A genetic mosaic mouse model illuminates the pre-malignant progression of basal-like breast cancer

Jianhao Zeng¹, Shambhavi Singh², Xian Zhou¹, Ying Jiang¹, Eli Casarez¹, Kristen A. Atkins³,⁴, Kevin A. Janes²,⁴, Hui Zong¹,⁴,*

¹Department of Microbiology, Immunology, and Cancer Biology, University of Virginia Health System, Charlottesville, VA 22908, USA
²Department of Biomedical Engineering, University of Virginia, Charlottesville, VA 22908, USA
³Department of Pathology, University of Virginia Health System, Charlottesville, VA 22908, USA
⁴University of Virginia Comprehensive Cancer Center, University of Virginia Health System, Charlottesville, VA 22903, USA

*Correspondence author: hz9s@virginia.edu    ORCiD: 0000-0002-4263-9633

Summary statement
A mouse model that reveals the process of basal-like breast cancer initiated from sporadic Brca1, p53-deficient cells enables spatially resolved analysis of the pre-malignant progression of mutant cells.

Keywords: Mouse genetic mosaic model, basal-like breast cancer, BRCA1, pre-malignancy, spatiotemporal analysis of tumor initiation

Abstract
Basal-like breast cancer (BLBC) is highly aggressive, often characterized by BRCA1 and p53 deficiency. Although conventional mouse models enabled the investigation of BLBC at malignant stages, its initiation and pre-malignant progression remain under-studied. Here, we leveraged a mouse genetic system known as Mosaic Analysis with Double Markers (MADM) to study BLBC initiation by generating rare GFP+ Brca1, p53-deficient mammary cells alongside RFP+ wildtype sibling cells. After confirming the close resemblance of mammary tumors arising in this model to human BLBC at both transcriptomic and genomic levels, we focused our studies on the pre-malignant progression of BLBC.
Initiated GFP+ mutant cells showed a stepwise pre-malignant progression trajectory from focal expansion to hyper-alveolarization and then to micro-invasion. Furthermore, despite morphological similarities to alveoli, hyper-alveolarized structures actually originate from ductal cells based on twin-spot analysis of GFP-RFP sibling cells. Finally, luminal-to-basal transition occurred exclusively in cells that have progressed to micro-invasive lesions. Taken together, our MADM model provides excellent spatiotemporal resolution to illuminate the pre-malignant progression of BLBC, and should enable future studies on early detection and cancer prevention for this devastating cancer.

Introduction

Breast cancer is the most frequently diagnosed cancer type and the second leading cause of cancer death in women (Siegel et al., 2022). Human breast cancer is a heterogeneous disease classified into six molecular subtypes with distinct prognosis: luminal A, luminal B, HER2-enriched, normal-like, claudin-low, and basal-like (Perou et al., 2000; Prat et al., 2010; Tobin et al., 2015). Basal-like breast cancer (BLBC) accounts for 15–20% of breast cancer cases and is the most aggressive subtype with earlier onset, increased chance of metastasis, and absence of hormonal-therapy targets (Fulford et al., 2007; Millikan et al., 2008; Tobin et al., 2015; Turner et al., 2004). BLBCs show a high prevalence of p53 mutations (~80%) and deficiency in homology-directed DNA repair (~50%); the latter is often caused by germline mutations in BRCA1/2, somatic epigenetic inactivation of BRCA1/2, or the loss of other essential genes for homology-directed DNA repair pathway (collectively termed as “BRCAness”) (Lord and Ashworth, 2016; McCabe et al., 2006; Network, 2012; Tian et al., 2019). Early detection and prevention of BLBC can fundamentally improve patient care, particularly for individuals with germline BRCA1 mutations who are at a high risk of BLBC. Therefore, it is imperative to gain a deep understanding of how BLBC initiates and progresses during the pre-malignant stages.

Genetically engineered mouse models (GEMMs) present an invaluable pre-clinical platform for studying human cancers. Conditional knockout of Brca1 and p53 in mouse mammary epithelial cells led to mammary tumors resembling human BLBC (Hollern et al., 2019; Liu et al., 2007; Molyneux et al., 2010; Xu et al., 1999). These models are useful for studying BLBC at the malignant stage; but are quite limited in examining cancer initiation and pre-malignant progression. First, conditional knockout models generate numerous rather than rare mutant cells at the cancer initiation stage; thus, they do not accurately mimic human cancer initiation from sporadic mutant cells (Liu et al., 2011; Muzumdar et al., 2007), which may impact pre-malignant development. Second, even if rare mutant cells can be generated, unequivocally pinpointing subtle aberrant behaviors of BRCA1 mutant cells at the pre-malignant stage remains challenging.

To overcome these limitations, we used a mouse genetic system known as Mosaic Analysis with Double Markers (MADM) developed by our lab (Fig. 1A). MADM consists of a pair of chimeric GFP and RFP coding sequences (separated by a loxP-containing intron) knocked into homologous...
chromosomes. Each knock-in cassette is syntenic with either the wildtype or mutant allele of one or more tumor suppressor genes on the same chromosome. From a non-labeled heterozygous animal, Cre/loxP-mediated inter-chromosomal mitotic recombination followed by X-segregation of chromosomes generates a homozygous mutant cell labeled with GFP and its sibling wildtype cell labeled with RFP. Mutant cells are rare (0.1%–1% or even lower) due to the low frequency of inter-chromosomal recombination (Muzumdar et al., 2007; Zong et al., 2005), thereby approximating sporadic cancer initiation. The permanent GFP labeling of rare mutant cells enables spatially resolved investigation of their aberrant behavior at any time point during tumorigenesis (Fig. 1A) (Liu et al., 2011; Yao et al., 2020; Zong et al., 2005). Furthermore, along with a GFP+ mutant cell, MADM simultaneously generates a sibling RFP+ wildtype cell, which serves as a perfect internal reference that enables the detection of subtle abnormalities of mutant cells. Here, we applied MADM to establish a mouse genetic mosaic model for BLBC, in which cancer initiates from sparse Brca1, p53-deficient cells in mammary glands. Our analyses of GFP+ mutant cells up to early malignancy led to important insights into cancer initiation and pre-malignant progression, creating future opportunities for early detection and cancer prevention studies.

Results
MADM model reveals the process of mammary tumorigenesis initiated by sporadic loss of Brca1 and p53

To establish a MADM-based mouse model of breast cancer, we prepared two stock mouse lines through a multi-generational breeding scheme (Fig. S1). For one stock line, we bred the Brca1 and p53 mutant alleles (Jacks et al., 1994; Xu et al., 1999) onto the MADM-TG allele (Hippenmeyer et al., 2010). For the other stock line, we introduced the MMTV-Cre transgene (Wagner et al., 2001) into the MADM-GT line (Hippenmeyer et al., 2010) to target mammary epithelial cells. Finally, we inter-crossed the two stock lines to generate MADM-Brca1-p53; MMTV-Cre mice (Fig. S1B, hereafter referred to as MADM-mutant mice), in which sparse GFP+ Brca1, p53-null cells are predisposed to become cancerous (Fig. 1A).

The rarity of GFP+ mutants in MADM not only closely mimics human cancer initiation but also enables clonal analysis of pre-malignant expansion (Greaves and Maley, 2012; Knudson, 1971; Muzumdar et al., 2007). We assessed the abundance of GFP+ mutant cells in mammary glands from MADM-mutant mice at three months old, an age shortly after the peak of MMTV-Cre expression (Buono et al., 2006; Wagner et al., 1997). Using whole-mount fluorescence imaging of mammary glands, we found a high abundance of heterozygous yellow cells (GFP and RFP double positive) generated from MADM recombination events in G1 or post-mitotic cells (G0) (Fig. S2A). These yellow cells clearly illuminate the entire mammary ductal system and confirm the in vivo recombination by MMTV-Cre.
Among many yellow cells, we observed sparse and scattered GFP+ mutant cells along mammary ducts (Fig. 1B). For a higher resolution view, we sectioned mammary glands and performed confocal imaging to visualize the abundance of GFP+ mutant cells. We found that GFP+ mutant cells were often singular (Fig. 1C) and accounted for ~2% of all mammary epithelial cells while RFP+ wildtype sibling cells accounted for ~0.6% (Fig. 1D). This difference in percentage could be caused by a growth advantage of mutant cells, or a survival disadvantage of wildtype cells, or a mix of both. Notably, almost all MADM-labeled cells were positive for cytokeratin 8 (CK8, marker for mammary luminal cells) but negative for cytokeratin 14 (CK14, marker for mammary basal cells) (Fig. S2B-F), suggesting that they arose from the luminal layer where the reported cell of origin for basal-like breast cancer resides (Lim et al., 2009; Molyneux et al., 2010; Shehata et al., 2012). We are aware that conventional Cre reporter showed that MMTV-Cre is expressed in both luminal and basal cells (Wagner et al., 2001). One possible explanation for this discrepancy is that MADM-labeling relies on inter-chromosomal recombination, which requires much higher concentration of Cre than intra-chromosomal recombination for conventional floxed alleles (Zong et al., 2005). Therefore, although both luminal and basal cells express MMTV-Cre, the level could be much lower in the basal cells (Rabieifar, 2019 #1225), leading to luminal-specific recombination in MADM.

To assess the progression of initiated mutant cells, we collected mammary glands from MADM-mutant mice at different time points (10 mice at each age) and evaluated the overall expansion of GFP+ cells through whole-mount imaging. At 3 months, the expansion of GFP+ mutant cells was barely noticeable, but from 6 months onward, GFP+ foci became visible, gradually expanded, and eventually formed GFP+ tumors (Fig. 1E). By examining a cohort of 15 mice with an endpoint of 12 months of age, we found 13 (87%) developed GFP+ tumors with a median latency of 11 months (Fig. 1F). Despite the rarity of cancer-initiating mutant cells, the MADM-mutant mice showed a tumor latency only slightly longer than the ~9 months latency of conditional knockout mouse model that initiates cancer with numerous mutant cells (Xu et al., 1999). This suggests that the abundance of cancer-initiating cells is not a rate-limiting factor for the kinetics of breast cancers driven by Brca1 and p53 deficiency. Among 43 GFP+ tumors from 33 mice, most mice had 1-2 tumors but rarely 3 or more tumors (Fig. S2G). Finally, we did not observe obvious tropism within the 5 pairs of mammary glands, except for a slightly higher incidence in the largest 4th pair and a relatively lower incidence in the 2nd pair (Fig. S2H).

**MADM-mutant mammary tumors resemble human basal-like breast cancer**

Human BLBCs are characterized by a high proliferation index, lack of estrogen receptor (ER), progesterone receptor (PR), and HER2 over-expression (Palacios et al., 2005; Perou et al., 2000; Rakha et al., 2008). We assessed 6 MADM tumors of their proliferation and hormone receptor expression by immunohistochemistry and found them highly positive for Ki67 (~70% cells) and mostly negative for ER, PR, and HER2 (Fig. 2A, S3A), matching these histopathological features of human
BLBCs. To further examine whether the MADM tumors resemble human BLBCs at the molecular level, we performed RNA sequencing of 12 MADM tumors and extracted a panel of 50 genes (PAM50) used to stratify breast tumor subtypes (Perou, 2011). We co-clustered PAM50 signatures of our MADM tumors with the profiles of 1104 human breast tumors from TCGA dataset annotated for five breast cancer subtypes. To mitigate overall differences in gene abundance across species, we identified a unique set of mouse-to-human orthologs across all TCGA and MADM tumors and normalized each sample to obtain relative expression values for each gene (see methods for details). We found MADM tumors clustered with the human basal-like subtype but not others (Fig. 2B). Another hallmark of human BRCA1-mutated BLBC is the high frequency of copy number variations (CNVs) for genomic loci containing oncogenes or tumor suppressors (Annunziato et al., 2019; Weigman et al., 2012). To determine whether MADM tumors also share this hallmark, we conducted whole-exome sequencing on six MADM tumors and paired normal somatic tissues. We found recurrent amplification of multiple chromosomal segments harboring oncogenes—such as Met, Myc, and Fgfr1—along with recurrent deletion of the tumor suppressor gene Rb1 (Fig. 2C), closely corresponding to human basal-like tumors (Fig. S3B) (Network, 2012). Collectively, the histopathological, transcriptomic, and genomic analyses of MADM tumors demonstrate that our MADM-mutant mice represent an authentic model for human BLBC.

**MADM resolves the characteristic morphological stages of pre-malignant progression**

We next investigated the progressive phenotypic alterations of MADM mutant cells throughout pre-malignancy with a cohort of mice at the intermediate ages between cancer initiation and tumor formation. We performed tissue clearing with the CUBIC method (Susaki et al., 2015) and then conducted whole-mount, 3D imaging using light-sheet microscopy (Fig. S4A) to look for morphological abnormalities by comparing green mutant ducts with yellow heterozygous ducts (Fig. 3A). At 3 months of age, we observed short stretches of mutant cells that occupied a continuous region without causing noticeable alterations in ductal morphology. At 6 months old, some GFP+ mutant cells extended side branches resulting in a slightly more complex morphology than yellow ducts. This hyper-proliferation of mutant cells preceding prominent changes in tissue organization is consistent with observations in BRCA1-mutant carriers (Martins et al., 2012; McKian et al., 2009). Upon further expansion at 8 months, some mutant branches developed extensive epithelial buds reminiscent of alveologenesis during early pregnancy (Richert et al., 2000) even though these are virgin MADM-mutant mice. The alveologenesis was pervasive in late-stage mammary glands and exhibited prominently distinct histology when compared with the internal control yellow heterozygous ducts (Fig. S4B, C). We further quantified the number of buds (alveoli) per 100-µm primary ducts and found that mutant ducts had over ten-fold more alveoli than the controls (Fig. S4D, E). Binning the data into four levels based on the number of alveoli per 100 µm primary ducts—level 0 (<5), level I (5-20), level II (20-50), and level III (>50) (Fig. S4F)—we
found that high-level alveologenesis (II-III) exclusively occurred within the GFP+ mutant regions (Fig. S4G), suggesting that this “hyper-alveolarization” is a characteristic feature of pre-malignant lesions. In 10-month-old mice, we occasionally observed tiny GFP+ spherical masses (<1 mm in diameter) that had lost ductal morphology but were not yet palpable, which were termed as “micro-invasion” hereafter (Fig. 3A).

To further determine whether the hyper-alveolarized ducts and the micro-invasions reflect a sequential progression of mammary cancer in MADM-mutant mice, we evaluated whether there is a temporal sequence in the occurrence of these structures. Since multiple expansion levels of mutant cells often co-exist in the same gland (Fig. S4B), we plotted the most-advanced GFP+ lesion observed in the mammary gland of each MADM-mutant animal at a series of ages, and observed a progressive emergence of focally expanded GFP+ cells, hyper-alveolarized GFP+ ducts, and micro-invasions, culminating with the formation of GFP+ tumors at ~1 year of age (Fig. 3B). Histological analysis further supported a temporal sequence of these pre-malignant structures (Fig. 3C). While mammary epithelial cells from control wildtype mice appeared normal across all ages (Fig. S5A), mutant cells in hyper-alveolarized ducts present at 8 months displayed abnormal nucleomegaly (about 1.5x larger than wildtype nuclei), small but conspicuous nucleoli, and loss of the basally-oriented nuclear polarity (Fig. S5B). The micro-invasions contained residual hyper-alveolarized lobular units along with more advanced micro-invasive carcinoma (measuring less than 1 mm) comprised of individual cells and larger cords that elicited a stromal response (Fig. S5C). Finally, the frank tumors were dominated by infiltrating cells with a desmoplastic stromal response; no remnants of alveoli were visible (Fig. 3C). The progressive nature from hyper-alveolarizations to micro-invasions to full-blown tumors was further supported by Ki67 staining, revealing a gradual increase in cell proliferation (Fig. 3D, E). Taken together, the MADM-mutant model revealed that cancer-initiating cells progress through a visually identifiable sequence of ductal morphology alterations before the emergence of mammary tumors.

Hyper-alveolarized structures arise from mutant ductal rather than alveolar regions

The mammary epithelium is composed of two anatomically and functionally distinct compartments—the alveolar regions that produce milk during lactation and the ductal regions that drain milk to the nipple (Fig. 4A) (Visvader, 2009; Visvader and Stingl, 2014). It was unclear whether the hyper-alveolarized structures consisting of mutant cells originated from alveolar or ductal mutants. Since MADM generates one RFP+ wildtype sibling cell alongside each original GFP+ mutant cell (Fig. 1A), we investigated this question using the “twin-spot” analysis that directly compares mutant cells with their wildtype sibling cells in a clone-by-clone fashion within ductal and alveolar compartments, respectively (Espinosa and Luo, 2008; Muzumdar et al., 2007; Terry et al., 2020).

For the twin-spot analysis, we selected mice at four months old because the focal expansion of GFP+ mutant cells was evident at this age, while hyper-alveologenesis was not yet present (Fig. 3B).
As a baseline, we first analyzed 40 twin spots in four MADM-wildtype mice that lacked \textit{Brca1} and \textit{p53} mutant alleles (Fig. S1B), and confirmed that GFP+/RFP+ cells did not expand prominently and remained similar in number in both ductal and alveolar regions (Fig. 4B, D, \textit{data for each individual mouse plotted in Fig. S6A}). In contrast, when we performed the twin-spot analysis in four MADM-mutant mice, we readily observed prominent expansion of GFP+ mutant clones over RFP+ wildtype clones in the ductal region (Fig. 4C, E, \textit{data for each individual mouse plotted in Fig. S6B}). Surprisingly, the mutant clones in alveoli, even though harboring the same \textit{Brca1} and \textit{p53} mutations as those in ducts, didn’t expand and showed no difference in clonal size from the neighboring RFP+ wildtype clones (Fig. 4C, E). To conduct the twin-spot analysis more rigorously, we compared the sizes of the GFP+ and RFP+ sibling clones in a strictly pairwise manner, and found that only GFP+ mutant clones within the ductal regions exhibited significantly larger sizes than their RFP+ sibling clones (Fig. 4F), indicating that the initial clonal expansion of \textit{Brca1}, \textit{p53}-null cells occurs exclusively in the ductal region. This finding is particularly interesting because, while multiple studies with human tissues and conventional mouse models have noted the outgrowth of alveolar-like buds during the pre-malignant development of breast cancer and interpreted it as aberrant alveolar cell expansion (Bach et al., 2021; McKian et al., 2009; Poole et al., 2006; Tao et al., 2017), our results showed that such outgrowth most likely originates from luminal progenitors in the ducts that are either intrinsically fated for the alveolar lineage (i.e., alveolar luminal progenitors within the ducts) or acquire the aberrant alveolar fate due to \textit{p53}/BRCA1-loss.

\textbf{The onset of luminal-to-basal transition coincides with the appearance of micro-invasion}

Basal-like breast cancers express basal-cell markers (CK5/14, α-SMA, P63, etc.) yet arise from luminal progenitor cells (Lim et al., 2009; Liu et al., 2007; Molyneux et al., 2010; Shehata et al., 2012). This luminal-to-basal transition is thought to be critical for cancer progression, as it promotes stemness and invasiveness of \textit{BRCA1}- or \textit{p53}-mutant mammary epithelial cells in vitro (Bai et al., 2022; Kim et al., 2011; Liu et al., 2008; Mizuno et al., 2010). However, in vivo mapping of luminal-to-basal transitions during tumorigenesis is lacking. To determine whether luminal-to-basal transition occurs in the two representative pre-malignant stages of MADM-mutant mice (hyper-alveolarized ducts and micro-invasions), we leveraged the single-cell resolution of MADM to carefully assess the mutant cells at each stage. Normal mammary luminal cells present a cuboidal shape, whereas basal cells show an elongated spindle shape (Rios et al., 2014). Within hyper-alveolarized ducts in ~8 months old MADM-mutant mice, mutant cells mostly exhibited a cuboidal shape with relatively homogeneous cell size (Fig. 5A upper panel). In contrast, mutant cells in micro-invasions in ~10 months old MADM-mutant mice showed heterogeneous morphologies, with a fraction adopting an elongated spindle shape reminiscent of basal cells (Fig. 5A lower panel). When we quantified the cell size and circularity of mutant cells in hyper-alveolarized ducts and micro-invasions from four MADM-mutant mice, we found that the mutant
cells in micro-invasions were significantly larger in size and lower in circularity than those in hyper-alveolarized ducts (Fig. 5B, C), implying a transition of cell state between these two stages.

To clarify whether the morphological change represents a luminal-to-basal transition, we assessed the expression of luminal cell marker (E-Cadherin) and two basal cell markers (α-SMA, CK5) in mutant cells by immunofluorescent staining. In hyper-alveolarized ducts, GFP+ mutant cells maintained the expression of E-cadherin and were negative for basal markers (Fig. 5D₁, D₂ upper panel). In micro-invasions, by contrast, some mutant cells co-expressed E-cadherin and basal cell markers (Fig. 5D₁, D₂ lower panel), suggesting an incomplete transition from the luminal to basal cell state. We quantified this observation by analyzing a total of ~700 cells in mammary tissues from four mice with hyper-alveolarized ducts or micro-invasions. In hyper-alveolarized ducts, E-Cadherin-positive mutant cells were almost exclusively negative for α-SMA and CK5, but in micro-invasions ~50% of mutant cells expressed both basal cell markers and E-Cadherin (Fig. 5E₁, E₂). The incompleteness of the luminal-to-basal transition in micro-invasions was further supported by the prevalent expression of another luminal marker CK8/18 in GFP+ mutant cells, comparable to that within hyper-alveolarized mutant ducts (Fig. S7A). Only sporadic estrogen receptor (ER) positive cells were found at both stages (Fig. S7B), which is consistent with the current understanding that ER- luminal progenitor cells serve as the cancer cell of origin of basal-like breast cancer with BRCA1 mutations (Lim et al., 2009). Taken together, the cellular heterogeneity and the incomplete luminal-to-basal transition in micro-invasions revealed by our study likely present an excellent therapeutic opportunity and warrant further comprehensive molecular and cellular analysis.

Discussion

In this study, we established a mouse genetic mosaic model for BLBC that share histopathological, transcriptomic, and genomic similarities with human basal-like breast cancers. Taking advantage of the spatial resolution provided by MADM, we identified multiple morphologically distinct pre-malignant stages, including focal expansion of mutant cells, hyper-alveolarization of mutant ducts, and micro-invasion. Surprisingly, the clonal analysis revealed that hyper-alveolarized mutant structures originate from ductal rather than alveolar cells. Further progression from hyper-alveolarized structures to micro-invasions results in loss of ductal organization and an incomplete luminal-to-basal transition, manifested by enlarged cell size, elongated cell shape, elevated cell proliferation, and the gaining of basal marker expression without losing luminal marker expression. Taken together, Our MADM-based mouse model presents a useful tool for studying the pre-malignancy BLBC, which should empower pre-clinical research on early detection and cancer prevention.

MADM-based cancer models offer several unique advantages. First, MADM creates a small number of homozygous mutant cells within a heterozygous animal, closing mimicking the sporadic loss of heterozygosity (Lohmussaar et al.) of tumor suppressor genes in cancer. For germline BRCA1
mutation carriers, cancer is often initiated by the sporadic loss of the wildtype \textit{BRCA1} allele (Cornelis et al., 1995; Maxwell et al., 2017; Nones et al., 2019). Second, the generation of mutant cells by MADM is coupled with permanent GFP labeling through a mitotic recombination event, allowing mutant cells to be unequivocally identified and tracked throughout the entire process of tumorigenesis (Muzumdar et al., 2007; Zong et al., 2005). Since both GFP and tdTomato are expressed at the high enough level to be visible under a fluorescence stereomicroscope within fresh, unstained mammary glands (Fig. 1E), MADM enables the gross evaluation of pre-malignant phenotypes and targeted sample collection prior to various downstream analyses that require unfixed tissues. Third, MADM generates sibling wildtype cells that are labeled with RFP, providing an internal control for the GFP+ mutant cells (Beattie et al., 2017; Terry et al., 2020). Without RFP+ wildtype siblings as a reference (Fig. 4), it would be difficult to distinguish the clonal expansion of GFP+ mutant ductal cells from stochastic neutral drift (Lopez-Garcia et al., 2010; Snippert et al., 2010). While our study is focused on breast cancer modeling, it should be noted that, due to the modular nature of the MADM system and the availability of a genome-wide library of MADM mice (Contreras et al., 2021), one could establish many other cancer models to study their initiation and pre-malignant progression upon the sporadic LOH of relevant tumor suppressor genes.

The observed progression trajectory in the MADM-mutant mice provided multiple insights for understanding the early genesis of BLBC. Although hyper-alveolarization was initially thought to reflect the aberrant expansion of mutant alveolar cells (Tao et al., 2017), our study showed that it actually originates from mutant ductal cells either intrinsically fated for or mis-differentiated toward the alveolar fate upon \textit{p53} and \textit{BRCA1} mutations. Corroborating our finding, the recent single-cell sequencing of mammary glands from \textit{Brca1}, \textit{p53}-deficient mice spanning various pre-malignant ages revealed dysregulation of transcription factors driving alveologenesis in luminal progenitor cells, causing aberrant alveolar outgrowth (Bach et al., 2021). Intriguingly, this finding may point to distinct roles of hormonal signaling between pre-malignancy and malignancy of BLBC: the hyper-alveolarization of mutant ducts resembles alveolar outgrowth during early pregnancy (Macias and Hinck, 2012), a process known to be regulated by the progesterone signaling (Brisken et al., 1998; Humphreys et al., 1997; Lydon et al., 1995); and progesterone receptors are overexpressed in \textit{Brca1}-deficient mammary epithelial cells of both human and mouse, and exposure to exogenous progesterone dramatically increases mammary gland volume in \textit{Brca1}-deficient mice (King et al., 2004; Ma et al., 2006; Poole et al., 2006). Therefore, although the expression of hormonal receptors are low/no in malignant BLBCs, progesterone signaling could play significant roles at pre-malignancy, which warrants further study and may offer a new avenue of cancer prevention (Nolan et al., 2016; Sigl et al., 2016; Trabert et al., 2020).

At the stage of micro-invasion, MADM-mutant cells undergo a partial luminal-to-basal transition, which could increase the stemness and invasiveness of \textit{Brca1} or \textit{p53}-deficient cells as reported in the literature (Bai et al., 2022; Kim et al., 2011; Liu et al., 2008; Luond et al., 2021; Mizuno et al., 2010) and was recently implicated as a critical step at the onset of basal-like tumorigenesis (Landragin et al.,
Our time course analysis of Brca1, p53-deficient cells in vivo showed that the loss of Brca1 and p53 does not immediately induce a luminal-to-basal transition even after the manifestation of hyper-alveolarization. Instead, mutant cells only reach an incomplete luminal-to-basal transition at ~10 months after mutant cells progress to micro-invasive lesions. While we cannot rule out cell-intrinsic mechanisms for this transition, our observation implies that the exposure of luminal cells to extrinsic stromal factors due to basement membrane breaching in micro-invasion could be the trigger for the luminal-to-basal transition, which should be investigated more thoroughly in the future.

While powerful, MADM has certain limitations. First, the tumor latency tends to be long, e.g. it takes ~8 months for MADM-mutant mice to progress into the pre-malignant stages in this study. If desired, additional clinically relevant mutations could be introduced to accelerate cancer development in our model (Annunziato et al., 2019). Compound mutations that are syntenic with Brca1 and p53 can be introduced using the same scheme as shown in Figure S1B; for mutations that are not syntenic, mutant alleles can be introduced into the MADM model through conventional breeding schemes (Muzumdar et al., 2016; Yao et al., 2020). Second, due to the involvement of many genetic elements in the MADM model, it tends to have mixed genetic background that precludes allotransplantation experiments. When necessary, one could backcross stock mice into desired pure genetic background to improve the versatility of this model. Third, MADM relies on mitotic recombination to generate mutant cells, and thus cannot be used to mutate post-mitotic cells. Finally, when constitutively expressed Cre transgene is use, e.g. MMTV-Cre for this study, the birth timing of mutant cells is not clear. When desired, one could use temporally controlled Cre lines such as Tet or CreER system to precisely control the timing of tumor initiation. Notwithstanding these limitations, the current MADM model for BLBC enables spatially resolved analysis throughout the pre-malignant progression, and can greatly facilitate studies of early detection and cancer prevention.

Materials and Methods

Animal

The following mouse lines were crossed to establish the MADM-mutant and MADM-wildtype mice: TG11ML (stock NO. 022976 JAX) (Henner et al., 2013), GT11ML (stock NO. 022977 JAX) (Henner et al., 2013), Brca1<sup>flox</sup> (strain NO. 01XC8; NCI) (Xu et al., 1999), p53<sup>KO</sup> (stock NO. 002101; JAX) (Jacks et al., 1994), MMTV-Cre (stock NO. 003553 JAX) (Wagner et al., 1997). The breeding schemes are shown in Fig. S1. We exclusively used female mice, primarily focusing on the 4th pair of mammary glands (the abdominal pair) for data collection, unless otherwise specified in figure legends. All animal work was performed in the University of Virginia Animal Vivarium. All procedures, including housing and husbandry were approved by the Institutional Animal Care and Use Committee (IACUC) at the University of Virginia, following national guidelines to ensure the humanity of all animal experiments.
Genotyping

For genotyping, the mouse toe was used to extract DNA for PCR. 120 µl of 50 mM NaOH was added to each toe and then incubated at 95°C for 20 min in the PCR machine, followed by adding 30 µl of 1 M Tris-HCl (pH 7.4) and mixing. 1 µl of toe solution was used for PCR template in a 20 µl PCR reaction. The PCR primer sequences are listed below.

1) *Eif* for MADM TG/GT cassettes: primer-1, 5-TGGAGGAGGACAAACTGGTCAC-3; primer-2, 5-TCAATGCGGCGGGGTGTGT-3; primer-3, 5-TTCCCTTTCTGCTTCATCTTT-3; PCR products, knock-in (KI) band, 230 bps and WT band, 350 bps. 2) *MMTV-Cre*: primer-1, 5-CACCCTGTATTAGCCGCG-3; primer-2, 5-GAGTCATCCTTAGCCGCGTA-3; PCR product, KI band, 300 bps. 3) *p53KO*: primer-1, 5-ACCGCTATCACGGACATAGC-3; primer-2, 5-CACAGCGTGGTGGTACCTTATG-3; primer-3, 5-GGTATACTCAGAGCCGGCCTG-3; PCR products, KI band, 700 bps and WT band, 450 bps. 4) *Brca1flox*: primer-1, 5-CTGGGTAGTTTGTAAGCATCC-3, primer-2, 5-TCTTATGCCCTCAGAAAACTC-3; PCR products, flox/flox band, 365 bps and WT band, 297 bps.

Immunofluorescence and immunohistochemistry

Mammary glands were harvested and fixed with 4% paraformaldehyde (PFA) at 4°C for 24 h. For immunofluorescence, tissues were then washed with PBS twice, soaked with 30% sucrose at 4°C for 48 h, and embedded in optimal cutting temperature (OCT). The tissues were sectioned at 20 µm thickness with Thermo NX50 Cryostat. For staining, slides were first blocked in 0.3% Triton-X 100 and 5% normal donkey serum in PBS for 20 min, then incubated with primary antibodies (CK8, Abcam ab182875, 1:200; CK14, Biolegend 905301, 1:400; E-cadherin, Biolegend 147301, 1:200; α-SMA, Sigma A5228, 1:500) diluted in blocking buffer at 4°C overnight. Secondary antibody incubation was performed for 1 h at room temperature in PBS. To stain nuclei, slides were incubated in DAPI solution (1 µg/mL in PBS) for 5 min before being mounted with 70% glycerol. Fluorescent images were acquired on Zeiss LSM 700/710 confocal microscope. Images were processed with Zen and Fiji. For immunohistochemistry, PFA fixed tissue was further processed for paraffin embedding and then sectioned at 4 µm thickness. After antigen retrieval, the primary antibody (Ki67, Epitomics 4203-1, 1:400) was incubated at 4°C overnight, HRP-conjugated secondary antibodies were then used, and 3,3’-diaminobenzidine (Vector Laboratories, SK-4100) was used to develop color.

Tissue clearing with the CUBIC method and 3D imaging

PFA-fixed mammary glands were cleared for large-scale 3D imaging with the standard CUBIC method (Susaki et al., 2015). Briefly, tissues were immersed in 50% reagent-1 (25 wt% urea, 25 wt% Quadrol, 15 wt% Triton X-100, 35 wt% dH₂O), shaken at 110rpm 37°C for 12 h, and then transferred to 100% reagent-1 with DAPI (1 µg/ml) for shaking until mammary glands became transparent. After reagent-1,
tissues were washed three times with PBS, 1 h each time with shaking to remove the reagent-1. Tissues were then shaken in 50% reagent-2 for 12h at 37°C, followed by 100% reagent-2 (25 wt% urea, 50 wt% sucrose, 10 wt% triethanolamine, 15 wt% dH2O), shaking for 48 h. The Zeiss Z.1 light-sheet microscopy system was used for acquiring images. Tissues were attached to the holder of the light-sheet microscope with super glue.

**Gene expression profiling**

About 50 mg of tissue from each mammary tumor was used for RNA extraction. The tissue was homogenized in 500 µl TRlitzol, then 100 µl chloroform was added and mixed thoroughly, followed by centrifugation (12,000 rcf) for 15 mins at 4°C. The upper layer aqueous phase containing the RNA was transferred to a new tube, and 5 µg of polyacrylamide was added, followed by an equal volume of 70% ethanol. Afterward, 700 µl of the sample was used as the input for RNA isolation using the RNeasy Mini Kit (QIAGEN) according to the manufacturer’s protocol. The quality of RNA samples was evaluated with Bioanalyzer (Agilent Technologies), and samples with an RNA Integrity Number (RIN) > 8 were used for library preparation. Libraries were prepared with a TruSeq Stranded mRNA Library Prep kit (Illumina). Libraries were multiplexed at an equimolar ratio, and 1.3 pM of the multiplexed pool was sequenced on a NextSeq 500 instrument with a NextSeq 500/550 Mid/High Output v2.5 kit (Illumina) to obtain 75-bp paired-end reads. From the sequencing reads, adapters were trimmed using fastq-mcf in the EAutils package (version ea-utils.1.1.2-779) with the following options: -q 10 -t 0.01 -k 0 (quality threshold 10, 0.01% occurrence frequency, no nucleotide skew causing cycle removal). Quality checks were performed using FastQC (version 0.11.8) and MultiQC (version 1.7). Data were aligned to the murine transcriptome (GRCm38.84) using HISATv2 (version 2.1.0) with options for paired-end reads. HISAT read counts were converted to transcripts and normalized to transcripts per million (TPM) using StringTie (version 2.1.6).

**PAM50 extraction and comparison with TCGA datasets.**

TCGA breast cancer expression data was obtained from the UCSC genome browser (Ciriello et al., 2015). Human orthologs for mouse genes were obtained from the Ensembl biomart in R using the getAttributes function. For both the human and murine datasets, we obtained the intersection of unique orthologs to obtain a set of 14,980 mice to human ortholog genes to evaluate co-expression. To enable cross-species comparisons, we performed a sample-wise column normalization to obtain new transcripts per million (TPM) estimates that accounted for gross differences in gene abundance between species. Hierarchical clustering of PAM50 genes was performed using “pheatmap” in R using Euclidean distance and “ward. D2” linkage.
Whole exome sequencing and copy number variation analysis

Genomic DNA was prepared from tumors developed in MADM-mutant mice and from the tails of the same mice as the control with a DNeasy Blood and Tissue Kit (Qiagen). Whole-exome sequencing at 100x coverage was performed as a contract service with Genewiz. Raw BCL files were converted to fastq files with bcl2fastq v.2.19 and adapter was trimmed with Trimmomatic v.0.38. Trimmed reads were mapped to the mouse reference genome, and somatic variants and copy number variations were called using the Dragen Bio-IT Platform (Illumina) in somatic mode and a panel of normals to remove technical artifacts. The filtered VCF was annotated with Ensembl Variant Effect Predictor (VEP) v95 for the Ensembl transcripts overlapping with the filtered variants. CNVs that passed quality control filters from Dragen were visualized using the GenVisR v.1.16.1 package in R.

Statistical analysis

Statistical analysis was performed with GraphPad Prism. Bar graphs were presented as the mean ± standard error of the mean (s.e.m.) unless otherwise annotated in the figure legend. The normality of data distribution was checked with qqplot in R. Depending on whether the data followed a normal distribution, the Student’s t-test or Mann-Whitney U test was used as indicated in the figure legends. The Chi-square test or Fisher’s Exact test was used to test frequency distribution as indicated in the figure legends. Statistical significance is noted by “not significant (n.s.)” p > 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001.

Acknowledgments

We thank Dr. Bing Xu and Xiaoyu Zhao for providing feedback on the manuscript. We thank Dr. Qinlei Gu for the artwork of the mammary glands distribution in mice. We also thank Dr. Pat Pramoonjago at the Biorepository and Tissue Research Facility, Sheri Vanhoose at the Research Histology Core, Dr. Stacey Criswell at the Advanced Microscopy Facility, and Shelly Verling at the vivarium for their assistance on the project. We thank Dr. Ammasi Periasamy at the UVA Keck Center for their assistance with the Zeiss Light-sheet Z.1 microscopy system.

Competing Interests

The authors declare no competing or financial interests.

Funding

This work was supported by the Basser Center for BRCA (H.Z.), Mary Kay Foundation (H.Z.), the Pinn Scholarship (H.Z.), the UVA Cancer Center Seed Grant (H.Z.), the National Cancer Institute #R01-CA194470 (K.A.J.), #R01-CA256199 (K.A.J. & H.Z.), #U54-CA274499 (K.A.J.), the UVA Cancer Center
Training Grant (J.Z.), and the UVA Wagner Fellowship (S.S.). The core facilities are supported by UVA Cancer Center Grant #P30-CA044579.

Data availability

Author contributions

References


Fig. 1. A MADM model that tracks the process of Brca1, p53-dependent mammary tumorigenesis from sporadic cancer-initiating cells to frank tumors

(A) Modeling breast cancer development from sporadic mutant cells to tumors with MADM. From a colorless heterozygous mouse, through inter-chromosomal recombination in mitotic cells at the G2 phase, MADM generates one GFP+ mutant cell and one RFP+ wildtype cells after X segregation (two recombinant sister chromatids segregate into different daughter cells); alternatively, Z segregation generates one colorless and one dual-colored (yellow) cell that are both heterozygous (two recombinant sister chromatids segregate into the same daughter cells). The schematic of MADM is reproduced with permission from (Zong et al., 2005), Cell Press.

(B) The MADM model induces sparse and scattered GFP+ mutant cells (arrows) in mouse mammary glands. The image is representative of four MADM-mutant mice collected at three months old for whole-mount fluorescence imaging. Scale bar =1 mm.

(C) High-resolution imaging of the GFP+ mutant cells (arrows) in mammary glands from three-month-old MADM-mutant mice. The image is representative of mammary tissue sections acquired from
four mice at three months old and imaged with wide-field fluorescence microscopy. Scale bar =100 μm.

(D) The proportion of MADM-labeled cells among all mammary epithelial cells in MADM-mutant mice at three months old. Sections of mammary glands were imaged with wide-field microscopy. The number of MADM-labeled and total mammary ductal cells was counted by GFP/RFP fluorescence and DAPI, respectively. Data are shown as the mean percentage ± s.d. from n =4 mice.

(E) Fluorescence imaging of whole mammary glands from a cohort of MADM-mutant mice at different ages, showing the progressive expansion of GFP+ mutant cells toward tumor formation. Arrow indicates GFP+ focal expansions. Images were collected on a fluorescence stereomicroscope and are representative of 10 mice for each age. Scale bar =500 μm.

(F) Percent tumor-free of MADM-mutant mice (n =15) after 12 months as the endpoint, showing a median latency of 11 months. Quantiles at the bottom show a narrow spreading of tumor latency.
Fig. 2. MADM-mutant mammary tumors resemble human basal-like breast cancer

(A) Whole-mount fluorescence image of GFP+ mammary tumors (upper left); H&E (upper middle), immunohistochemistry staining of Ki67 (upper right), estrogen receptor (ER, lower left with a mouse oviduct as a positive control in the inset), progesterone receptor (PR, lower middle with a mouse uterus as a positive control in the inset), and Her2 (lower right with a mouse HER2-amplified mammary tumor as a positive control in the inset). Representative images of tumors from 6 mice. Scale bar =100 μm.

(B) PAM50-based clustering of MADM tumors with human breast cancers previously subtyped by PAM50 analysis. 12 MADM tumors and 1104 human breast tumors from TCGA datasets were analyzed. Cross-species differences were normalized using a set of mouse-to-human orthologs (see methods for details).

(C) Copy-number variations (CNVs) in six MADM tumors were assessed by whole-exome sequencing. Gains in *Met, Fgfr1*, and *Myc*, and loss of *Rb1* are highlighted.
Fig. 3. MADM-mutant mammary glands reveal stereotyped morphological changes of pre-malignant ductal structures

(A) Progressive morphological changes in green mutant ducts compared to yellow internal control heterozygous ducts as mice age. Upper panel: Low-magnification 3D fluorescence imaging of mammary glands from a cohort of mice at different ages; Middle and lower panel: Higher magnification of the control yellow ducts and the green mutant ducts. Arrow in the left panel indicates a GFP+ mutant region within the mammary ducts. Mammary glands from MADM-mutant mice at each age (n =10) were cleared with the CUBIC method, and 3D high-resolution images were acquired by light-sheet microscopy. Scale bar =100 μm.

(B) Timeline of the first occurrence of each pre-malignant stage of GFP+ mutant foci. The morphology of all mutant ducts in MADM-mutant mice at each age (n =10) was examined by whole-mammary-gland fluorescence imaging. Mice are categorized based on the furthest stage of pre-malignancy reached. The proportional distribution of mice at each age is plotted.
(C) H&E staining of the progressive change of mammary ductal shape from normal to hyper-alveolarized, to micro-invasive, and to tumors. Paraffin slides of mammary glands from MADM-mutant mice at each age ($n=10$) were used for staining. Scale bar =200 μm.

(D) Ki67 staining of normal mammary ducts and MADM-mutant lesions at different stages. Ducts of each morphology were collected from four mice. Scale bar =100 μm.

(E) Quantification of Ki67 percent positivity ($\text{Ki67}+%$) of cells within normal ducts and MADM-mutant lesions at the hyper-alveolarized and micro-invasive stages. The dashed line indicates the average Ki67+% in frank tumors. Data are represented as mean ± s.e.m. from $n=4$ mice. ***<0.001 by Mann–Whitney test.
Fig. 4. Ductal but not alveolar cells show initial clonal expansion in MADM-mutant glands

(A) Schematic of ductal and alveolar regions in the mouse mammary gland.

(B) GFP+ and RFP+ clone pairs of MADM-wildtype mice are of similar sizes and do not show prominent expansion. 3D fluorescence images of cleared mammary glands from MADM-wildtype mice at four months old (n=4) were acquired as Z-stacks on a confocal microscope. The green arrows indicate the GFP+ cells, and the red arrows indicate RFP+ cells. Scale bar =100 μm.

(C) GFP+ mutant clones in the ductal region of MADM-mutant mice show prominent expansion, while their sibling RFP+ clones do not. In the alveolar region, neither GFP+ nor RFP+ clones expand. 3D fluorescence images of cleared mammary glands from MADM-mutant mice at four months old (n=4) were acquired as Z-stacks on a confocal microscope. The green arrows indicate the GFP+ cells, and the red arrows indicate RFP+ cells. Scale bar =100 μm.

(D) The clonal size distribution of GFP+ and RFP+ clones in ductal and alveolar regions of mammary glands from MADM-wildtype mice at four months old. Data were pooled from four mice, and the total number of clones is indicated for each group. n.s. p>0.05 by Fisher’s Exact test.

(E) The clonal size distribution of GFP+ and RFP+ clones in ductal and alveolar regions of mammary glands from MADM-mutant mice at four months old. Data were pooled from four mice, and the total number of clones is indicated for each group. n.s. p>0.05, ***p<0.0001 by Fisher’s Exact test.
(F) Quantification of the clonal size for paired GFP+ and RFP+ sibling cells in the same clones in the ductal and alveolar regions of mammary glands from both MADM-wildtype and MADM-mutant mice at four months. Clones from four mice were pooled. Data are represented as mean ± s.e.m. ***p<0.001 by paired t-test.
Fig. 5. MADM-mutant cells undergo a partial luminal-to-basal transition upon micro-invasion

(A) Morphology of GFP+ mutant cells in hyper-alveolarized ducts and micro-invasive lesions. Mammary gland sections were imaged by confocal microscopy. For each feature, samples from four mice were assessed. The dashed regions delineate single cells. Scale bar =50 μm.

(B) Size of mutant cells in hyper-alveolarized ducts and micro-invasions. Cells from four mice that represent each stage were plotted for each mouse. Data are represented as median ± interquartile range. **p<0.01, by nested t-test.

(C) The circularity of mutant cells in hyper-alveolarized ducts and micro-invasions. Cells from four mice that represent each stage were plotted for each mouse. Quartiles are shown. Data are represented as median ± interquartile range from n=50 cells. **p<0.01, by nested t-test.

(D) In hyper-alveolarized ducts, GFP+ mutant cells were exclusively positive for E-cadherin but negative (outside of the GFP+ signal) for α-SMA (D₁) or CK5 (D₂), whereas in micro-invasions, some mutant cells were E-cad+ α-SMA+ dual positive (D₁) or E-cad+CK5+ dual positive (D₂).
Frozen sections of mammary glands were stained and imaged by confocal microscopy. For each feature, samples from four mice were analyzed. Scale bar =50 μm.

(E) $E_1$: the proportion of E-cad+SMA- (luminal) and E-cad+SMA+ (partially transitioned) mutant cells; $E_2$ the proportion of E-cad+CK5- (luminal) and E-cad+CK5+ (partially transitioned) mutant cells; in hyper-alveolarized ducts and micro-invasions. For each data point, a total of ~700 mutant cells from mammary glands from four mice were assessed. Data are represented as mean ± s.d.
Fig. S1. The breeding scheme to build two stock mouse lines for a MADM-based breast cancer mouse model with Brca1, p53 deficiency

(A) Location of MADM TG/GT cassettes, p53, and Brca1 on mouse chromosome 11. The physical locations were indicated.

(B) The breeding scheme to incorporate p53 and Brca1 mutations into MADM-TG stock and the MMTV-Cre transgenes into the MADM-GT stock. Mating between the TG and GT stock produces MADM p53-Brca1 mice (MADM mutant) and the control MADM wildtype mice.
Fig. S2. Additional mechanism to generate yellow cells, MADM labeling specificity in the luminal cells, and distribution of tumor among glands.

(A) Cre-mediated inter-chromosomal recombination could also occur in G1 or post-mitotic cells (G0), which generate dual-colored yellow cells without altering genotype (heterozygous).

(B) MADM-colored cells locate in the CK8+ luminal layer but not in the CK14+ basal layer. Immunofluorescence staining of CK8 and CK14 was performed with sections of mammary glands from MADM-mutant mice at three months old (n =3). Scale bar =50 μm.

(C) Quantification of the percentage of MADM-colored cells that are CK8+ (luminal) or CK14+ (basal) from three MADM-mutant mice at three months old (n =3). Data are represented as mean ± s.d.

(D) Representative whole-mount fluorescence image of MADM mammary glands used for disassociation and flow analysis. Three mice at ~ 10 months old were used.

(E) MADM-colored cells were predominantly found in luminal cells but not in basal cells.

(F) The percentage of MADM-labeled cells that are luminal or basal cells. Each dot represents data from one mouse (n =3). Data are presented as mean ± s.d.

(G) The total number of tumors per mouse among a total of 33 MADM-mutant mice.

(H) The distribution of tumors among 5 pairs of mammary glands, based on the assessment of a total of 43 GFP+ tumors from 33 MADM-mutant mice.
Fig. S3. The hormone receptor status of MADM tumors and CNVs in human basal-like breast cancer

(A) Representative images of ER, PR, and Her2 status in three more MADM tumors showed a triple-negative phenotype. Immunohistochemistry staining was performed with tumor sections. Scale bar =100 μm.

(B) Analysis with TCGA datasets for human basal-like breast cancer, showing consistent amplification of FGFR1, MYC, loss of RB1, and over-expression of MET.
Fig. S4. Hyper-alveologenesis occurred specifically in mutant ducts.

(A) The pipeline for acquiring large-scale 3D images of mammary ducts with the CUBIC clearing method and light-sheet microscopy.

(B) Whole-mount fluorescence imaging of cleared mammary glands showing distinct morphology of mutant ducts. Mammary glands from 4 MADM-mutant mice at 8 months old mice were assessed. Scale bar =100 μm.

(C) H&E staining of hyper-alveolarized mutant ducts and normal-shape control ducts.

(D) The scheme to count the number of alveoli per 100 μm major ducts to quantify ductal-alveologenesis.

(E) Increased ductal-alveologenesis level of the green mutant ducts compared with the yellow heterozygous ducts (internal control) as shown by the quantification of ductal alveologenesis level from 8-month-old mice (n=4). Data are represented as mean ± s.e.m. Mann–Whitney test was used, ***<0.0001.

(F) Examples and criteria for categorizing mutant ducts into different alveologenesis levels.

(G) The distribution of ductal-alveologenesis levels among yellow heterozygous ducts (internal control) and green mutant ducts in glands from 8-month-old mice (n=6). 49 yellow control ducts and 23 green mutant ducts were imaged and quantified. A Chi-square test was used. **** p<0.0001.
Fig. S5. Morphology of mammary ducts in MADM-wildtype mice along aging, and detailed pathological features of hyper-alveolarized mutant ducts and micro-invasive lesions.

(A) Upper panel: whole-mount fluorescence imaging of mammary glands from MADM-wildtype mice at a cohort of ages, showing no expansions of GFP+ foci or prominent morphological change of mammary ducts. Scale bar =5 mm. Lower panel: H&E staining showing no morphological changes of mammary ducts along mice aging. Scale bar =200 μm. n=3 for each age.

(B) H&E staining of hyper-alveolarized mutant ducts and age-matched wildtype ducts. Mutant cells in hyper-alveolarized ducts exhibited 1) abnormal nucleomegaly, with nuclei approximately 1.5 times larger than those of wildtype cells (dashed circles), 2) small yet conspicuous nucleoli (arrows), indicative of proliferation, and 3) loss of the basolateral axis of nuclear polarity in some cells (arrowheads). Representative images from four mice. Scale bar =100 μm.

(C) H&E staining of micro-invasive lesions. The micro-invasive lesions encompass residual hyper-alveolarized units (white dashed circles), alongside with more advanced micro-invasive carcinoma (on the right), measuring less than 1 mm. Within these lesions, individual cells (red arrows) and cords of cells lacking the lumen (red dashed circles) invade the adipose area and elicit a stromal response marked by increased collagen deposition (red dashed lines, intense eosin stains). Inflammatory responses are evident at the border. Representative images from four mice. Scale bar =100 μm.
Fig. S6. Size distribution of GFP+ and RFP+ clones in ductal and alveolar regions of mammary glands from individual MADM-wildtype and MADM-mutant mice.

(A) Size distribution of GFP+ and RFP+ clones in ductal and alveolar regions of mammary glands in four individual MADM-wildtype mice.

(B) Size distribution of GFP+ and RFP+ clones in ductal and alveolar regions of mammary glands in four individual MADM-mutant mice.

*The total number of clones is indicated in parenthesis at the bottom.
Fig. S7. CK8/18 and ER staining of hyper-alveolarized mutant ducts and micro-invasive lesions.

(A) Cells in both hyper-alveolarized ducts and micro-invasive lesions are mostly CK8/18+. Frozen sections of mammary glands from four mice that represent each stage were stained and imaged by confocal microscopy. Scale bar =50 μm.

(A) The proportion of CK8/18+ cells among all GFP+ mutant cells. Each dot represent data from one mice. For each mouse, ~1000 cells from four imaging fields were quantified. Data are represented as mean ± s.d., n.s. >0.05 by Mann–Whitney test.

(B) Only sporadic ER+ cells are present in hyper-alveolarized ducts and micro-invasive lesions. Frozen sections of mammary glands from four mice that represent each stage were stained and imaged by confocal microscopy. Scale bar =50 μm.

(C) The proportion of ER+ cells among all GFP+ mutant cells. Each dot represent data from one mice. Data are represented as mean ± s.d., n.s. >0.05 by Mann–Whitney test.