Spring 2023

UNDERSTANDING THE IMPLICATIONS OF LINEAGE PLASTICITY IN BREAST CANCER EVOLUTION AND CHEMOTHERAPY RESPONSE

Gadisti Aisha Mohamed
fnu.gadisti.aisha.nurulhijjah.binti.mohamed.gr@dartmouth.edu

UNDERSTANDING THE IMPLICATIONS OF LINEAGE PLASTICITY IN BREAST CANCER EVOLUTION AND CHEMOTHERAPY RESPONSE

A Thesis

Submitted to the Faculty

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Molecular and Systems Biology

By Gadisti Aisha Nurulhijjah Binti Mohamed

Guarini School of Graduate and Advanced Studies

Dartmouth College

Hanover, New Hampshire

January 2023

Examining Committee:

Dr Diwakar R. Pattabiraman (chair)

Dr Yashi Ahmed

Dr Steven N. Fiering

Dr Wenjun Guo

F. Jon Kull, Ph.D.

Dean of the Guarini School of Graduate and Advanced Studies

Abstract

Intra-tumoral heterogeneity and the presence of a phenotypically diverse cell population within a single tumor represents a major hurdle in the understanding of tumor progression and dynamics, and complicates the effective diagnosis and management of this disease. One of the ways by which tumors gain intra-tumoral variation is through the acquisition of phenotypic or lineage plasticity, whereby tumor cells evolve away from the lineage of origin and gain altered profiles. These alterations may impart specific survival benefits to different subpopulations of cells, enabling them to proliferate faster, migrate away from the site of the primary tumor or evade drug-induced elimination, amongst others. Phenotypic plasticity and alterations of transcriptional profiles can be driven by either extrinsic signals or intrinsic cell autonomous mechanisms. Work presented in this thesis across three chapters has uncovered several molecular drivers altering cell-state plasticity in breast cancer, and their resulting effects on tumor development and progression. Lineage plasticity, driven by the transcription factor SOX10, allows breast tumors of the luminal lineage expressing lower Estrogen Receptor (ER) levels to gain basal-like characteristics, resulting in the evolution of these luminal-like tumors into a more basal-like subtype. Activation of Protein Kinase A (PKA) curtails cellular plasticity in a mouse mammary tumor model, preventing epithelial-mesenchymal transition (EMT) and metastasis, ultimately improving prognosis and survival. Eribulin treatment induces transcriptional reprogramming of breast tumor cell lines, forcing them to undergo mesenchymal-epithelial transition (MET). Together, these findings help to elucidate how cellular plasticity contributes to intra-tumoral heterogeneity of breast tumors, and how phenotypic diversity influences the progression, metastasis, and chemotherapy response of breast cancer. While these results have identified specific agents that act to promote phenotypic plasticity, the exact mechanisms by which they act, and the steps necessary for lineage evolution to occur are only partially understood. This work provides a foundation for further inquiry into the mechanisms driving phenotypic plasticity and resulting tumor heterogeneity, with the ultimate goal of developing better strategies to overcome this disease.

Acknowledgements

They say that it takes a village to raise a child, but I believe it also takes one to complete a PhD. I certainly would not have been able to succeed on this journey without the help and support of my very own village. First, to my supervisor, Dr Diwakar Pattabiraman, thank you for your guidance and leadership, for helping me to develop the necessary skills to be a good scientist and to view scientific questions in a different light. To Meredith, thank you for letting me vent my frustrations occasionally, and for being my lab partner through these last 4 years as we attempt to navigate the life of a PhD student while being part of a brand-new lab. To Nena, thank you for being the lab mother, for always being on top of things and making sure we thrive as a lab and as individuals. Life in the early days of the lab would not have been the same without you. To the rest of the Raman Noodles, Meisam and Hanxu, thank you for sticking out these last few months with me. Our time together as a lab may have been brief, but we had a good run!

Next, to my family. Separated by 15,000 kilometers, but always close to my heart. Abi and Umi, thank you for keeping me updated on everything regardless of my absence, and for all your emotional, spiritual, and material support. Angah, Hafiz, Fatim, Muhammad, Umar, Humairah, and Adam, thank you for sharing memes and pet updates, for the virtual movie watch parties, and for always thinking of me whenever you eat nice food. I love you all. To my extended family, the Jadis and the Sunardis, thank you for remembering me, for keeping me in your prayers, and for all the lively, chaotic, virtual family gatherings over zoom. I really miss our large family get-togethers.

Where would a village be without friends? Well, I am fortunate to have so many of them across several continents. Nani, Aida, Khairi, Rabbany, Izzati, thank you for always making time to catch up with me whenever I am back home, for your care and concern, and for entertaining this village idiot whenever I send ridiculous texts or DMs at weird hours (due to the time difference). To Soefie and the High Park Ave Crew, thank you for welcoming me into your home, for sharing your little family with me, and for making Toronto feel

like my kampung that I need to go back to for Hari Raya. To my roommates, Cristina, Varahi, Cameron, Depy, Dezi, and Jenn, thank you for feeding and entertaining me, especially during the pandemic. The harsh New England winters would have been too cold and miserable for this tropical city girl, if not for your friendship and companionship. To my Europe-based girls, Jasmin, Margarida, Katrina, and Almas, I miss you all dearly.

One of the greatest loves of my life is football (or, as they call it in this country, soccer). It brings me on such rollercoaster emotional journeys, day in and day out, and I wouldn't have it any other way. I would like to thank the Liverpool FC squad of 2018/2019 and 2019/2020 for realizing our dream of hoisting the UEFA Champions League and the English Premiere League trophies respectively. Jordan Henderson, my skipper. Jurgen Klopp, my gaffer. Thank you for bringing some light into my life in the dark days of the early pandemic, and for giving me something to look forward to every season as we keep on winning trophies (for the most part). As I write this thesis, I have also been captivated by the resilience, determination, and faith of the Moroccan National Team at the Qatar 2022 FIFA World Cup. My heart filled with pride as you dispatched big teams like Belgium, Spain, and Portugal, while continuously smashing the glass ceiling along the way. Your romp towards the semi-finals has given me so much joy and provided a much-needed distraction and temporary respite from the challenges of preparing this dissertation.

Finally, I would like to express my endless gratitude and appreciation to all the teachers and educational institutions that I have had the honor of learning from throughout my life, especially the two Madrasahs that have shaped me into the person that I am today. To the educators of Madrasah Al-Sagoff Al-Arabiah and Madrasah Al-Irsyad Al-Islamiah (now Madrasah Irsyad Zuhri Al-Islamiah), I would not be who I am today without the foundational knowledge and values that you have instilled upon me. Thank you for inspiring me to be better, for showing me how wide and vast the world is, and for reinforcing and strengthening my roots even as my leaves and branches spread out towards the glory of the sun. Rabbi zidni ilma. My Lord, advance me in knowledge.

Table of contents

List of tables

List of illustrations

Citations to previously published work

Chapter 2 is adapted from "Lineage plasticity enables low-ER luminal tumors to evolve and gain basal-like traits", Mohamed *et al* 2023, Under review at BMC Breast Cancer Research and available on Biorxiv
(https://www.biorxiv.org/content/10.1101/2022.11.25.517090v1)
Authors: Gadisti Aisha Mohamed[1,2,3,4], Sundis Mahmood[4], Min Kyung Lee[4], Nevena B. Ognjenovich[3], Owen M. Wilkins[4], Brock C. Christensen[5], Kristen E. Muller[5], Diwakar R. Pattabiraman[2,5]

Chapter 3 is adapted from "Limiting Self-Renewal of the Basal Compartment by PKA Activation Induces Differentiation and Alters the Evolution of Mammary Tumors", Ognjenovic *et al* 2020, available on Developmental Cell. 2020 Dec 7;55(5):544-557.e6.
(https://www.sciencedirect.com/science/article/pii/S1534580720307942)
Authors: Nevena B. Ognjenovic[3], Meisam Bagheri[2,3,4], Gadisti Aisha Mohamed[2,3,4], Ke Xu[4], Youdinghuan Chen[4], Mohamed Ashick Mohamed Saleem[4], Meredith S. Brown[3], Shivashankar H. Nagaraj[5], Kristen E. Muller[5], Scott A. Gerber[5], Brock C. Christensen[5], and Diwakar R. Pattabiraman[1,2,4,5]

Author contributions:
[1] Wrote manuscript; [2] Designed experiments; [3] Performed experiments; [4] Analyzed results; [5] Advised on subject material and results.

Chapter 1: Introduction

1.1 Overview of the problem

Tumor heterogeneity represents a significant hurdle in the diagnosis and management of cancer. Heterogeneous clonal populations within a single tumor may exhibit variations in gene expression, phenotypic features, and harbor different oncogenic mutations[1], providing them with unique tumorigenic and metastatic potential. Additionally, diverse cell populations may respond differently to various drug treatments, thus affecting clinical outcomes and contributing to chemotherapy resistance and tumor relapse[2,3]. While genetic mutations are known to contribute significantly to the emergence of a heterogeneous tumor[4,5], non-genetic factors have also been shown to play a role in intra-tumoral diversity[6,7]. Cellular plasticity, in which cancer cells alter their phenotypic identity by transdifferentiation and altered gene expression, has been implicated as a possible driver of non-genetic heterogeneity[8]. Understanding how cellular plasticity gives rise to heterogeneous tumors will provide better insights into how heterogeneous tumors develop, metastasize, and respond to treatment.

1.2 Tumor heterogeneity

One of the challenges in the management of cancer as a disease is its heterogeneity. Unlike other diseases such as Type I diabetes, which can be managed by the administration of insulin[9], or anaphylactic reactions which can be alleviated by epinephrine injections[10], there is no single cure or treatment that will be effective in eliminating all forms of cancers. Differences between and within tumors ensure that the disease presents differently in different patients, thus necessitating the use of a wide variety of treatments in an attempt to control this disease[11]. This heterogeneity also results in differing rates of development and metastasis[12], further complicating cancer management. Tumor heterogeneity can be categorized into 2 levels, inter-tumoral heterogeneity and intra-tumoral heterogeneity.

1.2.1 Inter-tumoral heterogeneity

Inter-tumoral heterogeneity refers to differences between tumors in different patients, or even between primary and secondary tumors. Tumors originating in the same organ can appear morphologically or histologically distinct, and even express different genes and biomarkers. In an effort to reduce the complexity of diagnosis and guide treatment decisions, tumors can be grouped into several different classifications or subtypes. The most common method of classifying solid tumors is the TNM Classification. This method uses information regarding the size of the tumor (T), whether the cancer has spread to the nearby lymph nodes (N), and whether it has metastasized to a distant site (M), in order to determine the stage of the disease in the patient. The TNM classification assigns each tumor into one of 5 stages, with the lowest stage, stage 0, indicating carcinoma-in-situ, which is a pre-cancerous lesion, while stage IV indicates metastatic cancer which has spread to at least 4 lymph nodes and to at least one secondary site[13]. This classification is commonly used in multiple carcinomas, such as oral, breast, gastric, and lung, and is useful for determining the course of treatment, and predicting outcomes[14].

An alternative to the TNM classification is the Barcelona Clinic Liver Cancer (BCLC) staging, which is specific for tumors originating in the liver. This classification system uses information regarding the size of the tumor, general health of the patient, and liver function, to assign the patient tumor into one of 5 stages. Stage 0 refers to a small tumor of less than 2cm and normal liver function, while stage D refers to a large tumor with sever liver impairment. Specific treatments are recommended for each stage of this disease, with tumor resection recommended for stage 0 patients, and the kinase inhibitor sorafenib recommended for stage C patients[15].

Besides using tumor stage to classify malignancies, tumors can also be stratified into subtypes based on gene expression. This method of classification uses large gene expression datasets from thousands of patient tumors to group tumors based on similarities in their gene expression profiles[16–18]. For example, glioblastoma, an aggressive

brain and spinal cord cancer, can be subdivided into 4 intrinsic subtypes, classical, mesenchymal, proneural, and neural. The classical subtype is associated with amplification of *EGFR*, the mesenchymal subtype showed reduced *NF1* and increased mesenchymal marker expression, the proneural subtype displayed high *PDGRFA* expression with mutations in *IDH1*, and the neural subtype typically showed high expression of neural markers such as *NEFL* and *GABRA1*[18].

1.2.2 Intra-tumoral heterogeneity

Variations of different clonal subpopulations within a single tumor is referred to as intra-tumoral heterogeneity. Each subpopulation within a heterogeneous tumor displays distinct gene expression and phenotypic profiles, which confers unique properties to each subpopulation, allowing them to grow at different rates[19], respond differently to drug treatments[20], and preferentially travel to different organs as they metastasize[21].

Distinct cell populations within a heterogeneous tumor may contain unique genetic mutations. In renal cell carcinoma, up to 69% of somatic mutations associated with this disease were found to be unique, and not shared across all regions of the tumor. Unique inactivating mutations in various tumor suppressor genes such as *SETD2*, *PTEN*, and *KDM5C* were also found in separate regions of a single tumor[22]. In colorectal cancers, consistent c-Kras mutations and TP53 deletions were found in adenocarcinomas, but these mutations were only found in discrete regions of adenomas[23], suggesting that these tumors initially display some degree of genetic heterogeneity, but the continued proliferation of the c-Kras mutated subclone results in the progression of the disease to the carcinoma stage.

Alternatively, intra-tumoral heterogeneity may take the form of phenotypic variability. Unlike permanent genetic aberrations, this type of heterogeneity involves differences in gene expression profiles of different clonal subpopulations within the tumor, allowing cells to easily transition into different cell states or phenotypes[24]. Phenotypic

heterogeneity in the form of the epithelial-mesenchymal transition (EMT) is thought to contribute to metastasis[25], but EMT heterogeneity has also been shown to help determine organotropism or preference of different tumor cells to colonize different organs[21]. Phenotypic heterogeneity may also facilitate drug resistance, as tumor subpopulations express different genes that may predispose them to avoid drug-induced elimination[26].

Metabolic plasticity between tumor cells also allows different tumor populations to grow at different rates. Mathematical modeling of metabolic heterogeneity within solid tumors suggests that, as tumors grow, hypoxia within the center of the tumor mass changes the metabolism of the tumor cells, leading to glycolysis and acidification, while cells on the tumor periphery remain metabolically normal[19]. The switch from aerobic respiration to anaerobic glycolysis thus slows down the growth of cells within the central tumor mass. These cells also develop acid resistance and gain a more aggressive and invasive phenotype in an attempt to escape the hypoxic environment. This ultimately influences treatment, as the metabolically normal cells are more readily eliminated with cytotoxic agents, allowing the cells with the more aggressive phenotype to be selected for during the treatment process[19].

Heterogeneity within a single tumor may complicate tumor diagnosis and management. Biopsies obtained from a small area of the tumor may provide insights into tumor phenotypes in the local area of the biopsy but may not represent the complexity of the tumor as a whole[27]. Samples obtained from biopsies may also be used to identify certain biomarkers that may potentially be used as drug targets or predict outcomes of treatment, however the selected drug may be successful in targeting only a select population of tumor cells[27].

1.3 Breast cancer heterogeneity and subtypes

Breast cancer is a complex disease with significant inter-tumoral variation and intra-tumoral heterogeneity, necessitating that these tumors be diagnosed and managed in different ways. Due to inter-tumoral diversity, breast tumors can be categorized into different groups or subtypes[16], in order to better understand how individual tumors may be different from others. Classification of breast tumors into subtypes also helps to determine the best course of treatment and aid in other diagnostic and prognostic predictions[28–30].

Clinical classification of breast cancer is aided by the expression of specific surrogate protein markers within the tumor. These are referred to as histological or immunohistochemistry (IHC) subtypes, where IHC is commonly used to quantify the expression of estrogen receptor (ER), progesterone receptor (PR), *HER2/Neu*, and Ki67[31]. Based on the expression of these markers, breast tumors are classified into 4 subtypes: luminal A, luminal B, Her2 positive, and basal-like[32].

The luminal A subtype consist of tumors that are ER and PR positive, with low Ki67 expression, indicating lower proliferative potential, and negative for *HER2/neu*[31]. 50-60% of all breast cancers fall under this classification, making it the most common subtype[32]. Treatments recommended for this subtype is antiestrogen therapy such as tamoxifen, to target the ER expressing population[33]. This subtype also has the best prognosis, with tumors presenting as lower grade, patients having higher survival and lower relapse rate, and less than 10% of patients developing secondary metastasis[34].

The luminal B subtype makes up 15-20% of all breast cancers[32], and consist of tumors that are ER and PR positive, with higher Ki67 expression indicating better proliferative potential, and similar to luminal A, *Her2/neu* negative[31]. This subtype may express lower levels of ER and is more aggressive than the luminal A subtype. Tumors also present as higher grade, thus patients diagnosed with this subtype tend to have worse prognosis and

higher relapse and metastasis rates[34]. Typical treatments for the luminal B subtype also include the antiestrogen therapies, as the ER positive expression suggest that these tumors may benefit from inhibiting this receptor[35].

The Her2 positive subtype consists of tumors that have high expression of human epidermal growth factor receptor-2 (*HER2*), commonly due to amplification of this gene[32]. Additionally, 40% of tumors in this subtype harbor mutations in p53[36], and about half also express low levels of ER[32]. Her2 positive tumors make up 15-20% of breast cancers, and treatment mainly involves targeting Her2, with drugs such as Trastuzumab proving to be effective against this subtype[37]. These tumors are also more aggressive, with patients facing poor prognosis of only 12% 10 year survival rate[32].

The basal-like subtype, also known as triple-negative breast cancer (TNBC) make up the smallest percentage (8-10%) of all breast cancer. This subtype is the most aggressive, with high rate of brain and lung metastasis[38]. Tumors in this subtype do not express ER or PR, and do not display Her2/neu amplification, thus making them triple-negative breast cancers, with additional Krt5 expression, a basal-marker, determining the basal-like nature of this subtype[32]. Common treatments used for basal-like or TNBCs are neoadjuvant chemotherapeutics followed by surgical resection of the tumor and additional adjuvant chemotherapies to prevent tumor recurrence.

Besides histological based subtypes, molecular gene expression profiles have also been used to determine breast cancer classification. Instead of using surrogate marker expression to classify tumors, the molecular or intrinsic subtypes use a larger gene expression signature to group breast tumors into 4 main subtypes, basal-like (corresponding to the basal-like or TNBC IHC subtype), *Erb-B2+* (corresponding to the Her2 positive IHC subtype), normal-breast-like, and luminal epithelial/ER+[16]. Additionally, the luminal epithelial/ER+ subtype can also be further subdivided into luminal A and luminal B[17] (corresponding to the luminal A and luminal B IHC subtypes respectively).

Intrinsic subtypes are more commonly used to stratify breast cancers in non-clinical research settings, however a smaller signature of 50 genes, also called the PAM50 gene set, may be used by clinicians to classify tumors based on these intrinsic subtypes[39]. Clinical outcomes for these subtypes also mirror the IHC subtype, with the basal-like subtype exhibiting the worst prognosis, followed by the *Erb-B2+* subtype[17].

Even though subtype stratification may help to determine prognosis and potential drug targets, significant intra-tumoral heterogeneity also exist in breast tumors, resulting in potential misdiagnosis and emergence of drug resistance. Heterogeneity in ER expression, specifically in some luminal B tumors, suggest that tamoxifen treatment may not be effective in eliminating a majority of the clonal subpopulations within the tumor. As low as 1% ER expressing nuclei is sufficient to classify tumors as ER/PR+ luminal tumors[40], with limited evidence to suggest that antiestrogen therapy will be successful in managing tumors with such low ER expression[35]. Furthermore, ER staining of biopsies may not reflect the extent of ER heterogeneity within the tumor, potentially resulting in errors in estimating ER expression in the tumor.

Intra-tumoral heterogeneity in *HER2/neu* amplification has also been observed[41]. Heterogeneity in Her2 expression in different areas of a single tumor has been reported, with IHC staining discrepancy of 1-50% between different regions of the tumor [42]. Some tumors showing regional or spatial Her2 variability were found to also exhibit genetic heterogeneity[43], defined as *HER2/neu* amplification ratio of greater than 2.2 in 5-50% of tumor cells[44]. This heterogeneity has important clinical implications, as patients with high Her2 heterogeneity were found to have a shorter disease-free survival than patients with relatively homogeneous Her2 expression[43].

Intra-tumoral heterogeneity in TNBC may take the form of variability in gene expression, most notably in differing upregulation of basal-markers[45]. Genetic heterogeneity in TNBC involving variability in *EGFR* and *CCDN1* amplification has also been reported[46]. High intra-

tumoral heterogeneity in TNBC has been shown to correlate with increased metastasis and poor metastasis-free survival[46].

1.4 Causes and consequences of tumor heterogeneity

Considering the important role tumor heterogeneity plays in the diagnosis and treatment of cancer, it is important to understand both how intra-tumoral variability arises and how it influences disease progression. Tumor development is commonly viewed as an evolutionary process[47], where cells that gain increased proliferative capabilities are able to multiply faster than neighboring populations, allowing them to outcompete other cells for resources. This mirrors Darwinian selection and evolution, in which species which gain a survival advantage are selected for by the environment over time[48], thus leading to the existence of the diverse array of flora and fauna that we see today. Similarly, natural selection of different tumor cells that have gained some survival advantage, whether genetically or phenotypically, contributes to the diversity of clonal subpopulations observed in a heterogeneous tumor.

1.4.1 Genetic sources of heterogeneity

Genetic mutations have long been considered an important contributor to genetic diversity, both in Darwinian evolution and in intra-tumoral heterogeneity[49]. Cancer often begins with mutations in 'gatekeeper' genes, either activating a proto-oncogene or inactivating a tumor suppressor protein, which represents a crucial first step in oncogenic transformation[50]. Proliferation of the initial transformed cell would thus generate a clone of cells, all carrying the same 'gatekeeper' mutation. Additional mutations can then occur in these transformed cells, which then expand to form new sub-clones. Accumulation of different mutations in different sub-clones, and subsequent clonal expansion, thus results in multiple subpopulations that are genetically divergent from the initial transformed cell, leading to the formation of a diverse tumor cell population[49,51].

Genetic heterogeneity has been observed in multiple cancers, including brain, breast, colon, esophagus, prostate, bladder, and myeloma[27]. The accumulation of genetic aberrations has been well studied in the tumorigenesis of colorectal carcinoma (CRC), where specific mutations in the *APC*, *KRAS*, *SMAD4* and *TP53* genes occur to transform and progress tumors from benign adenomas to metastatic carcinomas, in a process referred to as the adenoma-carcinoma sequence[52]. While these specific mutations characterize this disease and are largely present in most tumor cells, intra-tumoral genetic heterogeneity in CRC has also been observed. Loss of heterozygosity (LOH) in the *APC* and *DCC* locus has been described, and has been attributed to the discordance in the *APC* and *DCC* expression in 67% and 58% of cases respectively[53]. Point mutations in *KRAS* and *TP53* genes have also been observed, leading to genetic and functional heterogeneity of these genes within the tumor[53].

Histological differences within a tumor have been suggested as an indicator of genetic heterogeneity within CRC tumors[54], however morphological heterogeneity itself may not be sufficient to predict the genetic variability within the tumor. Next-generation sequencing of both morphologically homogeneous and heterogeneous tumors show the presence of genetic variation in all tumors, regardless of morphological status[55], suggesting that even histologically homogeneous tumors may harbor different genetic mutations. Genetic heterogeneity also complicates the diagnosis of CRC, as examination of single tumor regions results in errors estimating *KRAS* and *BRAF* expression in up to 30% of cases[54], suggesting that using tissues sampled across several regions of the tumor may be more accurate in estimating the extent of genetic mutations[56]. Finally, response to treatment can also be affected by the presence of genetic variations within the tumor. Failure to respond to Cetuximab, an anti-EGFR therapy normally administered to patients with *KRAS* wild-type tumors, has been associated with the presence of tumor subpopulations carrying *KRAS* mutations[57], suggesting that small genetic variations within the tumor is sufficient to induce therapeutic failure, and identifying these mutations prior to treatment may be beneficial in predicting treatment response.

1.4.2 Non-genetic heterogeneity

Tumor heterogeneity can also be caused by other, non-genetic factors. This includes epigenetic and metabolic differences, and the influence of the extracellular microenvironment[58]. Non-genetic heterogeneity involves differences in tumor cell properties and phenotypes that are not determined by genetic mutations.

A well-known theory of how tumors gain non-genetic, phenotypic variability is by the activity of cancer stem cells (CSCs). CSCs are cells that have self-renewal capabilities, and are hypothesized to be responsible for tumor maintenance and progression, as they are able to differentiate into the various other cell phenotypes present within the tumor[27]. Studies that support the existence of CSCs suggest that tumors harbor only a small percentage of these CSCs, and the majority of the cells within the tumor are considered to have limited self-renewal potential. Transplantation of a phenotypically distinct, $CD34^+CD38^-$, subpopulation of human acute myeloid leukemia cells into mice successfully reconstitute the tumor in the animal model, while transplantation of other phenotypes failed, indicating that the tumor self-renewal capabilities are limited to $CD34^+CD38^-$ cells[59]. Similarly, the $CD44^+CD24^-$ subpopulation of breast cancer cells have been found to be enriched in tumor-repopulating cells, indicating that the tumor cells with this phenotype are self-renewing CSCs[60].

Since the majority of the tumor consists of non-repopulating cells, anti-cancer therapeutic agents are usually effective in eliminating the bulk of the tumor, while therapy resistant CSCs may survive treatment and cause tumor relapse[61]. Studies in glioblastoma multiform cells treated with temozolomide (TMZ), a DNA alkylating chemotherapeutic agent routinely used to treat this disease, showed that a small population of cells with CSC-like properties were able to repopulate the tumor after drug treatment by entering a quiescent state[62]. TMZ is able to induce cell-cycle arrest, thus targeting proliferating cells, and by entering a quiescent or dormant state, CSCs are able to escape TMZ-induced elimination, before differentiating into a more proliferative state once treatment has

ended[62]. CSCs have also been shown to escape drug-induced elimination by increasing invasive potential[63], expressing drug efflux transporter proteins[64], and improving DNA repair capabilities[65].

1.4.3 Phenotypic and lineage plasticity

Although the CSC concept may suggest that only a small proportion of tumor cells are capable of differentiation, evidence suggest that cells with tumor-repopulating capabilities are in fact more abundant within the tumor population[66]. It is thus more useful to consider most tumor cells as phenotypically plastic, in which different subpopulations of cells have different degrees of 'stemness' and differentiation capabilities[67]. Phenotypic variations within a heterogeneous tumor is thus due to variations in gene expression profiles of these tumor subpopulations, which are not static, and can change depending on extrinsic cues and intrinsic cell-autonomous mechanisms[27]. This constant shift in the transcriptional state of tumor cells allows them to change their characteristics and differentiate away from their original lineage, thus taking on new phenotypic roles within the tumor[68].

Unlike the permanent heritable genetic mutations that underlie genetic heterogeneity, phenotypic and lineage plasticity are more transient and are thus thought to be non-heritable in nature[27]. Gene expression changes are thought to be stochastic and unstable, resulting in daughter cells that may not fully recapitulate the parental phenotype. While this is largely true in that stochastic gene expression patterns may allow daughter cells to phenotypically drift away from the parental lineage, stable and heritable forms of phenotypic variation have also been described. Melanoma cells displaying transcriptional variability have been shown to reprogram into a more transcriptionally stable state upon drug treatment, resulting in the formation of a new drug-resistant subpopulation[26], suggesting that, while not necessarily infinitely heritable, variability in gene expression patterns that define specific phenotypes may be stably inherited across several generations.

An example of a well-studied, relatively stable, and phenotypically plastic process is the epithelial-mesenchymal transition (EMT). In this process, epithelial cells lose apical-basal polarity, cell-to-cell contacts, and extracellular matrix attachment due to the dissolution of tight and adherens junctions and hemidesmosomes, allowing cells to gain a mesenchymal, invasive phenotype[69]. Multiple genes have been identified to initiate this process, particularly the activation of transcription factors *SNAIL*, *SLUG*, *TWIST* and *ZEB*, resulting in the reduction of epithelial markers such as E-cadherin, Mucin-1, and cytokeratins, and an increase in mesenchymal genes such as N-cadherin, vimentin and smooth muscle actin[69,70]. EMT and the reverse process, mesenchymal-epithelial transition (MET) are important processes involved in tumor metastasis and chemotherapy resistance. Epithelial and mesenchymal heterogeneity within the tumor allows mesenchymal subpopulations of cells to escape the site of the primary tumor, and the transient and reversible nature of this process allows these invading cells to undergo an MET at a secondary site, thus permitting the colonization of distant tissues and organs. Mesenchymal-like cells are also thought to be more resistant to chemotherapy, as these cells tend to exhibit lower proliferative potential, thus preventing elimination by drugs that target fast-cycling cells[71].

1.4.4 Tumor heterogeneity and cancer therapy

One of the consequences of tumor phenotypic heterogeneity is variability in the way tumor subpopulations respond to therapy. The presence of a heterogeneous population of cells within a tumor inevitably means that anti-cancer drugs may be more effective against certain populations more than others, as mentioned previously. Variation in gene expression patterns also allow cells to escape therapy in different ways. Transcriptional variability and high expression of resistance markers *AXL*, *EGFR*, and *NGFR* in subpopulations of melanoma cells have been shown to be predictive of vemurafenib resistance[26]. These cells appear to undergo a drug induced phenotypic change, where reduction in *SOX10* activity, and activation of TEAD and AP-1 signaling pathways results

in cellular dedifferentiation and epigenetic reprogramming, thus giving rise to a new, stable, drug resistant cell state[26]. Drug-induced trans-differentiation and epigenetic reprogramming have also been shown to give rise to diversity in cisplatin-resistant cells, but only in phenotypically homogeneous cell populations, while resistance in phenotypically heterogeneous tumors appear to arise by drug-induced selection of a pre-existing resistant subpopulation[20]. Intra-tumoral heterogeneity thus dictates the response of the tumor to a drug, but is itself influenced by the drugs, where selection of pre-existing resistant populations may decrease, while drug-induced phenotypic reprogramming may increase, cellular diversity.

Drug treatments have also been shown to induce lineage plasticity in several tumor models. In prostate cancer, treatment with androgen receptor (AR)- targeted therapies have been shown to induce reprogramming of AR-positive prostate adenocarcinoma into an AR-negative, neuroendocrine-like state that is resistant to therapy[72]. Eribulin, a commonly used third-line therapy against breast cancer, have also been shown to alter the intrinsic subtype of breast tumors, where 33.3% of cases showed a different intrinsic subtype upon treatment and subsequent surgical resection as compared to the original subtype at diagnosis[73]. This again indicates the ability of therapeutic drugs to induce epigenetic reprogramming that may bring about phenotypic changes in tumors upon treatment.

The emergence of drug resistant populations in heterogeneous tumors suggest that a single drug may not be sufficient in eliminating the tumor. One possible strategy to overcome this is by using combinatorial therapies that target multiple subpopulations of cells within tumors at the same time[74]. While this may eliminate more tumor subpopulations, rare drug resistant cells may still survive treatment, along with any cells that have undergone drug-induced phenotypic alterations resulting in resistance. Another possible strategy is to avoid administering the highest tolerable dose, and instead vary the dose of treatment to include high and low dose periods[75]. This would prevent the

complete elimination of the drug sensitive population, keeping the drug resistant population in-check, while also avoiding the emergence of a drug-dependent resistant cell state. An emerging strategy to avoid therapeutic resistance in heterogeneous tumors is to target or reduce heterogeneity itself[76]. Pre-treatment of tumor cells lines with histone deacetylase (HDAC) inhibitors prevents epigenetic reprogramming of a transient 'drug-tolerant' population, thus preventing the emergence of drug resistance in these cells[77]. Reducing epithelial and mesenchymal heterogeneity may also be a viable option to avoid drug resistance and prevent metastasis, with the chemotherapeutic drug eribulin showing promise in this regard due to its ability to reduce EMT and promote MET in TNBC[78].

Chapter 2: Luminal-to-basal plasticity allows low-ER tumors to acquire basal-like characteristics

Adapted from "Lineage plasticity enables low-ER luminal tumors to evolve and gain basal-like traits", Mohamed *et al* 2023, Under review at BMC Breast Cancer Research and available on Biorxiv (https://www.biorxiv.org/content/10.1101/2022.11.25.517090v1)

Data for Figures 1, 2, 3A, and 4A-D was collected by Sundis Mahmood and Kristen Muller

Data analysis for Figures 5, 6 and 9E was performed by Min Kyung Lee and Gadisti Aisha Mohamed

Experiments for Figures 3B-G, 4E-G, 7, 8, 9, and 10 were performed by Gadisti Aisha Mohamed

Computational analysis in Figures 9 and 10 was performed by Owen Wilkins

2.1 Abstract

Stratifying breast cancer into specific molecular or histological subtypes aids in therapeutic decision-making and predicting outcomes[16], however, these subtypes may not be as distinct as previously thought. Patients with luminal-like, estrogen receptor (ER)-expressing tumors have a better prognosis than patients with more aggressive, triple-negative or basal-like tumors. There is, however, a subset of luminal-like tumors with lower ER expression, which exhibit more basal-like features. We have found that breast tumors expressing lower levels of ER, traditionally considered to be luminal-like, represent a distinct subset of breast cancer characterized by the emergence of basal-like features. Lineage tracing of low-ER tumors in the MMTV-PyMT mouse mammary tumor model revealed that basal marker-expressing cells arose from normal luminal epithelial cells, suggesting that luminal-to-basal plasticity is responsible for the evolution and emergence of basal-like characteristics. This plasticity allows tumor cells to gain a new lumino-basal phenotype, thus leading to intra-tumoral lumino-basal heterogeneity. Single-cell RNA sequencing and analyses revealed SOX10 as a potential driver for this plasticity, which is known among breast tumors to be almost exclusively expressed in Triple Negative Breast Cancer (TNBC) and was also found to be highly expressed in low-ER tumors. These findings suggest that basal-like tumors may evolve from the evolutionary progression of luminal tumors with low ER expression.

2.2 Introduction

Breast cancer is a complex disease with multiple different biologic subtypes which have clinical implications on tumor development, prognosis, and treatment[28–30]. While traditional surrogate markers can be used to classify breast tumors into hormone-receptor positive, *HER2* amplified, and triple-negative subtypes, advancements in gene expression profiling have helped refine subtype stratification. Gene expression analyses across a diverse range of human breast carcinomas classified these tumors into four intrinsic subtypes: basal-like, *Erb-B2+*, normal-breast-like, and luminal epithelial/ER+[16]. Further refinement of these subtypes based on a larger sample size revealed that the ER-positive luminal epithelial subtype could be further divided into 2 subgroups: luminal A and luminal B, with luminal A expressing higher ER levels than luminal B. These intrinsic subtypes differ in their clinical outcomes, with the basal-like subtype exhibiting the worst prognosis, followed by *Erb-B2+*[17]. A smaller signature of 50 genes, PAM50, may be used by clinicians to classify tumors based on these intrinsic subtypes[39].

In the clinical environment, immunohistochemistry (IHC) based determination of surrogate protein marker expression is utilized to classify breast carcinomas into four subtypes: 1) estrogen receptor (ER) and progesterone receptor (PR) positive, and Ki67 low, 2) ER, PR, and Ki67 high, 3) *HER2/neu* amplified, and 4) ER, PR, and *HER2* negative, or triple-negative breast cancer (TNBC)[31]. These IHC-based subtypes correspond to the intrinsic subtypes luminal A, luminal B, *Erb-B2+*, and basal-like[17], respectively, providing biologic and clinically significant information used to guide treatment decisions[31,40]. Typically, patients with ER negative tumors, TNBC, and *HER2/neu* amplified, benefit from non-hormone-based forms of therapy: adjuvant or neoadjuvant chemotherapy for TNBC tumors, with the addition of anti-HER2 targeted therapy (i.e. Trastuzumab) for *HER2/neu* amplified tumors. Tumors are considered ER-positive when demonstrating 1% or greater ER expression[40], and patients typically receive treatment with an antiestrogen agent (i.e Tamoxifen) or aromatase inhibitor (i.e. letrozole, anastrozole). This low cutoff for ER-positive determination results in a heterogeneous group of tumors considered luminal-

like, as tumors with less than 10% ER-positive cells may exhibit different characteristics by comparison to those with >10% ER-positivity. There are limited data on the overall benefit of endocrine therapies for patients with low level (1-10%) ER expression, but given the possible benefit, patients are eligible for endocrine treatment[35]. Some studies suggest the majority of breast cancers with low ER expression show molecular features that are more similar to ER-negative, basal-like tumors than to ER-positive, endocrine sensitive tumors[79]. It is therefore essential to understand the underlying biology of breast cancer with low ER expression, in order to recognize their prognostic significance and identify ideal treatment regimens.

The normal mammary epithelium consists of cells from two different lineages: a luminal lineage characterized by the expression of Keratin 8 (Krt8), with more committed cells expressing ER and PR, and a basal lineage expressing Keratin 5 (Krt5) and/or Keratin 14 (Krt14)[80,81]. These lineages are derived from a bipotent mammary stem cell (MaSC) progenitor in the embryonic stage, but are maintained postnatally by unipotent luminal and basal progenitors[82–85]. Despite this lineage restriction, several studies have revealed the potential for lineage plasticity in the adult mammary gland in non-homeostatic settings. For example, lineage plasticity of the luminal and basal compartment allows them to regain multipotency in the adult mammary gland with luminal-derived basal cells (LdBCs) emerging in response to hormone stimulation during pregnancy[86], and basal cells repopulating mammary epithelium in response to injury or luminal cell ablation[87]. In the neoplastic setting, the luminal lineage has been identified as the cell of origin for *BRCA1*-mutant basal-like breast cancers, suggesting its involvement in the development of TNBC-like tumors typically observed in these patients[88]. Moreover, *BRCA1* and p53 deletions in the mouse luminal compartment result in tumors resembling typical human basal-like tumors[89]. In addition, claudin-low breast tumors, a mesenchymal subset of TNBCs, may also be derived from the luminal lineage[90]. These findings point to lineage plasticity being a core feature in the process of mammary tumorigenesis whereby luminal tumor cells gain the ability to stray from their lineage-of-origin. The heterogeneity of ER expression

within luminal-like tumors provides a starting point to study subpopulations within ER-positive tumors that may be more prone to plasticity and the acquisition of basal-like traits.

In this chapter, my work uncovers that luminal tumors with low-ER expression represent a distinct subtype with a higher tendency to gain basal-like traits. These tumors arise from luminal cells undergoing luminal-to-basal plasticity, leading to the emergence of cells that exhibit a lumino-basal phenotype. This plasticity of luminal tumor cells and presence of lumino-basal heterogeneity within breast tumors likely plays a critical role in their overall aggressive traits, especially their ability to progress and gain metastatic propensity.

2.3 Materials and methods

Dartmouth-Hitchcock Medical Center pathology database search

The pathology database (Cerner Millennium) at Dartmouth-Hitchcock Medical Center was retrospectively searched from January 2012 through August 2020 to identify all invasive breast cancer cases with low ER expression. Low ER expression was defined as a sample displaying 1-10% of cancer cells with ER expression by immunohistochemistry (IHC), according to the American Society of Clinical Oncology (ASCO) – College of American Pathologists (CAP) 2020 guidelines[35]. Pathology reports were reviewed to include all primary invasive breast cancers with low ER expression, and H&E and IHC slides were reviewed by a breast pathologist (KM). Pathologic characteristics were recorded from pathology reports and slide review and included tumor histologic type, tumor size, tumor grade (Nottingham combined histologic grade/modified Scarff-Bloom-Richardson grade), presence of ductal carcinoma in-situ (DCIS), lymphovascular invasion, axillary lymph node status, and response to neoadjuvant therapy, when administered. For patients with a pathologic complete response after neoadjuvant therapy, tumor characteristics were assessed on the pre-treatment core biopsy. A tumor was considered to exhibit basal-like histologic features when all of the following were present: solid sheets of tumor cells with a syncytial growth pattern, high-grade, pleomorphic cytological features, abundant

mitotic activity, tumor circumscription with pushing borders, prominent intra-tumoral and/or peripheral lymphocytic infiltrates, and tumor necrosis. Patient clinical features including age at diagnosis, treatment regimens, and follow-up status were recorded from electronic medical records.

Determination of ER, PR, and *HER2/neu* expression

ER, PR, and *HER2/neu* were performed on diagnostic core needle biopsies in all cases. Biomarkers were repeated on a subsequent surgical specimen at the request of treating clinicians in a minority of cases (n = 10). Immunohistochemical assays for ER and PR were performed on paraffin-embedded tissue sections fixed in 10% neutral buffered formalin for 6-72 hours using the polymer system technique with appropriate controls. The assays were performed according to the manufacturer's instructions using Anti-ER (Cell Marque, 249R-15-ASR, clone: SP1) and Anti-PR (Biocare Medical, ACA424B, clone: 16) antibodies. ER and PR were qualified (positive or negative) and quantified (% of tumor cells staining) by breast pathologists by "eyeballing" IHC stained slides. In addition, we evaluated for ER staining intensity (weak, moderate, or strong). *HER2/neu* analysis was performed using dual-probe FISH (Abbott Laboratories, PathVysion HER-2 DNA probe kit) to assess for gene amplification and results were interpreted in accordance with the ASCO/CAP HER2 testing guidelines[91].

Animal studies

All animal experiment IACUC protocols were approved by the Dartmouth College Committee on Animal Care. MMTV-PyMT mice ((Tg(MMTV-PyVT)$^{634Mul/LellJ}$ mice on a C57Bl/6J background, strain #: 022974)[92], Krt5-CreER mice (B6N.129S6(Cg)-*Krt5*$^{tm1.1(cre/ERT2)Blh}$/J, strain #: 029155)[82], Krt8-CreER mice (Tg(Krt8-cre/ERT2)^{17Blpn}/J, strain #: 017947)[82], and Rosa26-mTmG reporter mice (B6.129(Cg)-Gt(ROSA)26Sor$^{tm4(ACTB-tdTomato,-EGFP)Luo}$/J, strain #: 007676)[93] were purchased from The Jackson Laboratory. For tamoxifen induced mammary epithelial labelling, tamoxifen (Sigma-Aldrich, T5648-1G) was prepared by dissolving in commercially available corn oil for 5 hours at 37°C to a final

concentration of 30mg/ml. Krt5-CreERT/Rosa26-mTmG /MMTV-PyMT mice were administered 150mg/kg (100μl of tamoxifen stock for a 20g mouse) of the tamoxifen stock 3 times per week at week 3, while Krt8-CreERT/Rosa26-mTmG /MMTV-PyMT mice were administered 150mg/kg of the tamoxifen stock, 3 times per week at weeks 5 and 6. Mice were euthanized and tumors were harvested once tumors reached a volume of 1.5cm^3, usually at weeks 20-25. For analysis of the normal mammary gland, mice were euthanized at 8 weeks or age matched to tumor bearing mice.

Mammary gland dissociation

Mouse mammary fat pads were harvested and processed to obtain single-cell suspensions using established protocols[94] that were slightly modified. Mammary fat pads were digested in a solution of DMEM (Corning, 10-013-CV) with Hyaluronidase (Fischer Scientific, ICN10074091) and Collagenase A (Sigma-Aldrich, 10103586001) for 2 hours at 37°C with gentle agitation using a rotator. Red blood cells were subsequently removed with an ammonium chloride lyse (8.02g NH_4Cl, 0.84g $NaHCO_3$, 0.37g EDTA in 1L of water), and samples were agitated with Trypsin (Corning, 25-053-CI) and Dispase (Stem Cell Technologies, 7913) + DNAse I (Sigma-Aldrich, DN25-100mg) for 1 minute each to further dissociate the cells. Finally, samples were filtered through a 40mm cell strainer (Corning, 431750) to obtain a single-cell suspension.

Mammary gland whole mount preparation and Carmine Alum staining

Whole mammary glands were spread on a glass slide and fixed with Carnoy's fixative (60% ethanol, 30% chloroform, 10% glacial acetic acid) overnight at RT. Fixed tissue was rehydrated by washing with decreasing ethanol concentrations (70%, 50%, 30%, 10%) 2 times each for 10 minutes. Rehydrated tissue was then stained with Carmine Alum (Stem Cell Technologies, 07070) for 48-72 hours. Mammary glands were then dehydrated using increasing ethanol concentrations (70%, 95%, 100%) 2 times each for 15 minutes, and cleared in xylene overnight. Cleared mammary glands were then mounted with Permount

mounting medium (Fischer Chemical, SP15-100) and glass coverslips and allowed to dry overnight. Slides were imaged on the PerkinElmer Vectra3 slide scanner.

Tumor dissociation

Tumors harvested from euthanized mice were digested in DMEM containing 2 mg/ml Collagenase A and 100U/ml hyaluronidase at 37°C for 2 hours with gentle agitation using a rotator. Following digestion, samples were strained through 70mm (Corning, 431751) and 40mm cell strainers to obtain a single-cell suspension. Finally, red blood cells were removed with an ammonium chloride lyse, and cells were washed in PBS.

FFPE tissue processing

Harvested mammary glands, tumors and lungs were placed in tissue biopsy cassettes and fixed in 10% Neutral Buffered Formalin (Leica, 3800598) at 4°C overnight. The formalin was then removed and tissues were soaked in 70% ethanol at 4°C for at least 2 days before embedding in paraffin blocks. Hematoxylin & Eosin (H&E) staining was performed on sections cut from the paraffin blocks. Embedding, sectioning, and H&E staining were performed by Dartmouth-Hitchcock Pathology Shared Resources.

Flow Cytometry and Fluorescence assisted cell sorting (FACS)

Single-cell suspensions were first stained with fluorescently labelled antibodies. Tumor single-cell suspensions were stained with Alexa Fluor 700 anti-CD326 (Ep-CAM) antibody (Biolegend, 118239, clone: G8.8, 1:100 dilution), PE/Cyanine 7 anti-mouse CD31antibody (Biolegend, 102418, clone:390, 1:100 dilution), and PE/Cyanine 7 anti-mouse CD45 antibody (Biolegend, 103114, clone: 30-F11, 1:100 dilution) for 30 minutes on ice. Mammary gland single-cell suspensions were stained with all the above antibodies, with the addition of Super Bright 600 anti-CD49f (integrin alpha 6) antibody (Thermo Scientific, 63-0495-42, clone: GoH3, 1:100 dilution).

For staining intracellular keratins, single-cell suspensions were fixed for 15 minutes at RT in 2% paraformaldehyde (methanol free, Thermo Scientific, J19943-K2), and permeabilized for 15 minutes at RT in Intracellular Staining Perm Wash Buffer (Biolegend, 421002), before staining with Recombinant anti-Cytokeratin 8 antibody Alexa Fluor 647 (Abcam, ab192468, clone: EP1628Y, 1:100 dilution), and Recombinant anti-Cytokeratin 5 antibody (Abcam, ab236216, clone: SP27) conjugated to Dylight 405 (Abcam, ab201798), or anti-Cytokeratin 14 monoclonal antibody (Thermo Scientific, MA5-11599, clone: LL002), conjugated to Pacific Blue (Thermo Scientific, P30013) for 30 minutes at RT. Samples were washed and resuspended in PBS supplemented with 2% FBS before being analyzed for cell marker expression using BioRad ZE-5 cell analyzer. Compensation was performed with the aid of single-stained Ultracomp eBeads plus compensation beads (Invitrogen, 01-3333-42). Analysis and plot generation was performed on FlowJo.

For sorting of GFP-expressing cells, single-cell suspensions were only stained for extracellular markers, and DAPI (Sigma-Aldrich, 10236276001) was added at a dilution of 1:1000 after the final wash step in order to facilitate live-cell sorting. GFP-positive cells were sorted on FACSAria III cell sorter by first gating on DAPI-negative live cells, and CD31- and CD45-negative epithelial cells.

Immunohistochemistry (IHC) staining

Slides are cut at 4mm and air dried at RT before baking at 60°C for 30 minutes. Automated protocol performed on the Leica Bond Rx (Leica Biosystems) includes paraffin dewax, antigen retrieval, peroxide block and staining. Heat induced epitope retrieval using Bond Epitope Retrieval 2, pH9 (Leica Biosystems, AR9640) was incubated at 100 degrees Celsius for 20 minutes (for anti-Cytokeratin 5, Bond Epitope Retrieval 1, pH6.0 (Leica Biosystems, AR9961) was used instead). Primary antibody anti-ER (Cell Marque, 249-R-15-ASR, clone: SP1, 1:35 dilution), anti-p63 (Biocare Medical, CM163B, clone:4A4, 1:200 dilution), and anti-Cytokeratin 5 (Abcam, ab236216, clone: SP27, 1:100 dilution) was applied and incubated for 15 minutes at room temp. Primary antibody binding is detected and

visualized using the Leica Bond Polymer Refine Detection Kit (Leica Biosystems DS9800) with DAB chromogen and Hematoxylin counterstain. Slides were imaged using the PerkinElmer Vectra3 slide scanner, and PhenoChart. Staining was qualified and quantified by a breast pathologist (KM).

Multiplexed TSA staining

Staining was optimized based on the PerkinElmer OPAL Assay Development Guide (August 2017). Sample slides were baked at 60°C for 2 hours to remove paraffin wax, followed by 3 xylene washes of 10 minutes each. Slides were then rehydrated with decreasing concentrations of ethanol (100%, 95%, 70%, and 50%), followed by fixation in 10% Neutral Buffered Formalin (Leica, 3800598) for 30 minutes at RT. Antigen retrieval was performed in BOND Epitope Retrieval Solution 1 (Leica, AR9961) for 20 minutes at high pressure in a pressure cooker. After the slides were cooled, they were rinsed in PBS, and endogenous peroxide activity was blocked by treatment with 3% hydrogen peroxide for 10 minutes. Slides were washed in TBS + 0.1% tween (TBS-T) and blocked in Antibody Diluent/Block (Akoya Biosciences, ARD1001EA) for 30 minutes at RT. Primary antibody were added and slides were incubated at RT for 30 minutes. After TBS-T washes, secondary antibodies were added and incubated at RT for another 30 minutes, followed by TBS-T washes. Opal fluorophore was applied to slides for precisely 6 minutes, followed by TBS-T washes. Slides were then boiled for 2 minutes in a microwave at 100% power, followed by 15 minutes at 20% power in AR6 Buffer (Akoya Biosciences, AR600250ML) to affix Opal to target sites and remove primary and secondary antibodies. This process is repeated for each primary antibody used. After staining with the final antibody, Spectral DAPI (Akoya Biosciences, FP1490) was added, and slides were mounted with ProLong Diamond Antifade Mountant (Invitrogen, P36961) and glass coverslips.

The primary antibody and Opal pairs used are as follows:

For TMA samples: anti-Cytokeratin 14 (Abcam, ab119695, clone:SP53, 1:200 dilution) with Opal 620 (Akoya Biosciences, FP1495001KT, 1:500 dilution), anti-Cytokeratin 5

(Abcam, ab64081, clone: SP27, 1:300 dilution) with Opal 520 (Akoya Biosciences, FP1487001KT, 1:150 dilution), anti-ER (Cell Marque, 249-R-15-ASR, clone: SP1, 1:70 dilution) with Opal 650 (Akoya Biosciences, FP1496001KT, 1:500 dilution), and anti-Cytokeratin 8 (Abcam, ab53280, clone: EP1628Y, 1:400 dilution) with Opal 570 (Akoya Biosciences, FP1488001KT, 1:600 dilution).

For mouse tumor and lung samples: anti-Cytokeratin 14 (1:200 dilution) with Opal 620 (1:500 dilution), anti-Cytokeratin 5 (1:300 dilution) with Opal 690 (Akoya Biosciences, FP1497001KT, 1:150 dilution), anti-GFP (Cell Signaling, 2956, clone: D5.1, 1:150 dilution) with Opal 650 (1:500 dilution), and anti-Cytokeratin 8 (1:300 dilution) with Opal 570 (1:600 dilution).

Image processing, analysis, and phenotype training

Whole slide scans were imaged at 4x resolution using the PerkinElmer Vectra3 slide scanner, and Regions of interest (ROIs) were selected on PhenoChart. ROIs were then imaged at 20x resolution. Spectral unmixing was performed, and each Opal was assigned a color using the software InForm, which was also used to train the algorithm for phenotype quantification. Tissue and cell segmentation was performed (with the aid of DAPI as the nuclear marker, and Krt8, as the cytoplasmic marker), and cells were phenotyped based on marker expression, and validated by marker distribution (entire Cell Mean Fluorescent units extracted for each marker and normalized as a percentile of maximum and minimum fluorescence across all cells in all images).

Re-analysis of TCGA breast cancer cohort

Preprocessed protein expression data from RPPA assays and gene expression data from RNA-seq from breast cancer patients in The Cancer Genome Atlas (TCGA) were downloaded from Synapse (https://doi.org/10.7303/syn300013).

For ERα based analyses, 564 subjects of Luminal A and B primary breast tumors that were estrogen receptor-positive (ER+) with defined progesterone receptor (PR) status from immunohistochemistry staining were used. Binary classification of ERα levels were determined based on the distribution of ERα from RPPA assays. Subjects were defined as ERα low (n = 141) if the level of ERα were below the 25th percentile and as ERα high (n = 423) if the level of ERα were above the 25th percentile.

For *ESR1* based analyses, 719 subjects of Luminal A and B primary breast tumors that were ER+ with defined PR status from immunohistochemistry staining were used. Binary classification of *ESR1* levels were determined based on the distribution of *ESR1* from RNA-seq. Subjects were defined as *ESR1* low (n = 180) if the level of *ESR1* were below the 25th percentile and as *ESR1* high (n = 539) if the level of *ESR1* were above the 25th percentile.

Gene expression levels of 439 basal signature genes (obtained from previously published data[95]) were available in the TCGA data. Winsorized Z-scores for each gene were used to compare the differences in expression between ERα/*ESR1* high and low subjects. Log2 transformed counts were used to compare the differences for *POSTN, SPP1*, and *SOX10*.

scRNA-seq sample and library preparation

GFP expressing cells were collected from the FACSAria, resuspended in PBS + 0.05% BSA and brought to the Genomics Shared Resource for processing. Cell suspensions were counted on a Nexcelom K2 automated cell counter and loaded onto a Chromium Single Cell G Chip (10x Genomics Inc.) targeting a capture rate of 10,000 cells per sample. Single cell RNA-seq libraries were prepared using the Chromium Single Cell 3' v3.1 kit (10x Genomics) following the manufacturer's protocol. Libraries were quantified by qubit and peak size determined by Fragment Analyzer (Agilent). All libraries were pooled and sequenced on an Illumina NextSeq2000 using Read1 28bp, Read2 90bp to generate an average of 25,000 reads/cell.

scRNA-seq data analysis

Raw sequencing data were demultiplexed to create individual FASTQ files using Cell Ranger (v.6.0.1) *mkfastq* (10X Genomics)[96]. Cell Ranger *count* was used to map sequence reads to the reference genome (mm10-2020-A) and construct a matrix of raw read counts. R-package Seurat (v.4.0.4)[97] was used for downstream processing, normalization, and dataset integration. Raw read counts for cell-containing droplets were imported into R v.4.0.3 using Seurat function *Read10X*. Doublets were identified and removed using the simulation-based approach implemented by function *scDblFinder* from R-package scDblFinder[98], with the doublet rate argument (dbr) set based on the number of recovered cells from each experiment. Cells with ≤ 500 UMIs or ≤ 200 detected features were removed from further analyses. Cells were further filtered to exclude those identified as outliers (using function *isOutlier* from R-package scater v.1.18.6[99]) from the distribution of mitochondrial read counts (percentage reads mapped to mitochondrial genes). Genes with <10 assigned reads across all samples were also removed prior to downstream analysis. Read counts were normalized using *sctransform*[100] with default settings. Datasets were integrated using the anchor-based integration framework implemented in Seurat, using 3000 integration features and reciprocal principal components analysis (RPCA) for anchor selection. Unsupervised clustering was performed at multiple resolutions using default parameters in Seurat. Clustree v.0.4.4[101] was used to identify the optimal clustering solution for downstream analysis. Contaminating clusters expressing markers of lymphoid and myeloid lineages were identified and removed from the dataset, and the remaining cells were resubjected to unsupervised clustering analysis and dimensionality reduction. Cluster specific marker genes for each cluster were identified using the Seurat function *FindMarkers* with arguments "min.pct = 0.1, only.pos = TRUE" and default parameters. Differential expression analysis was also performed using *FindMarkers* with argument "min.pct = 0.1" and default parameters. Luminal progenitor (LP), mature luminal (ML), and basal (MS) gene expression signatures (obtained from Pal *et al.*[102]) were scored for enrichment at the individual cell-level using variance-adjusted Mahalanobis (VAM)[103]. VAM generates cell-specific scores, using the

gamma cumulative distribution function (CDF), between 0 (no enrichment) and 1 (highly enriched) for a given gene-set. Log-normalized counts (generated by Seurat function *NormalizeData*) were used as input to function *vamForSeurat,* which was run with default settings. Squared adjusted Mahalanobis distances were used to generate ternary plots, positioning each cell according to its combined expression of LP, ML, and MS gene signatures.

Raw scRNA-seq data

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002[104]) and are accessible through GEO Series accession number GSE214815 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc= GSE214815).

2.4 Results

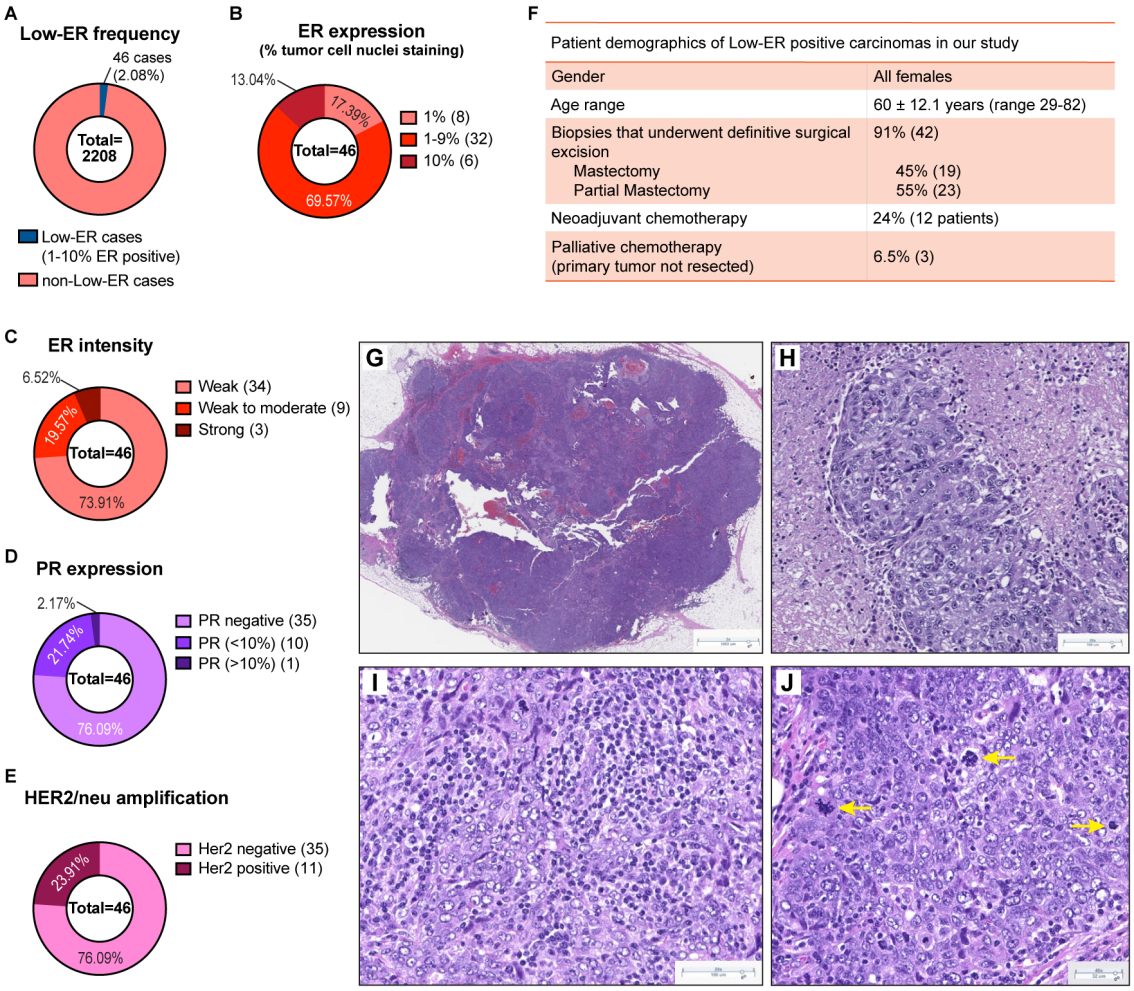
2.4.1 Low-ER breast tumors exhibit distinct basal-like features

Previous studies have observed that invasive breast carcinomas with low-ER expression, in which less than 10% of tumor cells express ER, share more similarities with TNBCs when compared to tumors that harbored more than 10% ER-expressing cells[105]. We analyzed newly diagnosed invasive breast carcinomas from the pathology database at Dartmouth-Hitchcock Medical Center (DHMC) from 2012-2020 (n=2208) and observed 46 (2.1%) that were classified as low-ER tumors containing between 1-10% ER-expressing tumor cells (Fig 1A). Most, (41 out of 46, 89%) were high-grade invasive carcinomas (Fig 2A) with 1-9% ER-expressing tumor cells (Fig 1B). The intensity of ER expression was also reduced in these low-ER tumors, with 93.5% showing moderate or weaker ER staining (Fig 1C and 2B-D). In contrast to high expressing ER tumors which typically also exhibit some degree of progesterone receptor (PR) expression, most low-ER tumors (76%) were PR negative (Fig 1D). The frequency of *HER2*-positivity was as expected, with 24% of cases harboring *HER2/neu* amplification (Fig 1E).

Treatments administered to patients harboring low-ER tumors are more similar to treatment regimens for patients with TNBC. In our cohort, 43% of patients received chemotherapy, 34% received radiation therapy, and only 23% received hormone therapy (Fig 2E). Interestingly, response rates to neoadjuvant chemotherapy in patients with low-ER tumors were similar to ER-negative tumors and significantly different from tumors with moderate and high ER-positive tumors[106]. Twelve patients (24%) received neoadjuvant chemotherapy, and most (75%) achieved a pathologic complete response (Fig 1F and 2A). The majority of patients (81%) had no evidence of the disease at follow-up (Fig 2F).

Microscopic examination of tumors with low-ER expression revealed histologic features commonly present in breast tumors with basal-like molecular profiles and carcinomas harboring *BRCA1* mutations. Twenty of 46 tumors (43%) showed well-circumscribed or pushing borders and were comprised of high-grade, pleomorphic tumor cells arranged in solid sheets with conspicuous mitotic activity, admixed necrosis, and prominent tumor infiltrating lymphoplasmacytic infiltrates (Fig 1G-J). The histologic findings in our cases are in agreement with several other recent studies that have shown that low-ER breast tumors show pathologic characteristics typical of ER-negative tumors with basal-like gene expression profiles[79,105,107].

These data indicate that low-ER tumors are a distinct subtype of breast cancer, separate from the typical, ER-expressing luminal-like subtypes. They display more similarities to basal-like or triple-negative tumors, especially with respect to biomarker expression, pathology and the types of treatments patients receive.



Patient demographics of Low-ER positive carcinomas in our study	
Gender	All females
Age range	60 ± 12.1 years (range 29-82)
Biopsies that underwent definitive surgical excision Mastectomy Partial Mastectomy	91% (42) 45% (19) 55% (23)
Neoadjuvant chemotherapy	24% (12 patients)
Palliative chemotherapy (primary tumor not resected)	6.5% (3)

Figure 1: Basal-like features of tumors expressing low ER. (A) Frequency of cases expressing low ER levels (<1% ER expressing nuclei) in newly diagnosed invasive breast carcinomas from the pathology database at Dartmouth-Hitchcock Medical Center (DHMC) from 2012-2020. **(B-E)** Breakdown of ER expression levels **(B)**, ER intensity levels **(C)**, PR expression **(D)**, and Her2/neu amplification **(E)**, in the 46 low-ER cases. **(F)** Patient demographics of the 46 low-ER cases in our cohort. **(G-J)** H&E staining of low-ER tumors showing solid tumor with pushing borders **(G)**, pleomorphic, high-grade nuclei with admixed necrosis **(H)**, prominent lymphoplasmacytic infiltrates **(I)**, and conspicuous mitotic activity, including atypical mitoses (yellow arrows) **(J)**.

A

Histological features of 46 Low-ER positive carcinomas	
Type	Invasive ductal carcinoma: 96% (44) Invasive lobular carcinoma: 4% (2)
Grade	High: 89% (41) Intermediate: 11% (5)
Size	2.97 ± 2.6 cm (range: 0.1 – 10.0 cm)
Ductal carcinoma in-situ present	55%
Lymphovascular invasion present	39%
Lymph node status	N0 = 30 (ITCs in two patients) N1 = 2 N2 = 0 N3 = 4
Response to neoadjuvant chemotherapy	Pathologic complete response = 75% (9/12) Pathologic partial response = 25% (3/12)* *Two had microinvasion only

Oncotype Recurrence Score (n=10) = 48.8 ± 12 (18-61)

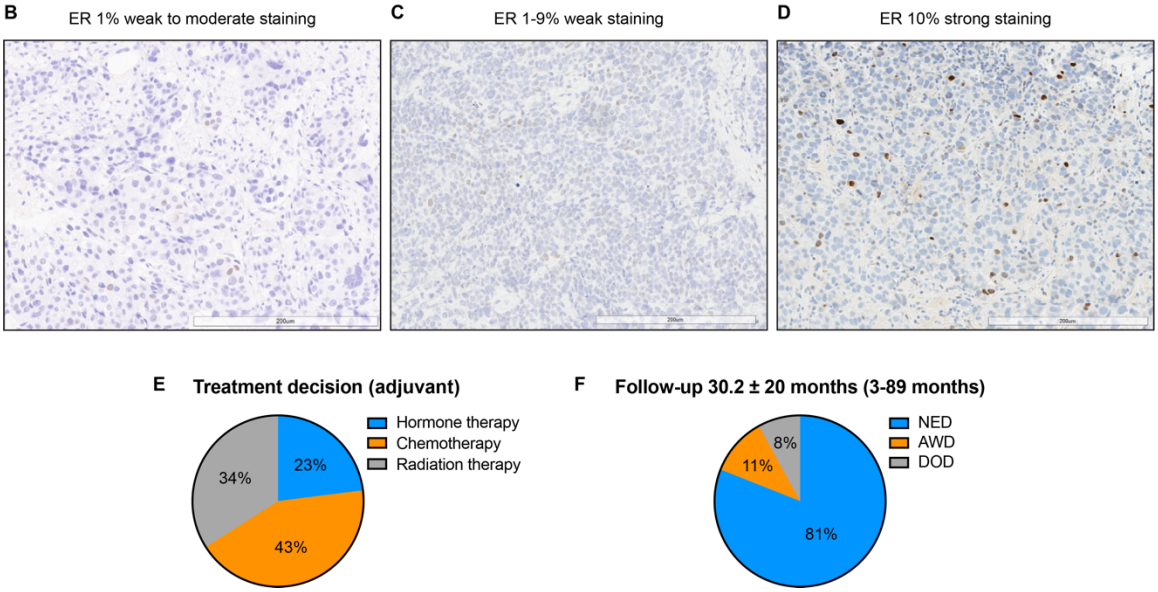


Figure 2: Histological features and treatment decisions for patients with low-ER tumors. **(A)** Histological features of the 46 low-ER cases in our cohort. **(B-D)** Representative IHC images showing the different levels and intensity of ER staining in the low-ER tumors. **(E-F)** Breakdown of treatment decisions **(E)**, and follow-up details **(F)**, of the 46 low-ER cases.

2.4.2 Low-ER tumors differ from luminal B tumors in their biomarker profiles

To investigate precisely how different low-ER tumors are from luminal tumors, 24 luminal B and 22 low-ER tumors were compiled into two tissue microarrays (TMAs). The luminal B tumors were selected based on a combination of pathologic characteristics including high tumor grade, high mitotic rate (>18 mitoses per 10 high power fields), and diffuse ER expression in tumor cells (all tumors showed >80% tumor cell nuclei with ER expression). Compared to the low-ER group, none of luminal B tumors showed histologic basal-like phenotypic characteristics. In the luminal B group, most patients (87.5%) received hormone therapy, or a combination of hormone therapy and chemotherapy (Fig 4A), which proved effective, with more than 91% of patients showing no evidence of disease at follow-up (Fig 4B). No patients received neoadjuvant chemotherapy (Fig 3A). Most tumors in the luminal B group showed strong PR positivity (Fig 4C) and all tumors were negative for HER2/neu amplification (Fig 4D), which are more typical features of a luminal-like breast cancer subtype. In comparison to low-ER tumors, none of the luminal B tumors contained the constellation of basal-like histologic features we observed in 43% of low-ER tumors.

The two TMAs were stained for expression of ER (luminal marker) and Krt5 (basal marker). As expected, immunohistochemistry (IHC) staining of these TMAs showed that all 24 luminal B samples were ER positive (>10% ER expressing cells), with almost 80% showing strong ER intensity, while the low-ER samples were mostly low-ER expressing, with weak or moderate ER intensity (Fig 3B, 3C, and 4E). Eight tumor cores in the low-ER TMA were observed to stain negative for ER, while one had 10-20% ER positive nuclei. The low-ER tumors were identified based on ER expression in the diagnostic biopsy, while TMA cores were obtained from the surgical specimens. These tumors are still biologically low-ER tumors, however due to focal ER expression and heterogeneity, the ER expression of these tumor cores may vary. When stained for Krt5, a basal marker used to identify basal-like breast cancer subtypes, none of the luminal B samples expressed Krt5, but 65% of the low-ER samples were $Krt5^{+}$ (Fig 3D and 3E). Staining for p63, another basal marker, also

showed higher positivity in low-ER tumors compared to luminal B (Fig 4F). These data support histological observations that low-ER tumors are more basal-like and express higher levels of basal markers than luminal B tumors.

We wanted to further investigate if the expression of basal markers within the low-ER tumors corresponded to a loss of luminal marker expression. Tyramide Signal Amplification (TSA) staining[108,109] was used with luminal marker ER and Krt8 antibodies in addition to basal markers Krt5 and Krt14. We utilized this staining method as it allows the use of multiple antibodies raised in the same species and have previously used it to stain TMAs containing hundreds of patient tumor samples[110,111]. All luminal B TMA cores strongly expressed both ER and Krt8, with no expression of either basal marker (Fig 4G). In contrast, low-ER TMA cores exhibited more heterogeneous ER, Krt5, and Krt14 expression (Fig 3F). Importantly, the low-ER TMA cores also expressed high Krt8 levels, suggesting luminal lineage identity is retained despite the reduction in ER and increase in Krt5 expression. Along these lines, most of the Krt5-expressing tumor cells also co-expressed Krt8, with only a small percentage of cells exclusively expressing the basal marker.

The low-ER TMA cores were further analyzed to identify the different cell types that these heterogeneous tumors were comprised of. While a few cells were found to only express Krt5, most of the Krt5-expressing cells co-expressed Krt8. Krt14 was less abundant in these tumors, with $Krt14^{+}$ cells also co-expressing both Krt5 and Krt8, indicating that most basal-like cells within these tumors express basal markers without losing their luminal identity i.e., exhibiting a lumino-basal phenotype (Fig 3F). Cells with fully basal phenotypes in which only Krt5 was expressed were rare and only found in 9 out of 13 low-ER tumor cores (Fig 3F and 3G). ER expression was expectedly weak and scarce, but was found both in cells expressing Krt8 only, and in cells co-expressing either Krt8 and Krt5 or Krt8 and Krt14 (Fig 3F).

To analyze the distribution of these various cell types within the low-ER tumors, we quantified each cell phenotype within the low-ER tumor cores. Of the 20 tumor cores analyzed, six cores predominantly consist of the luminal cells, seven cores predominantly consist of cells of the lumino-basal phenotype (Fig 3G), and seven cores were excluded from analysis due to an absence of basal marker expression. As expected, ER expression is more abundant in cells of the luminal phenotype as compared to those exhibiting a lumino-basal phenotype. The strictly basal phenotype was also not commonly found within these tumors, indicating that tumor cells rarely lost all luminal marker expression to become fully basal.

These results provide evidence that distinguishes low-ER tumors from luminal B tumors, both in terms of histopathology and luminal and basal marker expression. Furthermore, low-ER tumors are more heterogeneous in epithelial cell marker expression, with the emergence of a lumino-basal cell phenotype that could define the biological properties of this subtype.

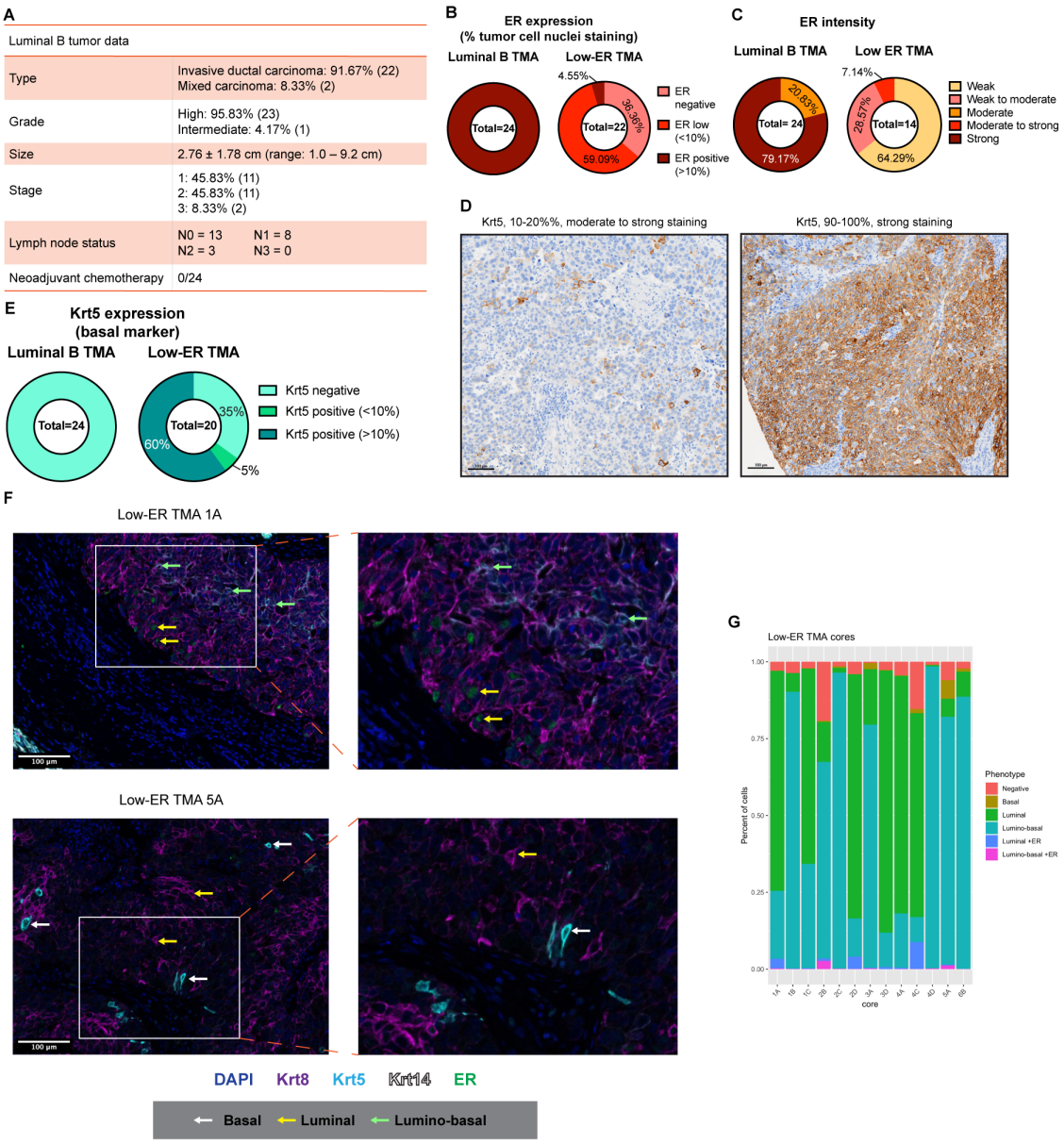


Figure 3: Low-ER tumors are distinct from luminal B tumors. (A) Details of the luminal B tumors included in TMAs as a comparison to Low-ER tumors. **(B-C)** ER expression **(B)** and intensity **(C)** levels of the luminal B tumors. **(D)** Representative images of Krt5 IHC staining in luminal B and low-ER TMAs. **(E)** Quantified Krt5 expression in luminal B and low-ER TMAs. **(F)** Representative images of TSA staining containing heterogeneous populations in low-ER TMAs. Samples were stained with Krt8 (purple), Krt5 (cyan), Krt14 (white), ER (green), and DAPI (blue). White arrows point to basal tumor cells, yellow arrows point to luminal tumor cells, and green arrows point to lumino-basal tumor cells. **(G)**

Quantification of different cell phenotypes found within 13 of the low-ER TMA cores which were found to harbor heterogeneous populations (TMA cores with homogenous populations were excluded).

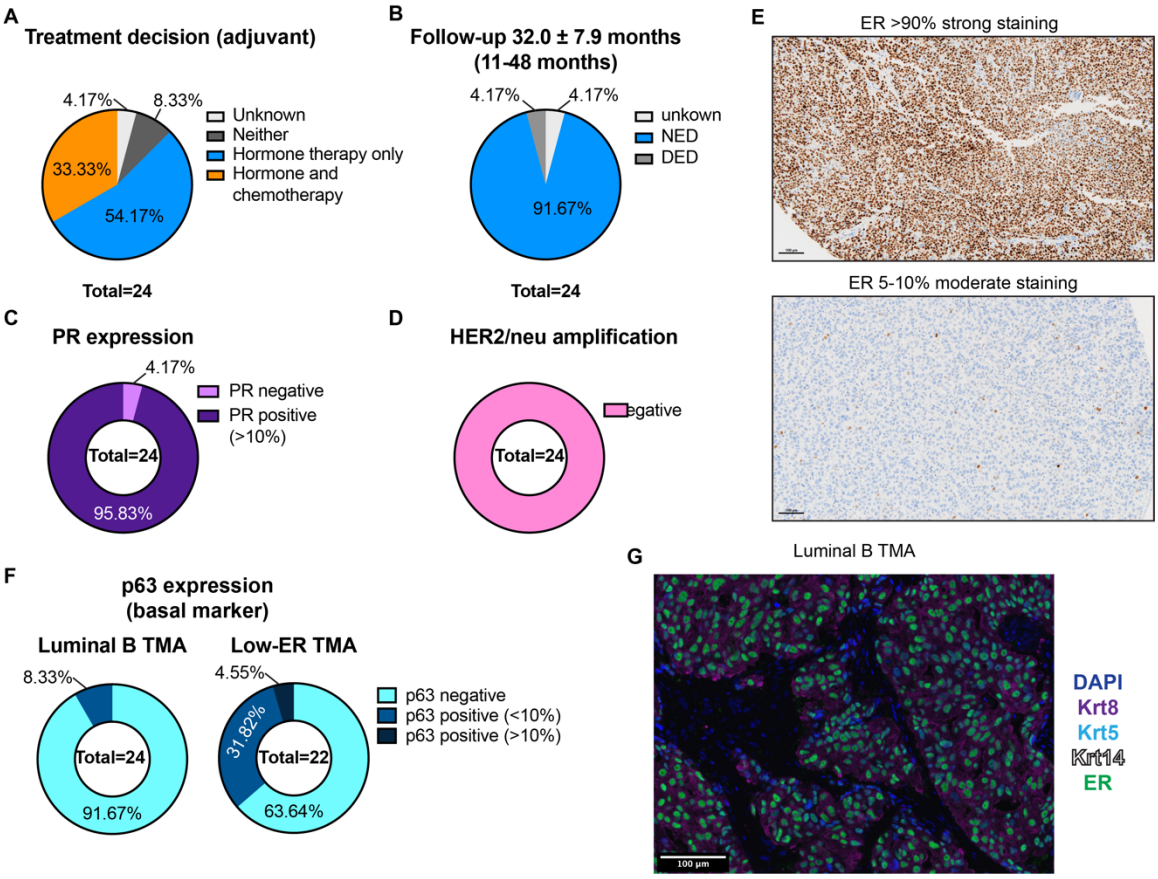


Figure 4: Luminal B tumors show features typical of luminal-like mammary tumors. (A-D) Breakdown of treatment decisions **(A)**, follow-up details **(B)**, PR expression **(C)**, and Her2/neu amplification **(D)**, of the 24 luminal B cases used in the TMA. **(E)** Representative IHC images showing the different levels and intensity of ER staining in the luminal B TMA cores. **(F)** p63 expression differences between luminal B and low-ER tumors. **(G)** Representative images of TSA staining showing homogeneous populations in luminal B TMAs. Samples were stained with Krt8 (purple), Krt5 (cyan), Krt14 (white), ER (green), and DAPI (blue).

2.4.3 Tumors with lower ER expression express a distinct basal signature

We sought to explore a larger set of ER-positive tumors, specifically to assess whether lower ER expression was associated with expression of a basal gene signature. We first analyzed 564 ER-positive breast cancer tumor cases from The Cancer Genome Atlas (TCGA), in which their ERα expression was quantified using Reverse Phase Protein Array (RPPA), which included both ER+/PR+, and ER+/PR- cases (Fig 6A). These cases were stratified into two groups; ERα low (141 cases, bottom quartile of ER expression), and ERα high (423 cases, top 75% of ERα expression) (Fig 5A) and analyzed their basal gene signature[95]. Unsupervised clustering of all basal signature genes revealed modules with higher relative expression in ERα low cases compared to ERα high cases (Fig 6B). Supervised clustering of these cases based on ERα expression revealed a statistically significant (Fig 5B) upregulation of basal signature gene expression in the ERα low cluster (Fig 5C) irrespective of PR status (Fig 6C). Similar results were observed when cases were stratified using *ESR1* mRNA levels instead of ERα levels (Fig 5D-F, 6D-F), whereby tumors expressing lower *ESR1* demonstrated an increased expression of genes conferring basal identity, suggesting that tumors gain basal-like traits upon concomitant reduction of ER levels.

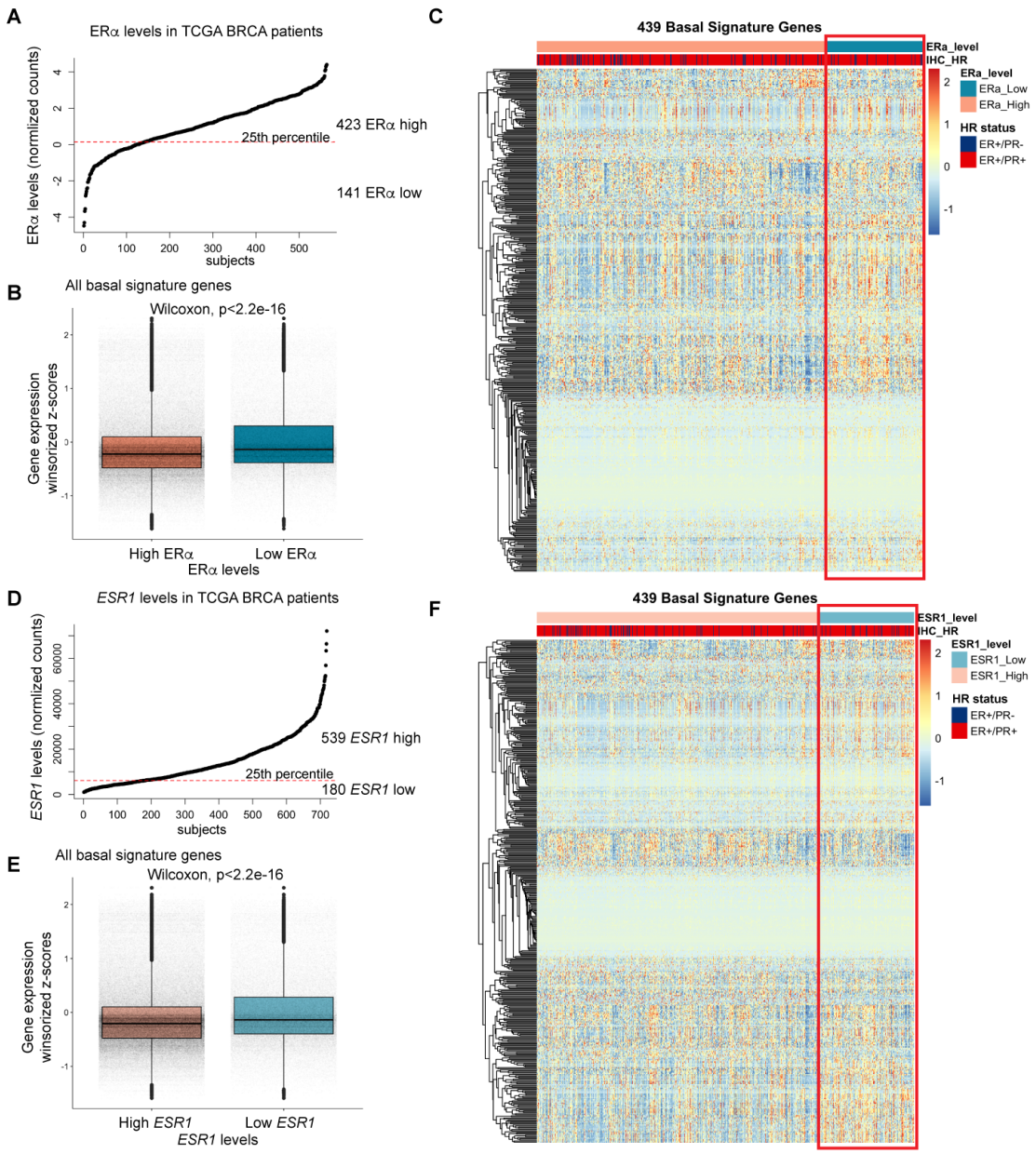


Figure 5: Distinct basal gene expression signature in tumors with lower ER. (A) Stratification of ER positive breast tumor cases from TCGA into ERα low and ERα high groups, based on ERα expression. **(B)** Boxplots comparing the distribution of basal signature gene expression in the ERα low and ERα high groups. **(C)** Heatmap with supervised clustering of the ER positive tumors highlighting higher basal signature gene expression in the ERα low group. **(D)** Stratification of ER positive breast tumor cases from TCGA into *ESR1* low and *ESR1* high groups, based on *ESR1* expression. **(E)** Boxplots

comparing distribution of basal signature gene expression in the *ESR1* low and *ESR1* high groups. **(F)** Heatmap with supervised clustering of the ER positive tumors highlighting higher basal signature gene expression in the *ESR1* low group.

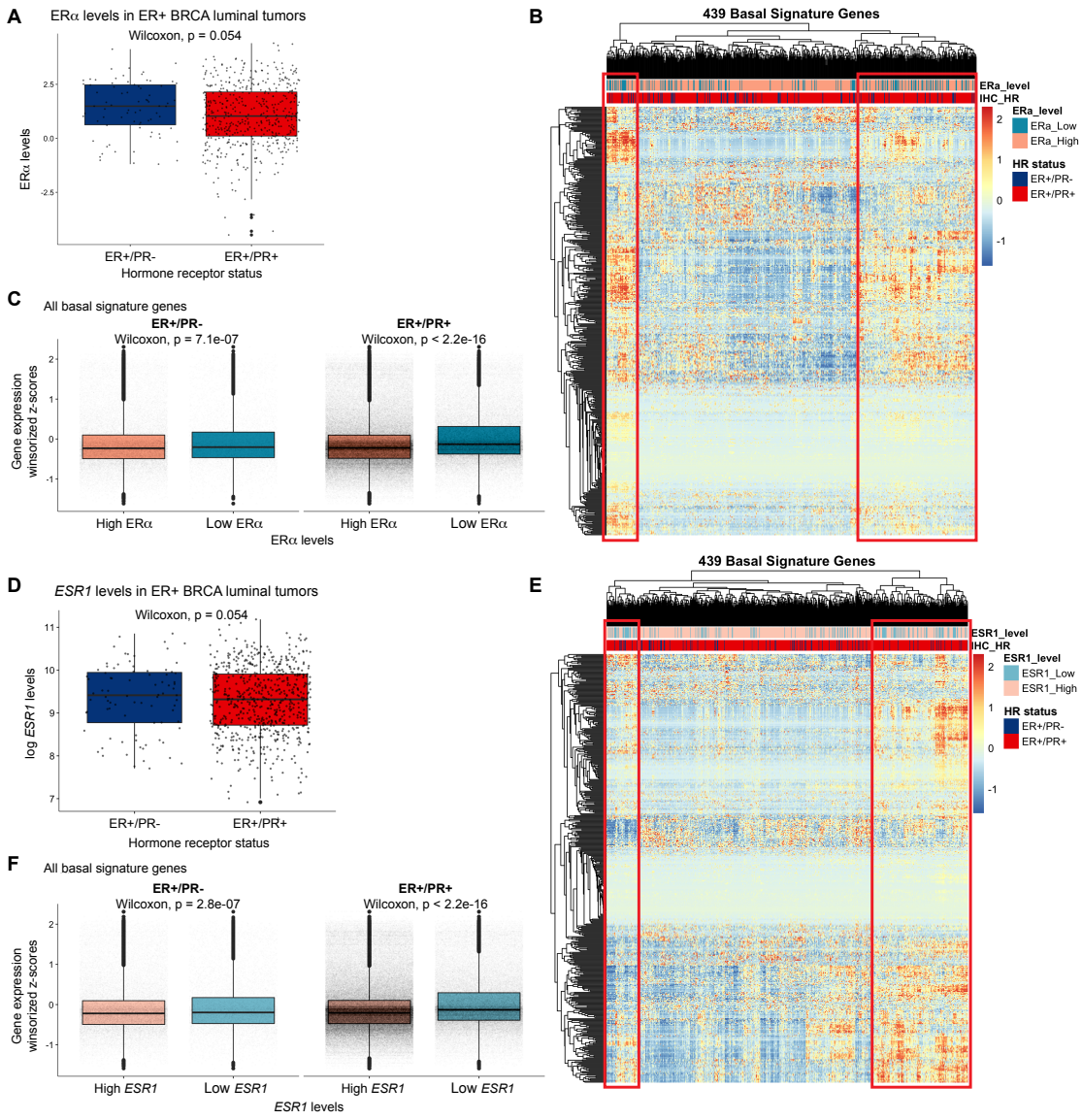


Figure 6: Unsupervised clustering show enrichment of basal signature genes in ER low tumors. (A) Boxplots showing distribution of ERα levels in the ER+/PR- and ER+/PR+ cases used in this analysis. **(B)** Heatmap with unsupervised clustering of the basal signature gene expression in ER positive tumors. **(C)** Boxplots showing the differences in distribution of basal signature gene expression stratified by ERα expression level and PR status. **(D)** Boxplots showing distribution of *ESR1* levels in the ER+/PR- and ER+/PR+ cases used in this analysis. **(E)** Heatmap with unsupervised clustering of the basal signature gene expression in ER positive tumors. **(F)** Boxplots showing the differences in distribution of basal signature gene expression stratified by *ESR1* expression level and PR status.

2.4.4 Tumor cell plasticity results in emergence of basal-like features in low-ER tumors

We reasoned that the emergence of basal-like characteristics in the low-ER luminal tumors may arise via two possible mechanisms. Firstly, an expansion of basal cells might occur during the later stages of tumorigenesis in low-ER tumors, or alternatively, cellular plasticity may reprogram luminal tumor cells and allow them to acquire basal-like traits. To identify which mechanism was at play, we carried out lineage tracing using a model that accurately captured aspects of low-ER breast tumors. MMTV-PyMT[92] is a mouse mammary tumor model that closely resembles human luminal B breast cancers[112] whereby late stage tumors lose ER expression[113]. IHC staining for ER revealed weak to moderate expression in MMTV-PyMT tumors, with half of the tumors displaying less that 10% ER expression (Fig 8A and 8B).

In order to trace the lineage of the MMTV-PyMT tumor cells, either Krt8 (luminal) or Krt5 (basal) specific, tamoxifen-inducible Cre-ERT promoters[82] were used to induce expression of GFP in an mTmG reporter mouse[93] to label luminal or basal cells, respectively (Fig 7A and 7B). The most efficient mammary epithelial GFP labelling was observed when tamoxifen induction was performed 3 days per week at postnatal week 3 for the Krt5-CreERT/Rosa26-mTmG model (Fig 8C), and at postnatal weeks 5 and 6 for the Krt8-CreERT/Rosa26-mTmG model (Fig 8D). Tumors that eventually arose in these tamoxifen-pulsed mice were harvested and analyzed for GFP expression by immunostaining and flow cytometry.

Flow cytometry analysis (Fig 8E) of tamoxifen-pulsed Krt5-CreERT/Rosa26-mTmG/MMTV-PyMT tumors revealed mostly GFP-negative cells (Fig 7C), indicating that MMTV-PyMT tumors did not originate from the Krt5-expressing basal lineage. In addition, most of the basal-marker expressing population (Krt5+ or Krt14+) did not inherit the GFP label from the basal lineage (mean GFP-positive basal cells= 1.67%) (Fig 7C and 7D). In contrast, a larger proportion (mean GFP-positive basal cells= 45.4%) of basal cells inherited the GFP label in the developing normal mammary gland (Fig 7E). These results indicate that the

basal-like tumor cells were not derived from the basal lineage, and that the basal-like traits emerging within MMTV-PyMT tumors did not arise from expansion of normal basal cells during the process of tumorigenesis.

Conversely, tamoxifen-pulsed Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT mice developed tumors that were primarily comprised of GFP-positive cells (Fig 7F), indicating that they originated from the luminal lineage. Strikingly, most of the basal-marker expressing tumor cells were found to have also inherited the GFP label (mean GFP-positive basal cells= 70.26%) (Fig 7F and 7G). In contrast, the GFP expression within the normal mammary gland was confined to a small percentage of basal cells (mean GFP-positive basal cells= 5.37%) (Fig 7H). This indicates that the basal-like cells within these tumors arose from the luminal lineage, providing evidence for luminal-to-basal plasticity whereby luminal cells acquire basal-like traits.

TSA staining confirmed the expression of GFP in luminal- or basal-like tumor cells. Luminal-like tumor cells were identified by Krt8 expression whereas basal-like tumor cells were identified by either Krt5 or Krt14 expression. In Krt5-CreERT/Rosa26-mTmG /MMTV-PyMT tumors, co-expression of Krt5 and GFP was restricted to cells in the tumor periphery (Fig 7I), suggesting that cells of the basal lineage were confined to the adjacent normal regions. Furthermore, Krt5 itself was primarily expressed in these adjacent normal regions, with very few Krt5-expressing cells within the main tumor. On the other hand, Krt14 expression was more abundant throughout the tumor (Fig 7I), indicating that Krt14 could serve as a more appropriate marker to track basal identity within these tumors.

GFP expression was more abundant throughout the Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT tumors (Fig 7I), reflecting their luminal origin, with most of these GFP-expressing cells co-expressing Krt8. Co-expression of GFP and Krt14 was also observed, with about 50% of basal marker expressing cells co-expressing GFP (mean= 47.79%) (Fig 8F), confirming the luminal lineage of these basal-like tumor cells. Interestingly, most of the

cells co-expressing GFP and Krt14 also co-expressed Krt8 (Fig 7I), suggesting that these tumor cells do not completely lose their luminal identity, but instead gain a lumino-basal phenotype. Quantification of these phenotypes within the tumor show that about 90% of the basal marker and GFP co-expressing cells were of the lumino-basal phenotype (mean=89.10%), with only about 10% of cells transitioning to a fully basal phenotype without Krt8 co-expression (mean=10.90%) (Fig 7J). These findings are consistent with the results from the flow cytometry analysis of these tumors which show that most of the tumor cells expressing the basal marker Krt5 or Krt14 descended from the luminal lineage.

2.4.5 Distant metastases are seeded by tumor cells of luminal origin

In addition to influencing the course of therapy, lineage plasticity could also play an important role in promoting metastasis within luminal-like tumors. To investigate whether one lineage is important for metastasis than the other, we analyzed lungs from Krt5-CreERT/Rosa26-mTmG/MMTV-PyMT and Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT tumor bearing mice. TSA staining of these lungs for Krt8, Krt14, and GFP revealed no GFP expression in all 7 metastases from Krt5-CreERT/Rosa26-mTmG /MMTV-PyMT mice (Fig 7K and 8G). In contrast, 16 out of 17 lung metastases from Krt8-CreERT/Rosa26-mTmG /MMTV-PyMT mice express GFP, indicating that the metastatic colony is seeded from a tumor cell of a luminal lineage.

All metastases, regardless of the model they arose from, exhibited Krt8 positivity (Fig 8G), consistent with the luminal nature of the primary tumor. In contrast, Krt14 was only observed in a fraction of the metastases (6 out of 17), primarily on the periphery of the metastatic colony. While this may suggest that luminal-to-basal plasticity may not be important in metastatic seeding, it is plausible that lumino-basal tumor cells may have lost their basal marker expression following metastatic colony formation.

Next, we addressed the requirement for lumino-basal cells for metastatic outgrowth, in contrast to its role in metastatic seeding. In our previous experiments above, lungs were

obtained from euthanized mice when they reached tumor burden (~1.0mm^3). Tumor burden was typically achieved about 4 weeks after an initial palpable tumor was observed, which may not allow sufficient time for the outgrowth of seeded metastases. In order to allow the lung metastases to continue to grow beyond this point, we surgically resected the primary tumor from a Krt8-CreERT/Rosa26-mTmG /MMTV-PyMT mouse, and allowed the lung metastases to develop until the surgically resected tumor began to regrow. This took an additional 3 weeks, which allowed the lung metastases to develop over a total of 7 weeks. Larger metastases were observed in this case, along with an increase in smaller metastatic nodules. Consistent with previous observations, all the metastases expressed GFP and Krt8, and only 20 out of 34 metastases exhibited Krt14 positivity (Fig 8G). Again, there is a possibility that lumino-basal tumor cells have lost their basal marker expression upon reaching the lung. As we are not able to study the metastatic process at multiple timepoints, it is unknown whether lumino-basal cells are important in the metastatic process.

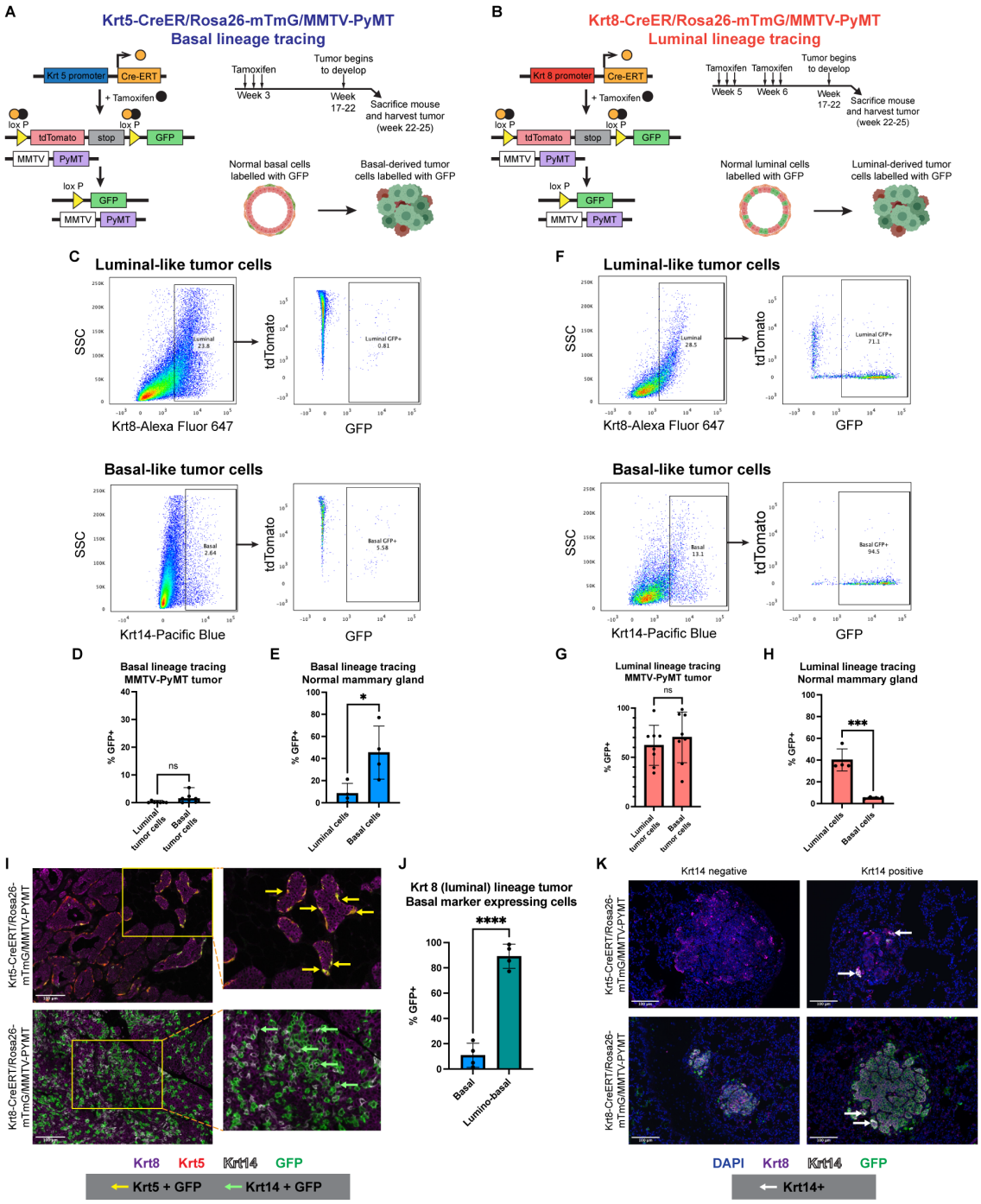
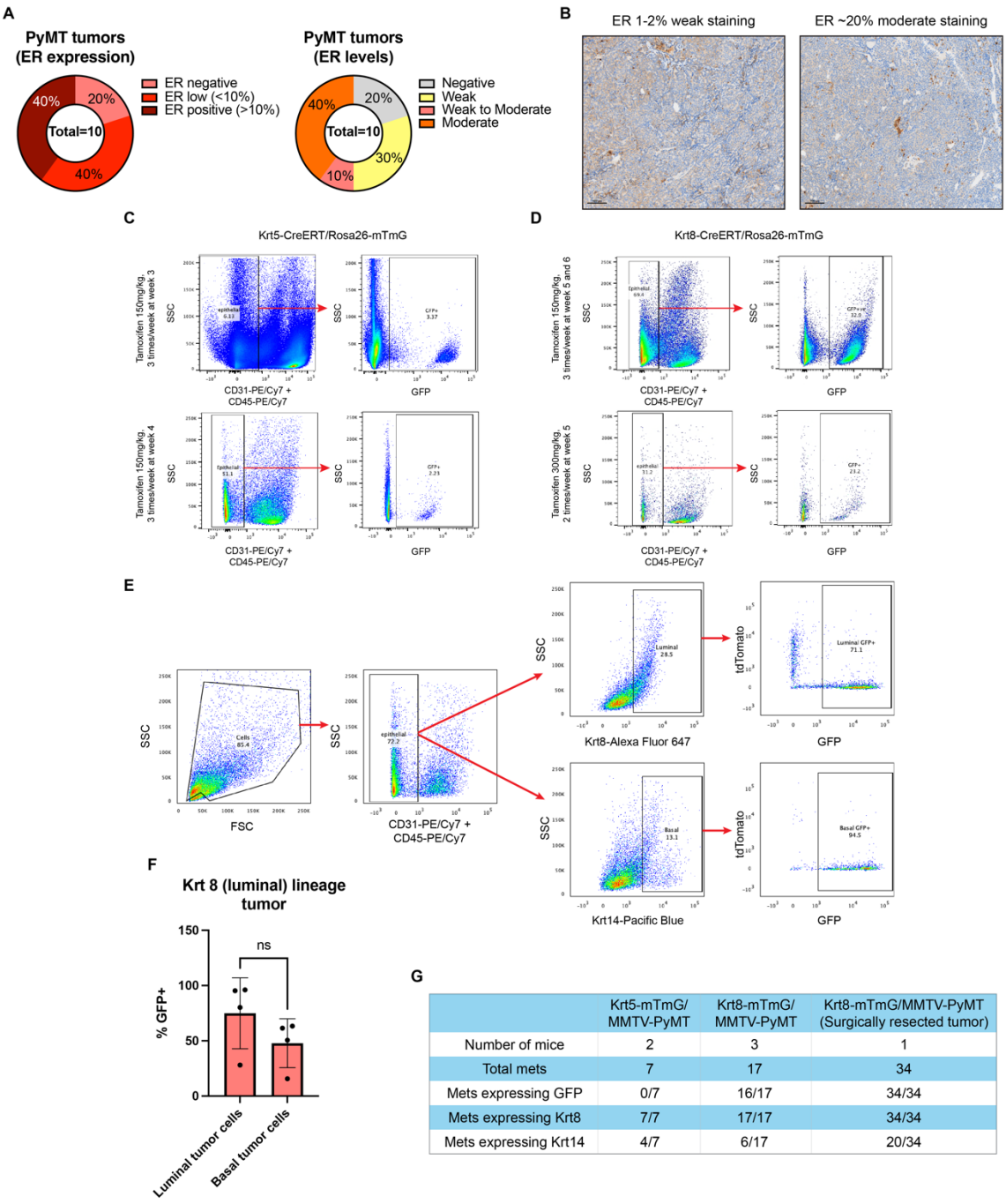


Figure 7: Basal-like MMTV-PYMT tumor cells arise from the luminal lineage. (A-B) Strategy to trace basal and luminal lineage tumor cells in MMTV-PyMT mice. Lineage-specific and tamoxifen-inducible Cre-ERT2 was used to specifically label basal **(A)** or luminal **(B)** epithelial cells with GFP. **(C)** Representative flow cytometry plots of luminal and basal tumor cells inheriting the basal lineage GFP label. **(D-E)** Percentage of basal and

luminal cells expressing basal lineage GFP in Krt5-CreERT/Rosa26-mTmG/MMTV-PyMT tumors (n=7, p=0.0513, unpaired t test) **(D)** and the normal Krt5-CreERT/Rosa26-mTmG mouse mammary gland (n=4, p=0.028, unpaired t test) **(E)**. **(F)** Representative flow cytometry plots showing luminal and basal tumor cells inheriting the luminal lineage GFP label. **(G-H)** Percentage of basal and luminal cells expressing luminal lineage GFP in Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT tumors (n=8, p=0.4938, unpaired t test) **(G)** and the normal Krt8-CreERT/Rosa26-mTmG mouse mammary gland (n=4, p=0.0005, unpaired t test) **(H)**. **(I)** Representative images of TSA staining of Krt5-CreERT/Rosa26-mTmG/MMTV-PyMT and Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT tumors. Samples were stained with Krt8 (purple), Krt5 (red), Krt14 (white), and GFP (green). Yellow arrows point to Krt5+/GFP+ tumor cells, and green arrows point to Krt14+/GFP+ tumor cells. **(J)** Quantification of lumino-basal tumor cells expressing GFP and strictly basal tumor cells expressing GFP from TSA-stained images of Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT tumors. **(K)** Representative images of TSA staining of lung metastases from Krt5-CreERT/Rosa26-mTmG/MMTV-PyMT and Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT mice. Samples were stained with Krt8 (purple), Krt14 (white), GFP (green), and DAPI (blue). White arrows point to Krt14-expressing cells.



	Krt5-mTmG/ MMTV-PyMT	Krt8-mTmG/ MMTV-PyMT	Krt8-mTmG/MMTV-PyMT (Surgically resected tumor)
Number of mice	2	3	1
Total mets	7	17	34
Mets expressing GFP	0/7	16/17	34/34
Mets expressing Krt8	7/7	17/17	34/34
Mets expressing Krt14	4/7	6/17	20/34

Figure 8: MMTV-PyMT as a model of low-ER mammary tumor. **(A)** ER expression in MMTV-PyMT tumors. **(B)** Representative IHC images of ER expression in MMTV-PyMT tumors. **(C-D)** Optimization of tamoxifen induced GFP labelling in Krt5-CreERT/Rosa26-mTmG **(C)** and Krt8-CreERT/Rosa26-mTmG **(D)** mouse mammary gland. **(E)** Flow cytometry gating strategy for identifying GFP expressing luminal and basal tumor cells. **(F)** Quantification of GFP expressing luminal and basal tumor cells from TSA stained images

of Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT tumors. **(G)** Counts of metastatic colonies from Krt5-CreERT/Rosa26-mTmG/MMTV-PyMT and Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT mice expressing Krt8, Krt14, and GFP.

2.4.6 Single cell RNA sequencing reveals SOX10 as a key driver of luminal-to-basal plasticity

To identify genetic drivers enabling the emergence of lineage plasticity, we carried out single-cell RNA sequencing to compare the gene-expression profiles of lineage-restricted luminal tumor cells and luminal-derived basal tumor cells. Tumors were harvested from two Krt8-CreERT/Rosa26-mTmG /MMTV-PyMT mice and sorted by flow cytometry (Fig 10A) for all GFP-expressing cells to specifically analyze both luminal- and basal-like tumor cells of luminal origin. Dimensionality reduction with uniform manifold approximation and projection (UMAP) revealed the presence of a large, connected cluster of cells, in addition to a smaller cluster that showed clear separation from the larger cluster. Unsupervised clustering identified a total of 18 clusters (Fig 9A), with similar results observed in the two mouse replicates (Fig 10B). Using previously described gene signatures[102], individual cells were scored to quantify their activity of luminal-progenitor, mature-luminal, and basal gene expression programs. Most cells scored highly for the luminal-progenitor signature (Fig 10C), in agreement with previous single-cell RNA analyses of MMTV-PyMT tumors[114], however cells from multiple clusters demonstrated concomitant luminal-progenitor and basal signatures (Fig 10D), suggesting that tumor cells do not fully establish a basal identity, and instead express a combination of both luminal progenitor and basal markers. Cluster 6 demonstrated the greatest combined luminal and basal signature scores, suggesting these cells likely express a lumino-basal phenotype (Fig 9B and 9C). Cluster 13 demonstrated high scores specifically for the mature luminal signature, while clusters 15 and 16 demonstrated high scores specifically for the basal signature, suggesting that these clusters contain luminal- and basal-like tumor cells, respectively (Fig 9B, 9C, and 10E).

To identify genes contributing to this lumino-basal plasticity, we compared the genes expressed in the lumino-basal cell cluster (cluster 6) against their possible ancestors, the lineage-restricted luminal cell (cluster 13). Analysis of the differentially expressed genes between the cluster 6 and cluster 13 uncovered upregulation of basal marker genes (*Krt5*,

Krt14, *Krt17*, and *Acta2*) (Fig 9D, light blue points) and simultaneous downregulation of luminal marker genes (*Krt7*, *Krt8*, *Krt18*, *Krt19*, and *Foxa1*) (Fig 9D, dark blue points) in cluster 6, supporting the proposed lumino-basal identity of this cluster. Of note, mesenchymal genes *Mmp2*, and vimentin (*Vim*) (Fig 9D, yellow points) were upregulated in cluster 6, while epithelial genes E-cadherin (*Cdh1*) and Epcam (Fig 9D, orange points) were downregulated, suggesting that the emergence of basal-like characteristics in these luminal-derived tumor cells may result from an epithelial to mesenchymal transition (EMT).

From the list of upregulated genes in cluster 6, we selected 3 that potentially regulate the luminal-to-basal plasticity observed in low-ER tumors (Fig 9D, red points). The gene *Spp1*, which encodes the protein osteopontin (*Opn*), was selected as it has the highest effect-size (avg_log2FC=5.42). Osteopontin is usually found as a component of bone, as it is an extracellular structural protein; however, intracellular osteopontin has been found to regulate EMT[115] via its interaction with the stemness marker *CD44*[116]. We also selected the gene *Postn*, which encodes periostin, an extracellular matrix protein that has been found to enable cell motility by binding to integrins $\alpha_v\beta_3$ and $\alpha_v\beta_5$[117]. Finally, we selected the transcription factor *Sox10*, as it has been implicated in the regulation of cell plasticity in mammary tumors[118]. Furthermore, three *Sox10* targets, *Nes*, *Mia* and *Pmp22*, identified using TRRUST v2 (Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining)[119], were found to be upregulated in cluster 6 as well (Fig 9D, green points), further supporting the role that this transcription factor may play in driving plasticity.

To assess the roles of each candidate gene on lumino-basal plasticity in low-ER tumors, we analyzed the expression of these genes in the ERα low vs ERα high stratified tumors obtained from TCGA (Fig 5). The expression of all 3 genes were found to be significantly higher in the ERα low group of tumors, as compared to the ERα high group (Fig 9E). We further analyzed protein expression of these potential plasticity drivers in the luminal B

vs low-ER tumor TMAs by using IHC staining. Periostin expression was mostly relegated to the stroma, with no expression in tumor cells (Fig 10F). Osteopontin could also be detected in the stroma, while tumor cells staining positive for this protein appear to be present on the edges of the tumor (Fig 10G); however, there was no difference in osteopontin expression between tumors in the luminal B or the low-ER TMAs (Fig 10H). On the other hand, SOX10 appears to have a more direct correlation to low-ER tumors (Fig 9F), with high (>10% of tumor cells) expression of SOX10 observed in 47.47% of low-ER tumors, while only 1 of the luminal B tumors (4.76%) expressed high levels of SOX10. 66.67% of luminal B tumors stained negatively for SOX10, compared to only 38.84% of low-ER tumors (Fig 9G), further indicating that SOX10 is more highly expressed in low-ER tumors, as compared to luminal B tumors. This suggests that SOX10 could be a potential driver of luminal-to-basal plasticity, expressed in both the MMTV-PyMT mouse tumors and in low-ER patient tumors.

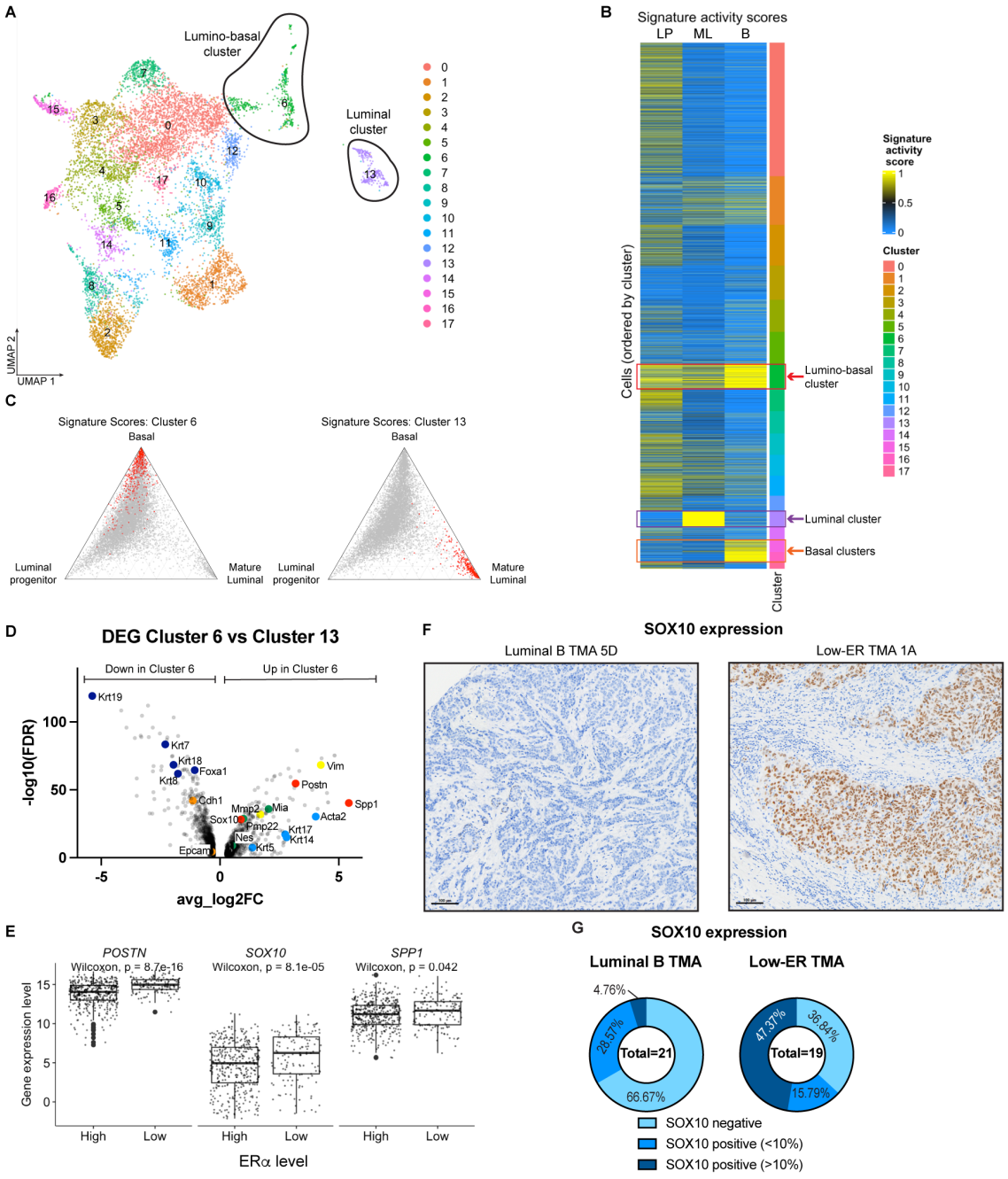


Figure 9: Sox10 is a potential driver of luminal-to-basal plasticity. (A) UMAP of Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT tumor cells sorted for GFP. Unsupervised clustering divided the cells into 18 different clusters. **(B)** Heatmap showing the basal, mature luminal, and luminal progenitor gene signature activity in each cluster. Cluster 13 (purple box) has high expression of mature luminal genes, while clusters 15 and 16 (orange box) have high expression of basal genes. Cluster 6 (red box) has high expression of both

mature luminal and basal genes, and is identified as the lumino-basal cluster. **(C)** Ternary plots showing the distribution of relative activity of luminal progenitor, mature luminal, and basal gene signatures across all tumor cells profiled by scRNA-seq. In the left plot, cells assigned to cluster 6 are highlighted in red, while cells assigned to cluster 13 are highlighted in the right plot. **(D)** Volcano plot showing genes upregulated and downregulated in cluster 6 as compared to cluster 13. Light blue points are basal marker genes, dark blue points are luminal marker genes, yellow points are mesenchymal genes, orange points are epithelial genes, red points are potential plasticity driver genes, and green points are SOX10 target genes. **(E)** Boxplots showing the differences in distribution of *POSTN*, *SOX10*, and *SPP1* gene expression between the ERα low and ERα high groups of tumors obtained from TCGA. **(F)** Representative images of SOX10 IHC staining in luminal B and low-ER TMAs. **(G)** Quantified SOX10 expression in luminal B and low-ER TMAs.

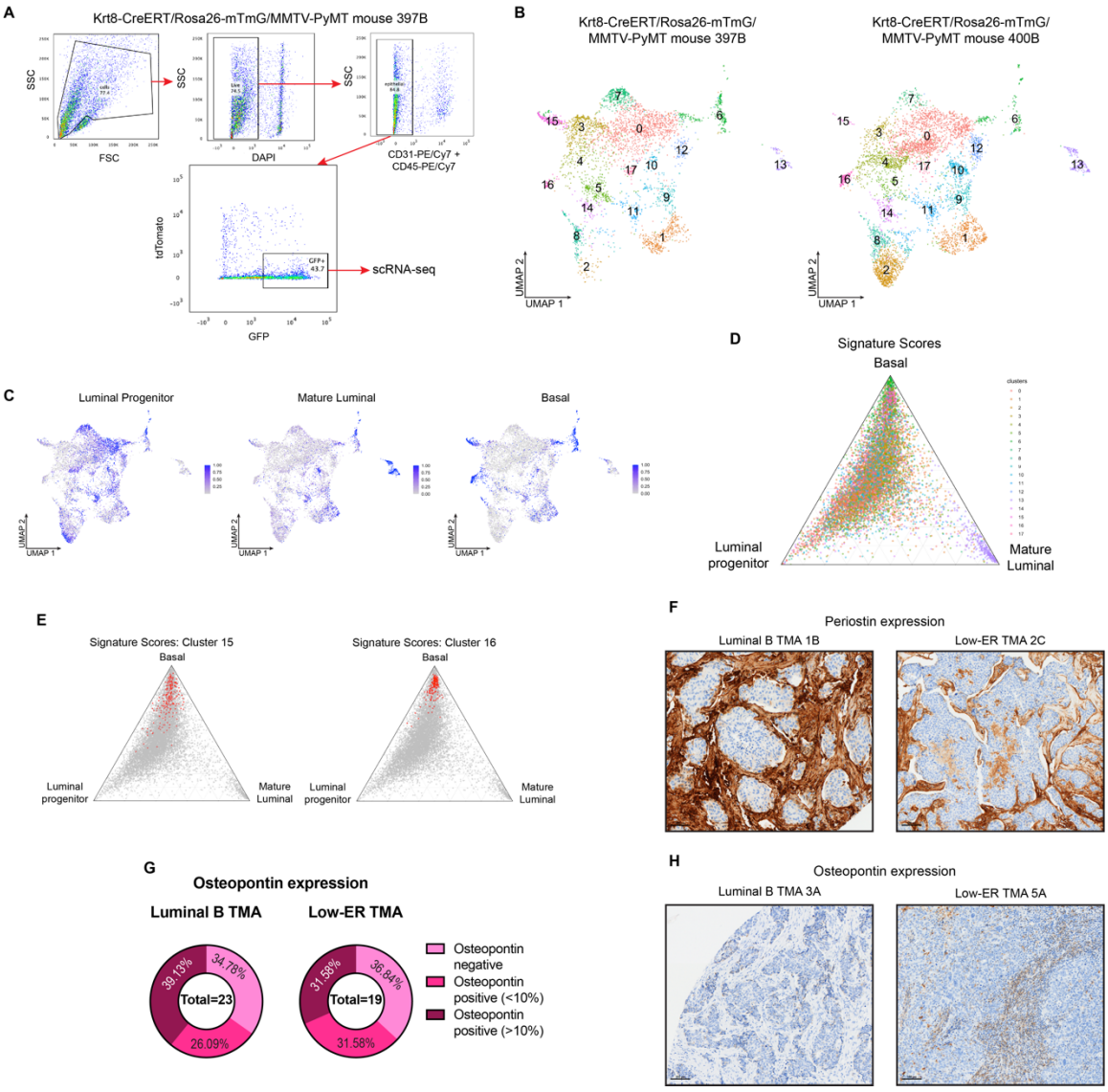


Figure 10: Osteopontin and Periostin expression are not correlative with a low-ER phenotype. (A) Flow cytometry plots showing the gating strategy and the GFP positive cell population harvested using fluorescence assisted cells sorting (FACS). **(B)** UMAP of Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT tumor cells sorted for GFP and split by sample of origin. Unsupervised clustering divided the cells into 18 different clusters. **(C)** UMAP projections of all tumor cells profiled with scRNA-seq. Cells are colored based on activity scores of luminal progenitor, mature luminal, and basal gene signatures. **(D)** Ternary plot showing the relative activity of luminal progenitor, mature luminal, and basal gene signatures across all tumor cells, colored by unsupervised cluster. **(E)** Ternary plots showing the distribution of relative activity of luminal progenitor, mature luminal, and

basal gene signatures across all tumor cells profiled by scRNA-seq. In the left plot, cells assigned to cluster 15 are highlighted in red, while cells assigned to cluster 16 are highlighted in the right plot. **(F-G)** Representative images of periostin **(F)** and osteopontin **(G)** IHC staining in luminal B and low-ER TMAs. **(H)** Quantified osteopontin expression in luminal B and low-ER TMAs.

2.5 Discussion

Our study demonstrates that low-ER breast carcinomas represent a distinct subset of luminal-like tumors, and should be classified as a separate tumor subtype from luminal A and B tumors for the purposes of therapy. We found that cells within the low-ER tumors undergo luminal-to-basal plasticity, which introduces lumino-basal heterogeneity, allowing them to gain basal-like characteristics. These findings potentially reframe the current intrinsic subtypes of breast cancer as connected to each other, instead of being distinct subsets, and that the different subtypes could essentially represent different stages of evolutionary progression of breast tumors as they deviate from their lineage-of-origin. This lineage divergence may influence breast cancer diagnosis and treatment, as luminal-like tumors, which typically exhibit better response to therapy, may evolve into more basal-like counterparts in response to drug-induced evolutionary pressures that could selectively eliminate their luminal-like ancestors while sparing basal-like tumor cells.

Lineage plasticity has also been previously described in other breast carcinoma studies. Li et al[120] found basal marker expression in luminal lineage-labelled cells, although they did not carry out further investigations on this observation. Hein et al[121] observed that oncogenic transformation of the luminal lineage resulted in a small percentage of tumor cells co-expressing Krt5, however, their model utilized HA-tagged Polyoma Middle T antigen (PyMT) and ErbB2 oncogenes to both induce transformation and label the luminal lineage, whereas our lineage-tracing strategy uncoupled the process of oncogenesis from lineage labelling. The use of the MMTV promoter to drive the expression of the PyMT oncogene allowed this protein to be expressed in both luminal and basal cells, as previously shown[114], while lineage labelling using keratin-driven *Cre* recombination, allows specific labelling of either luminal or basal cells. Consistent with our observations, they also found Krt5 expression to be restricted to the edges of the tumor. Finally, Koren et al[122] and Van Keymeulen et al[123] noted that PIK3CA mutations could induce oncogenesis and multipotency within mammary cells, resulting in heterogeneous, multi-

lineage tumors. These findings suggest that oncogenesis and lineage plasticity can occur simultaneously by the activity of various oncogenes. While we have identified SOX10 as the potential driver for lineage plasticity, it is unknown whether this transcription factor itself is activated by PyMT the driver oncogene used in our model. PyMT is known to induce oncogenic transform via interacting with, and activating c-Src[124], a non-receptor tyrosine kinase, thereby activating various other signaling molecules such as Shc[125] and PI3K[126]. It is thus possible that one of the cell signaling pathways activated may lead to an increase in SOX10 transcription and activity, and the PyMT oncogene may indirectly have an effect on SOX10 function. Further studies must be carried out to in order to identify if and how lineage plasticity can occur independently from oncogenesis, as the homogeneous tumor cell populations observed within some luminal-like tumors suggest that oncogenic transformation may not always lead to lumino-basal plasticity. Furthermore, our study showed specific unidirectional luminal-to-basal plasticity, and not simply multipotency, which has important implications when considering the potential evolution from a luminal-like to a basal-like tumor, as mentioned previously.

The presence of basal-like cells within breast tumors has important implications in tumor development and metastasis. We have previously shown that PKA-induced reduction of the lumino-basal subpopulation may be important in limiting metastasis and reducing chemotherapy resistance[114]. The collective migration of breast cancer has also been shown to involve leader cells with a re-activated basal program[127], suggesting that the successful establishment of metastasis by these invading tumor clusters may depend on leader cells that have undergone luminal-to-basal plasticity. It is important to note that, while we are not able to assess the requirement for lumino-basal plasticity in metastasis from our data, we cannot rule out the possibility that the cells that established these metastases could have gained lumino-basal features within the primary tumor, which may have been subsequently lost upon lung colonization. Metastatic dissemination has been shown to involve cells that have undergone partial EMT[128], suggesting that maintaining cellular plasticity is beneficial in metastasis. Circulating tumor clusters have

also been found to consist of multiple clones[129], suggesting that only a small percentage of lumino-basal cells may be required to successfully colonize distant sites. Further studies analyzing circulating tumor clusters in lineage-labelled mice, or studying early vs late metastases may help to address the importance of lumino-basal plasticity in metastasis. Alternatively, lineage ablation experiments can be carried out using mouse models with inducible and lineage-specific diphtheria toxin, to eliminate the lumino-basal population, which could assess if metastases can still develop without lumino-basal tumor cells.

Although the MMTV-PyMT mouse is an appropriate mammary tumor model that is commonly used in the study of breast carcinomas[113], the oncogenic mechanism and development of these murine mammary tumors may not fully reflect the typical progression of human carcinomas. The use of this model in our lineage tracing experiments may thus be a possible limitation in attempting to elucidate the origin of lumino-basal heterogeneity within low-ER tumors. The model, however, represents the closest to modeling low-ER breast cancers and to understand lineage plasticity and lumino-basal heterogeneity. Besides the use of this specific mouse model, alternative models may also be useful in attempting to study lumino-basal heterogeneity. Specifically, the *Brca1*$^{F22–24/F22–24}$;*p53*$^{+/–}$ mouse tumor model[130] more accurately reflects mammary epithelial transformation in human patients by introducing *Brca1* mutation and loss of *p53*. When crossed with *BLG-Cre* mice to induce oncogenesis in milk-producing luminal cells, this mouse model was shown to produce tumors that are basal-like and metaplastic. Substitution of *BLG-Cre* with Krt8- or Krt5-driven Cre could enable oncogenic transformation and lineage labelling to occur simultaneously. Lineage labelling of luminal progenitor cells in mice using ELF5-rtta in combination with TetO-Cre[131] could also help to address whether lumino-basal plasticity can specifically be observed in this luminal subpopulation. Finally, in order to confirm if lumino-basal plasticity is a crucial step in tumor development, lineage ablation experiments can also be carried out to assess whether eliminating the lumino-basal subpopulation would interrupt tumor development.

We have uncovered that SOX10 may be responsible for driving the lumino-basal plasticity seen in the low-ER tumors. This is in agreement with previous studies demonstrating the role of SOX10 in regulating cell-state plasticity in mammary tumors[118]. Interestingly, several recent studies have shown that SOX10 is preferentially expressed in triple negative and metaplastic breast carcinomas and has emerged as a useful immunohistochemical marker to utilize in breast pathology practice[132,133]. In addition, SOX10 is associated with developmental plasticity and bipotent progenitor identity in fetal mammary stem cells, suggesting that the activity of this transcription factor reflects the reactivation of the bipotent progenitor program in tumor cells. SOX10 has been shown to be expressed in TNBCs, and is associated with worse clinical outcomes in these patients[134], highlighting the similarities between this tumor subtype and the low-ER tumors in our study. SOX10 has also been shown to induce dedifferentiation and EMT[118], highlighting the increase in invasive potential of the basal-like progression of low-ER tumors, potentially leading to the worse prognosis and poor clinical outcomes observed in these patients.

Chapter 3: PKA activation curtails cellular plasticity by reducing basal tumor lineage

Adapted from "Limiting Self-Renewal of the Basal Compartment by PKA Activation Induces Differentiation and Alters the Evolution of Mammary Tumors", Ognjenovic *et al* 2020, available on Developmental Cell (https://www.sciencedirect.com/science/article/pii/S1534580720307942)

Experiments for Figures 11, 12A-D, and 13 were performed by Gadisti Aisha Mohamed and Nevena Ognjenovic

Experiments for Figures 12E-H were performed by Meisam Bagheri

Experiments for Figure 14 were performed by Diwakar Pattabiraman with computational analysis performed by Mohamed Ashick

3.1 Introduction

The epithelial-mesenchymal transition (EMT) is a cell developmental program which activates cellular plasticity to allow epithelial cells to gain a mesenchymal phenotype[25]. It is activated during embryonic development, along with the reverse process, mesenchymal-epithelial transition (MET), allowing cells to differentiate and give rise to the various cell phenotypes in the adult tissues and organs[70]. EMT is also important in post-embryonic wound healing, where Snai2[135] and fibroblast growth factor 2 (FGF2)[136] activate the EMT program in keratinocytes, allowing them to gain a mesenchymal-like phenotype in order to move over and cover wound surfaces[135,136]. In cancer, EMT plays a role in initiating metastasis by allowing tumor cells to escape the site of the primary tumor and colonize distant organs[25,70]. EMT plasticity promotes intra-tumoral heterogeneity, as some cells gain a mesenchymal phenotype while others remain epithelial. This heterogeneity thus allows tumors to disseminate as circulating tumor cell clusters (CTC clusters) containing a mixed population of cells expressing both epithelial and mesenchymal phenotypes[137]. CTC clusters promote metastasis by collective migration of both mesenchymal cells that allow intravasation and survival in the bloodstream, and epithelial cells that promote seeding in epithelial tissues of distant organs[138].

Previous studies have shown that treatment of mesenchymal tumor cell lines with forskolin and cholera toxin is able to inhibit or reverse EMT, thus preventing metastasis,

by elevating cyclic AMP (cAMP) levels and activating protein kinase A (PKA)[139]. The inactive PKA holoenzyme consist of 2 catalytic (C) subunits bound to 2 regulatory (R) subunits. Activating PKA involves the binding of cAMP to 2 regulatory sites each on the R subunit, thus releasing the C subunits and allowing them to phosphorylate substrate targets[140]. PKA has been shown to inhibit GLI and YAP1, both transcription factors involved in epidermal stem cell maintenance, and PKA overactivation results in the exhaustion of the stem cell compartment[141]. This suggests that PKA activation may disrupt cellular plasticity, and that the effects of EMT inhibition by PKA seen in the mesenchymal tumor cell lines may be due to this disruption. While the function of PKA in epidermal maintenance has been identified, the effects of PKA activity in the mammary context is unclear.

In this study, transgenic mouse models expressing constitutively active PKA were utilized to study the effects of PKA in both the normal and transformed mammary gland. The *Prkaca*$^{C\alpha R}$ mice express a mutant PKA C subunit with a tryptophan to arginine substitution on amino acid 196, which prevents it from binding to the R subunit[142,143]. The expression of the PKA-CαR mutant is cre recombinase-dependent, which allows its targeted expression in specific tissues[143]. In order to specifically express these mutants in the mammary gland, the *Prkaca*$^{C\alpha R}$ mice are crossed with MMTV-Cre mice, which express the cre recombinase specifically in mammary epithelial cells[144]. This would permit the study of PKA activation on the development of the normal mammary gland. To elucidate the effects of PKA activity on tumor development, progression, and metastasis, the *Prkaca*$^{C\alpha R}$/MMTV-Cre mice were additionally crossed to the MMTV-PyMT mouse[92], generating mammary tumors that also express the PKA-CαR constitutively active mutant.

3.2 Materials and Methods

Animal studies

All animal experiment IACUC protocols were approved by the Dartmouth College Committee on Animal Care. The PKA-CαR[143] mice were a gift from Stanley McKnight

(UWashington). Tg(MMTV-cre)4Mam/J line D (Strain #: 003553)[144], and MMTV-PyMT mice ((Tg(MMTV-PyVT)$^{634Mul/LellJ}$ mice on a C57Bl/6J background, strain #: 022974)[92] were purchased from The Jackson Laboratory. Mice were euthanized and tumors were harvested once tumors reached a volume of 1.5cm^3, usually at weeks 15-18. For analysis of the normal mammary gland, mice were euthanized at 8 weeks.

Mammary gland dissociation

Mouse mammary fat pads were harvested and processed to obtain single-cell suspensions using established protocols[94] that were slightly modified. Mammary fat pads were digested in a solution of DMEM (Corning, 10-013-CV) with Hyaluronidase (Fischer Scientific, ICN10074091) and Collagenase A (Sigma-Aldrich, 10103586001) for 2 hours at 37°C with gentle agitation using a rotator. Red blood cells were subsequently removed with an ammonium chloride lyse (8.02g NH_4Cl, 0.84g $NaHCO_3$, 0.37g EDTA in 1L of water), and samples were agitated with Trypsin (Corning, 25-053-CI) and Dispase (Stem Cell Technologies, 7913) + DNAse I (Sigma-Aldrich, DN25-100mg) for 1 minute each to further dissociate the cells. Finally, samples were filtered through a 40mm cell strainer (Corning, 431750) to obtain a single-cell suspension.

Mammary gland whole mount preparation and Carmine Alum staining

Whole mammary glands were spread on a glass slide and fixed with Carnoy's fixative (60% ethanol, 30% chloroform, 10% glacial acetic acid) overnight at RT. Fixed tissue was rehydrated by washing with decreasing ethanol concentrations (70%, 50%, 30%, 10%) 2 times each for 10 minutes. Rehydrated tissue was then stained with Carmine Alum (Stem Cell Technologies, 07070) for 48-72 hours. Mammary glands were then dehydrated using increasing ethanol concentrations (70%, 95%, 100%) 2 times each for 15 minutes, and cleared in xylene overnight. Cleared mammary glands were then mounted with Permount mounting medium (Fischer Chemical, SP15-100) and glass coverslips and allowed to dry overnight. Slides were imaged on the PerkinElmer Vectra3 slide scanner.

Tumor dissociation

Tumors harvested from euthanized mice were digested in DMEM containing 2 mg/ml Collagenase A and 100U/ml hyaluronidase at 37°C for 2 hours with gentle agitation using a rotator. Following digestion, samples were strained through 70mm (Corning, 431751) and 40mm cell strainers to obtain a single-cell suspension. Finally, red blood cells were removed with an ammonium chloride lyse, and cells were washed in PBS.

FFPE tissue processing

Harvested mammary glands, tumors and lungs were placed in tissue biopsy cassettes and fixed in 10% Neutral Buffered Formalin (Leica, 3800598) at 4°C overnight. The formalin was then removed, and tissues were soaked in 70% ethanol at 4°C for at least 2 days before embedding in paraffin blocks. Hematoxylin & Eosin (H&E) staining was performed on sections cut from the paraffin blocks. Embedding, sectioning, and H&E staining were performed by Dartmouth-Hitchcock Pathology Shared Resources.

Flow Cytometry and Fluorescence assisted cell sorting (FACS)

Single-cell suspensions were first stained with fluorescently labelled antibodies. Tumor single-cell suspensions were stained with APC anti-CD326 (Ep-CAM) antibody (Biolegend, 118214, clone: G8.8, 1:100 dilution), PE anti-CD49f (integrin alpha 6) antibody (Biolegend, 313612, clone: GoH3, 1:100 dilution), PE/Cyanine 7 anti-mouse CD31antibody (Biolegend, 102418, clone:390, 1:100 dilution), and PE/Cyanine 7 anti-mouse CD45 antibody (Biolegend, 103114, clone: 30-F11, 1:100 dilution) for 30 minutes on ice. DAPI (Sigma-Aldrich, 10236276001) was added at a dilution of 1:1000 after the final wash step in order to facilitate live-cell sorting.

Cells were sorted on the FACSAria III cell sorter for normal basal and luminal cells by first gating on DAPI-negative live cells, and CD31- and CD45-negative epithelial cells. Cells that were EpCam high and CD49f low were collected as luminal cells, while cells that were EpCam low and CD49f high were collected as basal cells.

Organoid assay

Dissociated cells from the mammary epithelium were cultured with advanced DMEM (Gibco, 12491015) with 5% Matrigel, 5% heat-inactivated FBS, 10 ng/ml EGF, 20 ng/ml bFGF, 4 mg/ml heparin, and 5 mM Y-27632. Cells were seeded at 1000 cells per well of a 96-well ultralow attachment plate (Corning, 29443-034). The number of organoids formed were counted 7-14 after seeding.

Immunohistochemistry (IHC) staining

Slides are cut at 4mm and air dried at RT before baking at 60°C for 30 minutes. Automated protocol performed on the Leica Bond Rx (Leica Biosystems) includes paraffin dewax, antigen retrieval, peroxide block and staining. Heat induced epitope retrieval using Bond Epitope Retrieval 2, pH9 (Leica Biosystems, AR9640) was incubated at 100 degrees Celsius for 20 minutes. Primary antibody anti-p-PKA substrate (Cell Signaling Technoloy, 9624, 1:200 dilution) was applied and incubated for 15 minutes at room temp. Primary antibody binding is detected and visualized using the Leica Bond Polymer Refine Detection Kit (Leica Biosystems DS9800) with DAB chromogen and Hematoxylin counterstain. Slides were imaged using the PerkinElmer Vectra3 slide scanner, and PhenoChart.

Immunofluorescence staining

Slides were rehydrated by incubating in Histoclear solution twice for 5 min each, followed by incubation in 100% ethanol twice for 5 min each, in 95% ethanol twice for 5min each, 70% ethanol twice for 5 min each, once in 35% ethanol for 5 min, and in water for 5 min. Pressure cooker-mediated heat-induced epitope retrieval was carried out in 250 ml of unmasking buffer containing sodium citrate at pH 6. After retrieval, slides were blocked for 30 min in PBS containing 3% normal horse serum after which they were incubated with primary antibody in blocking solution overnight at 4 C. Slides were washed twice with PBS and incubated with secondary antibody at room temperature for 1 hour in the dark. After two PBS washes, 20 ml of mounting medium was added, then slide contents

were topped with coverslips, and stored in the dark for 24 hours before imaging on a Zeiss LSM800 microscope and analyzed using the Zen Digital Imaging software.

Single-cell RNA-seq Library Construction and Sequencing

Single Cell Capture and Library Preparation: Immediately following dissociation, single cell suspensions were placed on ice and counted on a Luna automated cell counter. Cell concentrations from each sample were be normalized to 1000 cells/ul and loaded onto a Chromium Single Cell A Chip (10x Genomics Inc.) targeting a capture rate of 5,000 cells per sample. Single cell RNA-seq libraries were prepared using the Chromium Single Cell 3' v2 kit (10x Genomics) following the manufacturer's protocol. Libraries were quantified by qubit and peak size determined on a fragment analyzer instrument. All libraries were pooled and sequenced on an Illumina NextSeq500 High Output 26bp x 98bp run to generate an average of 50,000 reads/cell. Data Analysis: Raw sequencing were processed using the 10x Genomics Cell Ranger to generate quality metrics and primary data visualizations as well as gene expression matrices for downstream analysis in R using Seurat and other open-source packages.

Single-cell RNA-seq Data Processing

To investigate the transcriptional evolution of tumors at single cell level, we performed matched hyperplasia and tumor samples single-cell RNA-seq on the 10X genomics platform, and generated data at an average of ~78M reads per sample with cell numbers ranging from ~2600 to 5000. Single-cell transcriptome sequencing raw reads were quality filtered using *Fqtrim* (v0.9.7) tool[145]. Reads were trimmed and filtered for low quality bases, poly-A/T tails and N bases while retaining the paired-end integrity of the reads. Read1 of the pair containing barcodes was not considered for trimming but allowed to be filtered to maintain paired-end integrity. Sequencing and PCR errors in cell barcodes can convolute the process of differentiating reads per cell barcode; hence we used *UMItools* (v0.5.4) to distinguish the reads per cell (barcode) accommodating for technical errors[146].

A knee density method-based approach is used in *UMItools* to estimate the number of acceptable cell barcodes and then reads were assigned for each cell.

Clean and barcode classified reads were aligned against Mouse genome reference (GRCm38) using *STAR* aligner (v2.5.3a)[147] and output was restricted to uniquely aligned reads. De-duplication of transcripts with the same UMIs arising from PCR amplification were removed. Reads were assigned with position based annotation of genomic features using *featureCounts* module[148] in the *subread* package (v1.6.0). Taking advantage of the UMI information, read counts were extrapolated to quantify molecular level count for each transcript using directional-adjacency method-based count module in *UMItools*. Reads were grouped per cell (based on barcode) and then a gene expression matrix was generated with RNA molecule count of genes in rows for each cell in columns represented in GxC matrix format (where G is gene in rows and C is cells in columns).

Cell-Type Classification

Gene expression matrix was filtered for cells with <500 genes expressed and <500 total UMI counts genes and >0.25 percentage of reads aligned to Mitochondrial genome using Monocle2 (v2.6.4) R package[149–151]. Further, outliers of total mRNAs count for each cell was removed from the downstream analysis. Expression counts were normalized using negative binomial distribution of library sizes. Genes with mean expression value of 0.1 across cells were used for PCA based dimensionality reduction and then clustered using an unsupervised density Peaks algorithm in Monocle2 and projected using t-SNE method. To classify the population into various cell-types, we investigated the expression of curated and well-established cell-type specific marker genes' expression in our data and selected a specific list of markers that are expressed and/or not expressed for each cell-type. Following are the cell-type specific markers used for the classification in this study.

Cell type	Higher expression
Basal cells	*Krt14, Krt5, Snai2*
Luminal mature cells	*Pgr, Prlr, Foxa1, Gpx3, Esr1*
Luminal Progenitor cells	*CD14, Kit, Lalba*
Lumino-basal	*Krt14, Krt8, CD14*
EMT-like	*Trim29, Vim, Cldn6, Sox4*

Table 1: Specific markers used to classify cells in scRNA-seq data

3.3 Results

3.3.1 PKA activation leads to quantitative and qualitative changes in the normal mammary basal compartment

In order to study the effect of PKA activation in the development of the normal mammary gland, *Prkaca*$^{C\alpha R}$ mice were crossed with MMTV-Cre mice, which specifically expresses the Cre recombinase in the cells in the mammary gland, thus allowing the expression of the CαR subunit of PKA to be targeted to the mammary epithelial cells (Fig 11A). Mice that inherited only *Prkaca*$^{C\alpha R}$ or MMTV-Cre individually were used as controls (*Prkaca*^{WT}). Mammary glands were harvested after the onset of puberty at around 8 weeks. PKA activation appears to disrupt normal mammary gland development with interrupted ductal tree formation in *Prkaca*$^{C\alpha R}$ mammary glands (Fig 11B). A reduction in the number of Terminal Ductal Lobular Units (TDLU) (Fig 11C) was also observed, with a 65% reduction in the average number of TDLUs in *Prkaca*$^{C\alpha R}$ compared to *Prkaca*^{WT} mice (Fig 11D). This demonstrates that PKA activation prevents the mammary gland from developing normally, severely restricting the formation of the network of branched ducts that represent the typical structure of a normal mammary gland.

To investigate if PKA activation affects basal and luminal cells of the mammary gland equally, flow cytometry was performed to analyze the proportion of mammary epithelial cells in *Prkaca*$^{C\alpha R}$ and *Prkaca*^{WT} mammary glands. The flow cytometry analysis showed a reduction in the proportion of basal epithelial cells in *Prkaca*$^{C\alpha R}$ mice, with only 5.52% basal cells in the mammary gland, as compared to 11.9% basal cells in the *Prkaca*^{WT}

mammary gland, an almost 50% reduction (Fig 11E and F). In contrast, luminal cell proportions seem to be relatively consistent between the *Prkaca*$^{C\alpha R}$ (6.27%) and the *Prkaca*^{WT} (4.86%) mice. These results suggest the interruption of the mammary gland development by PKA activity primarily affects the basal epithelial lineage rather than the luminal epithelial lineage.

The basal epithelial compartment is known to harbor more mammary stem cells (source), and thus reduction of this epithelial lineage may result in the reduction in repopulating capacity of the mammary gland. To test this, organoid assays were performed, to analyze the ability of basal cells to form organoids in Matrigel. Basal epithelial cells were sorted from *Prkaca*$^{C\alpha R}$ and *Prkaca*^{WT} mice, and allowed to grow in media supplemented with 5% Matrigel, with 1000 cells deposited into each well of a 96-well plate. Basal cells sorted from *Prkaca*$^{C\alpha R}$ mice had a reduced ability to form organoids, with a 4-fold decrease in organoids formed per 1000 cells compared to basal cells from *Prkaca*^{WT} mice (Fig 11G and H). This indicates that PKA activity disrupts the self-renewal capabilities of the basal epithelial compartment, which thus causes the interruption of normal mammary gland formation.

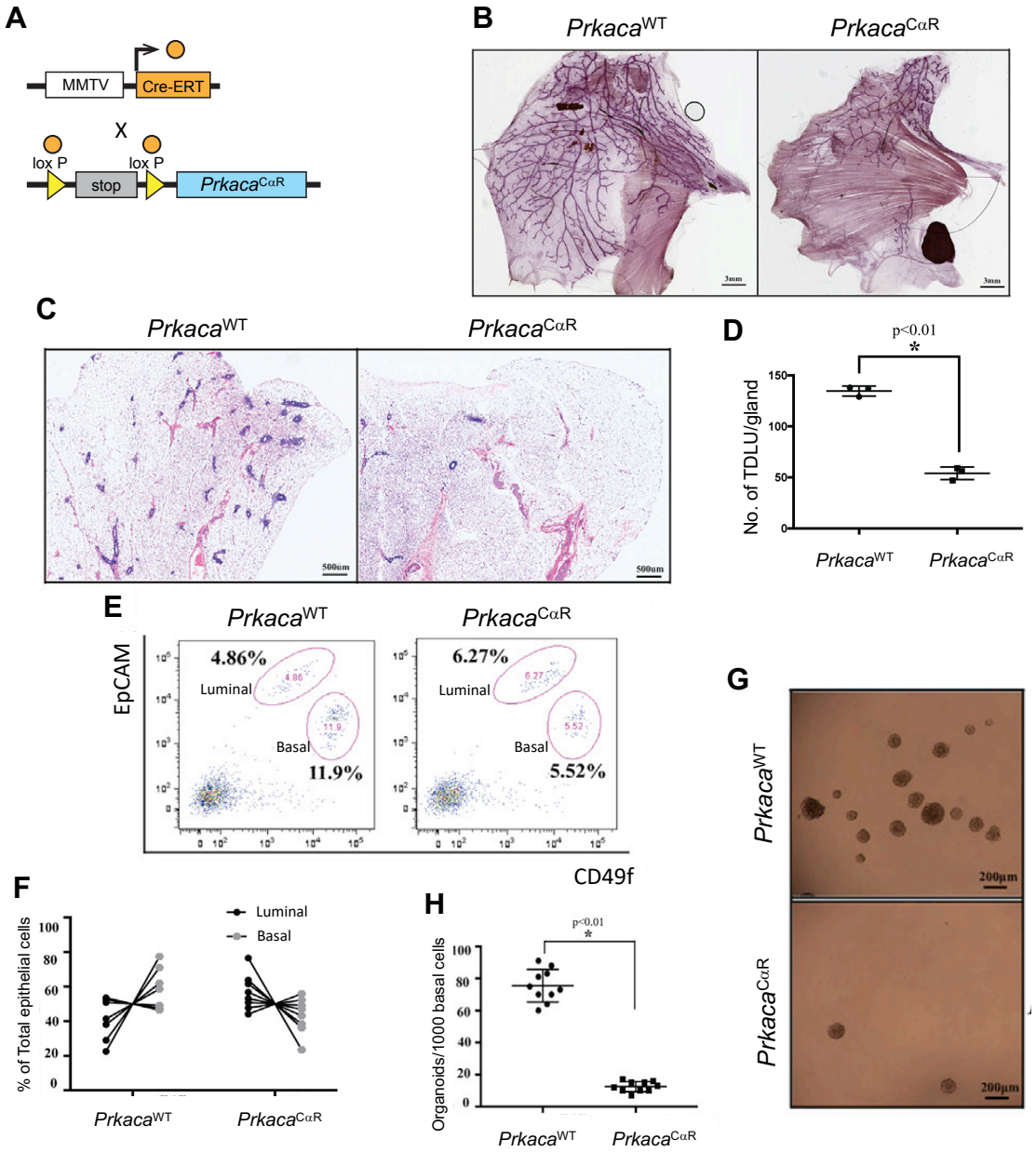


Figure 11: PKA activation reduces basal epithelial proportions and prevents self-renewal of basal epithelial compartment in the normal mammary gland. (A) Genetics of mice used to interrogate PKA activity in the normal mammary gland. *Prkaca*$^{C\alpha R}$ mice were crossed with MMTV-Cre mice to induce PKA activation specifically in the mammary gland. (B) Whole mount mammary gland stained with carmine alum shows interrupted ductal tree formation in *Prkaca*$^{C\alpha R}$ mice compared to in *Prkaca*^{WT} mice. (C) Representative H&E stained images of the mammary gland, showing reduction in the number of TDLUs in

Prkaca$^{C\alpha R}$ mice. (D) Quantification of TDLUs in *Prkaca*$^{C\alpha R}$ and *Prkaca*^{WT} mice. (E) Flow cytometry analysis of luminal and basal cells in the normal mammary gland of *Prkaca*$^{C\alpha R}$ and *Prkaca*^{WT} mice. (F) Proportions of luminal and basal quantified from flow cytometry analysis of of *Prkaca*$^{C\alpha R}$ and *Prkaca*^{WT} mice mammary glands. (G) Representative images of organoid assays to study the organoid forming capability of in *Prkaca*$^{C\alpha R}$ and *Prkaca*^{WT} basal cells. (H) Quantification of organoids formed by in *Prkaca*$^{C\alpha R}$ and *Prkaca*^{WT} basal cells.

3.3.2 PKA induced reduction of basal cells improve tumor prognosis and reduces metastasis

Prkaca$^{C\alpha R}$ and *Prkaca*^{WT} were crossed with MMTV-PyMT mice in order to study the effects of PKA activation in tumor development (Fig 12A). Tumors were allowed to develop until the tumor burden reaches 2cm^2, after which the mice were sacrificed to harvest and analyze the tumors and lung metastases. Histologically, PKA activation appears to affect the differentiation of mammary tumors, with *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumors showing a more differentiated phenotype, gland formation and nuclear characteristics intermediate grade Invasive Ductal Carcinomas (IDC), and vascularized stroma similar to papillary differentiation (Fig 12B). In contrast, *Prkaca*^{WT}/MMTV-PyMT tumors were more de-differentiated, and were comprised of solid sheets of pleiomorphic cells with high mitotic activity, similar to higher grade IDC (Fig 12B). Immunohistochemistry (IHC) staining for phospho-PKA substrates confirmed the activation of PKA in the *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumors, which showed higher levels of activated PKA target proteins (Fig 12C and D).

Prkaca$^{C\alpha R}$/MMTV-PyMT tumors also appear to have reduced basal-like tumor cells and mesenchymal tumor cells compared to the *Prkaca*^{WT}/MMTV-PyMT tumors. Immunofluorescence staining of the *Prkaca*^{WT}/MMTV-PyMT tumors showed most of the tumor cells expressing the luminal marker Krt8, with a proportion of cells expressing the basal marker Krt14 (Fig 12E). In contrast, the *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumors show a reduction in Krt14 staining, while Krt8 staining appears to remain (Fig 12F). This is consistent with the reduction in basal cell proportions in the normal mammary gland as a result of PKA activation. Staining for vimentin, a mesenchymal marker also appears reduced in the *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumors compared to the *Prkaca*^{WT}/MMTV-PyMT tumors (Fig 12G and H). The reduction in both Krt14 and vimentin expression suggest that luminal-to-basal plasticity may be synonymous with epithelial-to-mesenchymal plasticity, and that PKA's effect in the tumor context limits cellular plasticity thus preventing tumor cells from developing and evolving into different phenotypes.

The reduced mesenchymal and basal-like cells in the *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumors may suggest that the altered development of these tumors could also result in changes in tumor prognosis and metastasis. PKA activation appears to improve the overall survival of the mice, with *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumors developing slower, taking as much as 173 days to develop a tumor burden endpoint of $2cm^2$. In contrast, all of the *Prkaca*^{WT}/MMTV-PyMT mice reached the endpoint by day 142, 31 days earlier than the *Prkaca*$^{C\alpha R}$/MMTV-PyMT mice (Fig 13A and B). To evaluate the lung metastatic burden, lungs were sectioned and stained for H&E, which allows visualization of the macrometastatic colonies present in each lung (Fig 13D). Quantification of these macrometastases revealed that *Prkaca*$^{C\alpha R}$/MMTV-PyMT mice developed a lower number of lung metastases, with an average lung metastatic burden of less than 2, almost half as much as the *Prkaca*^{WT}/MMTV-PyMT mice (Fig 13C), despite taking a longer time to reach the same tumor burden (Fig 13A and B). These results show that PKA induced arrest of cellular plasticity interrupts regular metastatic evolution within tumors, thus preventing metastatic colonization and outgrowth and improving tumor prognosis.

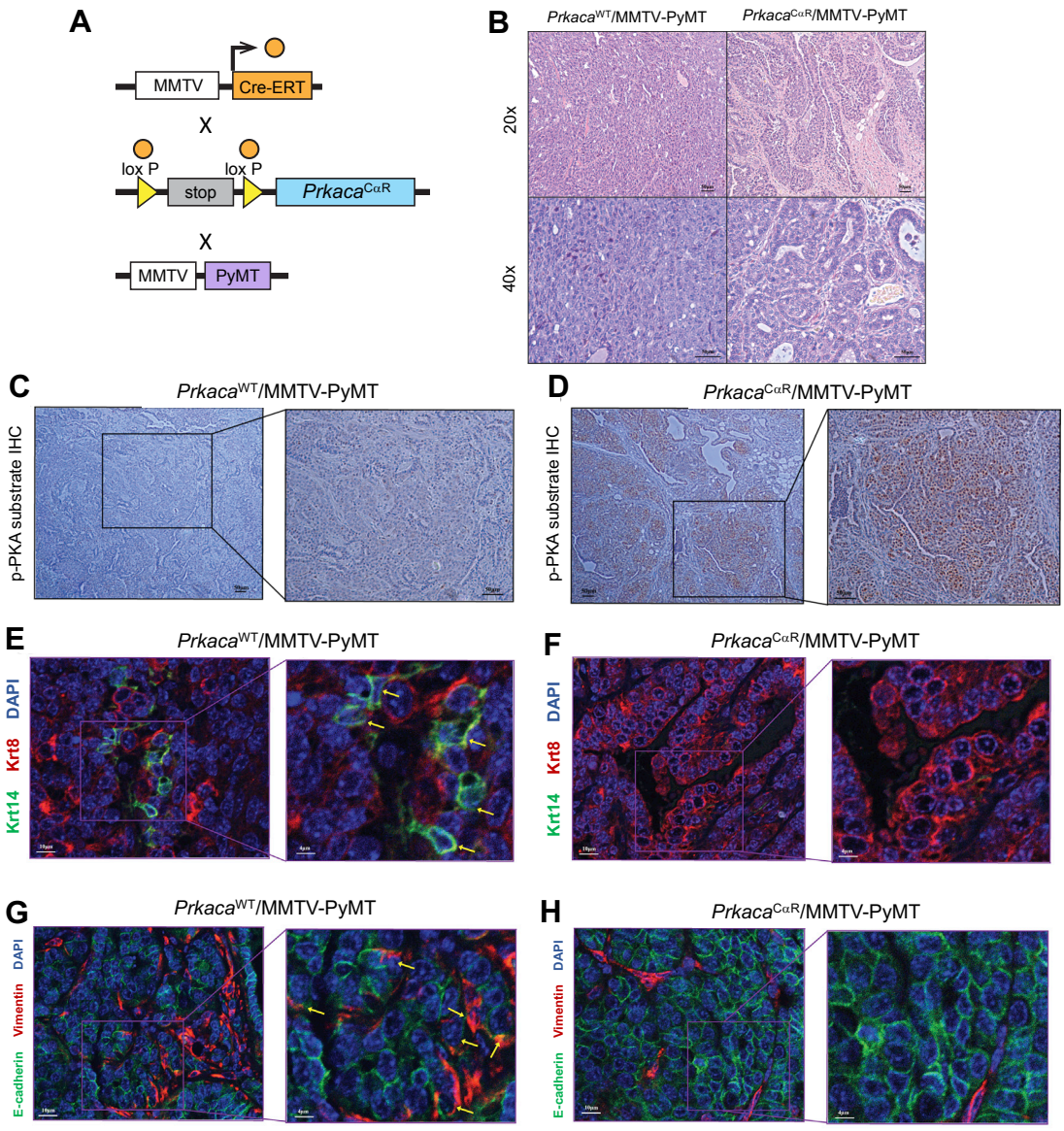


Figure 12: PKA activation prevents the de-differentiation of mammary tumors and results in reduction of basal-like and mesenchymal tumor cells. (A) Genetics of mice used to interrogate PKA activity in mammary tumors. The offspring of *Prkaca*$^{C\alpha R}$/MMTV-Cre crossed mice were further crossed with MMTV-PyMT mice to induce PKA activation and oncogenic transformation specifically in the mammary epithelial cells. **(B)** H&E staining to show the histological differences as a result of PKA activation in mammary tumors. *Prkaca*^{WT}/MMTV-PyMT tumors on the left display characteristics similar to high

grade IDC, while *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumors on the right show a differentiated phenotype, with characteristics similar to intermediate grade IDC. **(C and D)** IHC staining for phosphor-PKA substrates in *Prkaca*^{WT}/MMTV-PyMT **(C)**, and *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumors **(D)** to confirm the increase of PKA activity. **(E and F)** IF staining for luminal marker Krt8 and basal marker Krt14 in *Prkaca*^{WT}/MMTV-PyMT **(E)**, and *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumors **(F)**, showing reduction in Krt14 expression in tumors with increased PKA activity. **(G and H)** IF staining for mesenchymal marker vimentin and epithelial marker E-cadherin in *Prkaca*^{WT}/MMTV-PyMT **(G)**, and *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumors **(H)**, showing reduction in mesenchymal marker expression in tumors with increased PKA activity.

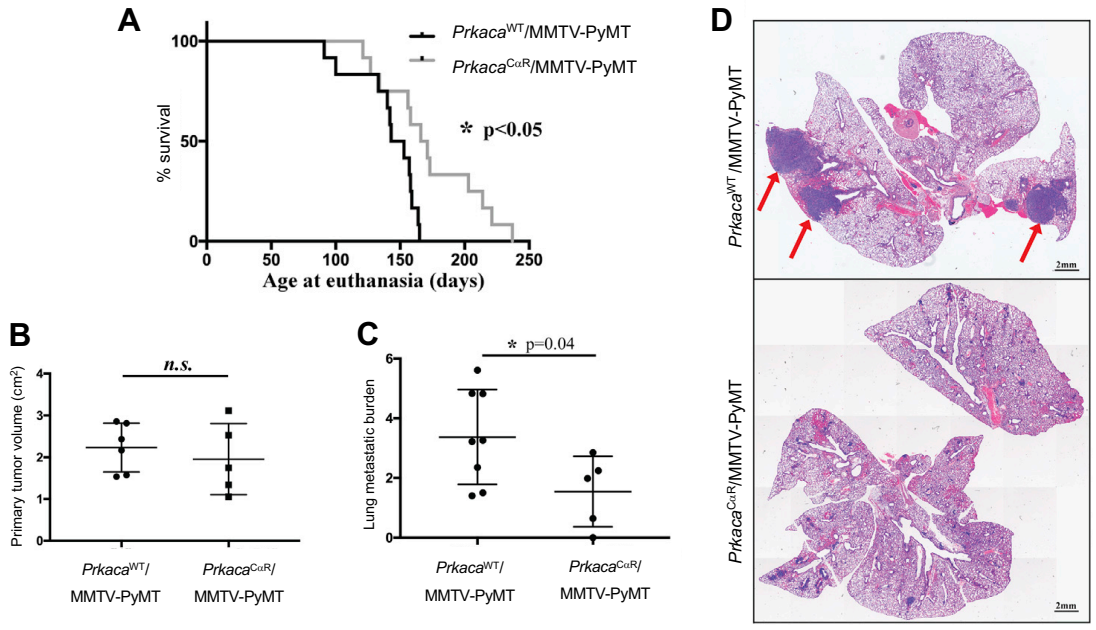


Figure 13: PKA activation improves tumor prognosis and reduces metastasis. (A) Kaplan-Meier survival curve showing improved survival of *Prkaca*$^{C\alpha R}$/MMTV-PyMT mice compared to *Prkaca*^{WT}/MMTV-PyMT mice. Each mouse was sacrificed once the tumors reach a size of about 2cm^2. **(B)** Comparison of tumor volume harvested between *Prkaca*$^{C\alpha R}$/MMTV-PyMT mice to *Prkaca*^{WT}/MMTV-PyMT mice. As each mouse was euthanized once the tumor burden was reached, there is no significant difference in the tumor volume harvestd. **(C)** Quantification of macrometastases per mouse, showing a larger tumor burden in *Prkaca*^{WT}/MMTV-PyMT mice as compared to *Prkaca*$^{C\alpha R}$/MMTV-PyMT mice. **(D)** Representative H&E image of harvested mouse lungs showing larger and more macrometastases in *Prkaca*^{WT}/MMTV-PyMT mice as compared to *Prkaca*$^{C\alpha R}$/MMTV-PyMT mice.

3.3.3 PKA activation alters mammary tumor evolution by limiting tumor cell plasticity

The distinct differentiation states of *Prkaca*^{WT}/MMTV-PyMT and *Prkaca*$^{C\alpha R}$ /MMTV-PyMT tumors (Fig 12B) show that PKA activation affects the development and evolution of mammary tumors. To track the changes in cell phenotypes in these tumors over time, early-stage hyperplastic glands were surgically resected from one of the two 4th inguinal mammary fat pads from *Prkaca*^{WT}/MMTV-PyMT and *Prkaca*$^{C\alpha R}$ /MMTV-PyMT at week 9, allowing the mouse to survive and develop mature carcinomas which were then harvested at week 22. Single-cell RNA sequencing was performed on the cells obtained from these two timepoints, and cell phenotypes were then annotated based on expression of well-known marker genes (Table 1).

Dimensionality reduction with t-distributed stochastic neighbor embedding (t-SNE) revealed hyperplastic tumors harboring mainly three cell phenotypes, basal, mature luminal (ML), and luminal progenitor (LP) (Fig 14A and C). As tumors progress, these tumor cells evolved, leading to the emergence of two additional phenotypes, lumino-basal (LB), and EMT-like (Fig 14B and D). While the proportions of cell phenotypes in *Prkaca*^{WT}/MMTV-PyMT and *Prkaca*$^{C\alpha R}$ /MMTV-PyMT tumors remain relatively similar at the hyperplastic stage, a divergence in the proportions of the LB and EMT-like cells were observed at the carcinoma stage, with a reduction of cells with these two phenotypes observed in *Prkaca*$^{C\alpha R}$ /MMTV-PyMT tumors. This indicates that PKA activation inhibits the evolution of the LB and EMT-like cells by preventing luminal-to-basal and epithelial-to-mesenchymal plasticity.

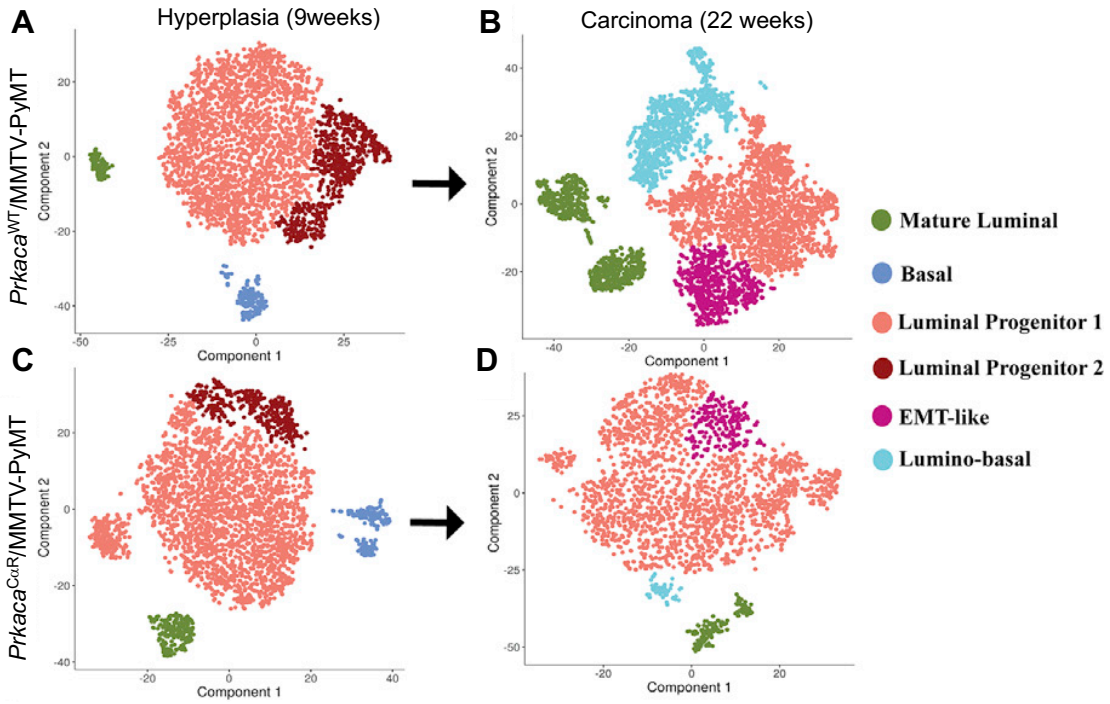


Figure 14: PKA activation alters the evolution of mammary tumors. (A-D) tSNE plots showing the different cell types in hyperplastic glands harvested at 9 weeks **(A and C)**, and mature carcinomas harvested at tumor burden of 22 weeks **(B and D)**. Tumor development appears to be similar at the early hyperplasia stage, with both *Prkaca*^{WT}/MMTV-PyMT **(A)** and *Prkaca*$^{C\alpha R}$/MMTV-PyMT **(C)** tumors having similar proportions of basal, mature luminal, and luminal progenitor cells, however the tumors evolve differently, as seen in the carcinoma stage where the *Prkaca*^{WT}/MMTV-PyMT tumor **(B)** has a larger proportion of mesenchymal and luminobasal cells compared to the *Prkaca*$^{C\alpha R}$/MMTV-PyMT tumor **(D)**.

3.4 Discussion

PKA is known to promote terminal differentiation and exhaustion of the stem cell compartment of the epidermis[141], and has been shown to promote MET and inhibit EMT in mammary tumor *in vitro* models[139]. In this study, PKA activity is shown to inhibit self-renewal of the basal compartment of the mammary gland, and disrupt cellular plasticity in mammary tumors This prevents the emergence of the lumino-basal population, reducing the metastatic potential of tumors, and affecting the ability of tumor cells to leave and colonize other organs, leading to improved survival. These results suggest that limiting the self-renewal and plasticity of tumor cells is a viable strategy to employ in order to help reduce the aggressiveness of the disease and thus improve prognosis and outcomes for patients with breast cancer.

The concurrent decrease in metastatic burden and lumino-basal cell population in PKA active tumors suggest that EMT may be synonymous with luminal-to-basal plasticity, where tumor cells with a more luminal phenotype represent an epithelial population, and the emergence of basal-like traits represent a mesenchymal shift in these epithelial cells. While both the luminal and basal populations retain epithelial markers, the lumino-basal population has been observed to express higher levels of mesenchymal markers *Vim* and *Mmp2*, while the luminal population expresses higher levels of epithelial markers *Epcam* and *Cdh1* (Chapter 2). The retention of epithelial markers in the basal and lumino-basal phenotypes suggest that these cells do not fully transition to a terminal mesenchymal state, but instead occupy a transitional or intermediate mesenchymal state along the EMT spectrum. Furthermore, genes associated with EMT promotion have also been shown to be important in basal cells. Slug, a transcription factor that inhibits E-cadherin expression and promotes EMT[152], has been found to also repress the expression of luminal markers Krt8/18, ER, and GATA3[153]. Smooth-muscle-actin (SMA), which has been found to be important in the dissemination of human lung adenocarcinoma[154], is also highly expressed in basal or myoepithelial cells[155]. Basal-like tumors have also been shown to have upregulated expression of mesenchymal genes *VIM*, *ACTA2*, and *CDH2*, and reduced

expression of the epithelial gene *CDH1*[156], suggesting that EMT is involved in the evolution of basal-like mammary carcinomas.

The results of this study suggest that inhibition of cellular plasticity and EMT by activating PKA is a potentially viable treatment strategy to prevent the development of metastatic breast cancer; however therapeutically activating a target may be more complicated than inhibiting or inactivating it. Alternatively, instead of targeting PKA directly, substrate proteins inhibited by PKA may be inhibited instead. Analog sensitive (AS) mutants can be used in order to elucidate the downstream effectors of PKA. AS mutants are kinases with a mutation in their ATP-binding site, allowing the kinase to utilize a bulky ATP-analog (N-6-benzyl-ATP)[157]. As the N-6-benzyl-ATP is larger than ATP, only AS mutants, and not other wild-type kinases, are able to utilize it. Radioactive tagging of the terminal phosphates on the N-6-benzyl-ATP and subsequent phosphorylation of substrates by PKA allows the tagged phosphate to be transferred to the target molecule, thus permitting the identification of PKA phosphorylated proteins. Downstream targets of PKA that may normally play a role in promoting cellular plasticity and are inhibited by PKA activity can then be identified, and may thus serve as an ideal target molecule for therapeutic intervention.

Chapter 4: Eribulin induced mesenchymal-epithelial transition (MET) enables tumor cells to gain sensitivity to subsequent chemotherapy

Experiments for Figure 17E were performed by Meisam Bagheri

Experiments for Figure 17-20 were performed by Gadisti Aisha Mohamed with computational analysis performed by Mohamed Ashick

4.1 Introduction

Chemotherapy resistance is one of the major causes of death of patients with breast cancer. Metastatic breast cancer patients treated with anthracycline- or taxane-based therapies show moderate response rates ranging from 30-70%, with a median time to progression of 6 to 10 months[158,159], after which treatment failure and tumor relapse frequently occurs. This relapse is largely driven by the emergence of resistant clones, which survive the initial treatment and repopulate the tumor.

The presence of various cell types within a heterogeneous tumor suggests that a single drug treatment, chemo- or targeted therapy, may not be efficient in eliminating all of the tumor cells at once. While a drug may be successful in eradicating most of the tumor cells, there may be some cells within the tumors that are resistant to the treatment, allowing them to survive, and repopulate the tumor once the sensitive population has been eliminated. Looking at the modes of resistance in a simplistic manner, drug resistant populations may emerge in one of two ways: Darwinian selection occurs when a pre-existing drug resistant population is selected for during drug treatment[160] (Fig 15A). This resistant population is able to tolerate and avoid drug-induced elimination, allowing it to survive the treatment. Other cells that are susceptible to the drug will be eliminated, allowing the resistant clone to then repopulate the tumor after treatment. On the other hand, Lamarckian induction occurs when the drug itself is able to trigger an epigenetic reprogramming of tumor cells to become resistant to the drug[20] (Fig 15B). This resistant phenotype is typically not found in the drug naïve population, and only emerges in response to drug treatment.

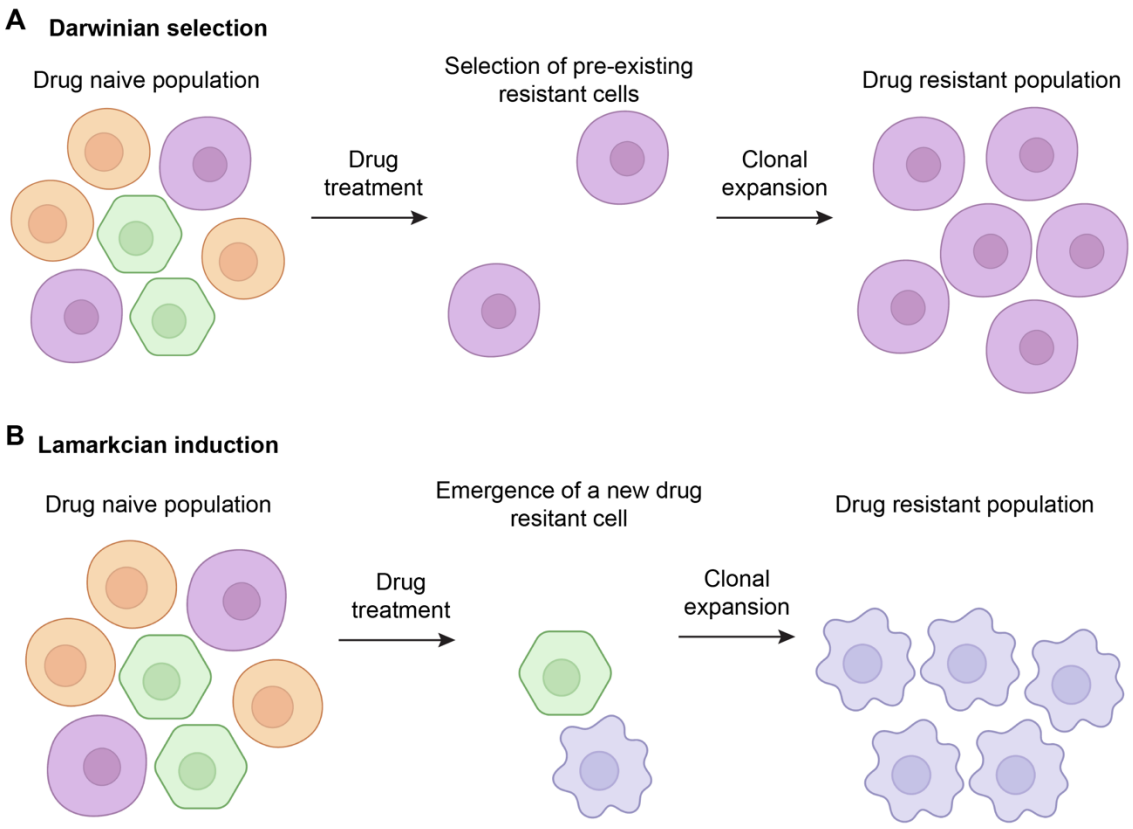


Figure 15: Modes of chemotherapeutic resistance. (A) Darwinian selection occurs when pre-existing drug resistant phenotypes are present in the drug naïve population. Drug treatment selects for these resistant phenotypes by eliminating other non-resistant clones, allowing the resistant population to repopulate. **(B)** Lamarckian induction occurs when the drug treatment itself induces epigenetic reprogramming of tumor cells, allowing a drug resistant clone to emerge after treatment. This drug resistant phenotype did not exist in the drug naïve population but arises after drug treatment to repopulate the tumor. (Created with BioRender.com)

Either of these modes of resistance may be used by tumor cells as a strategy to overcome therapeutic elimination. Darwinian selection has been implicated in the therapy resistance of metastatic melanoma, where pre-existing populations harboring MEK1 mutations (MEK1P124) respond poorly to BRAF inhibitors such as dabrafenib[161]. Mathematical modeling has also implicated Darwinian selection as the mode of resistance in the recurrence of *KRAS*-mutant colorectal carcinomas after treatment with the anti-EGFR agent panitumumab[162]. On the other hand, transcriptionally variable melanoma cells have been found to undergo epigenetic reprogramming upon treatment with the BRAF inhibitor Vemurafenib, suggesting that Lamarckian induction could be responsible for the emergence of Vemurafenib-resistant cells[26]. Finally, cisplatin treatments have been found to promote Darwinian selection of pre-existing resistant cells in a heterogeneous oral squamous cell carcinoma cell line population, while inducing Lamarckian induction in a homogeneous population, suggesting that chemotherapeutics can promote both these resistance mechanisms in tumors under different contexts[20].

Eribulin is a chemotherapeutic drug used as a third line therapy against breast cancer[163]. This drug works by preventing dynamic instability of microtubules, thereby inhibiting microtubule elongation and cell division[164]. Previous studies have shown that eribulin treatment is able to reverse epithelial-mesenchymal transition (EMT), and induce a mesenchymal-epithelial transition (MET) in triple-negative breast cancer (TNBC) cells and xenograft models, thus decreasing metastasis and prolonging survival[78]. This suggests that eribulin is able to induce epigenetic reprogramming, thus allowing cells to gain a new drug resistant phenotype, with the additional benefit of preventing tumor cell migration.

In order to study the mode of resistance that eribulin exerts against mammary tumors, lineage tracing using cells with expressed barcodes sequences were used (Fig 16). By identifying the barcode identities (IDs) of treatment naïve and drug resistant populations using single-cell RNA sequencing (scRNA-Seq), the lineage of drug resistant cells can be traced back to their pre-treatment ancestors. The gene expression profile of each cell

obtained from scRNA-Seq can be used to analyze transcriptomic similarity between the drug resistant cells and their pre-treatment ancestors (Fig 16B). Any changes that a cell undergoes upon drug treatment would manifest as changes in gene expression, which would imply an epigenetic or transcriptional reprogramming leading to the emergence of a *de novo* cell state. Alternatively, if no change in gene expression is observed, it would suggest that a selection event has occurred, amplifying a pre-existing resistant population that is transcriptionally stable. Cells that survive treatment due to Darwinian selection would be transcriptionally similar to the treatment-naïve cells expressing the same barcode, while those that survived due to Lamarckian induction would be transcriptionally distinct from their pre-treatment ancestors (Fig 16C).

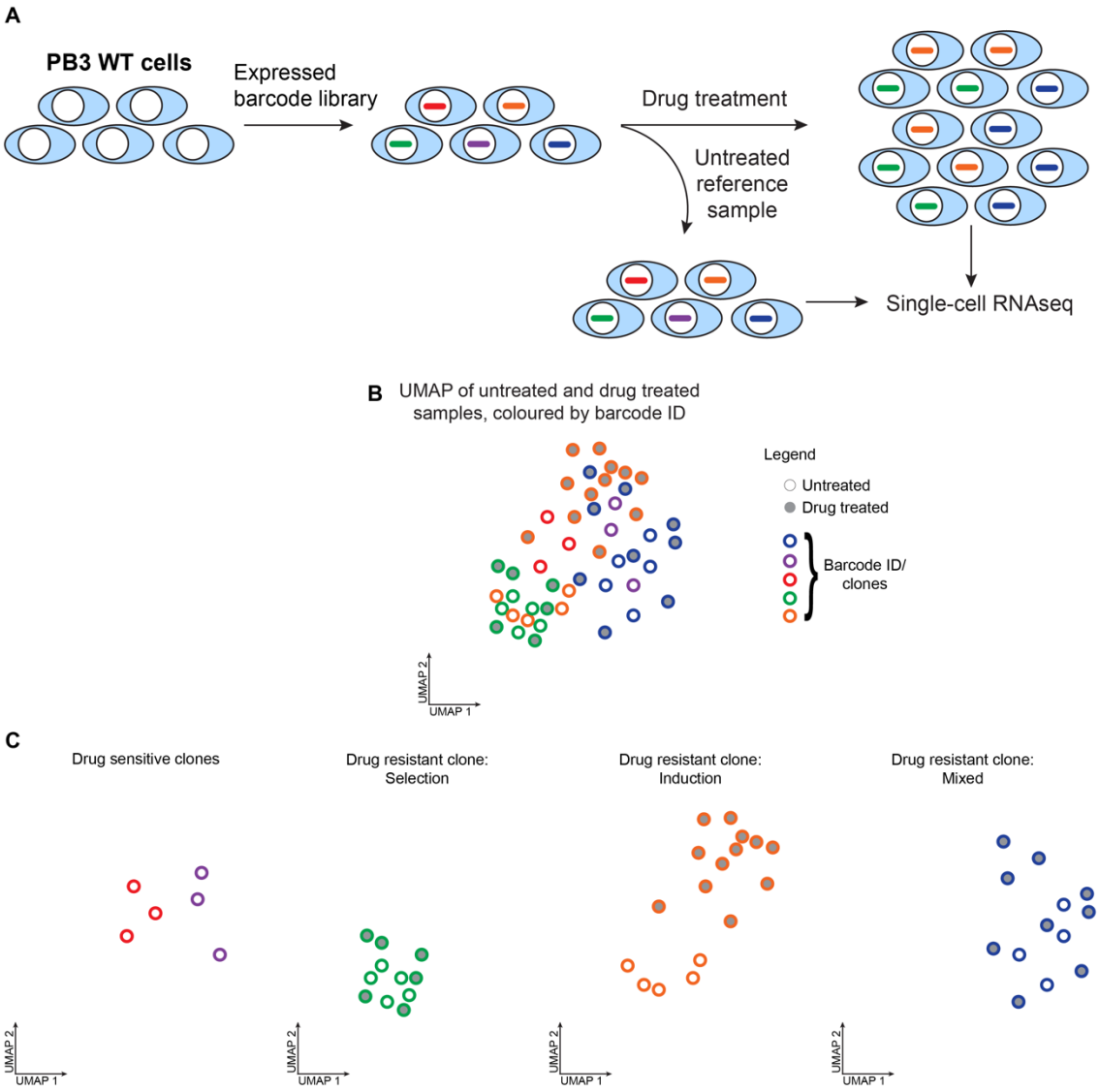


Figure 16: Strategy to elucidate the mode of resistance to eribulin and paclitaxel. (A) PB3 cells were labelled with expressed barcode sequences via lentiviral delivery. Successfully barcoded cells were then selected for and allowed to expand. A portion of the barcoded cells were then treated with a drug. scRNA-Seq was then performed on the resistant cells that emerged, along with a portion of untreated cells to identify the transcriptome and the barcode identity of each cell. **(B)** Example of scRNA-seq UMAP results showing distribution of treated and untreated cells colored by barcode ID. Each barcode ID represent a single clone within the population. **(C)** Determination of mode of resistance by identifying distribution of untreated and drug resistant clones with the same barcode. Barcode IDs that are found only in the untreated population and not in the drug

resistant population indicate drug sensitive clones that have been eliminated. Clones are resistant by selection if surviving drug treated cells cluster close to their untreated ancestors. Clones are resistant by induction if surviving drug treated clones cluster far apart from their untreated ancestors. A mixed mode of drug resistance can also occur if some of the surviving drug treated clones cluster close to the untreated ancestors, while some cluster further apart.

4.2 Materials and methods

Lentiviral barcoding and selection of successfully barcoded cells

PB3 cells were grown to an estimated 70% confluency, and trypsinized to obtain a single cell suspension. The cells were counted, and 200 000 cells were plated in a 6cm tissue culture dish, along with 6ug/ml Polybrene transfection reagent (Millipore, Cat #TR-1003-G, Lot #2129807), and were incubated at 37°C for one hour. A lentiviral barcode library (CloneTracker XP^{TM} 10M Barcode-3' Library with RFP-puro (Packaged), Cellecta #BCXP10M3RP-V), was then added at an MOI of 0.5, and the cells and virus were further incubated at 37°C overnight. After 24 hours, the media was replaced with fresh media, and the cells were allowed to grow for an additional two days. The cells were then trypsinized to obtain a single cell suspension, and flow sorted using the FACSAria III cell sorter to select for RFP expressing, successfully barcoded cells. The collected cells were then re-plated and allowed to grow until the plate reaches confluency. Any cells not used for further experiments were frozen down for long term storage.

Starting a founder population of barcoded cells

Since the 10x genomics platform can only run a maximum of 10 000 cells per sample, a founder population of at most 10 000 barcodes was first obtained to ensure consistency of barcode distribution across samples. This was done by counting 10 000 cells and plating them in one well of a 24 well plate. This founder population was allowed to grow by scaling up the plate size to ensure that a sufficient number of cells are obtained to carry out the drug treatment experiments.

Drug treatment of barcoded cells

Barcoded PB3 cells were counted, and 80 000 cells were seeded into each well of a 12 well plate. The cells were incubated at 37°C overnight, and 500nM eribulin or 100nM paclitaxel were added. Cells were incubated with the drug for 72 hours, before being replaced with fresh, drug free media. The cells were allowed to recover for 48 hours. Before the next round of treatment, the cells were re-plated into 6 well plates to allow

for more space for the cells to grow. As the paclitaxel treated cells seemed to be resistant to 100nM of the drug, the dosage was increased to 250nM for subsequent treatment rounds. The drug treatment was repeated for an additional two rounds, with a portion of cells frozen down after each treatment to represent the different treatment cycles. To drive the eribulin-induced epithelial change the eribulin treated cells were subjected to one additional round of treatment.

Single-cell omics

Untreated and Eri4 samples were initially run on the 10x Genomics single cell RNA-seq and single cell ATAC-seq platforms. The intermediate treatment cycles (Pac1-Pac3, and Eri1-Eri3) were run simultaneously on two 10x Genomics single cell RNA-seq runs, using multi-seq multiplexing to differentiate between samples. A second sample of untreated and Eri4 samples were run on 10x Genomics single cell Multiome ATAC + Gene Expression.

Barcode whitelist preparation

The lentiviral barcode library was purchased from Cellecta (CloneTracker XP^{TM} 10M Barcode-3' Library with RFP-puro (Packaged), Cellecta #BCXP10M3RP-V). A whitelist of barcodes was prepared by combining BC14 and BC30 (BC14sequence+BC30sequence) barcodes. In custom barcode only sequencing, barcodes are expected in the following format,

Custom barcode = 19bp constant + BC14 sequence + TGGT + BC30 sequence

These barcodes were prepared in various formats as per the requirement of downstream tools using in-house scripts. These reference files are used for feature barcode demultiplexing in downstream for custom barcodes FASTQ files.

Single cell/nuclei transcriptome data analysis

Eight libraries for single cell/nuclei transcriptome data were obtained in 3 batches namely Run1, Run2 and Run3. Each batch contains different treatment and dosage samples sequenced as listed in supplementary table S1 with their summary statistics. While Run1

and Run2 data were from single cell 3' V3 sequencing (scRNA) protocol, Run3 was single nuclei RNA+ATAC V1 sequencing (snRNA multiome) protocol. Hence, FASTQ and count matrix were generated using 10x Genomics cellranger-3.1.0, 10x Genomics cellranger-4.0.0[96] and 10x Genomics cellranger-arc-1.0.0[165] for Run1, Run2 and Run3 respectively. Cellranger reference mm10 was used as reference for all runs. For Run1 and Run2, whitelist custom barcodes (as detailed in previous section) were used in cellranger "--feature-ref" parameter for custom barcodes FASTQ. As cellranger-arc pipeline doesn't support "--feature-ref" barcoding, custom barcode FASTQ files were processed using CITE-seq-Count v1.4.3 tool[166].

Count matrices were analysed using Seurat v3.1.4 pipeline[167]. Briefly, low quality cells (number of genes expressed < 200, percentage of UMIs in mitochondrial and ribosomal genes individually > 50) are filtered out. Using selected variably expressing genes and significant PCs from PCA, clustering and UMAP projection were generated. Further, all the samples are merged into one object to illustrate the differences due to treated drug and their varying dosage effect. Post merging, batch effects due to run and technology differences were regressed out. Pseudotime analysis was done against all samples merged dataset using Monocle3 v0.2.1[150,151,168] while considering "Untreated" cells as the root cells. To represent the relative frequency of each barcode over pseudotime, Muller plot representation was used[169]. Relative frequencies were calculated as the ratio between number of cells associated with a barcode at a given pseudotime (binned) and the sum of all cells (from all barcodes) at the same pseudotime bin. Muller plot is accompanied with line chart illustrating the total number of cells at any given pseudotime.

Single cell/nuclei ATAC data analysis

Four libraries for single cell/nuclei ATAC data were obtained in 2 batches namely Run1 and Run3 using scATAC V1 and snATAC (multiome) V1 protocol respectively. Sample and summary statistics are provided in supplementary table S2. FASTQ and peak count matrix were generated using 10x Genomics cellranger-atac-1.2.0 and 10x Genomics cellranger-

arc-1.0.0 for Run1 and Run3 respectively. Cellranger reference mm10 was used as reference for both the runs. Custom barcode FASTQ files were processed using CITE-seq-Count v1.4.3 tool.

Peak count matrices were analysed using Signac v1.1.1 pipeline. Briefly, low quality cells (fragments in peak regions < 1000 and >75000, percentage of UMIs in peaks > 20, ratio of UMIs in blacklist regions to that of peak regions > 0.05, nucleosome signal > 10 and TSS enrichment < 2) are filtered out. Using selected top features and 2 to 30 PCs from LSI, clustering and UMAP projection were generated. Further, all the samples are integrated into one object using Harmony R package[170] to illustrate the differences due to treated drug and their varying dosage effect. In harmony, batch effects due to run and technology differences were regressed out.

Integration and Custom barcode classification

To assign each cell to a specific custom barcode, we have applied MULTIseqDemux method[171] from Seurat pipeline. Though the method has managed to differentiate the barcodes, this method was originally designed for multiple samples demultiplexing, where background estimation is appropriated for each cell from all the barcodes. Here, the enrichment level for each barcode can vary, and it is appropriate to calculate the background for each barcode from all the cells. We deployed an in-house developed R script using binomial distribution, that estimates background for each barcode from all the cells. We have considered only doublets and negatives annotated cells from MULTIseqDemux annotation. Using this approach, we were able to re-annotate and increase the number of singlets. Finally, RNA and ATAC datasets were integrated with TransferData module in Seurat pipeline.

Selection vs induction identification

To identify and illustrate the cells that undergo selection (Treated and Untreated cells clustering together) or induction (Treated and Untreated cells clustering further apart) process upon treatment, all cells in each barcode were analysed using Jaccard index and

Euclidean distance-based methods. Euclidean distance is calculated between median points of cells from a sample per barcode in their UMAP space. For differences within samples, we have used Jaccard index method from scclusteval v1.0[172], where cells of each barcode were re-clustered using density clustering method (with same variable genes from original integrated dataset). Based on the number of cells for each sample (treated or untreated) in various clusters, the Jaccard index is calculated. In density clustering, proximal cells are assigned to one cluster. Ideally, if all the cells are similar in a given scenario, then they will be clustered closely and assigned into one cluster. In complex scenarios, cells in barcode and then within sample, cells can form sub-clusters. Each sub-cluster can exhibit varying scenarios in selection vs induction process. Jaccard index will score each cluster as per the number of cells (out of total cells in a sample) in a given cluster for both the samples.

4.3 Results

4.3.1 Morphological changes in PB3 cells upon eribulin, but not paclitaxel treatment

PB3 cells, a heterogeneous cell line derived from MMTV-PyMT mammary tumors[173], were first transduced with a Lentiviral barcode library to label each cell with a unique, expressed barcode (Fig 17A). Cells were then treated with either 4 rounds of eribulin, or 3 rounds of paclitaxel. Paclitaxel was used as a control as it has a similar mechanism of action as eribulin. Brightfield images taken of barcoded PB3 cells upon eribulin treatment showed the emergence of an epithelial-like subpopulation of resistant cells, compared to the generally mesenchymal morphology of untreated cells (Fig 17B and D). These cells appear more compact and regularly shaped, with the cobblestone morphology typical of epithelial cells. Epithelial morphology was not observed in intermediate cell populations, where the cells underwent only 1 or 2 rounds of eribulin treatment (Fig 17D), suggesting that a change in phenotype has occurred in cells that are ultimately resistant to eribulin. In contrast, paclitaxel treated cells appear to maintain the elongated and irregular mesenchymal morphology of the original untreated population throughout the course of treatment (Fig 17C), suggesting that these cells did not undergo a phenotypic shift, and

thus cells that are resistant to paclitaxel may already be present in the population before treatment began. Expression of EMT markers were also different in eribulin-resistant cells, with an increase in the epithelial marker E-cadherin, and a reduction of the mesenchymal marker vimentin, as compared to controls (Fig 17E). No difference in EMT marker expression was observed in paclitaxel-resistant cells.

While these results may suggest eribulin resistance arises due to induction and paclitaxel resistance occurs due to selection, there is a possibility that a small population of epithelial eribulin-resistant cells may exist, undetected, in the original population, or that paclitaxel treatment may reprogram the cells into a different state where they still appear morphologically mesenchymal. It is thus necessary to study the evolution of drug resistant clones at the single-cell level to better understand the dynamics of treatment.

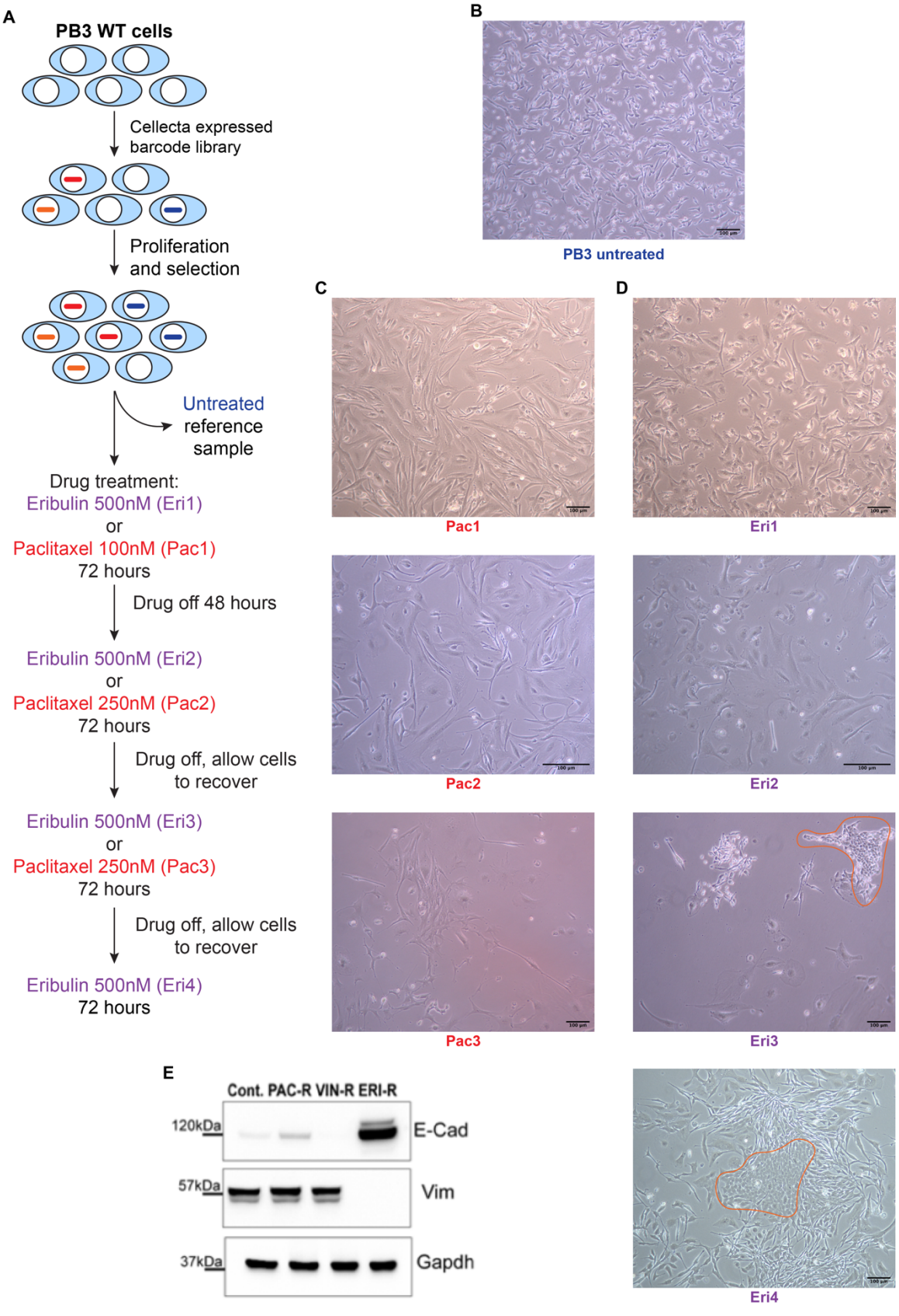

A
PB3 WT cells
Cellecta expressed barcode library
Proliferation and selection
Untreated reference sample
Drug treatment:
Eribulin 500nM (Eri1)
or
Paclitaxel 100nM (Pac1)
72 hours
Drug off 48 hours
Eribulin 500nM (Eri2)
or
Paclitaxel 250nM (Pac2)
72 hours
Drug off, allow cells to recover
Eribulin 500nM (Eri3)
or
Paclitaxel 250nM (Pac3)
72 hours
Drug off, allow cells to recover
Eribulin 500nM (Eri4)
72 hours
B
PB3 untreated
C
Pac1
Pac2
Pac3
D
Eri1
Eri2
Eri3
Eri4
E
Cont. PAC-R VIN-R ERI-R
120kDa
E-Cad
57kDa
Vim
37kDa
Gapdh

Figure 17: PB3 cells resistant to eribulin display epithelial morphology different from the original, untreated cells. (A) Overview of barcoding implementation and drug treatment schedules for generating eribulin- and paclitaxel-resistant PB3 cells. **(B)** Brightfield image showing mesenchymal morphology of untreated PB3 cells. **(C)** Brightfield images showing paclitaxel treated cells maintaining the mesenchymal morphology of PB3 cells before treatment. **(D)** Brightfield images showing the emergence of epithelial-like cells (outlined in orange) upon treatment with eribulin, with eribulin resistant populations consisting of a mix of cells displaying either epithelial and mesenchymal morphology. **(E)** Western blot showing expression of the epithelial marker E-cadherin (E-Cad) and the mesenchymal marker vimentin (Vim) in untreated PB3 control and drug-resistant cells. Eribulin resistant cells (ERI-R) showed an increase in E-cadherin expression, and decrease in vimentin expression, compared to untreated control cells (Cont.). No differences in E-cadherin and vimentin were observed in paclitaxel-resistant (PAC-R) and Vinorelbine-resistant (VIN-R) cells. Vinorelbine was used as an additional control, as it has similar mechanism of action as eribulin and paclitaxel.

4.3.2 Cells receiving eribulin treatments form distinct clusters on a UMAP

PB3 cells receiving each iteration of eribulin or paclitaxel treatment were harvested, along with an untreated reference sample, and were analyzed using single-cell RNA sequencing (Fig 18A). Dimensionality reduction with uniform manifold approximation and projection (UMAP) appears to show untreated cells clustering mostly towards the right, while eribulin resistant cells (Eri4) clustering separately on the left. The intermediate treatments Eri1, Eri2, and Eri3 cluster in between the untreated and Eri4 clusters, with Eri3 appearing as a distinct cluster separate from Eri1 and Eri2. Paclitaxel treated clusters appear to be less distinct, with Pac1, Pac2, and Pac3 cells clustering together towards the middle and bottom of the plot. These results suggest that eribulin resistance cells had evolved

Monocle pseudotime[151] orders the cells along an evolutionary trajectory, thereby helping to identify the cells in the most primitive and advanced transcriptomic states as they undergo treatment. In the eribulin pseudotime plot, untreated cells are in the most primitive transcriptional state, and Eri4 cells are in the most advanced transcriptional state (Fig 18B). Eribulin resistance seems to progress uniformly in the initial stages, with one single branch progressing through Eri1 and Eri2, before breaking off into 4 branches, showing diverse transcriptional fates of Eri3 and Eri4. This demonstrates potential induction where Eribulin induces transcriptional changes to allow resistant cells to be transcriptionally distinct from untreated cells. Three of the branches of the pseudotime points also seem to end on Eri3 cells, suggesting that the final transcriptional states of Eribulin resistant cells may already be established as early as the third treatment.

In the paclitaxel pseudotime plot, untreated cells are in the most primitive transcriptional state, and Pac2 are in the most advanced transcriptional state (Fig 18C). The Paclitaxel pseudotime plot also seems more simplified, with fewer branch points. The branch points also seem to end on different samples, with one ending on Pac1, two ending on Pac2, and

one ending on Pac3. This shows that there is no uniform directionality of Paclitaxel resistance, suggesting random selection mechanism of resistance

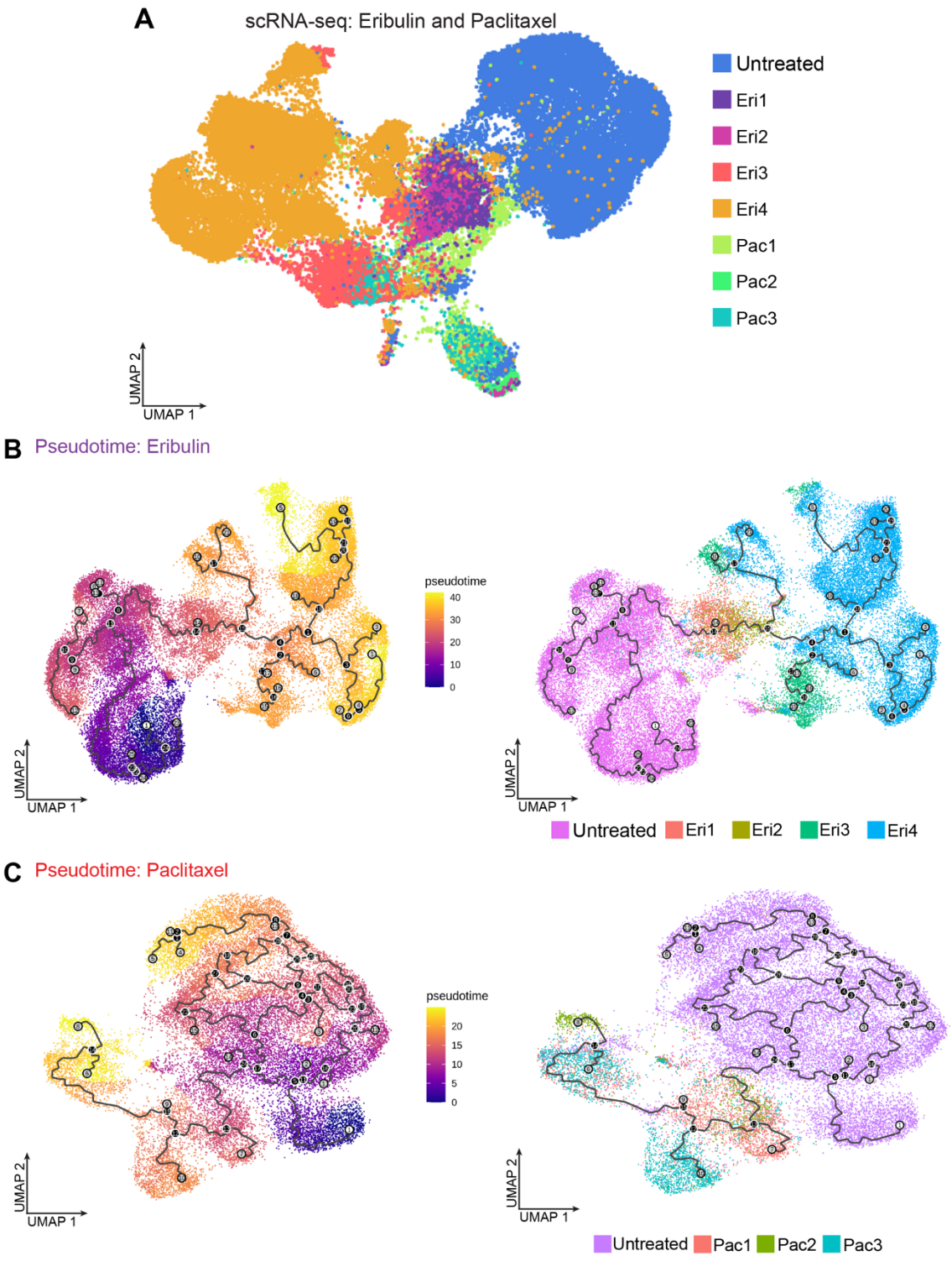


Figure 18: UMAP clustering and pseudotime analysis of eribulin and paclitaxel treated cells. (A) scRNA-seq UMAP projection of untreated, and eribulin and paclitaxel treated cells**. (B and C)** Monocle pseudotime projection showing the trajectory of tumor evolution upon treatment with eribulin **(B)** and paclitaxel **(C)** respectively. The main lineage path or

master path/node is indicated by the number 1 in the white circle. Sub-lineages from the master path (also called leaf nodes) are indicated by numbers in grey circles. Further sub-divisions or branches from the leaf nodes are indicated by numbers in black circles.

4.3.3 Eribulin treatment induces transcriptomic reprograming, allowing PB3 cells to gain drug resistance

Given that all the cells surviving from the subsequent drug treatments were descended from the initial founder population of barcoded cells, cells with the same barcode can thus be considered as a single clone, and treated cells are descendants of untreated cells with the same barcode (Fig 16). In order to understand the mode of resistance of each clone, 28 barcodes with the most abundant number of untreated and resistant cells from each treatment were analyzed individually. Calculating the Euclidean distance between the median position of untreated cells and the median position of resistant cells on the UMAP would provide a direct measure of distance between them, where shorter Euclidean distance signifies closer transcriptomic similarities between untreated and resistant populations, likely implying selection as the mode of resistance. On the other hand, a larger Euclidean distance signifies larger transcriptomic differences and imply induction as the primary mode of resistance.

Euclidean distance calculations revealed larger distances between untreated and resistant cells in eribulin resistant clones compared to paclitaxel resistant clones (Fig 19A and B). Eribulin resistant clones have a median Euclidian distance of 11.74 whereas paclitaxel resistant clones have a median Euclidian distance of 5.59 (Fig 19G), suggesting that eribulin treatment results in higher transcriptomic reprogramming of resistant cells compared to paclitaxel. Furthermore, barcodes with the smallest Euclidean distance score appear to originate from the bottom cluster of cells, suggesting that this region harbors clones that are resistant by selection, whereas resistant cells with the largest Euclidean distance score are found in the cluster of cells on the far left, suggesting that this region harbors clones that are resistant by induction (Fig 19A and B). Fourteen eribulin resistant, and only 3 paclitaxel resistant clones were found in the ‘induction’ region, while only 1 eribulin resistant, and 5 paclitaxel resistant clones were found in the ‘selection’ region, implying that eribulin resistance occurs predominantly via induction.

For clones that are not found within these two regions, Jaccard index was used to determine mode of resistance. For this method, density clustering was used to assign untreated and resistant cells into clusters based on proximity, and Jaccard index scores each cluster according to the number of untreated and resistant cells present[172]. A Jaccard index score of 1 indicates that the cluster consists of only resistant or untreated cell, while a Jaccard index score of <1 indicates that a mix of resistant and untreated cells can be found within the cluster. To determine if a clone is resistant by induction, untreated and resistant cells will thus have to occupy separate clusters, resulting in a Jaccard index score of 1 for the untreated cluster, and 1 for the resistant cluster (Fig 19C). For a clone that is resistant by selection, clusters will have a mix of untreated and resistant cells, resulting in a Jaccard index score of between 0 and 1 (Fig 19D).

Using a combination of the Euclidean distance calculation and Jaccard index scores, the mode of resistance for each clone was identified (Fig 19E), revealing induction as the predominant mode of resistance upon eribulin treatment, and selection induction as the predominant mode of resistance upon paclitaxel treatment. This indicates that pre-existing paclitaxel resistant cells may be present in the untreated population, while eribulin treatment itself is able to reprogram cells and allow them to evolve and gain resistance.

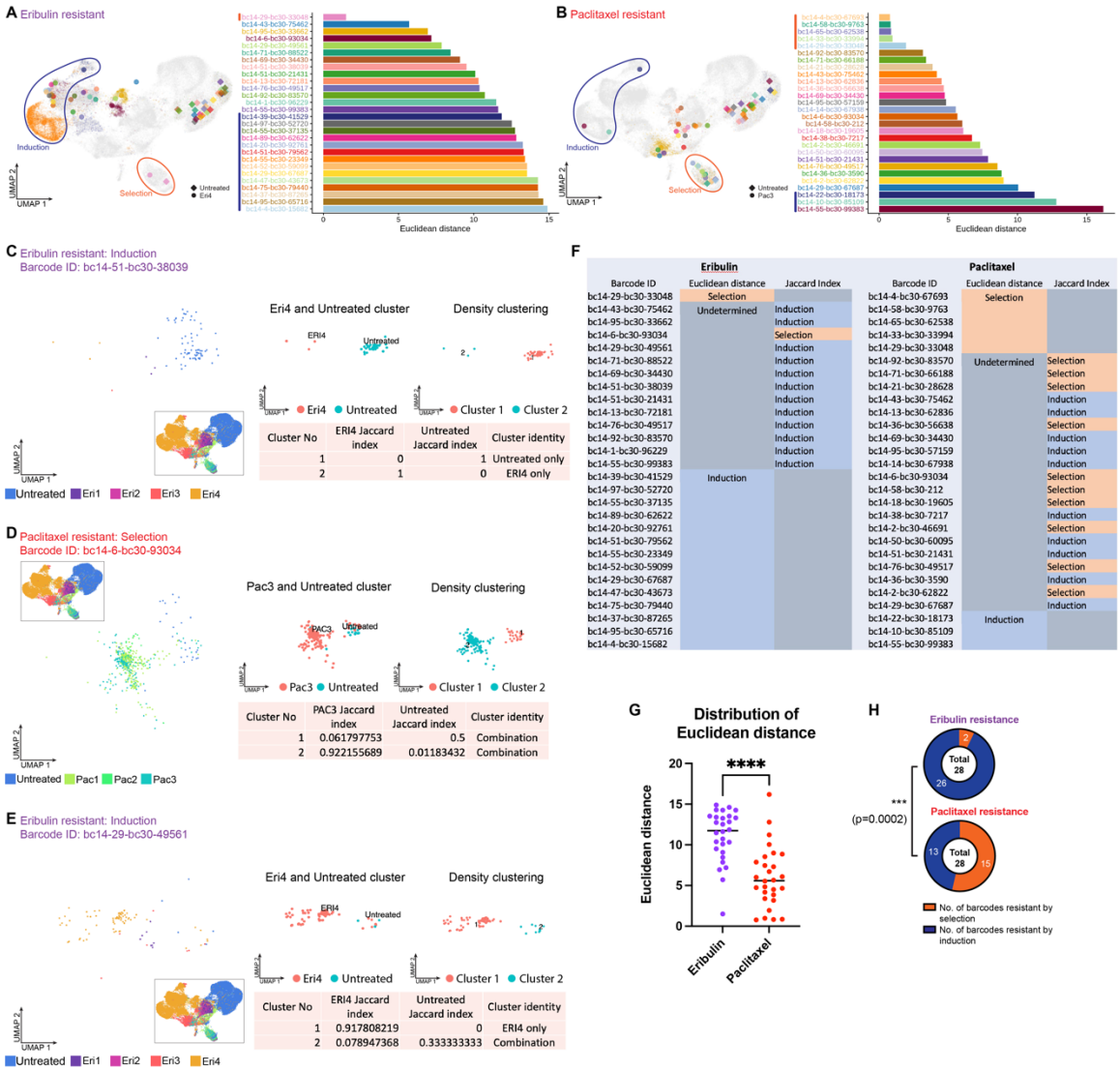


Figure 19: Determination of induction or selection mode of resistance. (A) Euclidean distance calculations between Eri4 and untreated cells with the same barcode **(A)**, and between Pac3 and untreated cells with the same barcode **(B)**. The UMAP overlay shows the median point of Eri4 or Pac3 (represented by •) and the median point of untreated cells (represented by ♦). **(C)** Example showing the determination of induction based on Jaccard index. Eri4 and untreated cells with this barcode fall into different clusters, as seen by the Jaccard index score. **(D)** Example showing the determination of selection based on Jaccard index. Pac3 and untreated cells with this barcode share the same clusters, as seen by the Jaccard index score. **(E)** Example showing a combination of induction and selection within the same barcode. For this barcode, some clusters show

resistance by induction (no untreated cells in the cluster), and some show resistance by selection (at least one untreated cell in the cluster). The cluster with the largest number of resistant cells is used to determine mode of resistance. In this example, cluster 1 has the largest number of Eri4 cells, and these cells are resistant by induction, hence, this barcode was assigned as resistant by induction. **(F)** Breakdown of induction or selection determination based on both Euclidean distance and Jaccard index calculations. **(G)** Distribution of Euclidean distance values of eribulin resistant and paclitaxel resistant barcodes. Median for eribulin: 11.74, median for paclitaxel: 5.597. $P<0.0001$, Mann-Whitney test. **(H)** Proportions of clones resistant by induction vs clones resistant by selection resistance for eribulin and paclitaxel treatments. $P=0.0002$, Chi-Square test

4.3.4 Single-cell ATACseq and Muller Plots support Lamarckian induction of eribulin-resistant cells

To confirm the results of the single-cell RNAseq data, and that Euclidean distance calculations can be used as a measure of epigenetic change, single-cell ATAC seq was performed on untreated and Eri4 (eribulin resistant) samples. Dimensionality reduction with UMAP of barcodes with 10 lowest Euclidean distance (representing selection, Fig 20A) show some Eri4 cells within the Untreated cluster, and some untreated cells in Eri4 cluster, suggesting similarities in accessible chromatin regions between the untreated and eribulin resistant samples. On the other hand, distinct segregation of Untreated and Eri4 cells were observed in barcodes with 10 largest Euclidean distance (representing induction, Fig 20B), suggesting differences in accessible chromatin regions between the two samples, and confirming that Euclidean distance calculation from RNAseq data is able to distinguish between induction and selection mode of resistance.

To further confirm Lamarckian induction of eribulin resistance and Darwinian selection of paclitaxel resistance, changes in clonal dynamics over the course of treatment were analyzed. Muller plots with the top 10 most abundant barcodes after eribulin and paclitaxel treatments were constructed and plotted against pseudotime, to identify if the dominating clones change over the course of the treatment. For the eribulin treated population, dominating clones differ at different points. In the early stages of drug treatment, the bulk of the population is made up of cells expressing 4 barcodes (bc14-55-bc30-99383, bc14-55-bc30-37135, bc14-39-bc30-41529, bc14-1-bc30-96229), suggesting that these 4 clones were able to resist initial eribulin treatment. Eventually these 4 clones succumbed to drug treatment, leading to the dominance of 3 other clones (bc14-95-bc30-33662, bc14-76-bc30-49517, bc14-6-bc30-93034) at pseudotime points 15 to 30. At around pseudotime point 23, another clone (bc14-55-bc30-23349) began to emerge, which then almost completely dominated the resistant population towards the end of the treatment. The emergence of different dominating clones at different timepoints suggest that these clones have had to adapt to survive the drug treatment. Drug induced

evolutionary pressures may force some clones to evolve in order to survive early eribulin treatment, but these changes failed to allow them to survive further treatment, thus allowing other clones to dominate.

Dominant clones as a result of paclitaxel treatment seem to be more consistent, with 3 clones dominating throughout the course of treatment (bc14-92-bc30-83570, bc14-76-bc30-49517, bc14-6-bc30-93034), with bc14-92-bc30-83570 appearing as the earliest dominant clone. The consistency in which dominating clones remain within the paclitaxel resistant populations suggest that these clones pre-exist within the untreated population, and that the drug treatment itself does not influence or induce the evolution of resistant cells.

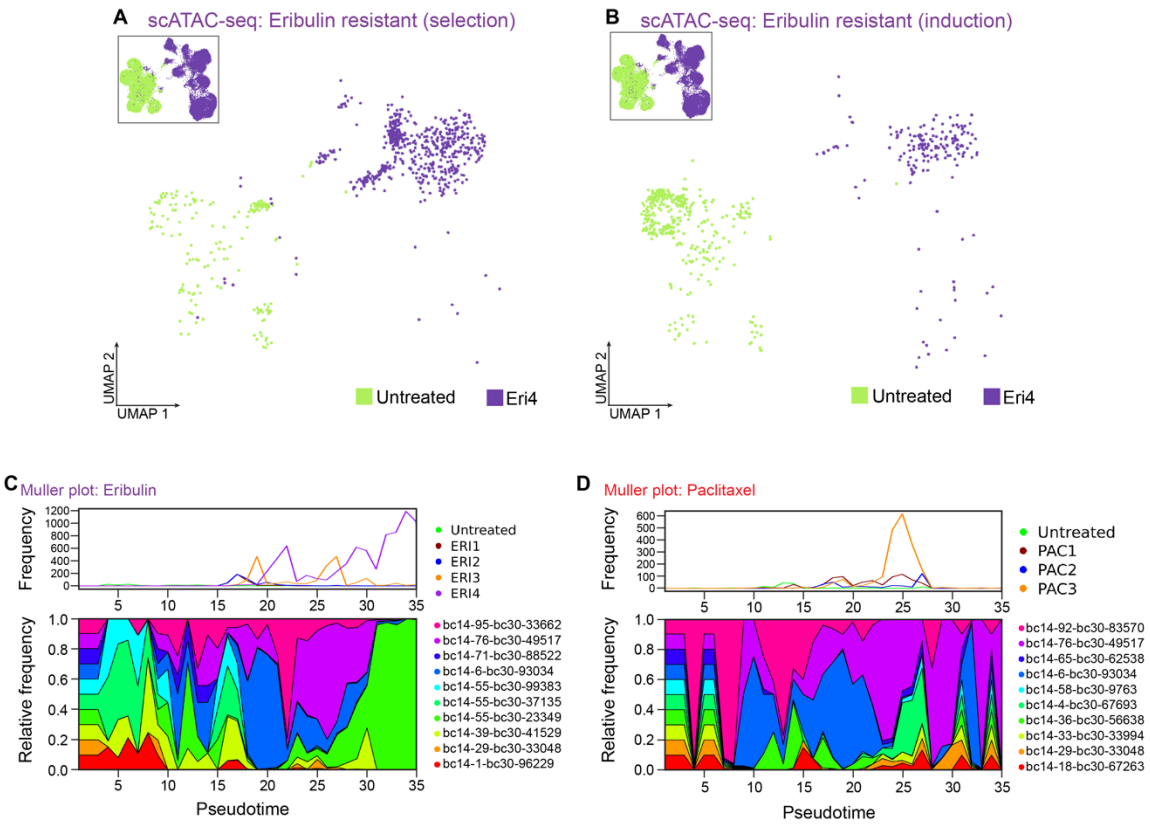


Figure 20: Further evidence of eribulin resistant cells gaining resistance while paclitaxel resistant populations are pre-existing. (A and B) scATAC-seq UMAP projection of Untreated and eribulin treated cells resistant by selection (lowest 10 Euclidean distance values) **(A)**, and eribulin treated cells resistant by induction (highest 10 Euclidean distance values) **(B)**. **(C and D)** Muller plot showing clonal diversity of the top 10 most abundant barcodes after eribulin **(C)** and paclitaxel **(D)** treatments.

4.4 Discussion

The results of this study reveal that a heterogeneous tumor cell population can respond, and gain resistance to, chemotherapeutic drugs via different modes. Analysis of individual drug resistant clones suggest that eribulin appears to induce transcriptional/epigenetic reprogramming in drug naive cells, thus allowing a new drug resistant phenotype to emerge, whereas paclitaxel appears to predominantly select for pre-existing resistant cells within the pre-treated population. These results confirm that the initiation of MET in response to eribulin treatment[78] is indeed a Lamarckian induction event that not only allows tumors to transition into a more epithelial state that is less likely to metastasize, but also to allow them to gain resistance to eribulin.

While Lamarckian induction appears to be the main mode of resistance employed in response to eribulin treatment, and Darwinian selection appears to be the mode of resistance of paclitaxel treated cells, both these modes of resistance were also observed to occur in response to both drug treatments. Two eribulin resistant clones were observed to have arisen due to selection, whereas 13 paclitaxel resistant clones were observed to have emerged due to induction. This suggests that a combination of Lamarckian induction and Darwinian selection can be employed by a heterogeneous cell population as successful strategies to evade drug elimination.

Molecular barcoding technology has been a useful tool used for lineage tracing, especially in drug treatment studies to elucidate the lineage and evolution of resistant clones. The expressed barcoding system has previously been used in various cancer models in order to differentiate heterogeneous cell populations, including studies to analyze clonal dynamics in NSCLC models upon erlotinib treatment[174], and studying clonal evolution and chemoresistance in patient-derived models of pancreatic cancer[175]. More recently, molecular barcodes were used to track clonal responses to next-generation therapeutics that both inactivate and degrade EGFR, compared to standard EGFR-targeting drugs that merely inhibit its kinase function[176]. Dual inhibitor-degrader therapeutics were found to

be less effective than a standard kinase-targeted inhibitor, and subsequent transcriptomic analysis between resistant and sensitive clones revealed a reduction in genes associated with ER processing in inhibitor-degrader resistant clones, suggesting that the downregulation of these genes play an important role in the adaptation or induction of drug resistance[176].

This study utilized molecular barcodes to trace the lineage of resistant cells *in vitro*, where drug treatments were administered in cell culture, however various other factors, such as the immunosuppressive tumor microenvironment, and tumor vascular support, may influence the response of tumor cells against the drug. It is thus useful to study drug induced resistance and evolution in an *in vivo* system. Unfortunately, as molecular barcoding with lentiviral vectors requires manipulation of the tumor cells before treatment, it is challenging to use them to study the effects of drug treatment on autochthonous tumors. A possible way to overcome this is to surgically implant tumor cells in an endogenous site in mice after *in vitro* lentiviral barcoding. The drugs could then be administered to the mice, allowing a more holistic *in vivo* analysis of how drug treatment, in combination with other natural biological processes, may affect drug resistance and evolution.

An alternative to using molecular barcodes that are introduced *in vitro*, *in vivo* cell labelling techniques can also be employed. These techniques involve the use of animal models that have been introduced with lineage barcodes consisting of multiple Cas9 target sites[177]. Over time, these target sites accumulate various mutations, amplifications or deletions that will be passed on to daughter cells, allowing the lineage of each cell to be determined after sequencing. MARC1 (mouse for actively recording cells) mice harboring these Cas9 target barcodes have been generated[178] and can be crossed with other tumor bearing mouse models, such as MMTV-PyMT, to allow tumor formation and development, and subsequent drug treatment and lineage analysis of resistant cells will

help to elucidate how drug treatment affects tumor development and evolution in an autochthonous, *in vivo* setting.

Identification of the mode of resistance to drugs such as eribulin or paclitaxel represents a first step in elucidating the dynamics of drug resistance. With the information gained from single-cell transcriptome and epigenome sequencing, similarities in genes expressed pre- and post- treatment can be identified, along with open chromatin regions that signify a propensity towards plasticity, which can provide an insight into possible mechanisms of induction. Similarities in gene expression patterns of cells resistant by selection can also be used as a predictive marker to response to different types of therapy and help to inform ideal combinatorial treatment regimens to target multiple potentially resistant phenotypes with the aim to prevent tumor repopulation and recurrence.

Chapter 5: Conclusions and future directions

This study aims to identify how lineage plasticity contributes to intra-tumoral heterogeneity of breast cancer, and its implications for tumor development, by elucidating the potential mechanisms that promote phenotypic alterations of tumor cells. SOX10 and PKA activity have been identified as intrinsic and cell-autonomous drivers that induce lineage evolution of luminal-like, low-ER tumors, and interrupt EMT and metastasis respectively, whereas the chemotherapeutic drug eribulin has been identified as an external selection pressure that forces tumor cells to undergo an MET. The identification of these intrinsic and extrinsic drivers of plasticity represents the first step in elucidating the mechanisms by which breast tumors evolve, metastasize, and survive treatment.

The identification of SOX10 as a possible intrinsic driver of luminal-to-basal transition suggests that this process is cell-autonomous, however, it is not known how SOX10 expression is activated within tumors. While stochastic and transient expression of this transcription factor may trigger expression of basal genes, the sustained expression of these basal-specific markers suggests that stochastic mechanisms of SOX10 activation are unlikely to play a major role. Furthermore, the high expression of SOX10 specifically in tumors expressing low ER suggests a possible link between reduction of ER activity and induction of SOX10 activity. This link is yet to be explored, along with possible environmental signals and extrinsic triggers that may reduce ER levels and increase SOX10 expression. The mechanisms by which SOX10 increases basal gene signature expression are also presently unknown, and elucidating this process could help identify potential methods to curtail luminal-to-basal plasticity and prevent the lineage evolution of luminal-like tumors.

SOX10 driven lineage plasticity of luminal cells was also found to involve an incomplete luminal-to-basal transition, in which cells do not transition into a fully basal phenotype. Instead, these cells express a combination of both luminal and basal markers, resulting in

the emergence of a hybrid lumino-basal phenotype. This suggests that complete lineage transition may not be necessary for tumor cells to gain a basal-like phenotype, leading to the declining prognosis associated with the basal-like subtype. Incomplete lineage transition or hybrid cell-states are commonly described in the study of EMT, where it is not necessary for tumor cells to fully recapitulate the phenotypic properties of a mesenchymal cell type in order for it to form metastases[128]. In fact, it appears to be more advantageous for cells to co-express properties of both epithelial and mesenchymal cells, to aid in both invasion away from the primary tumor, and colonizing the tissues of distant organs. Similarly, it may be more advantageous for the lumino-basal phenotype to maintain both luminal-like and basal-like characteristics, although the underlying implications of this phenomenon are yet to be explored. The normal mammary gland structure suggests that basal (myoepithelial) cells exhibit more mesenchymal characteristics, whereas luminal cells exhibit more epithelial-like traits, thus the luminal-to-basal transition may simply be a form of EMT, with the lumino-basal cell phenotype taking the form of a hybrid or intermediate EMT cell-state. Furthermore, transcriptomic analysis has revealed the expression of mesenchymal marker genes in basal cells, and epithelial marker genes in luminal cells. The EMT gene *SNAI2*, along with other mesenchymal markers, have been found to be highly expressed in both human and murine basal cells[179,180], while data from this study also show that the lumino-basal population expresses higher levels of mesenchymal markers *Vim* and *Mmp2*, with the luminal population expressing higher levels of epithelial markers *Epcam* and *Cdh1*.

The identification of the lumino-basal cell phenotype in the low-ER human tumor samples aids in the understanding of potential lineage evolution in human breast cancer subtypes. Well-known luminal and basal markers were used to define lineage-restricted luminal and basal cells and identify lumino-basal populations; however, these markers represent only a small subset of the genes expressed by luminal, basal and lumino-basal cells. While the luminal and basal gene expression signatures have been described in the normal mammary gland and tumors of both human and murine animal models[95,102], the lumino-

basal gene signature is yet to be defined. This study has attempted to identify a lumino-basal subpopulation by scRNA-seq analysis of mouse mammary tumor cells, but the small sample size of 2 mouse tumors is insufficient to determine a lumino-basal gene signature. Furthermore, while it may be useful to define a mouse tumor lumino-basal signature, it is more beneficial to identify a lumino-basal signature of human breast tumors, as this will aid in the understanding of the luminal-to-basal transition in the human disease context, while possibly providing new prognostic markers to predict the progression and therapeutic response of breast cancer in patients. In order to elucidate the human lumino-basal tumor cell signature, the lumino-basal population from a larger sample of human low-ER mammary tumors can be identified by scRNA-seq analysis. Comparing the gene expression profiles from this population against a lineage-restricted luminal population from the same samples would thus reveal genes enriched specifically in the lumino-basal cell population.

Preventing the metastatic spread of cancer is a problem that has long plagued clinicians and cancer researchers, as the likelihood of patients to survive a cancer diagnosis after the onset of metastatic disease are greatly reduced. The function of PKA to reverse EMT and promote MET thus presents a potential targetable avenue by which to prevent the spread of this disease and improve patient survival. Multiple pharmacological agents targeting PKA have been designed, such as the compounds H89 and KT 5720, however these drugs antagonize PKA activity, and have only been utilized to inhibit PKA in the research setting[181]. Furthermore, it is PKA activation that is desirable in this context in order prevent EMT, as PKA inhibition has been shown to promote tumorigenesis instead, and as such, designing PKA agonists would appear to be a beneficial strategy to reduce metastasis. Alternatively, other proteins within the PKA activation pathway may be targeted, as designing a compound to activate a protein may be more challenging than identifying a drug to inhibit it. This requires a more comprehensive understanding of the mechanism of action of PKA induced plasticity, and the signaling pathways activated to prevent EMT in tumor cells.

PKA is a protein kinase known to promote differentiation and exhaustion of the stem cell compartment in the epidermis[141], and promote MET in breast cancer cell lines[139], however the pathways activated by this kinase in order to curtail EMT in mammary tumors remains to be fully explored. Previous studies have provided clues as to how PKA functions to inhibit plasticity. In epidermal stem cells, PKA has been found to induce differentiation by inhibiting YAP and GLI transcription pathways[141], however there is no evidence for GLI activation in PKA induced MET of mammary tumor cells[139]. Instead, Sox4 and PHF2 have been implicated as important downstream effectors of PKA activity[114,139], suggesting that the mechanisms in which PKA induces MET and prevents EMT in the mammary tumor context are different from the processes which promote differentiation and exhaustion of the epidermal stem cell compartment. Sox4 is a transcription factor, whereas PHF2 is a histone lysine demethylase which regulates chromatin activation state, thus these 2 proteins may work in concert to regulate PKA-induced epigenetic changes by first remodeling the chromatin landscape, and subsequently activating genes that function to reverse EMT. Further studies have to be carried out to delineate the mechanisms of Sox4 and PHF2 activity in curtailing EMT in mammary tumors, along with other proteins that may play a role in this pathway.

Eribulin's ability to induce MET in tumor cells suggests that it may provide benefit as a treatment for early-stage breast cancer. This drug is currently used as a treatment for patients with advanced breast cancer, where development of metastatic disease has already occurred[163]. The results from this study suggests that eribulin may be better utilized as a first-line chemotherapy, or a pre-treatment therapy used to prevent the emergence of metastatic disease while the primary tumor is being managed. Furthermore, inducing an epithelial phenotype in tumors through eribulin pre-treatment may sensitize them to treatment with subsequent chemotherapeutic agents, as epithelial-like tumor cells are known to be more sensitive to chemotherapy than their more resistant mesenchymal counterparts. Further clinical studies must first be carried

out to identify if using eribulin before any other treatments would be beneficial for patient survival.

Identification of Lamarckian induction as a mode of resistance against eribulin treatment has shown how extrinsic factors can influence the plasticity and heterogeneity of tumor cells, however the intrinsic pathways that are activated in response to eribulin activity is still unknown. Eribulin functions to eliminate tumor cells by binding to microtubule ends and inhibiting growth, thus blocking dynamic instability of microtubules and preventing proliferation[164]. The altered transcriptomic output and MET as a result of eribulin treatment suggests that eribulin may have other functions besides microtubule binding, which may involve direct activation of transcriptional programs leading to the expression of epithelial genes and repression of mesenchymal genes, or histone modification effects that could alter the epigenetic state of the cell. It is thus essential to elucidate the mechanism by which eribulin promotes transcriptomic and phenotypic alteration, in order to evaluate the potential utility of this drug as an epigenetic modifier for use as a pre-treatment for cancer.

This work focuses on 3 specific areas of lineage plasticity and heterogeneity, namely low-ER heterogeneity, EMT heterogeneity, and drug induced transcriptomic changes that lead to drug resistance. Due to the nature of phenotypic heterogeneity and transcriptomic alterations, there are infinite ways in which gene expression profiles can vary, leading to the emergence of new tumor subpopulations with varying lineages and phenotypes. Identifying how each new phenotype emerges, and how it affects tumor development is therefore challenging, and the complexity of a heterogeneous tumor may never be fully appreciated. Adding to the challenge is the possibility that tumors may harbor extremely rare subpopulations of cells that may not be easily detected, but may play an essential role in metastasis or chemotherapy resistance. In recent years, elucidating the extent of intra-tumoral heterogeneity has been facilitated by the advent of single-cell sequencing technologies that have helped to uncover the unique genetic, transcriptomic, and

epigenetic properties of each cell within a tumor. This has paved the way towards identifying new and hybrid phenotypes, as used in this study to identify the lumino-basal cell population in low-ER tumors, as well as reveal potential mechanisms leading to the emergence of these phenotypes. While the enormous task of understanding tumor heterogeneity as a whole may seem daunting, seemingly small discoveries, like the ones presented in this study, will contribute towards the larger body of work that has already been, and yet to be, discovered, to hopefully provide a more meaningful appreciation and understanding of this complex disease.

References

1. Dexter DL, Kowalski HM, Blazar BA, Fligiel Z, Vogel R, Heppner GH. Heterogeneity of tumor cells from a single mouse mammary tumor. *Cancer Res*. 1978;38(10):3174-3181. http://www.ncbi.nlm.nih.gov/pubmed/210930. Accessed May 15, 2019.

2. Heppner GH, Dexter DL, DeNucci T, Miller FR, Calabresi P. Heterogeneity in drug sensitivity among tumor cell subpopulations of a single mammary tumor. *Cancer Res*. 1978;38(11 Pt 1):3758-3763. http://www.ncbi.nlm.nih.gov/pubmed/698935. Accessed May 15, 2019.

3. Saunders NA, Simpson F, Thompson EW, et al. Role of intratumoural heterogeneity in cancer drug resistance: molecular and clinical perspectives. *EMBO Mol Med*. 2012;4(8):675-684. doi:10.1002/emmm.201101131

4. Navin N, Krasnitz A, Rodgers L, et al. Inferring tumor progression from genomic heterogeneity. *Genome Res*. 2010;20(1):68-80. doi:10.1101/gr.099622.109

5. Rye IH, Trinh A, Saetersdal AB, et al. Intratumor heterogeneity defines treatment-resistant HER2+ breast tumors. *Mol Oncol*. 2018;12(11):1838-1855. doi:10.1002/1878-0261.12375

6. Shipitsin M, Campbell LL, Argani P, et al. Molecular Definition of Breast Tumor Heterogeneity. *Cancer Cell*. 2007;11(3):259-273. doi:10.1016/j.ccr.2007.01.013

7. Sharma A, Merritt E, Hu X, et al. Non-Genetic Intra-Tumor Heterogeneity Is a Major Predictor of Phenotypic Heterogeneity and Ongoing Evolutionary Dynamics in Lung Tumors. *Cell Rep*. 2019;29(8):2164-2174.e5. doi:10.1016/j.celrep.2019.10.045

8. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic plasticity as a driver of cancer formation, progression and resistance to therapy. *Cell Stem Cell*. 2019;24(1):65. doi:10.1016/J.STEM.2018.11.011

9. Medication & Treatments | ADA. https://diabetes.org/healthy-living/medication-treatments. Accessed December 20, 2022.

10. Anaphylaxis - Symptoms and causes - Mayo Clinic. https://www.mayoclinic.org/diseases-conditions/anaphylaxis/symptoms-causes/syc-20351468. Accessed December 20, 2022.

11. Janku F. Tumor heterogeneity in the clinic: is it a real problem? *Ther Adv Med Oncol*. 2014;6(2):43. doi:10.1177/1758834013517414

12. Caswell DR, Swanton C. The role of tumour heterogeneity and clonal cooperativity in metastasis, immune evasion and clinical outcome. *BMC Med*. 2017;15(1). doi:10.1186/S12916-017-0900-Y

13. Hartl D, Leboulleux S, Hadoux J, et al. TNM Classification. *Surg Thyroid Parathyr Gland*. February 2022:440-446. doi:10.1016/B978-0-323-66127-0.00047-8

14. Cancer Staging. https://www.cancer.org/treatment/understanding-your-diagnosis/staging.html. Accessed December 20, 2022.

15. Llovet JM, Ducreux M, Lencioni R, et al. EASL–EORTC Clinical Practice Guidelines: Management of hepatocellular carcinoma. *J Hepatol*. 2012;56(4):908-943. doi:10.1016/J.JHEP.2011.12.001

16. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nat 2000 4066797*. 2000;406(6797):747-752. doi:10.1038/35021093

17. Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A*. 2001;98(19):10869-10874. doi:10.1073/pnas.191367098

18. Verhaak RGW, Hoadley KA, Purdom E, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17(1):98-110. doi:10.1016/J.CCR.2009.12.020

19. Robertson-Tessi M, Gillies RJ, Gatenby RA, Anderson ARA. Impact of metabolic heterogeneity on tumor growth, invasion, and treatment outcomes. *Cancer Res*. 2015;75(8):1567. doi:10.1158/0008-5472.CAN-14-1428

20. Sharma A, Cao EY, Kumar V, et al. Longitudinal single-cell RNA sequencing of patient-derived primary cells reveals drug-induced infidelity in stem cell hierarchy. *Nat Commun 2018 91*. 2018;9(1):1-17. doi:10.1038/s41467-018-07261-3

21. Reichert M, Bakir B, Moreira L, et al. Regulation of Epithelial Plasticity Determines Metastatic Organotropism in Pancreatic Cancer. *Dev Cell*. 2018;45(6):696-711.e8. doi:10.1016/J.DEVCEL.2018.05.025

22. Gerlinger M, Rowan AJ, Horswell S, et al. Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N Engl J Med*. 2012;366(10):883-892. doi:10.1056/NEJMOA1113205/SUPPL_FILE/NEJMOA1113205_DISCLOSURES.PDF

23. Shibata D, Schaeffer J, Li ZH, Capella G, Perucho M. Genetic heterogeneity of the c-K-ras locus in colorectal adenomas but not in adenocarcinomas. *J Natl Cancer Inst*. 1993;85(13):1058-1063. doi:10.1093/JNCI/85.13.1058

24. Marusyk A, Almendro V, Polyak K. Intra-tumour heterogeneity: a looking glass for cancer? *Nat Rev Cancer*. 2012;12(5):323-334. doi:10.1038/nrc3261

25. Nieto MA, Huang RYYJ, Jackson RAA, Thiery JPP. EMT: 2016. *Cell*. 2016;166(1):21-45. doi:10.1016/J.CELL.2016.06.028

26. Shaffer SM, Dunagin MC, Torborg SR, et al. Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. *Nat 2017 5467658*. 2017;546(7658):431-435. doi:10.1038/nature22794

27. Marusyk A, Polyak K. Tumor heterogeneity: causes and consequences. *Biochim Biophys Acta*. 2010;1805(1):105-117. doi:10.1016/j.bbcan.2009.11.002

28. Hwang KT, Kim J, Jung J, et al. Impact of breast cancer subtypes on prognosis of women with operable invasive breast cancer: A Population-based Study Using SEER Database. *Clin Cancer Res*. 2019;25(6):1970-1979. doi:10.1158/1078-0432.CCR-18-2782/74282/AM/IMPACT-OF-BREAST-CANCER-SUBTYPES-ON-PROGNOSIS-OF

29. Hennigs A, Riedel F, Gondos A, et al. Prognosis of breast cancer molecular subtypes in routine clinical care: A large prospective cohort study. *BMC Cancer*. 2016;16(1):1-9. doi:10.1186/S12885-016-2766-3/FIGURES/2

30. Fallahpour S, Navaneelan T, De P, Borgo A. Breast cancer survival by molecular subtype: a population-based analysis of cancer registry data. *Can Med Assoc Open Access J*. 2017;5(3):E734-E739. doi:10.9778/CMAJO.20170030

31. Carey LA, Perou CM, Livasy CA, et al. Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295(21):2492-2502. doi:10.1001/JAMA.295.21.2492

32. Yersal O, Barutca S. Biological subtypes of breast cancer: Prognostic and therapeutic implications. *World J Clin Oncol*. 2014;5(3):412-424. doi:10.5306/wjco.v5.i3.412

33. Guarneri V, Conte P. Metastatic Breast Cancer: Therapeutic Options According to Molecular Subtypes and Prior Adjuvant Therapy. *Oncologist*. 2009;14(7):645-656. doi:10.1634/THEONCOLOGIST.2009-0078

34. Kennecke H, Yerushalmi R, Woods R, et al. Metastatic behavior of breast cancer subtypes. *J Clin Oncol*. 2010;28(20):3271-3277. doi:10.1200/JCO.2009.25.9820

35. Allison KH, Hammond MEH, Dowsett M, et al. Estrogen and Progesterone Receptor Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Guideline Update. *Arch Pathol Lab Med*. 2020;144(5):545-563. doi:10.5858/ARPA.2019-0904-SA

36. Tsutsui S, Ohno S, Murakami S, Kataoka A, Kinoshita J, Hachitanda Y. Prognostic significance of the coexpression of p53 protein and c-erbB2 in breast cancer. *Am J Surg*. 2003;185(2):165-167. doi:10.1016/S0002-9610(02)01203-5

37. Kunte S, Abraham J, Montero AJ. Novel HER2–targeted therapies for HER2–positive metastatic breast cancer. *Cancer*. 2020;126(19):4278-4288. doi:10.1002/CNCR.33102

38. Heitz F, Harter P, Lueck HJ, et al. Triple-negative and HER2-overexpressing breast cancers exhibit an elevated risk and an earlier occurrence of cerebral metastases. *Eur J Cancer*. 2009;45(16):2792-2798. doi:10.1016/J.EJCA.2009.06.027

39. Bernard PS, Parker JS, Mullins M, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160-1167. doi:10.1200/JCO.2008.18.1370

40. Hammond MEH, Hayes DF, Dowsett M, et al. American society of clinical oncology/college of american pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. *J Clin Oncol*. 2010;28(16):2784-2795. doi:10.1200/JCO.2009.25.6529

41. Cottu PH, Asselah J, Lae M, et al. Intratumoral heterogeneity of HER2/neu expression and its consequences for the management of advanced breast cancer [2]. *Ann Oncol*. 2008;19(3):596-597. doi:10.1093/annonc/mdn021

42. Turashvili G, Brogi E. Tumor Heterogeneity in Breast Cancer. *Front Med*. 2017;4(DEC):227. doi:10.3389/FMED.2017.00227

43. Seol H, Lee HJ, Choi Y, et al. Intratumoral heterogeneity of HER2 gene amplification in breast cancer: its clinicopathological significance. *Mod Pathol 2012 257*. 2012;25(7):938-948. doi:10.1038/modpathol.2012.36

44. Vance GH, Barry TS, Bloom KJ, et al. Genetic Heterogeneity in HER2 Testing in Breast Cancer: Panel Summary and Guidelines. *Arch Pathol Lab Med*. 2009;133(4):611-612. doi:10.5858/133.4.611

45. Chung W, Eum HH, Lee HO, et al. Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nat Commun*. 2017;8. doi:10.1038/NCOMMS15081

46. Yang F, Wang Y, Li Q, et al. Intratumor heterogeneity predicts metastasis of triple-negative breast cancer. *Carcinogenesis*. 2017;38(9):900-909. doi:10.1093/CARCIN/BGX071

47. Nowell PC. The clonal evolution of tumor cell populations. *Science*. 1976;194(4260):23-28. http://www.ncbi.nlm.nih.gov/pubmed/959840. Accessed May 19, 2019.

48. Darwin C. On the origin of species by means of natural selection, or, The preservation of favoured races in the struggle for life. *London J Murray*. 1859.

49. Vendramin R, Litchfield K, Swanton C. Cancer evolution: Darwin and beyond. *EMBO J*. 2021;40(18):e108389. doi:10.15252/EMBJ.2021108389

50. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. Cancer genome landscapes. *Science (80-)*. 2013;340(6127):1546-1558. doi:10.1126/SCIENCE.1235122/SUPPL_FILE/VOGELSTEIN.SM.COVER.PAGE.PDF

51. Meacham CE, Morrison SJ. Tumor heterogeneity and cancer cell plasticity. *Nature*. 2013;501(7467):328. doi:10.1038/NATURE12624

52. Fearon ER, Vogelstein B. A genetic model for colorectal tumorigenesis. *Cell*. 1990;61(5):759-767. doi:10.1016/0092-8674(90)90186-I

53. Baisse B, Bouzourene H, Saraga EP, Bosman FT, Benhattar J. Intratumor genetic heterogeneity in advanced human colorectal adenocarcinoma. *Int J Cancer*. 2001;93(3):346-352. doi:10.1002/IJC.1343

54. Molinari C, Marisi G, Passardi A, Matteucci L, De Maio G, Ulivi P. Heterogeneity in Colorectal Cancer: A Challenge for Personalized Medicine? *Int J Mol Sci*. 2018;19(12). doi:10.3390/IJMS19123733

55. Büttner J, Jöhrens K, Klauschen F, et al. Intratumoral morphological heterogeneity can be an indicator of genetic heterogeneity in colorectal cancer. *Exp Mol Pathol*. 2018;104(1):76-81. doi:10.1016/J.YEXMP.2018.01.007

56. Jones HG, Jenkins G, Williams N, et al. Genetic and Epigenetic Intra-tumour Heterogeneity in Colorectal Cancer. *World J Surg*. 2017;41(5):1375-1383. doi:10.1007/S00268-016-3860-Z/FIGURES/6

57. Laurent-Puig P, Cayre A, Manceau G, et al. Analysis of PTEN, BRAF, and EGFR status in determining benefit from cetuximab therapy in wild-type KRAS metastatic colon cancer. *J Clin Oncol*. 2009;27(35):5924-5930. doi:10.1200/JCO.2008.21.6796

58. Guo M, Peng Y, Gao A, Du C, Herman JG. Epigenetic heterogeneity in cancer. *Biomark Res 2019 71*. 2019;7(1):1-19. doi:10.1186/S40364-019-0174-Y

59. Lapidot T, Sirard C, Vormoor J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nat 1994 3676464*. 1994;367(6464):645-648. doi:10.1038/367645a0

60. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. *Proc Natl Acad Sci U S A*. 2003;100(7):3983-3988. doi:10.1073/PNAS.0530291100/ASSET/C251EFA2-0831-42E3-931C-CFA289B5DAF4/ASSETS/GRAPHIC/PQ0530291004.JPEG

61. Phi LTH, Sari IN, Yang YG, et al. Cancer Stem Cells (CSCs) in Drug Resistance and their Therapeutic Implications in Cancer Treatment. *Stem Cells Int*. 2018;2018. doi:10.1155/2018/5416923

62. Chen J, Li Y, Yu TS, et al. A restricted cell population propagates glioblastoma growth after chemotherapy. *Nat 2012 4887412*. 2012;488(7412):522-526. doi:10.1038/nature11287

63. Farmer P, Bonnefoi H, Anderle P, et al. A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer. *Nat Med 2008 151*. 2009;15(1):68-74. doi:10.1038/nm.1908

64. Donnenberg VS, Donnenberg AD. Multiple Drug Resistance in Cancer Revisited: The Cancer Stem Cell Hypothesis. *J Clin Pharmacol*. 2005;45(8):872-877. doi:10.1177/0091270005276905

65. Bao S, Wu Q, McLendon RE, et al. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. *Nat 2006 4447120*. 2006;444(7120):756-760. doi:10.1038/nature05236

66. Quintana E, Shackleton M, Sabel MS, Fullen DR, Johnson TM, Morrison SJ. Efficient tumour formation by single human melanoma cells. *Nat 2008 4567222*. 2008;456(7222):593-598. doi:10.1038/nature07567

67. Hill RP. Identifying Cancer Stem Cells in Solid Tumors: Case Not Proven. *Cancer Res*. 2006;66(4):1891-1896. doi:10.1158/0008-5472.CAN-05-3450

68. Quintanal-Villalonga Á, Chan JM, Yu HA, et al. Lineage plasticity in cancer: a shared pathway of therapeutic resistance. *Nat Rev Clin Oncol*. 2020;17(6):360. doi:10.1038/S41571-020-0340-Z

69. Serrano-Gomez SJ, Maziveyi M, Alahari SK. Regulation of epithelial-mesenchymal transition through epigenetic and post-translational modifications. *Mol Cancer 2016 151*. 2016;15(1):1-14. doi:10.1186/S12943-016-0502-X

70. Thiery JP, Acloque H, Huang RYJ, Nieto MA. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell*. 2009;139(5):871-890. doi:10.1016/J.CELL.2009.11.007

71. Anjomshoaa A, Nasri S, Humar B, et al. Slow proliferation as a biological feature of colorectal cancer metastasis. *Br J Cancer 2009 1015*. 2009;101(5):822-828. doi:10.1038/sj.bjc.6605229

72. Aggarwal R, Huang J, Alumkal JJ, et al. Clinical and genomic characterization of treatment-emergent small-cell neuroendocrine prostate cancer: A multi-institutional prospective study. *J Clin Oncol*. 2018;36(24):2492-2503. doi:10.1200/JCO.2017.77.6880

73. Pascual T, Oliveira M, Villagrasa P, et al. Neoadjuvant eribulin in HER2-negative early-stage breast cancer (SOLTI-1007-NeoEribulin): a multicenter, two-cohort, non-randomized phase II trial. *npj Breast Cancer 2021 71*. 2021;7(1):1-11. doi:10.1038/s41523-021-00351-4

74. Bhatia K, Bhumika, Das A. Combinatorial drug therapy in cancer - New insights. *Life Sci*. 2020;258:118134. doi:10.1016/J.LFS.2020.118134

75. Wang X, Zhang H, Chen X. Drug resistance and combating drug resistance in cancer. *Cancer Drug Resist*. 2019;2(2):141. doi:10.20517/CDR.2019.10

76. Marusyk A, Janiszewska M, Polyak K. Intratumor Heterogeneity: The Rosetta Stone of Therapy Resistance. *Cancer Cell*. 2020;37(4):471-484. doi:10.1016/J.CCELL.2020.03.007

77. Sharma S V., Lee DY, Li B, et al. A Chromatin-Mediated Reversible Drug-Tolerant State in Cancer Cell Subpopulations. *Cell*. 2010;141(1):69-80. doi:10.1016/J.CELL.2010.02.027

78. Yoshida T, Ozawa Y, Kimura T, et al. Eribulin mesilate suppresses experimental metastasis of breast cancer cells by reversing phenotype from epithelial–mesenchymal transition (EMT) to mesenchymal–epithelial transition (MET) states. *Br J Cancer 2014 1106*. 2014;110(6):1497-1505. doi:10.1038/bjc.2014.80

79. Iwamoto T, Booser D, Valero V, et al. Estrogen Receptor (ER) mRNA and ER-related gene expression in breast cancers that are 1% to 10% ER-positive by immunohistochemistry. *J Clin Oncol*. 2012;30(7):729-734. doi:10.1200/JCO.2011.36.2574

80. Mikaelian I, Hovick M, Silva KA, et al. Expression of Terminal Differentiation Proteins Defines Stages of Mouse Mammary Gland Development. *Vet Pathol*. 2006;43(1):36-49. doi:10.1354/vp.43-1-36

81. Sun P, Yuan Y, Li A, Li B, Dai X. Cytokeratin expression during mouse embryonic and early postnatal mammary gland development. *Histochem Cell Biol*. 2010;133(2):213-221. doi:10.1007/s00418-009-0662-5

82. Van Keymeulen A, Rocha AS, Ousset M, et al. Distinct stem cells contribute to mammary gland development and maintenance. *Nature*. 2011;479(7372):189-193. doi:10.1038/nature10573

83. Fu NY, Rios AC, Pal B, et al. Identification of quiescent and spatially restricted mammary stem cells that are hormone responsive. *Nat Cell Biol*. 2017;19(3):164-176. doi:10.1038/NCB3471

84. Wuidart A, Sifrim A, Fioramonti M, et al. Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat Cell Biol*. 2018;20(6):666-676. doi:10.1038/S41556-018-0095-2

85. Lilja AM, Rodilla V, Huyghe M, et al. Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland. *Nat Cell Biol*. 2018;20(6):677-687. doi:10.1038/S41556-018-0108-1

86. Song W, Wang R, Jiang W, et al. Hormones induce the formation of luminal-derived basal cells in the mammary gland. *Cell Res 2019 293*. 2019;29(3):206-220. doi:10.1038/s41422-018-0137-0

87. Centonze A, Lin S, Tika E, et al. Heterotypic cell–cell communication regulates glandular stem cell multipotency. *Nature*. 2020;584(7822):608-613. doi:10.1038/s41586-020-2632-y

88. Lim E, Vaillant F, Wu D, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med 2009 158*. 2009;15(8):907-913. doi:10.1038/nm.2000

89. Molyneux G, Geyer FC, Magnay F-A, et al. BRCA1 Basal-like Breast Cancers Originate from Luminal Epithelial Progenitors and Not from Basal Stem Cells. *Cell Stem Cell*. 2010;7(3):403-417. doi:10.1016/j.stem.2010.07.010

90. Rädler PD, Wehde BL, Triplett AA, et al. Highly metastatic claudin-low mammary cancers can originate from luminal epithelial cells. *Nat Commun*. 2021;12(1):1-16. doi:10.1038/s41467-021-23957-5

91. Wolff AC, Elizabeth Hale Hammond M, Allison KH, et al. Human epidermal growth factor receptor 2 testing in breast cancer: American society of clinical oncology/ college of American pathologists clinical practice guideline focused update. *J Clin Oncol*. 2018;36(20):2105-2122. doi:10.1200/JCO.2018.77.8738

92. Guy CT, Cardiff RD, Muller WJ. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol*. 1992;12(3):954-961. doi:10.1128/MCB.12.3.954-961.1992

93. Muzumdar MD, Tasic B, Miyamichi K, Li N, Luo L. A global double-fluorescent Cre reporter mouse. *genesis*. 2007;45(9):593-605. doi:10.1002/DVG.20335

94. Prater M, Shehata M, Watson CJ, Stingl J. Enzymatic dissociation, flow cytometric analysis, and culture of normal mouse mammary tissue. *Methods Mol Biol*. 2013;946:395-409. doi:10.1007/978-1-62703-128-8_25/FIGURES/00255

95. Pal B, Chen Y, Vaillant F, et al. A single-cell RNA expression atlas of normal, preneoplastic and tumorigenic states in the human breast. *EMBO J*. 2021;40(11):e107333. doi:10.15252/embj.2020107333

96. Zheng GXY, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun 2017 81*. 2017;8(1):1-12. doi:10.1038/ncomms14049

97. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol 2015 335*. 2015;33(5):495-502. doi:10.1038/nbt.3192

98. Germain P-L, Lun A, Macnair W, et al. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Research 2021 10979*. 2021;10:979. doi:10.12688/f1000research.73600.1

99. McCarthy DJ, Campbell KR, Lun ATL, Wills QF. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics*. 2017;33(8):1179-1186. doi:10.1093/BIOINFORMATICS/BTW777

100. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol*. 2019;20(1):1-15. doi:10.1186/S13059-019-1874-1/FIGURES/6

101. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience*. 2018;7(7):1-9. doi:10.1093/GIGASCIENCE/GIY083

102. Pal B, Chen Y, Milevskiy MJG, et al. Single cell transcriptome atlas of mouse mammary epithelial cells across development. *Breast Cancer Res*. 2021;23(1):1-19. doi:10.1186/S13058-021-01445-4/FIGURES/7

103. Frost HR. Variance-adjusted Mahalanobis (VAM): a fast and accurate method for cell-specific gene set scoring. *Nucleic Acids Res*. 2020;48(16):e94-e94. doi:10.1093/NAR/GKAA582

104. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207-210. doi:10.1093/NAR/30.1.207

105. Gloyeske NC, Dabbs DJ, Bhargava R. Low ER+ breast cancer: Is this a distinct group? *Am J Clin Pathol*. 2014;141(5):697-701. doi:10.1309/AJCP34CYSATWFDPQ

106. Landmann A, Farrugia DJ, Zhu L, et al. Low Estrogen Receptor (ER)–Positive Breast Cancer and Neoadjuvant Systemic ChemotherapyIs Response Similar to Typical ER-Positive or ER-Negative Disease? *Am J Clin Pathol*. 2018;150(1):34-42. doi:10.1093/AJCP/AQY028

107. Deyarmin B, Kane JL, Valente AL, et al. Effect of ASCO/CAP guidelines for determining ER status on molecular subtype. *Ann Surg Oncol*. 2012;20(1):87-93. doi:10.1245/S10434-012-2588-8

108. Roy S, Axelrod HD, Valkenburg KC, Amend S, Pienta KJ. Optimization of prostate cancer cell detection using multiplex tyramide signal amplification. *J Cell Biochem*. 2019;120(4):4804. doi:10.1002/JCB.28016

109. Lazarus J, Akiska Y, Lanfranca MP, et al. Optimization, Design and Avoiding Pitfalls in Manual Multiplex Fluorescent Immunohistochemistry. *J Vis Exp*. 2019;2019(149). doi:10.3791/59915

110. Brown MS, Abdollahi B, Wilkins OM, et al. Phenotypic heterogeneity driven by plasticity of the intermediate EMT state governs disease progression and metastasis in breast cancer. *Sci Adv*. 2022;8(31):8002. doi:10.1126/SCIADV.ABJ8002

111. Brown MS, Abdollahi B, Hassanpour S, Pattabiraman DR. Quantifying epithelial-mesenchymal heterogeneity and EMT scoring in tumor samples via tyramide signal amplification (TSA). *Methods Cell Biol*. 2022;171:149-161. doi:10.1016/BS.MCB.2022.06.003

112. Pfefferle AD, Herschkowitz JI, Usary J, et al. Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biol*. 2013;14(11):R125. doi:10.1186/gb-2013-14-11-r125

113. Lin EY, Jones JG, Li P, et al. Progression to Malignancy in the Polyoma Middle T Oncoprotein Mouse Breast Cancer Model Provides a Reliable Model for Human Diseases. *Am J Pathol*. 2003;163(5):2113. doi:10.1016/S0002-9440(10)63568-7

114. Ognjenovic NB, Bagheri M, Mohamed GA, et al. Limiting Self-Renewal of the Basal Compartment by PKA Activation Induces Differentiation and Alters the Evolution of Mammary Tumors. *Dev Cell*. 2020;55(5):544-557.e6. doi:10.1016/j.devcel.2020.10.004

115. Jia R, Liang Y, Chen R, et al. Osteopontin facilitates tumor metastasis by regulating epithelial–mesenchymal plasticity. *Cell Death Dis 2016 712*. 2016;7(12):e2564-e2564. doi:10.1038/cddis.2016.422

116. Zohar R, Suzuki N, Suzuki K, et al. Intracellular Osteopontin Is an Integral Component of the CD44-ERM Complex Involved in Cell Migration. *J Cell Physiol*. 2000;184:118-130. doi:10.1002/(SICI)1097-4652(200007)184:1

117. Gillan L, Matei D, Fishman DA, Gerbin CS, Karlan BY, Chang DD. Periostin Secreted by Epithelial Ovarian Carcinoma Is a Ligand for V 3 and V 5 Integrins and Promotes Cell Motility 1. *CANCER Res*. 2002;62:5358-5364. http://aacrjournals.org/cancerres/article-pdf/62/18/5358/2497408/ch1802005358.pdf. Accessed August 31, 2022.

118. Dravis C, Chung C-Y, Lytle NK, et al. Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. *Cancer Cell*. 2018;34(3):466-482.e6. doi:10.1016/j.ccell.2018.08.001

119. Han H, Cho JW, Lee S, et al. TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. *Nucleic Acids Res*. 2018;46(D1):D380-D386. doi:10.1093/NAR/GKX1013

120. Li Y, Lv Z, Zhang S, et al. Genetic Fate Mapping of Transient Cell Fate Reveals N-Cadherin Activity and Function in Tumor Metastasis. *Dev Cell*. 2020;54(5):593-607.e5. doi:10.1016/J.DEVCEL.2020.06.021

121. Hein SM, Haricharan S, Johnston AN, et al. Luminal epithelial cells within the mammary gland can produce basal cells upon oncogenic stress. *Oncogene 2016 3511*. 2015;35(11):1461-1467. doi:10.1038/onc.2015.206

122. Koren S, Reavie L, Couto JP, et al. PIK3CAH1047R induces multipotency and multi-lineage mammary tumours. *Nat 2015 5257567*. 2015;525(7567):114-118. doi:10.1038/nature14669

123. Van Keymeulen A, Lee MY, Ousset M, et al. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nat 2015 5257567*. 2015;525(7567):119-123. doi:10.1038/nature14665

124. Guy CT, Muthuswamy SK, Cardiff RD, Soriano P, Muller WJ. Activation of the c-Src tyrosine kinase is required for the induction of mammary tumors in transgenic mice. *Genes Dev*. 1994;8(1):23-32. doi:10.1101/GAD.8.1.23

125. Campbell KS, Ogris E, Burke B, et al. Polyoma middle tumor antigen interacts with SHC protein via the NPTY (Asn-Pro-Thr-Tyr) motif in middle tumor antigen. *Proc Natl Acad Sci*. 1994;91(14):6344-6348. doi:10.1073/PNAS.91.14.6344

126. Whitman M, Kaplan DR, Schaffhausen B, Cantley L, Roberts TM. Association of phosphatidylinositol kinase activity with polyoma middle-T competent for transformation. *Nat 1985 3156016*. 1985;315(6016):239-242. doi:10.1038/315239a0

127. Cheung KJ, Gabrielson E, Werb Z, Ewald AJ. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell*. 2013;155(7):1639-1651. doi:10.1016/J.CELL.2013.11.029/ATTACHMENT/2A1FD816-3063-4634-A607-E4C61311C346/MMC8.PDF

128. Lüönd F, Sugiyama N, Bill R, et al. Distinct contributions of partial and full EMT to breast cancer malignancy. *Dev Cell*. 2021;56(23):3203-3221.e11. doi:10.1016/j.devcel.2021.11.006

129. Tiede S, Kalathur RKR, Lüönd F, et al. Multi-color clonal tracking reveals intra-stage proliferative heterogeneity during mammary tumor progression. *Oncogene 2020 401*. 2020;40(1):12-27. doi:10.1038/s41388-020-01508-4

130. McCarthy A, Savage K, Gabriel A, Naceur C, Reis-Filho JS, Ashworth A. A mouse model of basal-like breast carcinoma with metaplastic elements. *J Pathol*. 2007;211(4):389-398. doi:10.1002/PATH.2124

131. Rios AC, Fu NY, Lindeman GJ, Visvader JE. In situ identification of bipotent stem cells in the mammary gland. *Nat 2014 5067488*. 2014;506(7488):322-327. doi:10.1038/nature12948

132. Cimino-Mathews A, Subhawong AP, Elwood H, et al. Neural crest transcription factor Sox10 is preferentially expressed in triple-negative and metaplastic breast carcinomas. *Hum Pathol*. 2013;44(6):959-965. doi:10.1016/J.HUMPATH.2012.09.005

133. Rammal R, Goel K, Elishaev E, et al. The Utility of SOX10 Immunohistochemical Staining in Breast PathologyStaining of Myoepithelial Cells, Distinction of Atypical Ductal Hyperplasia From Usual Ductal Hyperplasia, and Confirming Breast Origin in Triple-Negative Breast Cancer. *Am J Clin Pathol*. August 2022. doi:10.1093/AJCP/AQAC092

134. Saunus JM, De Luca XM, Northwood K, et al. Epigenome erosion and SOX10 drive neural crest phenotypic mimicry in triple-negative breast cancer. *npj Breast Cancer 2022 81*. 2022;8(1):1-16. doi:10.1038/s41523-022-00425-x

135. Arnoux V, Nassour M, L'Helgoualc'h A, Hipskind RA, Savagner P. Erk5 controls slug expression and keratinocyte activation during wound healing. *Mol Biol Cell*. 2008;19(11):4738-4749. doi:10.1091/MBC.E07-10-1078/ASSET/IMAGES/LARGE/ZMK0110887510010.JPEG

136. Koike Y, Yozaki M, Utani A, Murota H. Fibroblast growth factor 2 accelerates the epithelial–mesenchymal transition in keratinocytes during wound healing process. *Sci Reports 2020 101*. 2020;10(1):1-13. doi:10.1038/s41598-020-75584-7

137. Fabisiewicz A, Grzybowska E. CTC clusters in cancer progression and metastasis. *Med Oncol*. 2017;34(1):1-10. doi:10.1007/S12032-016-0875-0/TABLES/1

138. Genna A, Vanwynsberghe AM, Villard A V., et al. EMT-Associated Heterogeneity in Circulating Tumor Cells: Sticky Friends on the Road to Metastasis. *Cancers (Basel)*. 2020;12(6):1-38. doi:10.3390/CANCERS12061632

139. Pattabiraman DR, Bierie B, Kober KI, et al. Activation of PKA leads to mesenchymal-to-epithelial transition and loss of tumor-initiating ability. *Science (80-)*. 2016;351(6277). doi:10.1126/SCIENCE.AAD3680/SUPPL_FILE/TABLES9.XLSX

140. Taskén K, Aandahl EM. Localized Effects of cAMP Mediated by Distinct Routes of Protein Kinase A. *Physiol Rev*. 2004;84(1):137-167. doi:10.1152/PHYSREV.00021.2003/ASSET/IMAGES/LARGE/9J0140287111.JPEG

141. Iglesias-Bartolome R, Torres D, Marone R, et al. Inactivation of a Gαs–PKA tumour suppressor pathway in skin stem cells initiates basal-cell carcinogenesis. *Nat Cell Biol 2014 176*. 2015;17(6):793-803. doi:10.1038/ncb3164

142. Orellana SA, Stanley Mcknight G. Mutations in the catalytic subunit of cAMP-dependent protein kinase result in unregulated biological activity. *Proc Natl Acad Sci*. 1992;89(10):4726-4730. doi:10.1073/PNAS.89.10.4726

143. Niswender CM, Willis BS, Wallen A, et al. Cre recombinase-dependent expression of a constitutively active mutant allele of the catalytic subunit of protein kinase A. *Genesis*. 2005;43(3):109-119. doi:10.1002/GENE.20159

144. Wagner KU, Ward T, Davis B, Wiseman R, Hennighausen L. Spatial and temporal expression of the Cre gene under the control of the MMTV-LTR in different lines of transgenic mice. *Transgenic Res*. 2001;10(6):545-553. doi:10.1023/A:1013063514007

145. Pertea G. gpertea/fqtrim: fqtrim release v0.9.7. February 2018. doi:10.5281/ZENODO.1185412

146. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res*. 2017;27(3):491-499. doi:10.1101/GR.209601.116

147. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21. doi:10.1093/BIOINFORMATICS/BTS635

148. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930. doi:10.1093/BIOINFORMATICS/BTT656

149. Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mRNA quantification and differential analysis with Census. *Nat Methods 2017 143*. 2017;14(3):309-315. doi:10.1038/nmeth.4150

150. Qiu X, Mao Q, Tang Y, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods 2017 1410*. 2017;14(10):979-982. doi:10.1038/nmeth.4402

151. Trapnell C, Cacchiarelli D, Grimsby J, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol 2014 324*. 2014;32(4):381-386. doi:10.1038/nbt.2859

152. Bolós V, Peinado H, Pérez-Moreno MA, Fraga MF, Esteller M, Cano A. The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. *J Cell Sci*. 2003;116(3):499-511. doi:10.1242/JCS.00224

153. Nassour M, Idoux-Gillet Y, Selmi A, et al. Slug Controls Stem/Progenitor Cell Growth Dynamics during Mammary Gland Morphogenesis. *PLoS One*. 2012;7(12):e53498. doi:10.1371/JOURNAL.PONE.0053498

154. Lee HW, Park YM, Lee SJ, et al. Alpha-smooth muscle actin (ACTA2) is required for metastatic potential of human lung adenocarcinoma. *Clin Cancer Res*. 2013;19(21):5879-5889. doi:10.1158/1078-0432.CCR-13-1181/85833/AM/ALPHA-SMOOTH-MUSCLE-ACTIN-ACTA2-IS-REQUIRED-FOR

155. Prater MD, Petit V, Alasdair Russell I, et al. Mammary stem cells have myoepithelial cell properties. *Nat Cell Biol 2014 1610*. 2014;16(10):942-950. doi:10.1038/ncb3025

156. Sarrió D, Rodriguez-Pinilla SM, Hardisson D, Cano A, Moreno-Bueno G, Palacios J. Epithelial-Mesenchymal Transition in Breast Cancer Relates to the Basal-like Phenotype. *Cancer Res*. 2008;68(4):989-997. doi:10.1158/0008-5472.CAN-07-2017

157. Lopez MS, Kliegman JI, Shokat KM. The Logic and Design of Analog-Sensitive Kinases and Their Small Molecule Inhibitors. *Methods Enzymol*. 2014;548(C):189-213. doi:10.1016/B978-0-12-397918-6.00008-2

158. Sledge GW, Neuberg D, Bernardo P, et al. Phase III Trial of Doxorubicin, Paclitaxel, and the Combination of Doxorubicin and Paclitaxel as Front-Line Chemotherapy for Metastatic Breast Cancer: An Intergroup Trial (E1193). *J Clin Oncol*. 2003;21(4):588-592. doi:10.1200/JCO.2003.08.013

159. Alba E, Martín M, Ramos M, et al. Multicenter randomized trial comparing sequential with concomitant administration of doxorubicin and docetaxel as first-line treatment of metastatic breast cancer: A Spanish Breast Cancer Research Group (GEICAM-9903) phase III study. *J Clin Oncol*. 2004;22(13):2587-2593. doi:10.1200/JCO.2004.08.125

160. Burrell RA, Swanton C. Tumour heterogeneity and the evolution of polyclonal drug resistance. *Mol Oncol*. 2014;8(6):1095-1111. doi:10.1016/j.molonc.2014.06.005

161. Carlino MS, Fung C, Shahheydari H, et al. Preexisting MEK1P124 mutations diminish response to BRAF inhibitors in metastatic melanoma patients. *Clin Cancer Res*. 2015;21(1):98-105. doi:10.1158/1078-0432.CCR-14-0759

162. Diaz LA, Williams RT, Wu J, et al. The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nat 2012 4867404*. 2012;486(7404):537-540. doi:10.1038/nature11219

163. Cortes J, O'Shaughnessy J, Loesch D, et al. Eribulin monotherapy versus treatment of physician's choice in patients with metastatic breast cancer (EMBRACE): a phase 3 open-label randomised study. *Lancet*. 2011;377(9769):914-923. doi:10.1016/S0140-6736(11)60070-6

164. Smith JA, Wilson L, Azarenko O, et al. Eribulin binds at microtubule ends to a single site on tubulin to suppress dynamic instability. *Biochemistry*. 2010;49(6):1331-1337. doi:10.1021/BI901810U/ASSET/IMAGES/LARGE/BI-2009-01810U_0002.JPEG

165. Satpathy AT, Granja JM, Yost KE, et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol 2019 378*. 2019;37(8):925-936. doi:10.1038/s41587-019-0206-z

166. Roelli P, bbimber, Flynn B, santiagorevale, Gui G. Hoohm/CITE-seq-Count: 1.4.2. March 2019. doi:10.5281/ZENODO.2590196

167. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888-1902.e21. doi:10.1016/J.CELL.2019.05.031

168. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nat 2019 5667745*. 2019;566(7745):496-502. doi:10.1038/s41586-019-0969-x

169. Farahpour, F., Saeedghalati, M., & Hoffmann D. MullerPlot: Generates Muller Plot from Population/Abundance/Frequency Dynamics Data version 0.1.3 from CRAN. https://rdrr.io/cran/MullerPlot/. Accessed November 18, 2022.

170. Korsunsky I, Millard N, Fan J, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods 2019 1612*. 2019;16(12):1289-1296. doi:10.1038/s41592-019-0619-0

171. McGinnis CS, Patterson DM, Winkler J, et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat Methods 2019 167*. 2019;16(7):619-626. doi:10.1038/s41592-019-0433-8

172. Tang M, Kaymaz Y, Logeman BL, et al. Evaluating single-cell cluster stability using the Jaccard similarity index. *Bioinformatics*. 2021;37(15):2212-2214. doi:10.1093/BIOINFORMATICS/BTAA956

173. Dongre A, Rashidian M, Reinhardt F, et al. Epithelial-to-Mesenchymal Transition Contributes to Immunosuppression in Breast Carcinomas. *Cancer Res*. 2017;77(15):3982-3989. doi:10.1158/0008-5472.CAN-16-3292

174. Bhang HC, Ruddy DA, Krishnamurthy Radhakrishna V, et al. Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nat Med*. 2015;21(5):440-448. doi:10.1038/nm.3841

175. Seth S, Li CY, Ho IL, et al. Pre-existing Functional Heterogeneity of Tumorigenic Compartment as the Origin of Chemoresistance in Pancreatic Tumors. *Cell Rep*. 2019;26(6):1518-1532.e9. doi:10.1016/j.celrep.2019.01.048

176. Chang MT, Shanahan F, Nguyen TTT, et al. Identifying transcriptional programs underlying cancer drug response with TraCe-seq. *Nat Biotechnol 2021 401*. 2021;40(1):86-93. doi:10.1038/s41587-021-01005-3

177. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. Whole-organism lineage tracing by combinatorial and cumulative genome editing. *Science (80-)*. 2016;353(6298). doi:10.1126/SCIENCE.AAF7907/SUPPL_FILE/MCKENNA-SM.PDF

178. Kalhor R, Kalhor K, Mejia L, et al. Developmental barcoding of whole mouse via homing CRISPR. *Science (80-)*. 2018;361(6405). doi:10.1126/SCIENCE.AAT9804/SUPPL_FILE/AAT9804_KALHOR_SM.PDF

179. Guo W, Keckesova Z, Donaher JL, et al. Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell*. 2012;148(5):1015-1028. doi:10.1016/j.cell.2012.02.008

180. Lim E, Wu D, Pal B, et al. Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res*. 2010;12(2):1-14. doi:10.1186/BCR2560/FIGURES/7

181. Murray AJ. Pharmacological PKA inhibition: All may not be what it seems. *Sci Signal*. 2008;1(22). doi:10.1126/SCISIGNAL.122RE4/ASSET/B7D4C32A-2054-463B-A9BF-493FBDFDFB71/ASSETS/GRAPHIC/122RE4F3.JPEG