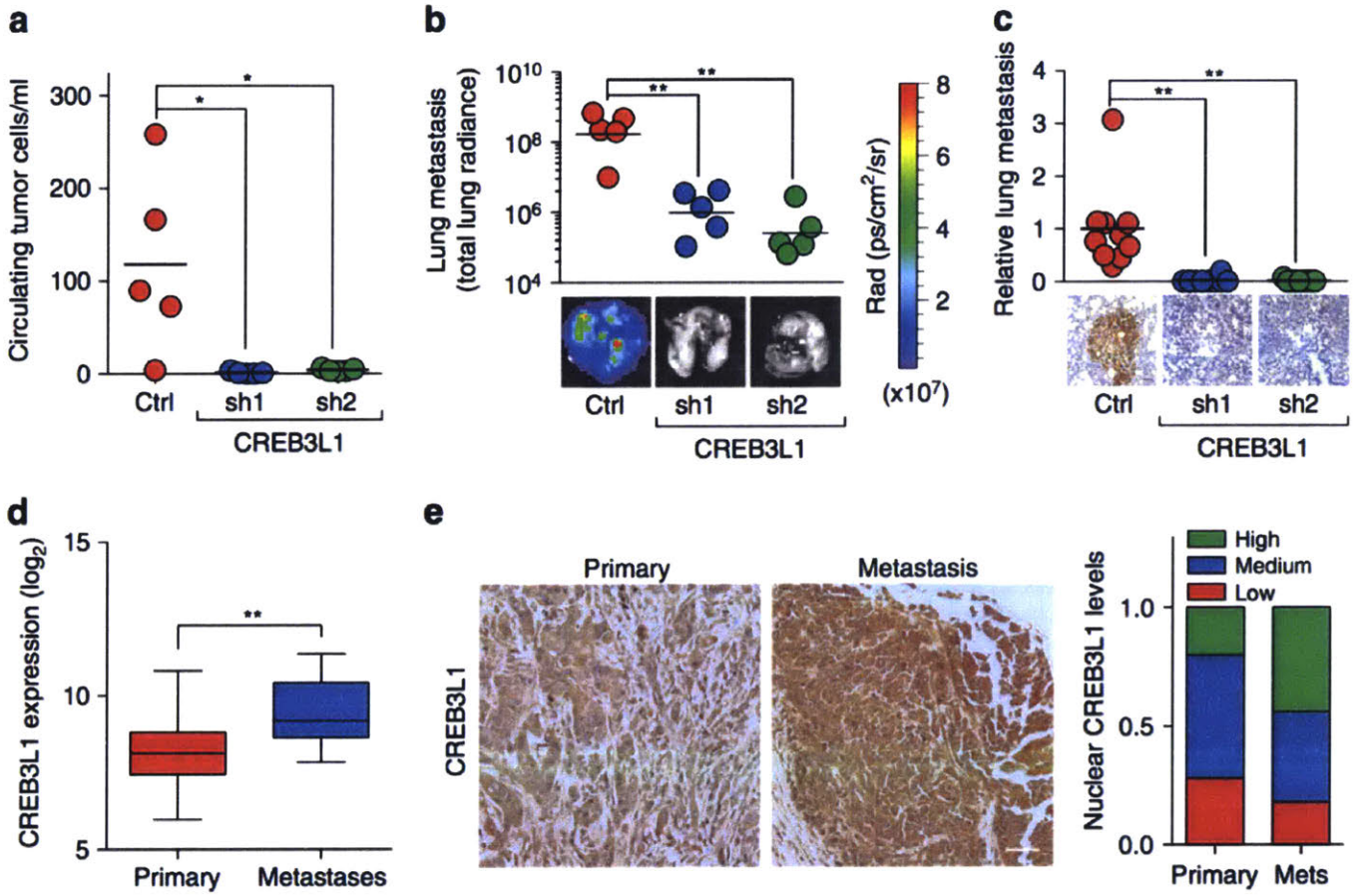


Fig. 3 CREB3L1 is required for metastasis and is upregulated in the metastases of breast cancer patients. **a** A quantification of circulating tumor cells (CTCs) per ml of blood from NOD/SCID mice 7 weeks after being transplanted with MDA.MB.231-Luc-LM2 cells transduced with shRNAs targeting LacZ or CREB3L1, $n = 5$. **b** Lung metastases of the animals in **a** were quantified (top) by radiance of the whole lung in each animal, $n = 5$. Total radiance was also imaged (bottom). **c** Lung metastases of the animals in **a** was gauged by immunohistochemistry staining for GFP-positive cancer cells in lung sections. GFP area per field was quantified (top) and imaged (bottom), $n = 10,10,9$ for Ctrl, sh1, and sh2, respectively (scale bar: 40 μm). **d** Expression of CREB3L1 in 529 primary breast cancer tissues and 45 breast cancer metastases (GSE20565, GSE20685, GSE7904, and GSE3744) shown as a box-and-whiskers plot. **e** Immunohistochemical staining for CREB3L1 in a tissue microarray containing primary breast cancer tissues (left, $n = 50$) and breast cancer metastases in lymph nodes (middle, $n = 50$). Nuclear CREB3L1 levels were quantified (right). Scale bar: 40 μm. Data are represented as mean ± SEM or the mean alone. * indicates $p < 0.05$; ** indicates $p < 0.01$ (Student's t-test)

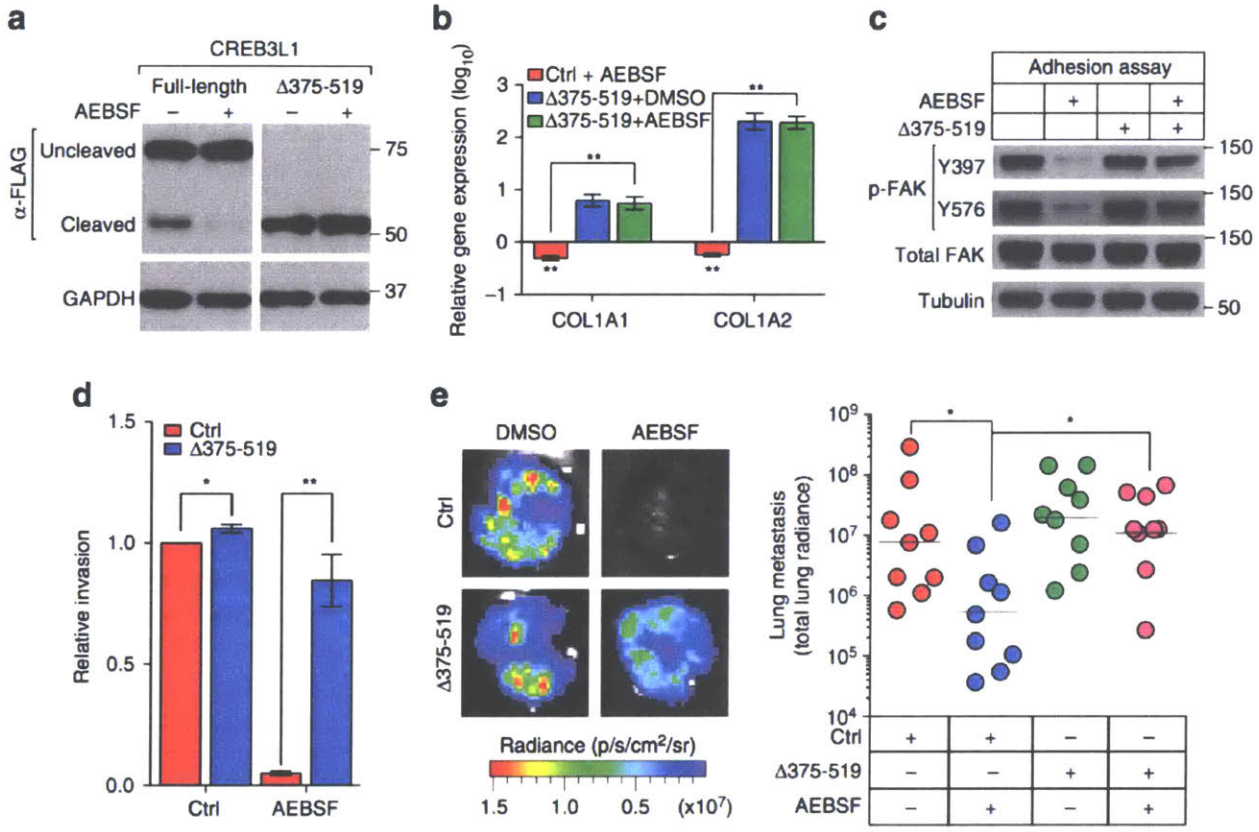


Fig. 4 Chemical inhibition of CREB3L1 blocks metastasis. **a** A western blot of CREB3L1 expression in full-length CREB3L1 or CREB3L1$^{\Delta 375-519}$ transduced SUM159 cells that were treated with solvent control or a S1P inhibitor, AEBSF. **b** SUM159 cells transduced with control or CREB3L1$^{\Delta 375-519}$ were treated with solvent control or AEBSF. qPCR analyses were performed to quantify the expression of ECM genes: *COL1A1* and *COL1A2*. The gene expression was normalized to the solvent-treated control transduction samples, $n = 3$. **c** Western blot of phosphorylated FAK (pY397 and pY576), total FAK, and Tubulin in cells of **b**. **d** Quantification of cell invasion of cells from **b** in a basement membrane-coated transwell assay, $n = 3$. **e** Two weeks after orthotopical transplantation with control or CREB3L1$^{\Delta 375-519}$ transduced MDA.MB.231-Luc-LM2 cells, NOD/SCID mice were treated with solvent control or AEBSF for another 4 weeks. Lung metastases of the animals were gauged by radiance from the whole lung in each animal, $n = 9$. Data are represented as mean ± SEM or geometric mean alone. * indicates $p < 0.05$; ** indicates $p < 0.01$ (Student's *t*-test)

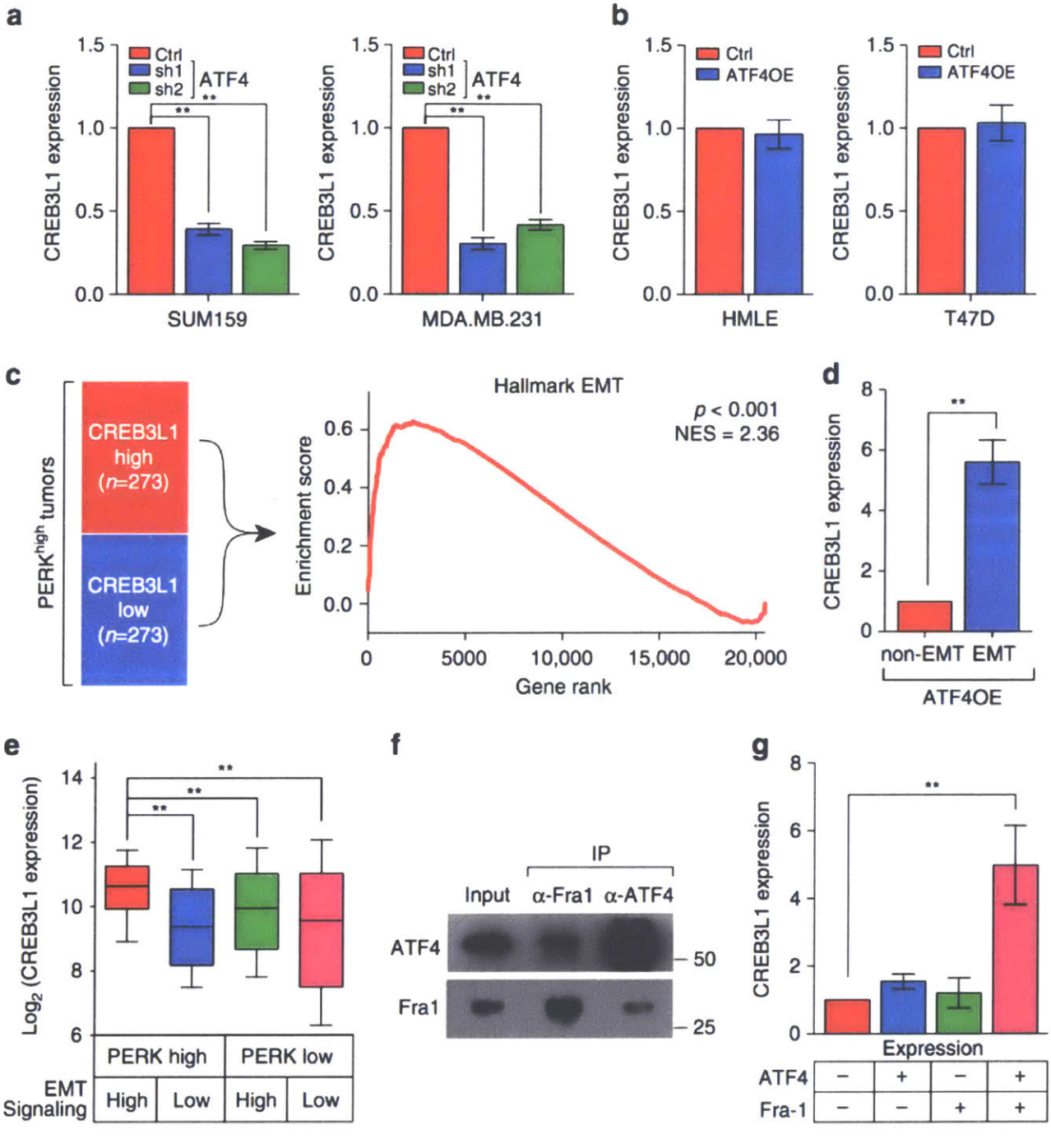

a
CREB3L1 expression
Ctrl
sh1
sh2
ATF4
**
SUM159
MDA.MB.231
b
ATF4OE
HMLE
T47D
c
PERK^high tumors
CREB3L1 high (n=273)
CREB3L1 low (n=273)
Hallmark EMT
p < 0.001
NES = 2.36
Enrichment score
Gene rank
d
non-EMT
EMT
ATF4OE
e
Log2 (CREB3L1 expression)
PERK high
PERK low
EMT Signaling
High
Low
f
IP
Input
α-Fra1
α-ATF4
ATF4
Fra1
50
25
g
Expression
ATF4
Fra-1

Fig. 5 PERK signaling requires an EMT to upregulate CREB3L1. **a** qPCR showing the relative expression of CREB3L1 in SUM159 and MDA.MB.231 cells transduced with shRNAs targeting luciferase or ATF4, $n = 3$. Data are represented as mean ± SEM. * indicates $p < 0.05$; ** indicates $p < 0.01$ (Student's *t*-test). **b** qPCR showing the relative expression of CREB3L1 in HMLE and T47D cells transduced with a control plasmid or ATF4-expressing plasmid (ATF4OE), $n = 3$. Data are represented as mean ± SEM. * indicates $p < 0.05$; ** indicates $p < 0.01$ (Student's *t*-test). **c** A set of 546 breast cancers (TCGA) expressing high levels of PERK pathway genes ($PERK^{high}$ tumors, see "Methods" section for details) were stratified into CREB3L1-high and CREB3L1-low cohorts, according to the expression of CREB3L1. Gene set enrichment analysis using the hallmark gene sets was performed to compare the CREB3L1-high and CREB3L1-low tumors (left). The top hallmark gene set was an EMT set (right). **D** qPCR showing the relative expression of CREB3L1 in HMLE cells overexpressing ATF4 with or without induction of an EMT, $n = 3$. Data are represented as mean ± SEM. * indicates $p < 0.05$; ** indicates $p < 0.01$ (Student's *t*-test). **e** Expression of CREB3L1 in breast cancers (TCGA) that have differential levels of PERK signaling and EMT signaling as indicated. Data are represented as a box-and-whiskers plot, $n = 122$. * indicates $p < 0.05$; ** indicates $p < 0.01$ (Mann–Whitney *U*-test). **f** HMLE cells were transduced with ATF4 and Fra-1 expressing constructs. Western blot for ATF4 or Fra-1 after immunoprecipitation of either ATF4 or Fra-1, which shows that these factors co-immunoprecipitate each other. **g** qPCR showing the relative expression of CREB3L1 in HMLE cells overexpressing control, ATF4, Fra-1, or a combination of ATF4 and Fra-1, $n = 3$. Data are represented as mean ± SEM. * indicates $p < 0.05$; ** indicates $p < 0.01$ (Student's *t*-test)

Chapter 3.9: Supplemental Data

Supplementary Table 1. The list of 388 PERK-ATF4 downstream genes.

SPP1
ATF5
SERPINF1
HMOX1
SNAI2
TMEM176A
GRB10
CLU
PTGIS
CCND2
CYP1B1
LTBP2
CEBPB
ATF4
FNDC3A
YPEL3
ASNS
PDGFRA
STEAP1
MAGED2
UAP1L1
RENBP
ETS2
SDPR
CCNG1
RBPJ
LRP1
S100A4
VEGFA
FLOT1
LOX

TGIF1
SCPEP1
IRF2BP2
TMEM176B
DDIT3
PTX3
SLIT2
CYB5R1
IL6ST
TGFBI
RPS28
YPEL5
ITM2B
SLC30A4
AGA
ERRFI1
ACTA2
CSF1
FOXG1
ROBO1
ALDH18A1
DNAJB9
RHBDD1
EMP2
EMB
LTBP3
HMGA2
SLC7A1
CARS
MDFIC
AEBP1
MMP2
IL17RC

FTH1
LRIG1
OGN
PIK3R1
LAMP2
WFS1
MAN2B2
GABARAPL1
HEXA
EGFR
AMDHD2
ATP1B1
SLC7A5
EID1
RAB11FIP5
CTSD
NFE2L1
TPP1
GM2A
PSPH
PKD2
SMAD3
SGCE
GAA
HEXB
ENO2
MOCS1
AKAP12
SQSTM1
TXNIP
CTSB
EYA4
ACADSB

ASAH1
CALCOCO1
ARHGAP29
SERINC3
TGFBR2
EIF4EBP1
CNKSR3
ARNT2
ABHD4
IAH1
CAV1
RPL37A
GRN
MAN2A1
ANKRD1
GNG12
HEBP1
ZFP36L2
BLOC1S1
GAS1
ITGB5
ANTXR1
SLC3A2
ADAM9
PDGFRB
SAT1
HIF1A
NARS
GPC4
MAP1LC3B
DDR2
BNIP3
SEMA3F

SLC12A2
AARS
SLC20A1
GNS
FOXC1
CDH11
TIMP2
SLC2A6
GHITM
IER3
IGF2R
COPZ2
RND3
SKAP2
TCEAL8
KIF3A
TBX15
EFNB2
NUCB1
MRC2
CPE
HERPUD1
MYL9
MPV17
FNIP1
RNF11
HERPUD2
USP11
SDC2
MVP
PLK2
CRIM1
APLP2

BGN
TEX264
P4HA2
GPR108
FRMD6
GARS
BID
INSIG2
FYN
ASCC2
ANXA4
GTPBP2
MKNK1
VAT1
MTHFD1L
MID1
HK2
LGALS3
CMTM3
RPL22L1
PLD3
CD47
POLD4
WARS
UGP2
IRX1
LGALS8
HDAC8
OSBPL2
GOT1
SH3BGRL
CCND1
MBNL2

FMR1
CXADR
TRAPPC6A
PSAP
UBE2W
PDRG1
GSTM1
CGNL1
RHOQ
HIST1H1C
COL6A2
ALDH3A2
UPRT
DERL1
HS6ST2
FEZ1
NFE2L2
TGFB1I1
BNIP3L
THRA
LIPA
AHNAK
NENF
SEL1L
PDIA4
OGT
COL1A2
PLA2G4A
MTHFD2
NR2F2
TAF15
GYPC
PLA2G12A

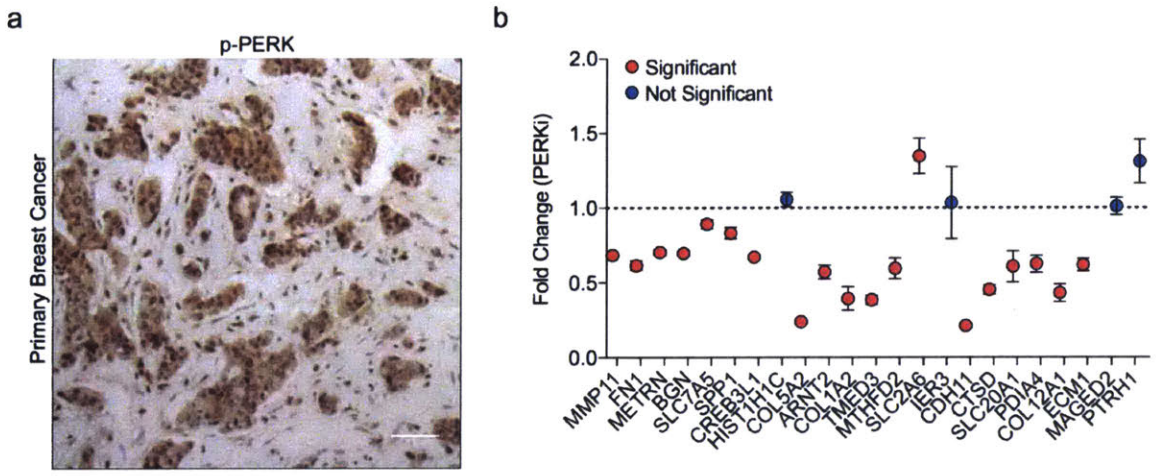
IMPACT
MEA1
TWIST1
SARS
PINK1
EMILIN1
KCTD15
ITM2C
EPDR1
PCMTD2
NPC2
RCN1
FGFR1
DYNLT3
IMP3
MARCKS
SIL1
TMEM106B
FN1
MMP11
GABARAP
STARD3NL
LAMP1
VIM
LONP1
ZFP36L1
DSTN
NPDC1
PSAT1
GHR
PHLDA3
PTRH1
XPOT

RNASET2
COPG2
OSTF1
PDK1
YARS
CREB3L1
SDC4
PGM2
CTDSP2
CTSA
COL6A1
RABAC1
SLC25A36
MSI2
APP
PTPRG
COL6A3
FBLN1
CRTAP
LRRFIP1
GRINA
ITFG1
SLC6A8
MCFD2
LAPTM4A
SYDE1
TARS
TBC1D17
RAP2C
LMO4
PNRC1
RTN4
PRNP

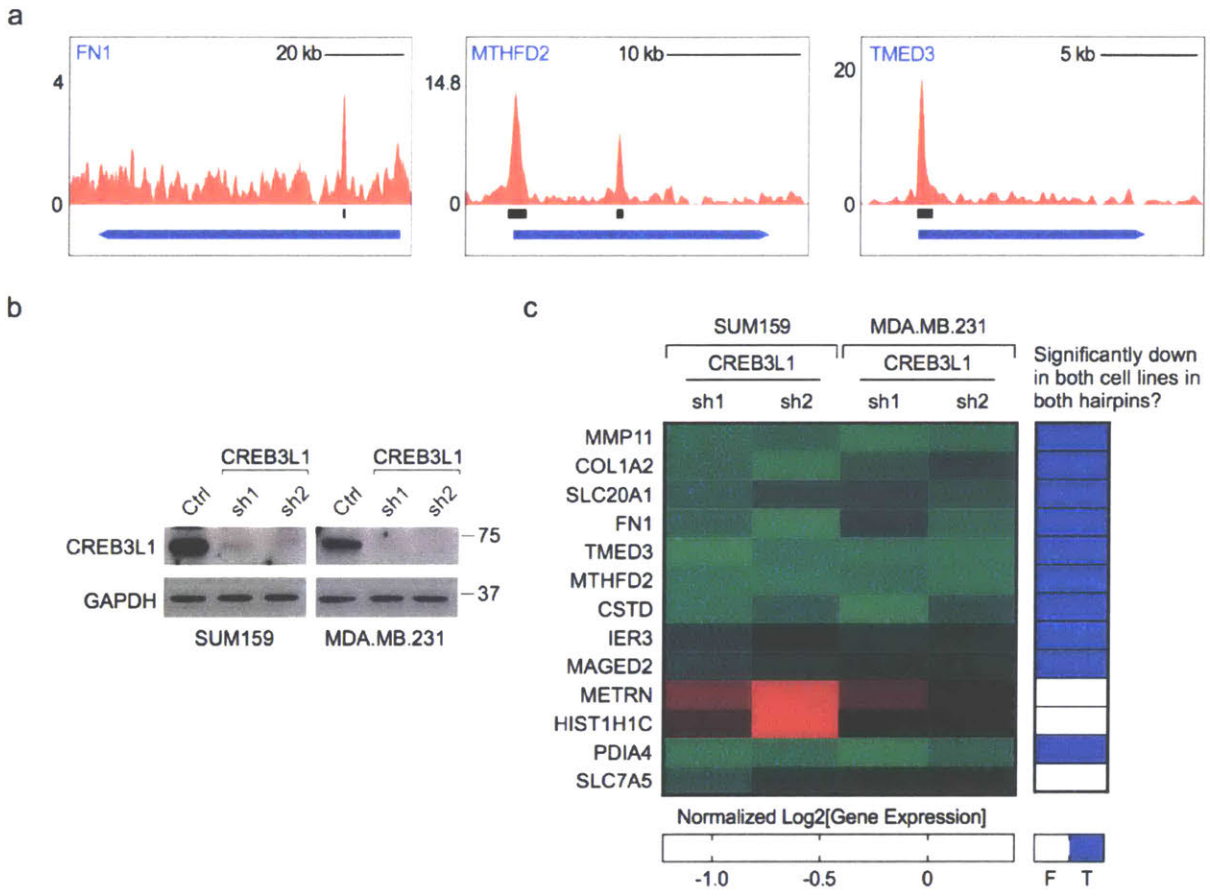
MXRA8
ECHS1
TCF4
MAOA
VEGFB
AKAP2
EEF2K
ATP6V1G1
TMEM98
YIPF3
SPRED1
BRI3
SPCS3
COL5A2
SH3PXD2B
SHMT2
TRIM35
LTBP4
NFIL3
P4HA1
RASSF8
ECM1
UBE2H
PHF10
WDR45
METRN
WIPI1
HSPB6
DHX40
NR1D2
RALBP1
LTBP1
CST3

FAT1
SLC1A5
ARID5B
ARL8A
KLF3
HSPA5
SOX4
TNFRSF1A
GNG5
ALDH2
MRPS6
LONP2
HSD17B10
TMED3
MAP1LC3A
PIGN
NPC1
BCAT2
ANXA7
PARP3
SOAT1
TBPL1
TAPBP
YIF1A
VPS26A
DPY19L1
RPL23
ZYG11B
NFU1
SUMF1
CHMP2B
CAPN2
PDCD4

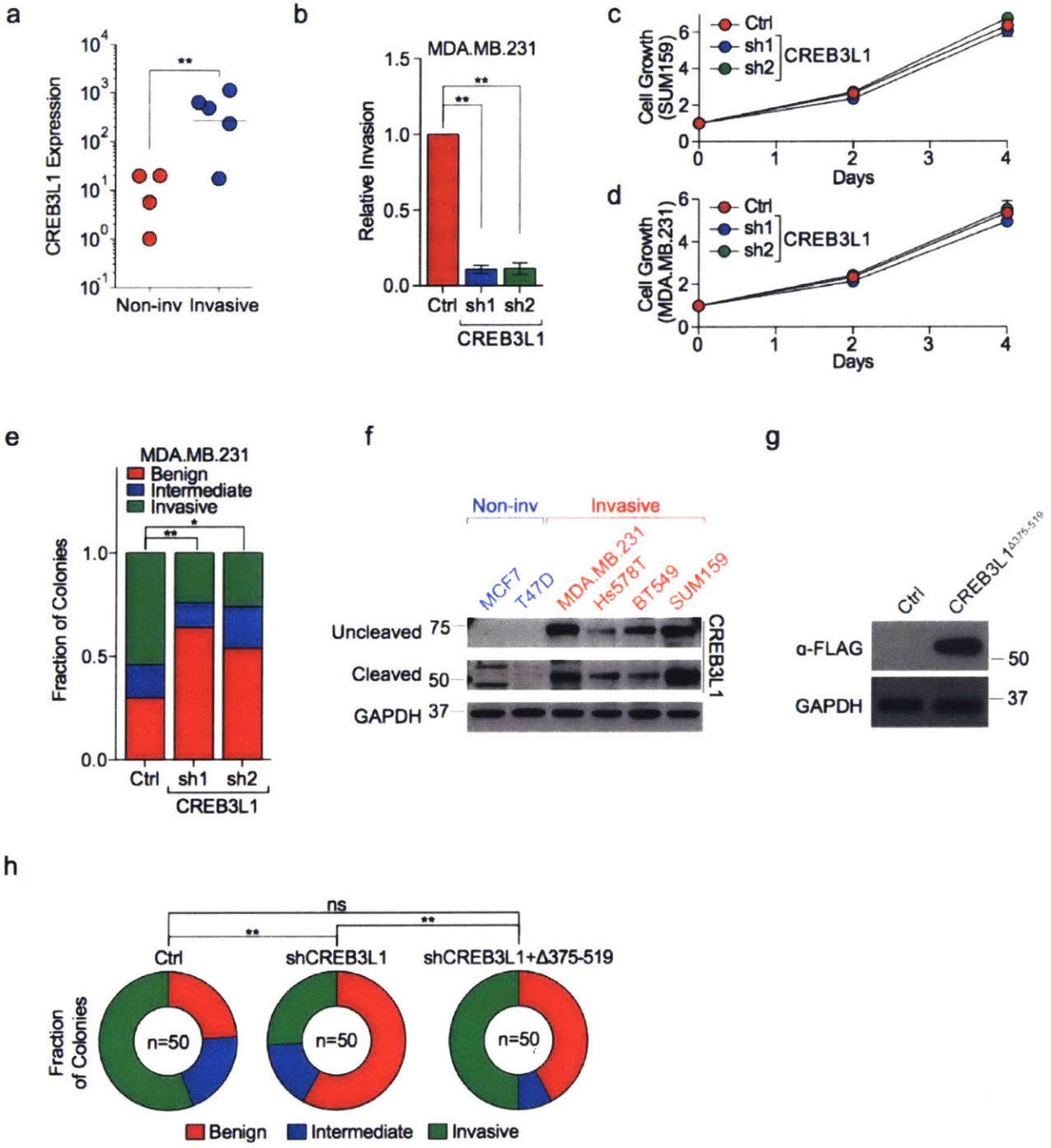
PCDHGC3
RCN3
PLEKHA3
CDKN1B
CHCHD7
JAK1
IARS
RPL15
COL12A1
ZFAND5
GLCE
FZD7
CCNDBP1
MAFG
CTSZ
ATP6V1F
PRCP
OSBPL3
DERL2
CRLS1
GPX4
CRYZL1
CAST
KRCC1
PHLDB2
AES
BRWD1



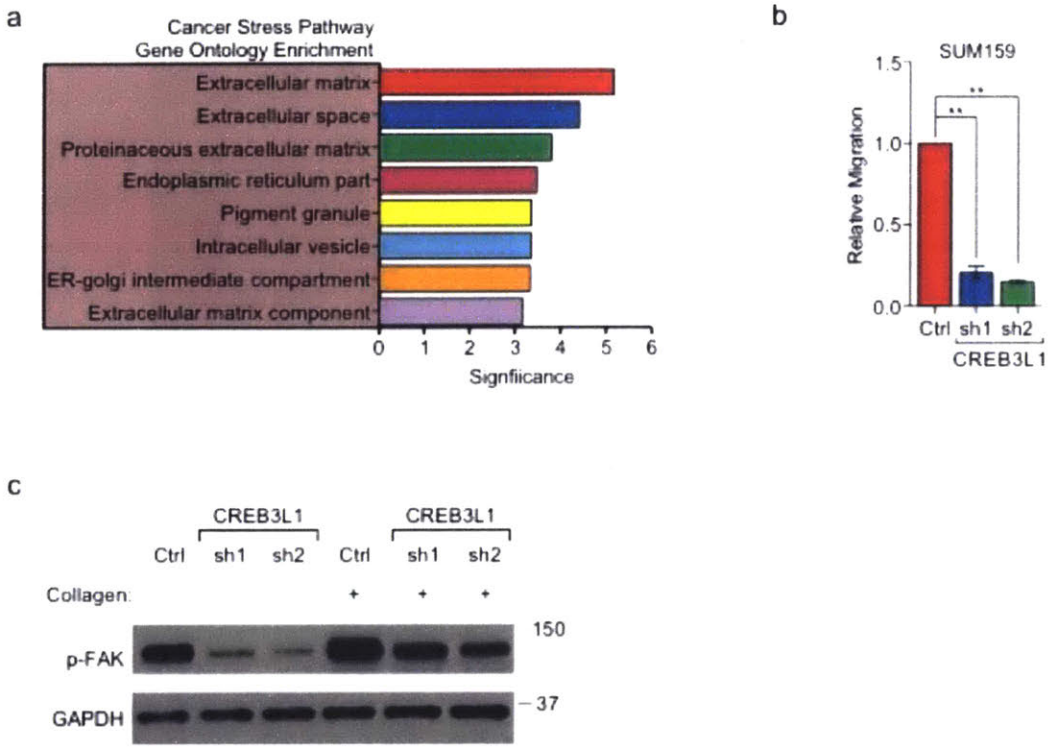
Supplementary Figure 1. PERK is phosphorylated in primary breast cancer and induces the expression of cancer-specific PERK genes (a) IHC staining for phosphorylated PERK in a set of 50 breast cancer samples (Biomax, #BR10010d). Phosphorylation of PERK was found in 35 out of 50 primary breast cancer tissue samples. Shown was a representative image of staining (scale bar: 40 μm). **(b)** qPCR showing the fold change of the expression of 23 cancer-specific PERK genes in SUM159 cells treated with GSK2656157, a PERK inhibitor, relative to vehicle control-treated cells, n=3. Data are represented as mean ± SEM. Red dot indicates $p<0.05$ (Student's *t*-test).



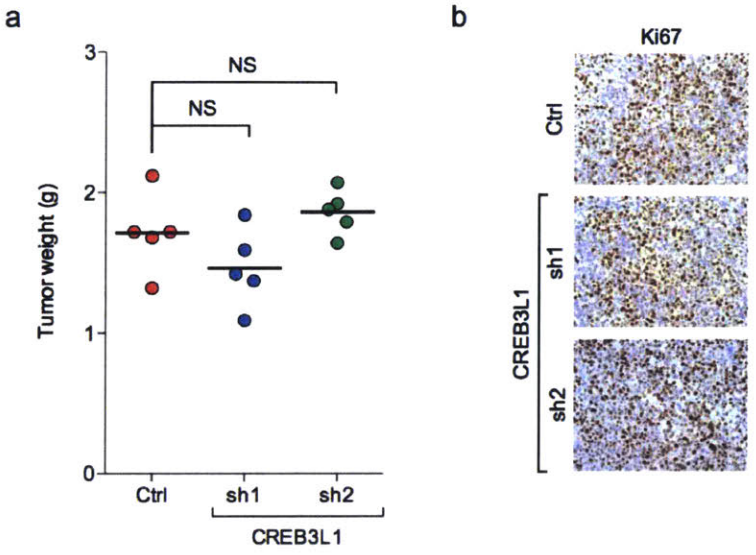
Supplementary Figure 2. CREB3L1 regulates the expression of Cancer-Specific PERK genes (a) Distribution of CREB3L1 ChIP-seq fold-enrichment over input control at the locus of three CSPS genes, FN1, MTHFD2, and TMED3. **(b)** Western blot showing the expression of CREB3L1 in SUM159 and MDA.MB.231 cells transduced with shRNAs targeting luciferase or CREB3L1. **(c)** Heatmap showing the fold change of the expression of 13 putative CREB3L1-target genes in SUM159 and MDA.MB.231 cells transduced with shRNAs targeting CREB3L1, relative to control shRNA-transduced cells, n=4. Genes that were significantly down-regulated by both shRNAs in both cell lines were marked in blue.


a
CREB3L1 Expression
Non-inv
Invasive
b
MDA.MB.231
Relative Invasion
Ctrl
sh1
sh2
CREB3L1
c
Cell Growth (SUM159)
Ctrl
sh1
sh2
CREB3L1
Days
d
Cell Growth (MDA.MB.231)
Days
e
MDA.MB.231
Benign
Intermediate
Invasive
Fraction of Colonies
f
Non-inv
Invasive
MCF7
T47D
MDA.MB.231
Hs578T
BT549
SUM159
Uncleaved 75
Cleaved 50
GAPDH 37
CREB3L1
g
Ctrl
CREB3L1Δ375-519
α-FLAG
50
GAPDH
37
h
ns
Ctrl
shCREB3L1
shCREB3L1+Δ375-519
Fraction of Colonies
n=50
Benign
Intermediate
Invasive

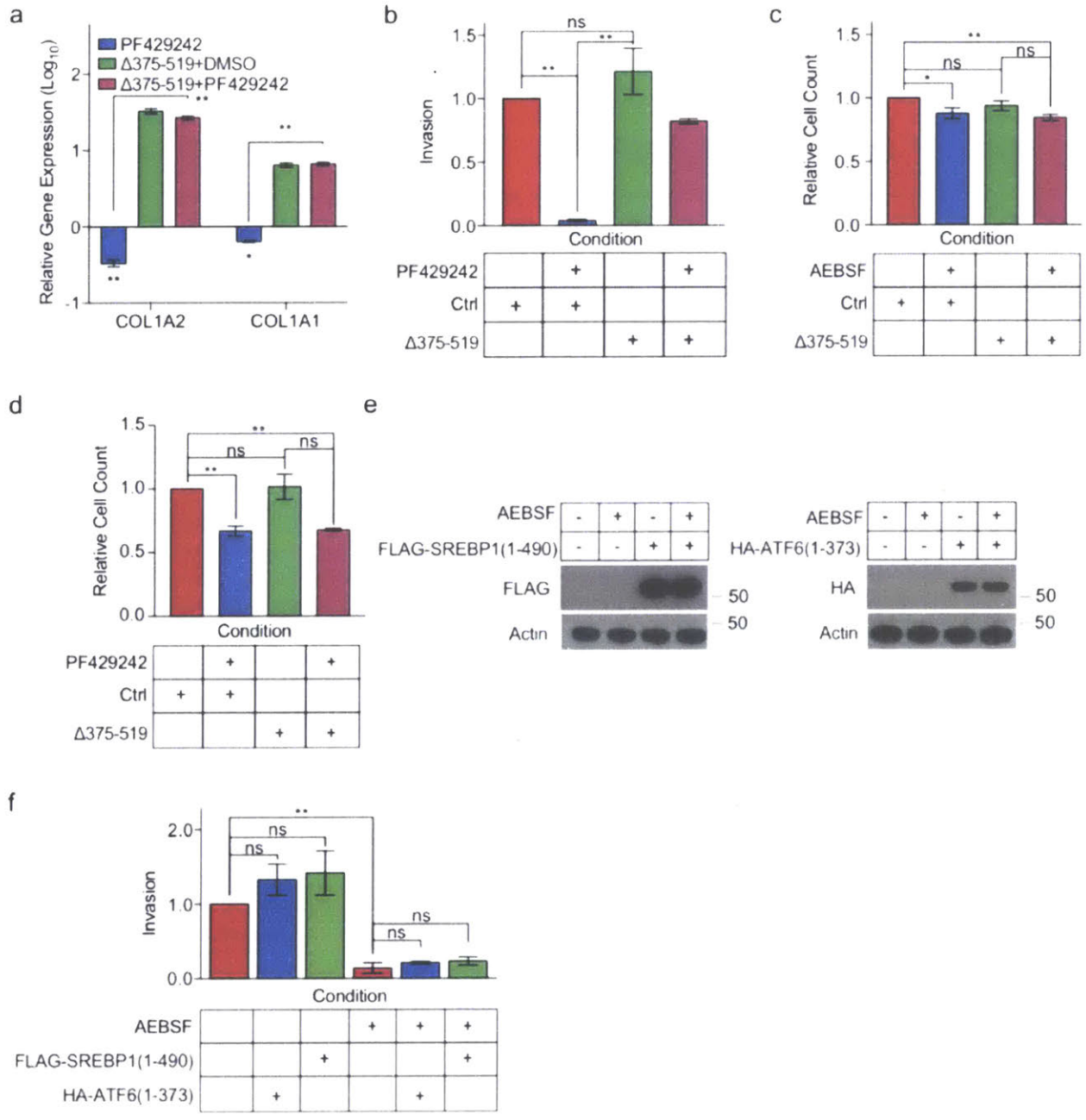
Supplementary Figure 3. CREB3L1 is up-regulated in invasive breast cancer cell lines and required for cell invasion (a) qPCR showing the relative expression of CREB3L1 in 4 luminal human breast cancer cells, MCF7, T47D, BT474, and ZR-75-3, and 5 basal-B cell lines, SUM159, MDA.MB.231, MDA.MB.157, BT549, and Hs578T, n=3. Data are represented as mean ± SEM. * indicates $p<0.05$; **indicates $p<0.01$ (Student's *t*-test). **(b)** Quantification of cell invasion of MDA.MB.231 cells transduced with shRNAs targeting luciferase or CREB3L1 in a basement membrane-coated trans-well assay, n=3. Data are represented as mean ± SEM. * indicates $p<0.05$; **indicates $p<0.01$ (Student's *t*-test). **(c)** and **(d)** Quantification of cell growth of SUM159 cells (c) or MDA.MB.231 cells (d) transduced with shRNAs targeting luciferase or CREB3L1, n=3. **(e)** Quantification of colony types formed by MDA.MB.231 cells transduced with shRNAs targeting luciferase or CREB3L1 in a seven day 3D matrigel invasion assay, n=50. $p=0.0022$ for sh1 and $p=0.0139$ for sh2, Chi-square test. **(f)** Western blot showing the uncleaved and cleaved CREB3L1 in MCF7, T47D, SUM159, MDA.MB.231, BT549, and Hs578T cells. **(g)** Western blot showing the expression of truncated CREB3L1 in HMLE cells overexpressing flag tagged CREB3L1Δ375-519. **(h)** Quantification of colony types formed by MDA.MB.231 cells transduced with an shRNA targeting luciferase, an shRNA targeting CREB3L1, or an shRNA targeting CREB3L1 and a construct expressing flag tagged CREB3L1Δ375-519 in a seven day 3D matrigel invasion assay, n=50. $p=0.0017$ comparing Ctrl vs shCREB3L1, $p=0.07$ comparing Ctrl vs shCREB3L1+Δ375-519, and $p=0.04$ comparing shCREB3L1 vs shCREB3L1+Δ375-519, Chi-square test.



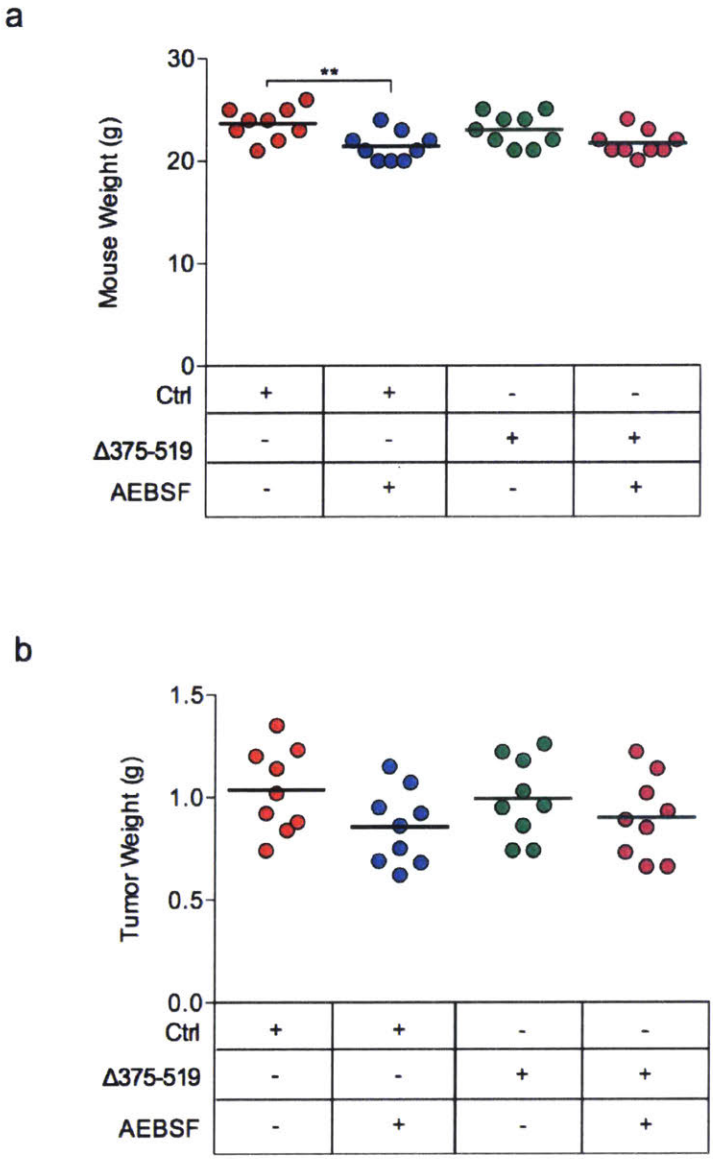
Supplementary Figure 4. CREB3L1 regulates ECM gene expression and activates FAK (a) Gene ontology analysis of cancer-specific PERK set regulated by CREB3L1 using cellular components set provided by the molecular signature database. Significance is represented as the negative log10 of the p-value. **(b)** Quantification of cell migration of SUM159 cells transduced with shRNAs targeting luciferase or CREB3L1 in a trans-well assay (not Matrigel-coated), n=3. **(c)** SUM159 cells transduced with shRNAs targeting luciferase or CREB3L1 were cultured in dishes coated with or without type I collagen for 24 hours prior to protein extraction and western blotting. The expression of phosphorylated FAK and GAPDH were examined by western blot. Data are represented as mean ± SEM. * indicates $p<0.05$; **indicates $p<0.01$ (Student's *t*-test).



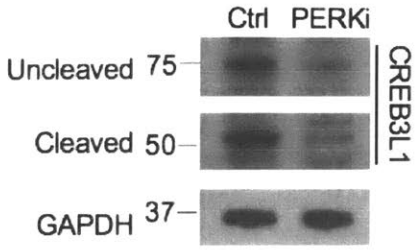
Supplementary Figure 5. Knockdown of CREB3L1 does not affect tumor growth *in vivo* (a) Tumor weights from mice that were inoculated with MDA.MB.231-LM2 cells transduced with a control hairpin or two hairpins targeting CREB3L1. Lines represent the mean, n=5. **(b)** IHC for Ki67 expression on tumors sections collected from the mice in (a).


a
PF429242
Δ375-519+DMSO
Δ375-519+PF429242
Relative Gene Expression (Log10)
COL1A2
COL1A1
b
Invasion
Condition
PF429242
Ctrl
Δ375-519
c
Relative Cell Count
Condition
AEBSF
Ctrl
Δ375-519
d
Relative Cell Count
Condition
PF429242
Ctrl
Δ375-519
e
AEBSF
FLAG-SREBP1(1-490)
FLAG
Actin
HA-ATF6(1-373)
HA
f
Invasion
Condition
AEBSF
FLAG-SREBP1(1-490)
HA-ATF6(1-373)

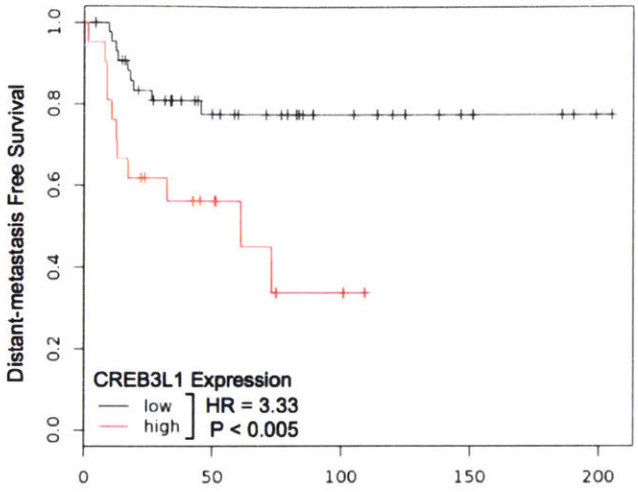
Supplementary Figure 6. Suppressing CREB3L1 with chemical inhibitors of proteases in vitro
(a) SUM159 cells were treated with solvent control or PF429292, in combination of transduction with control or CREB3L1Δ375-519. qPCR analyses were performed to quantify the expression of ECM genes *COL1A1* and *COL1A2*. The gene expression was normalized to the solvent treated control transduction samples, n=3. **(b)** Quantification of cell invasion of cells from (a) in a basement membrane-coated trans-well assay, n=3. **(c)** 16 hours after SUM159 cells were treated with solvent control or AEBSF, in combination of transduction with control or CREB3L1Δ375-519, cell number was counted and plotted for cell survival, n=3. **(d)** 16 hours after SUM159 cells were treated with solvent control or PF429242, in combination of transduction with control or CREB3L1Δ375-519, cell number was counted and plotted for cell survival, n=3. **(e)** Western blot showing expression of FLAG tagged SREBP1 (N terminus 1-490) or HA tagged ATF6 (N terminus 1-373) in SUM159 cells treated with the indicated condition. **(f)** Quantification of cell invasion of SUM159 cells treated with solvent control or AEBSF, in combination of transduction with control construct, FLAG tagged SREBP1(N terminus 1-490), or HA tagged ATF6 (N terminus 1-373), in a basement membrane-coated transwell assay, n=3. Data are represented as mean ± SEM. * indicates $p<0.05$; **indicates $p<0.01$ (Student's *t*-test).



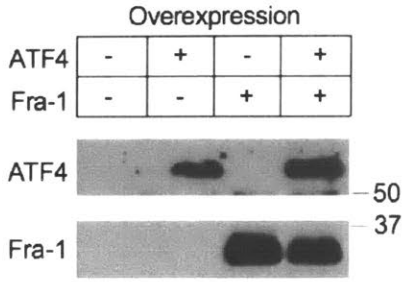
Supplementary Figure 7. Effect of chemical inhibition of CREB3L1 on the growth of mouse or the growth of primary tumors (a) Mouse weights following treatment with AEBSF or solvent control. Mice were inoculated with MDA.MB.231-Luc-LM2 cells transduced with a control plasmid or a constitutively active CREB3L1 (CREB3L1 Δ375-519), n=9. **(b)** Tumor weights from the conditions described in (a), n=9. * indicates $p<0.05$; **indicates $p<0.01$ (Student's *t*-test).



Supplementary Figure 8. Inhibition of PERK reduces the expression of CREB3L1 but does not change its proteolytic cleavage Western blot showing the expression of uncleaved and cleaved CREB3L1 in SUM159 cells treated with vehicle control or a PERK inhibitor, GSK2656157.



Supplementary Figure 9. Expression of CREB3L1 predicts distant metastasis-free survival in the mesenchymal subtype of triple-negative breast cancers Kaplan-Meier curves showing significant association of elevated CREB3L1 expression (red line) with shorter distant metastasis-free survival in a cohort of patients with triple-negative breast cancer of the mesenchymal subtype (KMplotter, mesenchymal subtype of TNBC).



Supplementary Figure 10. Overexpression of ATF4 and Fra-1 in HMLE cells Western blot showing the expression of ATF4 and Fra-1 in HMLE cells overexpressing a control construct, ATF4, Fra-1, or a combination of ATF4 and Fra-1.

Figure 2h

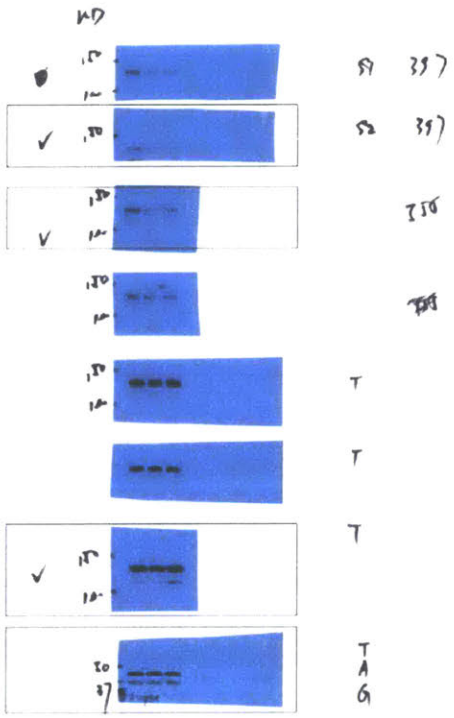


Figure 4a

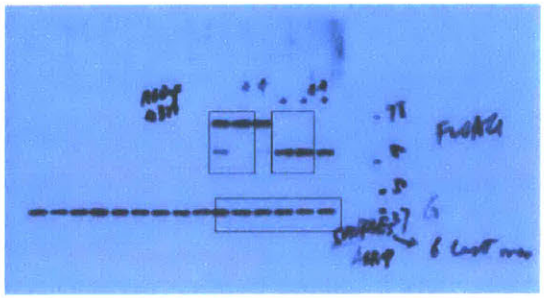


Figure 4a

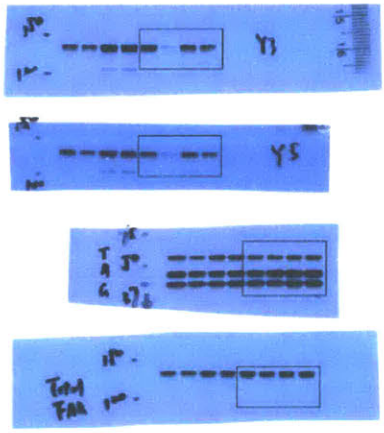
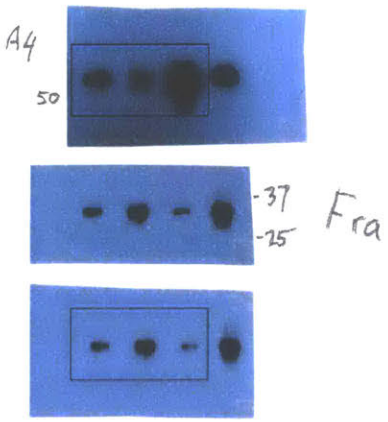
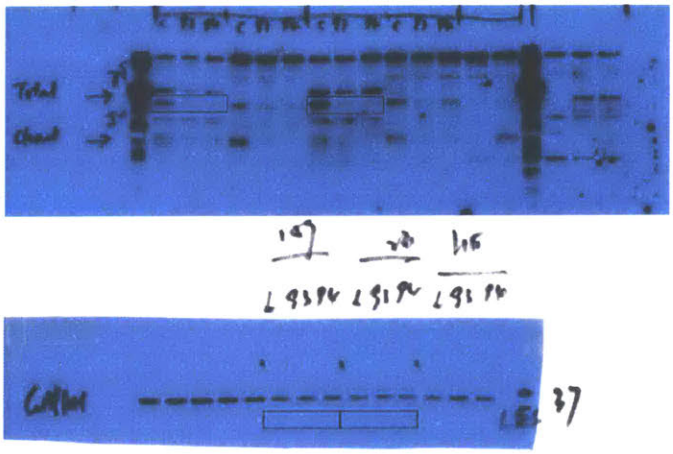


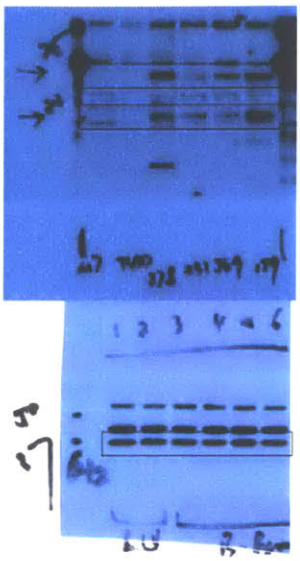
Figure 5f



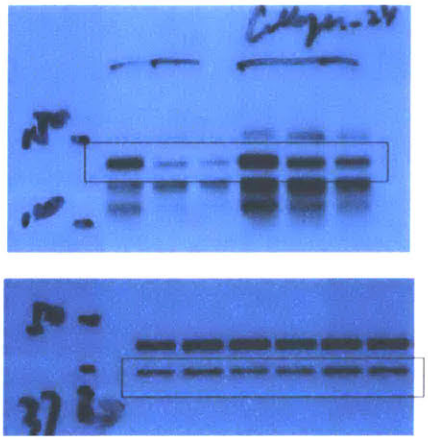
Supplementary Figure 2b



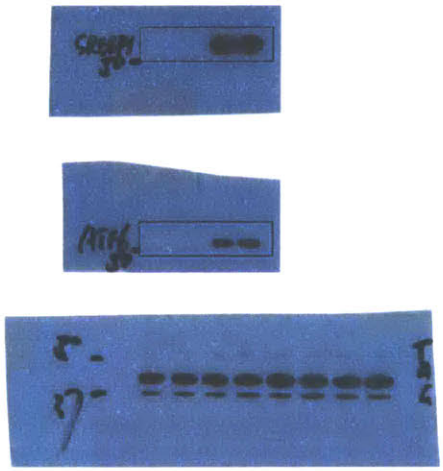
Supplementary Figure 3f



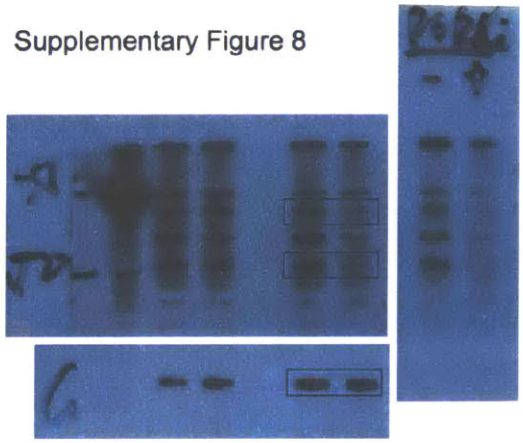
Supplementary Figure 4c



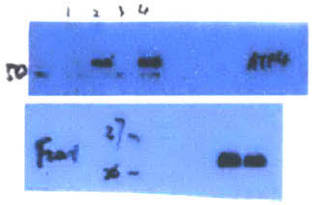
Supplementary Figure 6e



Supplementary Figure 8



Supplementary Figure 10



Supplementary Figure 11. Uncropped gel images of western blots Uncropped gel images of western blots showed in main figures and supplementary figures.

Chapter 4

Cell-matrix interaction drives multi-lineage differentiation of mammary stem cells through the discoidin domain receptor tyrosine kinase, DDR1

Author List: Dexter X. Jin*, Daniel H. Miller*, Todd M. Gierahn, Carman M. Li, Ethan S. Sokol, Yu-Xiong Feng, Robert A. Mathis, Charlotte Kuperwasser, J. Christopher Love, Piyush B. Gupta

*These authors contributed equally to this work.

DXJ performed experiments shown in Figure 1, Figure 2B-F, Figure 3, Figure 4A-C and 4E-F, and Figures S1-3. All authors contributed to the design of experiments, analysis of data, or editing of the manuscript.

Chapter 4.1: Abstract

Epithelial cells have a remarkable ability to self-organize into architecturally complex tissues comprising multiple cell lineages. Tissue morphogenesis and differentiation are regulated through interactions with the extracellular matrix (ECM). Despite this, little is known about the cell-matrix mechanisms that direct mammary stem cell (MaSC) fate decisions and differentiation. Here we show that the collagen receptor DDR1 coordinates stem cell differentiation and heterotypic Jagged1-Notch signaling to drive luminal differentiation and TDLU morphogenesis. In a 3D CRISPR organoid screen targeting all human kinases, we identified the collagen receptor DDR1 as a key component of the signaling cascade that triggers matrix-induced differentiation of bipotent mammary stem cells. By integrating functional assays with single-cell RNA sequencing (scRNA-seq) of primary human tissues cultured *ex vivo*, we show that DDR1 is required for stem cell differentiation into the basal cell lineage. Furthermore, we show that DDR1 also indirectly stimulates luminal Notch signaling to drive lobulogenesis by activating JAG1 expression in basal cells. These findings establish how ECM signals are transduced in a bilayered tissue from the basal to luminal cell lineages to coordinate duct and lobule growth. More generally, these findings suggest that DDR1 may coordinate collagen and Notch signaling to regulate multi-lineage differentiation and morphogenesis across diverse tissues.

Chapter 4.2: Introduction

The mammary gland undergoes dynamic changes beginning from its initial expansion during puberty, to fluctuations in proliferation and apoptosis during cyclical hormonal estrous cycles, ultimately to the robust expansion and terminal differentiation at pregnancy and lactation. Since the epithelial portion of the mammary gland must endure enormous cellular output to meet the demands of regenerative cycles and successive pregnancy and lactation, mammary stem cells (MaSCs) sustain the regenerative properties of mammary epithelial tissue (1-4).

The mammary epithelium is a complex bilayered network of ducts and lobules each comprised of an inner luminal layer of cells and an outer myoepithelial layer of cells. It also harbors stem and progenitor cells, which are the source of both differentiated luminal and myoepithelial cells. This bilayered epithelium is enveloped within a rich ECM, making the cell–matrix interface pivotal in controlling both cell function and tissue structure of the gland. While cell-matrix interactions have a critical role throughout mammary tissue formation, little is known about how these interactions affect mammary stem cell fate and differentiation. In addition, it is unclear how matrix signals are transduced in a bilayered tissue from the basal to luminal cell lineages to coordinate ductal and lobular growth.

Defining key mechanisms of stem cell-matrix interactions has relied on transgenic animal studies and limited tissue culture methods. Recently, we developed a novel 3D hydrogel culture that supports the growth and maturation of human mammary tissues derived from MaSCs within a matrix scaffold (5). The mammary tissues expanded in these cultures faithfully model the intricate architecture, cell lineages, and hormonal responses of human breast tissue (5). Since these hydrogel matrices contain only defined and physiologically relevant components and make it possible to access, visualize, and manipulate the expanding tissues, they provide a unique opportunity for delineating the genes and signaling networks that regulate stem cells-matrix interactions in human mammary tissue, as well as both lineage- and architecture-committed progenitors.

Using this new method, we performed a loss-of-function screen in which a cell line model of multi-potent MaSCs undergoes differentiation in 3D. Our screen revealed that the collagen receptor, discoidin domain receptor tyrosine kinase- DDR1, is required for MaSC differentiation. Using single-cell RNA sequencing (scRNA-seq) on patient-derived mammary tissues grown in hydrogels (5, 6), we show that DDR1 mediates the transition of MaSCs into unipotent basal progenitors. Furthermore, we demonstrate that lineage-committed basal cells express the Notch receptor ligand, JAG1, which activates Notch signaling in luminal progenitor cells, allowing for lobular maturation (7-9). Our findings elucidate a DDR1-JAG1-Notch axis critical for transducing matrix signal via stem cells to coordinate bilayered mammary epithelial differentiation and morphogenesis.

Chapter 4.3: Results and Discussion

A CRISPR screen in 3D culture identifies kinases involved in mammary epithelial cell differentiation

We previously reported that the MCF10A cell line serves as a model of human bipotent MaSCs (10). When seeded into 3D matrix hydrogels, MCF10A cells give rise to 3D organoids that mature and differentiate into complex ductal, lobular, and ductal-lobular tissue rudiments that resemble analogous structures within mammary gland tissue. Furthermore, as MCF10A cells mature and differentiate, they lose the ability to self-renew and reseed new cultures. As such, inhibition of MCF10A differentiation leads to the ability of organoids to retain their self-renewal activity and reseed secondary cultures (10) (Figure 1A). Because of this property, we have previously identified factors required for MaSCs to exit the bipotent state and differentiate (10).

To investigate cell-matrix interactions pivotal for MaSC differentiation, we exploited this loss of organoid re-seeding activity as a platform to perform a loss-of-function screen to identify novel pathways involved in mammary stem cell differentiation (Fig 1A). We chose to screen actionable targets using a custom pooled CRISPR sgRNA library containing 5070 guides targeting 507 kinases (11). We then transduced MCF10A cells with the pooled sgRNA library or control uninfected cells. Infected cells were seeded in hydrogels and allowed to form organoids. Primary organoids were dissociated, re-seeded and allowed to form secondary organoids. The number of secondary organoids increased dramatically in sgRNA library-infected MCF10A cells, compared to control cells (Fig 1B). To identify the kinases inhibiting differentiation, we sequenced the sgRNAs in the secondary organoids and in the original pool of 2D cultured cells. For each gene, we calculated the average log-fold change in representation of sgRNAs between secondary organoids and 2D cultured cells, and calculated an FDR-adjusted p-value using RIGER (Fig 1C).

Using this approach, we identified 33 kinases enriched in secondary organoids relative to controls (Table S1). We noted that several kinases (FYN (12), MAPK7 (13), and MAP3K7 (14)) have been implicated previously cellular differentiation validating the robustness of the approach. Indeed, gene-set enrichment analysis (GSEA) of the screen "hits" revealed enrichment in previously published gene-expression profiles of stem cells, further demonstrating that secondary organoid formation is indeed a valid surrogate marker for the MaSC phenotype (Fig S1).

Of the remaining kinase hits, by far the most significantly enriched gene was the collagen receptor, DDR1. In the mouse, DDR1 mRNA expression is highest during the two phases of mammary growth when morphogenesis and differentiation of mammary stem cells are required for mammary gland development: puberty and pregnancy (15). Interestingly, female mice lacking DDR1 fail to undergo lobular-alveolar differentiation during pregnancy and exhibit hyperproliferative and abnormal branching

of mammary ducts (16). These results combined with our screen finding suggest that DDR1 might be a key mediator of the matrix-epithelial cross talk during morphogenesis in the mammary gland and could be acting as a stem cell regulator. Because DDR1 had not been previously implicated in mammary stem/progenitor biology, we set out to delineate its function in mammary epithelial differentiation.

DDR1 is required for alveolar/lobular morphogenesis and stem/progenitor cell differentiation

To validate the ability of DDR1 loss to prevent differentiation, MCF10A cells were transduced with two additional independent sgRNAs against DDR1. Consistent with the primary screen, cells lacking DDR1 expression showed a significant increase in secondary organoid seeding in 3D (Fig 2A, B). Furthermore, using an independent chemical approach to inhibit DDR1 cleavage (DDR1i), MCF10A cells treated with DDR1i exhibited the same phenotype of increased secondary organoid formation as those transduced with sgRNAs against DDR1 (Fig S2). Thus, the increase in secondary organoid formation upon loss of DDR1 protein expression was confirmed chemically and genetically. We conclude that in MCF10As, DDR1 expression is required for mammary stem cell differentiation.

To investigate the role of DDR1 in primary human cells, we seeded freshly dissociated mammary epithelial cells (MECs) isolated from patient tissue and seeded them in 3D hydrogels. Cultures were treated in the presence or absence of DDR1i and analyzed after 12 days. While control tissues formed complex ductal-lobular structures, DDR1 inhibition led to a failure of lobular/alveolar formation and caused the development of breast tissues that contained only ducts (Fig 2C). Since luminal cells drive lobular/alveolar differentiation (17, 18), we examined whether DDR1 inhibition might be affecting the number of luminal or luminal progenitor cells. However, there was no reduction in the number of luminal cells in tissues developing in the absence of DDR1, rather we found a >2-fold increase in the number of luminal cells (Fig 2D). Therefore, we next examined if release from DDR1 inhibition might induce lobular/alveolar differentiation. Hydrogel-cultured tissues were treated with DDR1i for 12 days and then removed from the inhibitor for an additional 9 days. Indeed, release from DDR1 inhibition led to robust lobular/alveolar differentiation, (Fig 2E) indicating that the failure to form lobule/alveoli upon DDR1 inhibition was due to a defect in luminal progenitor cell differentiation.

We next examined whether defective mammary morphogenesis upon DDR inhibition might be due to changes in the number of progenitor cells. We assessed stem and progenitor cell number using colony forming assays on dissociated cells isolated from breast tissues grown in hydrogels in the presence or absence of DDR1i. In this assay stem and progenitor cells form colonies containing differentiated luminal or basal cells. DDR1 inhibition led to a > 2-fold increase in the proportion of colony-forming cells (Fig 2F) suggesting that that DDR1 regulates human breast stem/progenitor cell numbers as well as their differentiation.

DDR1 inhibition blocks differentiation of both stem and luminal progenitor cells

To determine how DDR1 inhibition is affecting breast stem cell biology, we performed scRNA-seq using Seq-Well (6) to generate a comparative cellular atlas at single-cell resolution on human breast tissues grown for 14 days in 3D hydrogels under three different conditions: 1) vehicle-treated control, 2) DDR1-inhibited (DDR1i), and 3) DDR1-inhibited for 12 days followed by release from inhibition for 2 days (DDR1i Rel). We first began by examining the molecular heterogeneity in 14-day human breast tissue outgrowths. Unsupervised analysis of 1467 cells with greater than 500 genes detected per cell revealed nine distinct clusters corresponding to different epithelial cell states in 3D human breast tissues (Figure 3A). Using unbiased clustering analysis, spatial reconstruction of single cell data, and integrated analysis across all three conditions, classified cells into 8 epithelial clusters based on a unique group of expressed genes (Fig 3A). Analysis using epithelial lineage-specific cytokeratins (Luminal: KRT8, KRT18, KRT19/ Basal: KRT14, KRT5) revealed six of the clusters were of the basal lineage (Clusters 0, 1, 2, 3, 5, and 7) while the remaining two clusters were of the luminal linage (Clusters 4 and 6; Fig S3A, B).

Gene set enrichment analysis (GSEA) further defined cells within Clusters 0, 1 as undifferentiated stem/progenitor cells within the basal lineage and Cluster 6 as undifferentiated stem/ progenitor cells within the luminal lineage. Additionally, since basal lineage Cluster 1 and luminal lineage Cluster 6 both displayed high scores for both stem cell and proliferation signatures, these findings indicated that these populations were active progenitor cells (Fig 3B, C, Fig S3C, Table 1). In contrast, Cluster 0 only expressed stem cell signatures suggesting it was not an actively proliferating basal stem cell population. The remaining Clusters (2, 3, 5 and 7) and Cluster 4 consisted of more mature lineage-restricted differentiated cells of the basal and luminal lineages, respectively (Fig 3B). Luminal Cluster 4 expressed high levels of LTF consistent with mature lobular luminal cells (Fig S3D).

To understand the lineage relationships between cell clusters, we constructed a pseudotemporally ordered tree using Monocle. This analysis revealed that Cluster 0 had multi-lineage potential giving rise to both basal and luminal cell lineages, while Cluster 1 was restricted to the basal lineage (Fig 3D). Likewise, Cluster 6 was restricted to the luminal lineage, giving rise to Cluster 4, which scored further along the luminal differentiation dimension (Fig 3E). Together, these results suggest that human breast tissues that grow in 3D retain a stem cell hierarchy with slow dividing bipotent stem cells reside at the apex of this hierarchy that give rise to lineage-restricted luminal and basal progenitors, whose progeny ultimately become mature luminal or basal cells.

Having assessed the epithelial hierarchy and their linear relationship to differentiated cells in normal human breast tissue, we next asked how these states were affected by DDR1 activity. Inhibition

of DDR1 blocked the differentiation of stem cells; there was a dramatic increase the proportion of bipotent stem cells, with a concomitant decrease in the number of mature basal cells (Fig 3F). In addition, inhibition of DDR1i activity trapped luminal cells in a proliferative progenitor state (Fig 3G, Fig S3D, Table 1). The fraction of luminal cells increased relative to vehicle treated and there was a significant increase in the fraction of proliferative luminal progenitors (Fig 3G). Upon release from DDR1 inhibition for the final two days of culture, expression levels of genes associated with mammary epithelial differentiation and early development were enriched (Fig 3H). The fraction of bipotent stem cells decreased and the number of basal progenitors increased (Fig 3E). In addition, we observed a trend towards an increase in mature luminal cells with concomitant decrease in luminal progenitors. Collectively, these results indicate that DDR1 is required for the differentiation of basal progenitors from bi-potent stem cells as well as the differentiation of mature luminal cells from luminal progenitor cells (Fig 3I).

DDR1 affects luminal progenitor cell differentiation by activating Notch

Knowing that DDR1 inhibition blocked differentiation of both MaSCs into basal progenitors as well as LPs into mature cells, we next set out to understand how DDR1 transduced its actions from the basal to luminal cell lineages to coordinate luminal cell maturation and lobular growth (Fig 2D, 3J). Notably, we found that NOTCH1 target genes are significantly upregulated upon release from DDR1 inhibition in DDR1i Rel tissues (Fig 4A). As Notch signaling is required for luminal cell and lobular differentiation (9), we next investigated what might be driving this change in activation of Notch signaling. Notch signaling is a conserved pathway that regulates cell fate decisions and is initiated by binding of a transmembrane ligand (Jagged (*JAG*) or Delta-like (DLL)) from one cell to a Notch receptor expressed on adjacent cells. We examined expression of the Notch receptor ligand, JAG1 and DLL and found that, upon inhibition of DDR1, the expression of JAG1 decreased dramatically in basal cells (Fig 4B). Conversely, release from DDR1 inhibition increased JAG1 expression only in basal progenitor and mature basal cells, but not stem cells or luminal cells (Fig 4C). This basal lineage-restricted upregulation of JAG1 was concomitant with luminal lineage-restricted upregulation of NOTCH1 target genes. This finding indicates that the matrix-cell interface stimulates JAG1 expression in basal cells through DDR1 activity, which in turns triggers Notch signaling in adjacent luminal cells.

To directly test this, we stimulated MCF10A cells with collagen to activate DDR1 and assess Notch activity. DDR1 activation and phosphorylation was induced upon collagen stimulation as was activation of Notch cleavage (Fig. 4D). Furthermore, addition of DDR1 inhibitor blocked Notch cleavage even in the presence of collagen. Similarly, collagen failed to induce Notch cleavage in cells lacking

DDR1 (DDR1KO) (Fig 4D). Together, these results demonstrate that DDR1 activation by collagen is required for Notch signaling.

Next, we examined whether failure of lobular/alveolar differentiation upon DDR1 inhibition was due to the lack of Notch activation. Accordingly, we treated patient-derived tissues grown in hydrogels with a γ-secretase inhibitor (GSI), which prevents Notch cleavage and nuclear translocation (19). Indeed, in the presence of GSI, led to a failure of lobular/alveolar formation and caused the development of breast tissues that contained only ducts, phenocopying the differentiation defect upon inhibition of DDR1 (Fig 4E, 2C). Taken together, these data indicate that DDR1 is required for the differentiation of bipotent mammary stem cells into lineage-committed basal progenitors. Lineage-restricted basal cells express the Notch ligand, JAG1, which is required for activation of Notch signaling in luminal progenitor cells. This activation of luminal Notch signaling, in turn, induces the maturation of luminal progenitors into mature lobular luminal cells (Fig 4F).

Chapter 4.4: Methods

Ethics statement

Primary tissues that would otherwise have been discarded as medical waste following surgery were obtained in compliance with all relevant laws, using protocols approved by the institutional review board at Maine Medical Center. All tissues were anonymized before transfer and could not be traced to specific patients; for this reason, this research was provided exemption status by the Committee on the Use of Humans as Experimental Subjects at the Massachusetts Institute of Technology. All patients enrolled in this study signed an informed consent form to agree to participate in this study and for publication of the results.

Cells culture and preparation of primary patient-derived tissue

MCF10A cells were obtained from ATCC and cultured in MEGM (Lonza CC-3150) supplemented with 100 ng/ml cholera toxin (Sigma-Aldrich), 1X GlutaMax, and 1X Penicillin and Streptomycin (Gibco).

Reduction mammoplasty tissue samples were mechanically dissociated and then incubated with 3 mg/ml collagenase (Roche Life Science, Indianapolis, IN, USA) and 0.7 mg/ml hyaluronidase (Sigma-Aldrich, St. Louis, MO, USA) at 37 °C overnight. Epithelial clusters were disrupted by trituration, washed, and depleted for fibroblasts.

3D culture

Hydrogels were prepared as previously described (5). Collagen gels were prepared as previously described (10).

CRISPR-Cas9 screen

Cells were transduced with Cas9 (No. 50661, Addgene) and, subsequently, with a kinase targeting sgRNA library (No. 51044, Addgene) as previously described (11). pCW-Cas9 was a gift from Eric Lander & David Sabatini (Addgene plasmid # 50661). Human CRISPR enriched pooled library was a gift from David Sabatini & Eric Lander (Addgene # 51044). Cells were then seeded and grown in collagen matrices for 10 days. Then the collagen pads were collected and incubated in 100 ug/ml collagenase in PBS at 37°C for 10 minutes. The structures were collected by centrifugation (500 RPM, 5 min), trypsinized with 0.25% trypsin for 20–25 minutes at 37°C. Cells were counted in trypan blue, spun down (500RPM, 5 min), and resuspended in MCF10A media; 7500 living cells were reseeded into a new collagen pad. After growth in the new collagen pads for 7 days, collagen pads were harvested to collect genomic DNA. Libraries were prepared as previously described and submitted for sequencing with a

HiSeq 2500. Subsequently, libraries were deconvoluted as previously described (11). To calculate sgRNA abundance for each sample, sgRNA reads were normalized to total reads from the sample. RIGER was used to calculate an FDR corrected significance via the "second best" metric as previously described (20). sgRNA representation in secondary organoids was calculated by subtracting the log2 transformed sgRNA abundance in 3D by the log2 transformed sgRNA abundance in 2D.

Colony assay

Patient-derived mammary tissues were grown in hydrogels for two weeks in the presence or absence of DDR1 inhibitor (2 μM; No. 5077, Tocris) (21). Then cells were dissociated to single cells and plated, cultured, fixed, and stained as previously described, with a modified culture time of one week (10).

Lentivirus production and infection

Lentivirus production was performed as previously described (22).

Immunofluorescence

Immunofluorescence was performed as previously described (5). Primary antibodies and stains used in this study were: DAPI (no. D1306, Life Technologies) and Phalloidin-AF647 (no. A22287, Life Technologies).

Microscopy

Images were captured using a Zeiss LSM 700 (immunofluorescence; Zeiss Microscopy, Thornwood, NY, USA), a Zeiss LSM 710 (immunofluorescence; Zeiss Microscopy), and a Zeiss Axiophot 25 (brightfield; Zeiss Microscopy).

Western Blotting

Western blotting was performed as previously described (28). Primary antibodies include anti-total DDR1 (no. 5583; Cell Signaling Technology), anti-phospho-DDR1 (no. 14531; Cell Signaling Technology), anti-cleaved NOTCH1 (no. 4147; Cell Signaling Technology), and anti–β-Tubulin (no. 5346; Cell Signaling Technology).

Flow Cytometry

Flow cytometry analysis was performed according to the manufacturer's protocol (BD Biosciences), with at least 10000 events captured per analysis. Antibodies used were as follows: PE-

conjugated anti-CD49f antibody (BDB555736, BD Biosciences), APC-conjugated anti-EPCAM antibody (BDB347200, BD Biosciences).

Seq-Well single-cell RNA sequencing

Libraries were prepared from single cells from dissociated hydrogel-grown tissues that were isolated and barcoded using Seq-Well as previously described (23). Sequencing was carried out on an Illumina NextSeq 500 to achieve paired end reads.

scRNA-seq analysis

Read alignment was performed as previously described (23). Briefly, reads were aligned to GRCh38, and individual reads from read 2 were tagged by their corresponding read 1 pair using the 12-bp cellular barcode and the 8-bp UMI. Afterwards, reads were deconvoluted to individual cells using Drop-seq tools (http://mccarrolllab.com/dropseq). Barcodes and UMIs were collapsed using a Hamming distance of 1 to obtain a digital gene expression matrix. For each cell within the digital gene expression matrix, UMI-collapsed gene expression was normalized by scaling by the total number of transcripts and multiplying by a factor of 10000. Scaled gene expression was then natural-log transformed.

The Seurat package was used to perform clustering analysis (24). First, cells with less than 500 genes and with mitochondrial content greater than 7.5% were removed from the analysis. Genes detected in less than 3 cells were dropped from the analysis. 7193 highly variable genes, selected using the MeanVarPlot function in Seurat with a low cutoff of 0.0125 and a high cutoff of 5 for dispersion and an average expression cutoff of 0.4, were used to perform principle component analysis. The top 8 principle components were used to perform t-distributed stochastic neighbor embedding (tSNE) analysis. Clusters were called using the FindClusters function with a resolution of 1 (24). Clusters corresponding the indicated clusters were classified using gene set enrichment analysis (GSEA) with the input being a pre-ranked list of gene enrichment in a specific cluster relative to all the remaining cells (25). NOTCH1 peaks were defined from the NOTCH1 dynamics peaks bed file from Wang et al (26). Genes associated with peaks were defined by HOMER using the annotatePeaks command (27). Additionally, to calculate p-values for individual genes, we applied the likelihood ratio test (LRT). To highlight cell cycle genes, we used genes from Macasko et al (23). To identify significantly enriched overlaps, we used the piano package to perform overlap analysis on Bonferroni corrected significant genes. Significant genes were selected from the DDR1-inh treated cells relative to the cells from the remaining conditions and from the DDR1-inh released cells relative to the remaining conditions. The Monocle package was used to perform lineage trajectory reconstruction using the default settings (28). Cells from clusters identified from the Seurat analysis were colored in lineage trajectory reconstruction based on their respective clusters.

Statistics

All statistics, excluding those performed for single cell and screen analyses, were performed using GraphPad Prism.

Chapter 4.5: Acknowledgements

We gratefully acknowledge Dr. Wendy Salmon and the Keck Imaging Facility at the Whitehead Institute for microscopy services. We thank the Maine Medical Center for human tissue samples and the donors. We thank Dr. Stuart Levine and the MIT BioMicro Center for high throughput sequencing services. We thank Tim Wang and Dr. David Sabatini for enriched CRISPR pools. We thank Dr. Joan Brugge for the helpful discussions. This research was supported in part through the NSFGRFP (1122374; ESS) and by the Whitehead Institute.

Chapter 4.6: References

1. Rios AC, Fu NY, Lindeman GJ, & Visvader JE (2014) In situ identification of bipotent stem cells in the mammary gland. *Nature* 506(7488):322-327.
2. Wang D, *et al.* (2015) Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* 517(7532):81-84.
3. Shackleton M, *et al.* (2006) Generation of a functional mammary gland from a single stem cell. *Nature* 439(7072):84-88.
4. Stingl J, *et al.* (2006) Purification and unique properties of mammary epithelial stem cells. *Nature* 439(7079):993-997.
5. Sokol ES, *et al.* (2016) Growth of human breast tissues from patient cells in 3D hydrogel scaffolds. *Breast Cancer Research* 18(19).
6. Gierahn TM, *et al.* (2017) Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nature methods* 14(4):395-398.
7. Bouras T, *et al.* (2008) Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. *Cell stem cell* 3(4):429-441.
8. Lafkas D, *et al.* (2013) Notch3 marks clonogenic mammary luminal progenitor cells in vivo. *The Journal of cell biology* 203(1):47-56.
9. Rodilla V, *et al.* (2015) Luminal progenitors restrict their lineage potential during mammary gland development. *PLoS biology* 13(2):e1002069.
10. Sokol ES, *et al.* (2015) Perturbation-expression analysis identifies RUNX1 as a regulator of human mammary stem cell differentiation. *PLoS computational biology* 11(4):e1004161.
11. Tim Wang JJW, David M. Sabatini, Eric S. Lander (2014) Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science* 343.
12. Zucchi I, *et al.* (2004) Association of rat8 with Fyn protein kinase via lipid rafts is required for rat mammary cell differentiation in vitro. *Proceedings of the National Academy of Sciences of the United States of America* 101(7):1880-1885.
13. Wang W, *et al.* (2013) Extracellular signal-regulated kinase 5 (ERK5) mediates prolactin-stimulated adult neurogenesis in the subventricular zone and olfactory bulb. *The Journal of biological chemistry* 288(4):2623-2631.
14. Zhang Y, O'Keefe RJ, & Jonason JH (2017) BMP-TAK1 (MAP3K7) Induces Adipocyte Differentiation Through PPARgamma Signaling. *Journal of cellular biochemistry* 118(1):204-210.
15. Barker KT, *et al.* (1995) Expression patterns of the novel receptor-like tyrosine kinase, DDR, in human breast tumours. *Oncogene* 10(3):569-575.

16. Vogel WF, Aszodi A, Alves F, & Pawson T (2001) Discoidin Domain Receptor 1 Tyrosine Kinase Has an Essential Role in Mammary Gland Development. *Molecular and Cellular Biology* 21(8):2906-2917.
17. Visvader JE & Stingl J (2014) Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes & development* 28(11):1143-1158.
18. Mona Shehata AT, Gemma Sharp, Nikola Novcic, I Alasdair Russell, Stefanie Avril, & Michael Prater PE, Carlos Caldas, Christine J Watson and John Stingl (2012) Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Research* 14(134).
19. Geling A, Steiner H, Willem M, Bally-Cuif L, & Haass C (2002) A gamma-secretase inhibitor blocks Notch signaling in vivo and causes a severe neurogenic phenotype in zebrafish. *EMBO reports* 3(7):688-694.
20. Luo B, *et al.* (2008) Highly parallel identification of essential genes in cancer cells. *Proceedings of the National Academy of Sciences of the United States of America* 105(51):20380-20385.
21. Kim H-G, *et al.* (2013) Discovery of a Potent and Selective DDR1 Receptor Tyrosine Kinase Inhibitor. *ACS Chemical Biology* 8(10):2145-2150.
22. Feng YX, *et al.* (2017) Cancer-specific PERK signaling drives invasion and metastasis through CREB3L1. *Nature communications* 8(1):1079.
23. Macosko EZ, *et al.* (2015) Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161(5):1202-1214.
24. Satija R, Farrell JA, Gennert D, Schier AF, & Regev A (2015) Spatial reconstruction of single-cell gene expression data. *Nature biotechnology* 33(5):495-502.
25. Subramanian A, *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* 102(43):15545-15550.
26. Wang H, *et al.* (2014) NOTCH1-RBPJ complexes drive target gene expression through dynamic interactions with superenhancers. *Proceedings of the National Academy of Sciences of the United States of America* 111(2):705-710.
27. Heinz S, *et al.* (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular cell* 38(4):576-589.
28. Qiu X, *et al.* (2017) Reversed graph embedding resolves complex single-cell trajectories. *Nature methods* 14(10):979-982.
29. Wong DJ, *et al.* (2008) Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell stem cell* 2(4):333-344.

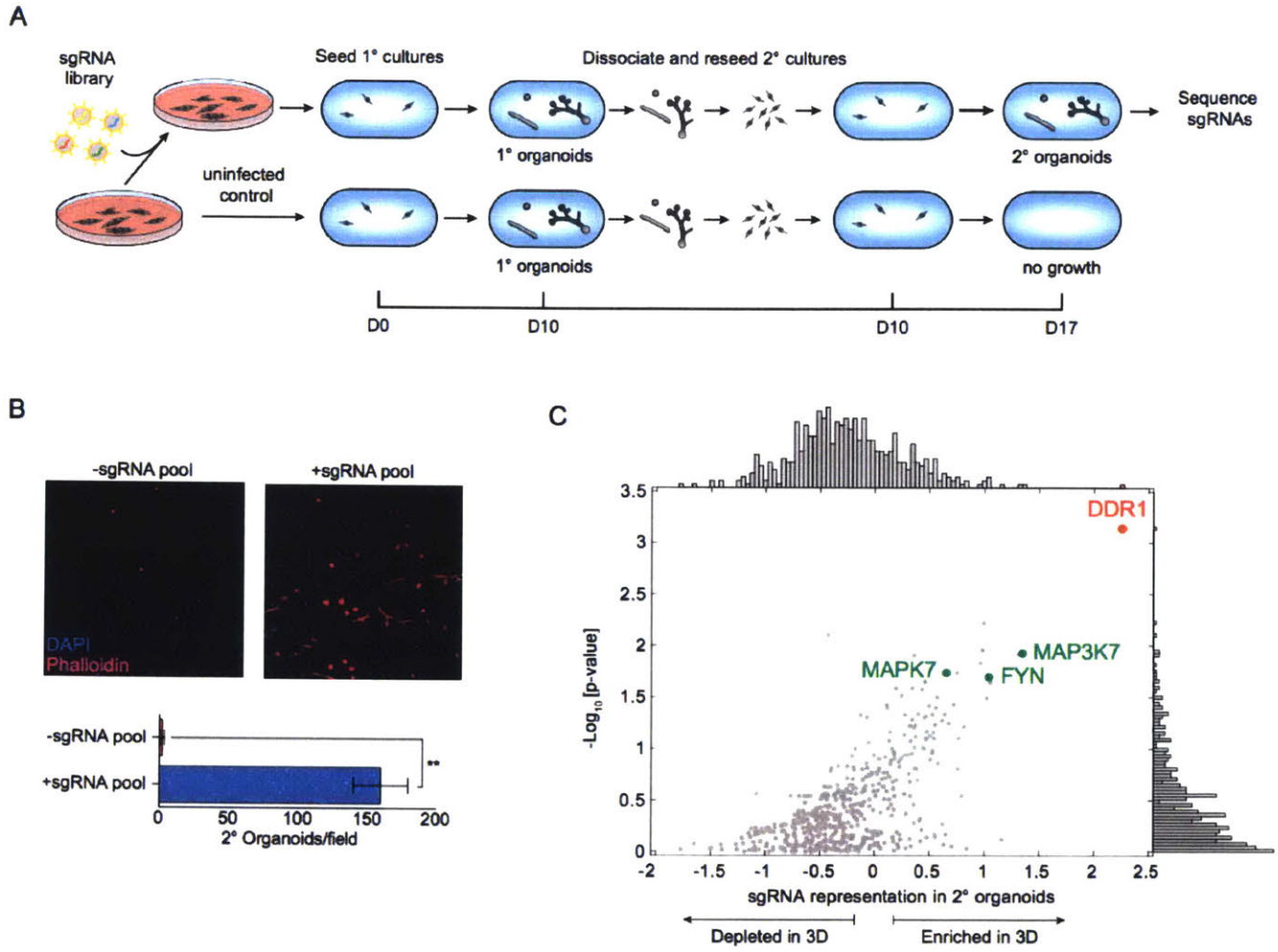


Figure 1. A CRISPR screen in mammary organoids for regulators of stem/progenitor cell differentiation. (A) Schematic of pooled CRISPR screening strategy in organoids. After ten days of 3D culture, primary organoids are extracted from the collagen matrix, dissociated to single cells, and reseeded into secondary cultures. Note that uninfected control cells differentiate in primary 3D culture, and are incapable of seeding secondary organoids. (B) Representative images (top) of phalloidin stained secondary organoids from MCF10A cells containing Cas9 without or with sgRNAs, and quantification (bottom) of the number of secondary organoids per field. (C) Screened kinases scored by significance relative to a null distribution using RIGER (y-axis) and by comparing the mean differential abundance of sgRNAs targeting the kinase (x-axis). DDR1 (red dot) was the top hit by both metrics. Adjacent histograms indicate the distribution of p-values (right y-axis) and mean differential (top x-axis). MAPK7, MAP3K7 and FYN (green dots) indicate significant genes that have been previously implicated in cellular differentiation. ** indicates $p<0.01$ (Student's *t*-test).

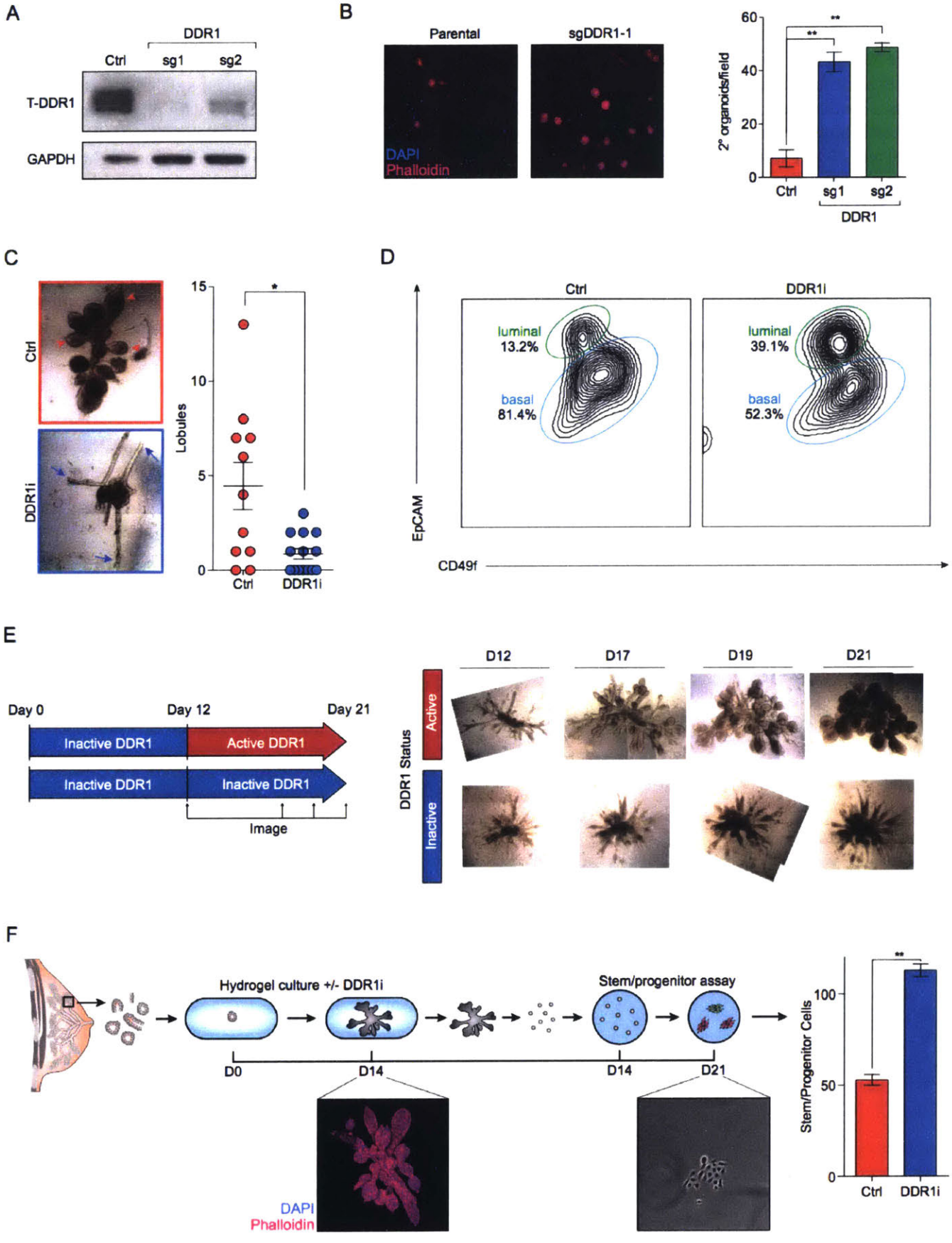
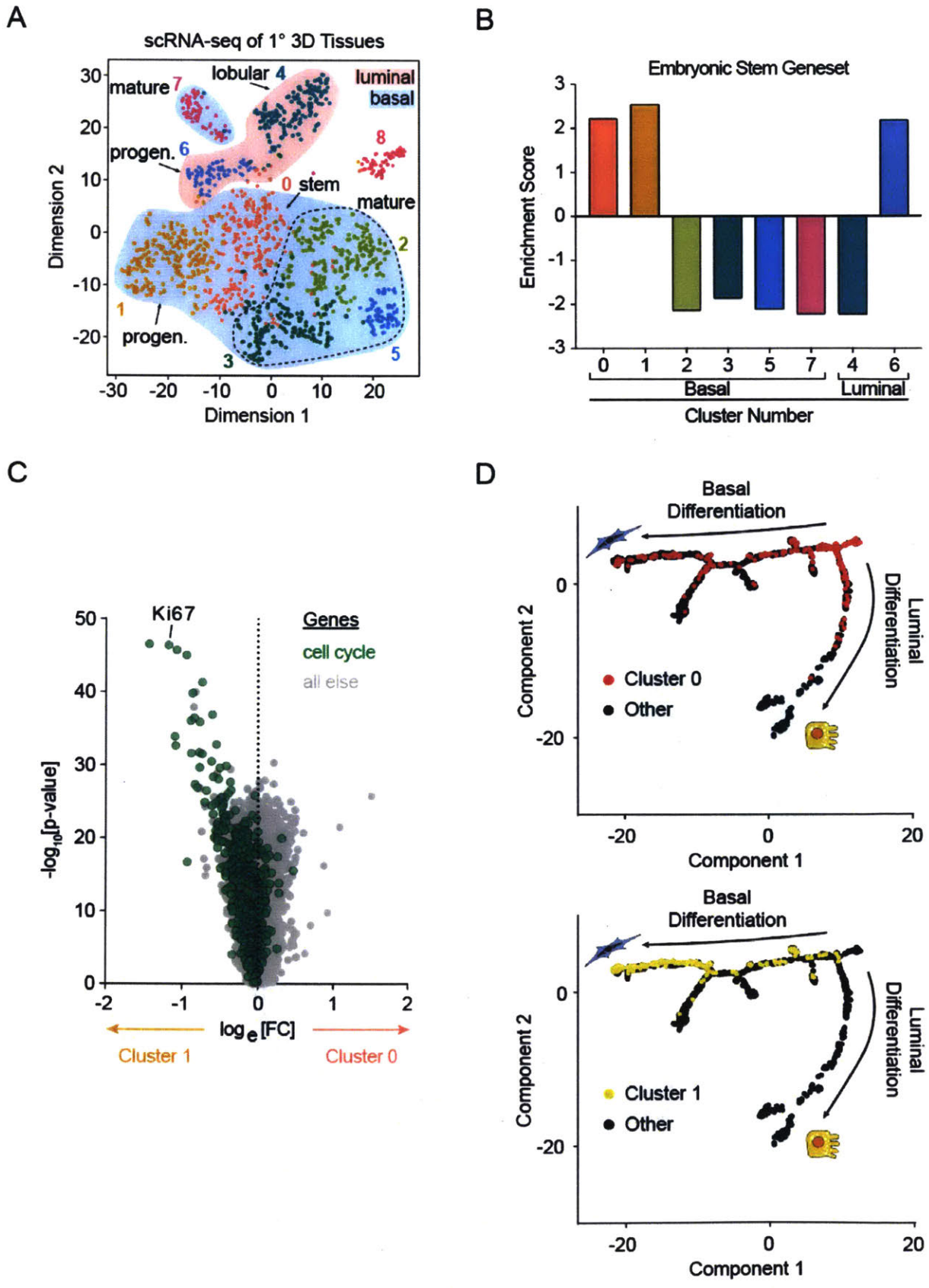

A
DDR1
Ctrl
sg1
sg2
T-DDR1
GAPDH
B
Parental
sgDDR1-1
DAPI
Phalloidin
2° organoids/field
60
40
20
0
**
**
Ctrl
sg1
sg2
DDR1
C
Ctrl
DDR1i
Lobules
15
10
5
0
*
Ctrl
DDR1i
D
Ctrl
DDR1i
luminal
13.2%
basal
81.4%
luminal
39.1%
basal
52.3%
EpCAM
CD49f
E
Day 0
Day 12
Day 21
Inactive DDR1
Active DDR1
Inactive DDR1
Inactive DDR1
Image
D12
D17
D19
D21
DDR1 Status
Active
Inactive
F
Hydrogel culture +/- DDR1i
Stem/progenitor assay
D0
D14
D14
D21
DAPI
Phalloidin
Stem/Progenitor Cells
100
50
0
**
Ctrl
DDR1i

Figure 2. DDR1 is required for mammary stem/progenitor cell differentiation. (A) Western blot for total DDR1 in parental cells or after knocking out DDR1 in pools of MCF10As with two independent sgRNAs. GAPDH was used as a loading control. (B) Fluorescence images (left) of secondary organoids stained with phalloidin (pink) and DAPI (blue) and quantification (right) of organoids from cells in (A). (C) Bright-field images (left) of control and DDR1-inhibited tissues grown in 3D hydrogel cultures for 13 days and quantification (right) of lobules per condition. Red arrowheads indicate lobules, blue arrows indicate ducts that terminated without a lobule. (D) Flow-cytometry analysis of EpCAM and CD49f expression on cells from patient-derived tissues grown in hydrogels for 12 days without DDR1-inhibition (left) and with DDR1-inhibition (right). (E) Schematic of DDR1-inhibition withdrawal experiments (left) and representative bright-field images from four time points after withdrawing DDR1-inhibitor. (F) Schematic (top) of progenitor assay in primary mammary epithelial cells with or without chemical inhibition of DDR1 and quantification (right) of stem/progenitor cells. Representative images of vehicle-treated 3D cultured patient tissues after 14 days and colonies after 7 days of culture are shown below. * indicates $p<0.05$; ** indicates $p<0.01$ (Student's *t*-test).


A
scRNA-seq of 1° 3D Tissues
mature 7
lobular 4
luminal
basal
progen. 6
8
0 stem
mature
2
1
progen.
3
5
Dimension 2
Dimension 1
B
Embryonic Stem Geneset
Enrichment Score
0 1 2 3 5 7 4 6
Basal
Luminal
Cluster Number
C
Ki67
Genes
cell cycle
all else
-log10[p-value]
log e [FC]
Cluster 1
Cluster 0
D
Basal
Differentiation
Luminal
Differentiation
Cluster 0
Other
Component 2
Component 1
Basal
Differentiation
Luminal
Differentiation
Cluster 1
Other
Component 2
Component 1

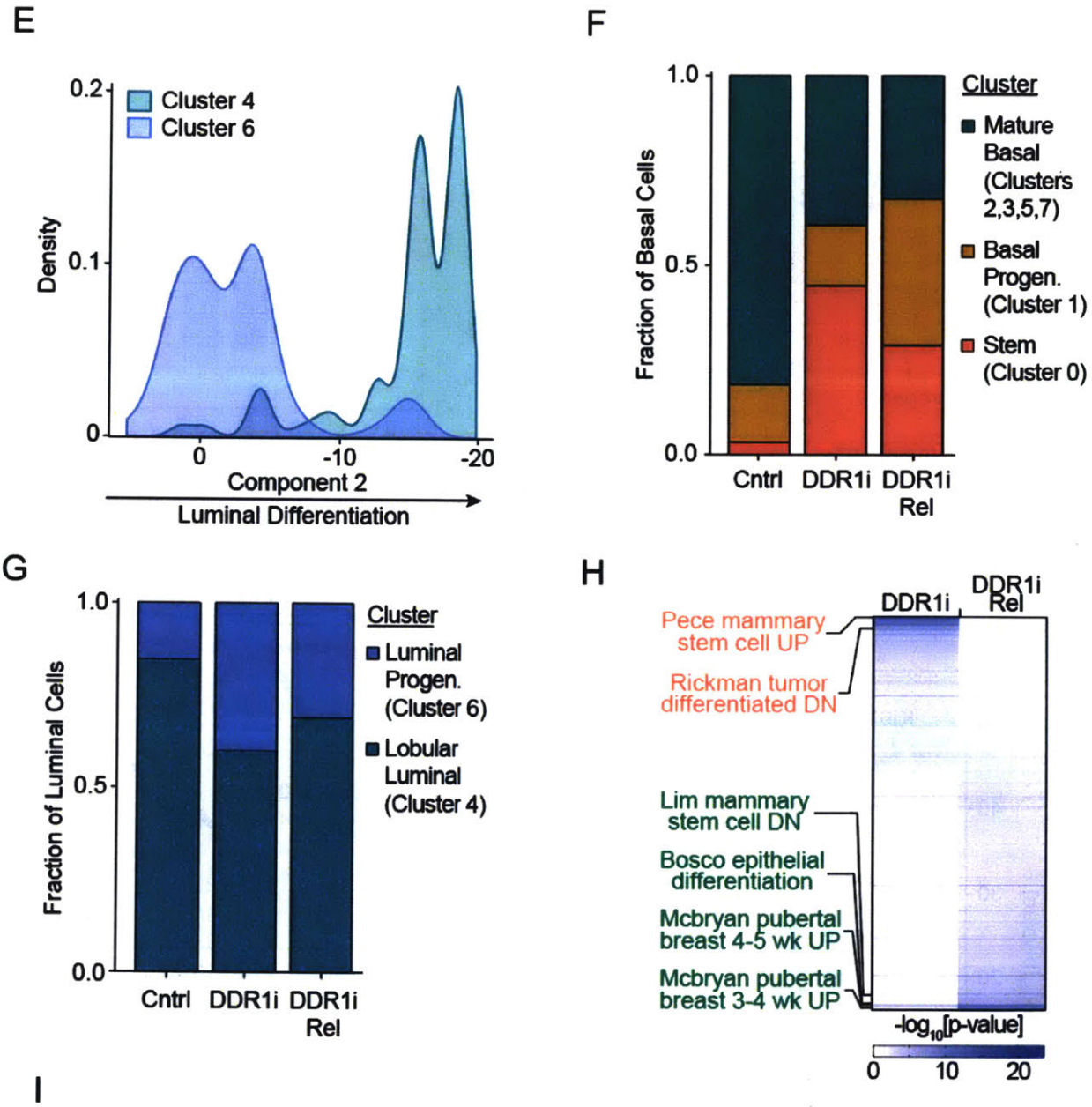

E
Cluster 4
Cluster 6
Density
0.2
0.1
0
0
-10
-20
Component 2
Luminal Differentiation
F
Fraction of Basal Cells
1.0
0.5
0.0
Cluster
Mature Basal (Clusters 2,3,5,7)
Basal Progen. (Cluster 1)
Stem (Cluster 0)
Cntrl
DDR1i
DDR1i Rel
G
Fraction of Luminal Cells
1.0
0.5
0.0
Cluster
Luminal Progen. (Cluster 6)
Lobular Luminal (Cluster 4)
Cntrl
DDR1i
DDR1i Rel
H
DDR1i
DDR1i Rel
Pece mammary stem cell UP
Rickman tumor differentiated DN
Lim mammary stem cell DN
Bosco epithelial differentiation
Mcbryan pubertal breast 4-5 wk UP
Mcbryan pubertal breast 3-4 wk UP
-log10[p-value]
0
10
20
I
DDR1
Basal Progenitor (Cluster 1)
Mature Basal (Clusters 2, 3, 5, 7)
MaSC (Cluster 0)
DDR1
Luminal Progenitor (Cluster 6)
Mature Luminal (Cluster 4)

Figure 3. scRNA-seq of primary tissues shows cell-state changes mediated by DDR1. (A) scRNA-seq data from all cells (n=1467) from patient-derived hydrogel-grown tissues projected onto two dimensions using t-SNE on the top eight principal components across 7193 variable genes. Each dot represents a single cell and is colored according to the cluster to which it is assigned. Luminal and basal epithelial clusters are shaded red and blue, respectively. (B) Bar chart of enrichment of a stem cell gene set calculated by GSEA in each of the basal and luminal epithelial clusters defined in (A) (29). Colors denote the clusters indicated in (A). (C) Volcano plot visualizing differential gene expression between cluster 0 and cluster 1 from (A). For each gene, the average expression difference is plotted against the statistical significance, as calculated by the likelihood ratio test for differential expression. Cell cycle genes are highlighted in green (23). MKI67 (Ki67) is labeled individually. (D) Inferred lineage relationships of all cells were projected onto two dimensions as differentiation trajectories, using Monocle. These trajectories represent the basal and luminal differentiation paths, as indicated. Cluster 0 (red; left panel) and Cluster 1 (yellow; right panel) cells are highlighted. (E) Distribution of cells from Cluster 4 and Cluster 6 along the luminal differentiation axis produced by Monocle analysis in (D) and plotted as a kernel density plot. (F) Stacked bar chart indicating the distribution of single cells from (A) among the indicated basal cell types from control, DDR1-inhibited, and DDR1-inhibitor released tissues. Cluster numbers are indicated on the right of the chart. (G) Stacked bar chart indicating the representation of luminal cell types from control, DDR1-inhibited, and DDR1-inhibitor released tissues. Cluster numbers are indicated to the right of the chart. (H) Heatmap of enriched gene sets using GSEA on DDR1i and DDR1i Rel tissues from (A). (I) Proposed model for DDR1's role in mammary epithelial cell differentiation.

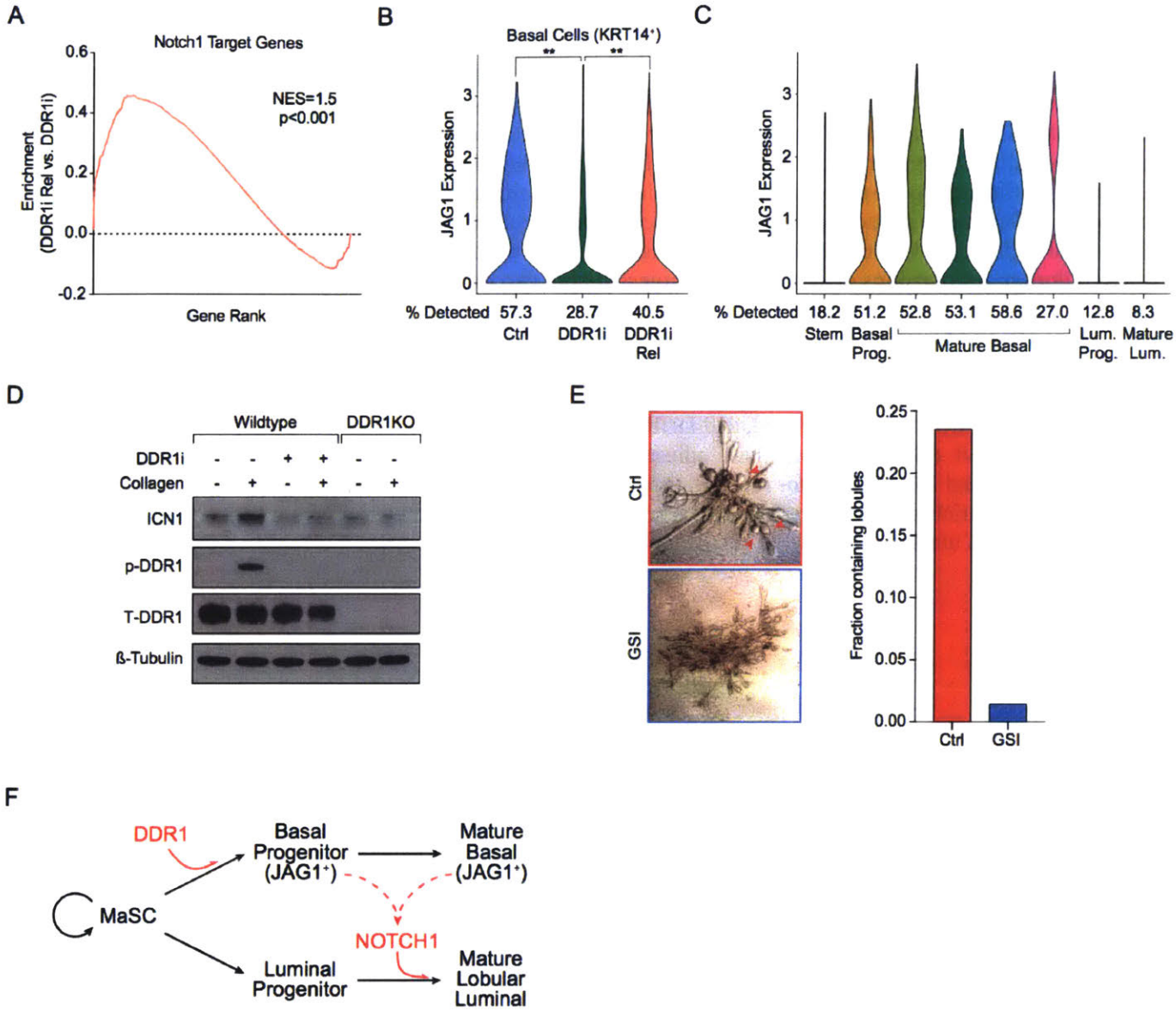
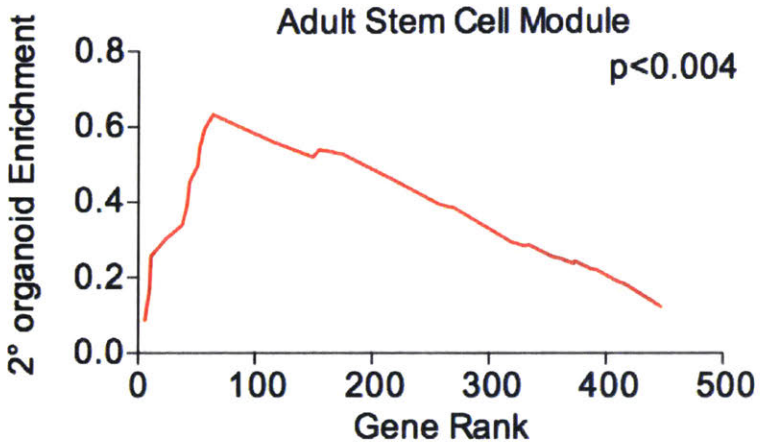


Figure 4. DDR1 signaling activates NOTCH1 to drive lobular differentiation. (A) Representative GSEA plot of NOTCH1 target gene expression in cells from Figure 3A from the DDR1I Rel tissues relative to DDR1i tissues (26). (B) Violin plot depicting JAG1 expression in basal cells from control, DDR1-inhibited and DDR1-inhibitor released tissues. The percentage of cells in which the JAG1 transcript was detected is indicated below the plot. (C) Violin plot showing the distribution of JAG1 expression in the epithelial clusters from Figure 3A. The percentage of cells in which JAG1 transcript was detected is indicated below the plot. (D) Western blot for phospho-DDR1, Total-DDR1, and cleaved Notch1 (ICN1) under the indicated conditions. Beta-tubulin was used as a loading control. (E) Bright-field images (left) of representative control and GSI-treated patient-derived tissues grown in hydrogels for 12 days, and quantification (right) of the fraction of tissues containing lobules. Red arrowheads indicate lobules. (F) Revised model for the role of DDR1 and Notch within the mammary differentiation hierarchy. Dashed red lines indicate interactions. ** indicates $p < 0.01$ (Student's *t*-test).

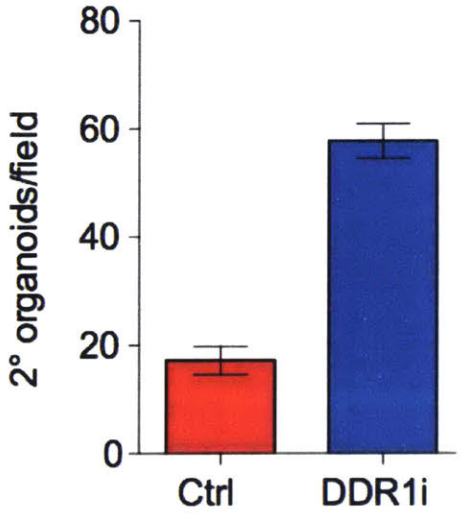
Cluster	Lineage	Stem Cells Genes	Prolif	Differentiation Potential	Cell State
0	Basal (KRT14$^+$)	Upregulated	-	Lum + Basal	Bipotent Stem
1	Basal (KRT14$^+$)	Upregulated	+	Basal	Basal Progenitor
2,3,5,7	Basal (KRT14$^+$)	Downregulated	-	Basal	Mature Basal
4	Luminal (KRT8$^+$)	Downregulated	-	Luminal	Mature Luminal
6	Luminal (KRT8$^+$)	Upregulated	+	Luminal	Luminal Progenitor

Table 1. Mammary cell states corresponding to scRNA-seq clusters. Summary of cluster identification using: lineage information derived from cytokeratin expression; stem cell gene expression enrichment determined by GSEA; proliferative potential based on Ki67 expression and cell cycle gene enrichment; and differentiation potential determined by the presence of cells within a cluster along lineage reconstruction paths by Monocle. Alongside the summaries of cluster characteristics are the identities of the mammary cell states to which these clusters were assigned.

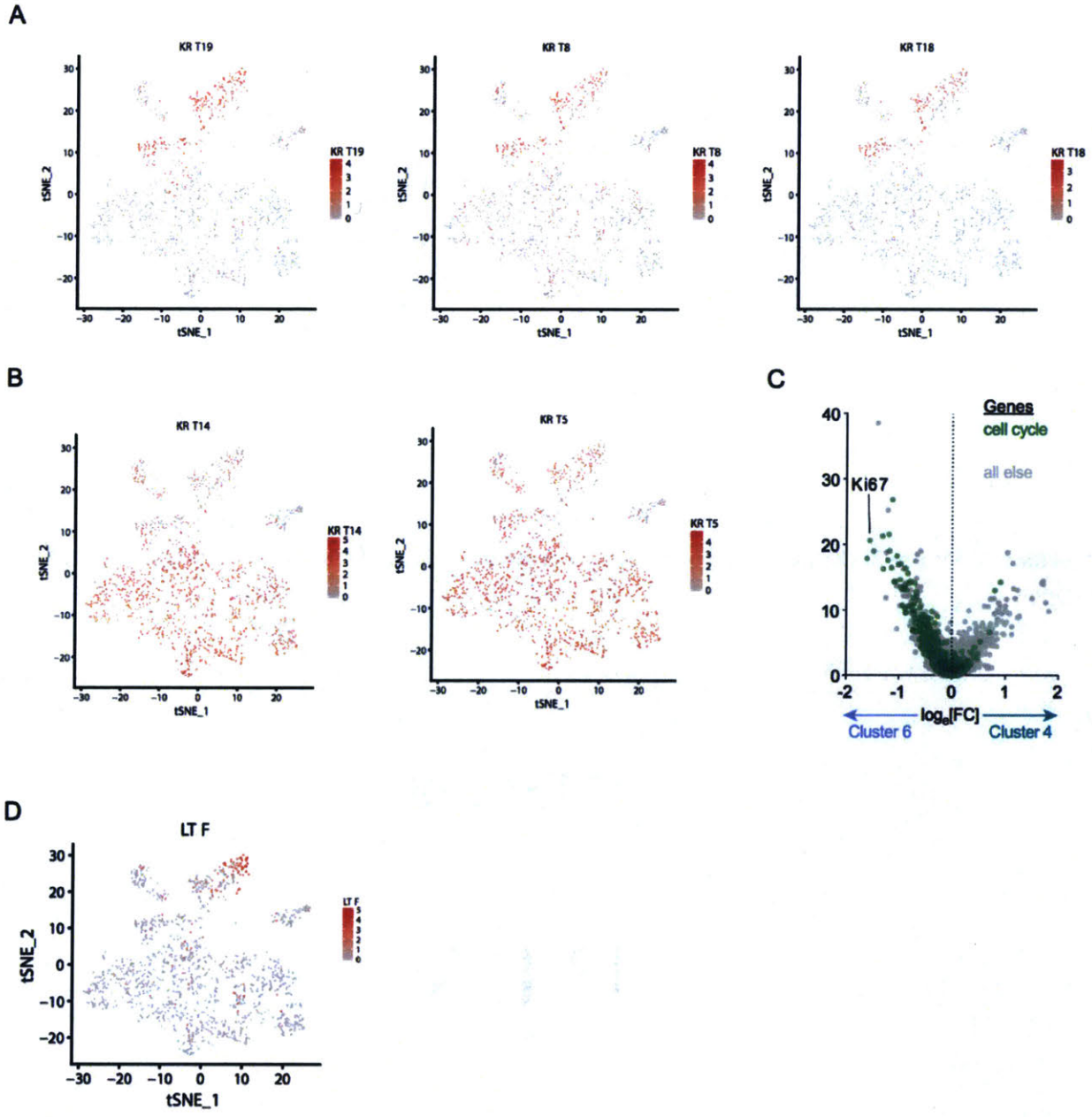
Chapter 4.8: Supplemental Data



Supplemental Figure 1. Representative GSEA plot for stem cell signatures using ranked screen hits.



Supplemental Figure 2. Quantification of secondary organoid seeding efficiency with or without DDR1 inhibition.



Supplemental Figure 3. scRNA-seq expression (A) Single cell expression of luminal cytokeratins overlayed as a heatmap on the tSNE plot from Fig. 3A. (B) Single cell expression of basal cytokeratins overlayed as a heatmap on the tSNE plot from Fig. 3A. (C) Volcano plot visualizing differential gene expression between cluster 4 and cluster 6. Ki67 is labeled individually. (D) Single cell expression of LTF overlayed as a heatmap on the tSNE plot from Fig. 3A.

Chapter 5
Conclusions

Chapter 5.1: Summary

This thesis set out to identify factors that are involved in breast cancer progression. To do so, I took two approaches: (1) to screen for factors directly associated with invasion and metastasis and (2) to search for factors involved in mammary differentiation. While the latter may seem separate, many of these sorts of factors have provided mechanistic insights into breast cancer progression (1, 2). At the time that this work had begun, it was noted that more than 50% of DCIS lesions never progress to invasive cancers (3, 4). Yet, because DCIS and IDC are so challenging to distinguish, the factors that promote DCIS invasion were poorly understood. Here, we show that SMARCE1 is required for the invasive progression of DCIS and other early-stage tumors. We also show that SMARCE1 drives invasion by regulating the expression of secreted proteases that degrade basement membrane. Functionally, we show that SMARCE1 promotes invasion of *in situ* cancers that grow within primary human mammary tissues and is also required for metastasis *in vivo*. Mechanistically, we find that SMARCE1 drives invasion by forming a SWI/SNF-independent complex with the transcription factor ILF3. In patients with early-stage cancers, SMARCE1 expression is a strong predictor of eventual relapse and metastasis. These findings establish SMARCE1 as a key driver of invasive progression in early-stage tumors.

Additionally, given the limited selection of treatments available for invasive breast cancers at the start of this thesis, it was of great interest to identify a more targeted approach towards treating invasive breast cancers. Previous studies have shown the benefits of targeting the PERK pathway, but chemical inhibitors of PERK cause on-target side effects (5). We overcome these challenges by identifying the transcription factor, CREB3L1, as an essential mediator of the pro-metastatic functions of PERK in breast cancer. We find that CREB3L1 specifically acts downstream of PERK in the mesenchymal subtype of triple-negative breast cancer. Functionally, we find that CREB3L1 is required for cancer cell invasion and metastasis using genetic and pharmacological approaches. We also find that CREB3L1 expression is predictive of distant metastasis in patients with the mesenchymal subtype of triple-negative breast cancer.

Finally, in this thesis, I have report that the collagen receptor DDR1 coordinates stem cell differentiation and heterotypic Jagged1-Notch signaling to drive luminal differentiation and TDLU morphogenesis. Functionally, we show that DDR1 is required for stem cell differentiation and luminal progenitor differentiation. By coupling functional assays with single-cell RNA sequencing (scRNA-seq) of primary human tissues cultured *ex vivo*, we show that, mechanistically, DDR1 promotes basal lineage commitment, which indirectly stimulates luminal Notch signaling to drive lobulogenesis by activating JAG1 expression in basal cells.

Chapter 5.2: Discussions and Future Directions

SMARCE1

In chapter 2 of this thesis, we discover SMARCE1 as a biomarker for breast cancers that will progress. In the process of functionally characterizing SMARCE1, we showed that SMARCE1 regulates an ECM-invasion module. Interestingly, we actually identified two modules. While we focused on the ECM-invasion module given its characteristic gene set, an interesting future direction would be to determine whether or not SMARCE1 regulates this second module. Fortunately, we laid the groundwork to analyze this second module in our study, so this question is answerable using the methods we presented. It would also be of great interest to characterize this module and determine whether this second module is important for aspects of tumorigenesis.

One outstanding question following the work presented in chapter 2 focuses on how SMARCE1 mechanistically activates its loci. While we show that ILF3 recruits SMARCE1 to target pro-invasive genes, it remains unclear how SMARCE1 does so. Prior work on SMARCE1 has indicated that it has an HMG domain which is involved in DNA binding (6). However, it has also been shown that SMARCE1 is dispensable for the remodeling capabilities of the SWI/SNF complex (6). Our work has indicated that SMARCE1 can function independently of the SWI/SNF complex. Therefore, it is still feasible that SMARCE1 can change chromatin accessibility. Since we showed that SMARCE1 is guided to its target loci by ILF3, it would also be interesting to determine if there is a significant overlap in putative SMARCE1 regulated loci and ILF3 targets.

CREB3L1

In chapter 3 of this thesis, we discover CREB3L1 as a potential therapeutic target due to its role in breast cancer progression and owing to its mechanism of activation. However, the chemical inhibitor used in our work is broad acting and treatment of patients with this chemical will present with many side effects. Therefore, it would be of great interest to identify other, more specific, ways to target CREB3L1. One potential approach to identify more specific ways of targeting CREB3L1 is to screen for drugs that may specifically interact with CREB3L1. Studies focused on SRBP1, a protein that is activated in a similar manner to CREB3L1, showed that additional factors provide specificity in targeting SRBP1 activation (7). Therefore, another manner to more directly target CREB3L1 would be to identify potential similar activation specification factors.

At first glance, screening for drugs that block a transcription factor may seem challenging, even with its rather unique mode of activation. However, work by Saito et al has shown that CREB3L1 and CREB3L2 actually secrete their C-terminal domains after they are activated (8). They also showed that it is possible to track the secreted C-terminal domain of CREB3L1, which provides an avenue to measure

CREB3L1 cleavage (8). One could easily screen for factors involved in CREB3L1 processing by tagging the C-terminal domain with a marker such as luciferase. This avenue also benefits in that the tracking could be performed at multiple time points in real time. Many other studies have had success in using these sorts of luciferase-based approaches to perform large-scale screens (9, 10).

Similar approaches can also be used to identify genes involved in CREB3L1 activation. However, since luciferase-based screens cannot be performed in a pooled format, it would be challenging to screen genome-wide. To limit factors that are presumptively involved in CREB3L1 activation, one could take advantage of an immunoprecipitation approach coupled with mass spectrometry to identify proteins bound to CREB3L1. This more limited set of genes could subsequently be individually targeted in the aforementioned luciferase tagged CREB3L1-expressing cells to find the genes that are functionally involved in CREB3L1 activation.

In this thesis, we also show that CREB3L1 can cell non-autonomously promote cell invasion in CREB3L1-lacking cells. In fact, we find that pro-invasive ECM are sufficient to promote invasion in CREB3L1-lacking cells. Recent studies have shown that during the process of invasion, more invasive cells can promote invasion of cells that are typically less invasive. These studies found that more invasive cells confer these properties by creating tracks of ECM that are pro-invasive (11). Given the parallels between our results and these results, it would be interesting to perform co-culture studies to see if typically non-invasive cells could be made to invade in the presence of cells that express high amounts of CREB3L1. Subsequently, one could follow these cells after ablating CREB3L1 to determine if CREB3L1 is necessary for such a phenotype.

Lastly, an important step to establish CREB3L1 as a bonafide therapeutic target is to verify that off-target effects are minimal. When we set out to identify factors downstream of PERK that could mediate its metastatic effects, we sought to avoid side effects such as pancreatic atrophy that are caused despite specifically targeting PERK (5). If the identification of a drug to specifically inhibit CREB3L1 activation is successful, it is essential to show that the *in vivo* side effects are minimal. To do so, one can replicate the *in vivo* drug treatment experiments performed in chapter 3 of this thesis, while also harvesting the pancreas of the treated and control mice to show that such side-effects that were seen in prior studies are not apparent in CREB3L1 inhibited conditions (5).

DDR1

In chapter 4 of this thesis, I explored the function of DDR1 in the normal mammary gland. Interestingly, recent studies have shown that inhibition of DDR1 alongside inhibition of Notch signaling can reduce tumor initiation and that overexpression of DDR1 is important for metastasis (12, 13). At first glance, these results may seem inconsistent with our work, which indicates that inhibition of DDR1

blocks stem and luminal progenitor cell differentiation. Given our model in normal development, one might hypothesize that inhibition of DDR1 in cancer, not activation of DDR1, would result in more aggressive tumors. However, the latter study which followed the role of DDR1 in metastasis had also performed mechanistic work to show that DDR1 functions non-canonically in the cancer context (12). One possibility that could reconcile our results with those of this study is that DDR1 is co-opted to activate novel targets (12). This hypothesis is rather tantalizing because luminal cells are typically thought to be the cells of origin for most breast carcinomas. Since luminal cells are usually not in contact with collagen, activation of DDR1 would be aberrant in luminal cells. While the work in this thesis does show that DDR1 affects luminal differentiation, it does so through heterotypic interactions across multiple lineages by way of JAG1-NOTCH1. Improper activation of DDR1 in luminal cells, which also express DDR1, could create novel networks that may explain such phenotypes as noted by Gao et al (12). Fitting this idea, a recent study that performed single-cell transcriptomic profiling of mouse mammary cells showed that TM4SF1, the factor which Gao et al reports is interacting with DDR1, is more expressed highly expressed in luminal cells (12, 14).

While our work mechanistically shows that DDR1 activates the JAG1-NOTCH1 axis to promote luminal differentiation, it does not focus on how DDR1 mechanistically directs basal fate commitment in stem cells. Since this role in directing basal fate was downstream of the kinase activity of DDR1, it is likely that one or more of the targets of DDR1 is responsible for basal differentiation. To best explore this mechanism, it would be ideal to identify direct targets of DDR1. One manner to do this is to apply an approach pioneered by Shokat et al (15). In this approach, one could mutate the gatekeeping residue of DDR1 to accept a thiolated ATP-analog. Then upon its activation, DDR1, and no other kinase, will phosphorylate its target genes with a thiophosphate instead of a standard phosphate. This thiophosphate can be specifically immunoprecipitated and the immunoprecipitated proteins can be identified by mass spectrometry. After identifying these target proteins, one can functionally show relevance to basal differentiation by performing a forward genetic screen for these genes in mammary cells in a single-cell approach using Perturb-Seq (16).

Chapter 5.3: Concluding Remarks

Given the topics discussed in the prior chapters of this thesis, it is tempting to speculate about potential mechanisms that may connect SMARCE1, CREB3L1, and DDR1. This is furthered by the intriguing fact that these genes are all connected by their relationship to the ECM. One may even posit that there exists a pathway which connects all three genes. The fact that SMARCE1 and CREB3L1 had strikingly similar phenotypes, spanning from their roles in invasion and CTC dissemination to colonization at distant sites, give this potential pathway some credence.

In such a hypothetical pathway, SMARCE1 – which had been identified as a member of the SWI/SNF complex – is likely to be placed it at the top, upstream of CREB3L1 and DDR1. Subsequently, given its role as a transcriptional activator, one might expect CREB3L1 to act upstream of DDR1. Fitting this idea, when CREB3L1 and SMARCE1 were immunoprecipitated with their associated chromatin in the ENCODE project in the same cell line, one can observe a SMARCE1 peak upstream of both CREB3L1 and DDR1. In contrast, CREB3L1 was not observed to bind nearby the SMARCE1 locus. Interestingly, while CREB3L1 was present at the DDR1 locus, it was actually found at a site downstream of DDR1 which was distinct from where SMARCE1 was bound. This suggests that there may be other factors that coordinate the activation of DDR1. While many experiments need to be completed to truly establish a potential pathway involving SMARCE, CREB3L1, and DDR1, connecting these three genes would provide valuable insight into the cellular control and sensing of the ECM.

Chapter 5.4: References

1. Kouros-Mehr H, Slorach EM, Sternlicht MD, & Werb Z (2006) GATA-3 Maintains the Differentiation of the Luminal Cell Fate in the Mammary Gland. *Cell* 127(5):1041-1055.
2. Kas SM, *et al.* (2017) Insertional mutagenesis identifies drivers of a novel oncogenic pathway in invasive lobular breast carcinoma. *Nature genetics* 49(8):1219-1230.
3. Alvarado M, Ozanne E, & Esserman L (2012) Overdiagnosis and overtreatment of breast cancer. *American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Meeting*, pp e40-45.
4. Sanders ME, Schuyler PA, Dupont WD, & Page DL (2005) The natural history of low-grade ductal carcinoma in situ of the breast in women treated by biopsy only revealed over 30 years of long-term follow-up. *Cancer* 103(12):2481-2484.
5. Atkins C, *et al.* (2013) Characterization of a Novel PERK Kinase Inhibitor with Antitumor and Antiangiogenic Activity. *Cancer Research* 73(6):1993.
6. Martens JA & Winston F (2003) Recent advances in understanding chromatin remodeling by Swi/Snf complexes. *Current Opinion in Genetics & Development* 13(2):136-142.
7. Brown MS & Goldstein JL (1999) A proteolytic pathway that controls the cholesterol content of membranes, cells, and blood. *Proceedings of the National Academy of Sciences of the United States of America* 96(20):11041-11048.
8. Saito A, *et al.* (2014) Chondrocyte Proliferation Regulated by Secreted Luminal Domain of ER Stress Transducer BBF2H7/CREB3L2. *Molecular cell* 53(1):127-139.
9. Zheng H, *et al.* (PKD1 Phosphorylation-Dependent Degradation of SNAIL by SCF-FBXO11 Regulates Epithelial-Mesenchymal Transition and Metastasis. *Cancer Cell* 26(3):358-373.
10. Gallagher CM, *et al.* (2016) Ceapins are a new class of unfolded protein response inhibitors, selectively targeting the ATF6α branch. *eLife* 5:e11878.
11. Chapman A, *et al.* (Heterogeneous Tumor Subpopulations Cooperate to Drive Invasion. *Cell Reports* 8(3):688-695.
12. Gao H, *et al.* (2016) Multi-organ Site Metastatic Reactivation Mediated by Non-canonical Discoidin Domain Receptor 1 Signaling. *Cell* 166(1):47-62.
13. Ambrogio C, *et al.* (2016) Combined inhibition of DDR1 and Notch signaling is a therapeutic strategy for KRAS-driven lung adenocarcinoma. *Nature medicine* 22(3):270-277.
14. Bach K, *et al.* (2017) Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. *Nature communications* 8(1):2128.

15. Hertz NT, *et al.* (2010) Chemical Genetic Approach for Kinase-Substrate Mapping by Covalent Capture of Thiophosphopeptides and Analysis by Mass Spectrometry. *Current protocols in chemical biology* 2(1):15-36.

16. Dixit A, *et al.* (2016) Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167(7):1853-1866.e1817.