

The pretenders: Basal breast cancer cells mimic epidermal keratinocytes and rely on retinoic acid signaling for survival, invasion and metastasis.

Andrea E. Doak

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2024

Reading Committee:
Kevin Cheung, Chair
Slobodan Beronja
Jon Cooper

Program Authorized to Offer Degree:
Molecular and Cellular Biology

© Copyright 2024

Andrea E. Doak

University of Washington

Abstract

The pretenders: Basal breast cancer cells mimic epidermal keratinocytes and rely on retinoic acid signaling for survival, invasion and metastasis.

Andrea E. Doak

Chair of Supervisory Committee:

Kevin Cheung

Department of Public Health Science, Fred Hutchinson Cancer Center

In breast cancer, the presence of cells expressing basal epithelial markers is associated with a high risk of tumor invasion, metastasis and therapy resistance. It has therefore been proposed that specifically killing basal breast cancer cells would be beneficial for patients. However, clinically feasible methods for specifically targeting this population of cells have not been reported. This, in part, can be attributed to the limited understanding of the proteins and signaling networks uniquely active in basal tumor cells. Basal cells are difficult to characterize due to their relative rarity in tumors and the transience of their existence. Rather than being a fixed cellular state, basal cells are typically formed

through cellular plasticity, in which non-basal cells convert into basal cells, which can then revert to their original state.

In this work, we enriched for basal cells using organotypic culture and then transcriptionally profiled each subpopulation of cells using single cell RNA-sequencing. We identified a unique transcriptional signature of invasive basal cells and discovered that it had significant overlap with the transcriptional signature of basal epidermal keratinocytes. Given that several transcription factors and pathways are known to regulate keratinocyte identity and function, we hypothesized that the same pathways are functional in basal tumor cells. To test our hypothesis, we first focused on retinoic acid signaling, which can be easily altered using pharmaceutical agents such as retinoids. As anticipated, activation of retinoic acid receptors suppressed transcription of several basal epidermal marker genes, but surprisingly lead to increased invasive behavior in 3D culture. Conversely, retinoic acid receptor antagonism prevented invasion and induced tumor cell death in 3D models of mouse and human breast cancers in which basal cells are present.

We additionally tested the role of KLF4, a transcription factor required for proper epidermal barrier formation, in basal breast cancer cell invasion and viability. In a similar manner to retinoic acid addition, knock-down of KLF4 lead to decreased transcription of epidermal markers and increased rates of invasion in 3D culture. Overexpression of KLF4 induced cell death. Confirming the translational relevance of our findings,

treatment with a retinoic acid receptor inhibitor in-vivo prevented tumor outgrowth and metastasis in-vivo.

Vitamin supplements, including those with the retinoic acid precursor vitamin A, are used by up to 70 percent of cancer patients. Further, retinoic acid has shown some clinical promise as an approach for treating breast cancer patients, although basal-like breast cancers appear to be resistant. Given our results, we wanted to determine whether basal breast cancers are not only resistant to retinoic acid, but are in fact more metastatic in the presence of excess vitamin A. To do so, we performed post-hoc analysis of two nutritionally annotated clinical cohorts of women tracked for over a decade. In these studies, vitamin A supplementation was associated with poor clinical outcomes specifically in patients with estrogen receptor-negative tumors, including earlier recurrence and mortality.

In total, we have reported a transcriptional program present in invasive basal cells and have identified factors that regulate this program. Further analysis of our transcriptional data may reveal novel markers and vulnerabilities of basal breast cancer cells.

Additional work is warranted to clarify patient populations in which vitamin A intake may be hazardous rather than beneficial or harmless.

Table of Contents

ACKNOWLEDGEMENTS	8
INTRODUCTION	12
BACKGROUND	13
BREAST CANCER STAGING	13
BREAST CANCER SUBTYPES	14
STUDYING METASTASIS	15
FIGURE 0.1 THE METASTATIC CASCADE	16
THE ROLE OF TUMOR HETEROGENEITY IN METASTASIS	17
PLASTICITY PROMOTES HETEROGENEITY AND METASTASIS	19
BASAL EPITHELIAL CELLS IN SOLID CANCERS	22
FIGURE 0.2 BASAL AND LUMINAL CELLS IN MAMMARY GLANDS AND BREAST CANCER	23
TABLE 1	25
BASAL EPITHELIAL CELLS IN COLLECTIVE METASTASIS	26
IDENTIFYING UNIQUE VULNERABILITIES OF EPITHELIAL BASAL CANCER CELLS	27
STUDY RATIONALE	29
CHAPTER 1: BASAL BREAST TUMORS HARBOR EPIDERMAL-LIKE PLASTICITY THAT IS VITAMIN A DEPENDENT AND THERAPEUTICALLY TARGETABLE	30
SUMMARY PARAGRAPH	31
MAIN	32
RESULTS	33
BASAL BREAST TUMOR CELLS CO-EXPRESS EPIDERMAL GENES IN MOUSE AND HUMAN BREAST CANCERS	33
CLASSICAL INDUCERS AND SUPPRESSORS OF EPIDERMAL DIFFERENTIATION REGULATE INVASION DYNAMICS AND CELL VIABILITY.	37
CORNIFICATION SIGNATURE IS PREDICTIVE OF ATRA INDUCED INVASION AND PROLIFERATION, AND RAR α INDUCED CELL DEATH IN MOUSE AND HUMAN BREAST TUMOR ORGANOID MODELS	41
RAR ACTIVATION IN BREAST TUMOR ORGANOID MODELS LIMITS EPIDERMAL DIFFERENTIATION AND KERATINIZATION	43
GENETIC EFFECTS OF PERTURBING EPIDERMAL DIFFERENTIATION FACTOR KLF4 IN BREAST CANCER ORGANOID MODELS	44
RAR SIGNALING ANTAGONISM MARKEDLY INHIBITS TUMOR GROWTH AND METASTATIC PROGRESSION IN VIVO	47
VITAMIN A SUPPLEMENT INTAKE IS ASSOCIATED WITH RECURRENCE AND MORTALITY IN ER-NEGATIVE BREAST TUMORS	51
DISCUSSION	55
METHODS	58
EXTENDED DATA	78
SUPPLEMENTARY TABLES	90
DISCUSSION AND FUTURE DIRECTIONS	95

THE TRANSCRIPTIONAL PROGRAM OF LEADER CELLS	95
EPIDERMAL-LIKE TRANSCRIPTION PROGRAM IN BREAST CANCER	96
RETINOIC ACID IN BREAST CANCER	98
VITAMIN A AND CANCER OUTCOMES	102
REFERENCES	104

Acknowledgements

This work, along with my completion of my degree, would not have been possible without the support and contributions of the people in my life. First, I would like to thank Kevin Cheung, my graduate mentor. Kevin, I am so grateful that you took me on as a graduate student and agreed to mentor me. During my time in the lab, I have grown so much a scientist and as a person. Thank you for walking me through planning a project, applying for funding, and presenting results. Thank you for sharing my joy in crafting hypotheses and theorizing about cellular biology, it has been a true pleasure to brainstorm with you and learn how to approach hard scientific problems. Thank you for setting up such a positive lab environment and for finding such fantastic people to fill it.

I want to thank my scientific mentors who helped me along my educational journey to eventually pursue a PhD. To Christy Fillman and Nancy Guild, thank you for showing me that science is more than just a laboratory assignment, and for giving me a place to discover something new. Shout-out to the entire SEAPHAGES project, which is still giving undergraduates an opportunity to participate in novel scientific research. To James Orth, thank you for taking me on as a graduate student and teaching me to be a resilient scientist. Thank you for showing me how important it is to laugh at the ridiculous setbacks in science and to celebrate every victory, no matter how small. Thank you for instilling in me what I expect will be a life-long love of cancer biology and microscopy. To Bob Doebele, I am so grateful for my time as a research assistant in your lab. My time in the Doebele lab opened my eyes to the broader scientific problems

being tackled in cancer biology, and taught me to think first and foremost about the patients during my work. I could not have asked for a better training environment, and I am forever thankful for the mentorship and friendship of Anh and Adriana. I would not have made it to graduate school, nor made it through graduate school without your guidance and support.

To the Cheung lab members, past and present, thank you for being such all-around wonderful humans. There are now too many of you to thank each individually, but know that each of you have taught me valuable skills and have made the process of getting a PhD a much better experience. I want to give a special thanks to Ami, for being my Cheung Lab graduate student buddy, for being there for me in tough times, for helping me work through hard scientific problems, for hunting down delicious food with me, and for introducing me to the Hutch United community.

Graduate school can be a daunting and uncertain endeavor, and I want to thank the MCB chairs, staff, and community for making it such a welcoming place. Thank you for giving me support and a sense of structure while navigating my PhD. I also could not have asked for a better cohort to begin this journey with. Thank you all for making Seattle a not-so-frozen place to move, and for giving me an instant community. I want to also thank the members of my committee: Julie, Cyrus, Boba, Emily, Jon, and Manu, for giving me advice, encouraging me, and brainstorming my project. Your feedback has been paramount in completing and interpreting the results of this project.

To the zoo squad: Molly, Hannah, Laila, Lews, Elya, and Aidan, I'm not sure how to put into words what you guys mean to me. The best thing to have come out of my time in Seattle is your friendship, and I am so lucky to have it. Thank you for all of the laughter, wine adventures, camping mishaps, and formula 1 watch parties. You guys not only made the PhD process manageable, but you made these past 6 years some of the best I have had. I can't wait to have our reunions!

Thank you to all of my Colorado friends, who have supported my graduate school journey even from the application stage. To Ryan, Ryan, Sara, Dan, and Alo: thank you guys for coming out to see us so often, and for making time to see us when we were in Denver. You guys made sure that Colorado always felt like home. I'm excited for all of our friendsgivings, Christmas parties, and for the big life events coming our way! It's so special to have such a dependable and supportive group of friends. Nina, Alba, and Nica, thank you guys for being such a big support group and sounding board. Thank you for always being there in the good times and the bad. Special thanks to Nina. Thank you for coming to visit us so often, for being there to vent to, for bringing so much laughter and wit into my life, and for your unwavering friendship. Another special shoutout to Luke. Luke, thank you for being such a caring, easy going, and fiercely loyal friend to Logan since middle school and to me since high school. Thanks for always being down to do anything, for going on Taco Bell runs with us, and for your fantastic AP calc math skills. Also thank you to Alba and Luke for moving to Seattle and getting married, making it easier to spend time with some of our favorite people. Thank you guys for letting me have dog time in my last year of grad school!

I would not have gotten to graduate school at all without the support of my family. To my parents: thank you for always encouraging my interest in science and for giving me the confidence that I can do anything I set my mind to. Thank you for showing me how to be resilient and how to find joy in tough times. Thank you for making room for Logan and I every time we came home, and for helping us navigate adulthood. Thank you to my siblings, Whitney and Ari, for checking in on me and for being such wonderful people to spend time with. Thank you to the Kimbrels, who I am lucky to call my new family. You guys bring out the most positive part of any situation, and have been such great buddies for exploring the PNW! Without all of my family, I would not be who I am or where I am today, and I am so grateful for all of you.

Finally, a huge thank you to my husband, Logan. Logan, I know that asking you to move across the country with me for school was a big ask, and I am so grateful you agreed to come! Thank you for picking Seattle as a place to live, you nailed the decision. Thank you for giving such incredible and truly undeserved support. I'm genuinely not sure how anyone makes it through a hard and demanding process like a PhD without someone like you. Thank you for being my food and travel buddy. Thank you for doing dishes and laundry when I was overwhelmed. Thank you for asking me to marry you in the middle of graduate school, it was the best decision I've ever made. I love you.

INTRODUCTION

Approximately one in eight women will receive a breast cancer diagnosis at some point in their lifetime ¹. Advances in proactive screening and targeted therapies have significantly improved outcomes, resulting in a current 5-year survival rate of approximately 90% ¹. Despite these advancements, it is projected that 42,780 patients will die from breast cancer in the United States in 2024 ¹. The majority of these deaths can be attributed to metastatic disease.

Treatment options for patients with widespread disease are limited compared to those with localized disease. This is particularly true for patients with metastatic breast cancer, especially those with metastatic triple-negative breast cancer ². Many triple-negative breast cancer tumors contain a unique population of tumor cells characterized by high expression of basal cytokeratin such as keratin 14 (K14) and keratin 5 (K5), in the absence of the myoepithelial marker smooth muscle actin (SMA) ³. The presence of these cells is associated with increased rates of metastasis, collective cellular behavior, and poor outcomes ⁴. Targeting these basal epithelial cells through inhibition of markers such as K14 and integrin beta-1 (ITGB1) has been shown to prevent metastasis in pre-clinical cancer models. However, clinically feasible methods for eliminating basal epithelial cells remain unknown. Given this, we aimed to transcriptionally characterize basal epithelial breast cancer cells to identify unique vulnerabilities of this cell type.

BACKGROUND

Breast cancer staging

Breast cancers are categorized by stage and subtype based on a variety of characteristics, which clinicians use to predict tumor behavior and guide treatment plans. The stage of breast cancer is primarily determined by the size of the tumor and the location of tumor cells. A smaller, relatively contained tumor will likely be assigned a lower stage. Conversely, if cancer cells have spread beyond the breast to distant sites, the cancer is considered metastatic and classified as stage IV⁵. Each increase in stage correlates with a reduced probability of survival².

Under ideal conditions, a contained breast tumor can be surgically removed through targeted surgery or a mastectomy. Localized cancer can also be effectively treated with radiotherapy⁶. However, surgery and radiation become significantly less effective when cancer cells have disseminated throughout the body. Unlike some cancer sites, the breast is not a vital organ, so disease within the breast or the loss of the breasts is not typically life-threatening. However, cancer growth in distant sites such as the lungs, liver, and brain can ultimately lead to the patient's death. Consequently, metastasis is responsible for the majority of breast cancer-related deaths.

Unfortunately, early detection of breast cancer does not always prevent metastasis. Even after achieving remission, patients remain at risk of disease recurrence, which often presents at distant sites. Nearly 20% of patients with early-stage breast cancer will be diagnosed with metastatic disease within the first five years following surgical tumor

removal ⁷. Passing this five-year mark is a significant milestone for cancer survivors; however, even then, 14% of patients who were originally diagnosed with stage I breast cancer will experience distant recurrence within 5-20 years after remission⁸.

Breast cancer subtypes

In addition to staging, breast cancer is often classified into subtypes based on the presence of three proteins: estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2). Tumors that are ER-positive, PR-positive, or positive for both ER and PR are considered hormone-receptor positive. Up to 70% of breast cancers fall into this category and can be treated with therapies such as aromatase inhibitors and selective estrogen receptor modulators (SERMs). Tumors that are HER2-positive may respond to HER2 inhibitors such as trastuzumab and can also be targeted with HER2 antibody-drug conjugates (ADCs) ⁶.

If a tumor does not express ER, PR, or HER2, it is classified as triple-negative breast cancer (TNBC). Historically, TNBC has been treated with cytotoxic chemotherapies, but recent advancements have led to the development of more targeted therapies. TNBC is increasingly being divided into subclasses, several of which respond to specific inhibitors of pathways such as receptor tyrosine kinase signaling and DNA repair⁹. In some cases, the use of immunotherapies such as PD-1 and PD-L1 inhibitors has proven successful ⁶.

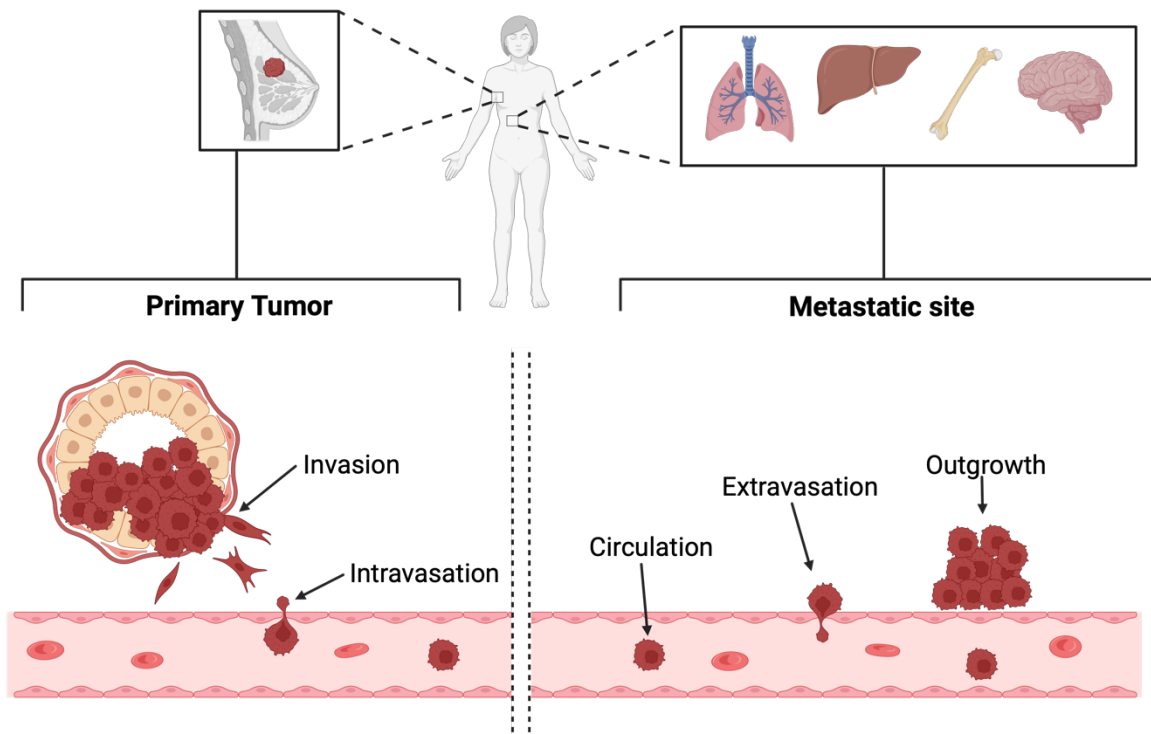
Despite these advances, the 5-year survival rate for stage IV TNBC remains at 12.8% ². This low survival rate reflects the unique challenge of treating TNBC in conjunction with

metastatic disease. There is a pressing need for safer and more effective therapies for these patients.

Studying metastasis

The process of cancer metastasis is traditionally described in a series of sequential steps termed “the metastatic cascade”. These steps include: 1) the invasion of tumor cells into adjacent tissue; 2) the intravasation of tumor cells into the bloodstream or lymphatic system; 3) the circulation of tumor cells within the vasculature; 4) the extravasation of tumor cells from a blood or lymph vessel into a new site; and finally, 5) cellular proliferation at the distant site^{10,11}. As with any cancer-related biological process, cancer cells can undertake a variety of versions of the metastatic cascade. However, to successfully colonize distant sites, they generally must complete the majority of these steps (Figure 0.1).

Figure 0.1 The metastatic cascade



Appropriately, an immense amount of effort has been spent to understand the biology of metastasis, with the hopes of finding methods for prevention and treatment of metastases. Various dogmas exist for investigating new therapeutic strategies aimed at targeting metastatic disease. One approach involves identifying treatments that inhibit or eliminate tumor cells universally, irrespective of their anatomical sites. Alternatively, a more targeted strategy focuses on selectively eradicating metastatic cells during and after completion of metastasis. Such a strategy is grounded in the observation that cancer cells either inherently possess or acquire specific traits that facilitate their dissemination and proliferation in remote anatomical locations¹². This hypothesis—that

metastatic cells exhibit distinct properties, thus presenting unique susceptibilities—serves as the guiding principle for the research delineated in this thesis.

The role of tumor heterogeneity in metastasis

Each cancer is unique. While cancers can be generally grouped by various criteria, as discussed in the previous section, the field is moving towards a personalized medicine approach. This involves individually assessing each cancer for sensitivity to different therapeutics. Beyond the heterogeneity between patients, there is also significant heterogeneity within individual cancers.

Tumors are composed of vast numbers of cells, ranging from millions to trillions. This large population of rapidly growing and evolving cells is highly heterogeneous. Within the tumor microenvironment, various cell types are present, including immune cells and stromal cells ¹³. Additionally, the transformed cells themselves exhibit extensive heterogeneity at multiple levels, including genetic, epigenetic, and phenotypic variations ¹⁴. In any diverse population, certain cells acquire traits that provide them with an evolutionary advantage, while others are at a disadvantage. Other cells may acquire specialized traits that enhance their survival under specific conditions.

One specific condition in which certain cells may become more adept at survival is the metastatic cascade. These cells are often referred to as metastasis-initiating cells (MICs) ¹². One of the earliest reports of this phenomenon comes from Fidler and Kripke in 1977. In their study, individual clones derived from the B16 melanoma line were injected into the circulatory system of mice, allowing for the seeding of pulmonary

metastases. Supporting their hypothesis that some cells are more likely to produce metastases, certain clones generated large numbers of metastases, while others produced few. When daughter cells from these clones were separated into new lines, the metastatic potential of the lines was maintained ¹⁵.

This foundational work has been followed by a series of publications reporting that certain cellular clones or cell lines are highly metastatic compared to others ¹⁶⁻¹⁹. These findings have prompted numerous studies aimed at identifying differences between highly and poorly metastatic cells, with the goal of discovering a core set of metastasis-driving characteristics ²⁰.

Instead of a single central mechanism for metastasis initiation, several distinct mechanisms have been identified. For instance, in a study of breast cancer patients, metastases were found to exhibit a higher overall genetic mutational burden, often including mutations in six cancer driver genes ²¹. This concept has been further developed using larger datasets of metastatic cancers across various types. A 2020 study confirmed that metastases possess widespread genetic mutations without a specific pattern, leading the authors to conclude that metastasis may coincide with genetic instability ²².

Tumor cells can also undergo epigenetic alterations that facilitate metastasis. For example, increased activation of the epigenetic modification protein EZH2 can downregulate tumor suppressor expression and promote the invasiveness and metastasis of melanoma cells ²³. Methylation signatures of cancer cells have been found to be predictive of metastasis in breast cancer cohorts ²⁴. This concept is further

supported by studies of transcriptional signatures in cancer metastasis, which can even predict the location of metastasis ^{25,26}. For instance, melanoma cells that express high levels of the lactate transporter MCT1 can uptake more lactate from their environment during circulation, which supports their metabolic activity and increases the likelihood of metastasis ²⁷.

Beyond changes in genetic landscapes and transcription, cells can exhibit signaling differences that enhance their resilience during metastasis. For example, cells with increased activity of HIF proteins, which protect the cell from death, are more likely to survive the hypoxic conditions encountered during circulation ²⁸. Additionally, breast cancer cell lines with heightened ERK and PI3K signaling are more invasive and metastatic ²⁹.

While no specific cell state has been ubiquitous or specific enough to justify a pan-cancer anti-metastasis therapy, it may be possible to identify the most likely metastatic cell populations on a per-patient basis and inhibit them.

Plasticity promotes heterogeneity and metastasis

Around the same time as studies highlighting metastasis-adept populations, Greenburg and Hay reported that under certain conditions, differentiated adult epithelial cells could convert into a mesenchymal cell state ³⁰. This process, known as Epithelial to Mesenchymal Transition (EMT), is now understood to occur without genetic mutation, but rather through regulation of signaling cascades and epigenetics ³¹. During EMT,

epithelial cells lose their cell-cell and cell-matrix attachments and often become migratory. It did not take long for this concept to be linked to tumor invasion and metastasis ³². Follow-up work confirmed that EMT is indeed common in tumors of epithelial origin and plays a significant role in metastasis ^{33,34}.

Thus, a new paradigm for metastasis was established: one in which epithelial cells undergo EMT, detach from the tumor, and migrate to blood vessels through which they travel to distant sites. At these distant sites, the cells can revert to a proliferative epithelial state in a process termed Mesenchymal to Epithelial Transition (MET). This model has largely stood the test of time, albeit with some refinements and caveats ³⁵.

A particularly important implication of the EMT model is that every tumor cell, regardless of genetics, may have the potential to become highly metastatic. To effectively target this form of metastasis, it is not sufficient to find cells that have already undergone EMT; preventing the conversion of epithelial cells to mesenchymal cells is crucial.

In a classic case of EMT, tumor cells completely detach from the tumor, move into the surrounding tissue, and transform into cells with a stretched out morphology and abundant membrane protrusions. These cells begin expressing mesenchymal markers such as Vimentin and N-cadherin while downregulating epithelial markers such as E-cadherin and epithelial cytokeratins ³⁵. As is the case with nearly any cancer-related biological phenomenon, there are exceptions to the theory of EMT-driven metastasis. In some cases, these cells are categorized as hybrid-EMT cells, which maintain both epithelial and mesenchymal properties, benefiting from the pro-metastatic and pro-

survival advantages of each ^{33,36}. Additionally, there are many tumors with little-to-no evidence of EMT that are still highly metastatic.

Cases of metastasis occurring in the absence of EMT have prompted efforts to identify alternative forms of cellular plasticity that may drive metastasis ^{37,38}. Such plasticity can take many forms. For example, colorectal cancer cells can become more or less stem-like by altering the expression of Lgr5, enabling them to shift cell states and proliferate ³⁹. Melanoma cells can transition towards a fatty acid oxidation state through changes in Yap signaling, which facilitates colonization of lymph nodes ⁴⁰. In some breast and melanoma cancers, tumor cells can adapt to become less dependent on anchorage by expressing genes typically found in hematopoietic cells ⁴¹. In pancreatic cancer, basal cells can upregulate podoplanin expression to support invasion and metastasis ⁴².

Cellular plasticity is often the result of a cell responding to changes in its environment, and can be induced through a variety of stimuli. The tumor microenvironment, composed of a multitude of cell types and extracellular proteins, can be at times a master regulator of tumorigenesis and cancer behavior ⁴³. This is particularly true in the case of cancer invasion, wherein the initiation, path, speed, and mode of migration are all altered in response to microenvironmental cues. Such cues include the matrix fibers and proteins, the stiffness of the matrix, signaling molecules secreted by nearby cells, systemic signaling molecules, and physical contact with neighboring cells ^{44,45}. The ways in which the environment alters tumor cell state are being uncovered at an astounding rate, and will no doubt our understanding of these interactions will continue to evolve. For instance, recent work has discovered that neurons within the tumor

microenvironment can signal to epithelial tumor cells, opening up a new field of cancer biology ⁴⁶.

The pro-metastatic plastic cell state examined in this work is characterized by the maintenance of an epithelial phenotype and high expression of the basal epithelial keratin 14 (K14). These cells are referred to as basal epithelial tumor cells or K14+ cells.

Basal epithelial cells in solid cancers

Before describing our cells of interest, it is essential to define the identity of "basal cells." Normal breast ducts are composed of a bi-layered epithelium. The inner, or "luminal" layer of cells lines the inside of the ducts and contains cells responsible for the secretion of milk. The outer, or "basal" layer consists of myoepithelial cells, which form a barrier around the luminal cells and can contract to propel milk through the ducts.

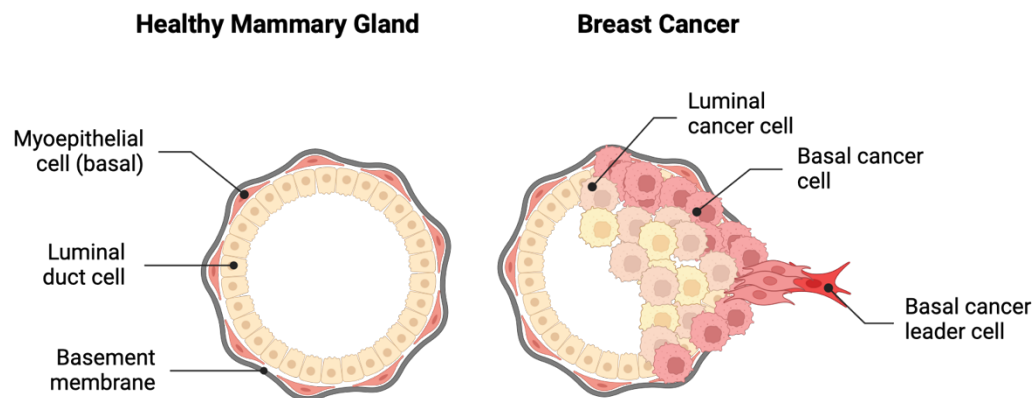
Myoepithelial cells are considered a form of basal epithelial cells since, similar to other epithelial tissues, these cells are in contact with the basement membrane and express canonical basal epithelial markers such as K14 and K5. Myoepithelial cells uniquely express smooth muscle actin (α SMA), a protein often associated with contractile smooth muscle cells, such as those in blood vessels ⁴⁷.

In the healthy mammary gland, myoepithelial cells are formed as cap cells - the leading cells during mammary gland branching - differentiate and lose motility. Upon injury, or conditions requiring mammary gland remodeling such as pregnancy, luminal cells are believed to be the primary source of new myoepithelial cells. However, some recent

work has shown that in some contexts, myoepithelial cells may de-differentiate and act as a source for luminal mammary gland cells ^{48,49}.

While investigating the identity of collectively invading cells, Cheung et al. discovered that invasive cells, particularly those at the leading edge of invasion, known as "leader cells," exhibit high production of K14. Unlike normal basal cells of the breast, these collectively invading cells do not express α SMA and occasionally co-express luminal cell markers ³ (Figure 0.2).

Figure 0.2 Basal and Luminal cells in mammary glands and breast cancer



Subsequent investigations revealed that K14-positive cells are not only present in invasive strands but also in circulating tumor cells and metastases, suggesting a potential role for K14 cells throughout the metastatic process. Indeed, when K14 levels were reduced in tumors using short hairpin RNA (shRNA)-mediated degradation, metastasis was significantly reduced in transplant models of breast cancer ⁴. This

finding indicates that K14 is not simply a marker of invasive basal cells, but that the intermediate filament plays a functional role in metastasis.

In addition to breast cancer, K14-positive (K14+) cells have been identified as a particularly aggressive cell population in various solid tumor types, including lung, bladder, and ovarian cancers. Moreover, K14+ cells have been associated with additional aggressive traits such as chemoresistance and recurrence (see Table 1). This is not a universal trait of solid tumors however, and there are cases in which luminal cells are the more aggressive cell type, particularly in prostate cancers⁵⁰⁻⁵².

Table 1

Cancer type	Experimental model	Basal cells associated with:	Basal cell markers	Ref.
Prostate	benign prostate cells and expression data	Proneural differentiation, therapy resistance, poor survival	Trop2+ CD49f(Itga6)hi	53
Non-small cell lung	tissue microarray	Poor survival, metastasis	K14	54
Lung adenocarcinoma	tissue microarray, in-vivo transplants	Poor survival, metastasis	K14	55
Breast	3D culture and in-vivo transplant	Invasion, metastasis	K14	56
Breast	3d culture	invasion	K14	57
Breast	tissue IHC	metastasis	K14	58
Ovarian	3D culture and tissue microarray	Invasion, poor survival	K14	59
Ovarian	tissue IHC	poor survival, chemoresistance	K5	60
Ovarian	3D culture	chemoresistance	K14	61
Bladder	expression data, in-vivo transplants,	Poor survival	K14, CD90, CD44, CD49f	62
Bladder	3D culture, tissue IHC	chemoresistance	EPCAM, K14, K5, p40	63
Oral	tissue IHC	poor survival	K14	64
Renal cell	tissue IHC	poor survival, high grade, disease stage	K14	65
Salivary gland	3D culture, tissue IHC	invasion, metastasis, recurrence	K14	66

While basal cells are observed as a stable population in certain instances, in many models the basal cell state is plastic and reversible. For example, in the PyMT model of breast cancer, tumor cells adopt a basal state specifically when in contact with a collagen I matrix. This basal state is subsequently lost when cells are suspended or transferred to a basement membrane extract such as Matrigel. In this case, cell state is controlled, at least in part, by transcriptional mechanisms. This is made obvious through

reporter systems using the Krt14 promoter to drive expression of GFP-bound actin, which is preferentially activated in basal cells in contact with Collagen I³.

Basal epithelial cells in collective metastasis

Unlike cells that have undergone full EMT, which tend to metastasize individually, basal epithelial cells typically maintain adhesion to groups of cells^{3,4}. This observation suggests that basal epithelial tumor cells play a role in collective metastasis, where the metastatic process is facilitated by a cohesive group of cells rather than individual cells.

Collective metastasis was recognized as early as the 1950s but has gained significant attention in recent years as a crucial cancer process. It is now understood that cells traveling in groups are more likely to survive circulation and rapidly develop into aggressive metastatic lesions compared to solitary cells. The efficiency of metastasis by tumor cell clusters can be attributed to several factors, elaborated upon in comprehensive reviews^{67,68}. These factors include the ability of clustered cells to provide mutual support through signaling, evade immune detection more effectively, and withstand mechanical stresses encountered during circulation.

The clinical significance of collective tumor cell behavior is underscored by its impact on patient outcomes. Patients harboring circulating tumor cell clusters face a significantly higher risk of cancer-related mortality, reported to be 2 to 15 times greater than that of patients without such clusters^{67,68}.

Several insightful studies have demonstrated that circulating tumor cell clusters often originate from groups of cells simultaneously exiting the primary tumor^{4,69}. These investigations emphasize the critical importance of comprehending collective behavior within the primary tumor to disrupt the dissemination of tumor cell clusters.

As previously mentioned, the potential involvement of K14+ cells in invasion and metastasis was elucidated through studies conducted by Cheung et al^{3,4}. These studies systematically explored the role of K14+ cells in various stages of the collective metastatic process, including invasion, circulation, and outgrowth. Basal epithelial cells were notably identified in collectively invading strands, particularly at the forefront of these strands. Moreover, K14+ cells were enriched within circulating tumor cell clusters and multiclonal metastases, subsequently diminishing in prevalence as more luminal-like cells proliferated at the distant site. Knockdown experiments targeting K14 alone effectively reduced collective invasion in 3D culture models and prevented metastasis in preclinical mouse models, invoking a supportive role of K14+ cells in metastasis.

Given the pivotal role of basal epithelial cells in collective metastasis, targeted inhibition of these cells holds promise for mitigating cancer spread through both cell-autonomous and non-cell-autonomous mechanisms.

Identifying unique vulnerabilities of epithelial basal cancer cells

To effectively inhibit K14+ cells, it is crucial to elucidate the cellular mechanisms underlying basal cell invasion, survival, and metastatic seeding. Keratin 14 expression

itself may play a pivotal role in conferring metastatic potential to basal epithelial cells. Keratins are integral to both cell-cell and cell-matrix adhesion processes ⁷⁰. Notably, studies have demonstrated that expression of keratin 14 and keratin 5 alone can restore cell-cell adhesion in cultured keratinocytes lacking all type-I keratins, highlighting keratin 14's potential as a central mediator of the collective behaviors observed in basal epithelial cells ⁷¹. However, the function of K14 likely involves complex interactions with various junctional and structural proteins, adapting dynamically to specific cellular contexts.

Of significant consideration in therapeutic development, basal keratins are indispensable components of epithelial cells and systemic inhibition could lead to severe side effects ⁷⁰. Addressing this challenge requires uncovering unique interactions involving keratin 14 in basal tumor cells or other proteins upregulated in unison with K14. Exploring the cellular mechanisms that promote the K14+ cell state may reveal novel traits exploitable in therapeutic strategies. If for instance, K14 is upregulated through pathways unique to cancers, then such pathways could be inhibited while sparing normal tissue.

Beyond K14, basal invasive cells have other reported molecular dependencies such as the transcription factor p63 ³. p63 plays a pivotal role in breast cancer, but similarly to K14 has proven to be a poor drug-target candidate. It is likely that there are additional proteins or protein functions specifically required by invasive basal cells that may be more specific to the context of cancer than known markers.

STUDY RATIONALE

In this study, we aimed to understand the transcriptional program of basal leader cells and the signaling pathways that drive it. We reasoned that this knowledge would enable us to identify new therapeutic targets to either inhibit the function of invasive basal cells or to prevent them from forming all together. Since leader cells are a rare and plastic population, studying them can be difficult. Organoid culture of MMTV-PyMT tumor cells has proven to be a method for enhancing the ratio of basal leader cells to luminal cells while maintaining similar invasive behavior to a tumor in-vivo³. Several approaches have been developed to isolate rare cells from organoid culture, but we felt each required extraneous biological perturbations to an already partial model of a tumor. These include use of tools such as photoconvertible tags that can be used to mark and isolate specific cell populations chosen by the investigator. Previously, K14+ cells were isolated for RNA sequencing using the Krt14promoter-GFP-actin reporter, which provided an incomplete view of basal markers without capturing heterogeneous populations^{3,4}. For these reasons, we chose to investigate the transcriptional program of basal leader cells using single cell RNA sequencing (scRNA-seq) across a time course of invasion. This approach allowed us to be unbiased in cell selection since all organoid cells were sequenced. Further, the addition of time points allowed us to investigate the continuum of cell states that exist as invasion progresses. The results of this sequencing data formed the basis for the studies described in this thesis.

Chapter 1: Basal breast tumors harbor epidermal-like plasticity that is vitamin A dependent and therapeutically targetable

Andrea E. Doak^{1,2}, Saki Suyama-Banjo¹, David R. Doody³, Alexandra R. Reers^{1,4}, Justin Hui^{1,2}, Brad A. Krajina¹, Yuji Kumagai¹, Nicole Rhoads¹, Jimin Park¹, Rebeca Alvarez⁵, Manu Setty^{6,7}, Eileen Rillamas-Sun⁸, Lia D'Addario⁹, Isaac J. Ergas⁹, Janise Roh⁹, Lawrence H. Kushi⁹, Heather Greenlee^{8,10,11}, Kathleen E. Malone³, Kevin J. Cheung^{1,11*}

¹ Translational Research Program, Public Health Sciences and Human Biology Divisions, Fred Hutchinson Cancer Center, Seattle, Washington 98109, USA.

² Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, Washington 98195, USA.

³ Epidemiology Program, Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington 98109, USA.

⁴ Molecular Medicine and Mechanisms of Disease Graduate Program, University of Washington, Seattle, Washington 98195, USA.

⁵ Department of Pathology, University of Washington, Seattle, Washington 98195, USA.

⁶ Computational Biology Program, Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington 98109, USA.

⁷ Basic Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington 98109 USA.

⁸ Cancer Prevention Program, Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, Washington 98109, USA.

⁹ Division of Research, Kaiser Permanente Northern California, Pleasanton, California 94588, USA.

¹⁰ Clinical Research Division, Fred Hutchinson Cancer Center, Seattle, Washington 98109, USA.

¹¹ Division of Hematology/Oncology, Department of Medicine, University of Washington School of Medicine, Seattle, Washington

SUMMARY PARAGRAPH

Tumor cells are highly plastic and transition into cell states that endow increased metastatic potential and drug resistance. This phenotypic plasticity is often considered a therapeutic challenge but could in principle introduce new vulnerabilities. In breast cancer, tumor cells that adopt basal epithelial phenotype are associated with collective migration and aggressive metastatic behavior. Here we show that basal breast tumor cells acquire a skin epidermis-like gene program rendering them vulnerable to regulators of epidermal differentiation, the natural process that forms the layer of dead skin cells on the body surface. Suppression of epidermal terminal differentiation pharmacologically via the vitamin A metabolite, all-trans retinoic acid, or genetically by KLF4 loss-of-function, markedly enhanced directed cell migration and proliferative potential in breast tumor organoid models. Conversely, retinoic acid receptor antagonists or KLF4 overexpression, which classically induce terminal cornification in skin, robustly blunted invasion, caused apoptosis, and increased keratin production. A transcriptional epidermal cornification signature was highly predictive of retinoic acid pathway responsiveness in mouse and human breast tumor models. In-vivo, treatment with retinoic acid receptor antagonists potently suppressed primary tumor growth, lung metastatic seeding, and outgrowth of established metastases. In independent epidemiologic studies, intake of vitamin A supplements was associated with increased risk of breast cancer recurrence. Our results reveal that basal phenotype breast tumor cells harbor a latent skin-like epidermal program. These highly metastatic cells thrive on vitamin A retinoids, and forced epidermal differentiation is a therapeutic strategy for their eradication.

MAIN

Highly plastic tumors harbor a continuum of cellular states, some of which confer selective advantages during critical phases of local invasion and metastatic spread⁷²⁻⁷⁵. A key goal of cancer biology research is to uncover the molecular architecture of cell plasticity and in turn, devise ways to control or eradicate these cells, thereby limiting metastasis and mortality^{38,76-79}.

Collective cell migration has emerged as a major mechanism for the local spread of cancer cells across solid tumor types and is associated with distant metastasis^{3,80-84}. During collective cell invasion, leading cancer cells display distinct molecular properties, such as states along epithelial-mesenchymal transition (EMT) axis, or expression of a basal epithelial lineage program^{74,85,86}. In addition to guiding collective cell migration, breast tumor cells with basal phenotype also support early metastatic seeding and colonization^{3,74,87-93}. In treatment contexts, basal-like tumor cells become enriched upon chemotherapy and endocrine therapy⁹⁴⁻⁹⁶ and act as a reservoir for drug resistance and tumor regrowth^{97,98}. And among breast cancer subtypes, basal-like tumor cells are particularly enriched in triple-negative breast cancers (TNBC)⁹⁹⁻¹⁰¹, which have poor clinical prognosis and propensity to metastasize. Together, these findings highlight the potential benefit of therapeutically targeting basal tumor cells to better eradicate breast tumors and prevent their successful metastasis. Yet compared with EMT states¹⁰²⁻¹⁰⁹, the key molecular vulnerabilities of basal tumor cells are not understood.

In the present study, we set out to investigate the dynamics of cell plasticity giving rise to basal leader cells in breast cancer. Through integrating mouse and human breast tumor organoid models, mouse models of metastasis, and clinical epidemiologic

evidence, we uncover a skin epidermal-like cell state, particularly enriched in triple-negative breast tumors, rendering these tumor cells exquisitely sensitive to retinoic acid receptor signaling mediated differentiation and keratinization-associated cell death.

RESULTS

Basal breast tumor cells co-express epidermal genes in mouse and human breast cancers

When breast tumor organoids are embedded in a 3D collagen I matrix, cells at the tumor matrix interface gain migratory features and invade collectively into the surrounding microenvironment^{3,82,83,110,111}. It has been shown previously that leader cells in both human and mouse breast cancers can arise from phenotypic plasticity in which cells acquire a basal epithelial molecular phenotype³.

Here we used single cell RNA sequencing (scRNA-seq) to interrogate the transcriptional dynamics of cell plasticity giving rise to basal leader cells in breast cancer. Mammary tumor organoids were isolated from the Mouse Mammary Tumor Virus Polyomavirus Middle-T (MMTV-PyMT) mouse model, which forms spontaneous aggressive mammary tumors and frequently metastasizes. Transcriptionally, MMTV-PyMT tumors show similarity with human luminal B subtype breast cancers and breast tumors with low estrogen receptor expression¹¹²⁻¹¹⁴. MMTV-PyMT tumor organoids embedded in 3D collagen rapidly formed cell protrusions that matured into multifocal invasive strands that were quantified by invasion index (**Fig. 1a-b; Ext data 1a**)¹¹⁰.

These events coincided with leading cells acquiring progressive expression of basal epithelial markers including cytokeratin-K14 (K14), indicating tumor cell plasticity

(**Fig. 1c**). Three time points were selected for single cell RNA sequencing (0, 16 and 72 hours). Tumor organoids were extracted from the matrix, dissociated into single cells, encapsulated and barcoded for downstream library preparation and RNA sequencing (10X Genomics), altogether yielding RNA profiles for 9,661 cells. Each biological replicate was analyzed separately and merged together, with similar results (Methods). Uniform manifold approximation and projection (UMAP) clustered tumor cells together by time point, while bifurcating into luminal and basal-like lineages. Further, *Krt14*⁺ cells became more abundant over time, indicating that the basal leader cell program is transcriptionally regulated (**Fig. 1d-e; Ext data 1b**).

To better define the basal leader cell state, we next identified genes whose expression was most correlated with *Krt14* transcript levels across all cells (**Fig. 1f, Supplementary Table 1**). To our surprise, gene set enrichment analysis of transcripts positively correlated with *Krt14* revealed enrichment of genes related to cornified envelope formation, epidermal differentiation, and epidermal morphogenesis, while transcripts inversely correlated with *Krt14* related to brown fat differentiation and cell motility (**Fig. 1g-i; Ext data 1b-c**). Likewise, unsupervised annotation using the Mouse Cell Atlas annotated *Krt14*-high cells as neonatal skin keratinocytes¹¹⁵ (**Ext data 1e**), indicating that basal tumor cells transcriptionally resemble cells of the epidermis. In contrast, *Krt14* transcript level were not correlated with either smooth muscle gene expression, a hallmark of mature myoepithelial cells, or mesenchymal gene expression, associated with EMT^{31,116,117} (**Fig 1j-k; Ext data 1d**).

Epidermal gene programs were top enriched gene sets in basal tumor cells in scRNA-seq datasets of invading organoids tumors from the highly invasive C3(1)Tag

mouse model of basal-like breast cancer ¹¹⁰ (**Fig1I, Ext data 1f**). Extending these results further, we found that *Krt14* was correlated with epidermal differentiation programs in scRNA-seq of in-vivo tumors from three mouse models of breast cancer, MMTV-PyMT, C3(1)Tag, and Pten-Myc ^{110,118-120} (**Fig. 1I; Ext data 2g, Supplementary Table 2**). Finally, we determined epidermal gene set enrichment in 47 human breast patient tumors from two human breast tumor scRNA-seq atlases ^{121,122} (**Fig 1I**). In both human datasets, basal tumor cells frequently co-expressed epidermal genes, particularly in tumors with triple-negative receptor status and PAM50 basal molecular subtype. (**Ext data1h-i**).

Figure 1

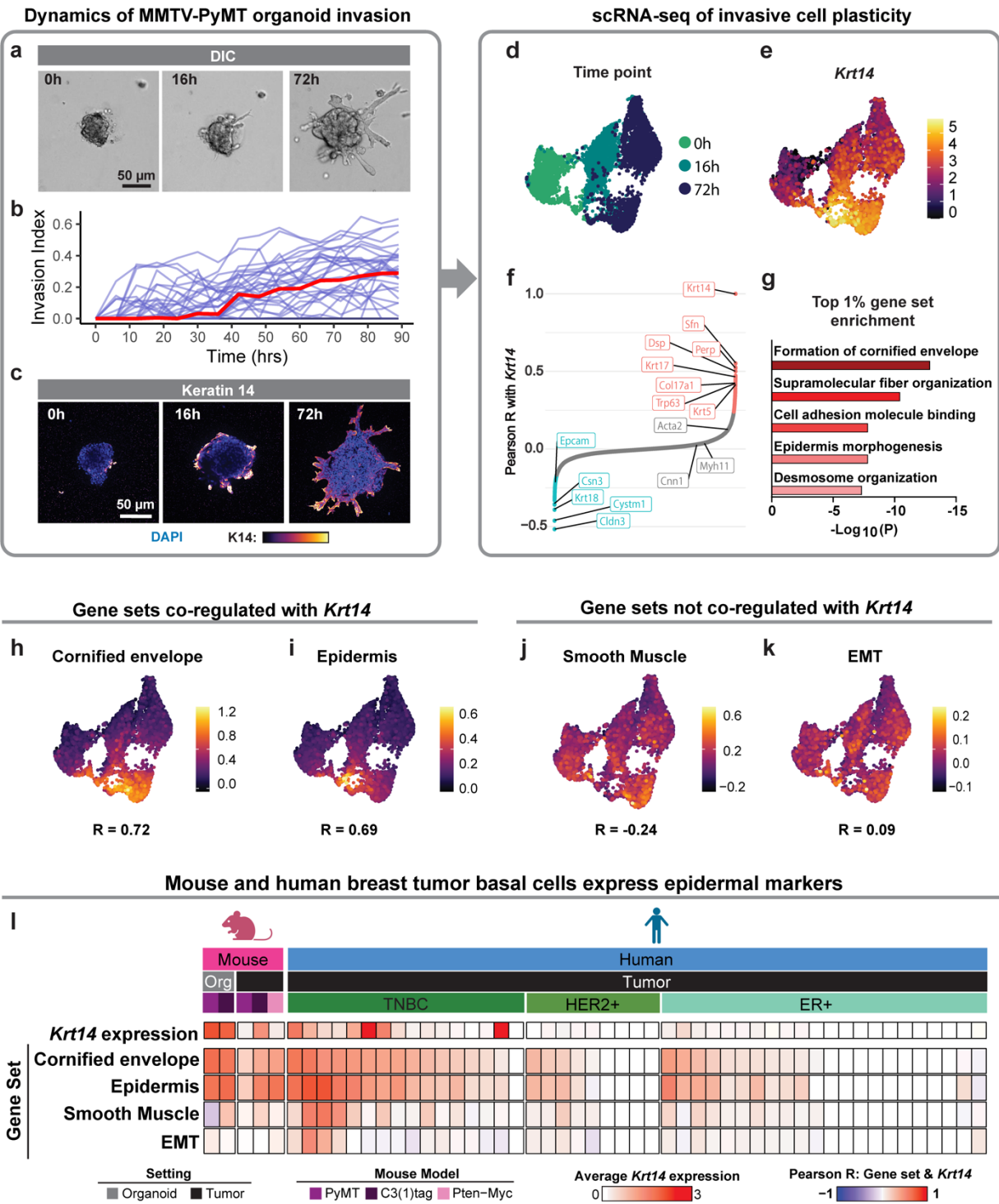


Figure legend on following page.

Figure 1: Basal breast tumor cells co-express epidermal genes in mouse and human breast cancers

a) Representative DIC microscopy time-lapse images of MMTV-PyMT mammary tumor organoids invading into 3D collagen I. **b)** Calculated invasion index of individually tracked organoids over time denoted in blue. N=37 organoids, 2 tumors. Median index denoted in red. **c)** Representative micrograph of Keratin 14 and DAPI immunofluorescence. **d-e)** scRNA-seq was performed on invading PyMT mammary tumor organoids at 0, 16 and 72 hours, n=2 biological replicates. UMAP for replicate 1 of cell-cycle regressed scRNA-seq data, labeled by time (d) and *Krt14* transcript expression (e). **f)** All detected genes sorted by correlation of expression with *Krt14* expression on per single cell basis. The 1% of genes with the highest correlation are colored in red, while the 1% with lowest correlation are colored in blue. Genes of interest are highlighted. **g)** Top 5 gene sets enriched for in the top 1% of genes correlating with *Krt14*, sorted by p-value. Calculated with Metascape. See **supplementary table 1** for gene set enrichments. **h-k)** UMAPs representing gene module scores calculated in Seurat using selected gene sets. Pearson R values denote correlation between each gene module score and *Krt14*. **l)** Heatmap showing *Krt14* and gene set co-expression in mouse and human tumors. scRNA-seq reported from two mouse organoid models, three mouse tumor models, and 47 human breast tumors. *Krt14* expression is colored by normalized *Krt14/KRT14* expression averaged across all tumor epithelial cells in the sample. Gene set values are colored by correlation of gene module score with *Krt14* across all epithelial tumor cells. See **supplementary table 2** for values.

Classical inducers and suppressors of epidermal differentiation regulate invasion dynamics and cell viability.

In skin, keratinocytes undergo a process of stratification and terminal differentiation, giving rise to a cornified rigid and crosslinked layer of dead cells at the outer surface of the epidermis^{123,124}. One major regulator of epidermal differentiation state is signaling by retinoic acid receptors (RARs), which are nuclear hormone receptors that bind vitamin A-derived metabolites to activate gene transcription¹²⁵⁻¹²⁹.

Natural and synthetic retinoic acid derivatives are used widely for a variety of dermatologic conditions to promote epidermal hyperproliferation and prevent terminal cornification^{130,131}. Conversely, vitamin A deficiency causes epidermal keratinocytes to undergo excessive cornification, which is associated with glandular-to-squamous metaplasia^{132,133}. Thus, vitamin A, through its metabolite, all-trans retinoic acid (AtRA),

activates RAR signaling and opposes epidermal differentiation, while deficiency of vitamin A reduces RAR signaling to promote differentiation (**Fig 2a**).

Given that basal breast tumor cells resemble an epidermal-like cell state, we next asked how basal breast tumor cells would respond to regulators of epidermal differentiation. We examined the effects of retinoic acid signaling on MMTV-PyMT tumor organoid dynamics using AtRA to induce retinoic acid receptor (RAR) activity, and the retinoic acid receptor inhibitor, pan-RAR reverse agonist BMS493 (RARI) to mimic retinoic acid deficiency (**Fig 2b**)¹³⁴, at typical dose ranges used to modulate keratin expression^{135,136}. Within 72 hours, both agents elicited striking responses. AtRA-treated organoids at doses ranging from 0.1 to 5 μ M produced greater numbers of collective invasion strands, that were also longer and narrower, when compared to control (**Fig 2c-d; Video 1**). In contrast, organoids treated with RARI had little to no collective strands at end point, and organoids dissociated into rounded-up single cells that were stationary (**Fig 2c,e; Video 1**).

To quantify individual cell dynamics, we next transduced tumor organoids with a nuclear H2B-GFP reporter, and tracked individual cells by confocal microscopy (**Ext data 2a, Video 2**). Surprisingly, single cell migration tracking revealed that AtRA did not increase net individual cell speed. Rather, AtRA increased cell persistence indicating improved net efficiency of directional collective migration (**Ext data 2b-k**). RARI treatment, however, led to decreased distance traveled and speed of individual cells, with evidence of nuclear blebbing and fragmentation (**Ext data 2b-k, Video 2**).

Since our observations in RARI treated organoids were consistent with cell death, we assessed viability and apoptosis of treated organoids using CellTiter-Glo and

Caspase3/7-Glo, respectively. Our results indicate that RARi treatment reduced viability and increased caspase activity in a dose dependent manner, with an IC₅₀ of 6.0 μ M (**Fig. 2g,i**). Interestingly, response curves for AtRA were biphasic, with increased viability and reduced caspase activity at lower doses, followed by reduced viability and increased caspase activity at high doses with IC₅₀ of 9.0 μ M (**Fig 2f,h**). These results indicate that AtRA treated organoids, specifically at low doses, are less prone to apoptosis and more highly proliferative or metabolically active-

To confirm that retinoic acid receptor activity regulates invasion, we expanded our drug panel to include two additional retinoic acid sources: 9-cis-retinoic acid (9-cis-RA) and retinyl palmitate. We observed increased invasion upon treatment with 9-cis-RA, which can bind and activate RARs directly, but as expected, no significant difference in invasion when organoids were exposed to retinyl palmitate, which is the storage form, normally enzymatic processed into retinoic acid by hepatic stellate cells (**Ext data 3a-b**). In our invasion scRNA-seq dataset, RAR transcript expression was broadly uniform throughout the uMAP (**Ext data 3c-e**). When treated with specific inhibitors for RAR α , RAR β and RAR γ isoforms using BMS614, LE135, and LY2955303, respectively ^{134,137-139}, tumor organoids had reduced invasion, indicating that RAR nuclear receptor inactivation, via one or more isoforms, is necessary for repression of invasion and tumor cell survival (**Ext data 3f-h**).

Figure 2

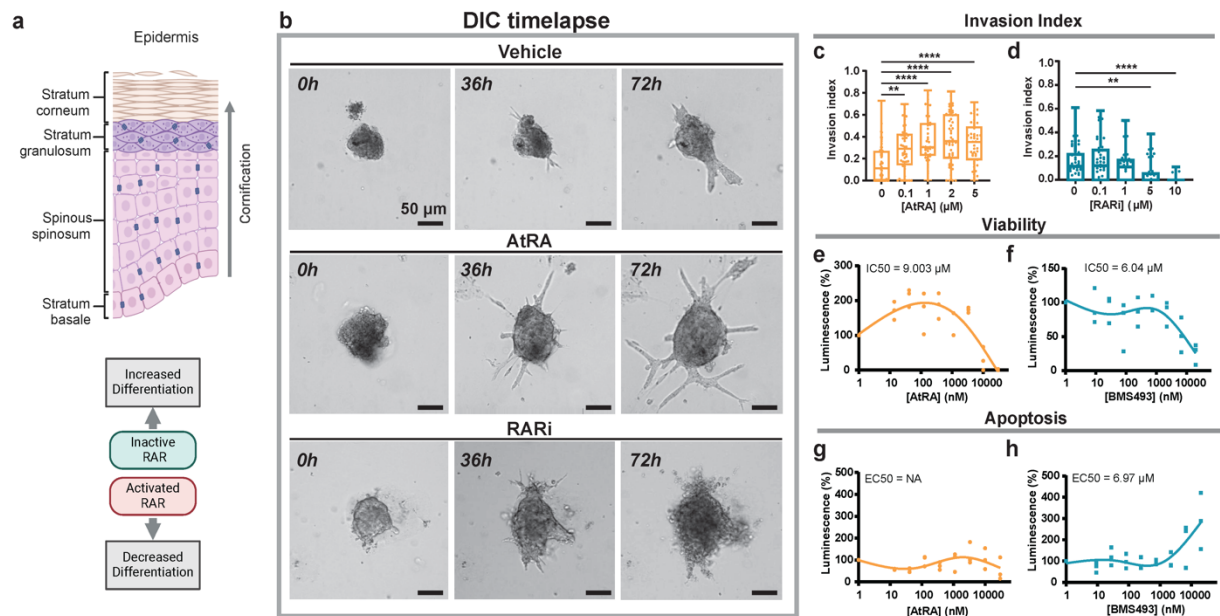


Figure 2. Retinoic acid signaling supports invasion and viability in MMTV-PyMT organoids

a) Schematic depicting the role of retinoic acid receptors in epidermal differentiation and cornification. Figure made with Biorender. **b)** Representative DIC microscopy time-lapse images of MMTV-PyMT mammary tumor organoids invading into 3D collagen I treated with either 1 μ M AtRA or 5 μ M RARi (BMS493). Due to the short half-life of retinoids, drugs were refreshed every 48 hours. **c)** Invasion index of MMTV-PyMT organoids exposed to increasing concentrations of AtRA. $n \geq 45$ organoids per dose, 3 biological replicates. P-values calculated with Dunnett's multiple comparison test after one way ANOVA. **d)** Invasion index of MMTV-PyMT organoids exposed to increasing concentrations of RARi. $n \geq 40$ organoids per dose, 3 biological replicates. P-values calculated with Dunnett's multiple comparison test after one way ANOVA. **e-f)** Relative viability determined by CellTiter-Glo luminescence of organoids exposed to various concentrations of AtRA (e) or RARi (f) for 72 hours. All three biological replicates plotted as points, while curve is drawn as a spline function. **g-h)** Relative caspase activity determined by Caspase-Glo luminescence of organoids exposed to various concentrations of AtRA (g) or RARi (h) for 72 hours. All three biological replicates plotted as points, while curve is drawn as a spline function.

For all figures with p-values indicated by asterisks, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. ns indicates non-significant comparisons. Boxplot represents median, 25th percentile and 75th percentile, whiskers extend from min to max.

Cornification signature is predictive of AtRA induced invasion and proliferation, and RAR α induced cell death in mouse and human breast tumor organoid models

Given the robust effects of retinoic acid signaling perturbations in the MMTV-PyMT organoid model, we next sought to determine the generality of this response. We chose two models for further study: C3(1)/tag, a mouse model of basal breast cancer, and MMTV-Neu a model of luminal breast cancer. By timelapse imaging, C3(1)/Tag organoids treated with AtRA displayed a pronounced increase in invasion, including in some cases shattering of organoids to individually highly migratory cells (**Fig. 3a-b and Video 3**). By CTG assay, C3(1)tag organoids had reduced cellular viability in response to RAR α with an IC₅₀ of 155 nM (**Fig 3c**). In contrast, MMTV-Neu, the luminal model, had minimal responses to either AtRA or RAR α (**Fig. 3a-c and Video 4**). Given the marked differences between models, we next hypothesized these models might have differences in epidermal gene expression. To answer this question, common mouse models of breast cancer were ranked by the epidermal cornification signature identified from scRNA-seq (**Fig 3d, Supplementary Table 2**)¹⁴⁰. This demonstrated that C3(1)/Tag organoids were among top scoring model for epidermal signature, whereas MMTV-Neu scored the lowest.

To extend these results further, we next asked whether epidermal cornification signature was predictive of AtRA and RAR α response in human breast tumors. Human patient derived breast tumor organoids with RNA-seq data were ranked by cornification score¹⁴¹ (**Fig 3e**). Two TNBC models were tested, TNBC216 and TNBC237, to represent high and moderate epidermal cornification scores, respectively. We treated each with AtRA and RAR α and determined invasion by collagen assay and CellTiter-Glo.

We found that invasion rates in both tumor organoid lines were responsive to AtRA and BMS493 (**Fig 3f-g, Ext data 4a-b, Videos 5, 6**). In a similar manner to the mouse model organoids, TNBC216 – which had the higher cornification score - was more sensitive to RARi than TNBC237 (**Fig 3h**). Taken together, these data indicate that cornification epidermal transcriptional signature is predictive of AtRA induced invasion and RAR inhibitor induced cell death in both mouse and human breast cancer organoid models.

Figure 3

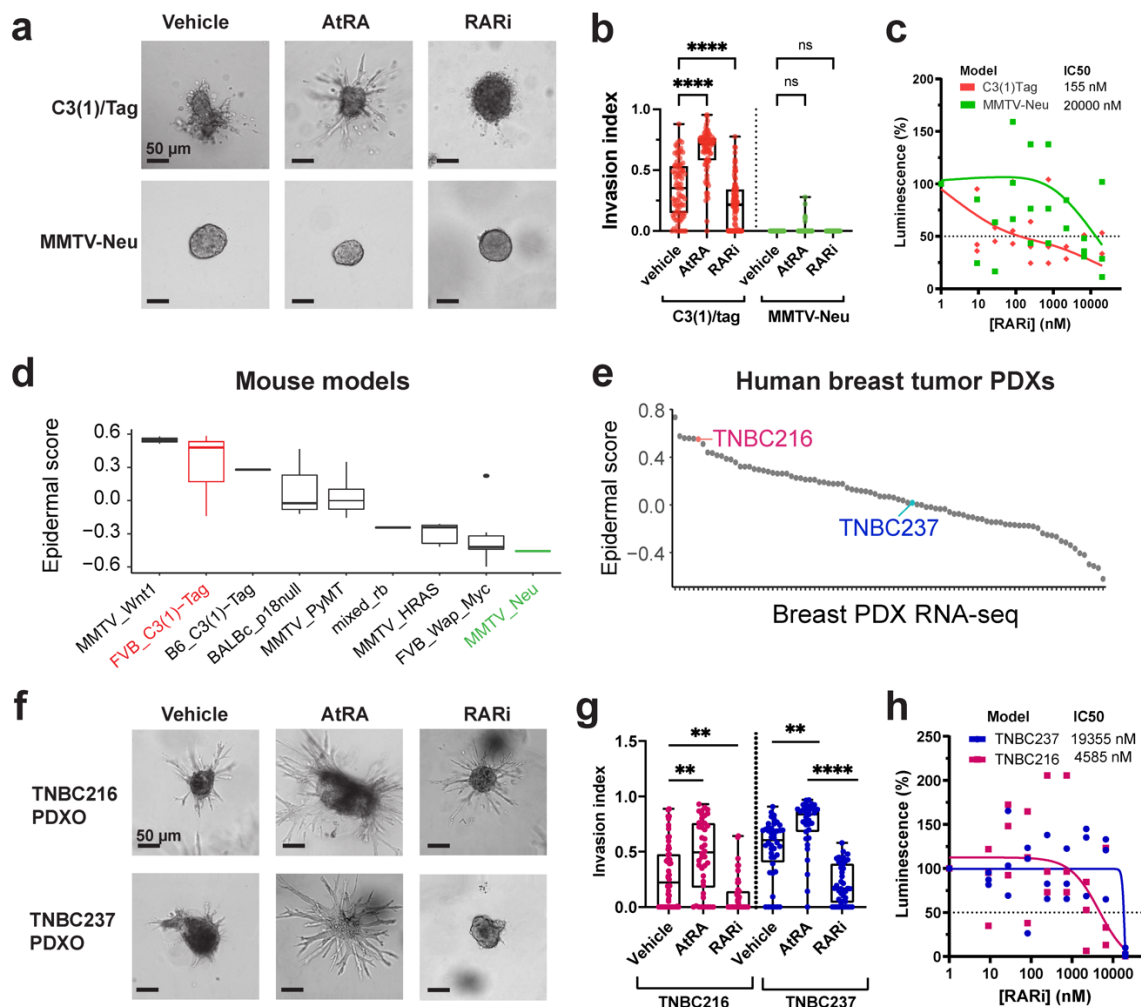


Figure 3. Epidermal gene expression predicts response to retinoic acid receptor signaling perturbations in mouse and human breast tumor organoids.

a) Representative endpoint DIC microscopy timelapse images of basal-like C3(1)/Tag or luminal MMTV-Neu mammary tumor organoids embedded in 3D collagen at treated with either 1 μ M AtRA or 5 μ M RARi (BMS493). **b)** Invasion indices of C3(1)/Tag and MMTV-Neu organoids treated with either 1 μ M AtRA or 5 μ M RARi. $n \geq 90$ organoids per condition. P-values calculated using Dunn's multiple comparison test after Kruskal-Wallis test. Boxplot represents median, 25th percentile and 75th percentile, whiskers extend from min to max. **c)** Relative CellTiter-Glo luminescence of C3(1)/Tag and MMTV-Neu organoids treated with a dose range of RARi. Curve was drawn with a spline equation. $n = 3$ passages of organoids on different days. **d)** Cornification scores calculated for mouse models of breast cancer using bulk tumor RNAseq. Tukey boxplots used. **e)** Cornification scores calculated for several PDX models of breast cancer using bulk tumor RNAseq. *Figure legend continues on following page.* **f)** Representative endpoint DIC microscopy timelapse stills of TNBC216 and TNBC237 tumor organoids embedded in 3D collagen at treated with either 1 μ M AtRA or 5 μ M RARi. **g)** Invasion indices of TNBC216 and TNBC237 organoids treated with either 1 μ M AtRA or 5 μ M RARi. $n \geq 20$ organoids per condition, 3 technical replicates. P-values calculated using Dunn's multiple comparison test after Kruskal-Wallis test. Boxplot represents median, 25th percentile and 75th percentile, whiskers extend from min to max. **h)** Relative CellTiter-Glo luminescence of TNBC216 and TNBC237 organoids treated with a dose range of RARi. Curve was drawn with a nonlinear inhibition equation in prism. $n = 3$ passages of organoids on different days. **note:** For all figures with p-values indicated by asterisks, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. ns indicates non-significant comparisons.

RAR activation in breast tumor organoids limits epidermal differentiation and keratinization

Given the specificity of RAR signaling in epidermal-like breast cancer models, we next asked how RAR signaling modulates gene expression on a molecular level. To address this question, we performed RNA-seq on AtRA and RARi treated PyMT organoids. As expected, canonical retinoic acid response gene Cyp26a1 and RAR target genes were strongly induced by AtRA treatment and reduced by RARi treatment in a dose dependent manner (**Ext Data 5a-b**). By PCA plot, AtRA treatment induced large scale changes in gene expression, whereas RARi induced more modest changes, likely due to widespread cell death associated with RARi (**Ext Data 5c**). Top gene programs induced by AtRA included genes involved in O-linked glycosylation, ER protein processing, and fat-soluble vitamin/retinol metabolism (**Fig 4a, Ext data 5f**). Notably,

tumor cells retained epithelial gene expression and did not induce markers of canonical EMT or myoepithelial cells (**Ext data 5d-e**). Top gene programs suppressed by AtRA included genes involved in epidermis development, barrier establishment, and hair follicle development (**Fig 4b**). These gene sets localized to invasive cells in our single cell RNAseq data, whereas the AtRA upregulated gene set localized to early timepoint cells (**Ext data 5g-j**). Consistent with the effects of AtRA in cultured keratinocytes¹⁴²⁻¹⁴⁴, genes associated with epidermal terminal differentiation were uniformly downregulated by AtRA (**Fig 4c**) which we further confirmed on an RNA level by qPCR for Krt14, Krt5, and Krt6a (**Ext Data 5k**, and by immunofluorescence for K14 and K17 (**Ext data 6d-e**). Conversely, RAR α treated organoids displayed increased expression of basal keratins K14 and K17 (**Ext data 6d-e**), and to a lesser extent AE13, an antibody reactive with hair keratins (**Ext data 6f**). Taken together, these observations reveal that retinoic acid signaling induces a non-EMT migratory program and limits epidermal differentiation and keratinization.

Genetic effects of perturbing epidermal differentiation factor KLF4 in breast cancer organoids.

To complement these pharmacologic perturbations, we next sought to test the effect of regulating epidermal differentiation state genetically. In MMTV-PyMT tumor organoids, top basal leader associated transcription factors included KLF4, which is required for formation of the cornified skin barrier (**Fig 4d**)¹⁴⁵⁻¹⁴⁸. In MMTV-PyMT organoids, KLF4 knockdown suppressed transcription of epidermal genes, including Krt6a, 14, and Tp63 (**Fig 4e-f**). K14 was also downregulated on a protein level (**Ext data 7a**). Functionally,

KLF4 knockdown markedly increased collective invasion (**Fig 4 g-h, and Video 7**). Conversely, despite many attempts, constitutive KLF4 overexpression under antibiotic selection was not tolerated in MMTV-PyMT tumor organoids. To rule out transduction inefficiency, we generated a doxycycline inducible KLF4 overexpression vector and stably integrated this vector in MMTV-PyMT organoids. Inducing KLF4 expression with doxycycline markedly reduced invasion, decreased viability (**Fig 4 e, g-i and Video 8**), and markedly increased epidermal gene expression (**Fig 4f**). It should be noted that MMTV-PyMT organoids lose invasive capacity if passaged or kept in culture for the amount of time necessary for transduction and selection, leading to a reduced level of invasion in controls. KLF4 knockdown did not lead to changes in Cyp26a1 expression, indicating that reductions in Klf4 do not likely impact RAR activity (**Ext data 7b**). However, AtRA treatment did lead to reductions in Klf4 transcripts, indicating RAR is a regulator of Klf4 (**Ext data 7c**). Together, these data reveal multiple routes for perturbing epidermal differentiation state that when applied to basal breast tumors potentially regulate their collective invasion and tumor cell viability.

Figure 4

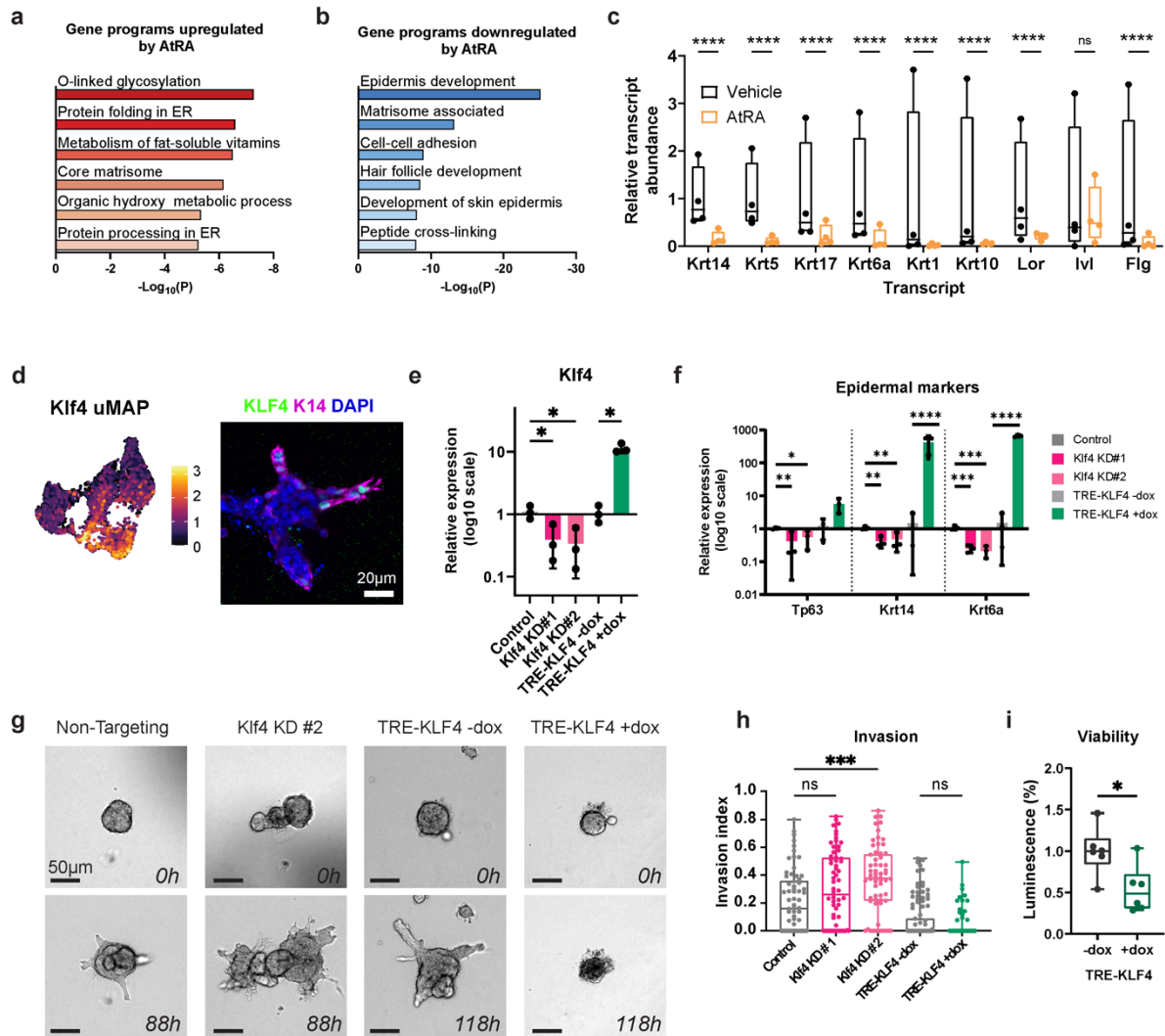


Figure legend on following page.

Figure 4. Regulation of collective invasion and viability by genetic and pharmacologic regulators of epidermal differentiation.

a-b) Top gene programs upregulated (a) and downregulated (b) by 1 μ M AtRA exposure for 72 hours in MMTV-PyMT organoids embedded in 3D collagen. Bulk RNAseq data gene set enrichment analysis performed with Metascape. **c)** Specific epidermal marker transcripts from bulk RNAseq of vehicle and AtRA treated organoids. For each gene transcript, counts were additionally normalized to the mean of controls. n = 4 biological replicates. P-values calculated from differential expression analysis with DESeq R package. **d)** uMAP showing transcription of Klf4 in scRNA-seq of invading MMTV-PyMT organoids, along with Klf4 stain in invading organoids. scale bar = 20 μ m. **e)** qPCR validation of Klf4 knockdown and overexpression. Bars represent mean of all replicates (points) with SD. N = 3 biological replicates. P-values calculated using fishers test after one-way ANOVA. **f)** Relative expression of selected epidermal markers in the context of Klf4 knockdown and overexpression. N = 3 biological replicates. P-values calculated using Fisher's test after one-way ANOVA. Bars represent mean of all replicates (points) with SD. **g)** Representative DIC images of organoids at start and endpoint in collagen in the context of Klf4 knockdown and overexpression. Scale bar = 50 μ m. **h)** Invasion indices for Klf4 knockdown and overexpression organoids. n $\geq$ 60 organoids per condition, 3 biological replicates. P-values calculated with Kruskal-wallis test following one-way ANOVA. **i)** CellTiter-Glo luminescence values of TRE-KLF4 organoids with and without doxycycline. n = 3 biological replicates, performed in triplicate. Controls for each technical replicate were compared to average of the two. P-values calculated with ratio paired T-test.

note: For all figures with p-values indicated by asterisks, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. ns indicates non-significant comparisons. Boxplots represents median, 25th percentile and 75th percentile, whiskers extend from min to max.

RAR signaling antagonism markedly inhibits tumor growth and metastatic progression in vivo

We next tested the effect of suppressing retinoic acid signaling in vivo. To do so, we first transplanted MMTV-PyMT tumor organoids into immunocompromised NSG mice and then upon reaching 200 mm² in size, began treatment with a 4-week course of vehicle or RARi (BMS493) 40mg/kg given intraperitoneally (i.p.) three times weekly. In vivo, shutting off RAR signaling markedly suppressed tumor outgrowth and prevented the formation of spontaneous metastasis (**Fig. 5a-c, Ext data 8a**). Importantly, RARi was well tolerated, with no differences in weight between treatment conditions and no marked differences in histology of contralateral normal mammary glands (**Ext Data 8b-c**). By H&E, tumor histology was consistent with poorly differentiated carcinoma in both

control and RARi treated tumors. However, RARi treated tumors were much smaller and had markedly fewer areas of necrosis, a histologic feature associated with metastatic dissemination (**Fig 5d-f**)¹⁴⁹. Given the effect of RARi on basal keratin expression in organoids, we next performed immunohistochemical staining of tumors for K14. This revealed a change in cellular organization and composition. In control tumors, K14+ cells were located at the basal edges of tumors, while in RARi tumors, K14+ cells were spread throughout the tumor (**Ext data 8d**). Numerically, total K14+ cells were less abundant but accounted for relatively higher percentage of cells (**Ext data 8e**), and mean K14 expression was decreased on individual cell basis with RARi treatment (**Ext data 8f**).

Extending these results further, we next performed a preclinical study of RARi in immunocompetent MMTV-PyMT mice harboring spontaneous tumors. In the immunocompetent setting, RARi led to symptoms of xerophthalmia, a common side effect of vitamin A deficiency¹⁵⁰, precluding long-term treatment. Notably, short term RARi also suppressed tumor growth and necrosis in immunocompetent MMTV-PyMT mice (**Ext data 8g-j**).

Given the strong impact of RARi on primary tumors, we next sought to define effects of RARi on metastatic colonization. We injected fluorescently tagged MMTV-PyMT clusters into the tail veins of NSG mice, and after 3 days, initiated treatment with RARi for 2 weeks and thereafter, assessed for metastatic burden. RARi markedly reduced the number and size of metastases (**Fig. 5g-h**). Next, to model the impact on already established metastases, we injected MMTV-PyMT tumor cells into tail veins of NSG mice, and thereafter allowed the metastases to grow for two weeks before

initiating a 3-week course of RARi. Likewise in the already-established metastasis setting, RARi treatment markedly reduced metastatic tumor burden (**Fig. 5g,i**).

Complementary orthotopic transplant and experimental metastasis studies were conducted treating mice with AtRA at a dose chosen in the range given to acute promyelocytic leukemia patients (12 mg/kg, 3x weekly). In these experiments, no effect of AtRA on tumor growth or metastasis was observed in these mice fed a vitamin A sufficient diet (**Ext data 8k-m**). In total, these data show that RAR signaling suppression, not activation, markedly inhibits tumor growth, metastatic seeding and metastatic progression in an aggressive mouse model of breast cancer.

Figure 5.

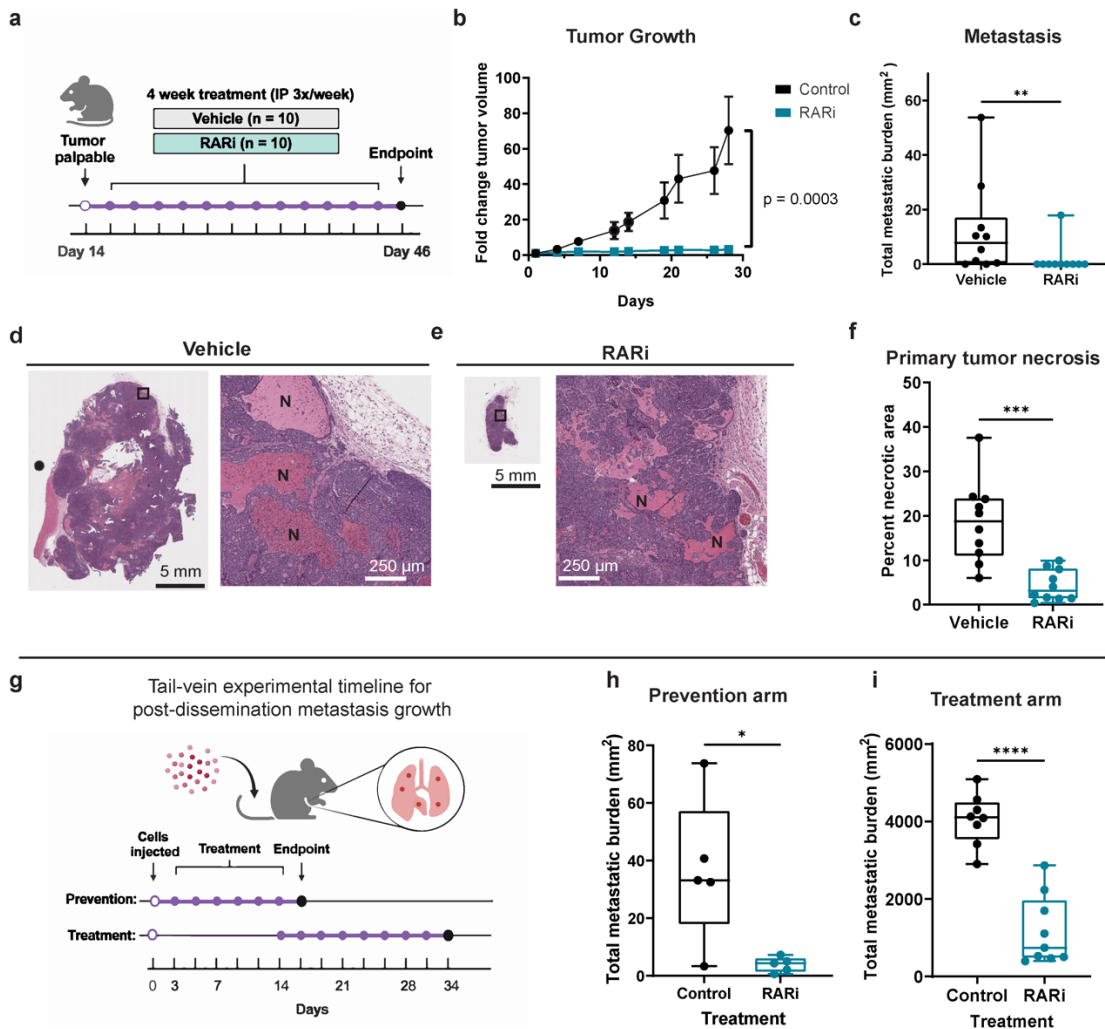


Figure 5 Retinoic acid receptor inhibition suppresses tumor growth and metastasis in vivo

a) Schematic depicting timeline for orthotopic transplant experiment. Made with Biorender. **b)** Fold change in tumor growth over time in mice treated with either vehicle or BMS493. $n = 10$ mice per treatment. P-values calculated using Mann-Whitney Test. **c)** Total macro-metastatic area visible on the outside of the lungs of mice by fluorescent detection of tumor cells at endpoint. $n = 10$ mice per treatment. P-values calculated using Mann-Whitney Test. **d-e)** Representative whole and zoomed images of H&E stained sections of tumors from mice treated with either vehicle (d) or BMS493 (e). Sections marked with "N" depict large necrotic regions. **f)** Percent of total tumor area that was necrotic in mice, quantified by nuclei density in qupath. $n = 10$ mice per treatment. P-values calculated using Mann-Whitney Test. **g)** Schematic depicting experimental timelines for prevention and treatment tail-vein experimental arms. **h)** Total macro-metastatic area visible on the outside of the lungs of mice by fluorescent detection of tumor cells at endpoint. $n = 5$ mice per treatment. P-values calculated using Mann-Whitney test. **i)** Total

macro-metastatic area visible on the outside of the lungs of mice by fluorescent detection of tumor cells at endpoint. n = 5 mice per treatment. P-values calculated using Mann-Whitney Test. **note:** For all figures with p-values indicated by asterisks, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. ns indicates non-significant comparisons. Boxplots represents median, 25th percentile and 75th percentile, whiskers extend from min to max.

Vitamin A supplement intake is associated with recurrence and mortality in ER-negative breast tumors

Given the robust effects of vitamin A signaling on basal breast tumor cell invasion and proliferation in our experimental models, we next asked whether vitamin A supplement intake was associated with recurrence and mortality in patients with history of breast cancer. To answer this question, we evaluated the clinical impact of vitamin A supplement intake within two observational cohort studies. The first was the Quilt Study, an NCI-funded, population-based prognostic cohort of women with invasive breast cancer. These study participants were ascertained through the Seattle/Puget Sound SEER cancer registry from 1997-1999, were interviewed multiple times, had extensive medical record reviews, and were monitored for cancer mortality and recurrence for over two decades¹⁵¹. Quilt Study participants diagnosed with breast cancer at ages 65-79 years (n=975) also previously participated in a case-control study that included interviews with questions on longitudinal supplement use up to the time of breast cancer diagnosis¹⁵². In this study, among ER negative breast cancer cases, we observed that vitamin A intake was associated with a 4.2-fold increased risk of breast cancer recurrence (95% CI, 1.5-11.6), compared to women who used other supplements. There was no association of vitamin A supplement intake with breast cancer recurrence among women with ER positive tumors (HR=1.0, 95% CI, 0.4-2.4). In alignment with our mechanistic studies in vitro and in vivo, the interaction of ER status with vitamin A

associated breast cancer recurrence was statistically significant (P-value of interaction=0.04), indicating that the increased risk associated with vitamin A was confined to women with ER-negative tumors. We also observed a 2.3-fold increased risk of breast cancer-related mortality (95% CI, 0.7-7.4) (**Fig. 6a-b, Supplementary table 3**) associated with vitamin A intake among women with ER negative tumors, whereas no such association was observed for women with ER positive tumors.

Extending these results further, we next examined incidence of breast cancer in relation to vitamin A intake. Notably, when the 975 cases were compared to the 1,007 age- and county-matched controls without breast cancer from the PACE Study ^{153,154}, Vitamin A use before diagnosis was associated with a 3.0-fold increased risk of ER-negative breast cancer (95% CI 1.6-5.8) but was not associated with risk of ER-positive breast cancer (odds ratio=0.9, 95% CI 0.6-1.3; p-value for difference in estimates < 0.01). Taken together, these findings provide epidemiologic support for the association of vitamin A intake, tumor incidence and metastatic potential specifically in ER-negative breast tumors.

The second observational cohort analyzed was the Pathways Study, an NCI-funded, prospective cohort of 4,504 Kaiser Permanente Northern California (KPNC) members with a history of breast cancer ¹⁵⁵. Women were diagnosed from 2005-2013, enrolled on average about two months of their diagnosis, and followed for breast cancer recurrence and mortality through 2021. At baseline at the time of enrollment and after 6 and 24 months of follow-up, participants completed questionnaires on their use of vitamin A and beta-carotene supplements. Analyses examined whether timing or duration of vitamin A and/or beta-carotene supplements before and after a breast

cancer diagnosis were associated with different risks of recurrence or mortality (**Figures 6c-f, Ext data 9a-b**). Among women who used any dietary supplements at baseline, 6 or 24 months, we observed a 1.56 increased risk of breast cancer recurrence (95% CI=1.01-2.42) in women who took vitamin A and/or beta-carotene before and after diagnosis and a 1.61 increased risk of recurrence (95% CI=1.08-2.41) in women who took vitamin A and/or beta-carotene for 3 or more years prior to their diagnosis relative to women who did not use vitamin A and/or beta-carotene supplements (**Supplementary Table 4**). Similar associations were observed examining only breast cancer related mortality [use before and after diagnosis: HR=1.45 (95% CI=0.89-2.37); use for 3 or more years: HR=1.45 (95% CI=0.91-2.31)] but estimates were not statistically significant, likely due to fewer events for mortality than overall recurrence. These findings suggest that longer duration of exposure to vitamin A and/or beta-carotene may increase risk of recurrence and death in women with history of breast cancer. In our examination by ER-status, we observed no differences in risk of recurrence between vitamin A and/or beta-carotene users among women with either ER-negative [HR=0.85 (95% CI=0.40-1.83)] or ER-positive [HR=1.26 (95% CI=0.93-1.69)] breast cancer (p-value for interaction=0.35). While these findings contrast with the Quilt study's findings, differences may be due to different ER-subtype ascertainment methods between cohorts and/or advances in breast cancer therapies from the time Quilt participants were treated to when Pathways participants were treated.

Figure 6

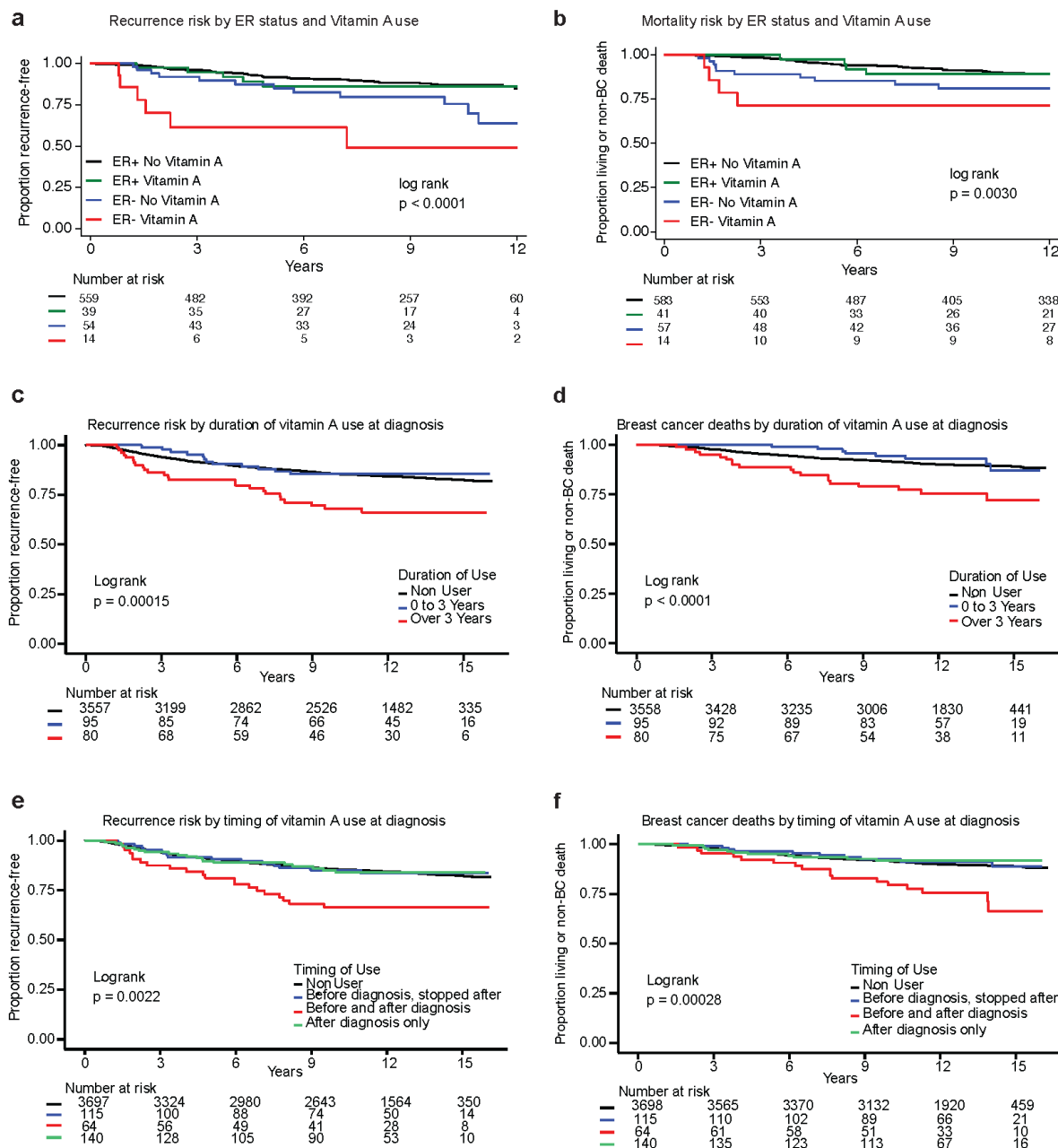


Figure 6. High dose vitamin A supplementation correlates with poor outcomes in ER-breast cancers

a-b) Kaplan-Meier curves for time to recurrence (a) or cancer related mortality (b) after breast cancer diagnosis, stratified by ER status and Vitamin A supplement use among PACE/Quilt Study participants ages 65-79 at diagnosis. Vitamin A supplement use included supplements named vitamin A, beta-carotene, and/or cod liver oil. *Figure legend continued on following page.*

c-d) Unadjusted Kaplan-Meier curves for time to breast cancer recurrence (c) or breast cancer mortality (d) stratified by years of vitamin A and/or beta-carotene supplement use prior to breast cancer diagnosis among Pathways Study participants.

e-f) Unadjusted Kaplan-Meier curves for time to breast cancer recurrence (e) or breast cancer mortality (f) stratified by starting, stopping, and continuing vitamin A and/or beta-carotene supplement use at time of breast cancer diagnosis among Pathways Study participants.

DISCUSSION

Tumor cell plasticity is linked to metastatic aggression and therapy resistance, but could also introduce latent vulnerabilities amenable to therapeutic attack. In this study, we employed single cell transcriptomics in invasive breast tumor organoids to reveal the molecular cartography of basal epithelial plasticity. This revealed that basal breast tumor cells harbor a transcriptional program resembling skin epidermis. In turn, this molecular map yielded a prediction for directing tumor cellular state transitions. By perturbing epidermal differentiation programs pharmacologically or genetically, we induced potent changes in basal tumor invasion, proliferation, and cell survival. We conclude that in basal breast tumors, RAR signaling is a molecular rheostat that promotes a hyper-migratory and proliferative phenotype, and when suppressed, induces terminal epidermal differentiation, hyperkeratinization, and acute cell death.

In addition, this study advances recent conceptual understanding of cell plasticity states and their relation to collective migration^{38,110,156-160}. AtRA induces increased directed cell migration and a hypermigratory phenotype, yet this molecular transition appears distinct from classical EMT^{35,91,161}. The importance of intermediate states along the EMT spectrum is relevant for understanding dynamics along the continuum of metastatic stages^{33,34,85,158,162}. Here, our data suggest the existence of cellular states along an epidermal differentiation spectrum that variably contribute to collective invasion, and that potentially links tumor cell migration to programs of epidermal wound

repair ^{143,163}. In addition, recent studies indicate the benefit of pushing mesenchymal cells past the point of no return- such as via PPAR γ agonists to drive metastatic cells toward adipocyte differentiation ^{103,109}. Our data show that antagonizing the RAR signaling pathway induces a similarly favorable effect by irreversibly directing basal breast tumor cells toward an epidermal differentiated fate, thereby limiting their invasive and metastatic potential. We suggest that patients with basal breast tumors, a molecular phenotype particularly enriched among poor-prognosis triple-negative breast cancers ^{101,164,165}, are candidates for RAR antagonists and other therapies under the umbrella of cornification inducing therapies. In other tumor types such as acute promyelocytic leukemia and pediatric neuroblastoma, retinoic acid is a potent differentiation-inducing agent ⁷⁶. We suggest different therapeutic paradigms ought to be applied when considering distinct tumor context, cell-of-origin, and developmental potential.

A question of particular importance is how environmental and dietary exposures might modulate differentiation state in breast tumors with basal phenotype. Over 70% of patients with history of breast cancer are reported to take dietary supplements, many of which include vitamin A and many at average doses exceeding Institute of Medicine guidelines ^{166,167}. Importantly, the health impact of retinoids and other vitamin A derivatives, which are plentiful in the developed world, has remained controversial. Our experimental and epidemiologic data provide a framework for thinking about vitamin A exposures, in which tumor molecular phenotype modifies this exposure risk. In tumor models with basal epidermal phenotype, AtRA had large tumor invasive and promoting effects; in tumors with luminal characteristics, the effect was negligible. Vitamin A may produce exposure dependent transcriptional signatures that when paired with other

genetic profiles ¹⁶⁸⁻¹⁷¹ better predict recurrence risk and therapy response. This study sets the stage now for future work to understand how these risks might be tailored on an individual level, requiring an understanding of how retinoid intake (both natural metabolites and synthetic derivatives) modify risk, at what levels, and in which molecular cancer contexts.

ACKNOWLEDGMENTS

We thank the patients participating in this study for their generosity. We thank our funding sources: Department of Defense Breast Cancer Research Program (BCRP; W81XWH-18-1-0098), Komen Career Catalyst Research Grant (CCR18548236), the NIH/NCI (R37CA234488), the Burroughs Wellcome Fund Career Award for Medical Scientists, the V Foundation (V Foundation Scholar Award), Seattle Translational Tumor Research, the Phi Beta Psi Society, the NIH/NCI (T32CA080416) Interdisciplinary Training in Cancer Research, R01CA105274 (PI: Kushi), U01CA195565 (MPIs: Kushi, Ambrosone) and Administrative Supplement U01CA195565-05S1 (Lead: Greenlee); R01CA241409 (MPIs: Kushi, Iyengar); R01 CA098858 (PI: Malone). We thank Kelsey Fleming at Fred Hutchinson Cancer Center for her assistance with managing our animal colony and performing in-vivo experiments. We additionally thank Tracy Goodpaster for her assistance with IHC experiments. This research was supported by the Experimental Histopathology, Genomics, and Comparative Medicine Shared Resources of the FHCC/University of Washington Cancer Consortium (P30 CA015704).

METHODS

Animal models

All mice were maintained under specific pathogen-free conditions, and experiments conformed to the guidelines as approved by the Institutional Animal Care and Use Committee of Fred Hutchinson Cancer Center (FHCC). FVB and MMTV-PyMT (FVB/N-Tg(MMTV-PyVT)634Mul/J) were obtained from Jackson Labs. NSG (NOD *skid* gamma) mice were bred by Fred Hutchinson Cancer Center shared resources.

Mouse organoid preparation and culture

Organoids were isolated from MMTV-PyMT, C3(1)TA_g, or MMTV-Neu mouse mammary tumors as previously described (Nguyen-Ngoc et al., 2015, Wrenn et al., 2020). Mice were harvested as the largest tumor neared 1-2 cm in diameter. Mammary tumors were dissected, minced with a scalpel, and then digested in a collagenase-trypsin solution for 30–60 minutes shaking at an angle at 100–150 rpm at 37°C. The digestion solution contained 2 mg/mL collagenase (Sigma C2139), 2mg/mL trypsin (GIBCO 27250–018), 5% fetal bovine serum, 5 µg/mL human insulin (Sigma-Aldrich I9278), 50 µg/mL gentamicin (GIBCO 15750–060) in 20 mL of DMEM/F12. After digestion, larger tumor fragments were removed by allowing them to settle in the tube, while the remaining solution was moved to a new tube. Single cells and debris were removed from the smaller fragment solution by centrifuging for 4 seconds at 450 g to isolate multicellular organoids, then resuspended in 10 mL DMEM/F12. This was repeated for a total of 4 washes. For suspension culture, clusters were cultured in non-adherent dishes in organoid media: DMEM/F12 (Thermo Fisher 10569044), 1% Penicillin-Streptomycin

(Sigma-Aldrich P4333), 1 % ITS-X(Thermo Fisher 51500056), 220 ng/mL hFGF-Basic (Peprotech 100-18C), and 2% (v/v) reduced growth factor Matrigel (#). Tips and tubes used to handle organoids were first coated in 2.5% BSA in DPBS to prevent loss of material.

Human organoid preparation

TNBC216 (J000106527/ BCM-3469) and TNBC237 (J000106528/ BCM-2147) patient derived xenografts (PDXs) were obtained from Jackson labs and maintained in mice by the preclinical modeling core lab at Fred Hutch Cancer Center. After tumor collection, tumors were processed into organoids using the same methodology as used for mouse tumor organoids, above.

All PDX organoids were cultured in suspension in media previously published (cite) containing Advanced DMEM/F12, 5% FBS, 1X HEPES, 1X Glutamax, 1 ug/mL hydrocortisone, 0.1 µg/mL gentamicin, 10 ng/mL human EGF, 10 µM Y-27632, 100 ng/mL FGF2, 1 mM NAC and 5% (v/v) reduced growth factor matrigel.

Preparation of rat-tail collagen

Our collagen preparation protocol was developed with guidance from Dr. Kayla Bayless (Bayless et al 2009), and adapted from Bornstein et al 1958 and Rajan et al 2006.

Freeze dried rat tails were purchased from Pel-freez and stored at -80°C. For collagen preps, two tails were thawed at room temperature in 70% ethanol. Sections of tail were progressively removed, pulling the collagen fibers from the tail. The collagen fibers were

cleaned in water, then cut into small pieces and dissolved in 0.1% acetic acid (Millipore Sigma A6283, dissolved in sterile ultra-pure water) for 8 days at 4°C, occasionally centrifuging the collagen and collecting the supernatant. All collected supernatant was split into 50 mL conical tubes and frozen in dry ice. The collagen was then lyophilized with the assistance of the Proteomics and Metabolomics Core at the Fred Hutchinson Cancer Center. The dried collagen was then weighed and reconstituted at a concentration of 4.5 mg/mL in 0.1% acetic acid.

3D collagen culture of organoids or spheroids for invasion

Organoids were embedded in 1.5 mg/mL collagen gels, which were prepared from the 4.5 mg/mL rat-tail collagen described above using protocols previously published (Cheung et al 2013). 4.5 mg/mL collagen was diluted to 1.8 mg/mL in 0.1% acetic acid. This was then mixed with 10X DMEM, 1 N NaOH, and ultrapure water at the following ratio: *1 mL collagen (1.8 mg/ml): 100 μ L 10X DMEM: 32 μ L NaOH: 66.66 μ L H₂O*. This mixture was allowed to polymerize at 4°C and was then mixed with organoids at a concentration of ~180 organoids/100 μ L. Glass-bottom 24 well plates were prepared by plating 40 μ L of collagen (without organoids) in the center of each well to be used. The plate was placed at 37°C for 5-10 minutes to allow the collagen disks to polymerize, after which 100 μ L of the collagen/organoid mixture was plated on top as a dome. This was allowed to polymerize for an additional 45-60 minutes at 37°C. After complete polymerization, 1 mL of organoid media (described in organoid culture section, without Matrigel) was added to each well.

In-culture drug studies

BMS493 (Medchemexpress, HY-108529) and all-trans retinoic acid (AtRA, Stem cell technologies 72262) were reconstituted in DMSO. For treatment, drugs were further diluted in DMSO to achieve a 1000x concentration of each desired dose, which was then diluted (1:1000) in the appropriate media for the cells being treated. This was then added to the media of 3D collagen cell cultures. Since retinoids are light-sensitive and unstable, treated cells were kept from bright light exposure and the drug containing media was changed every 48 hours.

3D culture DIC time-lapse imaging

Differential interference contrast (DIC) time-lapse images were captured hourly using a Leica SPE at 10X magnification, while temperature was maintained at 37°C and CO₂ at 5%.

Invasion index

After 3 days of invasion in collagen, organoids were fixed in 4% PFA in PBS for 12-15 minutes at room temperature and stored at 4°C. 10X magnification DIC images of fixed organoids were acquired on a Leica SPE microscope. Using Fiji, the area of the total organoid and the center of the organoid were outlined and measured. The center area was subtracted from the total area, leaving the invasive area of the organoid. This was divided by the total area to calculate the fraction of the organoid that was invasive, or the invasion index.

Time-lapse fluorescent single cell tracking

For single cell tracking, membrane tomato (mTmG) MMTV-PyMT organoids were sparsely transduced with a H2B-GFP lentiviral vector generously provided by the Paddison Lab at FHCC. Fluorescent time lapse images were acquired every 20 minutes using an Andor CSU-W confocal spinning disk on a Leica DMI8 inverted microscope at 20X magnification. Temperature was maintained at 37°C and CO₂ at 5%. Z-stacks were max-projected and stabilized in Image-J. CellProfiler ¹⁷² was used to track nuclei.

Viability and caspase assays

50 µL of collagen containing 30 organoids were plated into each well of opaque white 96-well plates following the 3d collagen protocol detailed above. After plating, cells were incubated with 95 µL of Matrigel-free organoid media. In a separate round bottom 96-well plate, compounds were serially diluted 3-fold from max concentration in media to achieve a 20X drug concentration. 5 µL of 20X drug was then added to each well. Maximum concentration of BMS493 was 20 µM while maximum concentration of AtRA was 30 µM. A no-cell control, vehicle control, and puromycin control were included. All conditions were plated in triplicate. After 72 hours, CellTiter-Glo 3D (Promega) and Caspase3/7-Glo 3D (Promega) kits were used according to instructions and luminescence was read on a BioTek Synergy H1 Hybrid Multi-Mode Reader. After reading, the no-cell control background was subtracted from all wells. Each well reading was normalized to the average luminescent signal of the vehicle control and then were averaged, representing one replicate. Vehicle control and puromycin control were used as upper and lower constraints, respectively, for IC₅₀ calculation.

3D culture immunofluorescence

At the appropriate end point for 3D collagen cell culture experiments, media was removed and PBS was added for 5-10 minutes. Afterwards, 1 ml 4% paraformaldehyde was added to each well of the 24-well plate that contained cells for 12 minutes at room temperature. Afterwards, collagen gels were washed twice with PBS and stored at 4°C until staining. To permeabilize the cells, 0.5% triton-x (diluted in DPBS) was added to each well for 1 hour. The gels were then blocked with 0.1% triton-x, 10% fetal bovine serum (FBS), and 1% bovine serum albumin (BSA) for 3 hours at room temperature or overnight at 4°C. Primary antibodies were then diluted as instructed in the same blocking solution and added to the gels for 3 hours at room temperature or overnight at 4°C. The gels were then washed with DPBS 3 times for 10 minutes each. Secondary antibodies were diluted in 5% host serum (in DPBS) and added to the collagen gels for 1 hour. The gels were then washed 3 times with DPBS again. Confocal images were acquired using an Andor CSU-W confocal spinning disk on a Leica DMI8 inverted microscope.

Single cell RNA sequencing

Organoids were produced as described above and cultured for 48 hours in 2% Matrigel organoid media. For the 0h timepoint, a subset of organoids were collected and dissociated to single cells by submersing in accutax (Stem Cell Technologies # 0792) for 10-15 minutes with occasional mixing by pipette. For the other two timepoints, organoids were plated in 3.75 mg/ml collagen I using the protocol above. At the

appropriate time, organoids were removed from collagen gels using collagenase(1 mg/ml, MilliporeSigma C2139) in DMEM/F12 for 5-10 minutes. Afterwards, organoids were dissociated to single cells in accutax. Dissociated cells were spun down, resuspended in PBS, and filtered through a 40 μ m size filter. Cells were then counted using a countess, and resuspended in freezing media (Organoid media with 20% FBS and 10% DMSO) at a concentration of 1 million cells per 500 μ L. Cells were shipped to Genewiz (Azenta Life Sciences) and sequenced on the 10X 3' platform. Alignment and pre-processing were performed by Genewiz using 10X protocols.

Single cell RNA sequencing data analysis

Seurat V4 ¹⁷³ was used for processing and analysis of count matrices generated by 10X. Cells containing fewer than 500 detected genes or more than 10% mitochondrial genes were filtered out to discard low quality and dead cells respectively. Data was then log-normalized using Seurat normalization with a scale factor of 2000. 2000 variable features were determined using Seurat vst method. Stromal cells were defined and filtered out using the markers Col1a1, Fabp4, and Epcam and removed from data to allow for analysis of epithelial tumor cells. After each replicate and time point were processed as stated above, timepoints from individual replicates were merged. For each timepoint-merged biological replicate, cell cycle regression was performed using Seurat, after which data was scaled using standard Seurat protocol. UMAP figures were made by performing Seurat principal component analysis with 20 principal components, followed by UMAP projection with uwot method. To find genes correlating with Krt14, biological replicates were merged, and then correlation was performed on a matrix of

gene expression in each cell. The highest and lowest 1% of genes correlating with Krt14 were used as input for gene set enrichment analysis with metascape. Gene module scores were made with gene set lists reported in metascape after analyzing genes highly correlated with Krt14 (**Supplementary table #1**). Mouse cell type annotation was performed using scMCA ¹¹⁵.

Public dataset analysis

Each additional dataset used for figure 1 and extended figures 1 and 2 was processed in the same manner as our original data unless pre-processing had already occurred. For heatmap production, each dataset sample was summarized by average Krt14 expression and the correlation of Krt14 with each gene set module score. Values were plotted using ComplexHeatmap ¹⁷⁴.

qPCR

Organoids were removed from collagen gels using collagenase (1 mg/ml, MilliporeSigma C2139) in DMEM/F12 at 37°C for 5-10 minutes. The organoids were then washed with PBS, and spun down at 300 xg for 5 minutes. RNA was then extracted from the organoid pellet using the RNAeasy kit (Qiagen) according to manufacturer instructions. All RNA was quantified using a Nanodrop and stored at -80°C. cDNA was produced using the SuperScript IV VILO kit according to instructions. All qPCR primers were ordered from Millipore-Sigma (KiCqStart predesigned primers). qPCR was run with PowerUp Sybr Green master mix following instructions using a primer concentration of 200 nM.

RNA-seq

Organoids were embedded in 3D collagen and treated with either BMS493 at a concentration of 5 μ M or AtRA at a concentration of 1 μ M for 72 hours. Doses were chosen to capture maximal phenotypic changes while cell viability was maintained. RNA was extracted from organoids as described above for qPCR. RNA was sent to Genewiz (Azenta) for sequencing and processed as follows:

Total RNA samples were quantified using Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and RNA integrity was checked with 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA). Samples were initially treated with TURBO DNase (Thermo Fisher Scientific, Waltham, MA, USA) to remove DNA contaminants. The next steps included performing rRNA depletion using QIAseq® FastSelect™-rRNA HMR kit (Qiagen, Germantown, MD, USA), which was conducted following the manufacturer's protocol.

RNA sequencing libraries were constructed with the NEBNext Ultra II RNA Library Preparation Kit for Illumina by following the manufacturer's recommendations. Briefly, enriched RNAs are fragmented for 15 minutes at 94°C. First strand and second strand cDNA are subsequently synthesized. cDNA fragments are end repaired and adenylated at 3'ends, and universal adapters are ligated to cDNA fragments, followed by index addition and library enrichment with limited cycle PCR. Sequencing libraries were validated using the Agilent TapeStation 4200 (Agilent Technologies, Palo Alto, CA, USA),

and quantified using Qubit 2.0 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA) as well as by quantitative PCR (KAPA Biosystems, Wilmington, MA, USA). The sequencing libraries were multiplexed and clustered on one lane of a flowcell. After clustering, the flowcell was loaded on the Illumina HiSeq 4000 instrument according to manufacturer's instructions. The samples were sequenced using a 2x150 Pair-End (PE) configuration.

Image analysis and base calling were conducted by the HiSeq Control Software (HCS). Raw sequence data (.bcl files) generated from Illumina HiSeq was converted into FASTQ files and de-multiplexed using Illumina's bcl2fastq 2.17 software. One mismatch was allowed for index sequence identification.

RNA-seq analysis.

Prior to alignment, RNA-seq reads were trimmed using Trim Galore to remove adapter sequences and low-quality bases, and read quality was confirmed using fastQC.

Trimmed reads were aligned to the *Mus musculus* genome using STAR¹⁷⁵. Paired reads with both mates aligning to the genome were counted at the gene level using feature Counts for downstream differential expression analysis.

Differential expression analysis was performed following the limma-voom procedure^{176,177}. Prior to differential expression analysis, very low abundance genes were filtered using the filterByExpr function in edgeR with default parameters. Effective library sizes were calculated using the trimmed mean of M-values (TMM) normalization method and

reads were processed in voom to generate log2-transformed values and precision weights for limma. Using the limma, normalized expression values were fit to a random effects linear model of the form $\text{Expr} \sim \text{Treatment} + \text{Mouse}$, where Treatment denotes the retinoic acid treatment received, and Mouse denotes the identity of the mouse from which the organoids were derived. The differential expression between treatment groups were then obtained as contrasts using the contrasts.fit function in limma. Empirical Bayes smoothing was applied on both the linear model and fitted contrast coefficients using edgeR. False-discovery rates for differential expression at the gene-level were calculated with the Benjamini-Hochberg correction procedure using the toptable function in limma. Normalized transcript abundances are reported as the TMM-normalized counts-per-million obtained from the voom function in limma.

In-vivo drug studies

BMS493 treatment was based upon the work described in Viragova et al 2023¹⁷⁸. A stock concentration of 50 mg/mL BMS493 (Medchemexpress, HY-108529) was prepared in DMSO and kept at -20°C in aliquots. On the day of treatment, the BMS493 was diluted to a final concentration of 2 mg/mL in 0.15M hydroxypropyl-β-cyclodextrin (Cayman Chemicals, 16169). This was then given to mice via intraperitoneal injection (I.P.) at a final dose of 40 mg/kg (1 mg/25 g, adjusted for weight of each mouse). AtRA treatment followed the same general protocol. Stock concentration of AtRA was 15 mg/mL in DMSO which was diluted to 0.6 mg/ml in 0.15M hydroxypropyl-β-cyclodextrin. This was then given to mice via intraperitoneal injection (I.P.) at a final dose of 12 mg/kg. Vehicle control was made by dissolving an equal volume of DMSO in 0.15M

hydroxypropyl- β -cyclodextrin. Mice were treated with BMS493, AtRA, or control 3 times a week (Monday, Wednesday, Friday) and monitored for negative effects. Side effects consistent with vitamin A deficiency were observed starting after one week of treatment including eye swelling and secretions.

In-vivo orthotopic tumor growth experiments

Tumor organoids were collected from MTMG MMTV-PyMT mice as described above. The day before transplant, tumor organoids were dissociated to single cells by suspending in accumax for 10 minutes. Cells were counted and allowed to re-aggregate into clusters for 24 hours in organoid media with 2% matrigel. The next day, spheroids were resuspended in a 1:1 matrigel:DMEM/F12 at a concentration of 3 million cells/100 μ L. 20 μ L of this mixture (containing 600,000 cells) was surgically transplanted into the right T4 mammary fat pad of 8 week old female NSG mice. After 3 weeks, tumors were palpable and treatment began as described above. Tumors were measured 2x/week and estimated tumor volume was calculated with the formula $V = (L*W^2)/2$.

In-vivo de-novo tumor growth experiments

Female MMTV-PyMT mice were monitored for palpable tumors, which typically appear at ages 8-10 weeks. At the beginning of the experiment, age-matched mice were split into two treatment groups. Mice were treated as described above three times a week. On each day of treatment, up to 4 tumors were measured per mouse (upper and lower, left and right). At the end of treatment the mice were euthanized. One tumor was flash

frozen in liquid nitrogen for RNA collection. The remaining tumors were collected and fixed in either 4% paraformaldehyde (PFA) or 10% neutral buffered formalin (NBF) for 72 hours. PFA fixed tumors were embedded in OCT and kept at -80 C until sectioning. NBF fixed tumors were processed, paraffin embedded, and sectioned by the Experimental Histopathology core at the Fred Hutchinson Cancer Center.

In-vivo tail-vein metastasis experiments

For tail-vein metastasis experiments, mTomato-MMTV-PyMT organoids were dissociated to single cells at day 0 using Accumax (20 minutes at 37°C). To generate clusters, single cells were plated in non-adherent dishes at 150,000 cells/mL in media +2% basement membrane-rich gel (v/v). After 24 hours, the clusters were resuspended in PBS at a concentration of 1 million cells/mL using cell counts from the day before. 200,000 cells in 200 uL of PBS were injected into Nod scid gamma (NSG) immunocompromised mice. For the de-novo study, mice were treated following the treatment protocol described above for 2 weeks, after which mice were euthanized and lungs imaged under a dissecting microscope for quantification of fluorescent (metastatic) area. Lungs were then fixed in 4% paraformaldehyde for 48 hours, then then transferred to 25% sucrose in DPBS overnight at 4°C before embedding in OCT and storing at -80°C.

2D histology

H&E and immunohistochemistry (IHC) stains were performed by the experimental histology core at Fred Hutchinson Cancer Center. Slides were scanned on a Ventana DP 200 slide scanner. Images were analyzed in qupath.

Lentiviral preparation

HEK293FT cells were purchased from ThermoFisher (R70007) and cultured in DMEM with 10% FBS, 1% penicillin/streptomycin, and 500 $\mu\text{g}/\text{mL}$ G418. Transfection was performed once cells reached 60-80% confluency. 1 hour before transfection, media was removed and changed to fresh G418-free media. Transfection was performed using Lipofectamine 3000 according to manufacturer protocol, using 3 μg MD2.G, 8 μg PsPax, and 12 μg of the shRNA containing plasmid. Media was again changed 16 hours after transfection. 48 hours after transfection, media was collected and filtered through 0.45 micron steri-flip filters, then aliquoted and stored at -80°C . Titers were calculated using lentivirus that had been through a single freeze-thaw cycle with Takara GoStix Plus according to protocol.

Lentiviral transduction for knockdown and expression

Organoids growing in suspension with 2% Matrigel were spun down and resuspended in organoid media supplemented with 8 $\mu\text{g}/\text{mL}$ protamine sulfate. Lentivirus was added to culture to achieve an MOI of 5-10 IFU/cell. 48 hours after transduction media was changed to fresh media containing selection agent if needed. Puromycin was used for selection at 1 $\mu\text{g}/\text{mL}$, blasticidin was used at 2 $\mu\text{g}/\text{mL}$, and gentimicin (G418) was used

at 400 µg/mL. Cells were grown in selection for at least 7 days before proceeding to experiments.

Epidemiological methods for Quilt Study

Quilt Study women who had participated in the PACE study were asked about supplement use at their in-person interview shortly after their breast cancer diagnosis (median time to interview 1.0 year). Questions (name/type, pills per week, duration) were asked relative to their diagnosis date, and separate sections gathered information about current use as of that time, and past use. Women reporting current use were asked to provide their bottles if available for verification of contents, while past use relied on recall. All supplements reported were reviewed by study team, and only those which definitely contained Vitamin A (including beta carotene, cod liver oil) were considered for the exposure variable. Women were categorized into three groups – those who took Vitamin A before their diagnosis, those who only took other supplements, and those who took no supplements; analysis was restricted to those who reported taking supplements. Due to the small sample size, we were unable to explore other aspects of use, such as duration. Women with distant-stage disease were excluded from analyses of recurrence, but not of mortality. Recurrence information was obtained from medical record reviews and from interviews of living women conducted as part of the Quilt study, and breast cancer mortality information was obtained via the cancer surveillance system (CSS) registry and medical record reviews. Women with no data available for relevant variables were excluded. Data analysis performed with Stata

Epidemiological methods for Pathways Study

Vitamin A and/or beta-carotene supplement use was self-reported by Pathways Study participants at study enrollment (baseline, on average about two months after diagnosis) and after 6- and 24-months of follow-up. At baseline, questions with categorical response options asked about a participant's use of the supplement before and since diagnosis, number of days per week taken, dose taken, and number of years taken. At follow-up, questions asked about number of days per week and dose taken in the previous six months (in 6-month follow-up questionnaire) or 12 months (in 24-month follow-up questionnaire). Women with AJCC stage IV breast cancer and/or with missing questionnaire data were excluded from the analyses. Analyses were also restricted to study participants who were dietary supplement users, as supplement users may differ in important ways from those who do not use supplements. Thus, women who reported they did not use any dietary supplements on the baseline, 6-, and 24-month questionnaires were excluded.

Use at diagnosis was grouped into four categories based on data reported from baseline and follow-up questionnaires: did not take vitamin A and/or beta-carotene, took before and after diagnosis, took before but stopped after diagnosis, took only after diagnosis. Among dietary supplement users, duration of use was grouped into three categories based on responses to number of years taken asked at baseline: did not take vitamin A and/or beta-carotene, took for >0-3 years, and took for >3 years. Recurrence and mortality data were ascertained via the KPNC Virtual Data Warehouse and electronic health record ¹⁸⁰. Recurrence was defined as a subsequent diagnosis of

breast cancer or death from breast cancer and breast cancer-specific mortality was defined as a death from breast cancer.

Distributions and means with standard deviations for the demographic and clinical characteristics of the Pathways Study cohort were calculated for exposure groups based on vitamin A and/or beta-carotene timing of use and duration of use (**Supplemental Table 4**). Kaplan Meier curves estimating the probability of recurrence or mortality were generated for the timing and duration of use exposure groups. Unadjusted and adjusted Cox proportional hazards models were used to estimate time to recurrence and breast cancer-specific mortality. Adjusted models included AJCC stage, age at diagnosis, receipt of chemotherapy (yes/no), receipt of radiation therapy (yes/no), and receipt of endocrine therapy (yes/no) as covariates. Statistical significance was based on an $\alpha=0.05$. Analyses were conducted using SAS v9.4 (SAS Institute, Cary, NC).

Similar statistical analyses were performed for the Quilt Study cohort with the following differences: Quilt cohort data included only ever/never use before diagnosis, SEER Summary Stage instead of AJCC was used; diagnosis year was included in models; and analyses were performed in Stata/SE v18. Unadjusted and adjusted hazard ratios (HR) and 95% confidence intervals (CI) were calculated using Cox proportional hazards methods, to estimate time to recurrence and breast cancer-specific mortality.

Table of primers, antibodies, plasmids, etc.

Resource	Source	Identifier
Antibodies		
Keratin 14	Biologend	906004
Keratin 17	Cell Signaling	4543S
Hair Keratin (AE13)	Santa Cruz	sc-57012
Loricrin	Abcam	ab85679
pHistone H3	Millipore	06-570
Cleaved caspase 3	Cell Signaling	9579
Phalloidin 568	Thermo Fisher	A12380
Recombinant DNA		
MD2.G	Gift from Didier Trono	Addgene #12259
PsPax2	Gift from Didier Trono	Addgene #12260
shKlf4 puro	Transomic	Available upon request
Tet-on Klf4 ORF	Vectorbuilder	Available upon request
EF1a Tet On (ETO)	Gift from Emily Hatch	Addgene #84776

Code and Data availability statement:

Raw data and code will be made publicly available upon publication. To recreate analyses, see github for R, cellprofiler, and qupath scripts. The following data will be deposited in the NCBI Gene Expression Omnibus (GEO): Fastq and processed Seurat object for scRNA-seq of invading MMTV-PyMT organoids, Fastq and processed counts

for RNAseq of MMTV-PyMT organoids. All other data is provided as supplementary files with this publication.

Additional scRNA-seq datasets used are available at the Gene Expression Omnibus accession numbers listed below:

- FVB mammary gland ¹⁸¹ : GSE103275
- C3(1)tag organoids ¹¹⁰: GSE149299
- MMTV-PyMT tumor ¹¹⁴: GSE214815
- C3(1)tag tumor ¹¹⁹: GSE182389
- Pten-Myc tumor ¹¹⁸: GSE215070
- Human BRCA dataset #1 ¹²¹: GSE161529
- Human BRCA dataset #2 ¹²²: GSE176078

QUILT study data has been provided as supplementary tables, but additional information is available upon request. For the Pathways Study, procedures for obtaining data, including the analytic dataset and associated SAS programs, are available upon request.

Ethics Declaration

A.D., S.S., and K.C. are inventors on a pending patent application related to this work.

Local researchers at our primary institution Fred Hutch and collaborators at Kaiser Permanente have been fully involved in this project. The clinical datasets in this study

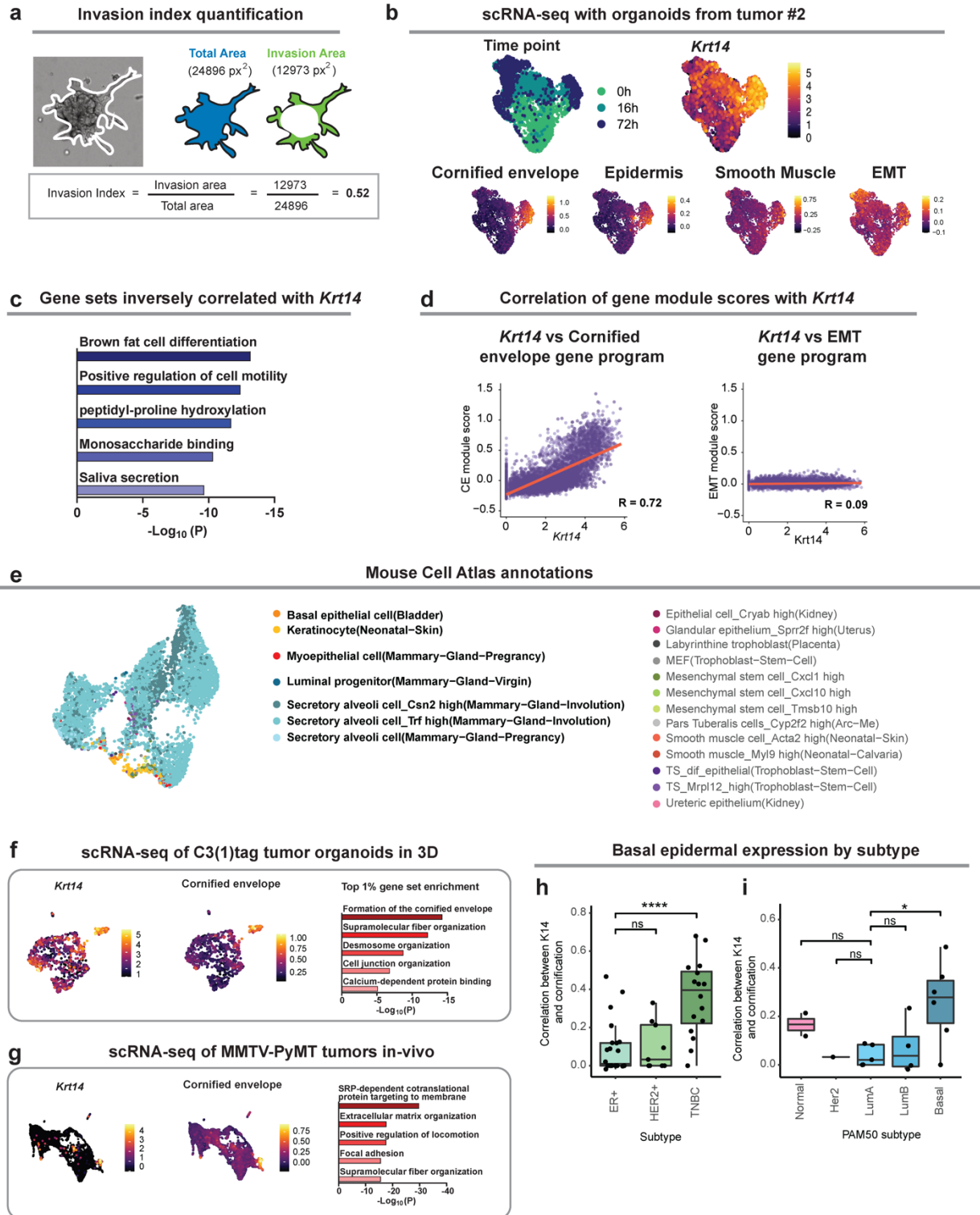
apply to women in the Pacific Northwest and Northern California at our respective institutions, making them highly locally relevant.

Author Contribution Statement

A.D., K.J., quilt/pathways authors conceived the study and designed experiments. A.D., S.S., N.R., and J.P, assisted in performing in-vivo experiments. B.K., A.D., and K.C., performed analysis of RNAseq data. A.D. performed scRNAseq and computational image analyses. A.D., A.R., J.H., and Y.K. performed cell culture and molecular biology experiments. R.A. consulted on pathological analysis of tissues. M.S. consulted on proper analysis of scRNAseq data. K.M. and D.D. conceived study design, analysis, and interpretation of QUILT analysis. E.R., L.D., I.E., J.R., L.K., and H.G. conceived study design, analysis, and interpretation of PATHWAYS analysis.

EXTENDED DATA

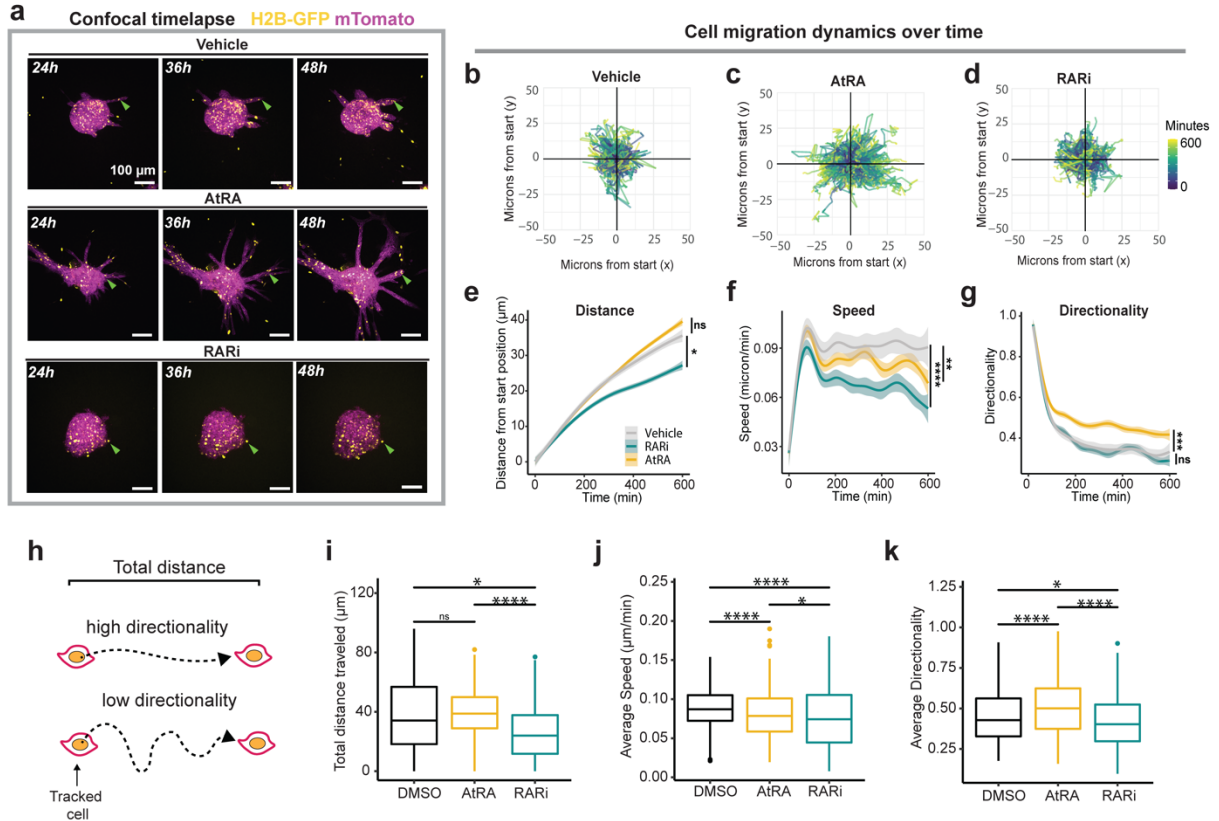
Extended data 1



Extended data 1: Epidermal identity of basal breast cancer cells

- a) Schematic depicting how invasion index is calculated from each DIC image. The invasion area is the total area minus the maximum convex area filling the inside of the organoid. The invasion index is the invasion area divided by the total area.
- b) UMAP of scRNA-seq data obtained from organoids from replicate 2, colored by time, *Krt14* expression, and gene module scores calculated in Seurat using gene sets from **supplementary table 2**.
- c) Top 5 gene sets enriched for in the bottom 1% of genes correlating with *Krt14*, sorted by p-value. Calculated with Metascape.
- d) Example correlation plots of two gene module scores with *Krt14* expression in MMTV-PyMT organoid cells from both tumors: formation of the cornified envelope (CE) , and EMT.
- e) Cell type assignment of individual cells from invading MMTV-PyMT organoids using unsupervised annotation and the Mouse Cell Atlas ¹¹⁵.
- f) UMAPs and GSEA of top 1% of genes correlating with *Krt14* in mouse C3(1)tag organoids embedded in 3D collagen I ¹¹⁰ . Cornified envelope and smooth muscle scores calculated using gene sets in supplementary table 1.
- g) UMAPs and GSEA using scRNA-seq data obtained from MMTV-PyMT de-novo tumors ¹¹⁴.
- h) Pearson R of correlation between cornified envelope score and *Krt14* stratified by breast cancer receptor subtype. P-value calculated by one sample t-test. **** indicates $p \leq 0.0001$.
- i) Pearson R of correlation between cornified envelope score and *Krt14* stratified by breast cancer PAM50 subtype. P-value calculated by one sample t-test. * indicates $p \leq 0.05$.

Extended data 2



Extended data 2: Migration dynamics

a) Representative confocal microscopy time-lapse images of mTmG MMTV-PyMT organoids invading in 3D collagen. Organoids were first sparsely transduced with lentivirus to express H2B-GFP (yellow) while mTomato membranes are shown in magenta.

b-d) Migration tracks of individual cells tracked in confocal time-lapses represented in a. P-values calculated with Wilcoxon test.

e) Cumulative distance from start point of cells tracked in confocal time-lapse. TP-values calculated with Wilcoxon test.

f) Speed of cells over time during migration, measured in microns/min. P-values calculated with Wilcoxon test.

g) Directionality of cells over time during migration, representing the propensity of cells to continue in the same direction as the frame before (Ext data 3a). P-values calculated with Wilcoxon test.

h) Schematic depicting cell migration tracks with high directionality and low directionality.

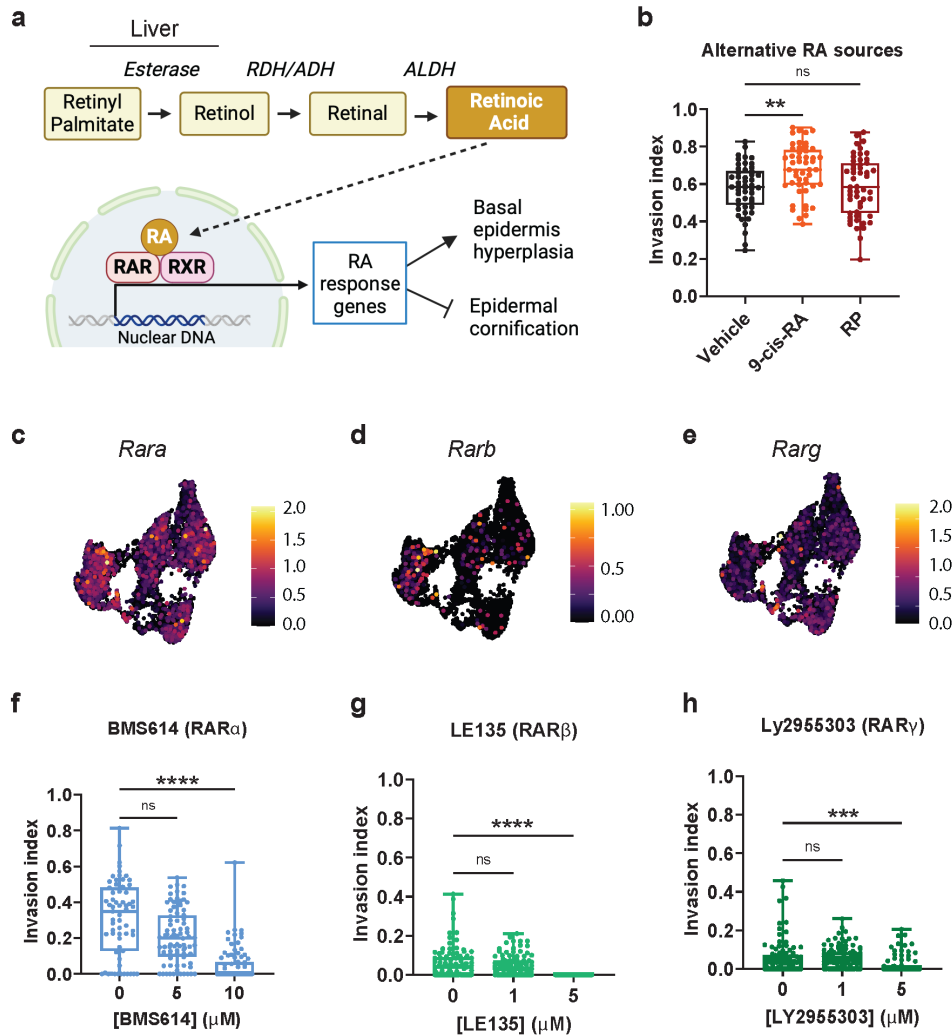
i) Average directionality of all cells across all timepoints. P-values calculated with Wilcoxon test.

j) Total distance traveled, in microns, of each cell at endpoint. P-values calculated with Wilcoxon test.

k) Average speed (microns/minute) of cells across all timepoints. P-values calculated with Wilcoxon test.

note: For all figures with p-values indicated by asterisks, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. ns indicates non-significant comparisons. For all figures $n = 23$ organoids, 352 cells. Boxplots are plotted in the style of Tukey.

Extended data 3



Extended data 3: Retinoic acid signaling pathway

a) Schematic of retinoic acid signaling pathway. Retinyl palmitate is often converted to retinol in the liver, then is further processed within cells of various tissues. Made with Biorender.

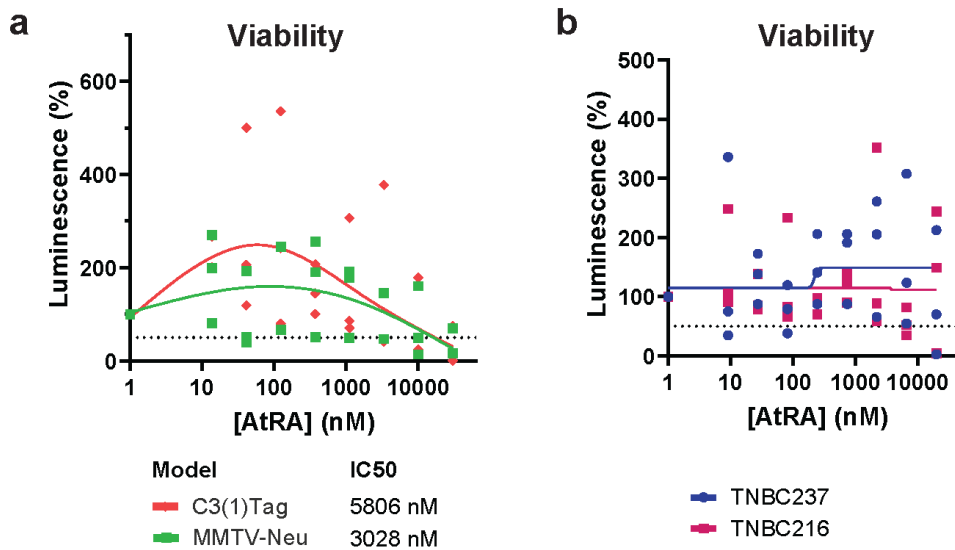
b) Invasion indices for MMTV-PyMT organoids treated with 9-cis retinoic acid (9-cis-RA) or retinyl palmitate (RP) both used at a concentration of 1 μM . $n = 3$ biological replicates, ≥ 90 organoids per dose. P-values calculated using Dunn's multiple comparison test after Kruskal-Wallis test.

c-e) Umaps showing transcription of *Rara* (c), *Rarb* (d), and *Rarg* (e) in scRNA-seq of invading MMTV-PyMT organoids.

f-h) Invasion indices for MMTV-PyMT organoids treated with RAR α (f), RAR β (g), and RAR γ (h) inhibitors across a dose range. n = 3 biological replicates, ≥ 90 organoids per dose. P-values calculated using Dunn's multiple comparison test after Kruskal-Wallis test.

note: For all figures with p-values indicated by asterisks, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. ns indicates non-significant comparisons. Boxplots represents median, 25th percentile and 75th percentile, whiskers extend from min to max.

Extended data 4

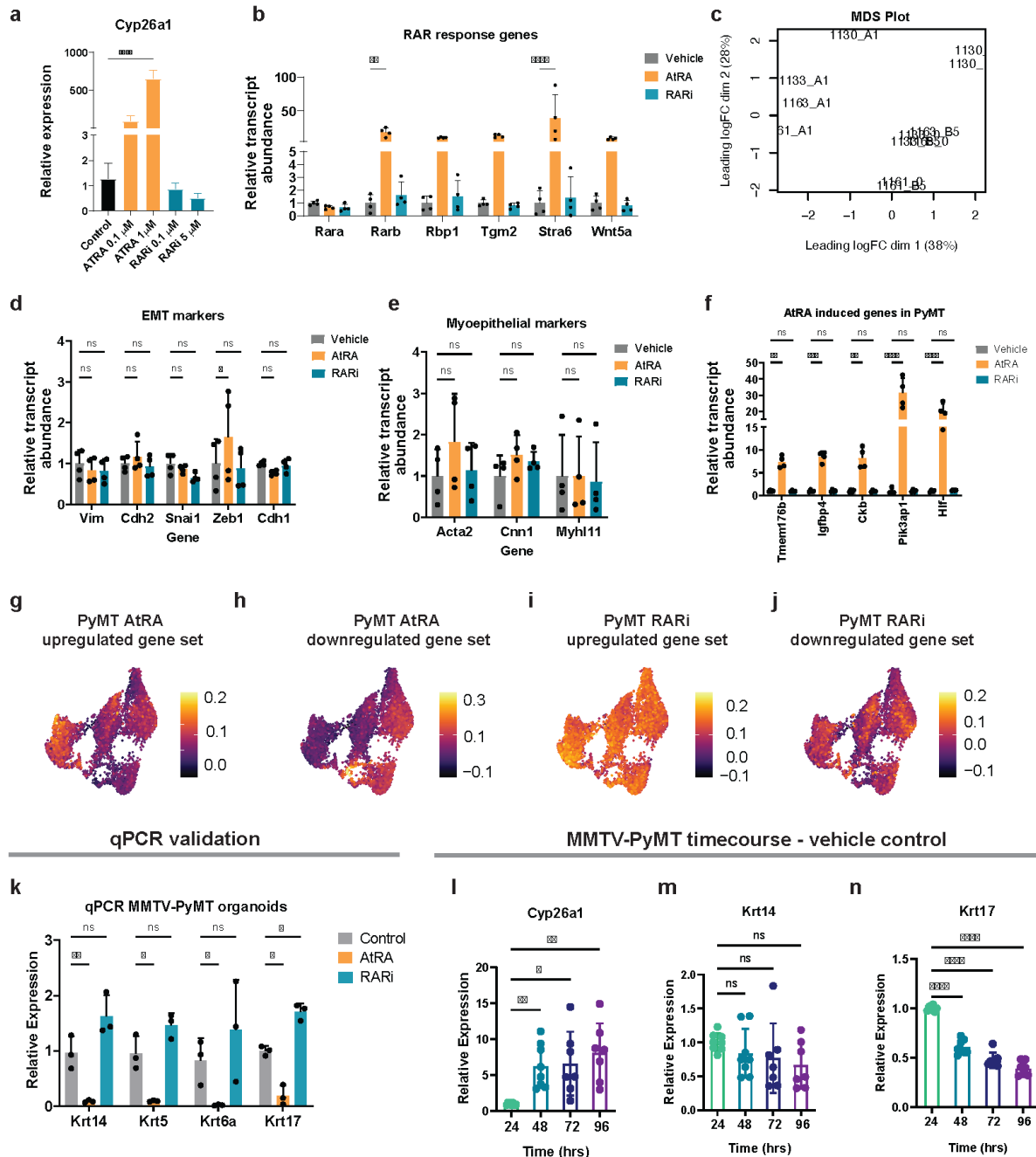


Extended data 4: Viability curves with AtRA in mouse and human organoids.

a) Relative CellTiter-Glo luminescence of C3(1)/Tag and MMTV-Neu organoids treated with a dose range of AtRA. Curve was drawn with a spline equation. n = 3 technical replicates.

b) Relative CellTiter-Glo luminescence of TNBC216 and TNBC237 organoids treated with a dose range of AtRA. Curve was drawn with a nonlinear inhibition equation in prism. n = 3 technical replicates.

Extended data 5



Extended data 5: Transcriptional changes induced by AtRA and RARi

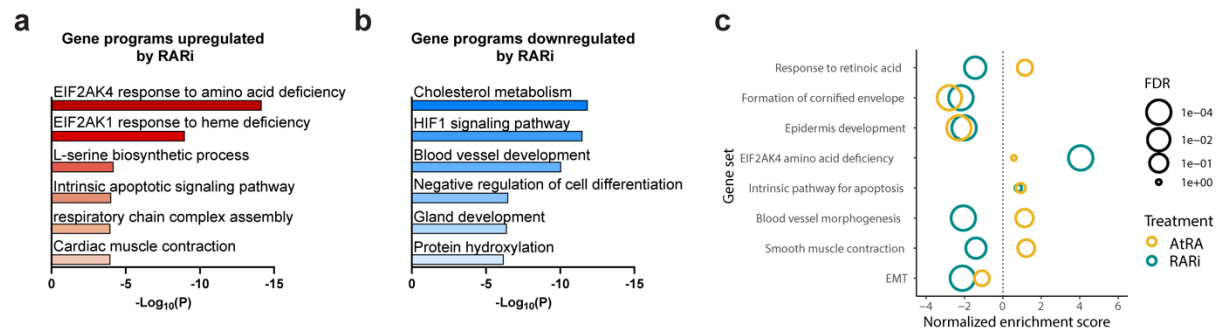
a) Relative abundance of Cyp26a1 transcripts in MMTV-PyMT organoids treated with a dose range of AtRA or RARi (BMS493) measured with qPCR. P-values calculated with Dunnett's multiple comparison test after one-way ANOVA. n = 3 biological replicates.

b) Relative abundance of retinoic response gene transcripts in MMTV-PyMT organoids treated with 1 μ M AtRA or 5 μ M RARi measured with bulk RNAseq. P-values calculated with Dunnett's multiple comparison test after two-way ANOVA. n = 4 biological replicates.

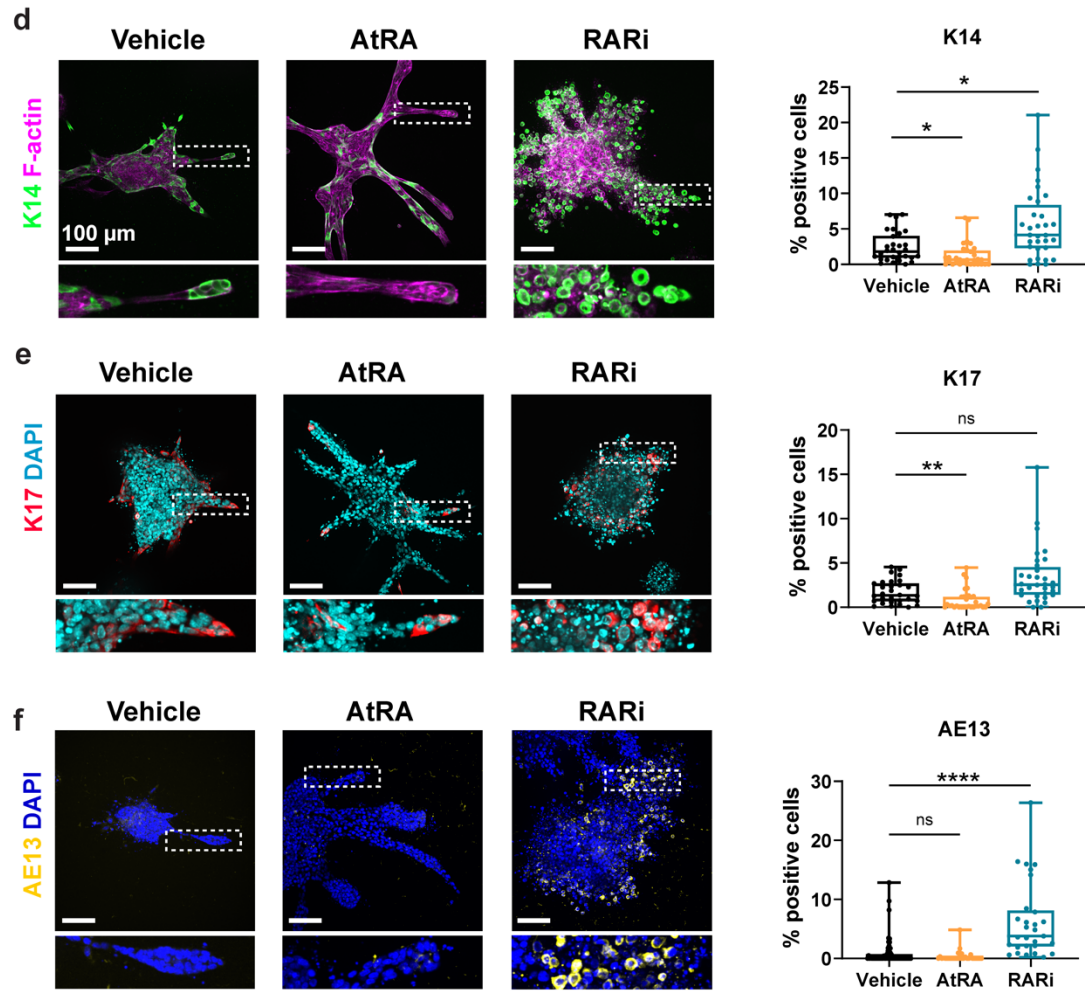
- c)** PCA plot of RNAseq samples, indicating that AtRA has a larger transcriptional effect than RARi. Mouse number is followed by either “A1” or “B5” to represent AtRA 1 μ M and BMS493 (RARi) 5 μ M, respectively.
- d)** Relative abundance of EMT marker gene transcripts in MMTV-PyMT organoids treated with 1 μ M AtRA or 5 μ M RARi measured with bulk RNAseq. P-values calculated with Dunnett’s multiple comparison test after two-way ANOVA. n = 4 biological replicates.
- e)** Relative abundance of myoepithelial marker gene transcripts in MMTV-PyMT organoids treated with 1 μ M AtRA or 5 μ M RARi measured with bulk RNAseq. P-values calculated with Dunnett’s multiple comparison test after two-way ANOVA. n = 4 biological replicates.
- f)** Selected genes with the highest transcriptional induction following AtRA treatment in PyMT organoids. Measured with bulk RNAseq. P-values calculated with Dunnett’s multiple comparison test after two-way ANOVA. n = 4 biological replicates.
- g-h)** Gene sets were created using the top 1% of up- and down-regulated transcripts in MMTV-PyMT organoids treated with AtRA. These gene sets were then used to calculate gene module scores for single cells in scRNA-seq of invading organoids. Gene module scores are depicted in UMAPs.
- i-j)** Gene sets were created using the top 1% of up- and down-regulated transcripts in MMTV-PyMT organoids treated with RARi. These gene sets were then used to calculate gene module scores for single cells in scRNA-seq of invading organoids. Gene module scores are depicted in UMAPs.
- k)** qPCR confirmation of changes in keratins after treatment with 1 μ M AtRA or 5 μ M RARi. P-values calculated with Dunnett’s multiple comparison test after two-way ANOVA. n = 3 biological replicates per condition.
- m)** Relative abundance of Krt14, Krt17, and Cyp26a1 transcripts over time in MMTV-PyMT organoids embedded in 3D collagen, measured with qPCR. P-values calculated with Dunnett’s T3 multiple comparison test after Welch’s ANOVA. n = 3 biological replicates per condition with two technical replicates.
- note:** For all figures with p-values indicated by asterisks, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. ns indicates non-significant comparisons. Bars represent mean of replicates (points) with SD.

Extended data 6

Bulk RNAseq 3D MMTV-PyMT organoids



Immunofluorescent detection of epidermal markers



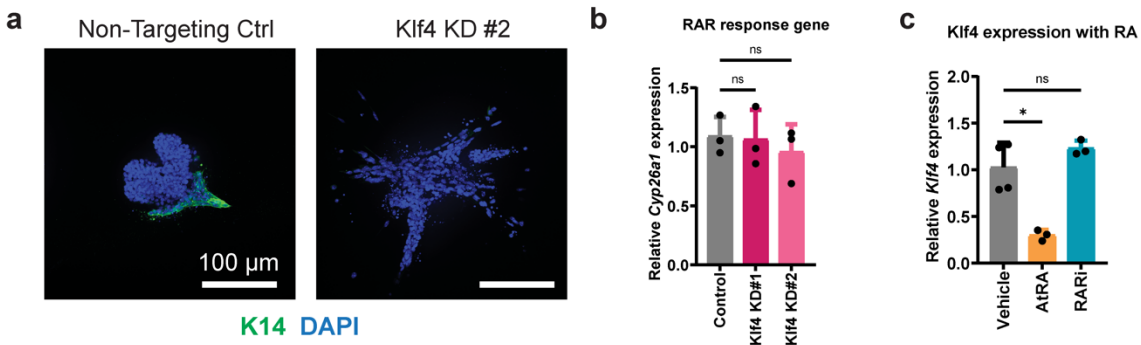
Extended data 6 – Epidermal gene expression changes induced by AtRA and RARi.

a-b) Top gene programs upregulated (a) and downregulated (b) by 5 μ M RARi exposure for 72 hours in MMTV-PyMT organoids embedded in 3D collagen. Bulk RNAseq data gene set enrichment analysis performed with metascape against human gene programs.

c) Enrichment scores of specific gene sets in MMTV-PyMT organoids treated with either AtRA or RARi for 72 hours in 3D collagen. Normalized enrichment scores determined by ranking transcripts by t-value and using pre-ranked gene set enrichment analysis.

d-f) Representative confocal images of immunofluorescent detection of Keratin 14 (d), Keratin 17 (e), AE13 hair keratin with quantification (f). All scale bars represent 100 microns. Plots indicate percent of positive cells on a per organoid basis from IF stains. n = 3 biological replicates, ≥ 30 organoids per condition, $\geq 60,000$ total cells. P-values calculated using Dunn's multiple comparison test after Kruskal-Wallis test.

Extended data 7



Extended data 7: Klf4 signaling

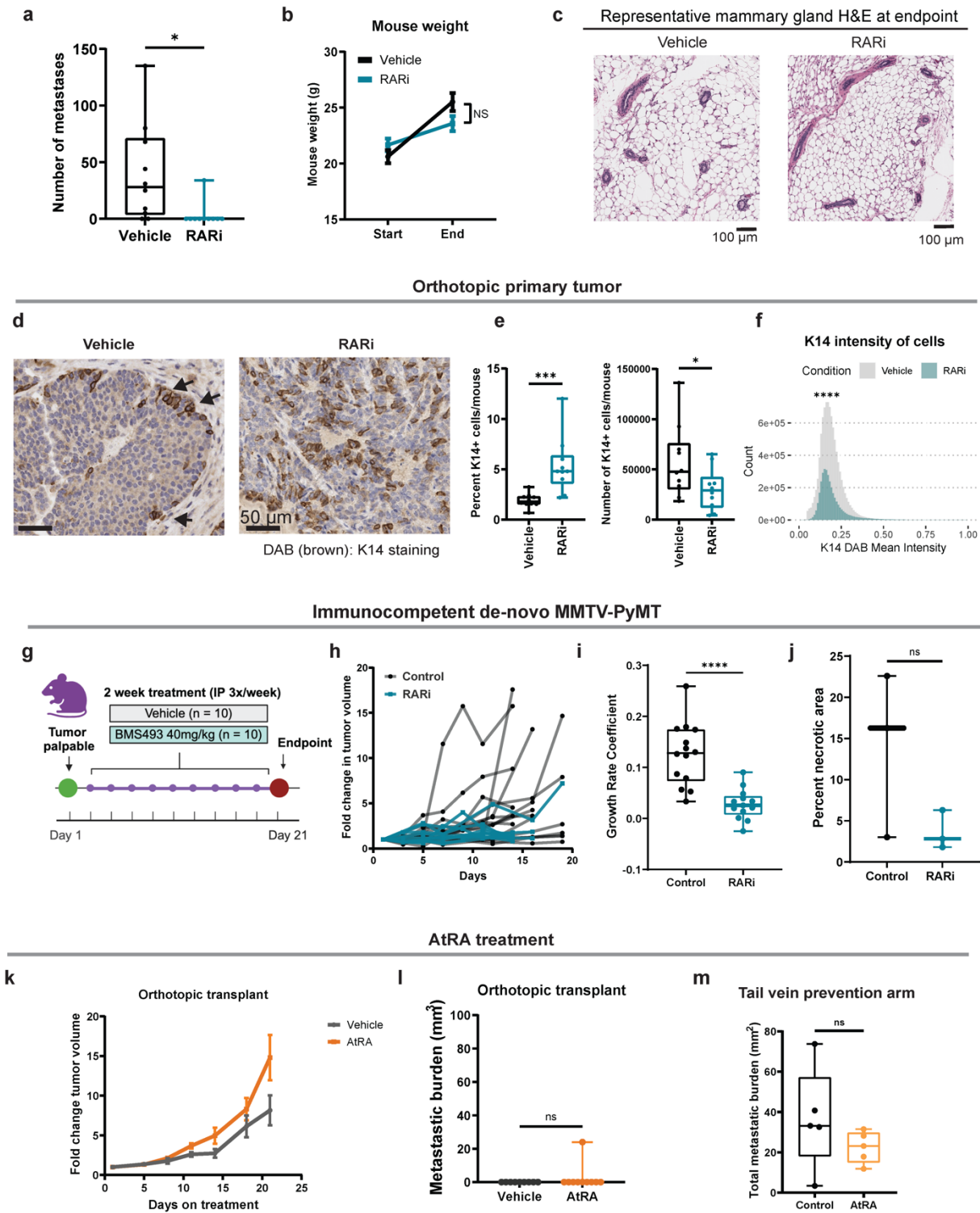
a) Representative images of K14 expression in PyMT organoids via immunofluorescence (green). Scale = 100 microns.

b) Relative expression of the RAR response transcript Cyp26a1 in Klf4 knockdown organoids. Detected with qPCR. n = 3 biological replicates. P-values calculated using Dunn's multiple comparison test after Kruskal-Wallis test.

c) Relative expression of Klf4 in PyMT organoids treated with AtRA and RARi. Detected with qPCR. n = 3 biological replicates. P-values calculated using Dunn's multiple comparison test after Kruskal-Wallis test.

note: For all figures with p-values indicated by asterisks, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. ns indicates non-significant comparisons. For all figures n = 23 organoids, 352 cells. Boxplots are plotted in the style of Tukey.

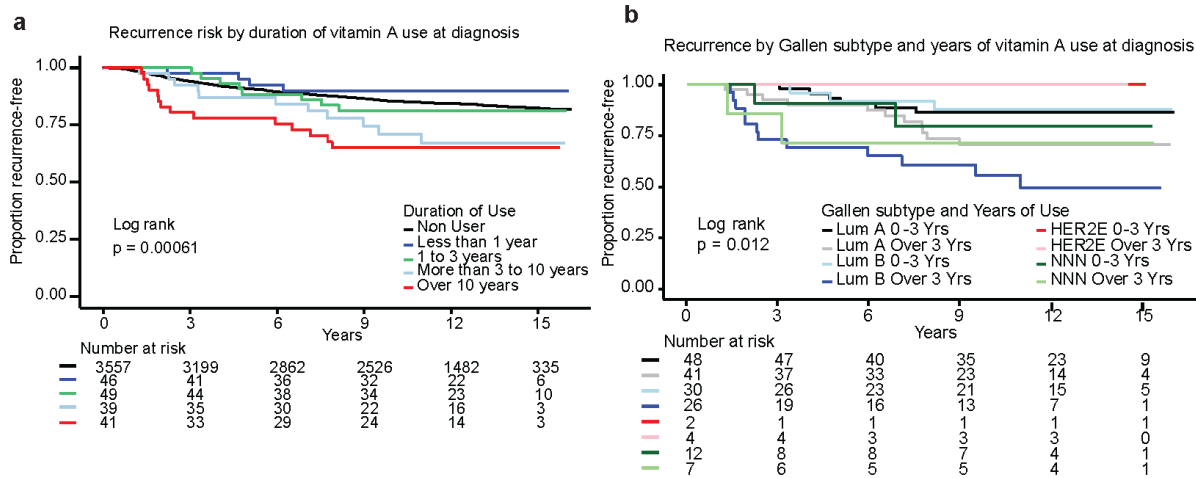
Extended data 8



Extended data 8: RARi and AtRA in-vivo

- a)** Number of macro-metastases in each mouse. $n = 10$ mice per treatment. P-values calculated with Mann-Whitney test.
 - b)** Mouse start and end weight in orthotopic tumor study, $n = 10$ mice per treatment. P-values calculated with Mann-Whitney test.
 - c)** Representative H&E images of mammary glands of mice treated with either vehicle or BMS493. No differences were apparent.
 - d)** Representative images of keratin 14 in tumors, stained with IHC. Scale bar represents 50 microns.
 - e)** Quantification of percent of cells scored as K14-positive (K14+) in tumor sections as calculated with QuPath. P-values calculated with Mann-Whitney test. $n = 10$ mice per condition, each with one tumor section stained.
 - f)** Violin plot of cellular K14 DAB intensity in tumors treated with either control or RARi. P-values calculated using Wilcoxon test. Medians values are displayed. $n = 10$ mice per condition, each with one tumor section stained.
 - g)** Schematic depicting timeline for treatment of de-novo MMTV-PyMT tumors in immunocompetent mice. Made with Biorender.
 - h)** Tumor measurements over time in MMTV-PyMT mice treated with either vehicle or BMS493. $n = 3$ mice per treatment, at least 5 tumors per treatment.
 - i)** Growth rate coefficient of de-novo MMTV-PyMT tumors treated with either vehicle or BMS493. Coefficients derived from fitted model of logarithmic tumor growth. P-values calculated with Mann-Whitney test. $n \geq 9$ tumors per treatment, ≥ 3 mice per treatment.
 - j)** Percent of total tumor area determined to be necrotic in de-novo MMTV-PyMT tumors after treatment with vehicle or BMS493. P-values calculated with Mann-Whitney test. $n = 3$ mice per treatment, one section each.
 - k)** Fold change in tumor growth over time in mice treated with either vehicle or AtRA. $n = 10$ mice per treatment. P-values calculated using Mann-Whitney Test. Duration of treatment was 3 weeks.
 - l)** Total macro-metastatic area visible on the outside of the lungs of mice by fluorescent detection of tumor cells at orthotopic experiment endpoint. $n = 10$ mice per treatment. P-values calculated using Mann-Whitney Test.
 - m)** Total macro-metastatic area visible on the outside of the lungs of mice by fluorescent detection of tumor cells at end of prevention tail-vein experiment. $n = 5$ mice per treatment. P-values calculated using Mann-Whitney Test.
- note:** For all figures with p-values indicated by asterisks, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$. ns indicates non-significant comparisons. Boxplots represents median, 25th percentile and 75th percentile, whiskers extend from min to max.

Extended data 9



Extended data 9: Consumption of vitamin A supplements and risk in breast cancer patients.

- a) Unadjusted Kaplan-Meier curves for time to breast cancer recurrence stratified years of vitamin A and/or beta-carotene supplement use across 5 categories in the Pathways Study.
- b) Unadjusted Kaplan-Meier curves for time to breast cancer recurrence stratified by St. Gallen subtypes and duration (0-3 years, 3 or more years) of vitamin A and/or beta-carotene supplement use prior to breast cancer diagnosis among Pathways Study participants.

SUPPLEMENTARY TABLES

Supplementary Table 1 : Gene correlation with Krt14 in scSeq

Gene	Correlation with Krt14	Gene	Correlation with Krt14	Gene	Correlation with Krt14	Gene	Correlation with Krt14
Krt14	1.0000	Ctsf	0.3415	Macf1	0.2880	Junb	0.2549
Sfn	0.5518	Moxd1	0.3397	Id4	0.2878	Ptms	0.2548
Serpib5	0.5255	Fermt1	0.3395	S100a11	0.2867	D16Ert472e	0.2545
Perp	0.5212	Cdh3	0.3390	Cadm1	0.2866	Rbm3	0.2523
Dsp	0.5004	Gas1	0.3378	Apoe	0.2863	Myl9	0.2519
Rflnb	0.4723	Rgs12	0.3376	Fat2	0.2861	Serpine2	0.2517
Krt17	0.4664	Marcks	0.3360	Apbb2	0.2855	Gpc6	0.2512
Trim29	0.4589	Hspa4l	0.3343	Fryl	0.2855	Aplp2	0.2512
Ctsc	0.4576	Ccn2	0.3338	Robo2	0.2849	Nebi	0.2504
Lmo1	0.4494	Slc27a3	0.3310	Slc25a5	0.2834	Ceacam1	-0.2511
Lama3	0.4404	Atp1b3	0.3296	Nfkbia	0.2825	Fdx1	-0.2553
Sdc1	0.4367	Samd5	0.3291	Phlda3	0.2810	Txnip	-0.2556
Palld	0.4334	Jag2	0.3277	Reep3	0.2801	Cox17	-0.2577
Hras	0.4299	Lima1	0.3273	Mt2	0.2795	Bik	-0.2587
Cbr2	0.4297	Postn	0.3270	Sfr1	0.2776	Egln3	-0.2589
Lgals7	0.4286	Adgrl3	0.3269	Plch2	0.2772	Rpl38	-0.2603
Col17a1	0.4252	Pkp1	0.3268	Svil	0.2769	Thsd4	-0.2607
Emb	0.4243	Cpne8	0.3262	Ucp2	0.2766	Smagp	-0.2612
Trp63	0.4242	Dsg3	0.3255	Vav3	0.2759	Ndr1	-0.2612
Ak1	0.4205	Pkp3	0.3239	Cblb	0.2749	Aqp5	-0.2621
Krt5	0.4193	Cav1	0.3234	Gjb3	0.2743	Snhg20	-0.2675
Sparc	0.4155	Bmp1	0.3233	Tpm2	0.2742	Sorbs2	-0.2676
Cst3	0.4145	Ifitm2	0.3227	Col18a1	0.2722	Itpk1	-0.2691
Jag1	0.4116	Ly6d	0.3226	Hs6st1	0.2715	Pnpla2	-0.2694
Pltp	0.4076	Bok	0.3208	Degs1	0.2710	Spint2	-0.2718
Sorcs2	0.4053	Ppp1r14c	0.3190	Adam12	0.2707	Elf5	-0.2723
Krt6a	0.4044	Jun	0.3177	Sostdc1	0.2702	Scd1	-0.2754
Pard6g	0.3905	Ephb6	0.3174	Cers6	0.2690	Fbxo36	-0.2771
Col7a1	0.3898	Tgfb1	0.3162	Plek2	0.2687	Nucb2	-0.2810
Htra1	0.3895	Steap4	0.3159	Rbp1	0.2680	Chchd10	-0.2828
Fst	0.3865	Tenm2	0.3157	Rbbp8	0.2671	Eif4ebp1	-0.2854
Anxa8	0.3865	Runx1	0.3122	S100a14	0.2661	Bin1	-0.2904
Aopep	0.3825	Tsc22d1	0.3117	Kif26b	0.2659	Cd68	-0.2910

Txndc17	0.3816	Wnt10a	0.3109	Mgst2	0.2658	Abhd5	-0.2916
Nxn	0.3800	Bcam	0.3104	Abrac1	0.2648	Lgals3	-0.2923
Pmepa1	0.3785	Angptl4	0.3074	Sgk1	0.2636	Bsg	-0.2951
Car12	0.3768	Wnt4	0.3072	Gprin3	0.2625	Gde1	-0.2985
Rbms3	0.3716	H3f3a	0.3071	Ctsd	0.2612	Aldoa	-0.2992
Pla2g7	0.3662	Kctd1	0.3065	Aqp3	0.2611	Proser2	-0.3006
Klf4	0.3654	Ntng1	0.3064	Bnc1	0.2610	Ntn1	-0.3012
Krt16	0.3625	Fxyd3	0.3027	Jup	0.2609	Litaf	-0.3079
Arl15	0.3622	Fos	0.3022	Marcksl1	0.2609	Syng1	-0.3080
Man1a	0.3618	Cntfr	0.3022	Hes1	0.2604	Hebp1	-0.3090
Slc6a6	0.3589	Pkp4	0.3005	Gsn	0.2601	Errf1	-0.3151
Ext1	0.3581	Ifitm3	0.3004	Igfbp4	0.2599	Ero1l	-0.3187
Bcl11b	0.3552	Col12a1	0.2982	Gjb4	0.2592	Wfdc18	-0.3214
Lmna	0.3536	Tiparp	0.2974	Schip1	0.2588	Plet1	-0.3253
Cald1	0.3529	Vsnl1	0.2950	Dynl1	0.2573	Cldn7	-0.3305
Prss12	0.3523	Arl4c	0.2934	Rnase4	0.2572	Csn3	-0.3553
Icam1	0.3477	Krt79	0.2924	Cpe	0.2572	Lcn2	-0.3582
2610035D17Rik	0.3472	Acpp	0.2910	Vwa1	0.2571	Krt18	-0.3909
Mir205hg	0.3469	Pxdc1	0.2910	Igfbp7	0.2570	Cystm1	-0.4615
Arhgap24	0.3451	Prss23	0.2908	Plec	0.2562	Cldn3	-0.5164
Cav2	0.3449	Vcan	0.2907	Clic4	0.2561		
Snai2	0.3422	Tmsb4x	0.2906	Arhgap42	0.2551		
		Limk2	0.2900	Plxdc2	0.2549		

Supplementary table 2: gene lists used for gene module scores

formation_of_cornified_envelope	epidermis_markers	smooth_muscle_contraction	EMT
R-MMU-6809371_Metascap	Pagenbase_mouse_epidermis_markers	R-MMU-445355_smooth_muscle_contraction	EMT_GO_term_summary_20230910_125021

Supplementary table 3: Characteristics by Vitamin A use among dietary supplement users only (n=705)

Demographics	Vitamin A Use		
	No Use	Used before Dx	
N (%)	650 (92.2)	55 (7.8)	
Age at diagnosis, years, Mean (SD)	71.9 (4.0)	72.5 (4.0)	p=0.25
Race and Ethnicity, n (%)			p=0.38
Non-Hispanic White	623 (95.8)	50 (90.9)	
Non-Hispanic Black	8 (1.2)	2 (3.6)	
Asian American	11 (1.7)	2 (3.6)	
Hispanic	3 (0.5)	1 (1.8)	
Pacific Islander	0	0	
American Indian / Alaska Native	3 (0.5)	0	
Unknown	2 (0.3)	0	
Education level, n (%)			p=0.34
High school or less	328 (50.5)	23 (41.8)	
Some college	200 (30.8)	22 (40.0)	
College graduate or higher	122 (18.8)	10 (18.2)	
Clinical			
Stage at diagnosis, n (%)			p=0.40
Local	515 (79.2)	42 (76.4)	
Regional	121 (18.6)	13 (23.6)	
Distant	13 (2.0)	0	
Unknown	1 (0.2)	0	
ER status, n (%)			p < 0.01
ER+	583 (89.7)	41 (74.5)	
ER-	57 (8.8)	14 (25.5)	
Unknown	10 (1.5)	0	
Had chemotherapy, n (%)	113 (17.4)	15 (27.3)	p=0.07
Had radiation therapy, n (%)	424 (65.2)	37 (67.3)	p=0.77
Had endocrine therapy, n (%)	451 (69.4)	33 (60.0)	p=0.14

Supplementary table 4.1: Characteristics by Start-Stop behavior of Vitamin A use among dietary supplement users only (n=4020)

Demographics	Vitamin A Use at Diagnosis			
	No Use	Before but stopped after diagnosis	Before and after diagnosis	After diagnosis only
N (%)	3698 (92.1)	115 (2.9)	64 (1.6)	140 (3.5)
Age at diagnosis, years, Mean (SD)	59.8 (12.0)	62.8 (10.8)	62.6 (10.9)	62.8 (12.0)
Race and Ethnicity, n (%)				
Non-Hispanic White	2466 (66.7)	82 (71.3)	47 (73.4)	103 (73.6)
Non-Hispanic Black	284 (7.7)	4 (3.5)	6 (9.4)	3 (2.1)
Asian American	475 (12.8)	10 (8.7)	5 (7.8)	18 (12.9)
Hispanic	391 (10.6)	17 (14.8)	5 (7.8)	9 (6.4)
Pacific Islander	9 (0.2)	0	0	0
American Indian / Alaska Native	73 (2.0)	2 (1.7)	1 (1.6)	7 (5.0)
Education level, n (%)				
High school or less	560 (15.2)	12 (10.4)	5 (7.8)	23 (16.4)
Some college	1285 (34.8)	45 (39.1)	24 (37.5)	45 (32.1)
College graduate or higher	1852 (50.1)	58 (50.4)	35 (54.7)	72 (51.4)
Clinical				
Stage at diagnosis, n (%)				
I	2037 (55.1)	70 (60.9)	34 (53.1)	83 (59.3)
II	1292 (34.9)	37 (32.2)	22 (34.4)	44 (31.4)
III	369 (10.0)	8 (7.0)	8 (12.5)	13 (9.3)
ER status, n (%)				
ER+	3097 (83.8)	95 (82.6)	55 (85.9)	122 (87.1)
ER-	599 (16.2)	20 (17.4)	9 (14.1)	18 (12.9)
Subtype IHC, n (%)				
Luminal A	2659 (75.0)	82 (73.9)	51 (83.6)	102 (75.0)
Luminal B	311 (8.8)	10 (9.0)	2 (3.3)	16 (11.8)
NNN	413 (11.6)	16 (14.4)	5 (8.2)	13 (9.6)
HER-2 OVER	164 (4.6)	3 (2.7)	3 (4.9)	5 (3.7)
Subtype Gallen, n (%)				
Luminal A	1845 (51.7)	61 (54.5)	29 (47.5)	60 (43.8)
Luminal B	1144 (32.1)	32 (28.6)	24 (39.3)	59 (43.1)
HER-2E	164 (4.6)	3 (2.7)	3 (4.9)	5 (3.7)
NNN	413 (11.6)	16 (14.3)	5 (8.2)	13 (9.5)
Had chemotherapy, n (%)	1334 (36.2)	34 (30.1)	15 (23.4)	49 (35.3)
Had radiation therapy, n (%)	1636 (44.2)	64 (55.7)	27 (42.2)	62 (44.3)
Had endocrine therapy, n (%)	2146 (58.4)	52 (45.2)	24 (38.1)	76 (54.7)

Supplementary table 4.2: Characteristics by duration of Vitamin A use at baseline among dietary supplement users only (n=3745)

	Duration of Vitamin A and/or Beta-Carotene Use at Baseline		
	No Use	>0-3 years	>3 years
Demographics			
N (%)	3557 (95.3)	95 (2.6)	80 (2.1)
Age at diagnosis, years, Mean (SD)	60.2 (12.0)	60.7 (10.6)	63.7 (11.0)
Race and Ethnicity, n (%)			
Non-Hispanic White	3414 (67.9)	60 (63.2)	63 (78.8)
Non-Hispanic Black	257 (7.2)	8 (8.4)	2 (2.5)
Asian American	446 (12.5)	12 (12.6)	5 (6.3)
Hispanic	363 (10.2)	12 (12.6)	9 (11.3)
Pacific Islander	7 (0.2)	0	0
American Indian / Alaska Native	70 (2.0)	3 (3.2)	1 (1.3)
Education level, n (%)			
High school or less	530 (14.9)	8 (8.4)	10 (12.5)
Some college	1243 (35.0)	44 (46.3)	23 (28.8)
College graduate or higher	1783 (50.1)	43 (45.3)	47 (58.8)
Clinical			
Stage at diagnosis, n (%)			
I	1984 (55.8)	60 (63.2)	42 (52.5)
II	1221 (34.3)	27 (28.4)	29 (36.3)
III	352 (9.9)	8 (8.4)	9 (11.3)
ER status, n (%)			
ER+	2981 (83.9)	79 (83.2)	69 (86.3)
ER-	574 (16.2)	16 (16.8)	11 (13.8)
Subtype IHC, n (%)			
Luminal A	2555 (74.9)	73 (79.4)	60 (77.9)
Luminal B	305 (8.9)	5 (5.4)	6 (7.8)
NNN	395 (11.6)	12 (13.0)	7 (9.1)
HER-2 OVER	158 (4.6)	2 (2.2)	4 (5.2)
Subtype Gallen, n (%)			
Luminal A	1767 (51.5)	48 (52.2)	41 (52.6)
Luminal B	1111 (32.4)	30 (32.6)	26 (33.3)
HER-2E	158 (4.6)	2 (2.2)	4 (5.1)
NNN	395 (11.5)	12 (13.0)	7 (9.0)
Had chemotherapy, n (%)	1252 (35.3)	29 (30.9)	17 (21.5)
Had radiation therapy, n (%)	1578 (44.4)	54 (56.8)	35 (43.8)
Had endocrine therapy, n (%)	2044 (57.9)	43 (45.3)	30 (38.0)

DISCUSSION AND FUTURE DIRECTIONS

The transcriptional program of leader cells

In this work, we initially set out to define the transcriptional program present in leader cells. Using single cell RNA sequencing, we were able to identify transcriptional programs that co-occur with upregulation of Krt14 in leader cells. We found this program is remarkably similar to that of basal epidermal cells, which lead us to investigate the role of skin-related therapeutics in basal breast cancers. However, this dataset contains a wealth of information that we have only begun to utilize. In our data, there are numerous markers upregulated in invasive basal cells that are associated with migration, cell-cell adhesion, cell-matrix adhesion, and extracellular matrix remodeling. There are additional genes that encode for proteins that are still only partly characterized. Any, or many, of these genes may encode for proteins essential for collective invasion and metastasis. We are eager to understand these further and prioritize future targets.

The transcriptional program that is present in leader cells may also enable reverse-engineering of the signaling pathways that induce invasion. Presumably, cell-cell and matrix-cell signaling pathways converge to activate particular transcription factors which then go on to activate and repress transcription leading to the program we have observed. More and more, datasets are becoming available that characterize the transcriptional output of various transcription factors, signaling pathways, and extracellular perturbations¹⁸². These can be used to elucidate the events that precede leader cell invasion. For instance, integrins and discoidin domain receptors can be

activated by extracellular collagen I, growth factor and Notch signaling can be activated by cell-cell signaling, and force activated pathways such as YAP signaling can be activated by stiff extracellular matrices and cellular forces^{111,183,184}. It is likely that breast cancer cells integrate signaling from these pathways and more as cellular fate decisions are made. Further investigation into the upstream pathways that dictate leader cell formation may reveal key steps that can be therapeutically regulated.

Our dataset also includes information on the transcriptional program of luminal tumor cells. We believe this data could additionally guide work to understand the “choice” a cell makes when becoming luminal or basal. It is possible that there are potential pathways to prevent growth of luminal cancer cells or encourage cells to become luminal rather than basal by investigating the transcriptional regulation of these two identities.

Epidermal-like transcription program in breast cancer

Our work has indicated that breast cancer cells are able to undergo a novel type of plasticity, one in which the cell’s transcriptional state shifts towards that of an epidermal keratinocyte. While we have not yet fully characterized this new cell state and its true similarity to the epidermis, it is possible that we are observing a form of developmental plasticity. Developmental plasticity is a common theme in cancer biology, in which cancer cells lose differentiation and regress towards a more embryonic or stem-like state¹⁸⁵. During development, the mammary gland is formed from the ectoderm as an epidermal appendage, meaning that the epidermis and the mammary gland share a

developmental past ^{47,186}. Further work is warranted to determine whether basal breast cancer cells are truly regressing, or whether they are entering a novel state. If developmental plasticity is occurring, its study could provide novel insights into how the healthy mammary gland forms and how cell state is programmed in breast cancer.

We have also made note of the apparent dual role of epidermal markers in breast cancer metastasis. While in this study blocking epidermal differentiation pathways in breast cancer organoids drives invasion, it has also previously been reported that Keratin 14 and p63, two protein markers of basal epidermal keratinocytes, are required for collective tumor organoid invasion ³. It is possible that there are two alternative migratory programs relying on different epidermal marker genes. It is additionally possible that breast cancer cells are most migratory when in a hybrid epidermal-like state or a unique epidermal-like state such as that found in wound healing, in which epidermal cells behave differently from homeostasis. The dual role of keratins in migration of cells has been demonstrated in other keratins, such as keratin 8¹⁸⁷. Further, keratin 6 is a classic marker of wound healing, a highly migratory keratinocyte program, yet loss of keratin 6 leads to enhanced rates of migration¹⁸⁸. There is still much to be understood regarding keratin dynamics, signaling, and structural properties, all of which may play a role in the phenotypes we have described¹⁸⁹.

Retinoic acid in breast cancer

Because of the plasticity and developmental regression often observed in cancers, differentiation therapy has long been proposed as an anti-cancer approach ⁷⁶. If, for instance, breast cancer cells could be re-differentiated into healthy breast cells in homeostasis, the cells would no longer pose as much of a health risk. The prospect of differentiation therapy in cancer became tangible when retinoic acid was used to successfully differentiate acute promyelocytic leukemia (APL) ¹⁹⁰. Retinoic acid has since been proposed as a pro-differentiation agent in a variety of cancers, and has found additional clinical success in neuroblastoma ^{76,191}.

Since then, additional differentiation therapies have been tested, mostly in pre-clinical models. These approaches do not necessarily need to convert tumor cells into healthy tumor cells of the same tissue. For instance, in one approach relevant to breast cancer, PPAR γ agonists can be used to differentiate breast cancer cells into adipocytes to prevent metastasis ^{103,109}.

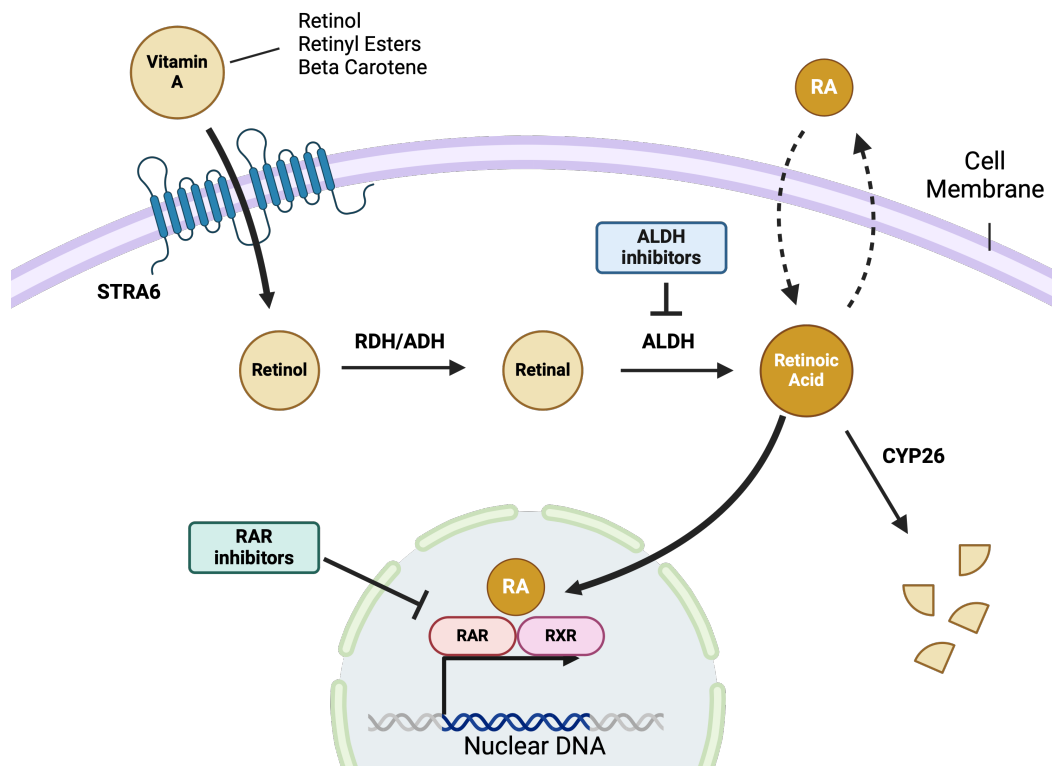
Our work proposes that basal breast cancers can be sensitive to the same pathways that drive skin differentiation. In normal keratinocytes, lack of retinoic acid or excess expression of Klf4 reduces proliferative potential and leads to excess keratinization. Conversely, excess retinoic acid or downregulation of Klf4 leads to reduced expression of differentiation markers and increased proliferation rates ^{132,144,148}. While it is unlikely that breast cancer cells are undergoing cornification, a carefully orchestrated form of

keratinocyte terminal differentiation ¹²³, they do appear to respond to perturbations in a similar manner to keratinocytes.

The findings presented in this thesis implicate RAR inhibitors as a potential novel therapy in breast cancers with epidermal marker gene expression. While pan-RAR antagonists would likely have toxic side effects, more specific inhibitors have already made it to clinical trials. Retinoic acid pathway proteins have been proposed as targets for male contraceptives and have had success in preclinical models. WIN18446, an inhibitor of Aldehyde dehydrogenase 1a2 (Aldh1a2) which enzymatically converts retinal to retinoic acid, was tested in patients with relative success, until it was abandoned for negative interactions with alcohol ¹⁹²⁻¹⁹⁴ (**Figure 7**). A compound that specifically inhibits RAR α , YCT-529, has just begun use in a Phase I trial as a male contraceptive (NCT06094283).

Future work will focus on carefully delineating which RAR is most important for basal breast cancer cell identity, and to screen specific inhibitors of key retinoic acid pathway proteins. It is our hope that one of these potential molecules will be safe for human use, while preventing cancer spread just as BMS493 has in mouse models.

Figure 7: Retinoic acid signaling



Of note, there are numerous reports of retinoic acid acting as an anti-cancer agent in breast cancer. Many of these studies have found that luminal, particularly estrogen receptor positive, breast cancers may have reduced viability when exposed to excess retinoic acid. However, basal and ER-negative models have been resistant to retinoic acid therapy¹⁹⁵⁻¹⁹⁷. Such findings align with the knowledge that retinoic acid acts in a context and tissue-dependent manner. For instance, retinoic acid prevents cornification and terminal differentiation of keratinocytes, but is required for proper spermatogenesis and neutrophil differentiation^{144,198,199}. It is therefore reasonable that retinoic acid functions differently in breast cancer in given contexts, particularly in cases in which

breast cancer cells transcriptionally resemble another tissue. We believe our work may enhance our understanding of how retinoic acid signaling acts in breast cancers, and in which cancers retinoic acid may have therapeutic potential.

Beyond understanding how RAR inhibition may be a possible route to breast cancer treatment, we are interested in understanding the effect of excess retinoic acid in breast cancer. In this work, we have shown that retinoic acid can promote breast cancer proliferation and invasion in cell culture. This invasion appears to occur in the absence of classical EMT. We believe it is important to fully characterize this form of invasion, and clarify whether or not it occurs in patients. If cells are utilizing aspects of the retinoic acid invasion program during metastasis, a novel mechanism of cancer migration may be uncovered. Understanding how such a pathway works could allow for a more targeted approach to prevent this behavior. In our work, treatment of mice with retinoic acid had no effect on tumor and metastatic burden. However, an important difference between our cell culture and in-vivo models is the absence of retinoic acid in our culture media, but the inclusion of Vitamin A in mouse diets. It is possible that in the context of a complete diet, there is sufficient retinoic acid present in-vivo to drive invasion and proliferation of tumor cells and that excess retinoic acid is redundant.

Additionally, we are interested in identifying the primary source of retinoic acid in tumor cells. Vitamin A is processed by the intestines and the liver into retinol, which circulates in the blood. Once transported into a cell, various enzymes process the retinol until it becomes a form of retinoic acid which can bind RARs²⁰⁰ (**Figure 7**). Retinoic acid can

then easily cross the cell membrane and can be passed between cells. We are interested in understanding whether retinoic acid in the tumor is derived from circulation, or whether tumors can act as a sink for retinoids to promote proliferative potential.

Vitamin A and cancer outcomes

Finally, our work has added to the growing evidence that diet and cancer outcomes are closely linked. Just as is the case for smoking, physical activity, and UV exposure, dietary habits have a clear role in rates of cancer occurrence and survival²⁰¹. For instance, high alcohol consumption is correlated with diagnosis rates in several types of cancers, and has been compared to tobacco use²⁰². Obesity and excess fatty acid have been estimated to be contributing factors in up to 20% of cancers²⁰³. Due to the metabolic dependence on glucose, Keto diet has been proposed as a supplemental treatment for cancer patients²⁰⁴. The role of vitamin supplements in cancer outcomes has been less clear. This may be for several reasons. Chief among them is the fact that supplements are largely unregulated by the FDA, and are sold as different chemical structures and in different concentrations. However, an estimated 70% or more of cancer patients take supplemental vitamins, often with the belief that vitamins will provide protection against cancer^{166,167}.

This belief may be attributed to the widely publicized anti-cancer effects of antioxidants such as vitamins A, C, and E, although studies investigating these claims have come to mixed conclusions. Reactive oxygen species are well known to induce DNA damage

and induce carcinogenesis, yet at the same time are often a cause of tumor cell death

205

We were fortunate to collaborate with scientists at Fred Hutchinson Cancer Center who conducted the QUILT study and scientists at Kaiser Permanente conducting the Pathways study^{153,154,166}. Our post-hoc analysis indicates that vitamin A intake may be associated with increased occurrence and poor outcomes, particularly in patients with hormone negative breast cancers. These findings underscore the need for fine-tuned analysis of vitamin supplement intake and cancer outcomes, with particular attention on specific populations. It is imperative that clinicians and patients are armed with knowledge of how supplements work before use to avoid any unwanted negative effects.

REFERENCES

- 1 Society, A. C. (American Cancer Society, 2024).
- 2 Surveillance Research Program, N. C. I. (2024).
- 3 Cheung, K. J., Gabrielson, E., Werb, Z. & Ewald, A. J. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell* **155**, 1639-1651 (2013).
<https://doi.org/10.1016/j.cell.2013.11.029>
- 4 Cheung, K. J. *et al.* Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proc Natl Acad Sci U S A* **113**, E854-863 (2016). <https://doi.org/10.1073/pnas.1508541113>
- 5 Giuliano, A. E. *et al.* Breast Cancer-Major changes in the American Joint Committee on Cancer eighth edition cancer staging manual. *CA Cancer J Clin* **67**, 290-303 (2017). <https://doi.org/10.3322/caac.21393>
- 6 Waks, A. G. & Winer, E. P. Breast Cancer Treatment: A Review. *JAMA* **321**, 288-300 (2019). <https://doi.org/10.1001/jama.2018.19323>
- 7 Sanpaolo, P., Barbieri, V. & Genovesi, D. Prognostic value of breast cancer subtypes on breast cancer specific survival, distant metastases and local relapse rates in conservatively managed early stage breast cancer: a retrospective clinical study. *Eur J Surg Oncol* **37**, 876-882 (2011). <https://doi.org/10.1016/j.ejso.2011.07.001>
- 8 Pan, H. *et al.* 20-Year Risks of Breast-Cancer Recurrence after Stopping Endocrine Therapy at 5 Years. *N Engl J Med* **377**, 1836-1846 (2017).
<https://doi.org/10.1056/NEJMoa1701830>
- 9 Bianchini, G., De Angelis, C., Licata, L. & Gianni, L. Treatment landscape of triple-negative breast cancer - expanded options, evolving needs. *Nat Rev Clin Oncol* **19**, 91-113 (2022). <https://doi.org/10.1038/s41571-021-00565-2>
- 10 Chaffer, C. L. & Weinberg, R. A. A perspective on cancer cell metastasis. *Science* **331**, 1559-1564 (2011). <https://doi.org/10.1126/science.1203543>
- 11 Gerstberger, S., Jiang, Q. & Ganesh, K. Metastasis. *Cell* **186**, 1564-1579 (2023).
<https://doi.org/10.1016/j.cell.2023.03.003>
- 12 Massague, J. & Ganesh, K. Metastasis-Initiating Cells and Ecosystems. *Cancer Discov* **11**, 971-994 (2021). <https://doi.org/10.1158/2159-8290.CD-21-0010>
- 13 Balkwill, F. R., Capasso, M. & Hagemann, T. The tumor microenvironment at a glance. *J Cell Sci* **125**, 5591-5596 (2012). <https://doi.org/10.1242/jcs.116392>
- 14 Marusyk, A., Almendro, V. & Polyak, K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer* **12**, 323-334 (2012). <https://doi.org/10.1038/nrc3261>
- 15 Fidler, I. J. & Kripke, M. L. Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197**, 893-895 (1977).
<https://doi.org/10.1126/science.887927>
- 16 Brattain, M. G., Strobil-Stevens, J., Fine, D., Webb, M. & Sarraf, A. M. Establishment of mouse colonic carcinoma cell lines with different metastatic properties. *Cancer Res* **40**, 2142-2146 (1980).

- 17 Clark, E. A., Golub, T. R., Lander, E. S. & Hynes, R. O. Genomic analysis of metastasis reveals an essential role for RhoC. *Nature* **406**, 532-535 (2000). <https://doi.org/10.1038/35020106>
- 18 Li, Y. *et al.* Establishment of cell clones with different metastatic potential from the metastatic hepatocellular carcinoma cell line MHCC97. *World J Gastroenterol* **7**, 630-636 (2001). <https://doi.org/10.3748/wjg.v7.i5.630>
- 19 Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* **352**, 169-175 (2016). <https://doi.org/10.1126/science.aaf2784>
- 20 Nguyen, D. X. & Massague, J. Genetic determinants of cancer metastasis. *Nat Rev Genet* **8**, 341-352 (2007). <https://doi.org/10.1038/nrg2101>
- 21 Angus, L. *et al.* The genomic landscape of metastatic breast cancer highlights changes in mutation and signature frequencies. *Nat Genet* **51**, 1450-1458 (2019). <https://doi.org/10.1038/s41588-019-0507-7>
- 22 Watkins, T. B. K. *et al.* Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* **587**, 126-132 (2020). <https://doi.org/10.1038/s41586-020-2698-6>
- 23 Zingg, D. *et al.* The epigenetic modifier EZH2 controls melanoma growth and metastasis through silencing of distinct tumour suppressors. *Nat Commun* **6**, 6051 (2015). <https://doi.org/10.1038/ncomms7051>
- 24 Fang, F. *et al.* Breast cancer methylomes establish an epigenomic foundation for metastasis. *Sci Transl Med* **3**, 75ra25 (2011). <https://doi.org/10.1126/scitranslmed.3001875>
- 25 Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nat Genet* **33**, 49-54 (2003). <https://doi.org/10.1038/ng1060>
- 26 Teng, S. *et al.* Tissue-specific transcription reprogramming promotes liver metastasis of colorectal cancer. *Cell Res* **30**, 34-49 (2020). <https://doi.org/10.1038/s41422-019-0259-z>
- 27 Tasdogan, A. *et al.* Metabolic heterogeneity confers differences in melanoma metastatic potential. *Nature* **577**, 115-120 (2020). <https://doi.org/10.1038/s41586-019-1847-2>
- 28 Semenza, G. L. Molecular mechanisms mediating metastasis of hypoxic breast cancer cells. *Trends Mol Med* **18**, 534-543 (2012). <https://doi.org/10.1016/j.molmed.2012.08.001>
- 29 Choi, C. & Helfman, D. M. The Ras-ERK pathway modulates cytoskeleton organization, cell motility and lung metastasis signature genes in MDA-MB-231 LM2. *Oncogene* **33**, 3668-3676 (2014). <https://doi.org/10.1038/onc.2013.341>
- 30 Greenburg, G. & Hay, E. D. Epithelia suspended in collagen gels can lose polarity and express characteristics of migrating mesenchymal cells. *J Cell Biol* **95**, 333-339 (1982). <https://doi.org/10.1083/jcb.95.1.333>
- 31 Brabletz, T., Kalluri, R., Nieto, M. A. & Weinberg, R. A. EMT in cancer. *Nat Rev Cancer* **18**, 128-134 (2018). <https://doi.org/10.1038/nrc.2017.118>

- 32 Boyer, B. & Thiery, J. P. Epithelium-mesenchyme interconversion as example of epithelial plasticity. *APMIS* **101**, 257-268 (1993). <https://doi.org/10.1111/j.1699-0463.1993.tb00109.x>
- 33 Pastushenko, I. & Blanpain, C. EMT Transition States during Tumor Progression and Metastasis. *Trends Cell Biol* **29**, 212-226 (2019). <https://doi.org/10.1016/j.tcb.2018.12.001>
- 34 Chaffer, C. L., San Juan, B. P., Lim, E. & Weinberg, R. A. EMT, cell plasticity and metastasis. *Cancer Metastasis Rev* **35**, 645-654 (2016). <https://doi.org/10.1007/s10555-016-9648-7>
- 35 Yang, J. *et al.* Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **21**, 341-352 (2020). <https://doi.org/10.1038/s41580-020-0237-9>
- 36 Jolly, M. K., Mani, S. A. & Levine, H. Hybrid epithelial/mesenchymal phenotype(s): The 'fittest' for metastasis? *Biochim Biophys Acta Rev Cancer* **1870**, 151-157 (2018). <https://doi.org/10.1016/j.bbcan.2018.07.001>
- 37 Jehanno, C., Vulin, M., Richina, V., Richina, F. & Bentires-Alj, M. Phenotypic plasticity during metastatic colonization. *Trends Cell Biol* **32**, 854-867 (2022). <https://doi.org/10.1016/j.tcb.2022.03.007>
- 38 Pérez-González, A., Bévant, K. & Blanpain, C. Cancer cell plasticity during tumor progression, metastasis and response to therapy. *Nat Cancer* **4**, 1063-1082 (2023). <https://doi.org/10.1038/s43018-023-00595-y>
- 39 Fumagalli, A. *et al.* Plasticity of Lgr5-Negative Cancer Cells Drives Metastasis in Colorectal Cancer. *Cell Stem Cell* **26**, 569-578 e567 (2020). <https://doi.org/10.1016/j.stem.2020.02.008>
- 40 Lee, C. K. *et al.* Tumor metastasis to lymph nodes requires YAP-dependent metabolic adaptation. *Science* **363**, 644-649 (2019). <https://doi.org/10.1126/science.aav0173>
- 41 Huh, H. D. *et al.* Reprogramming anchorage dependency by adherent-to-suspension transition promotes metastatic dissemination. *Mol Cancer* **22**, 63 (2023). <https://doi.org/10.1186/s12943-023-01753-7>
- 42 Wicki, A. *et al.* Tumor invasion in the absence of epithelial-mesenchymal transition: podoplanin-mediated remodeling of the actin cytoskeleton. *Cancer Cell* **9**, 261-272 (2006). <https://doi.org/10.1016/j.ccr.2006.03.010>
- 43 Yuan, S., Almagro, J. & Fuchs, E. Beyond genetics: driving cancer with the tumour microenvironment behind the wheel. *Nat Rev Cancer* **24**, 274-286 (2024). <https://doi.org/10.1038/s41568-023-00660-9>
- 44 Friedl, P. & Alexander, S. Cancer invasion and the microenvironment: plasticity and reciprocity. *Cell* **147**, 992-1009 (2011). <https://doi.org/10.1016/j.cell.2011.11.016>
- 45 Taddei, M. L., Giannoni, E., Comito, G. & Chiarugi, P. Microenvironment and tumor cell plasticity: an easy way out. *Cancer Lett* **341**, 80-96 (2013). <https://doi.org/10.1016/j.canlet.2013.01.042>
- 46 Hanahan, D. & Monje, M. Cancer hallmarks intersect with neuroscience in the tumor microenvironment. *Cancer Cell* **41**, 573-580 (2023). <https://doi.org/10.1016/j.ccell.2023.02.012>

- 47 Macias, H. & Hinck, L. Mammary gland development. *Wiley Interdiscip Rev Dev Biol* **1**, 533-557 (2012). <https://doi.org/10.1002/wdev.35>
- 48 Fu, N. Y., Nolan, E., Lindeman, G. J. & Visvader, J. E. Stem Cells and the Differentiation Hierarchy in Mammary Gland Development. *Physiol Rev* **100**, 489-523 (2020). <https://doi.org/10.1152/physrev.00040.2018>
- 49 Gudjonsson, T., Adriance, M. C., Sternlicht, M. D., Petersen, O. W. & Bissell, M. J. Myoepithelial cells: their origin and function in breast morphogenesis and neoplasia. *J Mammary Gland Biol Neoplasia* **10**, 261-272 (2005). <https://doi.org/10.1007/s10911-005-9586-4>
- 50 Agarwal, S. et al. Identification of Different Classes of Luminal Progenitor Cells within Prostate Tumors. *Cell Rep* **13**, 2147-2158 (2015). <https://doi.org/10.1016/j.celrep.2015.10.077>
- 51 Liu, X. et al. Low CD38 Identifies Progenitor-like Inflammation-Associated Luminal Cells that Can Initiate Human Prostate Cancer and Predict Poor Outcome. *Cell Rep* **17**, 2596-2606 (2016). <https://doi.org/10.1016/j.celrep.2016.11.010>
- 52 Lu, T. L. et al. Conditionally ablated Pten in prostate basal cells promotes basal-to-luminal differentiation and causes invasive prostate cancer in mice. *Am J Pathol* **182**, 975-991 (2013). <https://doi.org/10.1016/j.ajpath.2012.11.025>
- 53 Zhang, D. et al. Stem cell and neurogenic gene-expression profiles link prostate basal cells to aggressive prostate cancer. *Nat Commun* **7**, 10798 (2016). <https://doi.org/10.1038/ncomms10798>
- 54 Ooi, A. T. et al. Presence of a putative tumor-initiating progenitor cell population predicts poor prognosis in smokers with non-small cell lung cancer. *Cancer Res* **70**, 6639-6648 (2010). <https://doi.org/10.1158/0008-5472.CAN-10-0455>
- 55 Yao, S. et al. Keratin 14-high subpopulation mediates lung cancer metastasis potentially through Gkn1 upregulation. *Oncogene* **38**, 6354-6369 (2019). <https://doi.org/10.1038/s41388-019-0889-0>
- 56 Sonzogni, O. et al. Reporters to mark and eliminate basal or luminal epithelial cells in culture and in vivo. *PLoS Biol* **16**, e2004049 (2018). <https://doi.org/10.1371/journal.pbio.2004049>
- 57 Hwang, P. Y., Brenot, A., King, A. C., Longmore, G. D. & George, S. C. Randomly Distributed K14(+) Breast Tumor Cells Polarize to the Leading Edge and Guide Collective Migration in Response to Chemical and Mechanical Environmental Cues. *Cancer Res* **79**, 1899-1912 (2019). <https://doi.org/10.1158/0008-5472.CAN-18-2828>
- 58 Verma, A. et al. EZH2-H3K27me3 mediated KRT14 upregulation promotes TNBC peritoneal metastasis. *Nat Commun* **13**, 7344 (2022). <https://doi.org/10.1038/s41467-022-35059-x>
- 59 Bilandzic, M. et al. Keratin-14 (KRT14) Positive Leader Cells Mediate Mesothelial Clearance and Invasion by Ovarian Cancer Cells. *Cancers (Basel)* **11** (2019). <https://doi.org/10.3390/cancers11091228>
- 60 Ricciardelli, C. et al. Keratin 5 overexpression is associated with serous ovarian cancer recurrence and chemotherapy resistance. *Oncotarget* **8**, 17819-17832 (2017). <https://doi.org/10.18632/oncotarget.14867>

- 61 Karimnia, N. *et al.* Chemoresistance is mediated by ovarian cancer leader cells in vitro. *J Exp Clin Cancer Res* **40**, 276 (2021). <https://doi.org/10.1186/s13046-021-02086-3>
- 62 Volkmer, J. P. *et al.* Three differentiation states risk-stratify bladder cancer into distinct subtypes. *Proc Natl Acad Sci U S A* **109**, 2078-2083 (2012). <https://doi.org/10.1073/pnas.1120605109>
- 63 Wang, M. *et al.* Acquired semi-squamatization during chemotherapy suggests differentiation as a therapeutic strategy for bladder cancer. *Cancer Cell* **40**, 1044-1059 e1048 (2022). <https://doi.org/10.1016/j.ccell.2022.08.010>
- 64 Dmello, C. *et al.* Vimentin regulates differentiation switch via modulation of keratin 14 levels and their expression together correlates with poor prognosis in oral cancer patients. *PLoS One* **12**, e0172559 (2017). <https://doi.org/10.1371/journal.pone.0172559>
- 65 Wang, S. *et al.* Clinical utility of keratin 14 expression measurement in reflecting the tumor properties and prognosis in patients with renal cell carcinoma: a study with long-term follow-up. *Int Urol Nephrol* **56**, 2045-2053 (2024). <https://doi.org/10.1007/s11255-023-03923-4>
- 66 Gao, X. L. *et al.* Cytokeratin-14 contributes to collective invasion of salivary adenoid cystic carcinoma. *PLoS One* **12**, e0171341 (2017). <https://doi.org/10.1371/journal.pone.0171341>
- 67 Yamamoto, A., Doak, A. E. & Cheung, K. J. Orchestration of Collective Migration and Metastasis by Tumor Cell Clusters. *Annu Rev Pathol* **18**, 231-256 (2023). <https://doi.org/10.1146/annurev-pathmechdis-031521-023557>
- 68 Wrenn, E., Huang, Y. & Cheung, K. Collective metastasis: coordinating the multicellular voyage. *Clin Exp Metastasis* **38**, 373-399 (2021). <https://doi.org/10.1007/s10585-021-10111-0>
- 69 Aceto, N. *et al.* Circulating tumor cell clusters are oligoclonal precursors of breast cancer metastasis. *Cell* **158**, 1110-1122 (2014). <https://doi.org/10.1016/j.cell.2014.07.013>
- 70 Moll, R., Divo, M. & Langbein, L. The human keratins: biology and pathology. *Histochem Cell Biol* **129**, 705-733 (2008). <https://doi.org/10.1007/s00418-008-0435-6>
- 71 Kroger, C. *et al.* Keratins control intercellular adhesion involving PKC-alpha-mediated desmoplakin phosphorylation. *J Cell Biol* **201**, 681-692 (2013). <https://doi.org/10.1083/jcb.201208162>
- 72 Lambert, A. W. & Weinberg, R. A. Linking EMT programmes to normal and neoplastic epithelial stem cells. *Nat Rev Cancer* **21**, 325-338 (2021). <https://doi.org/10.1038/s41568-021-00332-6>
- 73 Obenauf, A. C. & Massagué, J. Surviving at a Distance: Organ-Specific Metastasis. *Trends in Cancer* **1**, 76-91 (2015). <https://doi.org/10.1016/j.trecan.2015.07.009>
- 74 Vilchez Mercedes, S. A. *et al.* Decoding leader cells in collective cancer invasion. *Nat Rev Cancer* **21**, 592-604 (2021). <https://doi.org/10.1038/s41568-021-00376-8>

- 75 Karras, P., Black, J. R. M., McGranahan, N. & Marine, J. C. Decoding the interplay between genetic and non-genetic drivers of metastasis. *Nature* **629**, 543-554 (2024). <https://doi.org/10.1038/s41586-024-07302-6>
- 76 de Thé, H. Differentiation therapy revisited. *Nat Rev Cancer* **18**, 117-127 (2018). <https://doi.org/10.1038/nrc.2017.103>
- 77 Löönd, F., Tiede, S. & Christofori, G. Breast cancer as an example of tumour heterogeneity and tumour cell plasticity during malignant progression. *Br J Cancer* **125**, 164-175 (2021). <https://doi.org/10.1038/s41416-021-01328-7>
- 78 Massagué, J. & Ganesh, K. Metastasis-Initiating Cells and Ecosystems. *Cancer Discovery* **11**, 971-994 (2021). <https://doi.org/10.1158/2159-8290.CD-21-0010>
- 79 Quintanal-Villalonga, Á. et al. Lineage plasticity in cancer: a shared pathway of therapeutic resistance. *Nat Rev Clin Oncol* **17**, 360-371 (2020). <https://doi.org/10.1038/s41571-020-0340-z>
- 80 Friedl, P., Locker, J., Sahai, E. & Segall, J. E. Classifying collective cancer cell invasion. *Nat Cell Biol* **14**, 777-783 (2012). <https://doi.org/10.1038/ncb2548>
- 81 Gaggioli, C. et al. Fibroblast-led collective invasion of carcinoma cells with differing roles for RhoGTPases in leading and following cells. *Nature Cell Biology* **2007** 9:12 9 (2007). <https://doi.org/10.1038/ncb1658>
- 82 Summerbell, E. R. et al. Epigenetically heterogeneous tumor cells direct collective invasion through filopodia-driven fibronectin micropatterning. *Science Advances* **6** (2020). <https://doi.org/10.1126/sciadv.aaz6197>
- 83 Zhang, J. et al. Energetic regulation of coordinated leader–follower dynamics during collective invasion of breast cancer cells. *Proceedings of the National Academy of Sciences* **116** (2019). <https://doi.org/10.1073/pnas.1809964116>
- 84 Clark, A. G. et al. Self-generated gradients steer collective migration on viscoelastic collagen networks. *Nat Mater* **21**, 1200-1210 (2022). <https://doi.org/10.1038/s41563-022-01259-5>
- 85 Saxena, K., Jolly, M. K. & Balamurugan, K. Hypoxia, partial EMT and collective migration: Emerging culprits in metastasis - PubMed. *Translational oncology* **13** (2020). <https://doi.org/10.1016/j.tranon.2020.100845>
- 86 Ye, X. et al. Upholding a role for EMT in breast cancer metastasis. *Nature* **2017** 547:7661 **547** (2017). <https://doi.org/10.1038/nature22816>
- 87 Padmanaban, V., Tsehay, Y., Cheung, K. J., Ewald, A. J. & Bader, J. S. Between-tumor and within-tumor heterogeneity in invasive potential. *PLOS Computational Biology* **16** (2020). <https://doi.org/10.1371/journal.pcbi.1007464>
- 88 Hwang, P. Y., Brenot, A., King, A. C., Longmore, G. D. & George, S. C. Randomly Distributed K14+ Breast Tumor Cells Polarize to the Leading Edge and Guide Collective Migration in Response to Chemical and Mechanical Environmental CuesLeader Cell Formation during Tumor Collective Migration. *Cancer research* **79**, 1899-1912 (2019).
- 89 Hwang, P. Y. et al. A Cdh3-β-catenin-laminin signaling axis in a subset of breast tumor leader cells control leader cell polarization and directional collective migration. *Developmental Cell* **58**, 34-50. e39 (2023).

- 90 Sonzogni, O. *et al.* Reporters to mark and eliminate basal or luminal epithelial cells in culture and in vivo. *PLoS biology* **16**, e2004049 (2018).
- 91 Lambert, A. W. *et al.* Δ Np63/p73 drive metastatic colonization by controlling a regenerative epithelial stem cell program in quasi-mesenchymal cancer stem cells. *Developmental cell* **57**, 2714-2730. e2718 (2022).
- 92 Lawson, D. A. *et al.* Single-cell analysis reveals a stem-cell program in human metastatic breast cancer cells. *Nature* **526**, 131-135 (2015).
- 93 Cheung, K. J. *et al.* Polyclonal breast cancer metastases arise from collective dissemination of keratin 14-expressing tumor cell clusters. *Proceedings of the National Academy of Sciences* **113**, E854-E863 (2016).
- 94 Chapman, M. P. *et al.* Modeling differentiation-state transitions linked to therapeutic escape in triple-negative breast cancer. *PLoS computational biology* **15**, e1006840 (2019).
- 95 Risom, T. *et al.* Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. *Nature communications* **9**, 1-17 (2018).
- 96 Li, Z. *et al.* ESR1 mutant breast cancers show elevated basal cytokeratins and immune activation. *Nat Commun* **13**, 2011 (2022). <https://doi.org/10.1038/s41467-022-29498-9>
- 97 Inayatullah, M. *et al.* Basal–epithelial subpopulations underlie and predict chemotherapy resistance in triple-negative breast cancer. *EMBO Molecular Medicine* **16**, 823-853 (2024).
- 98 Bado, I. L. *et al.* The bone microenvironment increases phenotypic plasticity of ER+ breast cancer cells. *Developmental cell* **56**, 1100-1117. e1109 (2021).
- 99 Perou, C. M. *et al.* Molecular portraits of human breast tumours. *nature* **406**, 747-752 (2000).
- 100 Burstein, M. D. *et al.* Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. *Clinical Cancer Research* **21**, 1688-1698 (2015).
- 101 Derakhshan, F. & Reis-Filho, J. S. Pathogenesis of Triple-Negative Breast Cancer. *Annu Rev Pathol* **17**, 181-204 (2022). <https://doi.org/10.1146/annurev-pathol-042420-093238>
- 102 Gupta, P. B. *et al.* Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138**, 645-659 (2009). <https://doi.org/10.1016/j.cell.2009.06.034>
- 103 Ishay-Ronen, D. *et al.* Gain Fat-Lose Metastasis: Converting Invasive Breast Cancer Cells into Adipocytes Inhibits Cancer Metastasis. *Cancer Cell* **35**, 17-32.e16 (2019). <https://doi.org/10.1016/j.ccell.2018.12.002>
- 104 Lu, D. *et al.* Salinomycin inhibits Wnt signaling and selectively induces apoptosis in chronic lymphocytic leukemia cells. *Proc Natl Acad Sci U S A* **108**, 13253-13257 (2011). <https://doi.org/10.1073/pnas.1110431108>
- 105 Ramesh, V., Brabletz, T. & Ceppi, P. Targeting EMT in Cancer with Repurposed Metabolic Inhibitors. *Trends in Cancer* **6**, 942-950 (2020). <https://doi.org/10.1016/j.trecan.2020.06.005>

- 106 Lengrand, J. *et al.* Pharmacological targeting of netrin-1 inhibits EMT in cancer. *Nature* (2023).
- 107 Pastushenko, I. *et al.* Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. *Nature* 2020 589:7842 **589** (2020). <https://doi.org/10.1038/s41586-020-03046-1>
- 108 Debaugnies, M. *et al.* RHOJ controls EMT-associated resistance to chemotherapy. *Nature* 2023 616:7955 **616** (2023). <https://doi.org/10.1038/s41586-023-05838-7>
- 109 Mueller, E. *et al.* Terminal differentiation of human breast cancer through PPAR gamma. (1998).
- 110 Grasset, E. M. *et al.* Triple-negative breast cancer metastasis involves complex epithelial-mesenchymal transition dynamics and requires vimentin. *Science Translational Medicine* **14**, eabn7571 (2022). <https://doi.org/10.1126/scitranslmed.abn7571>
- 111 Khalil, A. A. *et al.* A YAP-centered mechanotransduction loop drives collective breast cancer cell invasion. *Nat Commun* **15**, 4866 (2024). <https://doi.org/10.1038/s41467-024-49230-z>
- 112 Attalla, S., Taifour, T., Bui, T. & Muller, W. Insights from transgenic mouse models of PyMT-induced breast cancer: recapitulating human breast cancer progression in vivo. *Oncogene* **40**, 475-491 (2021). <https://doi.org/10.1038/s41388-020-01560-0>
- 113 Guy, C. T., Cardiff, R. D. & Muller, W. J. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol* **12**, 954-961 (1992). <https://doi.org/10.1128/mcb.12.3.954-961.1992>
- 114 Mohamed, G. A. *et al.* Lineage plasticity enables low-ER luminal tumors to evolve and gain basal-like traits. *Breast Cancer Res* **25**, 23 (2023). <https://doi.org/10.1186/s13058-023-01621-8>
- 115 Fei, L. *et al.* Systematic identification of cell-fate regulatory programs using a single-cell atlas of mouse development. *Nature Genetics* 2022 54:7 **54** (2022). <https://doi.org/10.1038/s41588-022-01118-8>
- 116 Lazard, D. *et al.* Expression of smooth muscle-specific proteins in myoepithelium and stromal myofibroblasts of normal and malignant human breast tissue. *Proc Natl Acad Sci U S A* **90** (1993). <https://doi.org/10.1073/pnas.90.3.999>
- 117 Petersen, O. W. & Van Deurs, B. Distinction between vascular smooth muscle cells and myoepithelial cells in primary monolayer cultures of human breast tissue. *In Vitro Cellular & Developmental Biology* 1989 25:3 **25** (1989). <https://doi.org/10.1007/BF02628464>
- 118 Doha, Z. O. *et al.* MYC Deregulation and PTEN Loss Model Tumor and Stromal Heterogeneity of Aggressive Triple-Negative Breast Cancer. *Nat Commun* **14**, 5665 (2023). <https://doi.org/10.1038/s41467-023-40841-6>
- 119 Thennavan, A., Garcia-Recio, S., Liu, S., He, X. & Perou, C. M. Molecular signatures of in situ to invasive progression for basal-like breast cancers: An integrated mouse model and human DCIS study. *npj Breast Cancer* **8**, 1-13 (2022). <https://doi.org/10.1038/s41523-022-00450-w>

- 120 Yeo, S. K. *et al.* Single-cell RNA-sequencing reveals distinct patterns of cell state heterogeneity in mouse models of breast cancer. *eLife* **9**, e58810 (2020). <https://doi.org/10.7554/eLife.58810>
- 121 Pal, B. *et al.* A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J* **40**, e107333 (2021). <https://doi.org/10.15252/embj.2020107333>
- 122 Wu, S. Z. *et al.* A single-cell and spatially resolved atlas of human breast cancers. *Nat Genet* **53**, 1334-1347 (2021). <https://doi.org/10.1038/s41588-021-00911-1>
- 123 Eckhart, L., Lippens, S., Tschachler, E. & Declercq, W. Cell death by cornification - PubMed. *Biochimica et biophysica acta* **1833** (2013). <https://doi.org/10.1016/j.bbamcr.2013.06.010>
- 124 Fuchs, E. & Green, H. Changes in keratin gene expression during terminal differentiation of the keratinocyte. *Cell* **19**, 1033-1042 (1980). [https://doi.org/10.1016/0092-8674\(80\)90094-x](https://doi.org/10.1016/0092-8674(80)90094-x)
- 125 Giguere, V. & Evans, R. M. Chronicle of a discovery: the retinoic acid receptor. *J Mol Endocrinol* **69**, T1-T11 (2022). <https://doi.org/10.1530/JME-22-0117>
- 126 Darmon, M. & Blumenberg, M. L. *Molecular biology of the skin : the keratinocyte*. (Academic Press, 1993).
- 127 Petkovich, M. & Chambon, P. Retinoic acid receptors at 35 years. *J Mol Endocrinol* **69**, T13-T24 (2022). <https://doi.org/10.1530/JME-22-0097>
- 128 Cunningham, T. J. & Duester, G. Mechanisms of retinoic acid signalling and its roles in organ and limb development. *Nat Rev Mol Cell Biol* **16**, 110-123 (2015). <https://doi.org/10.1038/nrm3932>
- 129 Niederreither, K., Subbarayan, V., Dolle, P. & Chambon, P. Embryonic retinoic acid synthesis is essential for early mouse post-implantation development. *Nat Genet* **21**, 444-448 (1999). <https://doi.org/10.1038/7788>
- 130 Beckenbach, L., Baron, J. M., Merk, H. F., Löffler, H. & Amann, P. M. Retinoid treatment of skin diseases. *Eur J Dermatol* **25**, 384-391 (2015). <https://doi.org/10.1684/ejd.2015.2544>
- 131 Fuchs, E. & Byrne, C. The epidermis: rising to the surface. *Current Opinion in Genetics & Development* **4** (1994). [https://doi.org/10.1016/0959-437X\(94\)90140-X](https://doi.org/10.1016/0959-437X(94)90140-X)
- 132 Logan, W. S. Vitamin A and keratinization. *Arch Dermatol* **105**, 748-753 (1972).
- 133 O'Connor, C., Varshosaz, P. & Moise, A. R. Mechanisms of Feedback Regulation of Vitamin A Metabolism. *Nutrients* **14** (2022). <https://doi.org/10.3390/nu14061312>
- 134 Germain, P. *et al.* Differential Action on Coregulator Interaction Defines Inverse Retinoid Agonists and Neutral Antagonists. *Chemistry & Biology* **16**, 479-489 (2009). <https://doi.org/10.1016/j.chembiol.2009.03.008>
- 135 Abashev, T. M., Metzler, M. A., Wright, D. M. & Sandell, L. L. Retinoic Acid signaling regulates Krt5 and Krt14 independently of stem cell markers in submandibular salivary gland epithelium. *Developmental dynamics : an official publication of the American Association of Anatomists* **246** (2017). <https://doi.org/10.1002/dvdy.24476>
- 136 Stellmach, V. *et al.* Retinoid-mediated transcriptional regulation of keratin genes in human epidermal and squamous cell carcinoma cells. *Proceedings of the National Academy of Sciences* **88** (1991). <https://doi.org/10.1073/pnas.88.11.4582>

- 137 Géhin, M. *et al.* Structural basis for engineering of retinoic acid receptor isotype-selective agonists and antagonists. *Chemistry & Biology* **6**, 519-529 (1999). [https://doi.org/10.1016/S1074-5521\(99\)80084-2](https://doi.org/10.1016/S1074-5521(99)80084-2)
- 138 Hughes, N. E. *et al.* Identification of potent and selective retinoic acid receptor gamma (RAR γ) antagonists for the treatment of osteoarthritis pain using structure based drug design. *Bioorg Med Chem Lett* **26**, 3274-3277 (2016). <https://doi.org/10.1016/j.bmcl.2016.05.056>
- 139 Li, Y., Hashimoto, Y., Agadir, A., Kagechika, H. & Zhang, X. k. Identification of a novel class of retinoic acid receptor beta-selective retinoid antagonists and their inhibitory effects on AP-1 activity and retinoic acid-induced apoptosis in human breast cancer cells. *J Biol Chem* **274**, 15360-15366 (1999). <https://doi.org/10.1074/jbc.274.22.15360>
- 140 Pfefferle, A. D. *et al.* Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biol* **14**, R125 (2013). <https://doi.org/10.1186/gb-2013-14-11-r125>
- 141 Petrosyan, V. *et al.* Immunologically "cold" triple negative breast cancers engraft at a higher rate in patient derived xenografts. *NPJ Breast Cancer* **8**, 104 (2022). <https://doi.org/10.1038/s41523-022-00476-0>
- 142 Kopan, R. & Fuchs, E. The use of retinoic acid to probe the relation between hyperproliferation-associated keratins and cell proliferation in normal and malignant epidermal cells. *J Cell Biol* **109**, 295-307 (1989). <https://doi.org/10.1083/jcb.109.1.295>
- 143 Tierney, M. T. *et al.* Vitamin A resolves lineage plasticity to orchestrate stem cell lineage choices. *Science* **383**, eadi7342 (2024). <https://doi.org/10.1126/science.adi7342>
- 144 Fuchs, E. & Green, H. Regulation of terminal differentiation of cultured human keratinocytes by vitamin A. *Cell* **25**, 617-625 (1981). [https://doi.org/10.1016/0092-8674\(81\)90169-0](https://doi.org/10.1016/0092-8674(81)90169-0)
- 145 Segre, J. A., Bauer, C. & Fuchs, E. Klf4 is a transcription factor required for establishing the barrier function of the skin. *Nat Genet* **22**, 356-360 (1999). <https://doi.org/10.1038/11926>
- 146 Jiang, W., Lobo-Ruppert, S. M. & Ruppert, J. M. Bad things happen in the basal layer: KLF4 and squamous cell carcinoma. *Cancer Biol Ther* **7**, 783-785 (2008). <https://doi.org/10.4161/cbt.7.5.6110>
- 147 Dai, X. & Segre, J. A. Transcriptional control of epidermal specification and differentiation. *Current Opinion in Genetics & Development* **14**, 485-491 (2004). <https://doi.org/10.1016/j.gde.2004.07.002>
- 148 Fortunel, N. O. *et al.* KLF4 inhibition promotes the expansion of keratinocyte precursors from adult human skin and of embryonic-stem-cell-derived keratinocytes. *Nat Biomed Eng* **3**, 985-997 (2019). <https://doi.org/10.1038/s41551-019-0464-6>
- 149 Yamamoto, A. *et al.* Metastasis from the tumor interior and necrotic core formation are regulated by breast cancer-derived angiopoietin-like 7. *Proc Natl Acad Sci U S A* **120**, e2214888120 (2023). <https://doi.org/10.1073/pnas.2214888120>

- 150 Sommer, A. Xerophthalmia and vitamin a status. *Progress in Retinal and Eye Research* **17** (1998). [https://doi.org/10.1016/S1350-9462\(97\)00001-3](https://doi.org/10.1016/S1350-9462(97)00001-3)
- 151 Korde, L. A., Doody, D. R., Hsu, L., Porter, P. L. & Malone, K. E. Bisphosphonate Use and Risk of Recurrence, Second Primary Breast Cancer, and Breast Cancer Mortality in a Population-Based Cohort of Breast Cancer Patients. *Cancer Epidemiology, Biomarkers & Prevention* **27** (2018). <https://doi.org/10.1158/1055-9965.EPI-17-0556>
- 152 Li, C. I. et al. Epidemiologic and molecular risk factors for contralateral breast cancer among young women. *British Journal of Cancer* **2003** 89:3 **89** (2003). <https://doi.org/10.1038/sj.bjc.6601042>
- 153 Li, C. I. et al. Reproductive and anthropometric factors in relation to the risk of lobular and ductal breast carcinoma among women 65-79 years of age. *Int J Cancer* **107**, 647-651 (2003). <https://doi.org/10.1002/ijc.11465>
- 154 Li, C. I. et al. The relationship between alcohol use and risk of breast cancer by histology and hormone receptor status among women 65-79 years of age. *Cancer Epidemiol Biomarkers Prev* **12**, 1061-1066 (2003).
- 155 Kushi, L. H., Kwan, M. L., Lee, M. M. & Ambrosone, C. B. Lifestyle factors and survival in women with breast cancer. *J Nutr* **137**, 236S-242S (2007). <https://doi.org/10.1093/jn/137.1.236S>
- 156 Clark, A. G. & Vignjevic, D. M. Modes of cancer cell invasion and the role of the microenvironment. *Curr Opin Cell Biol* **36**, 13-22 (2015). <https://doi.org/10.1016/j.ceb.2015.06.004>
- 157 Ilina, O. et al. Cell-cell adhesion and 3D matrix confinement determine jamming transitions in breast cancer invasion. *Nat Cell Biol* **22**, 1103-1115 (2020). <https://doi.org/10.1038/s41556-020-0552-6>
- 158 Luond, F. et al. Distinct contributions of partial and full EMT to breast cancer malignancy. *Dev Cell* **56**, 3203-3221 e3211 (2021). <https://doi.org/10.1016/j.devcel.2021.11.006>
- 159 Yoon, S. B. et al. Subpopulation commensalism promotes Rac1-dependent invasion of single cells via laminin-332. *J Cell Biol* **223** (2024). <https://doi.org/10.1083/jcb.202308080>
- 160 Zanutelli, M. R., Zhang, J. & Reinhart-King, C. A. Mechanoresponsive metabolism in cancer cell migration and metastasis. *Cell Metab* **33**, 1307-1321 (2021). <https://doi.org/10.1016/j.cmet.2021.04.002>
- 161 Lambert, A. W., Zhang, Y. & Weinberg, R. A. Cell-intrinsic and microenvironmental determinants of metastatic colonization. *Nat Cell Biol* **26**, 687-697 (2024). <https://doi.org/10.1038/s41556-024-01409-8>
- 162 Simeonov, K. P. et al. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell* **39**, 1150-1162 e1159 (2021). <https://doi.org/10.1016/j.ccell.2021.05.005>
- 163 Park, S. et al. Tissue-scale coordination of cellular behaviour promotes epidermal wound repair in live mice. *Nat Cell Biol* **19**, 155-163 (2017). <https://doi.org/10.1038/ncb3472>

- 164 Kumar, N. *et al.* The unique risk factor profile of triple negative breast cancer: a comprehensive meta-analysis. *J Natl Cancer Inst* (2024).
<https://doi.org/10.1093/jnci/djae056>
- 165 Foulkes, W. D., Smith, I. E. & Reis-Filho, J. S. Triple-negative breast cancer. *N Engl J Med* **363**, 1938-1948 (2010). <https://doi.org/10.1056/NEJMra1001389>
- 166 Greenlee, H. *et al.* Changes in vitamin and mineral supplement use after breast cancer diagnosis in the Pathways Study: a prospective cohort study. *BMC Cancer* **2014** 14:1 **14** (2014). <https://doi.org/10.1186/1471-2407-14-382>
- 167 Velicer, C. M. & Ulrich, C. M. Vitamin and Mineral Supplement Use Among US Adults After Cancer Diagnosis: A Systematic Review. *JCO* **26** (2016).
<https://doi.org/10.1200/JCO.2007.13.5905>
- 168 Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979-993 (2012). <https://doi.org/10.1016/j.cell.2012.04.024>
- 169 Koh, G., Degasperi, A., Zou, X., Momen, S. & Nik-Zainal, S. Mutational signatures: emerging concepts, caveats and clinical applications. *Nat Rev Cancer* **21**, 619-637 (2021). <https://doi.org/10.1038/s41568-021-00377-7>
- 170 Li, C. I. *et al.* Cancer Cell Intrinsic and Immunologic Phenotypes Determine Clinical Outcomes in Basal-like Breast Cancer. *Clin Cancer Res* **27**, 3079-3093 (2021).
<https://doi.org/10.1158/1078-0432.CCR-20-3890>
- 171 Bertucci, F. *et al.* Genomic characterization of metastatic breast cancers. *Nature* **569**, 560-564 (2019). <https://doi.org/10.1038/s41586-019-1056-z>
- 172 Stirling, D. R. *et al.* CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 433 (2021). <https://doi.org/10.1186/s12859-021-04344-9>
- 173 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021). <https://doi.org/10.1016/j.cell.2021.04.048>
- 174 Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847-2849 (2016).
<https://doi.org/10.1093/bioinformatics/btw313>
- 175 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21 (2013). <https://doi.org/10.1093/bioinformatics/bts635>
- 176 Law, C. W. *et al.* RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res* **5** (2016). <https://doi.org/10.12688/f1000research.9005.3>
- 177 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2010). <https://doi.org/10.1093/bioinformatics/btp616>
- 178 Viragova, S. *et al.* Inverse agonists of retinoic acid receptor/retinoid X receptor signaling as lineage-specific antitumor agents against human adenoid cystic carcinoma. *J Natl Cancer Inst* **115**, 838-852 (2023).
<https://doi.org/10.1093/jnci/djad062>
- 179 Stata Statistical Software v. Release 18 (StataCorp LLC, 2023).
- 180 Ross, T. R. *et al.* The HMO Research Network Virtual Data Warehouse: A Public Data Model to Support Collaboration. *EGEMS (Wash DC)* **2**, 1049 (2014).
<https://doi.org/10.13063/2327-9214.1049>

- 181 Pal, B. *et al.* Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nat Commun* **8**, 1627 (2017). <https://doi.org/10.1038/s41467-017-01560-x>
- 182 Manzoni, C. *et al.* Genome, transcriptome and proteome: the rise of omics data and their integration in biomedical sciences. *Brief Bioinform* **19**, 286-302 (2018). <https://doi.org/10.1093/bib/bbw114>
- 183 Meurette, O. & Mehlen, P. Notch Signaling in the Tumor Microenvironment. *Cancer Cell* **34**, 536-548 (2018). <https://doi.org/10.1016/j.ccell.2018.07.009>
- 184 Su, H. & Karin, M. Multifaceted collagen-DDR1 signaling in cancer. *Trends Cell Biol* **34**, 406-415 (2024). <https://doi.org/10.1016/j.tcb.2023.08.003>
- 185 Nieto, M. A. Epithelial plasticity: a common theme in embryonic and cancer cells. *Science* **342**, 1234850 (2013). <https://doi.org/10.1126/science.1234850>
- 186 Mikkola, M. L. & Millar, S. E. The mammary bud as a skin appendage: unique and shared aspects of development. *J Mammary Gland Biol Neoplasia* **11**, 187-203 (2006). <https://doi.org/10.1007/s10911-006-9029-x>
- 187 Iyer, S. V. *et al.* Understanding the role of keratins 8 and 18 in neoplastic potential of breast cancer derived cell lines. *PLoS One* **8**, e53532 (2013). <https://doi.org/10.1371/journal.pone.0053532>
- 188 Wang, F., Chen, S., Liu, H. B., Parent, C. A. & Coulombe, P. A. Keratin 6 regulates collective keratinocyte migration by altering cell-cell and cell-matrix adhesion. *J Cell Biol* **217**, 4314-4330 (2018). <https://doi.org/10.1083/jcb.201712130>
- 189 Yoon, S. & Leube, R. E. Keratin intermediate filaments: intermediaries of epithelial cell migration. *Essays Biochem* **63**, 521-533 (2019). <https://doi.org/10.1042/EBC20190017>
- 190 Huang, M. E. *et al.* Use of all-trans retinoic acid in the treatment of acute promyelocytic leukemia. *Blood* **72**, 567-572 (1988).
- 191 Reynolds, C. P., Matthay, K. K., Villablanca, J. G. & Maurer, B. J. Retinoid therapy of high-risk neuroblastoma. *Cancer Letters* **197**, 185-192 (2003). [https://doi.org/10.1016/S0304-3835\(03\)00108-3](https://doi.org/10.1016/S0304-3835(03)00108-3)
- 192 Kean, S. Why Don't We Have Male Birth Control Yet? **Health & Medicine** (2021). <<https://www.sciencehistory.org/stories/disappearing-pod/why-dont-we-have-a-male-birth-control-pill-yet/>>.
- 193 Heller, C. G., Moore, D. J. & Paulsen, C. A. Suppression of spermatogenesis and chronic toxicity in men by a new series of bis(dichloroacetyl) diamines. *Toxicol Appl Pharmacol* **3**, 1-11 (1961). [https://doi.org/10.1016/0041-008x\(61\)90002-3](https://doi.org/10.1016/0041-008x(61)90002-3)
- 194 Paik, J. *et al.* Inhibition of retinoic acid biosynthesis by the bisdichloroacetyldiamine WIN 18,446 markedly suppresses spermatogenesis and alters retinoid metabolism in mice. *J Biol Chem* **289**, 15104-15117 (2014). <https://doi.org/10.1074/jbc.M113.540211>
- 195 Centritto, F. *et al.* Cellular and molecular determinants of all-trans retinoic acid sensitivity in breast cancer: Luminal phenotype and RARalpha expression. *EMBO Mol Med* **7**, 950-972 (2015). <https://doi.org/10.15252/emmm.201404670>
- 196 Caricasulo, M. A., Zanetti, A., Terao, M., Garattini, E. & Paroni, G. Cellular and micro-environmental responses influencing the antitumor activity of all-trans retinoic acid

- in breast cancer. *Cell Commun Signal* **22**, 127 (2024).
<https://doi.org/10.1186/s12964-024-01492-2>
- 197 Garattini, E. *et al.* Retinoids and breast cancer: from basic studies to the clinic and back again. *Cancer Treat Rev* **40**, 739-749 (2014).
<https://doi.org/10.1016/j.ctrv.2014.01.001>
- 198 Lawson, N. D. & Berliner, N. Neutrophil maturation and the role of retinoic acid. *Exp Hematol* **27**, 1355-1367 (1999). [https://doi.org/10.1016/s0301-472x\(99\)00085-5](https://doi.org/10.1016/s0301-472x(99)00085-5)
- 199 Busada, J. T. & Geyer, C. B. The Role of Retinoic Acid (RA) in Spermatogonial Differentiation. *Biol Reprod* **94**, 10 (2016).
<https://doi.org/10.1095/biolreprod.115.135145>
- 200 Harrison, E. H. & Hussain, M. M. Mechanisms involved in the intestinal digestion and absorption of dietary vitamin A. *J Nutr* **131**, 1405-1408 (2001).
<https://doi.org/10.1093/jn/131.5.1405>
- 201 Mayne, S. T., Playdon, M. C. & Rock, C. L. Diet, nutrition, and cancer: past, present and future. *Nat Rev Clin Oncol* **13**, 504-515 (2016).
<https://doi.org/10.1038/nrclinonc.2016.24>
- 202 Connor, J. Alcohol consumption as a cause of cancer. *Addiction* **112**, 222-228 (2017). <https://doi.org/10.1111/add.13477>
- 203 Wolin, K. Y., Carson, K. & Colditz, G. A. Obesity and cancer. *Oncologist* **15**, 556-565 (2010). <https://doi.org/10.1634/theoncologist.2009-0285>
- 204 Weber, D. D. *et al.* Ketogenic diet in the treatment of cancer - Where do we stand? *Mol Metab* **33**, 102-121 (2020). <https://doi.org/10.1016/j.molmet.2019.06.026>
- 205 Goodman, M., Bostick, R. M., Kucuk, O. & Jones, D. P. Clinical trials of antioxidants as cancer prevention agents: past, present, and future. *Free Radic Biol Med* **51**, 1068-1084 (2011). <https://doi.org/10.1016/j.freeradbiomed.2011.05.018>