

RESEARCH

Open Access



Lineage plasticity enables low-ER luminal tumors to evolve and gain basal-like traits

Gadisti Aisha Mohamed¹, Sundis Mahmood², Nevena B. Ognjenovic¹, Min Kyung Lee³, Owen M. Wilkins^{4,5}, Brock C. Christensen^{1,3,5}, Kristen E. Muller^{2*} and Diwakar R. Pattabiraman^{1,5*}

Abstract

Stratifying breast cancer into specific molecular or histologic subtypes aids in therapeutic decision-making and predicting outcomes; however, these subtypes may not be as distinct as previously thought. Patients with luminal-like, estrogen receptor (ER)-expressing tumors have better prognosis than patients with more aggressive, triple-negative or basal-like tumors. There is, however, a subset of luminal-like tumors that express lower levels of ER, 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 intratumoral lumino-basal heterogeneity. Single-cell RNA sequencing 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 result from the evolutionary progression of luminal tumors with low ER expression.

Keywords Breast cancer, MMTV-PyMT, Lineage plasticity, Low-ER, Lumino-basal heterogeneity, Single-cell RNA sequencing, SOX10

Introduction

Breast cancer is a complex disease with multiple different biologic subtypes which have clinical implications on tumor development, prognosis, and treatment [1–3]. 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+ [4]. 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

*Correspondence:

Kristen E. Muller
Kristen.E.Muller@hitchcock.org
Diwakar R. Pattabiraman
raman@dartmouth.edu

¹ Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA

² Department of Pathology, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756, USA

³ Department of Epidemiology, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA

⁴ Department of Biomedical Data Science, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA

⁵ Dartmouth Cancer Center, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756, USA



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

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+* [5]. A smaller signature of 50 genes, PAM50, may be used by clinicians to classify tumors based on these intrinsic subtypes [6].

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) [7]. These IHC-based subtypes correspond to the intrinsic subtypes luminal A, luminal B, *Erb-B2+*, and basal-like, respectively [5], providing biologic and clinically significant information used to guide treatment decisions [7, 8]. 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 [8], 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 collection of tumors being considered luminal-like, as tumors with less than 10% ER-positive cells may exhibit different characteristics 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 [9]. Some studies suggest the majority of breast cancers with low ER expression show molecular features similar to ER-negative, basal-like tumors rather than ER-positive, endocrine sensitive tumors [10]. It is 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) [11, 12]. 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 [13–16]. 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 [17], and basal cells repopulating mammary epithelium in response to injury or luminal cell ablation [18]. 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 [19]. Moreover, *BRCA1* and p53 deletions in the mouse luminal compartment results in tumors resembling typical human basal-like tumors [20]. In addition, claudin-low breast tumors, a mesenchymal subset of TNBCs, may also be derived from the luminal lineage [21]. 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 study, we show 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.

Results

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 [22]. 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 (Additional file 1: Fig. S1A) 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 Additional file 1: S1B-D). In contrast to high expressing ER tumors which typically also exhibit some degree of progesterone receptor (PR) expression, most

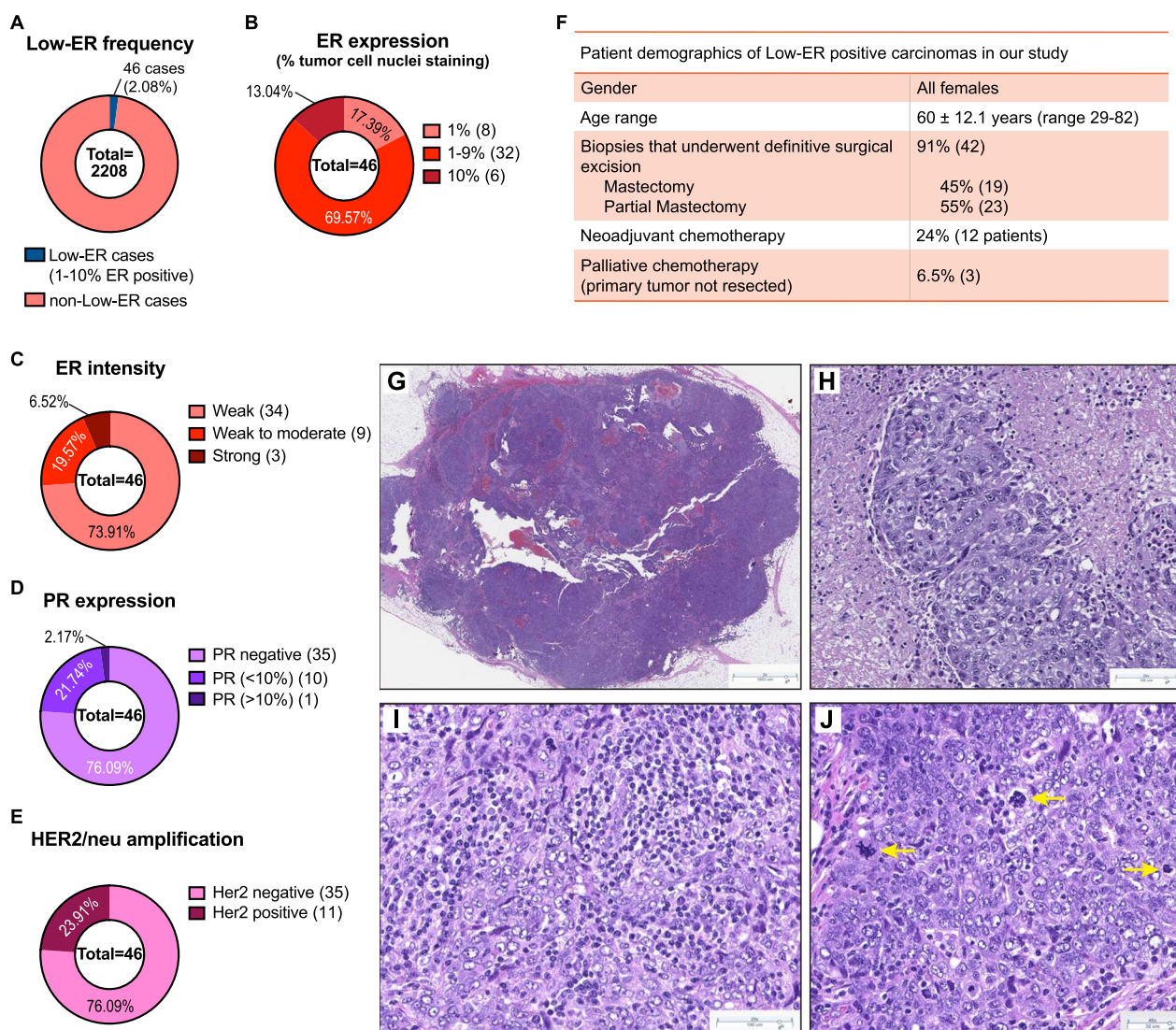


Fig. 1 Basal-like features of tumors expressing low ER. **A** Frequency of cases expressing low ER levels (< 1% ER-expressing nuclei) in newly diagnosed invasive breast carcinomas from the pathology database at Dartmouth-Hitchcock Medical Center (DHMC) from 2012–2020. **B–E** Breakdown of ER expression 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**)

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 (Additional file 1: Fig. S1E). 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 [23]. Twelve patients (24%) received neoadjuvant chemotherapy, and most (75%) achieved a pathologic complete response (Fig. 1F and Additional file 1: S1A). The majority of patients (81%) had no evidence of the disease at follow-up (Additional file 1: Fig. S1F); however, local or metastatic disease was detected in 6 patients. Biopsies were obtained from 4 of these patients, all of which tested negative for ER, PR (by IHC), and *HER2/neu* (by fluorescence in situ hybridization, FISH). While this may suggest that triple-negative tumors may evolve from tumors

that were initially identified as ER-expressing, analysis of a larger cohort of patients would be required to confirm these findings.

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 [10, 22, 24].

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.

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 (Additional file 1: Fig. S2A), which proved effective, with more than 91% of patients showing no evidence of disease at follow-up (Additional file 1: Fig. S2B). No patients received neoadjuvant chemotherapy (Fig. 2A). Most tumors in the luminal B group showed strong PR positivity (Additional file 1: Fig. S2C) and all tumors were negative for HER2/neu amplification (Additional file 1: Fig. S2D), 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. Instead, typical luminal B features include poorly differentiated tumors with irregular, infiltrative borders, scattered

tubule formation within a desmoplastic stroma, and tumor cells infiltrating as small solid nests (Fig. 2B, C).

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. 2D, E, and Additional file 1: S2E). 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 (Additional file 1: Fig. S2F), but 65% of the low-ER samples were Krt5⁺ (Fig. 2F, G). Staining for p63, another basal marker, also showed higher positivity in low-ER tumors compared to luminal B (Additional file 1: Fig. S2G). These data support histologic 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 [25, 26] 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 [27, 28]. All luminal B TMA cores strongly expressed both ER and Krt8, with no expression of either basal marker (Additional file 1: Fig. S2H and S2I). In contrast, low-ER TMA cores exhibited more heterogeneous ER, Krt5, and Krt14 expression (Fig. 2H and Additional file 1: S2J). 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,

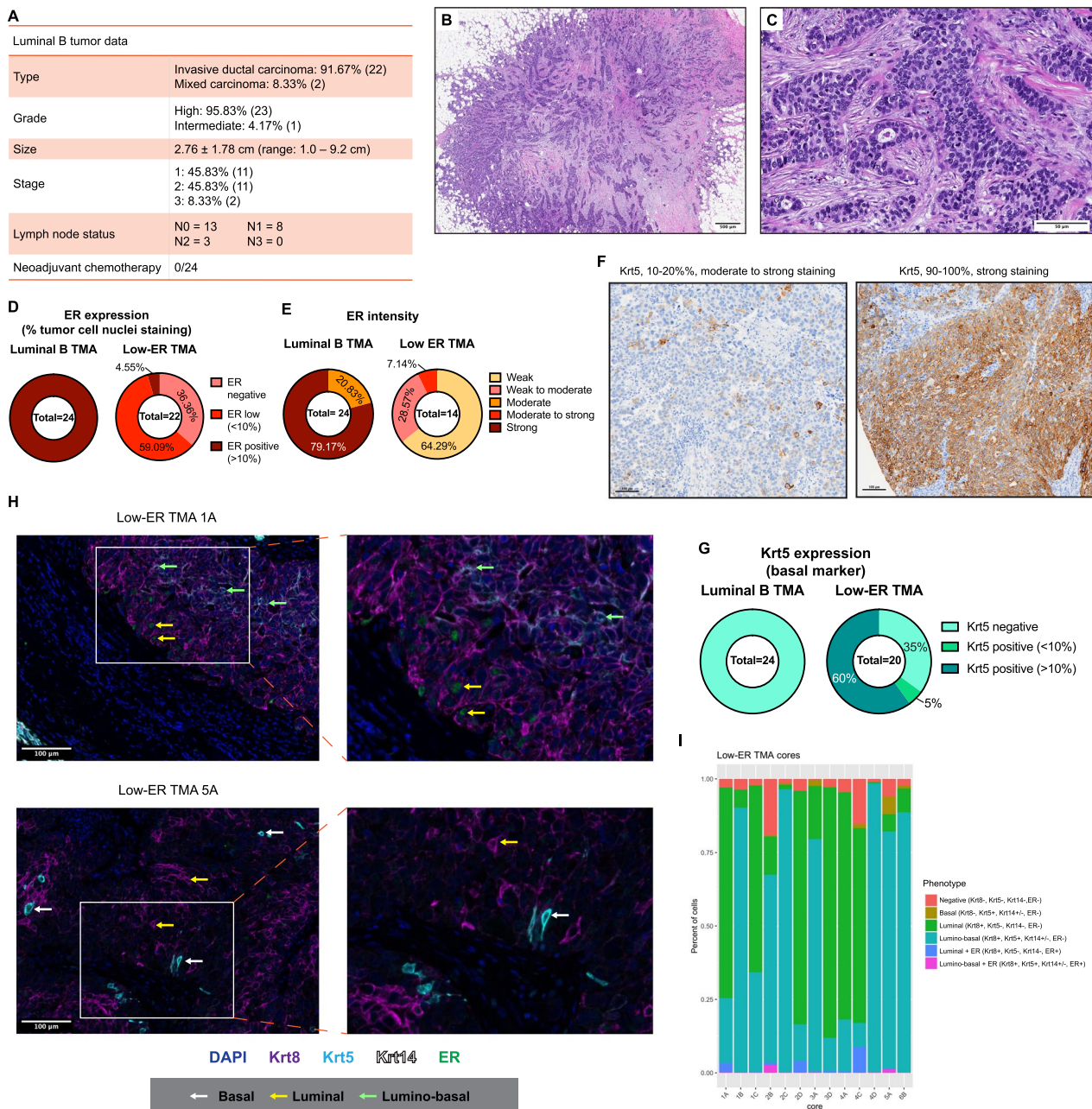


Fig. 2 Low-ER tumors are distinct from luminal B tumors. **A** Details of the luminal B tumors included in TMAs as a comparison to Low-ER tumors. **B, C** H&E staining showing typical features of a luminal B tumor (high-grade invasive ductal carcinoma ER+ (> 90%) PR+ (30%) HER2- high Ki67 tumor). **B** Lower power image showing a poorly differentiated carcinoma with irregular, infiltrative borders (1X H&E). **C** Tumor cells infiltrating as small solid nests with scattered tubule formation within a desmoplastic stroma. Note the lack of necrosis, solid growth, and a tumor associated lymphoplasmacytic infiltrate typically seen in breast carcinomas with basal-like characteristics (20X H&E). **D, E** ER expression (**D**) and intensity (**E**) levels of the luminal B tumors. **F** Representative images of Krt5 IHC staining in low-ER TMAs. **G** Quantified Krt5 expression in luminal B and low-ER TMAs. **H** 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. **I** 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)

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. 2H and Additional file 1: S2J). 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. 2H, I, and Additional file 1: S2J). 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. 2H and Additional file 1: S2J).

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. 2I), 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. All of the tumor cores contain a relatively small proportion of cells that stained negative for all the markers tested, which may represent stromal cells or cells that have undergone an epithelial–mesenchymal transition (EMT) and have lost cytokeratin expression.

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 biologic properties of this subtype.

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 ER α expression was quantified using Reverse Phase Protein Array (RPPA), which included both ER+/PR+, and ER+/PR– cases (Additional file 1: Fig. S3A). These cases were stratified into two groups; ER α low (141 cases, bottom quartile of ER α expression), and ER α high (423 cases, top 75% of ER α expression) (Fig. 3A) and analyzed their basal gene signature [29]. Unsupervised clustering of all basal signature genes revealed modules with higher relative expression in ER α low cases compared to ER α

high cases (Additional file 1: Fig. S3B). Supervised clustering of these cases based on ER α expression revealed a statistically significant (Fig. 3B) upregulation of basal signature gene expression in the ER α low cluster (Fig. 3C) irrespective of PR status (Additional file 1: Fig. S3C). Similar results were observed when cases were stratified using *ESR1* mRNA levels instead of ER α levels (Fig. 3D, E, Additional file 1: S3D-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. These differences in basal signature gene expression do not appear to influence the overall survival of patients with low ER expression (Additional file 1: Fig. S3G).

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 [30] is a mouse mammary tumor model that closely resembles human luminal B breast cancers [31] whereby late-stage tumors lose ER expression [32]. IHC staining for ER revealed weak to moderate expression in MMTV-PyMT tumors, with half of the tumors displaying less than 10% ER expression (Additional file 1: Fig. S4A and S4B).

In order to trace the lineage of the MMTV-PyMT tumor cells, either Krt8 (luminal) or Krt5 (basal) specific, tamoxifen-inducible CreERT promoters [13] were used to induce expression of GFP in an mTmG reporter mouse [33] to label luminal or basal cells, respectively (Fig. 4A, B). The most efficient mammary epithelial GFP labeling was observed when tamoxifen induction was performed 3 days per week on postnatal 3-week-old pups for the Krt5-CreERT/Rosa26-mTmG model (Additional file 1: Fig. S4C), and pups at postnatal weeks 5 and 6 of age for the Krt8-CreERT/Rosa26-mTmG model (Additional file 1: Fig. S4D). Tumors that eventually arose from these tamoxifen-pulsed mice were harvested and analyzed for GFP expression by immunostaining and flow cytometry.

Flow cytometry analysis (Additional file 1: Fig. S4E) of tamoxifen-pulsed Krt5-CreERT/Rosa26-mTmG/MMTV-PyMT tumors revealed they were comprised of mostly GFP-negative cells (Fig. 4C), indicating that MMTV-PyMT tumors did not originate from the

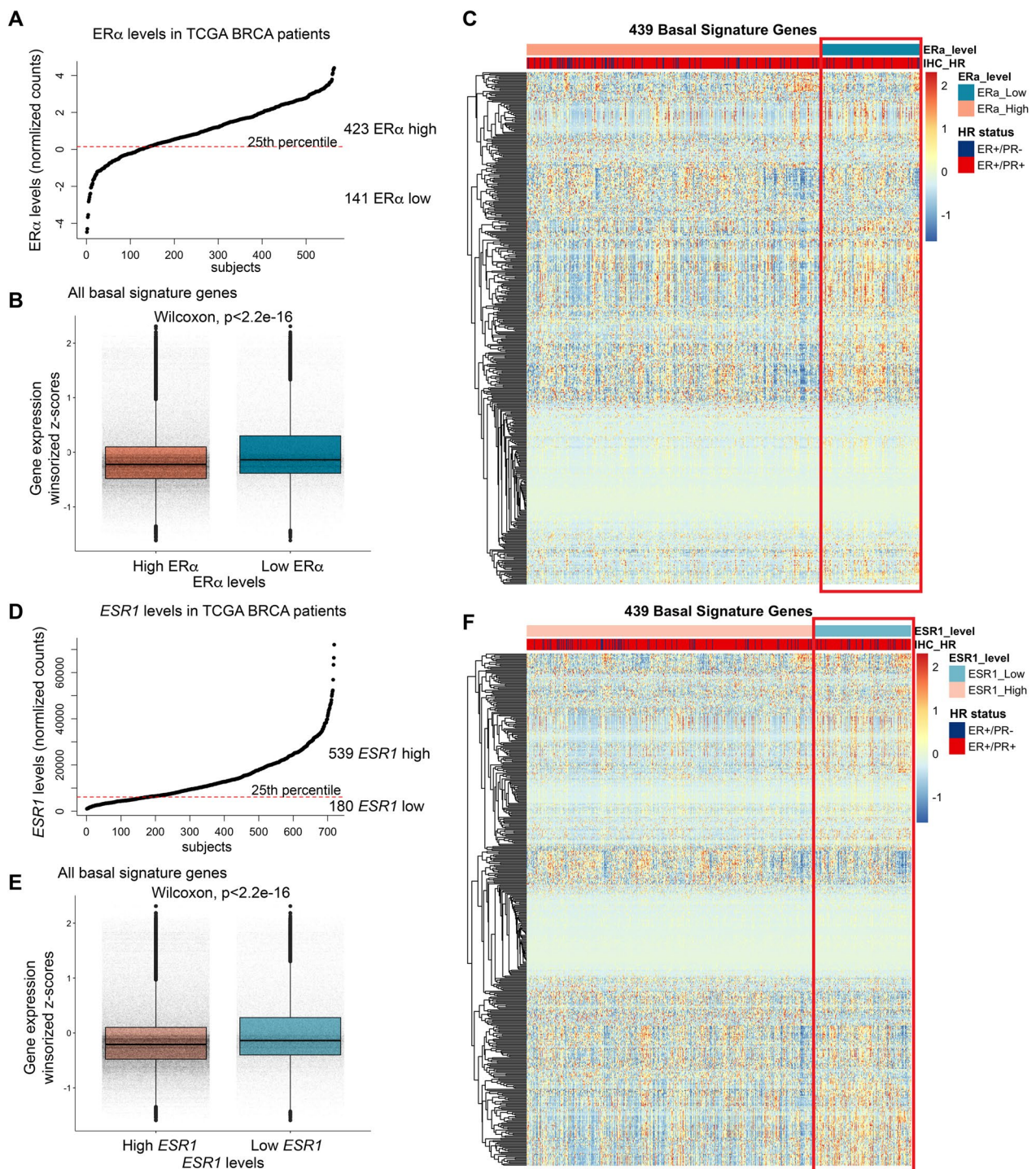


Fig. 3 Distinct basal gene expression signature in tumors with lower ER. **A** Stratification of ER-positive breast tumor cases from TCGA into ERα low and ERα high groups, based on ERα expression. **B** Boxplots comparing the distribution of basal signature gene expression in the ERα low and ERα high 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

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. 4C, D). The small proportion of GFP-positive luminal cells are likely descendants of Krt5-expressing luminal cells that may be found in the terminal duct lobular units (TDLU) [34]. 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. 4E). 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.

On the other hand, tamoxifen-pulsed Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT mice developed tumors that were primarily comprised of GFP-positive cells (Fig. 4F), 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. 4F, G). 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. 4H). 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 was also used to visualize and confirm 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. 4I), suggesting that cells of the basal lineage were confined to the adjacent normal regions of the tumor, and that the small proportion of GFP-positive basal cells found in this tumor (Fig. 4C) are likely long-lasting

myoepithelial cells present in adjacent normal tissues [35]. 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 (Additional file 1: Fig. S4F), 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. 4I and Additional file 1: S4G), reflecting their luminal origin, with most of these GFP-expressing cells co-expressing Krt8. No GFP expression was observed in basal lineage cells on the tumor periphery, confirming that the expression of GFP is restricted to luminal lineage cells (Additional file 1: Fig. S4G). Co-expression of GFP and Krt14 was also observed, with about 50% of basal marker-expressing cells co-expressing GFP (mean = 47.79%) (Additional file 1: Fig. S4H), 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. 4I), 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. 4J). 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.

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

(See figure on next page.)

Fig. 4 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 CreERT2 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

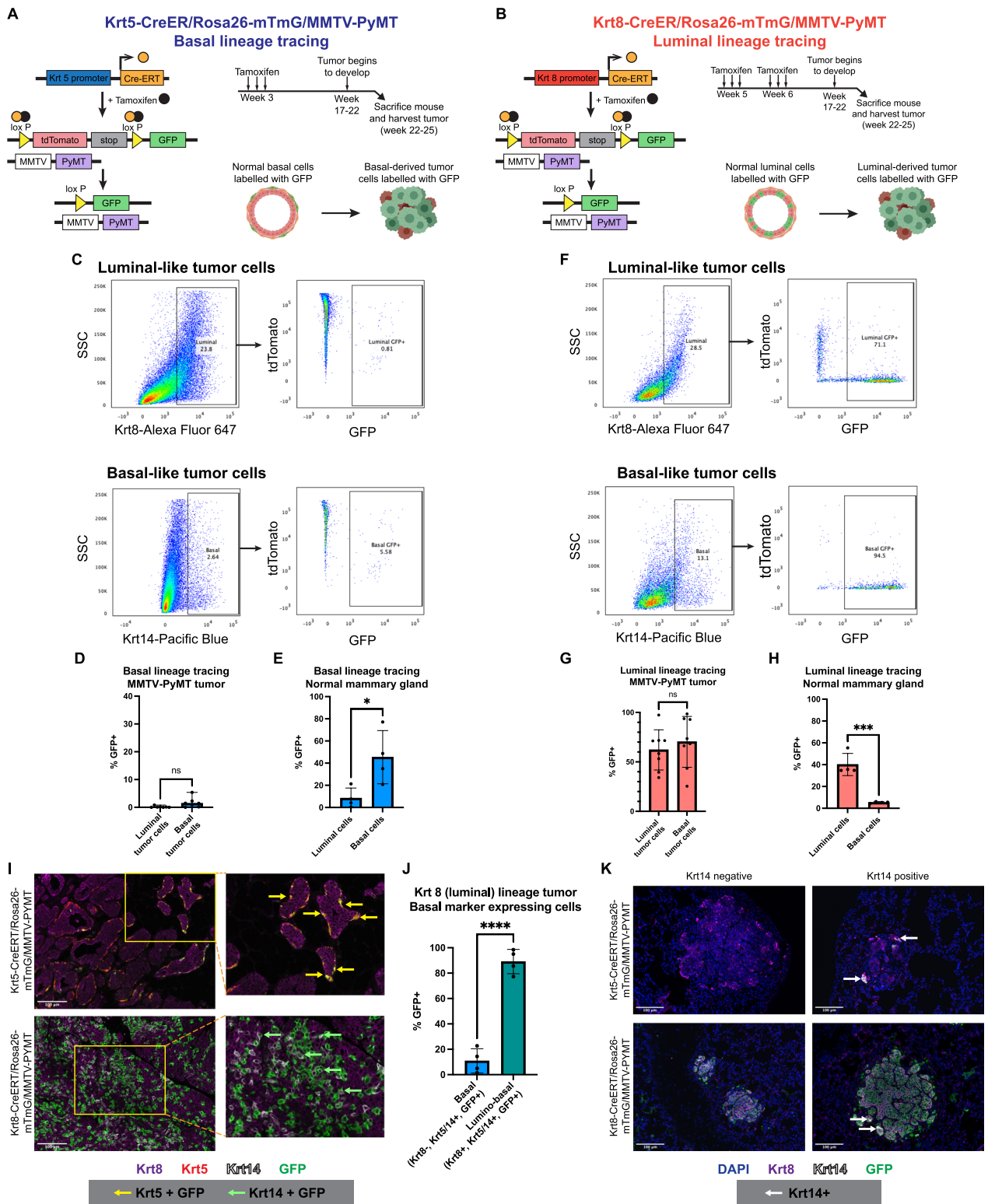


Fig. 4 (See legend on previous page.)

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. 4K and Additional file 1: S4J). In contrast, 16 out of 17 lung metastases from Krt8-CreERT/Rosa26-mTmG /MMTV-PyMT mice express GFP (Fig. 4K, Additional file 1: S4I and S4J), 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 (Additional file 1: Fig S4J), 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 (Fig. 4K and Additional file 1: S4I). 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 ($\sim 1.0\text{mm}^3$). 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

(Additional file 1: Fig. S4J). 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.

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 (Additional file 1: Fig. S5A) 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. 5A), with similar results observed in the two mouse replicates (Additional file 1: Fig. S5B). Using previously described gene signatures [36], 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 (Additional file 1: Fig. S5C), in agreement with previous single-cell RNA analyses of MMTV-PyMT tumors [37]; however, cells from multiple clusters demonstrated concomitant luminal progenitor and basal signatures (Additional file 1: Fig. S5D), 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. 5B, C). Cluster 13 demonstrated high scores specifically for the mature luminal signature, while clusters 15 and 16 demonstrated high scores specifically for the basal signature,

(See figure on next page.)

Fig. 5 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 ERa low and ERa 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

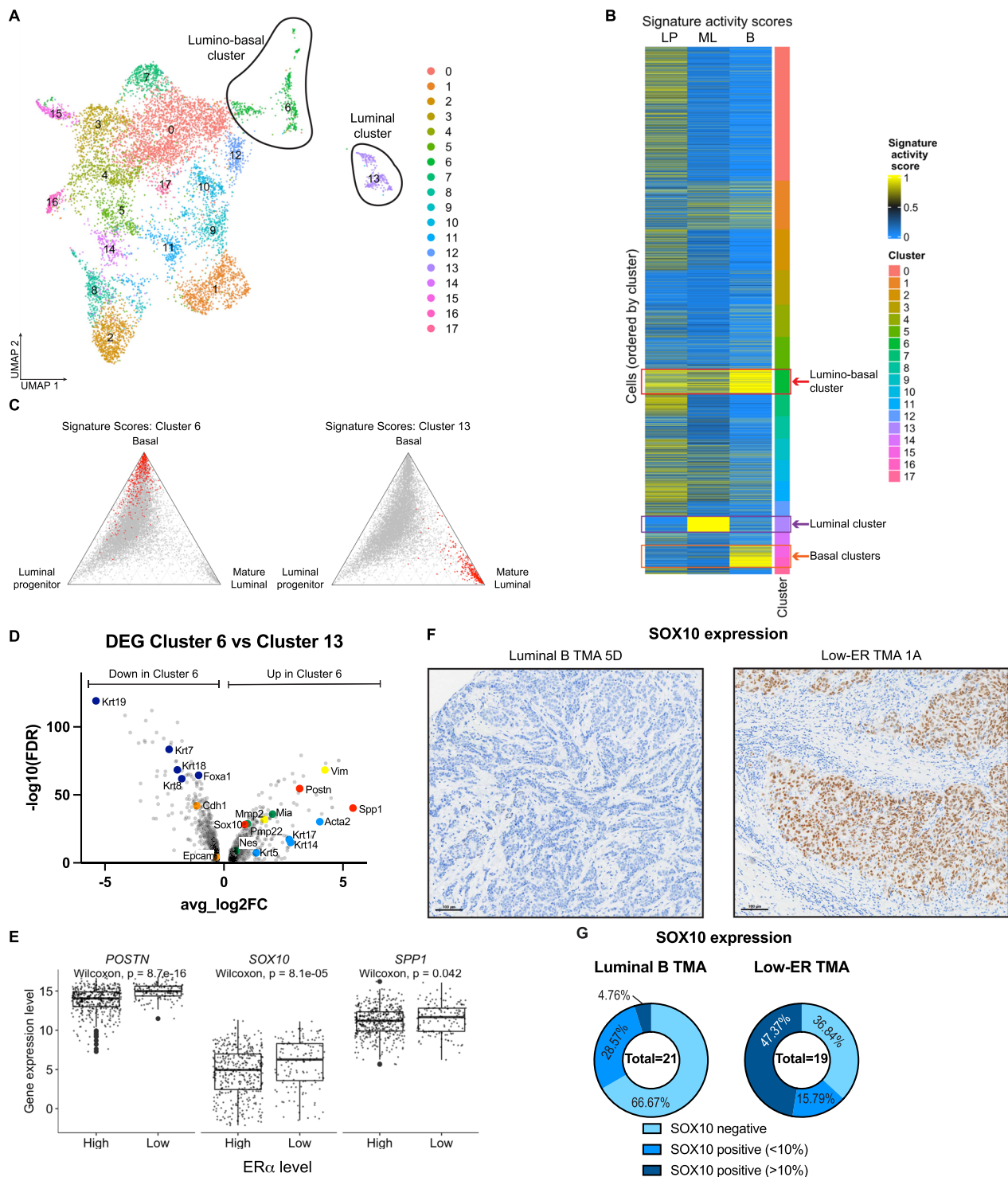


Fig. 5 (See legend on previous page.)

suggesting that these clusters contain luminal- and basal-like tumor cells, respectively (Fig. 5B, C, and Additional file 1: S5E).

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. 5D, light blue points) and simultaneous downregulation of luminal marker genes (*Krt7*, *Krt8*, *Krt18*, *Krt19*, and *Foxa1*) (Fig. 5D, dark blue points) in cluster 6, supporting the proposed lumino-basal identity of this cluster. Of note, mesenchymal genes *Mmp2*, and vimentin (*Vim*) (Fig. 5D, yellow points) were upregulated in cluster 6, while epithelial genes E-cadherin (*Cdh1*) and Epcam (Fig. 5D, 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 genes that may potentially regulate the luminal-to-basal plasticity observed in low-ER tumors (Fig. 5D, 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 [38] via its interaction with the stemness marker *Cd44* [39]. 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 $\alpha_v\beta_5$ [40]. Finally, we selected the transcription factor *Sox10*, as it has been implicated in the regulation of cell plasticity in mammary tumors [41]. Furthermore, three *Sox10* targets, *Nes*, *Mia*, and *Pmp22*, identified using TRRUST v2 (Transcriptional Regulatory Relationships Unraveled by Sentence-based Text mining) [42], were found to be upregulated in cluster 6 as well (Fig. 5D, 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 ER α low vs ER α high stratified tumors obtained from TCGA (Fig. 3). The expression of all 3 genes were found to be significantly higher in the ER α low group of tumors, as compared to the ER α high group (Fig. 5E). 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 tumor cells found expressing this protein (Additional file 1: Fig. S5F). 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 (Additional file 1: Fig. S5G); however, there was no difference in osteopontin expression between tumors in the

luminal B or the low-ER TMAs (Additional file 1: Fig. S5H). On the other hand, SOX10 appears to have a more direct correlation to low-ER tumors (Fig. 5F), 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. 5G), 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.

Finally, we sought to explore if the presence of a larger lumino-basal population might influence overall survival in patients. To this end, we again analyzed ER α low and ER α high tumors obtained from TCGA, stratified by *Krt5* (Additional file 1: Fig. S5I) and *SOX10* (Additional file 1: Fig. S5J) expression. In the ER α low group, high expression of both *Krt5* and *SOX10* correlate with a better prognosis, although this may be due to the smaller sample size of patients represented in this group (Additional file 1: Fig. S5I and S5J). In the ER α high group, high *Krt5* expression also appears to correlate with a better prognosis (Additional file 1: Fig. S5I), while *SOX10* expression on its own does not influence overall survival (Additional file 1: Fig. S5J). These results suggest that *Krt5* and *SOX10* may contribute to improved prognosis for patients with low and high ER expression; however, the influence of the lumino-basal population as a whole cannot be evaluated from this data, as the lumino-basal population is not defined by only these two genes. Identifying a lumino-basal gene signature, and whether high expression of this gene signature contributes to poor prognosis in breast cancer patients, may be more useful in elucidating the influence of the lumino-basal population on patient survival.

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. [43] found basal marker expression in luminal lineage-labeled cells, although they did not carry out further investigations on this observation. Hein et al. [44] 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 Erb-B2 oncogenes to both induce transformation and label the luminal lineage, whereas our lineage-tracing strategy uncoupled the process of oncogenesis from lineage labeling. 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 [37], while lineage labeling using keratin-driven *Cre* recombination, allows specific labeling 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. [45] and Van Keymeulen et al. [46] noted that PIK3CA mutations could induce oncogenesis and multipotency within mammary cells, resulting in heterogeneous, multilineage 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 transformation via interacting with, and activating c-Src [47], a non-receptor tyrosine kinase, thereby activating various other signaling molecules such as Shc [48] and PI3K [49]. 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 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.

This study has uncovered the emergence of a lumino-basal subpopulation within low-ER tumors, which aids in the understanding of lineage plasticity in the evolution of these tumors. Well-known luminal and basal gene signatures were used to identify this lumino-basal cell type in mouse tumors; however, a lumino-basal signature remains, as yet, undefined. The use of a small sample size of 2 mice was insufficient to provide enough data to define a murine lumino-basal gene signature. It would also be useful to identify a human lumino-basal signature that could aid in predicting patient outcomes and informing therapeutic decision-making. In order to define a robust lumino-basal signature, scRNA-seq would need to be performed on a larger sample set of human low-ER mammary tumors.

Our experiments on MMTV-PyMT mouse tumors have demonstrated that the emergence of the lumino-basal population in this model involves the expression of Krt14, but not Krt5. Krt14 is a myoepithelial/basal marker known to be expressed in several mammary stem cell or progenitor populations [50–52], whereas Krt5 appears to be more important for myoepithelial lineage commitment [53]. This suggests that the lumino-basal plasticity observed in these tumor cells involves a form of dedifferentiation of luminal-like cells into a more primitive cancer stem-like population. Furthermore, the emergence of basal-like characteristics in lumino-basal cells results from an incomplete luminal-to-basal transition, suggesting that a complete transition may not be necessary for tumor cells to gain the functional advantages endowed by a basal-like phenotype. An incomplete phenotypic transition is also commonly described in cells undergoing an epithelial–mesenchymal transition (EMT) [27, 54], which often assume a hybrid or intermediate state. Given that the lumino-basal population expresses higher levels of mesenchymal markers such as *Vim* and *Mmp2*, and the luminal population expresses higher levels of epithelial markers *Epcam* and *Cdh1* (Fig. 5D), 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, thus potentially allowing these cells to gain more aggressive and stem-like characteristics. The correlation between residence in a particular EMT state and exhibiting luminal vs. basal lineage traits is, as yet, poorly understood.

The basal-like characteristics observed in low-ER tumors suggest that these tumors may represent a distinct intrinsic subtype, separate from the luminal B or the basal-like subtypes described Perou et al. [4]. Further subclassification of breast cancer subtypes by Sorlie et al. [5], identified a separate luminal C subtype that appears to cluster closer to the basal-like subtype, suggesting that this subtype may be similar to the low-ER tumors

described in our study. Integrative clustering of genomic copy number and gene expression data performed by Curtis et al. [55] revealed 10 distinct breast cancer subgroups, one of which (IntClust4) appears to be similar to the low-ER tumors presented in our study. This cluster consists of a mix of ER-positive and ER-negative samples, including a proportion of ER-positive samples that were also defined as basal-like by PAM50, suggesting that the low-ER tumors exhibiting basal-like features identified in our study may align with this subgroup. Furthermore, significant lymphocytic infiltration was found in IntClust4, which was also observed in our low-ER tumor samples (Fig. 11), supporting the similarities between the low-ER tumors and IntClust4.

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 [37]. The collective migration of breast cancer has also been shown to involve leader cells with a reactivated basal program [56], 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 [54], suggesting that maintaining cellular plasticity is beneficial in metastasis. Circulating tumor clusters have also been found to consist of multiple clones [57], 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-labeled 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 which is commonly used in the study of breast carcinomas [32], 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. This 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 [58] 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 labeling to occur simultaneously. Additionally, lineage-specific labeling by *Cre* induction was performed at a relatively early stage of mouse development, which could potentially also result in the labeling of epithelial progenitor populations along with differentiated luminal and basal cells. It is thus possible, and likely, that the luminal-derived basal-like tumor cells in the *Krt8-CreERT/Rosa26-mTmG/MMTV-PyMT* mouse may have descended from a luminal progenitor ancestor instead of a differentiated luminal cell. To address this, lineage labeling of luminal progenitor cells could be carried out an *ELF5*-reporter that would specifically label this subpopulation [59]. This would allow the identification of tumor cells descending specifically from the luminal progenitor population, and determine if the lumino-basal population is derived from the luminal progenitor instead of differentiated luminal cells. 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.

The emergence of a lumino-basal population specifically in low-ER tumors suggest that the expression of basal-specific genes may be triggered by loss of ER function; however, attempts to visualize the co-expression of ER and basal cytokeratins in mouse tissue failed to yield robust results. The absence of anti-ER antibodies that are able to stain mouse tissues reliably precluded the TSA co-staining of ER with other keratin markers. Individual IHC staining for ER was suboptimal, with visible background and non-nuclear staining (Additional file 1: Fig. S4B), thus attempts to analyze co-staining of ER along with the luminal and basal cytokeratins proved to be challenging due to the added complexity of identifying co-staining across 4 markers. It is also possible that treatment of low-ER tumors with hormone therapies targeting ER may accelerate the expression of basal genes, thus enriching the lumino-basal population. Treatment

of MMTV-PyMT tumors with anti-ER therapies such as tamoxifen has been shown to delay tumor development, without completely inhibiting tumor progression [60]; however, the effect of this treatment on basal gene expression is unknown. It is thus crucial to first elucidate if basal gene expression requires ER downregulation, and second, the mechanisms by which this occurs, in order to understand the effects of ER loss on the emergence of basal genes and characteristics.

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 [41]. 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 [61, 62]. 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 [63], 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 [41], 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.

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 [9]. 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 intratumoral 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 h using the polymer system technique with appropriate controls. The assays were performed according to the manufacturer's instructions using Anti-ER (Cell Marque, 249-R-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 *HER2* DNA probe kit) to assess for gene amplification and results were interpreted in accordance with the ASCO/CAP *HER2* testing guidelines [64].

Animal studies

All animal experiment IACUC protocols were approved by the Dartmouth College Committee on Animal Care. MMTV-PyMT mice ((Tg(MMTV-PyV)T)^{634Mul/Lell}) mice on a C57Bl/6 J background, strain #: 022974) [30], *Krt5-CreER* mice (B6N.129S6(Cg)-*Krt5*^{tm1.1(Cre/ERT2)Blh}/J, strain #: 029155) [13], *Krt8-CreER* mice (Tg(*Krt8-Cre/ERT2*)^{17Blpn}/J, strain #: 017947) [13], and *Rosa26-mTmG* reporter mice (B6.129(Cg)-Gt(*ROSA*)26Sor^{tm4(ACTB-tdTomato,-EGFP)Luo}/J, strain #: 007676) [33] were purchased from The Jackson Laboratory. For tamoxifen induced mammary epithelial labeling, tamoxifen (Sigma-Aldrich, T5648-1G) was prepared by dissolving in commercially available corn oil for 5 h at 37 °C to a final concentration of 30 mg/ml. *Krt5-CreERT/Rosa26-mTmG* /MMTV-PyMT mice were administered 150 mg/kg (100 μ l of

tamoxifen stock for a 20 g mouse) of the tamoxifen stock 3 times per week at week 3, while Krt8-CreERT/Rosa26-mTmG /MMTV-PyMT mice were administered 150 mg/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 1.5cm³, 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 [65] 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, 10,103,586,001) for 2 h at 37 °C with gentle agitation using a rotator. Red blood cells were subsequently removed with an ammonium chloride lyse (8.02 g NH₄Cl, 0.84 g NaHCO₃, 0.37 g 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, DN25-100 mg) for 1 min each to further dissociate the cells. Finally, samples were filtered through a 40 mm cell strainer (Corning, 431,750) 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 min. Rehydrated tissue was then stained with Carmine Alum (Stem Cell Technologies, 07,070) for 48–72 h. Mammary glands were then dehydrated using increasing ethanol concentrations (70%, 95%, 100%) 2 times each for 15 min, 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 h with gentle agitation using a rotator. Following digestion, samples were strained through 70 mm (Corning, 431751) and 40 mm 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 labeled antibodies. Tumor single-cell suspensions were stained with Alexa Fluor 700 anti-CD326 (EPCAM) antibody (Biolegend, 118239, clone: G8.8, 1:100 dilution), PE/Cyanine 7 anti-mouse CD31 antibody (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 min 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 min at RT in 2% paraformaldehyde (methanol free, Thermo Scientific, J19943-K2), and permeabilized for 15 min 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 min 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 4 mm and air dried at RT before baking at 60 °C for 30 min. 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 min (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 min 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 h to remove paraffin wax, followed by 3 xylene washes of 10 min 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 min at RT. Antigen retrieval was performed in BOND Epitope Retrieval Solution 1 (Leica, AR9961) for 20 min 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 min. Slides were washed in TBS + 0.1% tween (TBS-T) and blocked in Antibody Diluent/Block (Akoya Biosciences, ARD1001EA) for 30 min at RT. Primary antibody were added and slides were incubated at RT for 30 min. After TBS-T washes, secondary antibodies were added and incubated at RT for another 30 min, followed by TBS-T washes. Opal fluorophore was applied to slides for precisely 6 min, followed by TBS-T washes. Slides were then boiled for 2 min in a microwave at 100% power, followed by 15 min 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 4 × resolution using the PerkinElmer Vectra3 slide scanner, and Regions of interest (ROIs) were selected on PhenoChart. ROIs were then imaged at 20 × 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 ER α -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 ER α levels were determined based on the distribution of ER α from RPPA assays. Subjects were defined as ER α low ($n = 141$) if the level of ER α were

below the 25th percentile and as ER α high ($n=423$) if the level of ER α were above the 25th percentile.

For *ESRI*-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 *ESRI* levels were determined based on the distribution of *ESRI* from RNA-seq. Subjects were defined as *ESRI* low ($n=180$) if the level of *ESRI* were below the 25th percentile and as *ESRI* high ($n=539$) if the level of *ESRI* were above the 25th percentile.

Gene expression levels of 439 basal signature genes (obtained from previously published data [29]) were available in the TCGA data. Winsorized Z-scores for each gene were used to compare the differences in expression between ER α /*ESRI* high and low subjects. Log₂ transformed counts were used to compare the differences for *POSTN*, *SPPI1*, 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 (10 \times 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 (10 \times 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 28 bp, Read2 90 bp 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) [66]. 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) [67] 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* [68], 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

[69]) 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* [70] 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 [71] 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. [36]) were scored for enrichment at the individual cell-level using variance-adjusted Mahalanobis (VAM) [72]. 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 [73] and are accessible through GEO Series accession number GSE214815 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE214815>).

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13058-023-01621-8>.

Additional file 1. Supplementary figures 1–5.

Acknowledgements

Part of the illustrations in Fig. 4 were created with BioRender.com. We thank Gary Ward, Jennifer Fields, Scott M. Palisoul, Dr Meredith Brown, Dr Pat Robison and Dr Fred W. Kolling IV, for sharing their technical expertise, and Dartmouth

Center for Comparative Medicine and Research (CCMR) for animal husbandry services. We acknowledge the following Shared Resources facilities: Dartlab (flow cytometry), Pathology shared resources, Microscopy shared resources, Genomics and Molecular Biology Shared Resource, and Dartmouth Cancer Center shared equipment, at the Dartmouth Cancer Center with the NCI Cancer Center Support Grant 5P30 CA023108-43. Single-cell studies were conducted through the Dartmouth Center for Quantitative Biology (COBRE, 5P20GM130454-03), in collaboration with the GMBSR with support from NIGMS (P20GM130454) and NIH S10 (S10OD025235) awards. This study was supported by awards 5R00CA201574-05 and 1R01CA267691 from the NIH to D.R.P.

Author contributions

This project was conceived and designed by G.A.M and D.R.P. Mouse experiments were carried out by G.A.M and N.B.O. Pathology database search and compilation of human patient data was performed by S.M and K.E.M. Analysis of TCGA data was performed by M.K.L and B.C.C. G.A.M carried out all other experiments. Single-cell RNA-seq data were analyzed by O.M.W with input from G.A.M. This manuscript was prepared by G.A.M and D.R.P, with contributions from all co-authors. All authors read and approved the final manuscript.

Funding

The use of shared resources at the Dartmouth Cancer Center were supported by the NCI Cancer Center Support Grant 5P30 CA023108-43. Single-cell studies were conducted through the Dartmouth Center for Quantitative Biology (COBRE, 5P20GM130454-03), in collaboration with the GMBSR with support from NIGMS (P20GM130454) and NIH S10 (S10OD025235) awards. This study was supported by awards 5R00CA201574-05 and 1R01CA267691 from the NIH to D.R.P.

Availability of data and materials

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

Declarations

Ethics approval and consent to participate

Research involving animals was carried out in accordance with The Institutional Animal Care and Use Committee (IACUC) and the Dartmouth College Committee on Animal Care (IACUC approval number: 2119).

Competing interests

The authors declare no competing interests.

Received: 12 October 2022 Accepted: 15 February 2023

Published online: 01 March 2023

References

- 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):1970–9. <https://doi.org/10.1158/1078-0432.CCR-18-2782/AM/IMPACT-OF-BREAST-CANCER-SUBTYPES-ON-PROGNOSIS-OF>.
- 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. <https://doi.org/10.1186/S12885-016-2766-3/FIGURES/2>.
- 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):E734–9. <https://doi.org/10.9778/CMAJO.20170030>.
- Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nat*. 2000;406(6797):747–752. <https://doi.org/10.1038/35021093>.
- Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*. 2001;98(19):10869–74. <https://doi.org/10.1073/pnas.191367098>.
- Bernard PS, Parker JS, Mullins M, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*. 2009;27(8):1160–7. <https://doi.org/10.1200/JCO.2008.18.1370>.
- Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA*. 2006;295(21):2492–502. <https://doi.org/10.1001/JAMA.295.21.2492>.
- 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):2784–95. <https://doi.org/10.1200/JCO.2009.25.6529>.
- 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):545–63. <https://doi.org/10.5858/ARPA.2019-0904-SA>.
- 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):729–34. <https://doi.org/10.1200/JCO.2011.36.2574>.
- 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. <https://doi.org/10.1354/vp.43-1-36>.
- Sun P, Yuan Y, Li B, Dai X. Cytokeratin expression during mouse embryonic and early postnatal mammary gland development. *Histochem Cell Biol*. 2010;133(2):213–21. <https://doi.org/10.1007/s00418-009-0662-5>.
- Van Keymeulen A, Rocha AS, Ousset M, et al. Distinct stem cells contribute to mammary gland development and maintenance. *Nature*. 2011;479(7372):189–93. <https://doi.org/10.1038/nature10573>.
- 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):164–76. <https://doi.org/10.1038/NCB3471>.
- Wuidart A, Sifrim A, Fioramonti M, et al. Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat Cell Biol*. 2018;20(6):666–76. <https://doi.org/10.1038/S41556-018-0095-2>.
- 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):677–87. <https://doi.org/10.1038/S41556-018-0108-1>.
- Song W, Wang R, Jiang W, et al. Hormones induce the formation of luminal-derived basal cells in the mammary gland. *Cell Res* 2019;29(3):206–220. <https://doi.org/10.1038/s41422-018-0137-0>.
- Centonze A, Lin S, Tika E, et al. Heterotypic cell–cell communication regulates glandular stem cell multipotency. *Nature*. 2020;584(7822):608–13. <https://doi.org/10.1038/s41586-020-2632-y>.
- 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;15(8):907–913. <https://doi.org/10.1038/nm.2000>.
- 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):403–17. <https://doi.org/10.1016/j.stem.2010.07.010>.
- 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. <https://doi.org/10.1038/s41467-021-23957-5>.
- Gloyeske NC, Dabbs DJ, Bhargava R. Low ER+ breast cancer: Is this a distinct group? *Am J Clin Pathol*. 2014;141(5):697–701. <https://doi.org/10.1309/AJCP34CYSATWFDPO>.
- Landmann A, Farrugia DJ, Zhu L, et al. Low estrogen receptor (ER)-positive breast cancer and neoadjuvant systemic chemotherapy response similar to typical ER-positive or ER-negative disease? *Am J Clin Pathol*. 2018;150(1):34–42. <https://doi.org/10.1093/AJCP/AQY028>.
- 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. <https://doi.org/10.1245/S10434-012-2588-8>.
- 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. <https://doi.org/10.1002/JCB.28016>.

26. Lazarus J, Akiska Y, Lanfranca MP, et al. Optimization, design and avoiding pitfalls in manual multiplex fluorescent immunohistochemistry. *J Vis Exp.* 2019;2019(149). <https://doi.org/10.3791/59915>
27. 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. <https://doi.org/10.1126/SCIADV.ABJ8002>.
28. 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;171:149–61. <https://doi.org/10.1016/BS.MCB.2022.06.003>.
29. 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. <https://doi.org/10.15252/emboj.2020107333>
30. 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):954–61. <https://doi.org/10.1128/MCB.12.3.954-961.1992>.
31. 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. <https://doi.org/10.1186/gb-2013-14-11-r125>.
32. 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. [https://doi.org/10.1016/S0002-9440\(10\)63568-7](https://doi.org/10.1016/S0002-9440(10)63568-7).
33. Muzumdar MD, Tasic B, Miyamichi K, Li N, Luo L. A global double-fluorescent Cre reporter mouse. *Genesis.* 2007;45(9):593–605. <https://doi.org/10.1002/DVG.20335>
34. Gusterson BA, Ross DT, Heath VJ, Stein T. Basal cytokeratins and their relationship to the cellular origin and functional classification of breast cancer. *Breast Cancer Res.* 2005;7(4):143–8. <https://doi.org/10.1186/bcr1041>.
35. Zeps N, Bentel JM, Papadimitriou JM, Dawkins HJS. Murine progesterone receptor expression in proliferating mammary epithelial cells during normal pubertal development and adult estrous cycle: Association with ER α and ER β status. *J Histochem Cytochem.* 1999;47(10):1323–30. https://doi.org/10.1177/002215549904701012/ASSET/IMAGES/LARGE/10.1177_002215549904701012-FIG2.JPEG.
36. 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. <https://doi.org/10.1186/S13058-021-01445-4/FIGURES/7>.
37. 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. <https://doi.org/10.1016/j.devcel.2020.10.004>.
38. Jia R, Liang Y, Chen R, et al. Osteopontin facilitates tumor metastasis by regulating epithelial–mesenchymal plasticity. *Cell Death Dis* 2016;7(12):e2564–e2564. <https://doi.org/10.1038/cddis.2016.422>
39. 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;184:118–30. [https://doi.org/10.1002/\(SICI\)1097-4652\(200007\)184:1](https://doi.org/10.1002/(SICI)1097-4652(200007)184:1).
40. 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.
41. 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. <https://doi.org/10.1016/j.ccell.2018.08.001>.
42. 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):D380–6. <https://doi.org/10.1093/NAR/GKX1013>.
43. 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. <https://doi.org/10.1016/J.DEVCEL.2020.06.021>.
44. Hein SM, Haricharan S, Johnston AN, et al. Luminal epithelial cells within the mammary gland can produce basal cells upon oncogenic stress. *Oncogene* 2015;35(11):1461–1467. <https://doi.org/10.1038/onc.2015.206>
45. Koren S, Reavie L, Couto JP, et al. PIK3CAH1047R induces multipotency and multi-lineage mammary tumours. *Nat* 2015;525(7567):114–118. <https://doi.org/10.1038/nature14669>
46. Van Keymeulen A, Lee MY, Ousset M, et al. Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature* 2015;525(7567):119–123. <https://doi.org/10.1038/nature14665>
47. 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. <https://doi.org/10.1101/GAD.8.1.23>.
48. 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):6344–8. <https://doi.org/10.1073/PNAS.91.14.6344>.
49. Whitman M, Kaplan DR, Schaffhausen B, Cantley L, Roberts TM. Association of phosphatidylinositol kinase activity with polyoma middle-T competent for transformation. *Nat* 1985;315(6016):239–242. <https://doi.org/10.1038/315239a0>
50. Sleeman KE, Kendrick H, Ashworth A, Isacke CM, Smalley MJ. CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells. *Breast Cancer Res.* 2006;8(1):R7. <https://doi.org/10.1186/BCR1371>.
51. Shackleton M, Vaillant F, Simpson KJ, et al. Generation of a functional mammary gland from a single stem cell. *Nature* 2006;439(7072):84–88. <https://doi.org/10.1038/nature04372>
52. Asselin-Labat ML, Shackleton M, Stingl J, et al. Steroid hormone receptor status of mouse mammary stem cells. *JNCI J Natl Cancer Inst.* 2006;98(14):1011–4. <https://doi.org/10.1093/JNCI/DJJ267>.
53. Deckwirth V, Rajakylä EK, Cattavarayane S, et al. Cytokeratin 5 determines maturation of the mammary myoepithelium. *iScience.* 2021;24(5). <https://doi.org/10.1016/j.isci.2021.102413>
54. 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. <https://doi.org/10.1016/j.devcel.2021.11.006>.
55. Curtis C, Shah SP, Chin SF, et al. The genomic and transcriptomic architecture of 2000 breast tumours reveals novel subgroups. *Nature* 2012;486(7403):346–352. <https://doi.org/10.1038/nature10983>
56. Cheung KJ, Gabrielson E, Werb Z, Ewald AJ. Collective invasion in breast cancer requires a conserved basal epithelial program. *Cell.* 2013;155(7):1639–51. <https://doi.org/10.1016/J.CELL.2013.11.029/ATTACHMENT/2A1FD816-3063-4634-A607-E4C61311C346/MMC8.PDF>.
57. Tiede S, Kalathur RKR, Lüönd F, et al. Multi-color clonal tracking reveals intra-stage proliferative heterogeneity during mammary tumor progression. *Oncogene* 2020;40(1):12–27. <https://doi.org/10.1038/s41388-020-01508-4>
58. 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):389–98. <https://doi.org/10.1002/PATH.2124>.
59. Rios AC, Fu NY, Lindeman GJ, Visvader JE. In situ identification of bipotent stem cells in the mammary gland. *Nature* 2014;506(7488):322–327. <https://doi.org/10.1038/nature12948>
60. Butt SA, Søgaard LV, Ardenkjaer-Larsen JH, et al. Monitoring mammary tumor progression and effect of tamoxifen treatment in MMTV-PyMT using MRI and magnetic resonance spectroscopy with hyperpolarized [1-13C]pyruvate. *Magn Reson Med.* 2015;73(1):51–8. <https://doi.org/10.1002/MRM.25095>.
61. 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):959–65. <https://doi.org/10.1016/J.HUMPATH.2012.09.005>.
62. Rammal R, Goel K, Elishaev E, et al. The utility of SOX10 immunohistochemical staining in breast pathology: staining 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.* 2022. <https://doi.org/10.1093/AJCP/AQAC092>.
63. 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;8(1):1–16. <https://doi.org/10.1038/s41523-022-00425-x>
64. 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. <https://doi.org/10.1200/JCO.2018.77.8738>
65. 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. https://doi.org/10.1007/978-1-62703-128-8_25/FIGURES/00255.
 66. Zheng GXY, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017;8(1):1–12. <https://doi.org/10.1038/ncomms14049>
 67. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;33(5):495–502. <https://doi.org/10.1038/nbt.3192>
 68. Germain P-L, Lun A, Macnair W, et al. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Research* 2021;10:979. <https://doi.org/10.12688/f1000research.73600.1>
 69. 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):1179–86. <https://doi.org/10.1093/BIOINFORMATICS/BTW777>.
 70. 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. <https://doi.org/10.1186/S13059-019-1874-1/FIGURES/6>.
 71. Zappia L, Oshlack A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. *Gigascience*. 2018;7(7):1–9. <https://doi.org/10.1093/GIGASCIENCE/GIY083>.
 72. 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. <https://doi.org/10.1093/NAR/GKAA582>.
 73. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res*. 2002;30(1):207–10. <https://doi.org/10.1093/NAR/30.1.207>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

