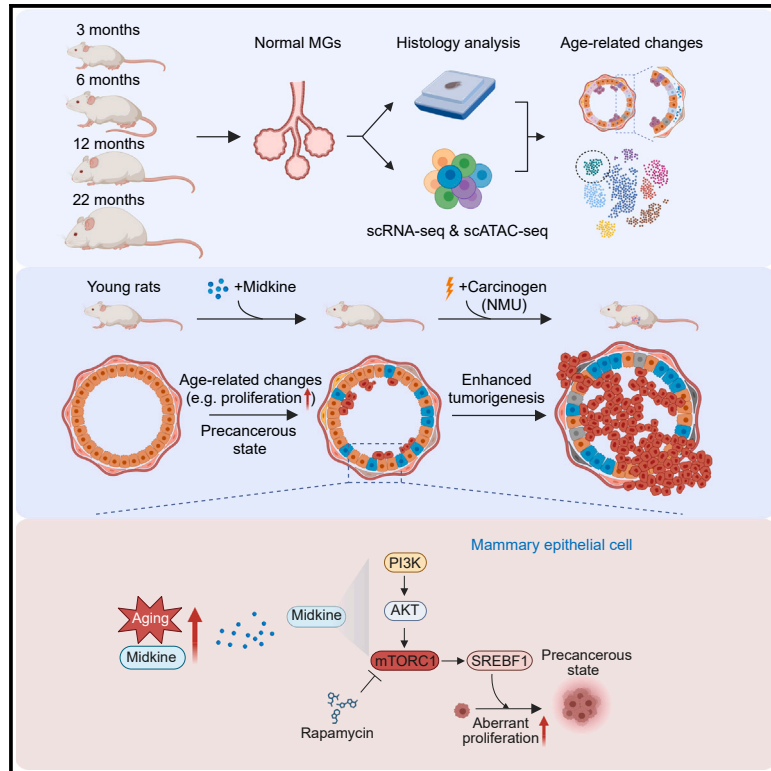Article

Midkine as a driver of age-related changes and increase in mammary tumorigenesis

Graphical abstract

Authors

Pengze Yan, Ernesto Rojas Jimenez, Zheqi Li, ..., So Yeon Park, Henry W. Long, Kornelia Polyak

Correspondence

kornelia_polyak@dfci.harvard.edu

In brief

Yan et al. reveal age-related changes in the rat mammary gland, identifying a unique aging-associated luminal progenitor population. Midkine treatment mimics the effects of aging and promotes mammary tumorigenesis. Elevated midkine levels are associated with increased breast cancer risk in younger women highlighting its role in age-related tumor initiation.

Highlights

- Cellular and molecular changes in the mammary gland with age

- Midkine is a key driver of aging-related changes and mammary tumorigenesis

- High midkine levels are associated with higher risk of breast cancer in younger women

CellPress

Article

Midkine as a driver of age-related changes and increase in mammary tumorigenesis

Pengze Yan,[1,2,3] Ernesto Rojas Jimenez,[1,2,3] Zheqi Li,[1,2,3] Triet Bui,[1,2,3] Marco Seehawer,[1,2,3] Jun Nishida,[1,2,3] Pierre Foidart,[1,2,3] Laura E. Stevens,[1,2,3] Yingtian Xie,[4] Miguel Munoz Gomez,[4] So Yeon Park,[5] Henry W. Long,[4] and Kornelia Polyak[1,2,3,4,5,6,7,*]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[2]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA
[3]Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA
[4]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA 02215, USA
[5]Department of Pathology, Seoul National University, Bundang Hospital, Seoul National University College of Medicine, Seongnam, Republic of Korea
[6]Harvard Stem Cell Institute, Cambridge, MA 02142, USA
[7]Lead contact
*Correspondence: kornelia_polyak@dfci.harvard.edu
https://doi.org/10.1016/j.ccell.2024.09.002

SUMMARY

Aging is a pivotal risk factor for cancer, yet the underlying mechanisms remain poorly defined. Here, we explore age-related changes in the rat mammary gland by single-cell multiomics. Our findings include increased epithelial proliferation, loss of luminal identity, and decreased naive B and T cells with age. We discover a luminal progenitor population unique to old rats with profiles reflecting precancerous changes and identify midkine (*Mdk*) as a gene upregulated with age and a regulator of age-related luminal progenitors. Midkine treatment of young rats mimics age-related changes via activating PI3K-AKT-SREBF1 pathway and promotes nitroso-N-methylurea-induced mammary tumorigenesis. Midkine levels increase with age in human blood and mammary epithelium, and higher *MDK* in normal breast tissue is associated with higher breast cancer risk in younger women. Our findings reveal a link between aging and susceptibility to tumor initiation and identify midkine as a mediator of age-dependent increase in breast tumorigenesis.

INTRODUCTION

Aging is a major risk factor for cancer as most tumors are diagnosed in individuals over 50 years of age.[1,2] Breast cancer is the most common cancer in women and the average age of diagnosis is 62,[3] yet studies involving aged animals remain limited and elderly patients are underrepresented in clinical trials. To better understand age-associated increase in breast tumor initiation and progression, we must dissect how aging impacts the normal mammary gland. Although histologic, mammographic, cellular, and molecular analysis of mammary tissues have already uncovered significant changes during aging in humans and mice,[4–11] their functional relevance in breast tumorigenesis has not been defined.

Normal mammary gland function requires tightly orchestrated interactions between epithelial and stromal cells[12] and perturbations of this contribute to tumorigenesis. The mammary epithelium is a bilayered ductal structure composed of luminal epithelial and basal/myoepithelial cells, and progenitor populations are considered the cell-of-origin in breast cancer.[7,13] Luminal epithelial cells, comprised of hormone-responsive luminal and secretory luminal progenitors, line the apical surface of the milk duct.[8,12] Encompassing the luminal cells, basal and myoepithe-

lial cells exhibit a dual nature of contractile muscle and epithelial characteristics.[8,14] Aging perturbs luminal-basal cell differentiation, leading to reduced epithelial lineage fidelity and the emergence of "intermediary" cells with mixed luminal and basal features, potentially increasing mammary tumorigenesis risk.[15–18] Nevertheless, the mechanisms by which aging perturbs epithelial differentiation and its impact on tumorigenesis remain unclear.

Rats are an ideal experimental model for aging and cancer, not only are they more physiologically similar to humans than mice,[19] but the nitroso-N-methylurea (NMU)-induced mammary tumor model stands as the only preclinical immunocompetent model for hormone-dependent estrogen receptor-positive (ER⁺) mammary tumors,[20,21] the most common subtype, especially in older postmenopausal women. NMU-induced rat mammary tumors resemble the histopathology and immune environment of human breast cancer,[22] and prior studies described that susceptibility to NMU-induced mammary tumors changes with age.[23]

Investigating the dynamics of naturally aging mammary glands is crucial toward understanding the complexities underlying age-associated increases in breast cancer susceptibility. Here, we describe age-associated changes in the histology and single-cell multiomic profiles of normal mammary glands from aged
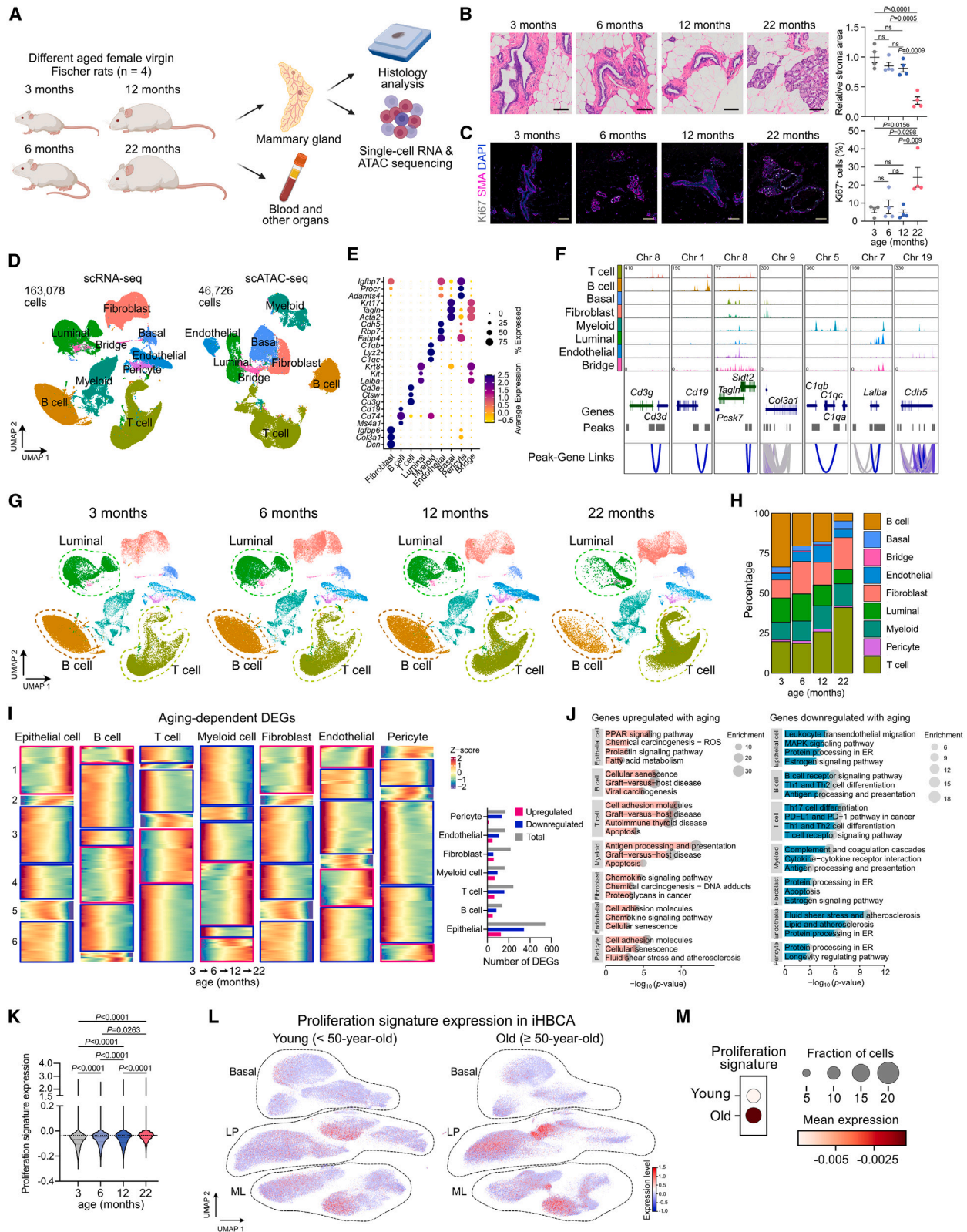
(legend on next page)

rats. We identified precancerous changes, including aberrant proliferation of epithelial cells, emergence of a unique luminal progenitor population, and decreasing proportion of naive B and T cells in the mammary glands. In follow up mechanistic studies, we discovered that midkine, a heparin-binding growth factor secreted by basal epithelial cells, is a mediator of an age-associated transcriptional shift and increased mammary epithelial cell proliferation. Using organoid cultures, we delineated that midkine induces mammary epithelial proliferation via the activation of the sterol regulatory element-binding transcription 1 (SREBF1) through the PI3K-AKT pathway. Furthermore, midkine treatment of young rats induced age-related changes and increased susceptibility to NMU-induced mammary tumors. Importantly, midkine levels increased with age in humans, both in the plasma and breast, and are higher in normal breast tissue of women with greater breast cancer risk, including *BRCA2* germline mutation carriers. Thus, our studies identify midkine as a clinically valuable marker of biological aging and cancer risk, as well as a promising therapeutic target for breast cancer prevention.

RESULTS

Single-cell profiles of the rat mammary gland at different ages

To investigate age-related changes in the cellular composition and molecular profiles of mammary tissues, we performed comprehensive characterization of mammary glands from virgin inbred Fischer 344 (F344) rats at ages 3, 6, 12, and 22 months (Figure 1A). Histologic analysis of hematoxylin-eosin-stained (H&E) sections demonstrated ductal structures in the 3-12-month age group and epithelial changes resembling gestational hyperplasia in 22-month-old rats accompanied by a gradual decrease of periductal stroma (Figure 1B). Immunofluorescence for the Ki67 proliferation marker, ER and progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) revealed significantly more Ki67$^+$ mammary epithelial cells in

22-month-old rats with a concomitant decrease in ER$^+$ and PR$^+$ cells and no change in HER2 (Figures 1C and S1A). Enzyme-linked immunosorbent assay (ELISA) for reproductive hormones in the peripheral blood showed significantly higher progesterone and prolactin in 6-month-old rats compared to other age groups, while estradiol levels gradually declined with age (Figure S1B).

Next, we performed single-cell transcriptome (scRNA-seq) and chromatin accessibility (scATAC-seq) sequencing on mammary glands from different aged rats to dissect gene expression and chromatin changes with age. We identified major cell populations using known cell-type-specific markers and performed unsupervised analyses including dimension reduction using uniform manifold approximation and projection (UMAP) as well as hierarchical clustering and gene ontology (GO) term enrichment (Figures 1D–1F and S1C-S1F, and Table S1). To assess age-related changes in cellular composition, we segregated the scRNA-seq UMAP data by age and quantified major cell clusters. The most notable changes included a pronounced decrease in B cells and transcriptional shifts in luminal epithelial and T cells with age (Figures 1G and 1H). Similar trends and increased open chromatin were observed in the scATAC-seq data (Figures S1G and S1H). To assess if these changes are clinically relevant, we re-analyzed the integrated human breast cell scRNA-seq data (integrated Human Breast Cell Atlas, iHBCA)[11] and similarly observed a transcriptomic shift in luminal cells and fewer B cells in normal breast tissues of women ≥50 years old compared to younger women (Figures S1I and S1J).

To identify differentially expressed genes (DEGs) among age groups, we aligned the data for each major cell type by age and classified the DEGs into 6 clusters based on their expression patterns (Figure 1I and Table S1). GO term analysis of upregulated genes in epithelial cells showed enrichment in PPAR and prolactin signaling, chemical carcinogenesis, and fatty acid metabolism (Figure 1J). Cellular senescence was enriched in aged B cells, while aged T cells and myeloid cells exhibited enrichment in apoptosis (Figure 1J). Aged fibroblast upregulated

---

Figure 1. Single-cell transcriptome and chromatin profiles of rat mammary glands at different ages

(A) Schematic overview of experimental design.

(B) Left, representative images of H&E-stained sections of mammary glands from rats with the indicated age. Scale bar, 100 μm. Right, quantification of stroma area. *n* = 4.

(C) Left, representative images of Ki67 and smooth muscle actin (SMA) immunofluorescence staining of mammary glands from rats with the indicated age. Scale bar, 50 μm. Right, quantification of Ki67$^+$ cells based on immunofluorescence. *n* = 4.

(D) UMAP plot depicting the major cell types of rat mammary glands based on scRNA-seq (left) and scATAC-seq (right) analysis.

(E) Dot plot showing the expression of known cell-type-specific markers in rat mammary tissue based on scRNA-seq data.

(F) Representative scATAC-seq peaks at genomic loci of known cell-type-specific markers.

(G) UMAP plots illustrating the dynamic changes of major cell clusters in mammary glands of different aged rats based on scRNA-seq data.

(H) Stacked bar plot showing the absolute percentage of each cell type in mammary glands of different aged rats based on scRNA-seq data. *n* = 4.

(I) Left, heatmaps showing the expression of aging-dependent DEGs in seven major cell types in different aged rats. Genes with increasing expression with age (upregulated DEGs) are highlighted in red, while those with decreasing expression (downregulated DEGs) are highlighted in blue. The numbers on the left side represent the order of gene clusters. Right, bar chart showing the number of DEGs.

(J) GO term analysis of upregulated (left) and downregulated (right) aging-dependent DEGs in each cell type. The dot size indicating the enrichment score of each term.

(K) Violin plot showing the expression of proliferation signature in the epithelial cells from different aged rats. The dashed line represents the mean expression value of the proliferation signature in the 22-month-old group.

(L) UMAP plot showing the expression of proliferation signature in breast epithelial cells from iHBCA split into young (*n* = 200) and old (*n* = 53) groups. Basal, basal epithelial cell; LP, luminal progenitor; ML, mature luminal.

(M) Dot plot showing the expression of proliferation signature in breast epithelial cells from iHBCA split into young and old groups.

Graphs (B and C) are presented as mean ± standard error of mean (SEM). *p* values were calculated by one-way ANOVA followed by Tukey's multiple comparisons test (B, C, and K). ns, not significant. Figure 1A was created with BioRender. See also Figures S1 and S2 and Table S1.
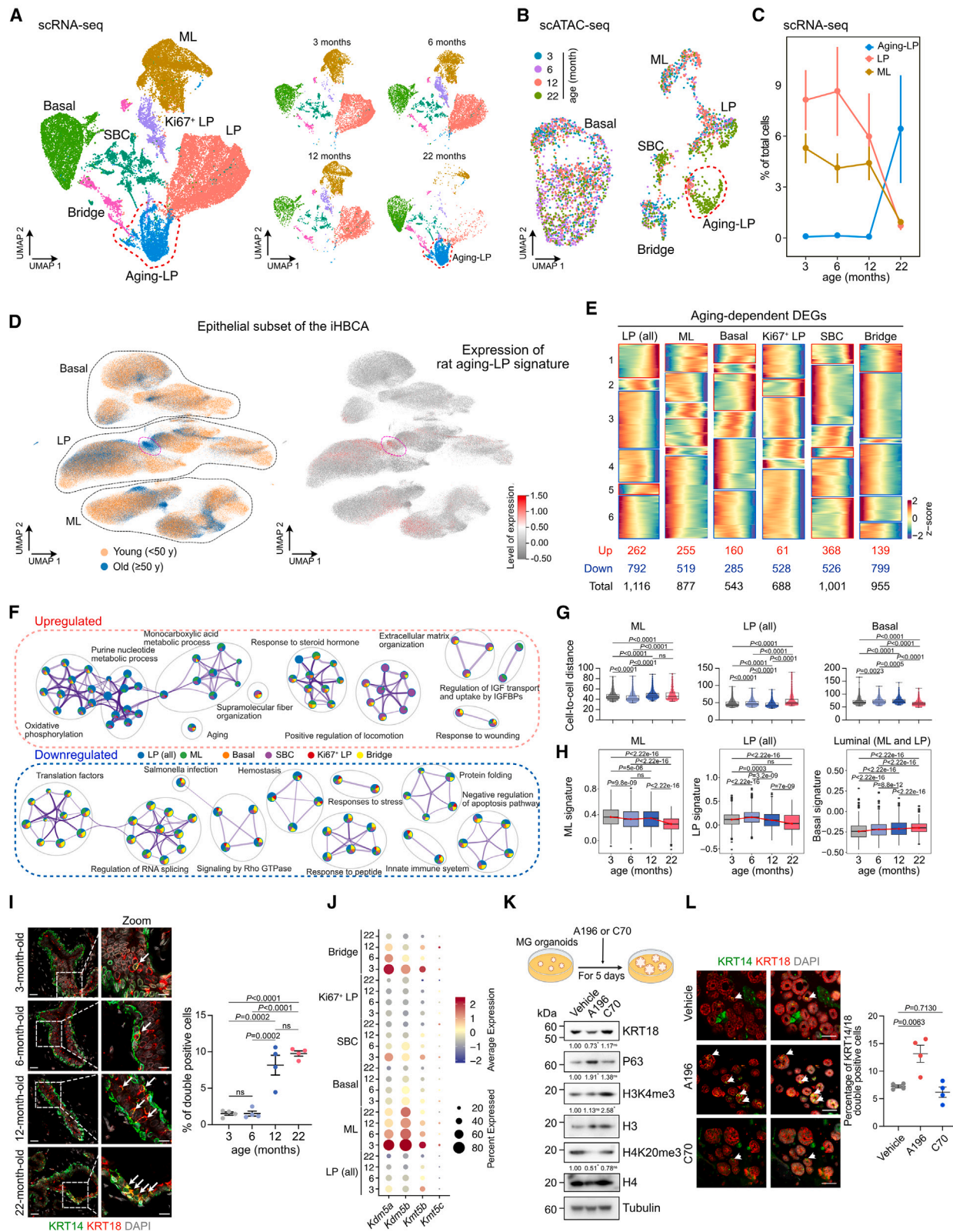
Figure 2. Age-related changes in the mammary epithelium

(A) UMAP plots colored by mammary epithelial cell subpopulations (left) or split by age (right) based on scRNA-seq data. Basal, basal epithelial cell; ML, mature luminal; LP, luminal progenitor; bridge, basal-luminal intermediate cell; SBC, secretory basal cell.

chemokine, chemical carcinogenesis, and proteoglycan in cancer pathways (Figure 1J), potentially promoting an inflammatory pro-tumorigenic environment.[24] Downregulated terms were associated with cellular functions including estrogen signaling in epithelial cells, B and T cell receptor signaling in B and T cells, respectively, underscoring the functional decline with age (Figure 1J). Proliferation markers were elevated in aged mammary epithelial cells in rat scRNA-seq data (Figure 1K) and also in luminal epithelial cells of older women (≥50 years) in the iHBCA dataset (Figures 1L and 1M).

Cellular and molecular changes in the microenvironment during aging

The microenvironment significantly impacts tumor initiation and progression,[25] and age-related alterations in stromal cells facilitate tumorigenesis. We assessed age-related changes in immune cells, fibroblasts, and vascular cells (Figure 1G). T and natural killer (NK) cells were classified into twelve distinct subclusters (Figures S2A and S2B). We observed an age-related loss of both CD4$^+$ and CD8$^+$ naive T cells and an accumulation of memory and effector T cells (Figure S2C). Pseudotime analysis of CD4$^+$ and CD8$^+$ T cell populations individually revealed a trajectory of transition from naive to central memory to effector states (Figures S2D and S2E), possibly due to age-associated decline of thymic activity.[26] B cells were subclustered into five subtypes, and naive B cells and overall B cell frequency declined in both mammary glands and bone marrow (Figures S2F–S2I), implying a systemic reduction in humoral immune responses during aging. Macrophages accounted for a large portion of the age-associated alterations in myeloid cells with an expansion of monocyte-derived macrophages. (Figures S2J–S2L). We validated age-related changes in immune cell populations by flow cytometry, confirming a decrease in total B cells, particularly naive B cells, a declining trend in both CD4$^+$ and CD8$^+$ naive T cells, and an increasing trend in monocyte-derived macrophages (Figure S2M). Lastly, we classified fibroblasts, endothelial cells, and pericytes into ten subclusters. While there were no obvious alterations in cellular composition (Figures S2N and S2O), we observed a decline in extracellular matrix gene expression (Figure S2P).

In summary, aging-related changes in the mammary microenvironment, including loss of naive T and B cells, expansion of monocyte-derived macrophages, and diminished collagen production, imply a shift to a more cancer-prone environment.

Changes in mammary epithelial cell transcriptomes with age

Mammary epithelial progenitors are considered the cell of origin of breast cancer,[7,13] therefore understanding the impact of age on normal epithelial cells is essential. Thus, we interrogated mammary epithelial cellular heterogeneity across ages, classifying them into seven subtypes: basal, mature luminal (ML), luminal progenitor (LP), proliferative Ki67$^+$ LP, basal-luminal intermediate cell (bridge), and secretory basal cells (SBCs) with stem cell features (Figure 2A and Table S2).[21] We also detected a distinct cluster positive for known LP markers specifically enriched in the oldest (22-month old) rats in both scRNA-seq and scATAC-seq data, termed aging-related LP (Aging-LP) (Figures 2A–2C and S3A–S3C). Furthermore, there was a marked decline in ML and normal LP cells with age (Figure 2C). We further analyzed the epithelial subset of the iHBCA dataset and confirmed an LP subcluster enriched in older women (≥50 years) with high expression of the rat Aging-LP signature (Figures 2D and S3D).

Next, we identified age-related DEGs in each epithelial subpopulation and observed variability in DEGs across cell types with only a few commonly enriched GO terms (Figures 2E, 2F, S3E, and S3F; Table S2). DEGs upregulated with age in each subcluster were enriched in oxidative phosphorylation, response to steroid hormones, and aging (Figure 2F). DEGs downregulated in older animals were associated with RNA processing and translation (Figure 2F and Table S2), implying an

---

(B) UMAP plots colored by age based on scATAC-seq data.

(C) Line chart showing the proportion of mammary epithelial cell subpopulations in different aged rats based on scRNA-seq data. $n = 4$.

(D) UMAP plots depicting the major epithelial subclusters from iHBCA colored by age (left) or the expression of our rat aging-LP signature (right).

(E) Heatmaps showing the expression of aging-dependent DEGs in six epithelial cell subtypes in different age rats (from 3 to 22 months). LPs includes both LPs and aging-LPs. Red and blue indicates genes with increasing (upregulated DEGs) and decreasing (downregulated DEGs) expression, respectively.

(F) Network visualizing representative GO terms and pathways of aging-dependent upregulated (top) and downregulated (bottom) DEGs in epithelial subclusters. LP (all) includes both LPs and aging-LPs. Nodes are pie charts colored by cell types, with slice size representing the fraction of genes from each list. Terms with >0.3 similarity are connected by lines.

(G) Violin plots showing cell-to-cell distances in the indicated cell populations.

(H) Boxplots showing the expression of ML signature (pct.1 ≥ 0.5, pct.2 ≤ 0.2) in ML, LP signature (pct.1 ≥ 0.5, pct.2 ≤ 0.5) in LP and basal signature (pct.1 ≥ 0.5, pct.2 ≤ 0.1) in luminal epithelial cells (LP and ML) from different aged rat. The red line represents the mean LP signature value across the different age groups, connecting the mean values for each group.

(I) Representative images of immunofluorescence staining for KRT14 and KRT18 epithelial cell markers in mammary glands from rats with the indicated age and graph depicting quantification of the KRT14 and KRT18 double-positive cells in each group. $n = 4$. Arrows indicate double-positive cells. Scale bar, 20 μm (left) and 10 μm (right).

(J) Dot plot showing the expression of H4K20 methyltransferase and H3K4 demethylase in epithelial subclusters from different aged rats based on scRNA-seq data.

(K) Western blot showing the expression of KRT18, P63, H3K4me3, total H3, H4K20me3, and total H4 in the organoids treated with A196 (5 μM) and C70 (10 μM) for 5 days $n = 3$. Tubulin was used as loading control. The relative contents of each target were labeled below the blots.

(L) Representative images of immunofluorescence staining for KRT14 and KRT18 in mammary organoids treated with A196 (5 μM) and C70 (10 μM) for 5 days and graph depicting the quantification of the KRT14 and KRT18 double-positive cells in each group. $n = 4$. Arrows indicate double-positive cells. Scale bar, 50 μm.

Graph (C) is presented as mean ± standard deviation (SD). Graphs (G and H): the boxes represent the upper 75% percentile (top), median (line) and lower 25% percentile (bottom), the whiskers range from minimum to maximum values and the bold dots represent outliers. Graphs (I and L) are presented as mean ± SEM. p values were calculated by one-way ANOVA followed by Tukey's multiple comparisons test (G and I), Kruskal–Wallis test (H) and Dunnett's multiple comparisons test (K and L). Figure 2K was created with BioRender. See also Figure S3, Table S2.
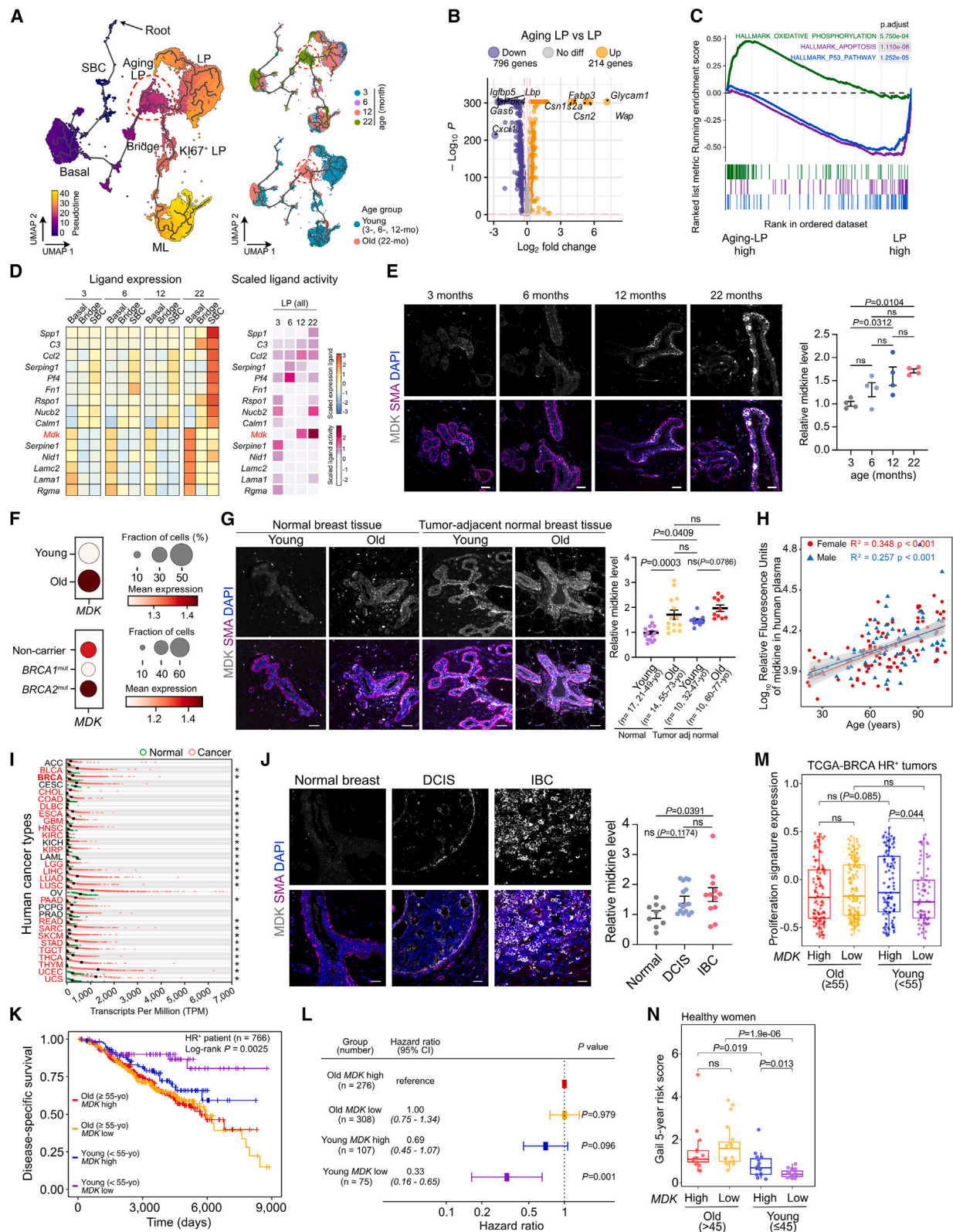
Figure 3. Trajectory and cell-cell communications analysis of mammary epithelial cells in different aged rats
(A) Pseudotemporal analysis of epithelial subclusters colored by pseudotime (left), age (top right) or age group (bottom right). Pseudotime is depicted from dark purple to light yellow. Basal, basal epithelial cell; ML, mature luminal; LP, luminal progenitor; Bridge, basal-luminal intermediate cell; SBC, secretory basal cell.

(legend continued on next page)

accumulation of dysfunctional proteins in the epithelial cells of aged rats, a hallmark of both aging and cancer.[27]

Increasing cellular transcriptomic heterogeneity and loss of lineage fidelity has been associated with aging,[28] and may contribute to tumor initiation. We calculated the cell-to-cell distances within each epithelial subpopulation and within each age group and found that cell-to-cell distance increased with aging in ML and LP cells but decreased in basal cells (Figure 2G). In addition, the expression of ML and LP signatures in the corresponding cell types decreased with age, while basal signatures increased in luminal epithelial cells (Figure 2H). Immunofluorescence staining for luminal (KRT18) and basal (KRT14) cell-specific markers confirmed their mutually exclusive expression in young (3-month old) rats and progressively increasing colocalization with age (Figure 2I), consistent with findings in human breast.[17,18,29] These results demonstrate an age-related increase in luminal epithelial cellular heterogeneity and declining lineage fidelity, reflecting perturbed epigenetic programs that may increase breast cancer risk.

We and others previously described that cellular transcriptomic heterogeneity and luminal identity are regulated by KDM5B histone H3 lysine 4 trimethyl (H3K4me3) demethylase activity,[30,31] and that H4K20me3 is one of the most variable histone modifications in triple-negative breast cancer (TNBC), linking cellular differentiation, senescence, aging, and cancer.[32] We first analyzed the expression of histone demethylases and methyltransferases regulating these histone marks in our scRNA-seq data. We found that expression of both H3K4me3 demethylases (*Kdm5a* and *Kdm5b*) and H4K20me3 methyltransferases (*Kmt5b* and *Kmt5c*) decreased in epithelial cells with age, especially in ML and bridge cells (Figure 2J). Correspondingly, H3K4me3

increased and H4K20me3 decreased with age with the most dramatic drop in H4K20me3 in the oldest group (Figures S3G and S3H). Treatment of young rat mammary organoids with KDM5 (C70) or KMT5B/5C (A196) inhibitors revealed decreased KRT18 luminal and increased P63 basal protein levels in A196-treated organoids, while C70 had no significant effect (Figure 2K). Immunofluorescence showed an increase in double-positive cells for KRT14 and KRT18 after A196 treatment (Figure 2L). These findings suggest that decreased KMT5B/5C activity and H4K20me3 levels might contribute to perturbed luminal differentiation and decreased mammary epithelial lineage fidelity during aging.

Alterations in mammary epithelial cell hierarchy and cell-to-cell communication with aging

To investigate how aging perturbs mammary epithelial differentiation, we performed Monocle3[33] analysis on our scRNA-seq data using the stem cell-like SBCs as starting point.[21] The pseudotime trajectory extended from SBCs to basal cells and continued through bridge cells to aging-LPs, then to LPs and to MLs via proliferating Ki67[+] LPs (Figure 3A). Our trajectory analysis placed aging-LPs before LPs, implying that aging-LPs may reflect less differentiated LPs (Figure 3A). Given that LPs are considered the cell-of-origin of most breast cancers regardless of subtype,[34] we explored molecular differences between aging-LPs and LPs. We identified a significant number of DEGs between the two distinct LP populations with several genes encoding milk proteins (e.g., *Wap* and *Csn2*) being top upregulated in aging-LPs (Figure 3B and Table S3), implying a shift to alveolar state. Gene set enrichment analysis (GSEA) unveiled genes downregulated in aging-LPs were associated with P53 and

---

(B) Volcano plot showing the DEGs between aging-LP and normal LP. Yellow dots represent upregulated genes in midkine treated group (log$_2$ fold change ≥ 0.25 and adjust p value ≤ 0.05). Purple dots represent downregulated genes in midkine treated group (log$_2$ fold change ≤ 0.25 and adjust p value ≤ 0.05).

(C) Gene set enrichment analysis (GSEA) showing the top enriched and depleted hallmarks in aging-LPs. p values were adjusted by the Benjamini-Hochberg (BH) method.

(D) Nichenet analysis showing the top 15 ligands from basal, SBC, and bridge to LP. Left, heatmap showing the ligand expression. Right, heatmap showing ligand activity predicted based on ligands and corresponding target gene expression.

(E) Left, representative images of immunofluorescence staining for MDK and SMA in mammary glands of different aged rat. Scale bar, 20 μm. Right, quantification of MDK expression in epithelial cells of different age rats. $n = 4$.

(F) Dot plot showing the expression of *MDK* in in breast epithelial cells from iHBCA split by age group (top) and germline mutation status (bottom).

(G) Left, representative images of immunofluorescence staining for MDK and SMA in normal human breast tissues from young (21–49 years old, $n = 17$), old (55–73 years old, $n = 14$) women and tumor adjacent normal breast tissues from young (32–47 years old, $n = 10$) and old (60–77 years old, $n = 10$) women. Scale bar, 50 μm. Right, quantification of MDK expression in epithelial cells of normal and tumor adjacent normal breast tissues.

(H) Scatterplot showing the Pearson correlation of MDK concentration in normal human plasma with age ($n = 171$). The proteome data of plasma from different aged humans were obtained from a previous study.[40]

(I) Dot plot showing the expression of *MDK* in various human tumor types (data from GEPIA, http://gepia.cancer-pku.cn). *, $p < 0.05$. Cancer names colored as red exhibit significantly higher *MDK* expression compared to their paired normal tissues.

(J) Left, representative images of immunofluorescence staining for MDK and SMA in normal ($n = 9$), DCIS ($n = 15$), and IBC ($n = 12$) patient samples. Scale bar, 20 μm. Right, quantification of MDK expression in the indicated groups.

(K) Kaplan-Meier plot depicting DSS of hormone receptor (HR)[+] breast cancer patients in METABRIC dataset. Patients were classified based on tumor *MDK* expression into high versus low split by upper quartile and patient age at diagnose (cutoff 55-year-old). $n = 766$. Hazard ratio, 95% CI and p value was calculated with log rank tests.

(L) Forest plot of showing hazard ratio and 95% confidence intervals (CIs) modeling contributions of the *MDK* expression and patient age at diagnosis to disease-specific survival in a Cox proportional hazards regression in METABRIC HR[+] breast cancers.

(M) Boxplot showing the expression of proliferation signature score of HR[+] breast cancer patients from TCGA dataset, which were segregated based on *MDK* expression into high versus low split by upper quartile and patient age at diagnose (cutoff 55-year old). $n = 446$.

(N) Boxplot showing the Gail 5-year risk scores of healthy women, which were segregated based on *MDK* expression into high versus low split by upper quartile and patient age at diagnose (cutoff 45-year-old). $n = 66$. This transcriptome data were obtained from a previous study.[48]

Graphs (E, G, and J) are presented as mean ± SEM. Graphs (M and N): the boxes represent the upper 75% percentile (top), median (line) and lower 25% percentile (bottom) and the whiskers range from minimum to maximum values. p values were calculated by one-way ANOVA Tukey's multiple comparisons test (E, G, and J), Pearson correlation test (H), Wilcoxon signed-rank test (I), Log rank test (K and L) and Kruskal-Wallis test (M and N). See also Figure S3 and Table S3.

apoptosis pathways, whereas upregulated genes were related to oxidative phosphorylation (Figure 3C). These findings suggest that aging-LPs are less differentiated, have diminished activity of tumor suppressive pathways, and are metabolically more active than LPs.

Luminal-basal cell interactions are critical for normal mammary gland development and function.[35,36] To assess age-related changes in cell-to-cell communication, we performed interactome analysis on all epithelial subtypes in our scRNA-seq data. Basal, SBC, and bridge cells were the major senders of ligands, whereas luminal epithelial cells (including LP and ML), basal and bridge cells were the major targets (Figure S3I). To explore mechanisms underlying the age-related transcriptional shift of LPs, we analyzed differential ligands and receptors between basal, SBC, and bridge cells and LPs at different ages using NicheNet.[37] We found that MDK was among the top differential ligands expressed by basal cells and had the most notable age-related increase in expression and activity (Figure 3D).

Identification of midkine as a biomarker of age and breast cancer

Mdk encodes midkine, a heparin-binding growth factor overexpressed in various tumors.[38] Midkine is highly elevated in the blood of breast cancer patients compared to healthy controls and further increases with disease progression.[39] We further investigated midkine expression across different ages and found an increase in midkine at both mRNA and protein levels with aging (Figures 3E and S3J). Our scATAC-seq data also showed an age-related increase in chromatin accessibility at the Mdk locus (Figure S3K). MDK expression was also significantly higher in normal breast epithelial cells from older ($\geq$50-year old) women and BRCA2 germline mutation carriers in the iHBCA dataset (Figure 3F).[11] To validate the scRNA-seq data, we performed immunofluorescence for MDK in normal breast tissues from healthy women and detected elevated expression in older (55-73-year old) versus younger (21-49-year old) women (Figure 3G). We also analyzed tumor-adjacent normal tissue from breast cancer patients and found higher expression of midkine compared to normal breast from healthy women in all age groups (Figure 3G), indicating that midkine could be a biomarker of breast cancer risk. Furthermore, plasma MDK levels in published proteomics datasets[40] showed significant positive correlation with age in healthy human individuals regardless of sex (Figure 3H).

Next, we analyzed MDK expression in tumors by comparing mRNA levels in normal and corresponding tumor tissues in the TCGA and GTEx datasets.[41,42] We found that MDK had significantly higher expression in 24 out of 31 tumor types including breast cancer (Figure 3I). To further investigate the expression of MDK during human breast tumor progression, we performed immunofluorescence staining for MDK in normal breast tissue, ductal carcinoma in situ (DCIS), and invasive breast cancer (IBC), and found higher MDK levels in DCIS and IBC compared to normal (Figure 3J).

To determine if MDK expression correlates with clinical outcome in breast cancer patients, we analyzed associations between MDK levels and disease-specific survival (DSS) of patients in hormone receptor positive (HR$^+$) breast cancer from Molecular Taxonomy of Breast Cancer International Consortium (METABRIC).[43] Due to the associations of age with survival and MDK expression, we divided patients into younger and older groups using age 55 as a cutoff. We found that older patients had shorter DSS regardless of MDK levels, while in younger patients, high MDK expression was associated with significantly shorter DSS (Figures 3K and 3L). In the TCGA-BRCA cohort, we correlated MDK expression with clinicopathologic features in HR$^+$ tumors, since DSS data were unavailable. We found that younger patients (<55 years old) with high MDK expression had a higher proliferation signature than those with low MDK, whereas no difference was observed in older patients ($\geq$55 years old) (Figure 3M).

To understand why high tumor MDK levels were associated with shorter survival in younger patients, we analyzed correlations between survival and signaling pathways in HR$^+$ tumors from younger and older patients. We found that the activity of hypoxia, PI3K-AKT-mTOR, and reactive oxygen species pathways was positively correlated with poor prognosis only in younger patients (Figure S3L). Since MDK can activate PI3K-AKT-mTOR signaling[38,44,45] and augment oxidative stress,[46] the activation of these pathways in tumors from younger women might explain why MDK levels are prognostic in younger but not in older patients.

Lastly, we evaluated if midkine expression in normal breast tissue of healthy women could predict risk of breast cancer by analyzing associations between MDK expression and Gail 5-year score,[47] a currently used predictor of breast cancer risk, in a previously reported RNA-seq dataset.[48] We found that younger ($\leq$45 years old) women with higher MDK levels had a significantly higher Gail 5-year risk score than those with lower MDK expression, and older age (>45 years old) was associated with higher Gail 5-year risk score regardless of MDK (Figure 3N).

Altogether, these data suggest MDK as a candidate biomarker of aging and a predictor of breast cancer risk and clinical outcomes in younger women.

Midkine treatment mimics aging-related changes in the mammary gland

Next, to test the functional relevance of midkine in age-related changes in the mammary gland, we treated young (3-4-week-old) rats with recombinant midkine for 4 weeks, collected tissues and performed histology analysis and scRNA-seq (Figure 4A). We analyzed 25,932 single-cell transcriptomes that grouped into eight major clusters based on known cell-type-specific marker genes (Figures 4B and S4A). The most noticeable differences between vehicle and midkine-treated tissues included an increase in epithelial cell fraction and a decrease in the relative proportion of B and naive T cells (Figures 4C and S4B–S4D). Within T cells, the populations of CD4$^+$ and CD8$^+$ naive T cells declined following midkine treatment, and the number of B cells also decreased, while both tissue-resident and monocyte-derived macrophages exhibited a slight increase (Figures S4C–S4E). Intriguingly, these MDK-induced changes resembled the age-related alterations we identified.

We further analyzed the pronounced MDK-induced changes in epithelial cells by subclustering into LP, ML, basal, and SBC, which revealed transcriptomic shifts in both luminal (comprising LP and ML) and basal epithelial cells (Figures 4D, S4F, and S4G). Trajectory analysis showed earlier pseudotime of midkine-treated LPs compared with vehicle rendering them more similar
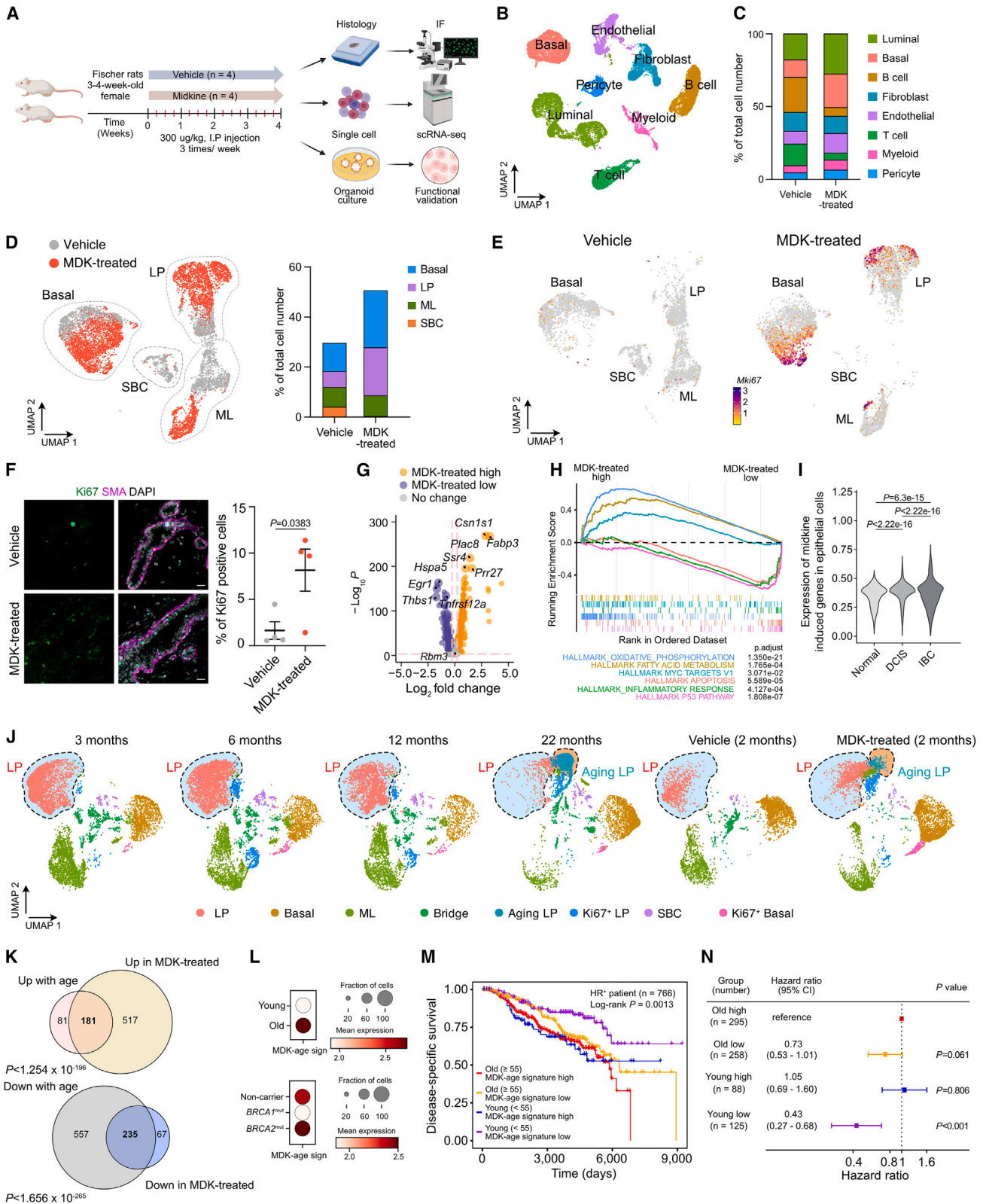
Figure 4. The effect of midkine on the mammary glands

(A) Overview of experiment designed to test the effect of midkine in rats.

(B) UMAP plot showing the major cell types in the mammary glands of control vehicle and midkine-treated rats.

(legend continued on next page)

to basal cells and mimicking the differences between aging-LP and LPs (Figure S4H). We also noted a significant increase in Ki67$^+$ epithelial cells in midkine-treated mammary epithelium again resembling the increased epithelial proliferation observed in aged mammary glands (Figure 1C and 4E–4F).

Since LPs were a major target cell for midkine based on our cell-to-cell communication analysis, we further explored their transcriptomic differences between vehicle and midkine-treated rats, revealing 698 significantly up and 302 significantly downregulated genes (Figure 4G and Table S4). GSEA demonstrated significant enrichment for metabolic pathways related to oxidative phosphorylation, fatty acid metabolism, and MYC targets v1 signaling in upregulated DEGs consistent with highly proliferative states. Whereas downregulated DEGs were mainly enriched in hallmarks of apoptosis, inflammatory response, and P53 pathway, implying muted tumor suppression (Figure 4H). GSEA of DEGs in ML and basal cells yielded similar results, suggesting a general midkine effect (Figures S4I–S4L and Table S4). We also tested the expression of these midkine-induced genes in our scRNA-seq dataset from normal, DCIS, and IBC tissue, and observed a significant progressive increase during tumor progression (Figure 4I).

To assess how midkine treatment mimics aging-related changes in LPs, we integrated all our epithelial scRNA-seq data, including data from rats of different ages and midkine-treated rats. UMAP and clustering analysis revealed a remarkable similarity between age and midkine treatment, with aging-LP cells enriched in both 22-month-old rats and midkine-treated 2-month-old rats (Figure 4J). To identify genes altered by both midkine treatment and age in LPs, we analyzed the overlap of DEGs and found that 181/262 (69%) genes upregulated with age overlapped with genes upregulated by midkine (defined as "MDK-age signature"), and 235/792 (30%) downregulated genes with age overlapped with genes downregulated in midkine-treated group (Figure 4K and Table S5). Similar to MDK itself, the expression of the MDK-age signature was also higher in LPs from older (≥50-year old) compared to younger

(<50-year old) women and from *BRCA2* germline mutation carriers relative to non-carriers in the iHBCA dataset (Figure 4L).

Lastly, we evaluated whether the "MDK-age signature" can predict clinical outcome in breast cancer patients using the METABRIC cohort. HR$^+$ breast cancer patients were stratified into four groups based on age (young and old, using age 55 as a cutoff) and signature enrichment (MDK-age high and low). We found that among young patients, lower expression of the "MDK-age" signature was associated with longer DSS, but there were no differences in DSS in older patients between high and low MDK-age signature groups (Figures 4M and 4N). Furthermore, the MDK-age signature better separated patients into outcome groups than *MDK* expression alone (Figure 3K).

Overall, our rat treatment data show that midkine is a major driver of aging-related changes in the mammary epithelium and that genes commonly upregulated by both midkine and aging have prognostic value in younger HR$^+$ breast cancer patients.

An MDK-SREBF1 interaction network orchestrates aging-related mammary epithelial changes

To identify regulators of aging-related changes in LPs, we constructed gene regulatory networks of transcription factors and their target genes using single-cell regulatory network inference and clustering (SCENIC).[49] We identified a series of transcriptional factors (TFs), including *Mlx*, *Creb3l1*, *Atf6b*, *Xbp1*, and *Srebf1*, with binding motifs enriched in the *cis*-regulatory elements of co-expressed genes in aging-LPs (Figure 5A). By analyzing the expression of the top 10 enriched TFs in LPs from different aged and midkine-treated rats, only *Srebf1* was consistently expressed and upregulated and acted as a top regulatory hub (red dot) with the highest number of target genes (pink dots), including the majority of the 181-gene MDK-age signature (Figures 5B–5D and S5A). Additionally, we analyzed SREBF1 TF activity in LPs in our scATAC-seq rat data and observed a gradual increase with aging and higher activity in aging-LP compared to LP (Figures 5E and 5F).

(C) Stacked bar plot showing the absolute percentage of each cell type in mammary glands from vehicle and midkine-treated rats (*n* = 2).

(D) Left, UMAP showing the distribution of epithelial subclusters colored by treatment. Right, stacked bar plot showing the percentage of epithelial subclusters in mammary glands from vehicle and midkine-treated rats (*n* = 2). Basal, basal epithelial cell; ML, mature luminal; LP, luminal progenitor; SBC, secretory basal cell.

(E) UMAP plot showing the expression of *Mki67* in mammary epithelial cells split by treatment.

(F) Left, representative images of immunofluorescence staining for Ki67 and SMA in mammary glands of vehicle or midkine-treated rats. Scale bar, 20 μm. Right, bar chart shows the percentage of Ki67 positive epithelial cells. *n* = 4.

(G) Volcano plot showing the DEGs between luminal progenitors from vehicle and midkine treated rats. Yellow dots represent the upregulated genes in midkine treated group (log$_2$ fold change ≥0.25 and adjust *p* value ≤0.05). Purple dots represent downregulated genes in midkine-treated group (log$_2$ fold change ≤0.25 and adjust *p* value ≤0.05).

(H) GSEA showing the significantly enriched and depleted hallmarks of LP DEGs between midkine and vehicle group. *p* values were adjusted by the BH method.

(I) Violin plot showing the expression of significantly upregulated genes in midkine treated LP in normal, DCIS, and IBC patient samples.

(J) UMAP plots showing mammary epithelial subclusters from 3-, 6-, 12-, 22-month old, and vehicle and midkine-treated rats of 2-month old.

(K) Venn diagram showing the overlap of genes between DEGs of LP from midkine-treated rats and age-dependent DEGs. Top, upregulated genes, bottom, downregulated genes.

(L) Dot plot showing the expression of MDK-age signature (sign) in breast epithelial cells from iHBCA split by age group (top) and germline mutation status (bottom).

(M) Kaplan-Meier plot depicting DSS of hormone receptor (HR)$^+$ breast cancer patients in the METABRIC dataset. Patients were classified based on MDK-age signature expression into high versus low split by upper quartile and patient age at diagnosis (cutoff 55-year old). *n* = 766. Hazard ratio, 95% CI and *p* value was calculated with log rank tests.

(N) Forest plot of showing hazard ratio and 95% CIs modeling contributions of the MDK-age signature expression and patient age at diagnosis to DSS in a Cox proportional hazards regression in METABRIC HR$^+$ breast cancer cohort.

Graph (F) is presented as mean ± SEM. *p* values were calculated by unpaired two-tailed t test (F), Kruskal Wallis test (I), hypergeometric test (K) and Log rank test (M and N). Figure 4A was created with BioRender. See also Figure S4, Tables S4 and S5.
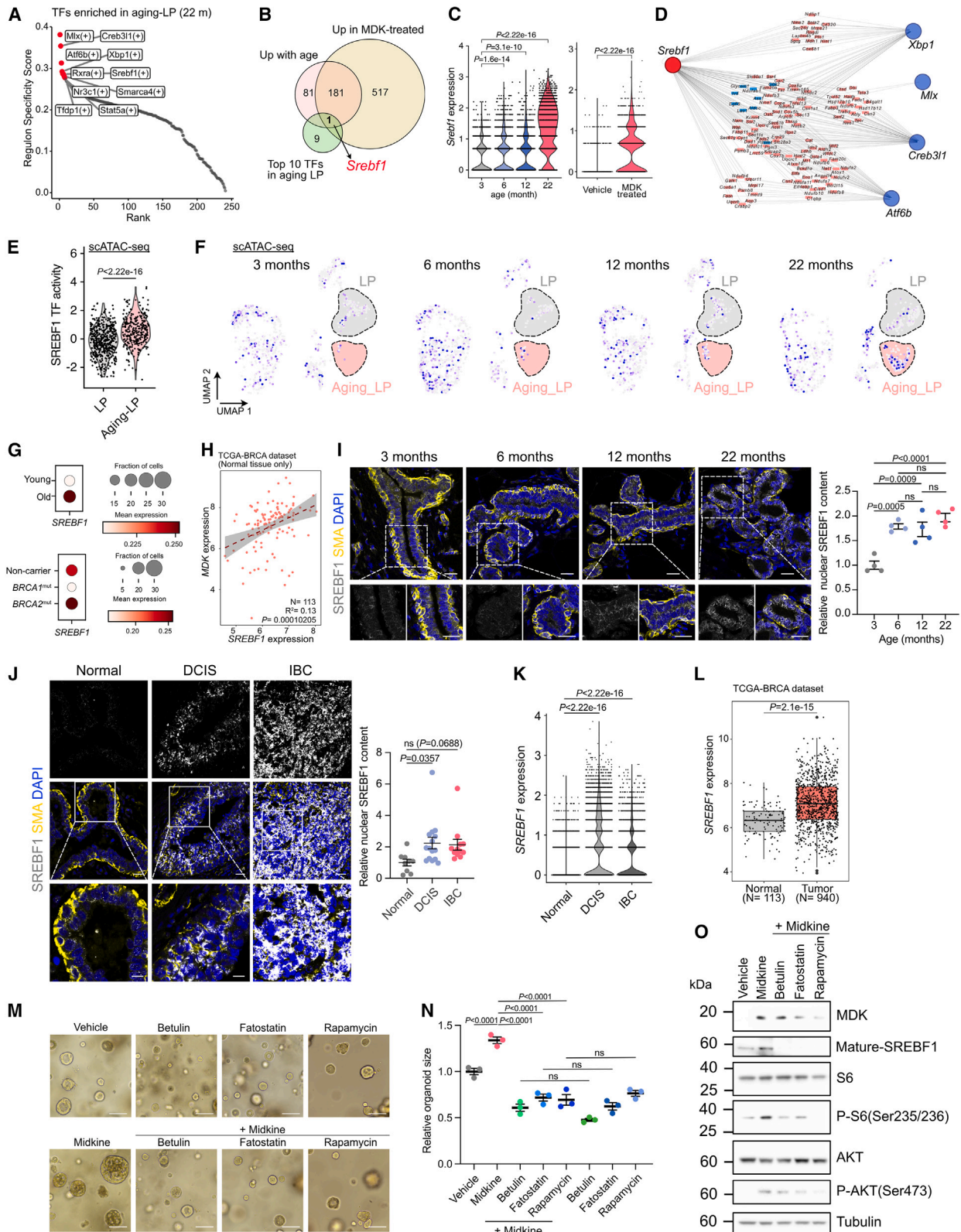
(legend on next page)

Notably, *SREBF1* was also more highly expressed in LPs from older (≥50-year old) versus younger (<50-year old) women and from *BRCA2* germline mutation carriers in the iHBCA dataset (Figure 5G). Furthermore, by exploring public RNA-seq datasets of normal human breast tissues we found a positive correlation between *SREBF1* and *MDK* expression (Figure 5H), suggesting that *Srebf1* might be a downstream mediator of midkine-induced aging-related changes in LPs and confirming the human relevance of the MDK-SREBF1 link.

Srebf1 is a proteolytically cleaved transcription factor whose active form translocates to the nucleus to regulate transcription.[50] Immunofluorescence for SREBF1 in rat mammary tissue showed increasing nuclear SREBF1 protein levels with age (Figure 5I). Immunofluorescence for SREBF1 in human breast showed increased nuclear localization in DCIS and IBC compared to normal non-cancerous breast tissues (Figure 5J). Consistently, *SREBF1* expression was higher in breast epithelial cells from DCIS and IBC compared to normal in our scRNA-seq data (Figure 5K) and was also increased in IBC compared to normal in the TCGA cohort (Figure 5L).

Next, we tested whether SREBF1 is a mediator of midkine-induced aging-associated alterations using a rat mammary organoid model. Mammary epithelial organoids were derived from young Fischer rats[8] and subjected to midkine treatment, with and without inhibitors blocking SREBF1 maturation (betulin and fatostatin),[51,52] and an mTOR pathway inhibitor (rapamycin) given the reported involvement of both midkine and SREBF1 in the PI3K-AKT-mTOR pathway.[53] Organoids treated with midkine were significantly larger compared to vehicle-treated, mirroring the growth promoting effects we observed *in vivo* (Figures 5M and 5N). Notably, this midkine-induced proliferation was blocked by both SREBF1 inhibitors and rapamycin treatment

(Figures 5M and 5N). Immunoblot analysis revealed that midkine treatment activated the AKT pathway, and upregulation of SREBF1 was blocked by rapamycin treatment (Figure 5O). Treatment with betulin and fatostatin also dampened the midkine-induced activation of the AKT pathway and rapamycin reduced MDK protein levels (Figure 5O). We also found that rapamycin and SREBF1 inhibitors reduced MDK at both mRNA and protein levels in the LA7 rat mammary tumor cell line (Figure S5B) and mammary organoids from both young and old rats (Figure S5C), implying a potential MDK-PI3K-AKT-mTOR-SREBF1 signaling loop.

Altogether, these data indicate that midkine-induced mammary epithelial proliferation is mediated by SREBF1 via the PI3K-AKT-mTOR pathway.

Midkine treatment promotes mammary tumor initiation

To investigate whether increasing levels of midkine with age might contribute to aging-related increased cancer risk, we tested the effect of midkine treatment on NMU-induced mammary tumorigenesis. We treated 3-4-week-old female virgin Fischer rats with midkine for two weeks, performed a single NMU injection at day 14, followed by an additional 2-week midkine treatment and a second NMU injection at day 28 (Figure 6A). Animals were sacrificed either when reaching endpoint or when the last rat in the midkine-treated group was euthanized. We found that rats treated with midkine developed more palpable mammary tumors with shorter latency than vehicle-treated controls (Figures 6A, 6B, and S6A). Tumors in the midkine-treated group had faster growth rates, leading to shorter survival compared to vehicle (Figures 6C and 6D). Tumors in the midkine-treated group had significantly higher tumor weight and volume at endpoint compared to control (Figure 6E). The smaller

---

Figure 5. SREBF1 is master transcriptional factor mediating age-related shift in luminal progenitors

(A) Regulon specificity score plot showing the core transcriptional factors regulating the transcriptome of aging-LP.

(B) Venn diagram showing the overlap of upregulated genes among DEGs of LP from midkine-treated rats and age-dependent DEGs as well as top 10 aging-LP enriched transcriptional factors.

(C) Violin plots showing the expression of *Srebf1* in LPs (all) from different age rats (left) and midkine-treated rats (right).

(D) Network visualization of potential transcriptional regulation in aging-LP. Red nodes represent potential transcriptional factors. Pink and blue nodes represent SREBF1 targets and non-targets, respectively.

(E) Violin plot showing the SREBF1 activity in aging-LP and normal LP based on scATAC-seq data.

(F) UMAP showing the SREBF1 transcriptional activity in epithelial subclusters from different aged rats. Aging LPs were highlighted with pink, LPs were highlighted with gray.

(G) Dot plot showing the expression of *SREBF1* in breast epithelial cells from iHBCA split by age group (top) and germline mutation status (bottom).

(H) Scatterplot showing the linear regression Pearson correlation of *Mdk* and *Srebf1* expression in normal tumor-adjacent breast tissues from breast cancer patients in the TCGA dataset.

(I) Representative images of immunofluorescence staining for SREBF1 and SMA in mammary glands of different aged rat (left) and plot depicting quantification of nuclear SREBF1 expression in mammary epithelial cells (right). Scale bar, 20 μm. $n = 4$.

(J) Representative images of immunofluorescence staining for SREBF1 and SMA in normal ($n = 9$), DCIS ($n = 15$), and IBC ($n = 12$) patient samples (left) and plot depicting quantification of nuclear SREBF1 expression in the indicated groups (right). Scale bar, 20 μm.

(K) Violin plot showing the expression of *Srebf1* in normal, DCIS, and IBC patient samples.

(L) Boxplot showing the expression of *Srebf1* in normal breast tissues and breast tumors based on TCGA dataset. Box and whisker plot, quartiles. Normal, $n = 113$; tumor, $n = 940$.

(M) Representative images of rat mammary organoids treated with midkine (300 ng/mL), betulin (5 μM), fatostatin (5 μM), and rapamycin (25 nM) for 5 days. Scale bar, 100 μm.

(N) Bar chart showing the quantification of organoid size in each treatment group. $n = 3$.

(O) Immunoblot showing the expression of MDK, mature SREBF1 (SREBF1 (M)), S6, p-S6(Ser235/236), AKT, p-AKT(Ser473) in the organoids treated with Midkine (300 ng/mL), betulin (5 μM), fatostatin (5 μM), and rapamycin (25 nM) for 5 days. Tubulin serves as loading control.

Graphs (I, J, and N) are presented as mean ± SEM. Graph (E) shows all points. Graph (L): the boxes represent the upper 75% percentile (top), median (line), and lower 25% percentile (bottom), the whiskers range from minimum to maximum values and the bold dots represent outliers. p values were calculated by Kruskal Wallis test (C [left] and K), Wilcoxon signed-rank test (C [right], E and L), Pearson correlation test (H), one-way ANOVA Tukey's test (I), one-way ANOVA Dunnett's test (J), and one-way ANOVA Sidak tests (N). See also Figure S5.
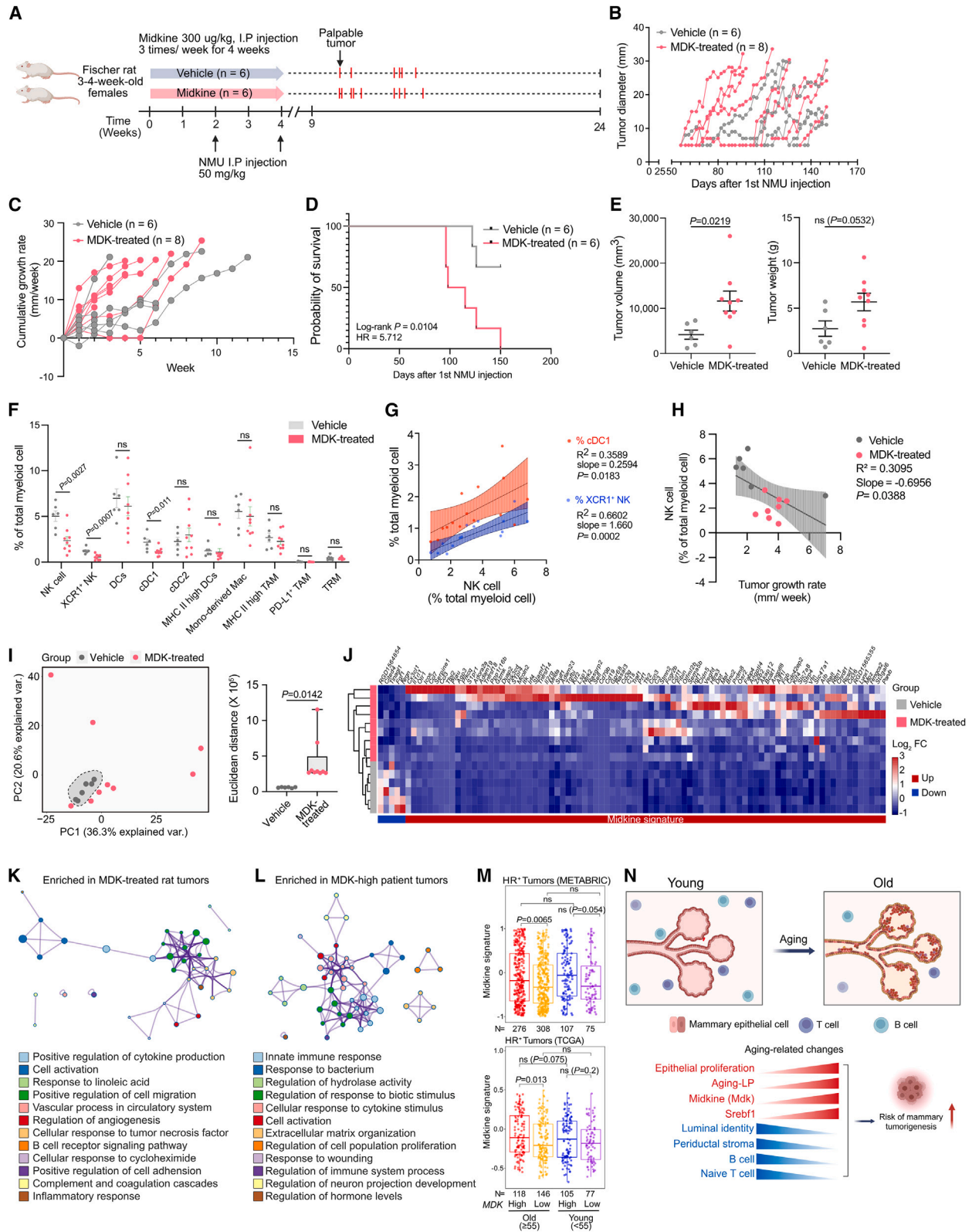
Figure 6. The effect of midkine on mammary tumorigenesis

(A) Overview of experiment designed to test the effect of midkine on mammary tumorigenesis.

(B) Spaghetti plot depicting the growth of mammary tumors from control vehicle and midkine-treated groups. Vehicle, *n* = 6. MDK-treated, *n* = 8.

(legend continued on next page)

size of some tumors was due to NMU treatment inducing multiple tumors per rat, requiring euthanasia when any tumor reached the maximum allowed size. Histologic analysis did not reveal noticeable differences in tumor cellularity, nuclear size, circularity, and receptor expression (ER, PR, and HER2), but we did observe an increasing trend in the number of Ki67$^+$ cells in tumors from the midkine group (Figures S6B–S6D). Furthermore, we detected a microscopic tumor in the normal-appearing mammary gland of an MDK-treated rat (Figure S6E).

To characterize the immune environment of mammary tumors from vehicle and MDK-treated animals, we performed polychromatic flow cytometry. We observed significantly fewer NK cells (including activated XCR1$^+$ NK cells) and conventional type 1 dendritic cells (cDC1) in tumors from the midkine-treated group, although there were no obvious differences in overall leukocyte composition (Figures 6F and S6F). The relative fraction of NK cells positively correlated with XCR1$^+$ NK cells and cDC1 (Figure 6G), consistent with prior reports suggesting a role for NK cells in recruiting cDC1 and promoting anti-tumor immune responses.[54] Multivariate correlation analysis highlighted that higher NK and CD8$^+$ T cell proportions, coupled with lower Treg levels, were associated with slower tumor growth rates (Figure S6G). Specifically, simple linear regression analysis underscored NK cell proportion as significantly negatively correlated with tumor growth rate, indicating a potential impact on tumor growth (Figures 6H and S6H).

To dissect transcriptomic differences, we performed bulk RNA-seq on tumors from vehicle ($n = 6$) and midkine ($n = 9$) treated groups. Principal-component and sample-to-sample correlation analyses of RNA-seq data demonstrated that a subset of tumors in the midkine group was different from vehicle (Figures 6I and S6I). Additionally, tumors from midkine-treated rats exhibited higher Euclidean distances compared to vehicle (Figure 6I), indicating greater heterogeneity. DEG analysis revealed higher expression inflammation-related genes, such as chemokines (e.g., *Ccl3*, *Ccl7*, and *Ccl11*), complement factors

(e.g., *C1s* and *C3*), and cell proliferation (e.g., *Igf1*, *Rasal3*, and *Klf4*), and very few consistently downregulated genes in midkine-treated tumors (Figure 6J and Table S6). GO term analysis showed enrichment of inflammation- and cancer progression-related terms in the upregulated DEGs associated with positive regulation of cytokine production, inflammatory response, and cell activation in midkine-treated tumors (Figure 6K). Consistently between rat and human, similar terms were enriched in MDK-high clinical breast tumors, including immune response, response to cytokines, cell activation, and regulation of cell proliferation (Figure 6L). We defined a "midkine signature" based on the genes upregulated in tumors from midkine-treated rats (Figure 6J). Analysis of this signature in METABRIC and TCGA datasets showed that HR$^+$ breast tumors with high *MDK* expression levels also had elevated midkine signature regardless of patient age (Figure 6M). Overall, these data provide strong evidence that midkine promotes mammary tumor initiation and progression, implying that higher midkine levels in older individuals may favor tumorigenesis.

DISCUSSION

Aging is one of the strongest risk factors for breast cancer,[1,2] yet the mechanisms by which aging promotes tumor initiation and progression remain elusive. Here, we describe single-cell multiomic profiling of mammary glands of different age rats. Unlike most murine models, rats develop ER$^+$ mammary tumors[16,20,55] which is the most common subtype diagnosed in older, postmenopausal women.[56] We observed an age-related increase in mammary epithelial cell proliferation, a decline in luminal epithelial cell identity, and a decrease in naive B and T cell proportions. We also discovered a distinct subset of luminal progenitors specific to old rats with molecular profiles reflecting precancerous changes, providing a tentative link between aging and mammary tumor initiation. Furthermore, we provide functional evidence for potential mechanisms linking age-related molecular and

(C) Spaghetti plot showing the cumulative growth rate of mammary tumors from control vehicle and midkine-treated groups. Vehicle, $n = 6$. MDK-treated, $n = 8$.

(D) Kaplan-Meier survival plot for control vehicle and midkine-treated rats. p value was calculated with log rank tests. Vehicle, $n = 6$. MDK-treated, $n = 6$.

(E) Dot plots showing the mammary tumor volume (left) and weight (right) at the endpoint of control vehicle and midkine-treated groups. Vehicle, $n = 6$. MDK-treated, $n = 9$. One small tumor in midkine-treated group was discovered at the time of tissue dissection, it did not have a corresponding growth curve data.

(F) Polychromatic FACS analyses of myeloid cells in mammary tumors from control vehicle and midkine-treated groups. Vehicle, $n = 6$. MDK-treated, $n = 9$.

(G) Scatterplot showing the positive correlation of percentage of NK cell with the percentage of cDC1 (red), and XCR1$^+$ NK (blue) of total myeloid cells. The shaded area surrounding the lines represents the 95% confidence interval. Vehicle, $n = 6$. MDK-treated, $n = 9$. Best-fit R^2 and p values were computed by simple linear regression.

(H) Scatterplot showing the negative correlation of percentage of NK cell of total myeloid cells with the tumor growth rate of all tumors. The shaded area surrounding the lines represents the 95% confidence interval. Vehicle, $n = 6$. MDK-treated, $n = 8$. Best-fit R^2 and p values were computed by simple linear regression.

(I) Left, principal component analysis plot of mammary tumors from vehicle and midkine-treated groups. Right, boxplot showing the Euclidean distance of tumors in each group. Vehicle, $n = 6$. MDK-treated, $n = 9$.

(J) Heatmap illustrating the expression of DEGs (adjust p value <0.05, |Log$_2$FC| > 0.5) between the tumors from vehicle and midkine-treated groups.

(K) Network showing representative GO terms enriched in tumors from midkine-treated groups (Midkine signature). Enriched terms with p values <0.00001 are rendered in a network plot, with terms connected by edges if similarity >0.3.

(L) Network showing representative GO terms enriched in tumors from MDK-high patients of METABRIC dataset. Enriched terms with p values <0.00001 are rendered in a network plot, with terms connected by edges if similarity >0.3.

(M) Boxplot showing midkine signature expression in HR$^+$ breast tumors from METABRIC (top) and TCGA-BRCA (bottom) datasets. Patients were classified by tumor *MDK* expression (high vs. low, upper quartile) and age at diagnosis (cutoff 55-year old). $n = 766$ (METABRIC) and 446 (TCGA). Box and whisker plot, quartiles.

(N) Summary of the observed phenotypic and molecular changes in the mammary gland with age.

Graphs (E and F) are presented as mean ± SEM. Graphs (I and M): the boxes represent the upper 75% percentile (top), median (line) and lower 25% percentile (bottom) and the whiskers range from minimum to maximum values. p values were calculated by Log rank (Mantel-Cox) test (D), unpaired two-tailed t test (E, F, and I) and Kruskal Wallis test (M). Figures 6A and 6N were created with BioRender. See also Figure S6 and Table S6.

phenotypic changes. Specifically, we identified midkine, one of the top upregulated genes with age, as a mediator of aging-associated mammary changes through the activation of the PI3K-AKT-SREBF1 signaling axis. We also showed that midkine-treated rats had increased susceptibility to NMU-induced mammary tumors (Figure 6N). In humans, midkine levels increase with age in the normal breast tissue of women and in the peripheral blood of both men and women. Higher *MDK* expression in normal breast tissue is associated with a higher risk of breast cancer in younger women, while higher *MDK* levels in HR$^+$ tumors predict shorter DSS in younger patients.

Our results demonstrating increasing transcriptomic heterogeneity and decreasing lineage fidelity with age align with previous findings in the mammary epithelium[8,9,18,57] and other cell types like T cells[28,58] and oocytes.[59] Cell states are maintained by epigenetic programs and perturbations in DNA methylation and chromatin patterns have been commonly observed with aging and tumorigenesis.[60–63] Our finding of increased H3K4me3 in the mammary epithelium is consistent with the higher H3K4me3 levels observed in aged hematopoietic stem cells in mice.[64] We also uncovered a loss of H4K20me3 in the oldest age group and showed that inhibition of the H4K20 methyltransferases promotes luminal-to-basal shift *in vitro*. H4K20me3 has not previously been linked to aging but is known to be important for maintaining genomic stability and heterochromatin,[32] thus, its decline with age may contribute to increasing transcriptomic heterogeneity and decreasing lineage fidelity.

Aging is associated with substantial alterations in the stroma, including declining tissue stiffness[65] and immune function, increasing susceptibility to age-related diseases and cancer.[66] Consistently, we observed a loss of naive T and B cells and diminished collagen production in aged rat mammary glands, suggesting a more cancer-prone environment.[8,67–70] Many age-related tumorigenic changes involve non-cell-autonomous mechanisms and tissue landscape alterations, and most anti-aging compounds (e.g., rapamycin) and interventions (e.g., exercise and caloric restriction) act through modulating the microenvironment.[71] Intriguingly, while some of the changes we observed in the mammary gland were gradual, including the decrease in B cells and LP, others, like the increase in aging-LP population, were mainly detected in the oldest group. This could be due to the limited evaluation of age groups between 12 and 22 months or may indicate that gradual alterations culminate in sudden phenotypic shifts. Similarly, functional decline and cancer incidence show sharp increases around ages 55–60 in women, coinciding with reproductive cessation and suggesting evolutionary selection.[71,72]

Our single-cell profiles revealed midkine as a candidate driver of aging-related changes in the mammary epithelium, confirmed through *in vitro* and *in vivo* studies. Specifically, midkine treatment potently increased mammary epithelial cell proliferation and induced aging-related transcriptional changes. Midkine was originally identified as a product of a retinoic acid-responsive gene in embryonal carcinoma cells.[73,74] Along with pleiotrophin, it belongs to the neurite growth factor family and has been implicated in various neurological and inflammatory diseases.[75] We found that plasma midkine levels in healthy individuals positively correlate with age, regardless of sex, establishing midkine as a putative aging biomarker. Midkine is upregulated in various

human cancer types and is proposed as a potential tumor marker.[75] Concordantly, we found higher midkine levels in DCIS and IBC patient samples compared to normal breast tissue and younger breast cancer patients with high *MDK*-expressing ER$^+$ tumors had significantly shorter DSS, suggesting that *MDK* expression can be an age-dependent prognostic marker in ER$^+$ breast cancer. Furthermore, higher *MDK* expression in normal breast tissue of younger women was associated with higher Gail score, implying that MDK is also an age-dependent predictor of breast cancer risk.[76]

Germline mutation carriers of cancer predisposition genes are commonly diagnosed with cancer at a much younger age than the general population. Emerging evidence indicate commonalities between aging and germline predisposition for cancer. For example, changes in luminal progenitors and increase of mammary epithelial cells with "mixed" luminal-basal phenotypes is observed during aging but also detected in *BRCA1* germline mutation carriers at a much younger age.[29,77] Human epidemiologic data also indicate that adults with early onset cancer show signs of "accelerated aging".[78] Our finding of higher *MDK* expression in luminal progenitors from *BRCA2* germline mutation carriers supports this hypothesis and suggests that midkine may especially be important in ER$^+$ breast cancer development, the predominant breast cancer subtype diagnosed in these women.[79]

Utilizing the NMU-induced rat mammary tumor model, we showed that midkine treatment promotes tumor initiation and progression potentially by increasing epithelial cell proliferation that expands the target population for NMU, and results in more heterogeneous and rapidly growing tumors. The higher transcriptomic heterogeneity in tumors from midkine-treated rats supports this hypothesis that multiple different epithelial cell types may serve as the cell of origin of cancer. However, midkine treatment also altered the microenvironment as evidenced by a decrease in infiltrating T and B cells in the normal mammary gland and fewer NK cells in tumors from midkine-treated animals, both of which may contribute to increased tumor initiation and progression. Consistent with our findings, breast cancer patients with higher tumor NK signature have better prognosis[54] and in human gastric cancer, midkine has associated with lower NK cell cytotoxicity and enhanced tumor development.[80]

In summary, our work provides a comprehensive understanding of mammary tissue aging at the single-cell level and illuminates the role of midkine in driving age-related cellular and molecular changes that promote mammary tumorigenesis. Our discovery of a luminal progenitor population with precancerous molecular profiles specific to older rats that is also induced by midkine treatment is consistent with the hypothesis that progenitor cells are the normal cell-of-origin for cancer[7,13] and provide a biological link between aging and cancer. Based on our findings, midkine is both a biomarker of breast cancer risk and a target for cancer prevention.

Limitations of the study

In our study we showed that midkine treatment promotes mammary tumorigenesis but did not test whether a midkine inhibitor (iMDK) would block age-related increase in tumorigenesis. This limitation is due to the lack of specificity and toxicity of iMDK, the only currently available midkine inhibitor. iMDK reduces

MDK level through an unidentified mechanism but it also inhibits PI3K signaling. Furthermore, iMDK needs to be dissolved in high concentration of DMSO, which creates a problem for *in vivo* administration.

Another limitation is the lack of functional characterization of the of age and midkine treatment-related changes in immune cell populations in mammary tumorigenesis. However, the primary focus of our study was understanding how age-related changes in the mammary epithelium promote tumorigenesis and the role of midkine in this. Thus, while we recognize the importance of altered immune cell populations, characterizing their functional relevance is beyond the scope of the current study and it is a crucial area for future research.

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact Kornelia Polyak, Dana-Farber Cancer Institute, 450 Brookline Ave., SM1070B, Boston, MA 02215, USA. E-mail: kornelia_polyak@dfci.harvard.edu.

Materials availability
Materials and reagents can be requested from the lead contact upon reasonable request. This paper did not generate unique reagents.

Data and code availability
All data needed to evaluate the conclusions in the paper are present in the paper and/or the supplemental information. All raw genomic data was deposited to Gene Expression Omnibus (GEO) and the accession numbers are listed in the key resources table. This paper does not report original code. Any additional information required to reanalyze the data reported in this work is available from the lead contact upon request.

AUTHOR CONTRIBUTIONS

K.P. and P.Y. conceived the project and designed experiments. P.Y. performed the experiments and bioinformatic analysis of scRNA-seq, scATAC-seq, and bulk RNA-seq data. Z.L. and E.R.J. analyzed public gene expression data. T.B. helped with FACS analysis. E.R.J., Z.L., M.S., J.N., P.F., and L.E.S. assisted with rat tissue collection. Y.X. assisted with scATAC-seq data preprocessing. M.M.G. performed scATAC-seq library construction. S.Y.P. provided human tissue samples. P.Y. and K.P. wrote the manuscript with input from all authors. K.P. and H.L. supervised the work.

DECLARATION OF INTERESTS

K.P. serves on the Scientific Advisory Board of Ideaya Biosciences and Scorpion Therapeutics, holds equity options in Scorpion Therapeutics and Ideaya Biosciences, and receives sponsored research funding from Novartis where she also consults. L.E.S. is current employee of Astra-Zeneca. H.W.L. receives research funding from Novartis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animal model
  - Human breast tissue samples
- METHOD DETAILS
  - Rat treatment and tissue harvesting
  - Mammary organoid and cell culture
  - Immunoblotting
  - Immunofluorescence staining
  - Histology
  - Plasma hormone ELISA
  - Flow cytometric analysis
  - Quantitative real-time PCR
  - Single cell RNA-seq library preparation
  - Single cell ATAC-seq library preparation
  - Bulk RNA-seq library preparation
  - Single-cell RNA-seq data analysis
  - Single-cell ATAC-seq data analysis
  - Bulk RNA-seq data analysis
  - Survival analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Statistical analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.ccell.2024.09.002.

REFERENCES

1. Peto, J. (2001). Cancer epidemiology in the last century and the next decade. Nature *411*, 390–395. https://doi.org/10.1038/35077256.

2. Edwards, B.K., Howe, H.L., Ries, L.A.G., Thun, M.J., Rosenberg, H.M., Yancik, R., Wingo, P.A., Jemal, A., and Feigal, E.G. (2002). Annual report to the nation on the status of cancer, 1973-1999, featuring implications of age and aging on U.S. cancer burden. Cancer *94*, 2766–2792. https://doi.org/10.1002/cncr.10593.

3. Bidoli, E., Virdone, S., Hamdi-Cherif, M., Toffolutti, F., Taborelli, M., Panato, C., and Serraino, D. (2019). Worldwide Age at Onset of Female Breast Cancer: A 25-Year Population-Based Cancer Registry Study. Sci. Rep. *9*, 14111. https://doi.org/10.1038/s41598-019-50680-5.

4. Azam, S., Sjölander, A., Eriksson, M., Gabrielson, M., Czene, K., and Hall, P. (2019). Determinants of Mammographic Density Change. JNCI Cancer Spectr. *3*, pkz004. https://doi.org/10.1093/jncics/pkz004.

5. Pelissier Vatter, F.A., Schapiro, D., Chang, H., Borowsky, A.D., Lee, J.K., Parvin, B., Stampfer, M.R., LaBarge, M.A., Bodenmiller, B., and Lorens, J.B. (2018). High-Dimensional Phenotyping Identifies Age-Emergent Cells in Human Mammary Epithelia. Cell Rep. *23*, 1205–1219. https://doi.org/10.1016/j.celrep.2018.03.114.

6. McCormack, V.A., Perry, N.M., Vinnicombe, S.J., and Dos Santos Silva, I. (2010). Changes and tracking of mammographic density in relation to Pike's model of breast tissue aging: a UK longitudinal study. Int. J. Cancer *127*, 452–461. https://doi.org/10.1002/ijc.25053.

7. Zhang, M., Lee, A.V., and Rosen, J.M. (2017). The Cellular Origin and Evolution of Breast Cancer. Cold Spring Harb. Perspect. Med. *7*, a027128. https://doi.org/10.1101/cshperspect.a027128.

8. Li, C.M.C., Shapiro, H., Tsiobikas, C., Selfors, L.M., Chen, H., Rosenbluth, J., Moore, K., Gupta, K.P., Gray, G.K., Oren, Y., et al. (2020). Aging-Associated Alterations in Mammary Epithelia and Stroma Revealed by Single-Cell RNA Sequencing. Cell Rep. *33*, 108566. https://doi.org/10.1016/j.celrep.2020.108566.

9. Gray, G.K., Li, C.M.C., Rosenbluth, J.M., Selfors, L.M., Girnius, N., Lin, J.R., Schackmann, R.C.J., Goh, W.L., Moore, K., Shapiro, H.K., et al. (2022). A human breast atlas integrating single-cell proteomics and transcriptomics. Dev. Cell *57*, 1400–1420.e7. https://doi.org/10.1016/j.devcel.2022.05.003.

10. Bai, H., Liu, X., Lin, M., Meng, Y., Tang, R., Guo, Y., Li, N., Clarke, M.F., and Cai, S. (2024). Progressive senescence programs induce intrinsic vulnerability to aging-related female breast cancer. Nat. Commun. *15*, 5154. https://doi.org/10.1038/s41467-024-49106-2.

11. Reed, A.D., Pensa, S., Steif, A., Stenning, J., Kunz, D.J., Porter, L.J., Hua, K., He, P., Twigger, A.J., Siu, A.J.Q., et al. (2024). A single-cell atlas enables mapping of homeostatic cellular shifts in the adult human breast. Nat. Genet. *56*, 652–662. https://doi.org/10.1038/s41588-024-01688-9.

12. Polyak, K., and Kalluri, R. (2010). The role of the microenvironment in mammary gland development and cancer. Cold Spring Harb. Perspect. Biol. *2*, a003244. https://doi.org/10.1101/cshperspect.a003244.

13. Tharmapalan, P., Mahendralingam, M., Berman, H.K., and Khokha, R. (2019). Mammary stem cells and progenitors: targeting the roots of breast cancer for prevention. EMBO J. *38*, e100852. https://doi.org/10.15252/embj.2018100852.

14. Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84–88. https://doi.org/10.1038/nature04372.

15. Dong, Q., Gao, H., Shi, Y., Zhang, F., Gu, X., Wu, A., Wang, D., Chen, Y., Bandyopadhyay, A., Yeh, I.T., et al. (2016). Aging is associated with an expansion of CD49f(hi) mammary stem cells that show a decline in function and increased transformation potential. Aging (Albany NY) *8*, 2754–2776. https://doi.org/10.18632/aging.101082.

16. Russo, J., Gusterson, B.A., Rogers, A.E., Russo, I.H., Wellings, S.R., and van Zwieten, M.J. (1990). Comparative study of human and rat mammary tumorigenesis. Lab. Invest. *62*, 244–278.

17. Miyano, M., Sayaman, R.W., Shalabi, S.F., Senapati, P., Lopez, J.C., Angarola, B.L., Hinz, S., Zirbes, A., Anczukow, O., Yee, L.D., et al. (2021). Breast-Specific Molecular Clocks Comprised of ELF5 Expression and Promoter Methylation Identify Individuals Susceptible to Cancer Initiation. Cancer Prev. Res. *14*, 779–794. https://doi.org/10.1158/1940-6207.CAPR-20-0635.

18. Miyano, M., Sayaman, R.W., Stoiber, M.H., Lin, C.H., Stampfer, M.R., Brown, J.B., and LaBarge, M.A. (2017). Age-related gene expression in luminal epithelial cells is driven by a microenvironment made from myoepithelial cells. Aging (Albany NY) *9*, 2026–2051. https://doi.org/10.18632/aging.101298.

19. Iannaccone, P.M., and Jacob, H.J. (2009). Rats. Dis. Model. Mech. *2*, 206–210. https://doi.org/10.1242/dmm.002733.

20. Russo, J., and Russo, I.H. (2000). Atlas and histologic classification of tumors of the rat mammary gland. J. Mammary Gland Biol. Neoplasia *5*, 187–200. https://doi.org/10.1023/a:1026443305758.

21. Alečković, M., Cristea, S., Gil Del Alcazar, C.R., Yan, P., Ding, L., Krop, E.D., Harper, N.W., Rojas Jimenez, E., Lu, D., Gulvady, A.C., et al. (2022). Breast cancer prevention by short-term inhibition of TGFβ signaling. Nat. Commun. *13*, 7558. https://doi.org/10.1038/s41467-022-35043-5.

22. Gil Del Alcazar, C.R., Trinh, A., Alečković, M., Rojas Jimenez, E., Harper, N.W., Oliphant, M.U.J., Xie, S., Krop, E.D., Lulseged, B., Murphy, K.C., et al. (2022). Insights into Immune Escape During Tumor Evolution and Response to Immunotherapy Using a Rat Model of Breast Cancer.

23. Thompson, T.A., Haag, J.D., and Gould, M.N. (2000). ras gene mutations are absent in NMU-induced mammary carcinomas from aging rats. Carcinogenesis *21*, 1917–1922. https://doi.org/10.1093/carcin/21.10.1917.

24. Bhowmick, N.A., Neilson, E.G., and Moses, H.L. (2004). Stromal fibroblasts in cancer initiation and progression. Nature *432*, 332–337. https://doi.org/10.1038/nature03096.

25. Howell, A., Landberg, G., and Bergh, J. (2009). Breast tumour stroma is a prognostic indicator and target for therapy. Breast Cancer Res. *11*, S16. https://doi.org/10.1186/bcr2435.

26. Elyahu, Y., and Monsonego, A. (2021). Thymus involution sets the clock of the aging T-cell landscape: Implications for declined immunity and tissue repair. Ageing Res. Rev. *65*, 101231. https://doi.org/10.1016/j.arr.2020.101231.

27. López-Otín, C., Pietrocola, F., Roiz-Valle, D., Galluzzi, L., and Kroemer, G. (2023). Meta-hallmarks of aging and cancer. Cell Metabol. *35*, 12–35. https://doi.org/10.1016/j.cmet.2022.11.001.

28. Martinez-Jimenez, C.P., Eling, N., Chen, H.C., Vallejos, C.A., Kolodziejczyk, A.A., Connor, F., Stojic, L., Rayner, T.F., Stubbington, M.J.T., Teichmann, S.A., et al. (2017). Aging increases cell-to-cell transcriptional variability upon immune stimulation. Science *355*, 1433–1436. https://doi.org/10.1126/science.aah4115.

29. Shalabi, S.F., Miyano, M., Sayaman, R.W., Lopez, J.C., Jokela, T.A., Todhunter, M.E., Hinz, S., Garbe, J.C., Stampfer, M.R., Kessenbrock, K., et al. (2021). Evidence for accelerated aging in mammary epithelia of women carrying germline BRCA1 or BRCA2 mutations. Nat. Aging *1*, 838–849. https://doi.org/10.1038/s43587-021-00104-9.

30. Zou, M.R., Cao, J., Liu, Z., Huh, S.J., Polyak, K., and Yan, Q. (2014). Histone demethylase jumonji AT-rich interactive domain 1B (JARID1B) controls mammary gland development by regulating key developmental and lineage specification genes. J. Biol. Chem. *289*, 17620–17633. https://doi.org/10.1074/jbc.M114.570853.

31. Hinohara, K., Wu, H.J., Vigneau, S., McDonald, T.O., Igarashi, K.J., Yamamoto, K.N., Madsen, T., Fassl, A., Egri, S.B., Papanastasiou, M., et al. (2018). KDM5 Histone Demethylase Activity Links Cellular Transcriptomic Heterogeneity to Therapeutic Resistance. Cancer Cell *34*, 939–953.e9. https://doi.org/10.1016/j.ccell.2018.10.014.

32. Agredo, A., and Kasinski, A.L. (2023). Histone 4 lysine 20 tri-methylation: a key epigenetic regulator in chromatin structure and disease. Front. Genet. *14*, 1243395. https://doi.org/10.3389/fgene.2023.1243395.

33. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

34. Chaffer, C.L., and Weinberg, R.A. (2010). Cancer cell of origin: spotlight on luminal progenitors. Cell Stem Cell *7*, 271–272. https://doi.org/10.1016/j.stem.2010.08.008.

35. Forster, N., Saladi, S.V., van Bragt, M., Sfondouris, M.E., Jones, F.E., Li, Z., and Ellisen, L.W. (2014). Basal cell signaling by p63 controls luminal progenitor function and lactation via NRG1. Dev. Cell *28*, 147–160. https://doi.org/10.1016/j.devcel.2013.11.019.

36. Faraldo, M.M., Teulière, J., Deugnier, M.A., Taddei-De La Hosseraye, I., Thiery, J.P., and Glukhova, M.A. (2005). Myoepithelial cells in the control of mammary development and tumorigenesis: data from genetically modified mice. J. Mammary Gland Biol. Neoplasia *10*, 211–219. https://doi.org/10.1007/s10911-005-9582-8.

37. Browaeys, R., Saelens, W., and Saeys, Y. (2020). NicheNet: modeling intercellular communication by linking ligands to target genes. Nat. Methods *17*, 159–162. https://doi.org/10.1038/s41592-019-0667-5.

38. Filippou, P.S., Karagiannis, G.S., and Constantinidou, A. (2020). Midkine (MDK) growth factor: a key player in cancer progression and a promising therapeutic target. Oncogene *39*, 2040–2054. https://doi.org/10.1038/s41388-019-1124-8.

39. Ibusuki, M., Fujimori, H., Yamamoto, Y., Ota, K., Ueda, M., Shinriki, S., Taketomi, M., Sakuma, S., Shinohara, M., Iwase, H., and Ando, Y. (2009). Midkine in plasma as a novel breast cancer marker. Cancer Sci. *100*, 1735–1739. https://doi.org/10.1111/j.1349-7006.2009.01233.x.

40. Tanaka, T., Biancotto, A., Moaddel, R., Moore, A.Z., Gonzalez-Freire, M., Aon, M.A., Candia, J., Zhang, P., Cheung, F., Fantoni, G., et al. (2018). Plasma proteomic signature of age in healthy humans. Aging Cell *17*, e12799. https://doi.org/10.1111/acel.12799.

41. Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70. https://doi.org/10.1038/nature11412.

42. GTEx Consortium (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.

43. Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., et al. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature *486*, 346–352. https://doi.org/10.1038/nature10983.

44. Chen, S., Liao, J., Li, J., and Wang, S. (2024). GPC2 promotes prostate cancer progression via MDK-mediated activation of PI3K/AKT signaling pathway. Funct. Integr. Genomics *24*, 127. https://doi.org/10.1007/s10142-024-01406-y.

45. Hu, B., Qin, C., Li, L., Wei, L., Mo, X., Fan, H., Lei, Y., Wei, F., and Zou, D. (2021). Midkine promotes glioblastoma progression via PI3K-Akt signaling. Cancer Cell Int. *21*, 509. https://doi.org/10.1186/s12935-021-02212-3.

46. Cai, Y.Q., Lv, Y., Mo, Z.C., Lei, J., Zhu, J.L., and Zhong, Q.Q. (2020). Multiple pathophysiological roles of midkine in human disease. Cytokine *135*, 155242. https://doi.org/10.1016/j.cyto.2020.155242.

47. Gail, M.H., Brinton, L.A., Byar, D.P., Corle, D.K., Green, S.B., Schairer, C., and Mulvihill, J.J. (1989). Projecting individualized probabilities of developing breast cancer for white females who are being examined annually. J. Natl. Cancer Inst. *81*, 1879–1886. https://doi.org/10.1093/jnci/81.24.1879.

48. Kang, T., Yau, C., Wong, C.K., Sanborn, J.Z., Newton, Y., Vaske, C., Benz, S.C., Krings, G., Camarda, R., Henry, J.E., et al. (2020). A risk-associated Active transcriptome phenotype expressed by histologically normal human breast tissue and linked to a pro-tumorigenic adipocyte population. Breast Cancer Res. *22*, 81. https://doi.org/10.1186/s13058-020-01322-6.

49. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.

50. Brown, M.S., and Goldstein, J.L. (1997). The SREBP pathway: regulation of cholesterol metabolism by proteolysis of a membrane-bound transcription factor. Cell *89*, 331–340. https://doi.org/10.1016/s0092-8674(00)80213-5.

51. Tang, J.J., Li, J.G., Qi, W., Qiu, W.W., Li, P.S., Li, B.L., and Song, B.L. (2011). Inhibition of SREBP by a small molecule, betulin, improves hyperlipidemia and insulin resistance and reduces atherosclerotic plaques. Cell Metabol. *13*, 44–56. https://doi.org/10.1016/j.cmet.2010.12.004.

52. Shao, W., Machamer, C.E., and Espenshade, P.J. (2016). Fatostatin blocks ER exit of SCAP but inhibits cell growth in a SCAP-independent manner. J. Lipid Res. *57*, 1564–1573. https://doi.org/10.1194/jlr.M069583.

53. Porstmann, T., Santos, C.R., Griffiths, B., Cully, M., Wu, M., Leevers, S., Griffiths, J.R., Chung, Y.L., and Schulze, A. (2008). SREBP activity is regulated by mTORC1 and contributes to Akt-dependent cell growth. Cell Metabol. *8*, 224–236. https://doi.org/10.1016/j.cmet.2008.07.007.

54. Böttcher, J.P., Bonavita, E., Chakravarty, P., Blees, H., Cabeza-Cabrerizo, M., Sammicheli, S., Rogers, N.C., Sahai, E., Zelenay, S., and Reis e Sousa, C. (2018). NK Cells Stimulate Recruitment of cDC1 into the Tumor Microenvironment Promoting Cancer Immune Control. Cell *172*, 1022–1037.e14. https://doi.org/10.1016/j.cell.2018.01.004.

55. Nicotra, R., Lutz, C., Messal, H.A., and Jonkers, J. (2024). Rat Models of Hormone Receptor-Positive Breast Cancer. J. Mammary Gland Biol. Neoplasia *29*, 12. https://doi.org/10.1007/s10911-024-09566-0.

56. Jenkins, E.O., Deal, A.M., Anders, C.K., Prat, A., Perou, C.M., Carey, L.A., and Muss, H.B. (2014). Age-specific changes in intrinsic breast cancer subtypes: a focus on older women. Oncologist *19*, 1076–1083. https://doi.org/10.1634/theoncologist.2014-0184.

57. Kumar, T., Nee, K., Wei, R., He, S., Nguyen, Q.H., Bai, S., Blake, K., Pein, M., Gong, Y., Sei, E., et al. (2023). A spatially resolved single-cell genomic atlas of the adult human breast. Nature *620*, 181–191. https://doi.org/10.1038/s41586-023-06252-9.

58. Lu, J., Ahmad, R., Nguyen, T., Cifello, J., Hemani, H., Li, J., Chen, J., Li, S., Wang, J., Achour, A., et al. (2022). Heterogeneity and transcriptome changes of human CD8(+) T cells across nine decades of life. Nat. Commun. *13*, 5128. https://doi.org/10.1038/s41467-022-32869-x.

59. Castillo-Fernandez, J., Herrera-Puerta, E., Demond, H., Clark, S.J., Hanna, C.W., Hemberger, M., and Kelsey, G. (2020). Increased transcriptome variation and localised DNA methylation changes in oocytes from aged mice revealed by parallel single-cell analysis. Aging Cell *19*, e13278. https://doi.org/10.1111/acel.13278.

60. Wang, K., Liu, H., Hu, Q., Wang, L., Liu, J., Zheng, Z., Zhang, W., Ren, J., Zhu, F., and Liu, G.H. (2022). Epigenetic regulation of aging: implications for interventions of aging and diseases. Signal Transduct. Targeted Ther. *7*, 374. https://doi.org/10.1038/s41392-022-01211-8.

61. Lu, Y., Brommer, B., Tian, X., Krishnan, A., Meer, M., Wang, C., Vera, D.L., Zeng, Q., Yu, D., Bonkowski, M.S., et al. (2020). Reprogramming to recover youthful epigenetic information and restore vision. Nature *588*, 124–129. https://doi.org/10.1038/s41586-020-2975-4.

62. Berdyshev, G.D., Korotaev, G.K., Boiarskikh, G.V., and Vaniushin, B.F. (1967). [Nucleotide composition of DNA and RNA from somatic tissues of humpback and its changes during spawning]. Biokhimiia *32*, 988–993.

63. Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89–92. https://doi.org/10.1038/301089a0.

64. Sun, D., Luo, M., Jeong, M., Rodriguez, B., Xia, Z., Hannah, R., Wang, H., Le, T., Faull, K.F., Chen, R., et al. (2014). Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal. Cell Stem Cell *14*, 673–688. https://doi.org/10.1016/j.stem.2014.03.002.

65. Selman, M., and Pardo, A. (2021). Fibroageing: An ageing pathological feature driven by dysregulated extracellular matrix-cell mechanobiology. Ageing Res. Rev. *70*, 101393. https://doi.org/10.1016/j.arr.2021.101393.

66. Mogilenko, D.A., Shchukina, I., and Artyomov, M.N. (2022). Immune ageing at single-cell resolution. Nat. Rev. Immunol. *22*, 484–498. https://doi.org/10.1038/s41577-021-00646-4.

67. Qian, B.Z., Li, J., Zhang, H., Kitamura, T., Zhang, J., Campion, L.R., Kaiser, E.A., Snyder, L.A., and Pollard, J.W. (2011). CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. Nature *475*, 222–225. https://doi.org/10.1038/nature10138.

68. Franklin, R.A., Liao, W., Sarkar, A., Kim, M.V., Bivona, M.R., Liu, K., Pamer, E.G., and Li, M.O. (2014). The cellular and molecular origin of tumor-associated macrophages. Science *344*, 921–925. https://doi.org/10.1126/science.1252510.

69. Amens, J.N., Bahçecioglu, G., and Zorlutuna, P. (2021). Immune System Effects on Breast Cancer. Cell. Mol. Bioeng. *14*, 279–292. https://doi.org/10.1007/s12195-021-00679-8.

70. Xu, S., Xu, H., Wang, W., Li, S., Li, H., Li, T., Zhang, W., Yu, X., and Liu, L. (2019). The role of collagen in cancer: from bench to bedside. J. Transl. Med. *17*, 309. https://doi.org/10.1186/s12967-019-2058-1.

71. Laconi, E., Marongiu, F., and DeGregori, J. (2020). Cancer as a disease of old age: changing mutational and microenvironmental landscapes. Br. J. Cancer *122*, 943–952. https://doi.org/10.1038/s41416-019-0721-1.

72. DeGregori, J. (2018). Adaptive Oncogenesis : A New Understanding of How Cancer Evolves inside Us (Harvard University Press).

73. Kadomatsu, K., Tomomura, M., and Muramatsu, T. (1988). cDNA cloning and sequencing of a new gene intensely expressed in early differentiation stages of embryonal carcinoma cells and in mid-gestation period of mouse embryogenesis. Biochem. Biophys. Res. Commun. 151, 1312–1318. https://doi.org/10.1016/s0006-291x(88)80505-9.

74. Matsubara, S., Tomomura, M., Kadomatsu, K., and Muramatsu, T. (1990). Structure of a retinoic acid-responsive gene, MK, which is transiently activated during the differentiation of embryonal carcinoma cells and the mid-gestation period of mouse embryogenesis. J. Biol. Chem. 265, 9441–9443.

75. Neumaier, E.E., Rothhammer, V., and Linnerbauer, M. (2023). The role of midkine in health and disease. Front. Immunol. 14, 1310094. https://doi.org/10.3389/fimmu.2023.1310094.

76. Velentzis, L.S., Freeman, V., Campbell, D., Hughes, S., Luo, Q., Steinberg, J., Egger, S., Mann, G.B., and Nickson, C. (2023). Breast Cancer Risk Assessment Tools for Stratifying Women into Risk Groups: A Systematic Review. Cancers 15, 1124. https://doi.org/10.3390/cancers15041124.

77. Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A., et al. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat. Med. 15, 907–913. https://doi.org/10.1038/nm.2000.

78. Tian, R., Tica, S., Hong, D., Oduyale, O., Zong, X., Ren, D., Govindan, R., and Cao, Y. (2024). Rising accelerated aging in recent generations associated with elevated risk of early-onset cancers. In Proceedings of the 115th Annual Meeting of the American Association for Cancer Research.

79. Metcalfe, K., Lynch, H.T., Foulkes, W.D., Tung, N., Olopade, O.I., Eisen, A., Lerner-Ellis, J., Snyder, C., Kim, S.J., Sun, P., and Narod, S.A. (2019). Oestrogen receptor status and survival in women with BRCA2-associated breast cancer. Br. J. Cancer 120, 398–403. https://doi.org/10.1038/s41416-019-0376-y.

80. Zhao, S., Wang, H., Nie, Y., Mi, Q., Chen, X., and Hou, Y. (2012). Midkine upregulates MICA/B expression in human gastric cancer cells and decreases natural killer cell cytotoxicity. Cancer Immunol. Immunother. 61, 1745–1753. https://doi.org/10.1007/s00262-012-1235-3.

81. Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., McQuaid, S., Gray, R.T., Murray, L.J., Coleman, H.G., et al. (2017). QuPath: Open source software for digital pathology image analysis. Sci. Rep. 7, 16878. https://doi.org/10.1038/s41598-017-17204-5.

82. Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. 10, 1523. https://doi.org/10.1038/s41467-019-09234-6.

83. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. https://doi.org/10.1101/gr.1239303.

84. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell 184, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

85. Villanueva, R.A.M., and Chen, Z.J. (2019). ggplot2: Elegant Graphics for Data Analysis (2nd ed.). Meas. Interdiscip. Res. Perspect. 17, 160–167. https://doi.org/10.1080/15366367.2019.1565254.

86. Morgan, M., Pages, H., Obenchain, V., and Hayden, N. (2016). Rsamtools: Binary alignment (BAM), FASTA, variant call (BCF), and tabix file import, pp. 677–689. R package version 1.

87. Stuart, T., Srivastava, A., Madad, S., Lareau, C.A., and Satija, R. (2021). Single-cell chromatin state analysis with Signac. Nat. Methods 18, 1333–1341. https://doi.org/10.1038/s41592-021-01282-5.

88. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS 16, 284–287. https://doi.org/10.1089/omi.2011.0118.

89. Jin, S., Guerrero-Juarez, C.F., Zhang, L., Chang, I., Ramos, R., Kuan, C.H., Myung, P., Plikus, M.V., and Nie, Q. (2021). Inference and analysis of cell-cell communication using CellChat. Nat. Commun. 12, 1088. https://doi.org/10.1038/s41467-021-21246-9.

90. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. https://doi.org/10.1186/s13059-014-0550-8.

91. Wolf, F.A., Angerer, P., and Theis, F.J. (2018). SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15. https://doi.org/10.1186/s13059-017-1382-0.

92. Satpathy, A.T., Granja, J.M., Yost, K.E., Qi, Y., Meschi, F., McDermott, G.P., Olsen, B.N., Mumbach, M.R., Pierce, S.E., Corces, M.R., et al. (2019). Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. Nat. Biotechnol. 37, 925–936. https://doi.org/10.1038/s41587-019-0206-z.

93. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation 2, 100141. https://doi.org/10.1016/j.xinn.2021.100141.

94. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. USA 102, 15545–15550. https://doi.org/10.1073/pnas.0506580102.

95. Huynh-Thu, V.A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring regulatory networks from expression data using tree-based methods. PLoS One 5, e12776. https://doi.org/10.1371/journal.pone.0012776.

96. Cornwell, M., Vangala, M., Taing, L., Herbert, Z., Köster, J., Li, B., Sun, H., Li, T., Zhang, J., Qiu, X., et al. (2018). VIPER: Visualization Pipeline for RNA-seq, a Snakemake workflow for efficient and complete RNA-seq analysis. BMC Bioinf. 19, 135. https://doi.org/10.1186/s12859-018-2139-9.

97. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

98. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47. https://doi.org/10.1093/nar/gkv007.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-Ki67	Abcam	Cat# ab16667; RRID: AB_302459
Mouse monoclonal anti-SMA	Thermo Fisher Scientific	Cat# MS113P; RRID: AB_64000
Rabbit polyclonal anti-Progesterone Receptor	Abcam	Cat# ab16661; RRID: AB_443421
Mouse monoclonal anti-Estrogen Receptor alpha	Thermo Fisher Scientific	Cat# MA5-13304; RRID: AB_11002193
Mouse monoclonal anti-ErbB2/HER2	Abcam	Cat# ab16901; RRID: AB_443537
Rabbit monoclonal anti-Midkine	Abcam	Cat# ab52637; RRID: AB_880698
Rabbit polyclonal anti-Midkine	Proteintech	Cat# 11009-1-AP; RRID: AB_2250619
Rabbit polyclonal anti-SREBF1	Proteintech	Cat# 14088-1-AP; RRID: AB_2255217
Rabbit monoclonal anti-Phospho-S6 (Ser235/236)	Cell Signaling Technology	Cat# 4858; RRID: AB_916156
Rabbit monoclonal anti-S6 ribosomal protein	Cell Signaling Technology	Cat# 2217; RRID: AB_331355
Rabbit polyclonal anti-Phospho-Akt (Ser473)	Cell Signaling Technology	Cat# 9271; RRID: AB_329825
Rabbit polyclonal anti-AKT	Cell Signaling Technology	Cat# 9272; RRID: AB_329827
Mouse monoclonal anti-alpha-Tubulin	Sigma-Aldrich	Cat# T5168; RRID: AB_477579
Mouse monoclonal anti-Cytokeratin 18	Abcam	Cat# ab668; RRID: AB_305647
Rabbit polyclonal anti-Keratin 14	BioLegend	Cat# 905301; RRID: AB_2565048
Rabbit polyclonal anti-Histone H4 (tri methyl K20)	Abcam	Cat# ab9053; RRID: AB_306969
Mouse monoclonal anti-Histone H3	Active Motif	Cat# 39763; RRID: AB_2650522
Rabbit polyclonal anti-Cytokeratin 17	Abcam	Cat# ab53707; RRID: AB_869865
Mouse monoclonal anti-Histone H4	Abcam	Cat# ab31830; RRID:AB_1209246
Rabbit polyclonal anti-H3K4me3	Abcam	Cat# ab8580; RRID:AB_306649
Rabbit monoclonal anti-P63	Abcam	Cat# ab214790; RRID: AB_10971840
Mouse monoclonal anti-CD45	BD Biosciences	Cat# 740371; RRID: AB_2740103
Mouse monoclonal anti-MHC Class II	Thermo Fisher Scientific	Cat# 46-0463-82; RRID: AB_10736599
Mouse monoclonal anti-CD45R (B220)	Thermo Fisher Scientific	Cat# 25-0460-82; RRID: AB_2573352
Mouse monoclonal anti-CD3	Bio-Rad	Cat# MCA772A700; RRID: AB_2073345
Mouse monoclonal anti-CD19	Abcam	Cat# ab24936; RRID: AB_726191
Mouse monoclonal anti-CD161	Bio-Rad	Cat# MCA1427A647; RRID: AB_322589
Mouse monoclonal anti-CD4	BioLegend	Cat# 201518; RRID: AB_1186084
Mouse monoclonal anti-CD8	BD Biosciences	Cat# 740139; RRID: AB_2739895
Mouse monoclonal anti-TCR gamma/delta	BioLegend	Cat# 202605; RRID: AB_492904
Rat monoclonal anti-FOXP3	Thermo Fisher Scientific	Cat# 48-5773-82; RRID: AB_1518812
Mouse monoclonal anti-CD11b	Bio-Rad	Cat# MCA275PB; RRID: AB_566459
Rabbit-polyclonal anti-EMR1	Bioss	Cat# bs-7058R-PE
Rat monoclonal anti-Ly6g	Abcam	Cat# ab25024; RRID: AB_470400
Mouse monoclonal anti-CD11c	Abcam	Cat# ab11029; RRID: AB_297683
Mouse monoclonal FITC anti-rat CD3	BioLegend	Cat# 201403; RRID: AB_2073344
Mouse monoclonal APC anti-rat TCR α/β	BioLegend	Cat# 201110; RRID: AB_313919
Mouse monoclonal APC/Cyanine7 anti-rat CD4	BioLegend	Cat# 201518; RRID: AB_1186088

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse monoclonal BV711 anti-rat CD8a	BD Biosciences	Cat# 740724; RRID: AB_2740402
Mouse monoclonal PerCP/Cyanine5.5 anti-rat CD45RA	BioLegend	Cat# 202318; RRID: AB_2565944
Mouse monoclonal BV786 anti-rat CD45RC	BD Biosciences	Cat# 742491; RRID: AB_2740824
Mouse monoclonal PE/Cyanine7 anti-rat CD45	BioLegend	Cat# 202214; RRID: AB_2174546
Armenian hamster monoclonal BV605 anti-human/mouse/rat CD278 (ICOS)	BioLegend	Cat# 313538; RRID: AB_2687079
Mouse monoclonal BV650 anti-Rat CD25	BD Biosciences	Cat# 742755; RRID: AB_2741020
Rat monoclonal PE anti-FOXP3	Thermo Fisher Scientific	Cat# 12-5773-82; RRID: AB_465936
Mouse monoclonal InVivoMAb anti-mouse/human/rat PD-L1	Bio X Cell	Cat# BE0383; RRID: AB_2927520
Rabbit monoclonal anti-CD163	Abcam	Cat# ab182422; RRID: AB_2753196
Mouse recombinant PE anti-rat CD68	BioLegend	Cat# 201004; RRID: AB_2936606
Mouse monoclonal BV421 anti-rat CD11b/c	BD Biosciences	Cat# 743977; RRID: AB_2741898
Mouse monoclonal BV650 anti-rat RT1B	BD Biosciences	Cat# 744129; RRID: AB_2742019
Mouse monoclonal BV605 anti-rat Integrin αE2 (CD103)	BD Biosciences	Cat# 745138; RRID: AB_2742741
Mouse monoclonal APC/Cyanine7 anti-mouse/rat XCR1	BioLegend	Cat# 148224; RRID: AB_2783117
Mouse monoclonal BV786 anti-rat CD314 (NKG2D)	BD Biosciences	Cat# 744734; RRID: AB_2742444
Mouse monoclonal BV711 anti-rat CD161a	BD Biosciences	Cat# 744053; RRID: AB_2741956
Mouse monoclonal PE/Cyanine7 anti-rat CD45R(B220)	Thermo Fisher Scientific	Cat# 25-0460-82; RRID: AB_2573352
Mouse monoclonal FITC anti-mouse/rat/human CD27	BioLegend	Cat# 124208; RRID: AB_1236466
Biological samples		
Human normal breast tissue samples	Seoul National University Bundang Hospital (Seoul, Korea)	
Human DCIS and IBC tissue samples	Seoul National University Bundang Hospital (Seoul, Korea)	
Human normal breast tissue samples	Brigham and Women's Hospital (Boston, USA)	
Chemicals, peptides, and recombinant proteins		
Recombinant rat midkine protein	Novus Biologicals	Cat# NBP2-35273
N-Nitroso-N-methylurea	MedChemExpress	Cat# HY-34758; CAS# 684-93-5
Betulin	MedChemExpress	Cat# HY-N0083; CAS# 473-98-3
Fatostatin	MedChemExpress	Cat# HY-14452; CAS# 125256-00-0
Rapamycin	Selleckchem	Cat# S1039; CAS# 53123-88-9
A-196	MedChemExpress	Cat# HY-100201; CAS# 1982372-88-2
C70	Xcess Bio	Cat# M60192; CAS# 1596348-32-1
Advanced DMEM/F-12	Gibco	Cat# 12634028
GlutaMAX™ Supplement	Gibco	Cat# 35050061
HEPES Buffer (1 M, pH 7.4)	Boston BioProducts	Cat# BBH-74; CAS# 7365-45-9
Recombinant Human R-Spondin-1	Peprotech	Cat# 120-38
Amphotericin B	Gibco	Cat# 15290026; CAS# 1397-89-3
B-27™ Supplement (50X), serum free	Gibco	Cat# 17504044
N-Acetyl-L-cysteine	Sigma-Aldrich	Cat# A9165; CAS# 616-91-1
Nicotinamide	Sigma-Aldrich	Cat# N0636; CAS# 98-92-0
SB 202190	Sigma-Aldrich	Cat# S7067; CAS# 152121-30-7

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Recombinant Human Keratinocyte Growth Factor (KGF or FGF-7)	Peprotech	Cat# 100-19
Recombinant Human FGF-10	Peprotech	Cat# 100-26
Recombinant Human Heregulinβ-1	Peprotech	Cat# 100-03
Animal-Free Recombinant Human EGF	Peprotech	Cat# AF-100-15
Insulin, human recombinant, zinc solution	Thermo Fisher Scientific	Cat# 12585014
DMEM high glucose	Corning	Cat# 10-013-CV
Hydrocortisone	Sigma-Aldrich	Cat# H0135
Corning™ Matrigel™ Growth Factor Reduced (GFR) Basement Membrane Matrix, Phenol Red-Free, LDEV-Free	Corning	Cat# 356238; Lot# 2025001
Clarity Max Western ECL Substrate	Bio-Rad	Cat# 1705062
LI-COR Intercept T20 (TBS) Antibody Diluent	LI-COR	Cat# 92765001
Collagenase type IV	Worthington Biochemical Corporation	Cat# LS004189; CAS# 9001-12-1
Hyaluronidase	Sigma-Aldrich	Cat# H3506; CAS# 37326-33-3
Critical commercial assays		
Progesterone ELISA Kit	Cayman Chemical	Cat# 582601; RRID: AB_2811273
Prolactin (rat) EIA Kit	Cayman Chemical	Cat# 589701
Estradiol ELISA Kit	Cayman Chemical	Cat# 501890; RRID: AB_2832924
RNeasy Micro Kit	QIAGEN	Cat# 74106
PrimeScript RT Reagent Kit	TAKARA	Cat# RR037A
TB Green® Premix Ex Taq™ II	TAKARA	Cat# RR82WR
LIVE/DEAD™ Fixable Aqua Dead Cell Stain Kit, for 405 nm	Thermo Fisher Scientific	Cat# L34966
Chromium Next GEM Single Cell 5' HT Kit v2	10x Genomics	PN: 1000356
Chromium Next GEM Single Cell ATAC Kit v2	10x Genomics	PN: 1000406
Deposited data		
scRNA-seq, scATAC-seq and bulk-RNA-seq data	This study	GSE251835
scRNA-seq data of iHBCA	Reed et al.[11]	https://cellxgene.cziscience.com/e/55003f67-c494-46f1-83fb-902745646379.cxg/
RNA-seq data of normal breast tissues from healthy women	Kang et al.[48]	https://xenabrowser.net/datapages/?cohort=Normal%20Breast%20(Benz%202020)
Experimental models: Cell lines		
LA7	American Type Culture Collection (ATCC)	Cat# CRL-2283; RRID:CVCL_3838
Experimental models: Organisms/strains		
Rat: Fisher 344 (3-4-week-old rats)	Envigo	N/A
Rat: Fisher 344 (3-, 6-, 12- and 22-month-old rats)	National Institute on Aging	N/A
Oligonucleotides		
Mdk-F: GAGCCGACTGCAAGTACAA	This study	N/A
Mdk-R: ATTGTAGCGCGCCTTCTT	This study	N/A
Actb-F: CTCCCTGGAGAAGAGCTATGA	This study	N/A
Actb-R: AGGAAGGAAGGCTGGAAGA	This study	N/A

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Software and algorithms		
QuPath (version: 0.4.0)	Bankhead et al.[81]	https://qupath.readthedocs.io/en/0.4/index.html
ZEISS ZEN lite	ZEISS	https://www.zeiss.com/microscopy/en/products/software/zeiss-zen-lite.html
GraphPadPrism (version: 10.0.3)	GraphPad Software Inc.	https://www.graphpad.com/updates/prism-900-release-notes
FlowJo (version: 10.8.2)	FlowJo, LLC	https://www.flowjo.com/
Cell Ranger (version: 6.1.2)	10x Genomics	https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger
Cell Ranger-ATAC (version 2.0.0)	10x Genomics	https://software.10xgenomics.com/single-cell-atac/software/pipelines/2.0/what-is-cell-ranger-atac
Metascape	Zhou et al.[82]	https://metascape.org/gp/index.html#/main/step1
Cytoscape (version: 3.9.1)	Shannon et al.[83]	https://cytoscape.org
Seurat (version: 4.3.0)	Hao et al.[84]	https://satijalab.org/seurat/
Image Lab (version: 6.1.0)	Bio-Rad	https://www.bio-rad.com/en-us/product/image-lab-software?ID=KRE6P5E8Z
monocle3 (version: 1.2.9)	Cao et al.[33]	https://cole-trapnell-lab.github.io/monocle3/
ggplot2 (version: 3.4.0.9000)	Randle Aaron et al.[85]	https://ggplot2.tidyverse.org
Rsamtools (version: 2.14.0)	Morgan et al.[86]	https://bioconductor.org/packages/release/bioc/html/Rsamtools.html
SCENIC (version: 1.3.1)	Aibar et al.[49]	https://scenic.aertslab.org
Nichenetr (version: 1.1.1)	Browaeys et al.[37]	https://github.com/saeyslab/nichenetr
Signac (version: 1.9.0)	Stuart et al.[87]	https://stuartlab.org/signac/index.html
clusterProfiler (version 4.2.1)	Yu et al.[88]	https://github.com/YuLab-SMU/clusterProfiler
CellChat (version 1.6.1)	Jin et al.[89]	https://github.com/sqjin/CellChat
DESeq2 (version: 1.40.2)	Love et al.[90]	http://www.bioconductor.org/packages/release/bioc/html/DESeq2.html
Scanpy (version: 1.9.6)	Wolf et al.[91]	https://scanpy.readthedocs.io/en/stable/#

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animal model

Female Fischer rats were purchased from Envigo at 3-4 weeks of age. Aged rats were obtained from NIA. Animal experiments were performed by P.Y. and E.R.J. according to protocol 15-005 approved by the Dana-Farber Cancer Institute Animal Care and Use Committee. All rats were housed in individually ventilated, solid bottom, polysulfone 135 sq. in. microisolator cages. The cages were used in conjunction with the Optimice® rack systems with integrated automatic watering. Temperature and humidity in the rodent facility were controlled at 20°C and a target range of 35–55% relative humidity. A standard photoperiod of 12 h light/12 h dark was controlled by an automated system.

Human breast tissue samples

Fresh normal breast tissue samples were collected from women undergoing reduction mammoplasty or prophylactic mastectomy at Brigham and Women's Hospital following informed consent using protocols DF-HCC #93-085 and #10-458 approved by the Dana-Farber Cancer Institute Institutional Review Board. Normal and tumor adjacent normal breast tissues, as well as formalin fixed paraffin embedded DCIS and IBC tissue samples were obtained from Seoul National University Bundang Hospital (Seoul, Korea) using Institutional Review Board-approved protocols. All human tissue samples were de-identified prior to transfer to the laboratory.

METHOD DETAILS

Rat treatment and tissue harvesting

All animal experiments were conducted in an AAALAC-accredited SPF rodent-only barrier facility at Dana-Farber Cancer Institute (Boston, MA) in strict accordance with protocol 15-005 approved by the Dana-Farber Cancer Institute Animal Care and Use Committee. F344 rats of different age were purchased from Envigo or provided by the National Institute of Aging. Tissues were collected as previous described.[22] Briefly, two small pieces of mammary tissues were collected for histology analysis, fixed overnight in 4% formalin, stored in 70% ethanol, followed by paraffin embedding, sectioning, and hematoxylin and eosin staining by the Rodent Pathology Core of Harvard Medical School. The remaining tissue was digested in Dulbecco's modified eagle medium/nutrient mixture F-12 (DMEM/F12) medium containing 2 mg/mL collagenase type IV, 2 mg/mL hyaluronidase, and 2 mg/mL bovine serum albumin (BSA) at 37°C for 1 to 2 hours, and frozen in freezing medium (containing 90% fetal bovine serum (FBS) with 10% dimethyl sulfoxide (DMSO)) in liquid nitrogen. Bone marrow was collected by flushing the femurs and frozen in freezing medium for subsequent flow cytometry analysis. Blood was extracted from heart, collected in purple-capped K2EDTA-coated tubes and mixed immediately to prevent blood clots from forming. The samples were then centrifuged at 2000 RPM for 7 min at 4°C. The plasma supernatant was transferred to another tube and stored at -80°C.

For midkine treatment experiment rats were randomly assigned to vehicle and midkine treatment groups. Rat midkine recombinant protein was dissolved in ddH$_2$O and stored at -80°C for up to one month. Midkine was injected intraperitoneally 3 times/week at 300 μg/kg for four consecutive weeks. All rats were sacrificed immediately after the final dose, and mammary tissue were collected for histology, flow cytometry and single cell sequencing analysis.

For NMU-induced mammary tumor experiment, Fischer rats were treated with rat recombinant midkine protein at 300 μg/kg 3 times/ week for two weeks and followed by a single intraperitoneal injection (i.p.) of 50 mg/kg NMU. Then, rats were treated with midkine protein for another two weeks with same injection frequency, followed by another single NMU injection. Rats were checked three times/ week, and tumor growth was monitored using caliper measurements. Maximum tumor size burden allowed for rats is 4 cm and this was not exceeded in any of the experiments. All the remaining rats were sacrificed upon the last tumor in midkine treatment group reached the size limit. Rats were euthanized by CO$_2$ inhalation. Tissues and tumors were saved for histology, flow cytometry and transcriptomic analysis.

Mammary organoid and cell culture

Mammary gland organoids were isolated and cultured as previous described.[8] In brief, freshly collected tissue was minced into small pieces and dissociated to single cells as described above. After 1 h digestion, the tissue was filtered through a 500 μm strainer, and the cells were harvested by centrifugation at 1,500 rpm for 5 mins. The resulting pellet was resuspended in an appropriate volume of Matrigel and promptly dropped into 24-well plate. After a 30-min solidification, 700 μL of full organoid culture medium (containing Advanced DMEM/F12 supplemented with 1% GlutaMax, 1% HEPES (1 M), 1% penicillin/streptomycin, 0.1% Amphotericin B, 1X B27, 500 ng/mL recombinant R-spondin, 1.25 mM N-acetylcystine, 10 mM Nicotinamide, 1 μM SB202190, 5 ng/mL FGF7, 20 ng/mL fibroblast growth factor 10 (FGF10), 5 nM Heregulin beta-1, 5 ng/mL epidermal growth factor (EGF)) was added on top of the 3D matrix. Subsequently, the medium was changed every 3 days. The rat breast cancer cell line LA7 was obtained from ATCC. This cell line was maintained in DMEM high glucose supplemented with 5% FBS, 0.005 mg/mL insulin, 50 nM hydrocortisone and 20 mM HEPES, cultured in a humidified incubator at 37°C with 5% CO$_2$ and were regularly tested and verified to be free of mycoplasma contamination. Reagents used in organoid culture are listed in key resources table.

Immunoblotting

Cells were lysed in SDS buffer containing 4% SDS, 100 mM Tris-HCl (pH 7.0) and 4% 2-mercaptoethanol. Proteins were resolved in SDS-polyacrylamide gels (4-12%) and transferred to PVDF membranes using a Tris-glycine buffer system. Membranes were blocked with 5% milk powder in Tris Buffered Saline with 0.1% Tween 20 (TBST) for 1 h at room temperature (RT), followed by incubation with primary antibodies in antibody diluent (LI-COR) overnight at 4°C. Then, after being washed for three times with TBST, the membrane was incubated with secondary antibodies at RT for 1 h. Finally, immunoreactive bands were detected with Clarity Max Western ECL Substrate (Bio-Rad) and imaged using ChemiDoc MP imaging system (Bio-Rad). Antibodies used in immunoblotting are listed in key resources table.

Immunofluorescence staining

Following deparaffinization and rehydration, formalin-fixed paraffin-embedded (FFPE) slides underwent antigen retrieval in Tris/EDTA buffer (pH 6 or 9, Dako) for 45 minutes within a steamer. A blocking solution (10% goat serum) was applied for 1 hour at RT. Incubation with primary antibody was conducted overnight at 4°C within a moist chamber. Subsequently, corresponding secondary antibodies were applied for 1 hour at RT. Slides were then mounted with VectaShield HardSet Antifade Mounting Medium with DAPI (Vector Laboratories). The images were taken using a Zeiss 980 Confocal or THUNDER Imager 3D, then analyzed using QuPath (version 0.4.0). For immunofluorescence staining of KRT14 and KRT18, organoids in the 3D matrigel structures were embedded in Histogel (Epredia) to make blocks. Then the blocks were sectioned and stained following the FFPE protocol described above. Antibodies used in immunofluorescence staining are listed in key resources table.

Histology

FFPE tissue sections were deparaffinized and stained with H&E by the Rodent Histopathology Core in Harvard Medical School. Normal mammary gland and mammary tumor histology were determined by an experienced rodent pathologist. H&E slides were imaged using the Olympus VS120 Virtual Slide Microscope and visualized and quantified using QuPath software. Stroma area in normal mammary gland and tumor cell cellularity were quantified using H&E images in Qupath.

Plasma hormone ELISA

Plasma was collected as described in the above and stored in -80°C before used for ELISA. ELISA was performed using diluted and undiluted plasma according to the manufacturer's instructions for estradiol, prolactin and progesterone ELISA kits (Cayman Chemical). Absorbances were measured using BioTek Synergy 2 microplate reader (Agilent). The concentration of each sample was determined by using the equation obtained from the standard curve plot.

Flow cytometric analysis

Immune cell compositions in mammary glands, bone marrows and tumors were analyzed using flow cytometry. Mammary glands, tumors and bone marrows were harvested and dissociated as mentioned above. Lymphocytes and myeloid cells were analyzed using separate panels. Briefly, single-cell suspensions were initially stained with Live/Dead Aqua and Fc blocking antibody at 4°C for 20 mins. For lymphocyte panel, cells were stained with cell surface markers in FACS buffer (contained 0.5% BSA, 1% FBS and 4 mM EDTA) on ice for 1 h and followed by staining with intracellular marker post-fixation and permeabilization. The myeloid panel involved an initial incubation with unconjugated primary antibodies and corresponding secondary antibodies, succeeded by staining with the remaining conjugated antibodies. Unstained cells were used as negative controls, single antibody staining and Fluorescence Minus One (FMO) controls were used for gating controls. Cells were analyzed using BD LSR Fortessa cell analyzer. FlowJo (Becton Dickinson, version 10.8.2) was used for analysis and graphing. Gating strategies are shown in Data S1–S3. Antibodies used in flow cytometry are listed in key resources table.

Quantitative real-time PCR

RNA was isolated using RNeasy Mini Kit (Qiagen). 2 μg RNA per reaction was used for reverse transcription with PrimeScript RT Reagent Kit (Takara). TB Green Premix Ex Taq II (Takara) was used for real-time PCR on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad). Primer sequences for *Actb* and *Mdk* are indicated in key resources table.

Single cell RNA-seq library preparation

For whole mammary gland scRNA-seq, viable cells were resuspended in PBS with 0.04% BSA at a cell concentration of 1000 cells/μL. 10,000 cells were loaded onto a 10x Genomics Chromium™ instrument (10x Genomics) according to the manufacturer's instructions. The scRNA-seq libraries were prepared according to the Chromium Next GEM Single Cell 5' HT v2 protocol (10x Genomics). Quality control for amplified cDNA libraries and final sequencing libraries were performed using Bioanalyzer High Sensitivity DNA Kit (Agilent). The pooled libraries were sequenced on Illumina NovaSeq S4 platform.

Single cell ATAC-seq library preparation

Single cell ATAC-seq experiments were conducted on the 10x Chromium platform using the Chromium Next GEM Single Cell ATAC Kit v2 following established procedures.[92] Briefly, digested single cells from mammary gland of different aged rats were subjected to nuclei isolation according to the protocol of the manufacturer. Approximately 15,000 nuclei were selected for tagmentation. After tagmentation, the nuclei were processed for capture using the 10x Chromium controller. Gel emulsion generation was followed by linear amplification and DNA purification, performed according to the manufacturer's protocol. The resulting DNA was used for library construction as described on the manufacturer's website. Finally, libraries were quantified by quantitative PCR and were sequenced on an NovaSeq 6000 sequencer.

Bulk RNA-seq library preparation

Tumors were collected at the time of dissection and snap frozen in liquid nitrogen. RNA was isolated using RNeasy Mini Kit (Qiagen) with on-column DNA digestion. Libraries preparation with poly(A) selection and 150-bp paired-end sequencing on an Illumina HiSeq platform were performed by GENEWIZ (USA; www.GENEWIZ.com).

Single-cell RNA-seq data analysis

Data processing and cell type annotation

Sequenced reads were processed and aligned to the rat reference genome Rnor_6 using 10x Genomics Cell Ranger 6.1.2. We applied stringent filters to eliminate cells with (1) nFeature_RNA < 500, (2) percent_mito > 15%, and (3) percent_ribo < 5%. This pre-filtering resulted in the detection of 17,818 genes in 166,433 cells. The filtered data were then imported into Seurat (version 4.3.0)[84] and normalized using SCTransform and decomposed using PCA. Neighborhood graph computing was performed FindNeighbors() using the first 20 dimensions and clusters were determined using FindClusters() at a resolution of 0.4. UMAP embeddings were then computed using the first 20 PCA dimensions as input. Clusters 15 and 27 were removed from further analysis due to those clusters only appeared in one out of sixteen samples. Multiple classic cell type specific markers were used jointly to

annotate each cluster. Go term analysis of each cell type specific genes were also employed to further confirm the accuracy of annotation. Cell subtypes were distinguished by evaluating the expression of well-known markers and the top DEGs defining each cluster. To assess differential expression, the "FindMarkers" function in Seurat was employed with the default Wilcoxon rank-sum test. The lists of differentially expressed genes were typically refined based on a corrected p value threshold ('p_val_adj' $\leq$ 0.05) and absolute Log_2(Fold change) $\geq$ 0.25.

Details of each scRNA-seq sample:

scRNA-seq sample	Age	Analyzed cell number
3M-1	3 months old	11,175
3M-2	3 months old	13,129
3M-3	3 months old	7,672
3M-4	3 months old	10,139
6M-1	6 months old	11,837
6M-2	6 months old	6,209
6M-3	6 months old	12,693
6M-4	6 months old	10,678
12M-1	12 months old	9,734
12M-2	12 months old	10,927
12M-3	12 months old	6,119
12M-4	12 months old	10,997
22M-1	22 months old	10,174
22M-2	22 months old	9,024
22M-3	22 months old	10,195
22M-4	22 months old	1,2376
Total		16,3078

Generation of aging-LP signature
The aging-LP, Basal, LP and ML signatures (Table S2) were generated by considering positive differentially expressed genes ('p_val_adj' $\leq$ 0.05) with certain pct.1 and pct.2 cutoffs among all epithelial subtypes. The proliferation signature was composed of the following genes: *Mki67*, *Top2a*, *Pcna*, *Cdk6*, *Ccnb1*, *Ccne1*, *Plk1*, *Mybl2*, *Foxm1*, *Bub1*, *Cdc6*, *Ccna2*, *Ccnf*, and *E2f3*. The expression of the signature in a cell was quantified by using the "AddModuleScore" function of the Seurat R package.

Identification of aging-dependent DEGs
To identify aging-dependent DEGs, we used the function of "FindAllMarkers" in Seurat to identify age-dependent DEGs of different ages (3-, 6-, 12-, 22-month-old) for each cell type. Cell types with less than 3 cells were removed before differential expression analysis. The $\log_2$fold change of each DEG and the adjusted p value were tested by non-parametric bilateral Wilcoxon rank sum test. Only those with absolute Log_2(Fold change) $\geq$ 0.25 and 'p_val_adj' $\leq$ 0.05 were considered as aging-dependent DEGs. To generate the age-dependent heatmap, first, we used the age-dependent DEGs calculated to define an expression matrix, and the cells of each cell type are sorted by age scale. The ComplexHeatmap package was then used to visualize the spline smooth expression pattern of aging-dependent DEGs and the genes were divided into 6 groups based on their expression pattern. Finally, the up or down aging-dependent DEGs was determined based on general expression trend. The DEGs are listed in Tables S1 and S2.

Gene ontology (GO) and gene set enrichment analysis (GSEA)
GO analysis of DEGs was performed through Metascape (https://metascape.org/)[82] and visualized with GraphPad Prism (Version 10.0.3). GSEA for the DEGs were conducted using clusterProfiler package.[93] Hallmark gene set from MsigDB[94] (version 7.5.1) were used for GSEA. p values were adjusted by the Benjamini–Hochberg method.

Pseudotime analysis
Pseudotime analyses for T cells (CD4$^+$ and CD8$^+$ T cells) and epithelial cells were performed with the R package monocle3 (version 1.2.9).[33] Briefly, scRNA-seq data was visualized with UMAP, cells were clustered, and the graph was learned based on the reduced dimensions using the learn graph function. Subsequently, cells were ordered using the "order_cells" function where root nodes were set to the SBC (for epithelial trajectory analysis) and naïve T cell (for T cell trajectory analysis). Finally, the visualization functions "plot_cells" were used to plot each cell type along the same pseudotime trajectory.

Cell-cell communication analysis
Cell-cell communication analysis was conducted using CellChat[89] (version 1.6.1) and NicheNet[37] (version 1.1.1) packages. CellChat was used to quantify the total number of interactions among all the epithelial subtypes, then the major sources of ligands to LP were

further analyzed using NicheNet. In brief, the total epithelial expression data were imported into CellChat, and a CellChat object was made using "createCellChat" function. After annotating the object with relevant labels and identifying overexpressed genes, the communication probability was inferred using "computeCommunProb" function. Cell–cell communications for each cell signaling pathway were generated with "computeCommunProbPathway" function. The circle plot of number of interactions were generated using "netVisual_circle" function. The interaction among Basal, SBC, Bridge and LP were further studied using "NicheNet" package. The Basal, SBC and Bridge cells were chosen as "sender" populations, and LP (including aging-LP) was set as "receiver" population. For ligands and receptor interactions, genes which are expressed in larger than 10% cells of clusters were considered. Top 50 ligands and top 250 targets of differential expressed genes of "sender" and "receiver" were extracted for paired ligand-receptor activity analysis. Finally, expression and scaled activity of top 15 ligands were plotted using "make_ligand_activity_target_exprs_plot" function, and the circle plot showing top 15 ligand-receptor pairs were plotted using "make_circos_lr" function.

Transcriptional regulatory network analysis

Core regulatory transcription factors of aging-LP were predicted based on the scRNA-seq data by the GENIE3 (version 1.22.0)[95] and RcisTarget (version 1.20.0) R packages of the SCENIC workflow (version 1.3.1).[49] GENIE3 inferred gene regulatory networks from gene expression matrix of luminal progenitors (including aging-LP) from all four age groups. RcisTarget was used to identify enriched transcription factor-binding motifs and to predict candidate target genes (regulons) based on the RcisTarget database containing motifs with genome-wide rankings. The transcription factor and target gene networks were visualized with Cytoscape (version 3.9.1).[83]

Single-cell RNA-seq analysis of iHBCA dataset

The iHBCA dataset was retrieved from CELLxGENE (https://cellxgene.cziscience.com). Milk-derived epithelial cells in this dataset were removed before analysis. iHBCA dataset was analyzed using Scanpy. Samples were classified based on age (young, <50 years old; old ≥50 years old) or germline mutation types. The results were illustrated using Python code around Scanpy.

Single-cell ATAC-seq data analysis

The scATAC-seq data was initially processed using the Cell Ranger ATAC pipeline (version 2.0.0). This involved converting BCL files into fastq format and demultiplexing reads using "cellranger-atac mkfastq." The subsequent step, "cellranger-atac count," aligned Tn5-cut sites to the rat genome (Rnor_6), removed duplicate reads, and filtered out background cells. The resulted peak-by-cell matrices and barcoded fragment files were then imported into the Signac package in R.[84] Within Signac, peak calling was performed, and the data underwent quality control checks. Latent semantic indexing (LSI) was used to reduce noise, and UMAP embeddings were computed to identify clusters representing different cell types. Known marker genes were used to annotate these clusters, and motifs were added to the peaks. Finally, "RunChromVar" function was used to assess motif activity across all cells. Details of each sample:

scATAC-seq sample	Age	Analyzed cell number
3M-1	3 months old	5,476
3M-2	3 months old	8,684
6M-1	6 months old	5,178
6M-2	6 months old	5,619
12M-1	12 months old	2,834
12M-2	12 months old	3,402
22M-1	22 months old	5,947
22M-2	22 months old	9,586
Total		46,726

Bulk RNA-seq data analysis

RNA-seq datasets were analyzed using VIPER pipeline.[96] In brief, paired-end reads were aligned to the rat reference rn6 genome. Genes with 0 counts across all samples were filtered out and the remaining counts were then normalized using TMM with edgeR.[97] Log$_2$transformed TMM-normalized counts per million [log$_2$(TMM-CPM + 1)] were used for analysis. Batch effects among different biological replicates were removed using removeBatchEffect function from the LIMMA package.[98] Principle component analysis were performed using prcomp function provided by Stats package. DEGs were identified by using DESeq2[90] with cutoff of Log$_2$(Fold change) $\geq$ 0.5, padj $\leq$ 0.05 for all the samples.

Survival analysis

For the survival analysis in METABRIC data set, normalized probe intensity values were obtained from Synapse (Syn1688369). For genes with multiple probes, probes with the highest inter-quartile range (IQR) were selected to represent the gene. Patients were assigned an ER$^+$/HER2$^-$ status based on the downloaded clinical annotation files with "Negative" annotation for HER2 and "Positive"

annotation for ER based on IHC evaluation. Enrichment levels of *MDK* and "MDK-age" signature in each tumor sample were quantified by "GSVA" R package after converting all gene names to human homology using "gprofiler2" R package. Old and young patients were defined using the cutoff of 55 years old. Top (high) and bottom (low) 25% patients ranked by *MDK* expression or "MDK-age" signature" enrichment level were further overlayed by age groups (old and young) and all four groups were subjected to survival comparison using the disease-specific survival outcome. Survival analysis was performed using Cox proportional hazards regression models using "survminer" package and censored to 9,000 days. Patients in "Old high" group was used as a reference, harzard ratios with 95% confidence intervals and log-rank *p* values were reported for other subgroups. The Hallmark gene signature collection enrichment scores were calculated by "GSVA" package from each tumor. In addition, the proliferation signature score of each tumor from TCGA dataset was calculated by the average enrichment score of "G2M_Checkpoint", "E2F_Targets" and "Mitotic_Spindle" signatures from the Hallmark collection.

For breast cancer risk analysis, RNA-seq data with RSEM TPM values from 151 normal human mammary glands and the associated Gail 5-year risk scores were obtained from Kang et al.[48] Old and young patients were defined using the cutoff of 45 years old. *MDK* high and low groups were defined based on the top and bottom 25% populations.

QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis

Bioinformatic statistical tests are performed in R or Python. Statistical tests are described in the relevant sections above. Statistical analyses were performed by using GraphPad Prism 10 software or R. All the *p* values were labeled in the figures. $p < 0.05$ is considered statistically significant.