Luminal epithelial cells integrate variable responses to aging into stereotypical changes that underlie breast cancer susceptibility

Rosalyn W. Sayaman[1,2,4,6,7,8]*, Masaru Miyano[1,2,8], Parijat Senapati[3], Arrianna Zirbes[1,5], Sundus Shalabi[1,5], Michael E. Todhunter[1,2], Victoria Seewaldt[1], Susan L. Neuhausen[1], Martha R. Stampfer[7], Dustin E. Schones[3], and Mark A. LaBarge[1,2,7,9]*

[1] City of Hope, Department of Population Sciences, Beckman Research Institute, Duarte, CA, 91010, USA

[2] City of Hope, Center for Cancer and Aging, Beckman Research Institute, Duarte, CA, 91010, USA

[3] City of Hope, Department of Diabetes Complications and Metabolism, Beckman Research Institute, Duarte, CA, 91010, USA

[4] City of Hope, Cancer Metabolism Training Program, Beckman Research Institute, Duarte, CA, 91010, USA

[5] City of Hope, Irell and Manella Graduate School of Biological Sciences, Duarte, CA, 91010, USA

[6] Current: Department of Laboratory Medicine, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA 94143, USA

[7] Lawrence Berkeley National Lab, Biological Sciences and Engineering, Berkeley, CA 94720, USA

[8] These authors contributed equally

[9] Lead contact

* Correspondence: mlabarge@coh.org (MAL), rwsayaman@gmail.com (RWS)

Abstract:

Effects from aging in single cells are unpredictable, whereas aging phenotypes at the organ- and tissue-levels tend to appear as stereotypical changes. The mammary epithelium is a bilayer of two major phenotypically and functionally distinct cell lineages, the luminal epithelial and myoepithelial cells. Mammary epithelia exhibit substantial stereotypical changes with age that merits attention because they are putative breast cancer-cells-of-origin. We hypothesize that effects from aging that impinge upon maintenance of lineage fidelity increases susceptibility to cancer initiation. We identified two models of age-dependent changes in gene expression, directional changes and increased variance, which contributed to genome-wide loss of lineage fidelity. Age-dependent variant responses were common to both lineages, whereas directional changes were almost exclusively detected in luminal epithelia and implicated downregulation of chromatin and genome organizers such as *SATB1*. Epithelial expression of gap junction protein *GJB6* increased with age, and modulation of *GJB6* expression in heterochronous co-cultures revealed that it provided a communication conduit from myoepithelial cells that drove directional change in luminal cells. Age-dependent luminal transcriptomes comprised a prominent signal detectable in bulk tissue during aging and transition into cancers. A machine learning classifier based on luminal-specific aging distinguished normal from cancer tissue and was predictive of breast cancer subtype. We speculate that luminal epithelia are the ultimate site of integration of the variant responses to aging in their surrounding tissue and that their emergent aging phenotype both endows cells with the ability to become cancer-cells-of-origin and embodies a biosensor that presages cancer susceptibility.

Keywords

aging; breast cancer; cancer susceptibility; lineage fidelity; luminal epithelia; variance; gene expression; differential expression; differential variability; machine learning; PAM50 subtype predictors; *SATB1*; *GJB6*

Introduction

The stereotyped aging phenotypes exhibited by organisms, organs, and tissues represent the integration of accumulated, stochastically incurred damages to individual cells that result in commonly understood hallmarks of aging (López-Otín et al., 2013; Todhunter et al., 2018). Age-associated directional changes in transcriptomes of whole tissues are well documented (de Magalhães et al., 2009; Glass et al., 2013; Peters et al., 2015; Volkova et al., 2005). These directional molecular changes explain, at least in part, the noticeable phenotypic changes that accompany aging. However, while increased susceptibility to a plethora of diseases, including cancers is a prominent consequence of aging, the manifestation and onset of diseases vary between same-aged individuals. Indeed, variance between individuals arises in the contexts of tumors, diet, and aging (Bashkeel et al., 2019; de Jong et al., 2019; Slieker et al., 2016; Xie et al., 2011). We propose that the variability across individuals may itself be an important molecular phenotype of aging, and individuals with outlier expression profiles provide an avenue for understanding biological processes that explain the differential cancer-susceptibility of aged individuals.

The breast is an excellent model system for examining aging at the molecular and cellular levels because normal tissue across the adult lifespan is available from common cosmetic and prophylactic surgeries. Cultured pre-stasis human mammary epithelial cells (HMEC) supports growth of all lineages from women across the lifespan (Garbe et al., 2009; Labarge et al., 2013) and enable detailed and reproducible molecular studies of cancer progression (Stampfer et al.,

2013). Well-established lineage-specific markers and cell-sorting protocols facilitate experimentation at lineage-specific resolution. Furthermore, breast tissue provides an ideal model for studying aging-associated cancer susceptibility as 82% of new breast cancers are diagnosed in women ≥50y (DeSantis et al., 2019). Directional changes in gene expression with age were reported in whole breast tissue, including changes associated with breast cancer biological processes (Lee & Lee, 2017; Yau et al., 2007). However, aging is also associated with significant shifts in proportions of breast cell lineages, including epithelial and stromal populations (Benz, 2008; Garbe et al., 2012), so it is unclear how tissue-level molecular changes in normal aging reflect changes in cell-intrinsic and microenvironment states. Lineage-specific analyses are necessary to unravel such mechanisms.

The mammary epithelium, the origin of breast carcinomas, is a bilayer of two major phenotypically and functionally distinct cell lineages. Myoepithelial cells (MEPs) are basally located, contractile and have tumor suppressive properties (Pandey et al., 2010). Luminal epithelial cells (LEPs) are apically-located, secretory and include subpopulations of hormone receptor positive cells (Booth & Smith, 2006). We previously demonstrated loss of lineage fidelity as an aging phenomenon – where the faithfulness of expression of established lineage-specific markers diminishes with age without loss of the lineage-specificity of other canonical markers and the gross phenotypic and histological differences between LEPs and MEPs (Miyano et al., 2017). We hypothesize that the aging mechanisms that impinge upon the genome-wide maintenance of lineage fidelity are drivers of susceptibility to cancer initiation in breast tissue.

Here we demonstrate how age-dependent directional and variant transcriptional responses integrate in breast epithelia and explain how these changes could lead to increased susceptibility to cancer initiation. Through transcriptomic profiling of primary LEPs and MEPs we showed that loss of lineage fidelity in gene expression with age was a genome-wide phenomenon. We identified two models mediating loss of lineage fidelity in breast epithelia with age: (i) via directional changes identified through differential expression (DE); or (ii) via an increase in

variances identified through differential variability (DV) analysis. Age-dependent DE explained part of the observed loss of lineage fidelity, while our model of the overall increase in variances with age also accounted for a comparable fraction of this loss. Directional changes in expression with age strikingly occurred almost exclusively in luminal cells, whereas changes in variance were found in both epithelial lineages. The genome-wide directional changes in LEPs involved dysregulation of chromatin and genome organizers such as *SATB1* with age, which we also detected in bulk tissue. Loss of lineage fidelity led to enrichment of genes and biological processes commonly dysregulated in cancers, and alteration of the LEP-MEP interactome that was significantly modulated by apical cell-cell junction proteins, such as *GJB6*. Modulation with shRNA *GJB6* in MEPs was sufficient to reduce the rate of molecular aging of adjacent LEPs as determined with a breast specific biological clock. Using machine learning, we showed that genes with age-dependent directional and variable changes in normal LEPs had predictive value in distinguishing normal breast tissue from breast cancers and classifying breast cancer PAM50 subtypes. Age-dependent changes in LEPs reflected dysregulation of biological processes that are convergent with breast cancer. The degree and variability of age-dependent changes across individuals may explain the differential susceptibility of specific individuals to breast cancer initiation, and to the development of specific breast cancer subtypes.

Results

Genome-wide loss of lineage-specific expression in breast epithelia with age

LEPs and MEPs were enriched from 4th passage finite-lifespan HMEC from reduction mammoplasties from two age cohorts: younger <30y women considered to be premenopausal (m_{LEP}=16, m_{MEP}=16, samples, n=11 subjects, age range 16-29y) and older >55y women considered to be postmenopausal (m_{LEP}=11, m_{MEP}=11, n=8, age range 56-72y) (**Figure 1—table supplement 1**). We analyzed the expression of 17,328 genes with comparable dynamic ranges

and consistent lineage-specific expression between primary organoid and 4th passage LEPs and MEPs in both age cohorts (linear regression R^2=0.88 to 0.91, p<0.0001) (**Figure 1—figure supplement 1A-1D**). To understand how lineage fidelity of the two epithelial cell types (**Figure 1A**) changes with age, we performed differential expression (DE) analysis comparing LEP and MEP expression in younger <30y and older >55y women. DE genes between LEPs and MEPs decreased with age (adj. p<0.05, <0.01, <0.001) (**Figure 1—figure supplement 1E-1F**). Restricting analysis to genes with strong lineage-specific bias (DE adj. p<0.001, lfc≥1), we found 4,040 genes (23% of all genes analyzed) with highly significant lineage-specific DE in younger women– of which 59% were LEP-specific and 41% were MEP-specific. In contrast, 3,345 genes had highly lineage-specific DE in older women – of which 56% were LEP-specific and 44% were MEP-specific. Shifts in lineage-specific expression with age were illustrated in the strata-plot in (**Figure 1B**). Loss of lineage-specific expression with age occur genome-wide and was detected in 1,022 genes – a majority of which (65.5%) were LEP-specific.

Lineage fidelity is the loss of the faithful expression of lineage-specific markers with age. Statistically, we described this loss as a phenomenon whereby the magnitude of gene expression differences that distinguish LEPs from MEPs decreased with age, which is seen as shifts in distributions of fold changes between lineages to smaller values in the older cohort (Kolmogorov-Smirnov two-sample test, KS p<0.0001) (**Figure 1C**). We found that 76% of LEP-specific genes and 63% of MEP-specific genes had higher fold changes between lineages in younger cells compared to older cells (**Figure 1D**). These percentages indicated loss of lineage fidelity was not restricted to genes that lost lineage-specific expression. Indeed, within the subset of genes where lineage-specific DE was maintained with age by significance threshold, the majority – 67% of LEP-specific genes and 54% of MEP-specific genes, still showed larger fold differences between LEPs and MEPs in younger women (**Figure 1E**). These data expand on our earlier findings that identified loss of lineage fidelity in a limited set of lineage-specific probes (Miyano et al., 2017),

and underscore the genome-wide nature of this phenomenon whereby gene expression differences that distinguish the major epithelial lineages of the breast decrease with age.

Loss of lineage fidelity with age leads to disrupted lineage-specific signaling

Because loss of lineage-specific expression could upset the relative balance of ligands and receptors in each lineage, we explored how loss of lineage fidelity could lead to disrupted or dysregulated cell-cell communication between neighboring cell types. We defined the breast interactome as a set of possible ligand-receptor interactions between cell populations based on the DE of cell-specific ligands and their cognate receptors in younger women. Using published ligand-receptor pairs (LRPs) (Ramilowski et al., 2015), we identified 224 candidate lineage-specific LRPs in younger LEPs and MEPs based on the DE of 62 LEP-specific and 66 MEP-specific ligands, and 45 LEP-specific and 47 MEP-specific cognate receptors (**Figure 2—figure supplement 1A**). Protein-protein interaction (PPI) functional enrichment of lineage-specific LRPs identified top KEGG canonical biological pathways (FDR $p<0.001$) (**Figure 2—figure supplement 1B-1C**), with ligands and receptors related to cytokine-cytokine receptor interaction, PI3K-Akt, MAPK and Rap1 signaling commonly enriched in LEPs and MEPs. Enrichment of cytokine, immune and infection-related pathways further suggested lineage-specific interactions between epithelial and immune cells. LEP-specific LRPs were enriched for cell adhesion molecules (CAMs) involved in cell-cell and cell-ECM interactions, and axon guidance molecules (AGMs), while MEP-specific LRPs were enriched for ECM-receptor interaction and focal adhesion LRPs.

Loss of lineage fidelity with age led to disruption of 74 LRPs based on the loss of lineage-specific expression of ligands and/or their cognate receptors (**Figure 2A**). For each lineage, we considered KEGG canonical biological pathways (FDR $p<0.01$) that were likely to exhibit dysregulated signaling either through direct disruption of the LRPs due to loss of lineage-specific

signaling of the ligand or receptor, or indirect loss of signaling homeostasis via dysregulation of its cognate pair (**Figure 2—figure supplement 1B-1C**). Loss of lineage-specific expression of LEP LRPs with age was enriched for canonical pathways involved in (1) cell-cell and cell-ECM interactions including CAMs, AGMs, and adherens junctions, and (2) cytokine, immune and infection-related pathways. Loss of lineage-specific expression of MEP LRPs with age were associated with (1) pathways in cancer; (2) pathways involved with MAPK, EGFR, NOTCH and PI3K-AKT signaling; and (3) MEP contractility. These findings suggest that loss of lineage fidelity with age has the potential to affect a wide range of biological processes regulating lineage-specific function and signaling, including potential dysregulation of cancer-related processes and immune-specific signaling by the epithelia.

Models of loss of lineage fidelity in breast epithelia

To understand the changes within each cell population that contribute to the observed aging-associated loss of lineage fidelity, we explored two models that could explain the decrease in DE between LEPs and MEPs with age. The first model took into account age-dependent directional changes either through stereotypic, up- or down-regulation, leading to loss of lineage-specific expression – e.g., LEPs acquire MEP-like expression patterns and/or MEPs acquire LEP-like expression patterns in the older cohort (**Figure 3Ai**). The second model considered aging-associated increase in variances in the expression of lineage-specific genes in LEPs and/or MEPs from older women, leading to a loss of detection of DE between lineages (**Figure 3Aii**). We describe the contributions of each in the following sections.

The luminal lineage is a hotspot for age-dependent directional changes

There was an extreme lineage bias in the numbers of DE genes between younger and older cells, with the majority of age-dependent changes occurring in LEPs. In LEPs, 471 genes

were DE as a function of age; in contrast, in MEPs only 29 genes were DE with age (adj. $p<0.05$) (**Figure 3B**). Moreover, we identified age-dependent changes that showed lineage independence with five genes: *LRRC4, PSORS1C1, SCNN1B, ZNF518B,* and *ZNF521* commonly DE across cell types, leaving only 24 genes changed with age exclusively in MEPs. That stereotypic directional changes associated with aging were almost exclusively found in LEPs suggests that this lineage could serve as a primary indicator of aging – a proverbial canary in the coalmine.

Age-dependent differential upregulation (251 genes) and downregulation (220 genes) of LEP gene expression (adj. $p<0.05$) occurred at comparable frequencies (**Figure 3C**). In LEPs, 82% of the genes that were DE changed in a direction towards acquiring MEP-like patterns with age (**Figure 3C**). Although changes in MEPs were far fewer, we note that shifts in expression and methylation in older MEPs led to LEP-like patterns (**Figure 3—figure supplement 1A**).

Because dysregulation of regulatory factors like transcription factors (TFs) could lead to further dysregulation of downstream targets, we compared TF expression between younger and older cells in each lineage. Expression of key TFs (Lambert et al., 2018) were significantly altered in older cells, with 50 TFs showing age-dependent DE in LEPs, and 4 TFs DE in MEPs (adj. $p<0.05$), the majority of which have known roles in breast cancer progression. Of these DE TFs in LEPs, 88% changed expression in older LEPs towards the direction of MEP-like expression (**Figure 3D**). These included highly expressed TFs in younger LEPs that were down-regulated with age, such as: LEP-specific TF *ELF5* (Miyano et al., 2021), *GRHL2*, *SGSM2*, *HES4*, *ZNF827*, and genome organizer *SATB1* (**Figure 3—figure supplement 1Bi-v**). Loss of *GRHL2* and *SGSM2* are associated with down-regulation of E-cadherin and epithelial-to-mesenchymal transition (EMT) in mammary epithelial cells (Lin et al., 2019; Xiang et al., 2012). *HES4* is a canonical target gene of Notch1 which plays an important role in normal breast epithelial differentiation and cancer development (Kontomanolis et al., 2018). *ZNF827* mediates telomere homeostasis through recruitment of DNA repair proteins (Vilas et al., 2018). And *SATB1* has genome organizing functions in stem cells and tumor progression (Kohwi-Shigematsu et al.,

2013). Several TFs also gained expression in older LEPs, including: *SP3* and *ZNF503* (**Figure 3—figure supplement 1Bvi-vii**). *SP3* silencing inhibits Akt signaling and breast cancer cell migration and invasion (Mansour, 2020). *ZNF503* inhibits *GATA3* expression, a key regulator of mammary LEP differentiation, and down regulation is associated with aggressive breast cancers (Kouros-Mehr et al., 2006; Shahi et al., 2017). Age-dependent dysregulation of TFs in LEPs may presage larger-scale changes through TF binding of gene regulatory regions of downstream targets in older LEPs.

To understand the contribution of age-dependent directional changes in driving loss of lineage fidelity (**Figure 3Ai**), we examined the overlap of age-dependent DE genes with genes that lost lineage-specific DE. Only 9% of the loss in lineage-specific DE was explained by age-dependent DE in LEPs or MEPs at adj. $p<0.05$. If we considered all genes with at least 2-fold change DE with age, these age-dependent directional changes accounted for only 21% of loss of lineage-specific expression events, leading to a significant decrease in the magnitude of expression fold changes between LEP- and MEP-specific expression in the older cells (**Figure 3E**). These findings suggest that other mechanisms play a substantial role in regulating lineage fidelity, and that molecular changes associated with aging are not limited to stereotyped directional changes.

Aging-associated increase in variance contributes to loss of lineage fidelity

Next, we explored the alternate model that incorporated measures of variance as an explanation for the loss of lineage-specific expression in older epithelia (**Figure 3Aii**). Gene expression means and variances (**Figure 3—figure supplement 1C-1F**) of LEPs and MEPs from younger cells were categorized into quantiles and corresponding categories in older cells were then assessed. Gene expression means shifted minimally between younger and older cells (**Figure 3—figure supplement 1C-1D**), whereas shifts in variances occurred at a much higher frequency (**Figure 3—figure supplement 1E-1F**). Though the dynamic ranges of gene

expression in LEPs and MEPs changed as a function of age, these changes were not stereotyped across individuals – i.e., different aged individuals had different sets of genes that deviated from the range of expression seen in younger samples.

Differential variability (DV) analysis identified 137 genes in LEPs and 128 genes in MEPs with significant age-dependent DV (adj. $p<0.05$) (**Figure 3F**). Twelve regulatory TFs in either LEPs or MEPs that had tuned windows of expression in younger cells were dysregulated in older cells through a significant increase in variance (adj. $p<0.05$) (**Figure 3—figure supplement 1G-1H**) and included *HES4, GLI1* and *KDM2B* in LEPs, and *HES6* in MEPs. *HES4*, which was also DE with age, is a known Notch target. *GLI1* activates the hedgehog pathway in mammary stem cells (Bhateja et al., 2019). Estrogen-regulated *HES6* is known to enhance breast cancer cell proliferation (Hartman et al., 2009). And lastly, *KDM2B (FBXL10)*, which was down-regulated in a subset of older women (**Figure 3—figure supplement 1Bviii**), is a histone demethylase ZF-CxxC protein that binds unmethylated DNA. These analyses suggested that age-dependent variability in expression across individuals can lead to differential outcomes as different downstream targets could be modulated in different individuals.

To understand how our model of age-dependent variability affected lineage-specific expression, we focused on genes that lost lineage-specific expression with age and that showed at least a 2-fold increase in variance in the older cohort (**Figure 3G**). Genes with 2-fold increases in variances with age explained 27% of the observed loss of lineage-specific expression events, on a par with the proportion (21%) explained by genes that had 2-fold changes in DE. Both of our models of directional and variant changes with age led to a significant decrease in the differential magnitude of LEP- and MEP-specific expression in the older cells (**Figure 3E, 3G**).

These data suggest that increased variances in transcription are considerable drivers of the loss of lineage fidelity in breast epithelia. The observed variances across the older cohort may underlie the age-dependent dysregulation of susceptibility-associated biological processes in specific individuals.

Hallmark pathways associated with cancer are dysregulated with age in luminal and myoepithelial lineages

Gene set enrichment analysis (GSEA) identified Molecular Signatures Database (MSigDB) hallmark gene sets that were dysregulated with age, including gene sets known to be dysregulated in breast cancers that were enriched in older LEPs and MEPs (**Figure 3H-3I**).

Seventeen hallmark gene sets were significantly modulated in LEPs (adj. $p<0.05$) based on DE (**Figure 3H** top). Three immune-related gene sets were enriched in younger LEPs and included genes up-regulated in response to interferon IFN-alpha and -gamma, and during allograft rejection. In contrast, 14 gene sets were enriched in older LEPs, which included: genes regulated by MYC; genes encoding cell-cycle related targets of E2F TFs and involved in the G2/M checkpoint; genes upregulated by mTORC1 complex activation and during unfolded protein response; and genes involved in the p53 and protein secretion pathways.

Twenty hallmark gene sets were significantly modulated in MEPs (adj. $p<0.05$) based on DE (**Figure 3I** top). Five gene sets were enriched in younger MEPs, including: MYC and E2F targets; and genes defining EMT. In contrast, 15 gene sets were enriched in older MEPs, including: genes involved in p53 pathways; genes down-regulated by KRAS activation; genes mediating programmed cell death by caspase activation (apoptosis); immune-related gene sets upregulated in response to IFN-alpha, IFN-gamma and by IL-6 via STAT3, genes regulated by NF-kB in response to TNF and genes encoding components of the innate complement system; and genes encoding components of apical junction complex.

Five gene sets were significantly modulated (adj. $p<0.05$) based on DV and were enriched in older LEPs (**Figure 3H** bottom). These included: E2F targets that were similarly enriched via DE; genes defining responses to inflammation; and genes upregulated in response to IFN-alpha and IFN-gamma – gene sets that in contrast were enriched via DE in LEPs of younger women.

Seven gene sets were significantly modulated (adj. $p<0.05$) based on DV and enriched in older MEPs (**Figure 3I** bottom). These included: genes involved in DNA repair and G2/M checkpoint; genes regulated by NF-kB in response to TNF – a gene set similarly enriched via DE; as well as MYC targets, E2F targets, and genes defining EMT – gene sets that in contrast were enriched via DE in MEPs from younger women.

Several enriched gene sets were involved in processes that were disrupted with age either via DE or DV in LEPs and MEPs, and such overlaps likely suggest integration of directional and variant responses and reflect their convergent impact in common biological processes. Furthermore, the divergence in the age-dependent DE and DV enrichment of cellular processes, such as MYC gene targets and genes involved in immunomodulatory signaling, suggests the genes that become variable with age are associated with pathways that are otherwise important in maintaining lineage-specificity and -function in younger cells.

Age-dependent directional changes in the luminal lineage are indicators of aging breast tissue

Because LEPs dominated the age-specific signal amongst epithelia, we examined if the age-dependent DE contribution of the luminal lineage was detectable in bulk normal primary breast tissue (GSE102088, n=114) (Song et al., 2017). Genome-wide analysis identified 97 genes to be DE between younger <30y and older >55y tissues (adj. $p<0.05$), the relatively smaller number of genes compared to age-dependent DE observed in LEPs likely due the cellular heterogeneity found in bulk tissue. To characterize the contribution of the LEP lineage to aging biology of the breast, we next performed gene set enrichment analysis (GSEA) to assess enrichment of LEP-specific age-dependent DE genes at tissue level. We found significant enrichment of differentially upregulated genes identified in young <30y LEPs in tissue from younger women (adj. $p=0.012$) (**Figure 4A**) and differentially upregulated genes identified in old

>55y LEPs in tissue from older women (adj. p=0.006) (**Figure 4B**). These GSEA results indicate that while age-dependent changes in other cell populations may confound detection of the LEP-specific signal, age-dependent changes in LEPs were still reflected in bulk tissue.

We then explored the GSEA leading edge genes – genes with the largest contribution to the significant enrichment of the LEP-specific age-dependent genes in bulk tissue. Of the leading edge genes, we found genome organizer *SATB1*, which showed significant LEP-specific expression compared to MEPs (adj. p<0.001, lfc≥1) (**Figure 4Ci**), to have the strongest signal in bulk tissue (**Figure 4Cii**). *SATB1* was significantly downregulated in both LEPs and breast tissue of older relative to younger women (adj. p<0.05) (**Figure 4Ci-ii**). This decrease with age in *SATB1* was also detected in normal breast tissue of women with cancer in the TCGA cohort (n=111, Wilcoxon adj. p<0.001) (**Figure 4—figure supplement 1A**). In TCGA breast cancers (n=1,089), PAM50 Luminal A (LumA), Luminal B (LumB) and Her2-enriched (Her2) breast cancer subtypes had the lowest expression of *SATB1* relative to PAM50 Basal-like (Basal) and Normal-like (Normal) intrinsic subtypes (KW adj. p<0.0001) (**Figure 4Ciii**). Moreover, in primary tumors with matched normal tissue (n=114 tumor and n=109 normal), we found *SATB1* to be significantly downregulated in PAM50 LumA, LumB and Her2 breast cancers relative to their matched normal tissue (**Figure 4—figure supplement 1B**). Together, these results suggest that *SATB1*-mediated genome organization may play a regulatory role in the maintenance of the luminal lineage and the observed genome-wide dysregulation with age and breast cancer.

Since we expected the signal in bulk tissue to be muted due to cellular heterogeneity, we also explored leading edge genes with nominally significant DE between younger and older tissue (unadj. p<0.05). Of the 251 genes upregulated in younger LEPs, 35 genes (14%) show nominally significant differential upregulation in young tissue (**Figure 4D**), including EMT-associated *GRHL2* (**Figure 3—figure supplement 1Ai**) and LEP-specific TF *ELF5* which we had previously shown to be predictive of accelerated aging in genetically high risk LEPs (Miyano, Sayaman et al., 2021). Of the 220 genes upregulated in older LEPs, 26 genes (12%) show nominally

significant differential upregulation in old tissue (**Figure 4E**), including GATA3 inhibitor *ZNF503* (**Figure 3—figure supplement 1Avii**). Of the 61 genes we identified to be commonly dysregulated between younger and older LEPs and breast tissue, 17 were LEP-specific and 14 were MEP-specific in our lineage-specific DE analysis.

Common age-dependent DE genes between LEPs and bulk tissue showed significant PPI network enrichment (PPI enrichment p=0.014), including a 51-gene network that involved 11 DE TFs (**Figure 4F**) and 10 genes with high connectivity in the network (degree >10) that are potential nodes of integration. These include genes downregulated in the older group: TF *BCL11A* – a subunit of the BAF (SWI/SNF) chromatin remodeling complex (Kadoch et al., 2013), TF *SOX4* – involved in determination of cell fate, TF *GRHL2*, *MLLT3* – a chromatin reader component of the super elongation complex (SEC) (Moustakim et al., 2018); and genes upregulated in the older group: *DPP4 (CD26) – a* cell surface receptor involved in the costimulatory signal essential T-cell activation (Ikushima et al., 2000), *HSPA12A* – heat shock protein associated with cellular senescence, and *KITLG* – a ligand for luminal progenitor marker c-KIT in breast (Kim & Villadsen, 2018). Community detection algorithm identified 7 communities (**Figure 4F**). Functional network enrichment (FDR<0.05) showed that Community 3 anchored by TFs *BCL11A* and *SOX4* was enriched for genes associated with transcriptional regulation. *SATB1*, *BCL11A*, *MLLT3* and *ZNF521* were linked to chromosomal rearrangement and were downregulated in LEPs and breast tissues of older women. These genes showed breast cancer subtype-specific expression, and *BCL11A* and *ZNF521* in particular were downregulated in PAM50 LumA, LumB and Her2 tumors relative to matched normal tissue (**Figure 4—figure supplement 2Ai-iii, 2Bi-iii**). Community 4 members *RARB*, *TIMP3* and *CDH13* have been implicated as tumor suppressor gene targets of DNA methylation and epigenetic regulation in cancers. In addition, community 7 members *PHC1* and *PCGF3* are components of the Polycomb group (PcG) multiprotein polycomb repressor complex (PRC)- PRC1-like complex which is required for maintenance of the transcriptionally repressive state of many genes throughout development. *PHC1* and *PCGF3* were downregulated

in LEPs and breast tissues of older women and showed breast cancer subtype-specific expression, with *PCGF3* downregulated in PAM50 LumB and Her2 tumors relative to matched normal tissue (**Figure 4—figure supplement 2Aiv-v, 2Biv-v**).

Taken together, genes commonly DE in younger and older LEPs and breast tissue either reflect stereotypic aging-associated molecular changes across different breast cell populations or are driven by LEP-specific changes, suggesting that age-dependent molecular changes in LEPs contribute to essential processes involved in the aging biology of the entire breast, and that are dysregulated in cancers.

Genes encoding for cell-cell junctional proteins are dysregulated in aging epithelia

We showed previously that MEPs can impose aging phenotypes on LEPs – with LEPs from younger women acquiring expression patterns of older LEPs when co-cultured on apical surfaces of MEPs from older women (Miyano et al., 2017). This non-cell autonomous mechanism of aging requires direct cell-cell contact between LEPs and MEPs, suggesting that cell-cell junctional proteins play a role in age-dependent dysregulation in LEP-MEP signaling. Indeed, we identified apical junction-associated genes to be significantly enriched with age in MEPs (**Figure 3I** top).

We explored age-dependent dysregulation of a curated set of genes encoding for membrane components of adherens junctions, tight junctions, gap junctions, desmosomes, and cell adhesion molecules (CAMs) in LEPs and MEPs to identify candidate genes that may regulate communication between the lineages. Because age-dependent changes involve both DE and DV, we performed the non-parametric Lepage test to jointly monitor the central tendency and variability of gene expression of 198 genes encoding for cell-surface junction proteins between the younger and older cohorts. We found 42 genes were modulated in LEPs and/or MEPs with age (Lepage test $p<0.05$) (**Figure 5—figure supplement 1A**). These include genes that were

modulated via a significant directional change with age such as the desmosomal cadherins genes, *DSG3* (desmoglein) and *DSC3* (desmocollin), which have been previously shown to be expressed in both LEPs and MEPs (Garrod & Chidgey, 2008) (**Figure 5—figure supplement 1Bi-ii**); and the genes encoding for essential tight junction components, *CLDN10* and *CLDN11* (**Figure 5—figure supplement 1Biii-iv**).

Gap junction *GJB6* (Connexin-30), which is expressed in both LEPs and MEPs in the normal mammary gland and forms homo- (LEP-LEP) and hetero-cellular (LEP-MEP) channels (Teleki et al., 2014) (**Figure 5A**), is of specific interest as it showed modulation via an increase in variance in older MEPs (p=0.02) and nominal increase in variance in older LEPs (p=0.06) (**Figure 5B**). As such, modulation of *GJB6* provided an avenue for exploring changes that could occur in both lineages and in only a subset of older women that may thus lead to differential susceptibility across aged individuals. To understand the transcriptional regulation of the GJB6 junctional protein, we explored the ChIP-seq (Cistromics) mammary gland data from The Signaling Pathways Project (SPP) Ominer database (**Figure 5C**). Nine TFs had binding signals within +/- 10 of the TSS of *GJB6*, including progesterone receptor PGR, MYC and the LEP-specific TF *ELF5* – which we have previously shown to be regulated via direct LEP-MEP interactions in co-culture studies (Miyano et al., 2017).

Gap Junction protein *GJB6* is a mediator of the non-cell autonomous mechanism of aging in breast

Because changes in MEPs are predominantly associated with DV rather than DE, we hypothesized that LEPs serve as integration nodes for dysregulation in MEPs where variant changes converge via common pathways leading to directional changes in genes downstream of these pathways (**Figure 5D**). We identified TF *ELF5* to be one such target; indeed, expression of the highly LEP-specific TF was dynamic and responsive to microenvironment changes (Miyano

et al., 2017), and serves as an independent biological clock in breast (Miyano et al., 2021). *ELF5* was downregulated in younger LEPs when co-cultured on apical surfaces of older MEPs for 10 days (**Figure 5E**), concordant with the observed phenomenon of *ELF5* downregulation in LEPs with age.

We asked whether knockdown or inhibition of *GJB6* expression in the subset of older MEPs with higher expression relative to younger MEPs could restore proper signaling between LEPs and MEPs. To test this, we used our established heterochronous co-culture system and used recovery of LEP expression of *ELF5* as a readout (**Figure 5F**). If bringing variant *GJB6* under tighter control prevents chronologically older MEPs from imposing older biological ages in younger LEPs, then *ELF5* levels should not decrease in co-culture. FACs-enriched LEPs from younger women were co-cultured for 10 days on older MEPs treated with either shGJBG or scramble shRNA (shCtrl) (**Figure 5F, Figure 5—figure supplement 1C**). When co-cultured on top of older MEP-shGJB6 relative to MEP-shCtrl, LEP-expression of *ELF5* was maintained at higher levels (**Figure 5H**), consistent with higher expression levels in younger women. LEP-expression of *ELF5* likewise showed a stepwise (though non-significant) increase when older MEP feeder layers were pre-treated with increasing concentrations of a non-specific gap junction inhibitor 18 alpha-glycyrrhetinic acid (18αGA) (**Figure 5—figure supplement 1D**). Thus, reducing the level or variance of *GJB6* prevented older MEPs from imposing an older biological age in younger LEPs as determined by *ELF5* expression.

These data suggest that variance is a driver of stereotypical aging phenotypes at the tissue level, and that constraining specific changes caused by an increase in molecular noise during aging – such as in cell-cell communication nodes, may prevent the spread of age-related cues amongst epithelia.

Age-dependent dysregulation in LEPs shape predictors of normal breast tissue and PAM50 subtypes

Our GSEA and literature review of genes with age-dependent changes in LEPs revealed enrichment for pathways and genes commonly dysregulated in breast cancers. Unsupervised hierarchical clustering of TCGA matched normal and primary tumor samples (n=1,201) based on the expression of these LEP-specific age-dependent DE and DV genes identified four main sample clusters (**Figure 6A**): (i) cluster 1 represented predominantly by PAM50 LumA and Her2 breast cancer subtypes; (ii) cluster 2 by PAM50 LumB and LumA subtypes; (iii) cluster 3 by PAM50 Basal subtype; and (iv) cluster 4 by matched normal samples. This suggests that age-dependent changes in LEPs may reflect dysregulation of biological processes that play a role in initiation of tumors from normal tissue, and in the etiology of breast cancer subtypes. We therefore assessed whether DE and DV genes that change in LEPs with age can be used as biomarkers that can classify normal tissue from cancer and predict breast cancer subtypes.

Using 75% of TCGA data for training and cross-validation (n=873), we built an elastic net machine learning (ML) classifier of normal breast tissue and PAM50 breast cancer subtypes based on the 536 age-dependent DE and DV genes identified in LEPs that were represented in our ML datasets. The best performing model selected during cross-validation had a mean balanced accuracy of 0.91, mean sensitivity=0.86, and mean specificity=0.96. Our ML classifier proved predictive in the remaining 25% of TCGA test data which the model had not seen (n=288, mean balanced accuracy=0.93, mean sensitivity=0.88, mean specificity=0.97), and in an independent test set composed of normal tissue from GTEx sand breast cancer tissue from GSE81540 (n=3,364, mean balanced accuracy=0.87, mean sensitivity=0.77, mean specificity=0.94) (**Figure 6B-6D**). We further assessed our ML model performance in the two test sets using three measures of the area under the curve (AUC) for multi-class prediction: (i) AUC of each group vs. the rest; (ii) micro-average AUC calculated by stacking all groups together; and

(iii) macro-average AUC calculated as the average of all group results (Wei & Wang, 2020). We found all per group vs. rest AUC to be > 0.9, and micro-average and macro-average AUC > 0.95 in both the TCGA (**Figure 6C**) and GTEx/GSE81540 test sets as annotated in (**Figure 6D**). In addition to accurately classifying PAM50 subtypes, LEP-specific aging biomarkers identified normal from cancer tissue 100% and 93.3% of the time respectively in the TCGA and GTEx/GSE81540 test sets.

We next identified the genes that contributed most to the ML predictor. We identified 127 genes with scaled variable importance contribution of ≥ 25% in the prediction of at least one class (**Figure 7A**), with 18% of predictors deriving from DV analysis. Of these, estrogen receptor *ESR1* downregulated in older LEPs and transmembrane protein *TMEM45B* upregulated in older LEPs are part of the 50-gene PAM50 subtype predictors with prognostic significance (Parker et al., 2009). We highlight the top five genes with the highest variable importance for each class: (i) Normal tissue – *FN1, ABCA10, HAS3, KLHL13* and *ACVR1C*; (ii) PAM50 LumA – *FAM198B, HSD17B1,* SEC chromatin reader component *MLLT3* discussed previously*, SERPINE1* and *PLIN2*; (iii) PAM50 LumB – *GPR108, TRIM29,* desmosomal cadherin *DSG3* discussed previously*, ESR1* and *ZNRF3*; (iv) PAM50 Her2 – *ESR1, TMEM45B, FA2H, CDK12* and *TMEM63C*; and (iv) PAM50 Basal – *SLC25A37*, *PGBD5*, *BAIAP2L2*, *TBX19* and *GPR161* (**Figure 7—figure supplement 1A-1E**). Top predictors of normal tissue and of PAM50 Her2 and PAM50 Basal subtypes showed larger differences in median expression relative to other groups. In contrast PAM50 LumA and LumB top predictors showed large differences in median expression compared to non-luminal subtypes but exhibited relatively smaller but nonetheless significant differences in median expression relative to each other (**Figure 7—figure supplement 1A-1E**). While expression of these predictive genes may not be specific to the LEP lineage alone, our findings suggest that age-dependent dysregulation of these genes in LEPs could disrupt lineage-specific signaling and the homeostatic control mechanisms of key biological processes that have been implicated in breast cancers.

Our results illustrate that age-dependent changes in LEPs embody biology that is relevant and contributes to tissue-level biology predictive of breast cancer subtypes. These changes may reflect age-related dysregulation convergent with development of frank tumors. The degree and variability of these age-dependent changes across individuals may explain the differential susceptibility of specific individuals to breast cancer initiation, and the development of specific breast cancer subtypes.

Discussion

Our analyses have shown that aging involves integration of directional and variant responses that reshape the transcriptomic landscape of the two main epithelial lineages of the breast, the LEPs and MEPs. These changes lead to a loss in lineage fidelity with age where faithfulness of lineage-specific expression is diminished. This is seen via the genome-wide loss of tuned windows of expression of lineage-specific genes and a decrease in the magnitude of differential expression between the genes that define LEPs and MEPs. Our approach delineated the contribution of each epithelial lineage and identified two models mediating this loss of lineage fidelity in breast epithelia with age – either via directional changes, as measured by differential gene expression; or via an increase in variance, as measured by differential variability analysis. Aging-dependent expression variances occur in both lineages, whereas the overwhelming majority of directional expression changes with age occurred in LEPs. This is a striking finding when one considers the two lineages arise from common progenitors, and with their juxtaposition in tissue allowing for direct cell-cell communication.

We hypothesize that LEPs are the nexus of integration for variant responses that lead to stereotyped directional age-dependent outcomes in mammary gland – the likely result of a dynamic process of iterative feedback between LEPs and MEPs, and other cell types in the breast. LEPs from older women still maintain canonical LEP-specific features, whereas they

exhibit genome-wide loss of lineage fidelity that implicate dysregulation of genes with known roles in breast cancer. This suggests that susceptibility to cancer entails loss of proper specification of the luminal lineage, and that age-dependent molecular changes in LEPs contributes to this loss. Age-dependent directional changes in LEPs are detectable in bulk tissue and implicated downregulation of chromatin and genome organizers such as *SATB1*, suggesting means by which loss of lineage fidelity may be perpetuated genome-wide. The pathways affected by transcriptomic changes during aging are commonly linked with breast cancer, and the age-dependent changes in LEPs reflect relevant biology that distinguish normal tissue from breast cancers and is predictive of PAM50 breast cancer subtypes. Together our findings illustrate how age-dependent changes in LEPs contribute to the aging biology of breast tissue, and we propose that this biology reflects dysregulation convergent with processes associated with breast cancers.

Aging studies of gene expression in human tissues have been largely restricted to analysis of bulk tissue, lacked cell-type specific resolution, and were focused on directional changes with age using DE analysis. Bulk analyses make it impossible to separate impacts of aging that are driven by the intrinsic changes that occur molecularly within each lineage versus compositional changes that reflect shifts in cell type proportions. Lineage-specific analyses provide intermediate resolution between bulk RNA-seq and single-cell RNA-seq and allows for cost effective analysis of cell population-level responses and interactions. As such, lineage-resolution analyses also provide an avenue to validate computational deconvolution methods that have emerged to extract cell-type specific contributions in bulk tissue (Shen-Orr & Gaujoux, 2013; Titus et al., 2017). We provided evidence that tissue-level changes with age are driven not just by changing compositions of the breast (Benz, 2008; Garbe et al., 2012), but by intrinsic molecular changes in the underlying cell populations. Indeed, while bulk tissue expression reflects cellular heterogeneity, we were able to identify age-dependent changes in bulk tissue that mirror the DE in LEPs with age, suggesting that LEPs contribute, if not drive, the certain emergent properties of breast tissue.

In addition to down regulation of genome organizer *SATB1* in LEPs and breast tissue (adj. $p<0.05$), we also identified another 60 genes that showed concordant directional changes in LEPs (adj. $p<0.05$) and in primary tissue (unadj. $p<0.05$) despite the difference in platforms (RNA-seq vs. microarray). Network and community analyses showed enrichment of genes involved in chromosomal rearrangement, including *BCL11A* – a subunit of the BAF (SWI/SNF) chromatin remodeling complex (Kadoch et al., 2013), *MLLT3* – a chromatin reader component of the SEC (Moustakim et al., 2018), and *PHC1* and *PCGF3* – components of the PcG multiprotein PRC-1-like complex required for developmental maintenance of repressed but transcriptionally poised chromatin configuration through alteration of chromatin accessibility, folding and global architecture of nuclear organization (Illingworth, 2019). Of note, *SATB1* was downregulated in PAM50 LumA, LumB and Her2 breast cancer subtypes relative their matched normal tissue and was expressed the lowest in luminal subtype cancers – the subtype most associated with aging. *SATB1* has genome organizing functions in tumor progression (Kohwi-Shigematsu et al., 2013), and has been described as a key regulator of EMT in cancers (Naik & Galande, 2019); however its role in normal breast epithelia remain to be elucidated. *BCL11A*, *MLLT3, ZNF521, PHC1* and *PCGF3* also show subtype-specific dysregulation in breast cancers, with *BCL11A*, *ZNF521 and PHC1* specifically downregulated in luminal subtype cancers. We speculate that chromatin and genome organization play a key role in the maintenance of the luminal lineage, and that their dysregulation may mediate loss of lineage fidelity observed genome-wide.

The striking phenotypic changes in LEPs are starkly juxtaposed to MEPs, which have so far revealed few obvious signs of changes with age. Nevertheless, heterochronous bilayers of MEPs and LEPs suggest that the chronological age of MEPs controls the biological age of LEPs, illustrating that MEPs do change with age and revealing the existence of a non-cell autonomous mechanism that integrates aging-imposed damage across the tissue (Miyano et al., 2017). Here, we showed that changes in MEPs largely involved changes in gene expression variances with age. Aging-associated increases in variances in both lineages drove a large fraction of the

observed loss of lineage fidelity in epithelia of older women, comparable to the contribution of age-dependent directional changes. In our opinion, changes in variance are an underappreciated component of aging analyses. DE analysis is a ubiquitous statistical tool used in the analysis of expression profiling studies, whereas only more recently have changes in variance been systematically analyzed (Bashkeel et al., 2019; de Jong et al., 2019; Slieker et al., 2016; Xie et al., 2011), along with the development of differential variance analytical tools (Phipson et al., 2016; Phipson & Oshlack, 2014; Ran & Daye, 2017).

For changes to be detected as significant in DE analysis, the following assumptions must be met: (i) the biological phenomenon causes dysregulation that is directional – e.g., genes are either up- or down-regulated; and (ii) dysregulation occurs at the same time – e.g., in the same genes in the same pattern, across the majority of individuals in the group of interest – i.e., it is stereotypic. While some aging processes may be deterministic, like telomere shortening, other processes may be stochastic, born out of the accumulation of random physicochemical insults that manifest as an increase in noise in the system (Todhunter et al., 2018). In the latter case, the signal itself is the noise in the system. Another way to view this type of dysregulation is by observing the deviation from a set range. A change in the dynamic range of expression, for instance, of regulatory genes such as TFs that have very tuned or narrow windows of expression, can lead to dysregulation as expression deviates from the set range. This noise can lead to decoupling of tightly regulated networks. While both increase and decrease in dynamic range with age do occur, here we specifically focused on increases in variance and its effect on the loss of lineage fidelity with age, hypothesizing that genes that have very tightly tuned windows of expression in younger healthy individuals and that see large increases in variance in older subjects, are good candidates for susceptibility factors that could be predictive of breast cancer risk. Accordingly, in single-cell studies, aged cells were shown to have increased transcriptional variability and loss of transcriptional coordination compared to younger cells of the same tissue (Enge et al., 2017; Kowalczyk et al., 2015; Levy et al., 2020; Martinez-Jimenez et al., 2017),

suggesting that increase in cellular heterogeneity with age underlies the population-level increases in variances between the individuals we observed. As such, the molecular signals of aging cells may not be fully captured as stereotyped directional changes – rather a large fraction of age-associated changes will be reflected as increases in measured variance in the molecular signal across an aged cohort.

We identified potential ligand-receptor pairs and junctional proteins including tight junctions, desmosomes, and gap junction components, that mediate dysregulated cell-cell and cell-microenvironment signaling within the epithelium. We provided experimental validation for the role of gap junction protein, Connexin-30 (*GJB6*) in mediating the ability of MEPs to impose an aging phenotype on LEPs (Miyano et al., 2017). It is unclear whether this occurs chemically through passage of ions or small molecules through gap junction channels, indirectly via gap junction-mediated structural proximity of LEPs and MEPs, or via signaling complexes with connexin-interacting proteins including cytoskeletal elements, tight and adherens junctions, and enzymes like kinases and phosphatases (Dbouk et al., 2009). How this occurs will require further exploration.

GSEA further identified age-dependent enrichment of gene sets in LEPs and MEPs that were commonly dysregulated in breast cancers including gene sets related to: inflammation and immunosenescence, processes synonymous with aging and cancer progression (Fulop et al., 2017); cell-cycle related targets of E2F transcription factors, which are thought to play a role in regulating cellular senescence (Lanigan et al., 2011); and targets of the oncogene *MYC.* Aging associated changes in the immune response were further implicated in our ligand-receptor pair analysis, where known immune-associated ligands and receptors exhibited loss of lineage-specific expression in breast epithelia. We showed previously that *in situ* innate and adaptive immune cell infiltration of the breast epithelia and interstitial stroma change with age (Zirbes et al., 2021) how this is linked to the dysregulation of epithelial signaling remain to be explored. Whereas our age-specific analyses did not identify oncogenes that were DE between younger

and older epithelia, gene set enrichment in LEPs and MEPs revealed a putative example of priming. We did not detect changes in Myc expression with age, but Myc targets were among the gene sets that were significantly enriched through DE or DV analysis. *Myc* is amplified or overexpressed in ~35% of breast cancers and exerts pleiotropic effects across the genome (Xu et al., 2010). In the context of cancer progression, Myc is able to induce telomerase activity, which enables bypass of the replicative senescence barrier in mammary epithelial cells (Garbe et al., 2014). While we do not have evidence for the direct involvement of Myc in this context, we speculate that secondary events such as demethylation at Myc binding sites at target genes could explain the enrichment of Myc relevant signatures.

Using machine learning, we built a predictive elastic net model using age-dependent DE and DV genes identified in LEPs that was able to classify normal breast tissue from breast tumors and predicted breast cancer subtypes in test sets of publicly available normal and cancer tissue transcriptomes from more >3,000 women. This illustrates how tissue-level predictive biomarkers of breast cancer with subtype-specific expression relative to matched normal samples are dysregulated with age at the cell population-level as we observed in the luminal lineage. The contribution of non-epithelial cell types to the age-dependent expression of these genes in bulk tissue remain the subject of future studies. Given that the mammary epithelium is the origin of breast carcinomas and age is the most significant risk factor for breast cancers, age-dependent changes in the transcriptomic landscape of luminal cells may be a key contributing factor to the tissue-level dysregulation of cell-cell and cell-microenvironment signaling in breast cancers and may reflect relevant biology convergent with the development of frank tumors. Indeed, the variance in expression of these genes across aged individuals may reflect the differential susceptibility of certain individuals to specific breast cancer subtypes.

Conclusions

Our studies culminate in the exploration of how directional and variant responses are integrated in breast tissue of older women that contribute to aging biology. We show that increased variance in the transcriptomic profiles of mammary epithelial lineages across individuals is a substantial outcome of aging and is likely central to our understanding of the increased susceptibility to breast cancers with age. Strikingly, LEPs are able to integrate age-dependent signals from MEPs and almost exclusively exhibit the stereotyped directional changes seen in aging epithelia that comprise a prominent signal detected in bulk tissue. Moreover, age-dependent directional and variant changes in LEPs can shape the tissue-level expression of predictive biomarkers that classify normal tissue and breast cancer subtypes, illustrating how age-dependent dysregulation may play a key role in the transition into frank cancers. We demonstrate how an increase in molecular noise during aging may lead to sufficient variance in the transcriptomes between aged individuals and propose that this is a mechanism for differential susceptibility to development of breast cancers. Because cancer susceptibility indicates a state that could be more easily pushed towards cancer initiation, we can consider the variances between aged individuals to occupy multiple metastable states, some of which represent susceptible phenotypes that can be perturbed and pushed towards cancer. We speculate that these are examples of age-dependent priming states susceptible to malignant transformation. Therefore, the degree to which breast-cancer associated genes are variably expressed across the different cell populations of the breast and across different individuals may explain why breast cancers develop in only a subset of women in a subtype-specific manner.

List of abbreviations

18αGA	18 alpha-glycyrrhetinic acid
adj. *p*	Adjusted *p*-value (test statistic)
AGM	Axon guidance molecule

AUC	Area under the receiving operator characteristic curve
BH	Benjamini-Hochberg
CAM	Cell adhesion molecule
ChIP	Chromatin immunoprecipitation
cor	correlation
DE	Differential expression or differentially expressed (in context)
DV	Differential variability or differentially variable (in context)
EMT	Epithelial-to-mesenchymal transition
FDR	False discovery rate
GEO	Gene Expression Omnibus
GSEA	Gene set enrichment analysis
GTEx	The Genotype-Tissue Expression Project
HMEC	Human mammary epithelial cells
KS	Kolmogorov-Smirnov test
KW	Kruskal-Wallis test
LEP	Luminal epithelial cells
lfc	$\log_2$ fold change (test statistic)
LRP	Ligand-receptor pair
MEP	Myoepithelial cells
ML	Machine learning
MSigDB	Molecular Signatures Database
PAM50 Basal	PAM50 Basal-like intrinsic subtype
PAM50	PAM50 Her2-enriched intrinsic subtype
PAM50 LumA	PAM50 Luminal A intrinsic subtype
PAM50 LumB	PAM50 Luminal B intrinsic subtype
PAM50 Normal	PAM50 Normal-like intrinsic subtype

PcG	Polycomb group
PPI	Protein-protein interaction
PRC	Polycomb repressor complex
rlog	Regularized log
SEC	Super elongation complex
SPP	Signaling Pathways Project
STRING	Search Tool for the Retrieval of Interacting Genes/Proteins
TCGA	The Cancer Genome Atlas
TF	Transcription factor
TSS	Transcription start site

Methods

Experimental Model and Subject Details

The FACS-enriched luminal epithelial (LEP) and MEP myoepithelial (MEP) cells from finite lifespan, non-immortalized human mammary epithelial cells (HMECs) grown to 4th passage serve as our experimental model system. HMECs were derived from normal breast tissue organoids collected from reduction mammoplasties (RM) prepared at Lawrence Berkeley National Laboratory (Berkeley, CA) with approved IRB for sample distribution and collection from specific locations. Protocols have already been established by our lab for the propagation and maintenance of these cells *in vitro* which allow for highly-reproducible source material (Garbe et al., 2009; Labarge et al., 2013).

We examined genome-wide transcription in 54 primary luminal epithelial (LEP) and MEP myoepithelial (MEP) samples from 19 women across a range of ages (**Figure 1—table supplement 1**). These epithelial lineages were isolated from human mammary epithelial cells (HMECs) from two age cohorts: younger <30y women considered to be premenopausal (age

range 16-29y, m_{LEP}=16, m_{MEP}=16 samples, n=11 subjects) and older >55y women considered to be postmenopausal (age range 56-72y, m_{LEP}=11, m_{MEP}=11 samples, n=8 subjects).

Materials Design Analysis

Group allocation: Samples were allocated based on age demographics of women donating tissue. As menopausal status was not available for all subjects, we used age cutoffs to identify two cohorts in this aging study. The two age groups were defined as younger <30y women considered to be premenopausal, and older >55y women considered to be postmenopausal; samples from middle-aged >30y and <55y women were excluded from analysis. Only finite lifespan cell strains derived from reduction mammoplasties from healthy women were included; strains from prophylactic mastectomies (women with high-risk breast cancer mutations or family history) or normal-adjacent to tumor tissue were excluded. No group blinding or masking was used.

Replicates: For the purposes of this study using finite-lifespan cell strains derived from primary tissue, subject-level biological replicates (n) refer to data derived from tissue from different subjects. Sample-level biological replicates (m) are data derived from the same subject but from separate cell culture/co-culture, FACS isolation, sample and library preparation, and RNA-sequencing experiments. Sample replicates are typically bridge samples used across different experiment batches. Technical replicates (l) are data derived from a single subject from a single sample pre-processing experiment (e.g., qPCR). When sample replication for the same subject can be modeled, analyses are done at sample level; if not, analyses are done at subject level by taking the mean value of sample replicates (see **Method Details** and **Quantification and Statistical Analysis** section). For clarity, subject (n), sample (m), and techinical (l) replicates for each analysis are annotated in each figure. The datasets generated and analyzed during the current study include RNA-sequencing count data publicly available as part of GSE182338 (Miyano et al., 2021; Sayaman et al., 2021; Shalabi et al., 2021). Criteria for inclusion of samples

and gene transcripts included in the analyses, and all exclusion criteria are described in the **Method Details** section.

Sample-size estimation: Sample size for RNA-seq analyses was restricted to available organoids in the established HMEC bank that were isolated from reduction mammoplasties of healthy women who fall under the age range of interest: younger <30y and older >55y, that could be expanded in primary culture. Power analyses for the limma::voom DE pipeline were conducted post-hoc using the R package ssizeRNA::check.power function (Bi & Liu, 2019) (see **Supplementary Methods**). Power calculation for lineage-specific DE between LEP and MEP with DE genes defined at BH adj. $p<0.001$ and fold-change ≥2 yielded (i) an average power (ave.pw)=0.91 and fdr average (fdr.ave)= 0.00062 in younger <30y (n=11 subjects); and (ii) ave.pw=0.90 and fdr.ave= 0.00047 in older >55y (n=8). Power calculation for age-dependent DE between young <30y and old >55y with DE genes defined at BH adj. $p<0.05$ yielded a range of (iii) ave.pw=0.69-0.76 and fdr.ave= 0.040-0.045 in LEPs (for n=8 and n=11 respectively); (iv) ave.pw=0.87-0.89 and fdr.ave= 0.083-0.047 in MEPs (for n=8 and n=11 respectively).

Statistical reporting: Statistical analysis methods are described in full in the **Quantification and Statistical Analysis** section. Exact p-values are shown in figures when feasible, otherwise significance levels are annotated. Exact p-values and summary statistics at defined significance level thresholds are provided in **Supplementary Tables** (made available after publication); full summary statistics are available upon request.

Method Details

Breast tissue collection and HMEC culture: Primary HMECs were initiated and maintained according to previously reported protocols using M87A medium containing cholera toxin and oxytocin at 0.5 ng/ml and 0.1nM, respectively (Garbe et al., 2009; Labarge et al., 2013). For

experiments, 4th passage HMECs were cultured to sub-confluence prior to FACS-sorting. HMEC strains used in this study for RNA-seq are provided (**Figure 1—table supplement 1**).

Flow cytometry: FACS-enriched LEPs and MEPs were isolated from 4th passage finite-lifespan HMEC from reduction mammoplasties from two age cohorts: younger <30y women considered to be premenopausal (age range 16-29y) and older >55y women considered to be postmenopausal (age range 56-72y). LEP and MEP enrichment was performed across multiple studies (Miyano et al., 2021; Sayaman et al., 2021; Shalabi et al., 2021; Todhunter et al., 2021). Enrichment was conducted by FACS using well-established LEP-specific (CD227 or CD133) and MEP-specific (CD271 or CD10) cell-surface markers. Protocols were validated to sort similar populations regardless of antibody combination. Briefly, breast epithelial cells were stained and sorted following standard flow cytometry protocol. Primary HMEC strains for RNA-seq were stained with anti-human CD227-FITC (BD Biosciences, clone HMPV) or anti-human CD133-PE (BioLegend, clone7), and anti-human CD271-APC (BioLegend, clone ME20.4). Primary HMEC strains for Infinium 450K array stained with anti-human CD227-FITC (BD Biosciences, clone HMPV) and anti-human CD10-PE (BioLegnend, clone HI10a).

Cell co-cultures: In co-culture study (Miyano et al., 2017), FACS-enriched MEPs from 4th passage HMEC were re-plated on 6-well plates and cultured until the cells were confluent. The cells were treated with Mitomycin C (Santa Cruz Biotechnology, sc-3514) at 10µg/ml for 2.5h. In co-culture with shGJB6 study, FACS-enriched control and shGJB6 transduced MEPs from older >55y women were plated on 6-well plates and cultured until the cells were confluent. FACS-enriched 4th passage LEPs from younger <30y women were seeded directly on the mitomycin C-treated or shRNA transduced MEP layer. LEPs from co-cultures were separated after 10 days for gene expression qPCR analysis by FACS using anti-human CD133-PE (BioLegend, clone7) and anti-human CD271-APC (BioLegend, clone ME20.4). For Gap junction inhibition assay, cells were

cultured with indicated concentration of 18-alpha-Glycryrhetinic acid (Sigma, G8503) for 7 days; LEPs from co-culture were then separated using FACS with anti-CD227-FITC (BD Biosciences, 559774, clone HMPV) and anti-CD10-PE (Biolegend, 312204, clone HI10a).

RNA isolation and qPCR: Total RNAs were isolated from enriched LEPs and MEPs with Quick-RNA Microprep Kit (Zymo Research, R1050). For RNA-seq, isolated RNAs were submitted to Integrative Genomic Core at City of Hope (IGC at COH) for library preparation and sequencing. For qPCR, cDNAs were synthesized with iScirpt Reverse Transcription Supermix (BioRad, 1708840) according to the manufacturer's manual. Quantitative gene expression analysis was performed by CFX384 real-time PCR (BioRad) with iTaq Universal SYBR Green Supermix (BioRad, 1725125). Data were normalized to RPS18 or TBP by relative standard curve method.

Forward and reverse primer sequences generated in this study are indicated below:

GJB6 forward and reverse primers:

5'-CTACAGGCACGAAACCACTCG-3', 5'ACCCCTCTATCCGAACCTTCT-3'

ELF5 forward and reverse primers:

5'-TAGGGAACAAGGAATTTTTCGGG-3', 5'-GTACACTAACCTTCGGTCAACC-3'

TBP forward and reverse primers:

5'-GAGCTGTGATGTGAAGTTTCC-3', 5'-TCTGGGTTTGATCATTCTGTAG-3'

RPS18 forward and reverse primers:

5'-GGGCGGCGGAAAATAG-3', 5'-CGCCCTCTTGGTGAGGT-3'

Sequences for shGJB6 and shCtrl were ggatacttgctccattcatac and gcttcgcgccgtagtctta, respectively. shCtrl (CSHCTR001LVRU6GP) and shGJB6 Lenti-virus vector (HSH06069132LVRU6GP) were purchased from GeneCopoeia.

RNA-sequencing: Transcriptomic profiling of LEPs and MEPs from two age cohorts: younger <30y (m=32 LEP and MEP samples, n=11 subjects) and older >55y (m=22, n=8) women (**Figure 1—table supplement 1**) was performed via RNA-sequencing as part of the LaBarge sequencing collection GSE182338 (Miyano et al., 2021; Sayaman et al., 2021; Shalabi et al., 2021). Briefly, RNA sequencing libraries were prepared with Kapa RNA mRNA HyperPrep Kit (Kapa Biosystems, Cat KR1352) or KAPA stranded mRNA-seq (Kapa Biosystems, Cat KK8420) according to the manufacturer's protocol using 100 ng of total RNA from each sample for polyA RNA enrichment. Sequencing was performed on Illumina HiSeq 2500 with single read mode, and real-time analysis was used to process the image analysis. RNA-sequencing reads were trimmed using Trimmomatic (Bolger et al., 2014), and processed reads were mapped back to the human genome (hg19) using TOPHAT2 software (Kim et al., 2013). HTSeq (Anders & Huber, 2010) and RSeQC (Wang et al., 2012) were applied to generate the count matrices.

RNA-sequencing data pre-processing was conducted in *DESeq2* (Love et al., 2014) and *edgeR* (Robinson et al., 2010) on the entirety of the LaBarge sequencing collection GSE182338 (m=120 LEP and MEP samples, n=48 subjects) as described in (Miyano et al., 2021; Sayaman et al., 2021; Shalabi et al., 2021) including samples not included in this study. RNA-seq transcript Ensembl IDs were mapped to corresponding gene symbols, Entrez IDs and Uniprot IDs using *EnsDb.Hsapiens.v86* (*v2.99.0*) database (Rainer, 2017). We restricted analysis to the 17,328 genes with comparable dynamic ranges and consistent lineage-specific expression between primary organoid and 4th passage LEPs and MEPs in both age cohorts (linear regression $R^2 \geq 0.88$ to 0.91, $p<0.0001$) (**Figure 1— figure supplement 1A-1D**). ComBat batch-adjusted regularized log (rlog) expression values (Johnson et al., 2007; Leek et al., 2020; Love et al., 2014) were used for visualization and downstream analysis.

Public Data Sets: For differential expression analysis in bulk normal primary breast tissue, GSE102088 (Song et al., 2017) microarray data (n=114) were downloaded from the Gene

Expression Omnibus (GEO) database using the GEOquery (Davis & Meltzer, 2007). For machine learning, three data sets were used: (1) TCGA RNA-seq FPKM data from matched normal or PAM50 Normal, Luminal A (LumA), Luminal B (LumB), Her2 and Basal subtype breast cancer tissues (n=1,201) were downloaded using TCGAbiolinks (Colaprico et al., 2016) package; (2) GTEx RNA-seq count data from female subjects (n=180) were downloaded using the recount3 (Collado-Torres et al., 2017; Wilks et al., 2021) and FPKM transformed; and (3) GSE81540 (Brueffer et al., 2020; Brueffer et al., 2018; Dahlgren et al., 2021) RNA-seq FPKM data from PAM50 Normal, LumA, LumB, Her2 and Basal subtype breast cancer tissues (n=3,184) were downloaded from GEO.

Quantification and Statistical Analysis

Differential Analyses: For differential analyses of LEP and MEP samples, a combination of lineage and age group was modeled. Differential expression (DE) was performed in limma voom (Law et al., 2014; Ritchie et al., 2015) on sample-level data from 17,328 genes with eBayes moderation and RNA-seq batch modeled as a covariate, and with adjustment for biological replicates. Differential variability (DV) was performed in MDSeq (Ran & Daye, 2017) on batch-adjusted subject-level data from 14,601 genes whose variances could be estimated after outlier removal. For lineage-specific DE analyses, contrasts between LEP and MEP in younger <30y and in older >55y women were performed. Lineage-specific DE thresholds were set at Benjamini-Hochberg (BH) adjusted $p<0.001$ and $\log_2$ fold changes, lfc ≥1 in each age cohort. LEP-specific and MEP-specific genes were defined as those with lineage-specific DE in younger <30y women. For age-dependent analyses, contrasts between <30y and >55y LEPs and <30y and >55y MEPs were performed, and age-dependent directional or variant changes were defined at DE or DV BH adj. $p<0.05$ in each lineage.

Age-dependent DE analysis of normal primary breast tissue was performed on publicly available GSE102088 microarray data (n=114 subjects, <30y n=35, >30y<55y n=68, >55y n=11)

(Song et al., 2017) in limma with eBayes moderation. Significant DE between age groups in bulk tissue were defined at BH adj. $p<0.05$ and nominal significance at unadj. $p<0.05$.

Gene Set Enrichment Analysis (GSEA): Fast gene set enrichment analysis (fgsea) (Korotkevich et al., 2021) was used to identify age-dependent enrichment of Molecular Signatures Database (MSigDB) hallmark gene sets (Liberzon et al., 2015) in LEPs or MEPs using DE and DV rank-ordered test statistics. Enriched gene sets were defined as those with enrichment BH adj. $p<0.05$. For bulk tissue GSEA analysis (GSE102088, <30y n=35, >55y n=11), gene sets were constructed from age-dependent genes in LEPs: (i) 251 genes that were differentially upregulated in young <30y LEPs; and (ii) 220 genes that were that were differentially upregulated in old >55y LEPs. Age-dependent enrichment was assessed in bulk tissue using DE rank-ordered test statistics; enrichment was similarly defined at BH adj. $p<0.05$.

Lineage-specific Ligand-Receptor Pair Interactions and Functional Network Analysis: Ligand-receptor pairs (LRPs) (Ramilowski et al., 2015) gene symbols were mapped to Ensembl IDs using *EnsDb.Hsapiens.v86* database (Rainer, 2017). Lineage-specific LRPs were defined based on either the LEP-specific or MEP-specific (DE adj. $p<0.001$ and fold-change ≥2) expression of either the ligand and/or its cognate receptor in the younger cohort. Lineage-specific LRP interactions were considered to be lost in the older cohort when lineage-specific DE of the ligand and/or its cognate receptor was lost in the older cells (not DE at adj. $p<0.001$, lfc≥1). Functional network enrichment of LRPs in the younger cohort and LRPs lost in the older cohort were performed using the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (https://string-db.org/) and enriched KEGG pathways (false discovery rate, FDR $p<0.05$) were reported.

Age-dependent DE Protein-protein Interactions and Functional Network Analysis: Protein-protein interaction (PPI) analysis was performed using the STRING database (https://string-db.org/). All

possible PPI are considered using all active interaction sources and setting minimum require interaction score to the lowest confidence threshold of 0.150. Network was visualized in igraph (Csárdi & Nepusz, 2006), and only the largest fully connected main network of genes was plotted. Community detection was performed on this main network using optimal community structure algorithm in igraph (Brandes et al., 2008). Each community was then analyzed in STRING for functional network enrichment (FDR $p<0.05$) and common functional terms were summarized and reported.

Unsupervised hierarchical clustering: Unsupervised hierarchical clustering in heatmaps (gplots::heatmap.2, pheatmap) (Kolde, 2019; Warnes et al., 2020) were implemented using hclust Ward's clustering criterion (ward.D2) agglomerative method with Euclidean distances as distance metric.

Machine Learning: Machine learning (ML) multi-class prediction of normal breast tissue and PAM50 breast cancer subtypes was performed in caret (Kuhn, 2008) using an elastic net model (glmnet) (Friedman et al., 2010) based on tissue expression of 536 mapped age-dependent DE and DV genes identified in LEPs. ML was carried out in three large publicly available RNA-seq datasets of normal and cancer breast tissue: GTEx, TCGA and GSE81540 (Brueffer et al., 2020; Brueffer et al., 2018; Dahlgren et al., 2021) with analysis restricted to tissues from women annotated as normal or PAM50 LumA, LumB, Her2 and Basal subtypes. The ML model was trained using 10-fold cross-validation with 3 repeats in 75% of TCGA data (n=873) using a hybrid subsampling technique (SMOTE) (Torgo, 2010), and optimizing for mean balanced accuracy. Model performance was then evaluated in the 25% of TCGA (n=288) and an independent dataset of normal tissues from GTEx and breast cancer tissues from GSE81540 (n=3,364). ML multi-class prediction performance was evaluated in each test set using the MultiROC package (Wei & Wang, 2020): (i) macro-average area under the ROC curve (AUC), calculated as the average

of all group results; (ii) micro-average AUC, calculated by stacking all groups together; and (iii) AUC of each group vs. the rest. Gene predictors were identified as genes with scaled variable importance contribution to the predictive model. Genes with scaled variable importance ≥25% in prediction of at least one class were visualized; gene expression the top 5 predictors in each class were further analyzed in the TCGA breast cancer cohort.

For a full description of the **Materials Design Analysis**, **Method Details**, and **Quantification and Statistical Analysis** sections, see **Supplementary Methods**.

Declarations

Availability of data and materials: Human mammary epithelial cells (HMECs) derived from subjects included in this study are available upon request. Forward and reverse primer sequences generated in this study are provided in the Methods section. The datasets generated and analyzed during the current study include RNA-sequencing count data publicly available as part of GSE182338 (Miyano et al., 2021; Sayaman et al., 2021; Shalabi et al., 2021). The gene expression data that support the findings of this study are available from GSE102088 (Song et al., 2017); GSE81540 (Brueffer et al., 2020; Brueffer et al., 2018; Dahlgren et al., 2021); The Cancer Genome Atlas (TCGA) Research Network: https://www.cancer.gov/tcga; and The Genotype-Tissue Expression (GTEx) Project: https://gtexportal.org/. Analysis was conducted using standard R/Bioconductor packages and statistical tests implemented in R. All package versions, model design, and parameters are described in detail in **Supplementary Methods**. Summary statistics at defined significance levels are provided in **Supplementary Tables** (made available after publication); full summary statistics are available upon request.

Competing interests: The authors declare that they have no competing interests.

Funding: The investigators are grateful for support from the National Cancer Institute (NCI) Cancer Metabolism Training Program Postdoctoral Fellowship T32CA221709 (RWS). From the National Institutes of Health U01CA244109, R33AG059206, R01EB024989, R01CA237602; the Department of Defense/Army Breast Cancer Era of Hope Scholar Award BC141351 and Expansion Award BC181737, Conrad N. Hilton Foundation, Yvonne Craig-Aldrich Fund for Cancer Research, and City of Hope Center for Cancer and Aging (MAL). Research reported in this publication included work performed in the Integrative Genomics and Bioinformatics, and Analytical Cytometry Cores supported by the National Cancer Institute of the National Institutes of Health under grant number P30CA033572. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions: Conceptualization, RWS, MM, MAL; Methodology, RWS, MM; Software Programming, RWS; Validation, MM, RWS, PS; Formal Analysis, RWS, MM; Investigation, RWS, MM, MRS, SS, PS, MET, AZ; Resources, MAL, MRS, MM; Data Curation, MRS, MM, RWS, MET, SS; Writing – Original Draft Preparation, RWS, MAL, MM; Writing – Review & Editing Preparation, RWS, MAL, MM, DES, MRS, PS, SLN, AZ, MET, VS, SS; Visualization, RWS, MM; Supervision, MAL; Project Administration, MAL; Funding Acquisition, MAL.

Acknowledgments: We would like to acknowledge the contributions of Dr. James Garbe to the HMEC bank; our current and previous research associates Jennifer Lopez, Jessica Bloom and Jonathan Lee for their technical support; our patient advocates, Susan Samson and Sandy Preto; and the City of Hope Bioinformatics Core led by Dr. Xiwei Wu, specifically Dr. Min-Hsuan Chen who generated the RNA-sequencing raw count data, and Dr. Jinhui Wang who prepared the RNA-

sequencing library and performed the sequencing on the Illumina HiSeq2500 platform. The results shown and referenced here are based in part upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga; and The Genotype-Tissue Expression (GTEx) Project: https://gtexportal.org/ supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS.

References

Anders, S., & Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol*, *11*(10), R106. https://doi.org/10.1186/gb-2010-11-10-r106

Bashkeel, N., Perkins, T. J., Kærn, M., & Lee, J. M. (2019). Human gene expression variability and its dependence on methylation and aging. *BMC Genomics*, *20*(1), 941. https://doi.org/10.1186/s12864-019-6308-7

Benz, C. C. (2008). Impact of aging on the biology of breast cancer. *Crit Rev Oncol Hematol*, *66*(1), 65-74. https://doi.org/10.1016/j.critrevonc.2007.09.001

Bhateja, P., Cherian, M., Majumder, S., & Ramaswamy, B. (2019). The Hedgehog Signaling Pathway: A Viable Target in Breast Cancer? *Cancers (Basel)*, *11*(8). https://doi.org/10.3390/cancers11081126

Bi, R., & Liu, P. (2019). *ssizeRNA: Sample Size Calculation for RNA-Seq Experimental Design*. In R package version 1.3.2.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120. https://doi.org/10.1093/bioinformatics/btu170

Booth, B. W., & Smith, G. H. (2006). Estrogen receptor-alpha and progesterone receptor are expressed in label-retaining mammary epithelial cells that divide asymmetrically and retain their template DNA strands. *Breast Cancer Res*, *8*(4), R49. https://doi.org/10.1186/bcr1538

Brandes, U., Delling, D., Gaertler, M., Gorke, R., Hoefer, M., Nikoloski, Z., & Wagner, D. (2008). On Modularity Clustering. *IEEE Transactions on Knowledge and Data Engineering*, *20*(2), 172-188. https://doi.org/10.1109/TKDE.2007.190689

Brueffer, C., Gladchuk, S., Winter, C., Vallon-Christersson, J., Hegardt, C., Hakkinen, J., George, A. M., Chen, Y., Ehinger, A., Larsson, C., Loman, N., Malmberg, M., Ryden, L., Borg, A., & Saal, L. H. (2020). The mutational landscape of the SCAN-B real-world primary breast cancer transcriptome. *EMBO Mol Med*, *12*(10), e12118. https://doi.org/10.15252/emmm.202012118

Brueffer, C., Vallon-Christersson, J., Grabau, D., Ehinger, A., Hakkinen, J., Hegardt, C., Malina, J., Chen, Y., Bendahl, P. O., Manjer, J., Malmberg, M., Larsson, C., Loman, N., Ryden, L., Borg, A., & Saal, L. H. (2018). Clinical Value of RNA Sequencing-Based Classifiers for Prediction of the Five Conventional Breast Cancer Biomarkers: A Report From the Population-Based Multicenter Sweden Cancerome Analysis Network-Breast Initiative. *JCO Precis Oncol*, *2*. https://doi.org/10.1200/PO.17.00135

Colaprico, A., Silva, T. C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T. S., Malta, T. M., Pagnotta, S. M., Castiglioni, I., Ceccarelli, M., Bontempi, G., & Noushmehr, H.

(2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res*, *44*(8), e71. https://doi.org/10.1093/nar/gkv1507
Collado-Torres, L., Nellore, A., & Jaffe, A. (2017). recount workflow: Accessing over 70,000 human RNA-seq samples with Bioconductor [version 1; peer review: 1 approved, 2 approved with reservations]. *F1000Research*, *6*(1558). https://doi.org/10.12688/f1000research.12223.1
Csárdi, G., & Nepusz, T. (2006). The igraph software package for complex network research. *InterJournal*, *Complex Systems*, 1695. https://igraph.org
Dahlgren, M., George, A. M., Brueffer, C., Gladchuk, S., Chen, Y., Vallon-Christersson, J., Hegardt, C., Hakkinen, J., Ryden, L., Malmberg, M., Larsson, C., Gruvberger-Saal, S. K., Ehinger, A., Loman, N., Borg, A., & Saal, L. H. (2021). Preexisting Somatic Mutations of Estrogen Receptor Alpha (ESR1) in Early-Stage Primary Breast Cancer. *JNCI Cancer Spectr*, *5*(2), pkab028. https://doi.org/10.1093/jncics/pkab028
Davis, S., & Meltzer, P. S. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics*, *23*(14), 1846-1847. https://doi.org/10.1093/bioinformatics/btm254
Dbouk, H. A., Mroue, R. M., El-Sabban, M. E., & Talhouk, R. S. (2009). Connexins: a myriad of functions extending beyond assembly of gap junction channels. *Cell Commun Signal*, *7*, 4. https://doi.org/10.1186/1478-811x-7-4
de Jong, T. V., Moshkin, Y. M., & Guryev, V. (2019). Gene expression variability: the other dimension in transcriptome analysis. *Physiol Genomics*, *51*(5), 145-158. https://doi.org/10.1152/physiolgenomics.00128.2018
de Magalhães, J. P., Curado, J., & Church, G. M. (2009). Meta-analysis of age-related gene expression profiles identifies common signatures of aging. *Bioinformatics*, *25*(7), 875-881. https://doi.org/10.1093/bioinformatics/btp073
DeSantis, C. E., Ma, J., Gaudet, M. M., Newman, L. A., Miller, K. D., Goding Sauer, A., Jemal, A., & Siegel, R. L. (2019). Breast cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, *69*(6), 438-451. https://doi.org/10.3322/caac.21583
Enge, M., Arda, H. E., Mignardi, M., Beausang, J., Bottino, R., Kim, S. K., & Quake, S. R. (2017). Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns. *Cell*, *171*(2), 321-330.e314. https://doi.org/10.1016/j.cell.2017.09.004
Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, *33*(1), 1 - 22. https://doi.org/10.18637/jss.v033.i01
Fulop, T., Larbi, A., Dupuis, G., Le Page, A., Frost, E. H., Cohen, A. A., Witkowski, J. M., & Franceschi, C. (2017). Immunosenescence and Inflamm-Aging As Two Sides of the Same Coin: Friends or Foes? *Front Immunol*, *8*, 1960. https://doi.org/10.3389/fimmu.2017.01960
Garbe, J. C., Bhattacharya, S., Merchant, B., Bassett, E., Swisshelm, K., Feiler, H. S., Wyrobek, A. J., & Stampfer, M. R. (2009). Molecular distinctions between stasis and telomere attrition senescence barriers shown by long-term culture of normal human mammary epithelial cells. *Cancer Res*, *69*(19), 7557-7568. https://doi.org/10.1158/0008-5472.can-09-0270
Garbe, J. C., Pepin, F., Pelissier, F. A., Sputova, K., Fridriksdottir, A. J., Guo, D. E., Villadsen, R., Park, M., Petersen, O. W., Borowsky, A. D., Stampfer, M. R., & Labarge, M. A. (2012). Accumulation of multipotent progenitors with a basal differentiation bias during aging of human mammary epithelia. *Cancer Res*, *72*(14), 3687-3701. https://doi.org/10.1158/0008-5472.can-12-0157
Garbe, J. C., Vrba, L., Sputova, K., Fuchs, L., Novak, P., Brothman, A. R., Jackson, M., Chin, K., LaBarge, M. A., Watts, G., Futscher, B. W., & Stampfer, M. R. (2014).

Immortalization of normal human mammary epithelial cells in two steps by direct targeting of senescence barriers does not require gross genomic alterations. *Cell Cycle*, *13*(21), 3423-3435. https://doi.org/10.4161/15384101.2014.954456
Garrod, D., & Chidgey, M. (2008). Desmosome structure, composition and function. *Biochim Biophys Acta*, *1778*(3), 572-587. https://doi.org/10.1016/j.bbamem.2007.07.014
Glass, D., Viñuela, A., Davies, M. N., Ramasamy, A., Parts, L., Knowles, D., Brown, A. A., Hedman, A. K., Small, K. S., Buil, A., Grundberg, E., Nica, A. C., Di Meglio, P., Nestle, F. O., Ryten, M., Durbin, R., McCarthy, M. I., Deloukas, P., Dermitzakis, E. T., . . . Spector, T. D. (2013). Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol*, *14*(7), R75. https://doi.org/10.1186/gb-2013-14-7-r75
Hartman, J., Lam, E. W., Gustafsson, J. A., & Ström, A. (2009). Hes-6, an inhibitor of Hes-1, is regulated by 17beta-estradiol and promotes breast cancer cell proliferation. *Breast Cancer Res*, *11*(6), R79. https://doi.org/10.1186/bcr2446
Ikushima, H., Munakata, Y., Ishii, T., Iwata, S., Terashima, M., Tanaka, H., Schlossman, S. F., & Morimoto, C. (2000). Internalization of CD26 by mannose 6-phosphate/insulin-like growth factor II receptor contributes to T cell activation. *Proc Natl Acad Sci U S A*, *97*(15), 8439-8444. https://doi.org/10.1073/pnas.97.15.8439
Illingworth, R. S. (2019). Chromatin folding and nuclear architecture: PRC1 function in 3D. *Curr Opin Genet Dev*, *55*, 82-90. https://doi.org/10.1016/j.gde.2019.06.006
Johnson, W. E., Li, C., & Rabinovic, A. (2007). Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*, *8*(1), 118-127. https://doi.org/10.1093/biostatistics/kxj037
Kadoch, C., Hargreaves, D. C., Hodges, C., Elias, L., Ho, L., Ranish, J., & Crabtree, G. R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet*, *45*(6), 592-601. https://doi.org/10.1038/ng.2628
Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., & Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, *14*(4), R36. https://doi.org/10.1186/gb-2013-14-4-r36
Kim, J., & Villadsen, R. (2018). Expression of Luminal Progenitor Marker CD117 in the Human Breast Gland. *J Histochem Cytochem*, *66*(12), 879-888. https://doi.org/10.1369/0022155418788845
Kohwi-Shigematsu, T., Poterlowicz, K., Ordinario, E., Han, H. J., Botchkarev, V. A., & Kohwi, Y. (2013). Genome organizing function of SATB1 in tumor progression. *Semin Cancer Biol*, *23*(2), 72-79. https://doi.org/10.1016/j.semcancer.2012.06.009
Kolde, R. (2019). *pheatmap: Pretty Heatmaps*. In R package version 1.0.12.
Kontomanolis, E. N., Kalagasidou, S., Pouliliou, S., Anthoulaki, X., Georgiou, N., Papamanolis, V., & Fasoulakis, Z. N. (2018). The Notch Pathway in Breast Cancer Progression. *ScientificWorldJournal*, *2018*, 2415489. https://doi.org/10.1155/2018/2415489
Korotkevich, G., Sukhov, V., Budin, N., Shpak, B., Artyomov, M. N., & Sergushichev, A. (2021). Fast gene set enrichment analysis. *bioRxiv*, 060012. https://doi.org/10.1101/060012
Kouros-Mehr, H., Slorach, E. M., Sternlicht, M. D., & Werb, Z. (2006). GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell*, *127*(5), 1041-1055. https://doi.org/10.1016/j.cell.2006.09.048
Kowalczyk, M. S., Tirosh, I., Heckl, D., Rao, T. N., Dixit, A., Haas, B. J., Schneider, R. K., Wagers, A. J., Ebert, B. L., & Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation programs upon aging of hematopoietic stem cells. *Genome Res*, *25*(12), 1860-1872. https://doi.org/10.1101/gr.192237.115
Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5), 1 - 26. https://doi.org/10.18637/jss.v028.i05

Labarge, M. A., Garbe, J. C., & Stampfer, M. R. (2013). Processing of human reduction mammoplasty and mastectomy tissues for cell culture. *J Vis Exp*(71). https://doi.org/10.3791/50011

Lambert, S. A., Jolma, A., Campitelli, L. F., Das, P. K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T. R., & Weirauch, M. T. (2018). The Human Transcription Factors. *Cell*, *172*(4), 650-665. https://doi.org/10.1016/j.cell.2018.01.029

Lanigan, F., Geraghty, J. G., & Bracken, A. P. (2011). Transcriptional regulation of cellular senescence. *Oncogene*, *30*(26), 2901-2911. https://doi.org/10.1038/onc.2011.34

Law, C. W., Chen, Y., Shi, W., & Smyth, G. K. (2014). voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*, *15*(2), R29. https://doi.org/10.1186/gb-2014-15-2-r29

Lee, G., & Lee, M. (2017). Classification of Genes Based on Age-Related Differential Expression in Breast Cancer. *Genomics Inform*, *15*(4), 156-161. https://doi.org/10.5808/gi.2017.15.4.156

Leek, J. T., Johnson, W. E., Parker, H. S., Elana J. Fertig, E. J., Jaffe, A. E., Zhang, Y., Storey, J. D., & Collado Torres, L. (2020). *sva: Surrogate Variable Analysis*. In R package version 3.35.2.

Levy, O., Amit, G., Vaknin, D., Snir, T., Efroni, S., Castaldi, P., Liu, Y. Y., Cohen, H. Y., & Bashan, A. (2020). Age-related loss of gene-to-gene transcriptional coordination among single cells. *Nat Metab*, *2*(11), 1305-1315. https://doi.org/10.1038/s42255-020-00304-4

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*, *1*(6), 417-425. https://doi.org/10.1016/j.cels.2015.12.004

Lin, J. H., Lee, W. J., Wu, H. C., Wu, C. H., Chen, L. C., Huang, C. C., Chang, H. L., Cheng, T. C., Chang, H. W., Ho, C. T., Tu, S. H., & Ho, Y. S. (2019). Small G protein signalling modulator 2 (SGSM2) is involved in oestrogen receptor-positive breast cancer metastasis through enhancement of migratory cell adhesion via interaction with E-cadherin. *Cell Adh Migr*, *13*(1), 120-137. https://doi.org/10.1080/19336918.2019.1568139

López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M., & Kroemer, G. (2013). The hallmarks of aging. *Cell*, *153*(6), 1194-1217. https://doi.org/10.1016/j.cell.2013.05.039

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Mansour, M. A. (2020). SP3 is associated with migration, invasion, and Akt/PKB signalling in MDA-MB-231 breast cancer cells. *J Biochem Mol Toxicol*, e22657. https://doi.org/10.1002/jbt.22657

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., de la Roche, M., Marioni, J. C., & Odom, D. T. (2017). Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science*, *355*(6332), 1433-1436. https://doi.org/10.1126/science.aah4115

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., Sedrak, M. S., Stampfer, M. R., Seewaldt, V. L., & LaBarge, M. A. (2021). Breast-Specific Molecular Clocks Comprised of ELF5 Expression and Promoter Methylation Identify Individuals Susceptible to Cancer Initiation. *Cancer Prev Res (Phila)*, *14*(8), 779-794. https://doi.org/10.1158/1940-6207.CAPR-20-0635

Miyano, M., Sayaman, R. W., Stoiber, M. H., Lin, C. H., Stampfer, M. R., Brown, J. B., & 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*(10), 2026-2051. https://doi.org/10.18632/aging.101298

Moustakim, M., Christott, T., Monteiro, O. P., Bennett, J., Giroud, C., Ward, J., Rogers, C. M., Smith, P., Panagakou, I., Diaz-Saez, L., Felce, S. L., Gamble, V., Gileadi, C., Halidi, N., Heidenreich, D., Chaikuad, A., Knapp, S., Huber, K. V. M., Farnie, G., . . . Fedorov, O. (2018). Discovery of an MLLT1/3 YEATS Domain Chemical Probe. *Angew Chem Int Ed Engl*, *57*(50), 16302-16307. https://doi.org/10.1002/anie.201810617

Naik, R., & Galande, S. (2019). SATB family chromatin organizers as master regulators of tumor progression. *Oncogene*, *38*(12), 1989-2004. https://doi.org/10.1038/s41388-018-0541-4

Pandey, P. R., Saidou, J., & Watabe, K. (2010). Role of myoepithelial cells in breast tumor progression. *Front Biosci (Landmark Ed)*, *15*, 226-236. https://doi.org/10.2741/3617

Parker, J. S., Mullins, M., Cheang, M. C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J. F., Stijleman, I. J., Palazzo, J., Marron, J. S., Nobel, A. B., Mardis, E., Nielsen, T. O., Ellis, M. J., Perou, C. M., & Bernard, P. S. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol*, *27*(8), 1160-1167. https://doi.org/10.1200/jco.2008.18.1370

Peters, M. J., Joehanes, R., Pilling, L. C., Schurmann, C., Conneely, K. N., Powell, J., Reinmaa, E., Sutphin, G. L., Zhernakova, A., Schramm, K., Wilson, Y. A., Kobes, S., Tukiainen, T., Ramos, Y. F., Göring, H. H., Fornage, M., Liu, Y., Gharib, S. A., Stranger, B. E., . . . Johnson, A. D. (2015). The transcriptional landscape of age in human peripheral blood. *Nat Commun*, *6*, 8570. https://doi.org/10.1038/ncomms9570

Phipson, B., Maksimovic, J., & Oshlack, A. (2016). missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*, *32*(2), 286-288. https://doi.org/10.1093/bioinformatics/btv560

Phipson, B., & Oshlack, A. (2014). DiffVar: a new method for detecting differential variability with application to methylation in cancer and aging. *Genome Biol*, *15*(9), 465. https://doi.org/10.1186/s13059-014-0465-4

Rainer, J. (2017). *EnsDb.Hsapiens.v86: Ensembl based annotation package*. In R package version 2.99.0.

Ramilowski, J. A., Goldberg, T., Harshbarger, J., Kloppmann, E., Lizio, M., Satagopam, V. P., Itoh, M., Kawaji, H., Carninci, P., Rost, B., & Forrest, A. R. (2015). A draft network of ligand-receptor-mediated multicellular signalling in human. *Nat Commun*, *6*, 7866. https://doi.org/10.1038/ncomms8866

Ran, D., & Daye, Z. J. (2017). Gene expression variability and the analysis of large-scale RNA-seq studies with the MDSeq. *Nucleic Acids Res*, *45*(13), e127. https://doi.org/10.1093/nar/gkx456

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., & Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*, *43*(7), e47. https://doi.org/10.1093/nar/gkv007

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139-140. https://doi.org/10.1093/bioinformatics/btp616

Sayaman, R. W., Miyano, M., Senapati, P., Shalabi, S., Zirbes, A., Todhunter, M. E., Seewaldt, V., Neuhausen, S. L., Stampfer, M. R., Schones, D. E., & LaBarge, M. A. (2021). *Epigenetic changes with age primes mammary luminal epithelia for cancer initiation*. https://www.biorxiv.org/content/biorxiv/early/2021/02/13/2021.02.12.430777.full.pdf

Shahi, P., Wang, C. Y., Lawson, D. A., Slorach, E. M., Lu, A., Yu, Y., Lai, M. D., Gonzalez Velozo, H., & Werb, Z. (2017). ZNF503/Zpo2 drives aggressive breast cancer progression by down-regulation of GATA3 expression. *Proc Natl Acad Sci U S A*, *114*(12), 3169-3174. https://doi.org/10.1073/pnas.1701690114

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., Seewaldt, V. E., & LaBarge, M. A. (2021). Evidence for accelerated aging in mammary epithelia of women carrying germline BRCA1 or BRCA2 mutations. *Nature Aging*, *1*(9), 838-849. https://doi.org/10.1038/s43587-021-00104-9

Shen-Orr, S. S., & Gaujoux, R. (2013). Computational deconvolution: extracting cell type-specific information from heterogeneous samples. *Curr Opin Immunol*, *25*(5), 571-578. https://doi.org/10.1016/j.coi.2013.09.015

Slieker, R. C., van Iterson, M., Luijk, R., Beekman, M., Zhernakova, D. V., Moed, M. H., Mei, H., van Galen, M., Deelen, P., Bonder, M. J., Zhernakova, A., Uitterlinden, A. G., Tigchelaar, E. F., Stehouwer, C. D., Schalkwijk, C. G., van der Kallen, C. J., Hofman, A., van Heemst, D., de Geus, E. J., . . . Heijmans, B. T. (2016). Age-related accrual of methylomic variability is linked to fundamental ageing mechanisms. *Genome Biol*, *17*(1), 191. https://doi.org/10.1186/s13059-016-1053-6

Song, M. A., Brasky, T. M., Weng, D. Y., McElroy, J. P., Marian, C., Higgins, M. J., Ambrosone, C., Spear, S. L., Llanos, A. A., Kallakury, B. V. S., Freudenheim, J. L., & Shields, P. G. (2017). Landscape of genome-wide age-related DNA methylation in breast tissue. *Oncotarget*, *8*(70), 114648-114662. https://doi.org/10.18632/oncotarget.22754

Stampfer, M. R., LaBarge, M. A., & Garbe, J. C. (2013). An Integrated Human Mammary Epithelial Cell Culture System for Studying Carcinogenesis and Aging. In *Cell and Molecular Biology of Breast Cancer* (pp. 323–361). https://doi.org/https://doi.org/10.1007/978-1-62703-634-4_15

Teleki, I., Szasz, A. M., Maros, M. E., Gyorffy, B., Kulka, J., Meggyeshazi, N., Kiszner, G., Balla, P., Samu, A., & Krenacs, T. (2014). Correlations of differentially expressed gap junction connexins Cx26, Cx30, Cx32, Cx43 and Cx46 with breast cancer progression and prognosis. *PLoS One*, *9*(11), e112541. https://doi.org/10.1371/journal.pone.0112541

Titus, A. J., Gallimore, R. M., Salas, L. A., & Christensen, B. C. (2017). Cell-type deconvolution from DNA methylation: a review of recent applications. *Hum Mol Genet*, *26*(R2), R216-r224. https://doi.org/10.1093/hmg/ddx275

Todhunter, M. E., Miyano, M., Moolamalla, D. S., Filippov, A., Sayaman, R. W., & LaBarge, M. A. (2021). Volume-constrained microcontainers enable myoepithelial functional differentiation in highly parallel mammary organoid culture. *iScience*, *24*(4), 102253. https://doi.org/10.1016/j.isci.2021.102253

Todhunter, M. E., Sayaman, R. W., Miyano, M., & LaBarge, M. A. (2018). Tissue aging: the integration of collective and variant responses of cells to entropic forces over time. *Curr Opin Cell Biol*, *54*, 121-129. https://doi.org/10.1016/j.ceb.2018.05.016

Torgo, L. (2010). *Data Mining with R, learning with case studies*. Chapman and Hall/CRC. http://www.liaad.up.pt/~ltorgo/DataMiningWithR

Vilas, C. K., Emery, L. E., Denchi, E. L., & Miller, K. M. (2018). Caught with One's Zinc Fingers in the Genome Integrity Cookie Jar. *Trends Genet*, *34*(4), 313-325. https://doi.org/10.1016/j.tig.2017.12.011

Volkova, M., Garg, R., Dick, S., & Boheler, K. R. (2005). Aging-associated changes in cardiac gene expression. *Cardiovasc Res*, *66*(2), 194-204. https://doi.org/10.1016/j.cardiores.2004.11.016

Wang, L., Wang, S., & Li, W. (2012). RSeQC: quality control of RNA-seq experiments. *Bioinformatics*, *28*(16), 2184-2185. https://doi.org/10.1093/bioinformatics/bts356

Warnes, G. R., Bolker, B., Bonebakker, L., Gentleman, R. H., W., Andy Liaw, A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., & Venables, B. (2020). *gplots: Various R Programming Tools for Plotting Data*. In R package version 3.0.3.

Wei, R., & Wang, J. (2020). *multiROC: Calculating and Visualizing ROC and PR Curves Across Multi-Class Classifications*. In (Version R package version 1.1.1) https://CRAN.R-project.org/package=multiROC

Wilks, C., Zheng, S. C., Chen, F. Y., Charles, R., Solomon, B., Ling, J. P., Imada, E. L., Zhang, D., Joseph, L., Leek, J. T., Jaffe, A. E., Nellore, A., Collado-Torres, L., Hansen, K. D., & Langmead, B. (2021). recount3: summaries and queries for large-scale RNA-seq expression and splicing. *bioRxiv*, 2021.2005.2021.445138. https://doi.org/10.1101/2021.05.21.445138

Xiang, X., Deng, Z., Zhuang, X., Ju, S., Mu, J., Jiang, H., Zhang, L., Yan, J., Miller, D., & Zhang, H. G. (2012). Grhl2 determines the epithelial phenotype of breast cancers and promotes tumor progression. *PLoS One*, *7*(12), e50781. https://doi.org/10.1371/journal.pone.0050781

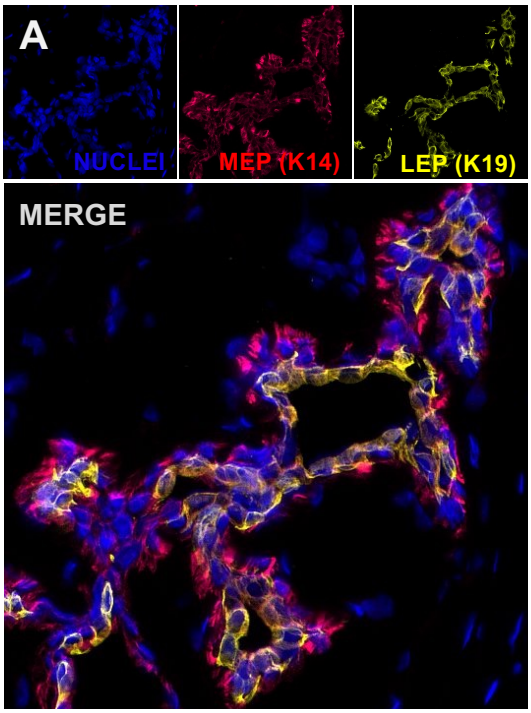
Xie, H., Wang, M., de Andrade, A., Bonaldo Mde, F., Galat, V., Arndt, K., Rajaram, V., Goldman, S., Tomita, T., & Soares, M. B. (2011). Genome-wide quantitative assessment of variation in DNA methylation patterns. *Nucleic Acids Res*, *39*(10), 4099-4108. https://doi.org/10.1093/nar/gkr017

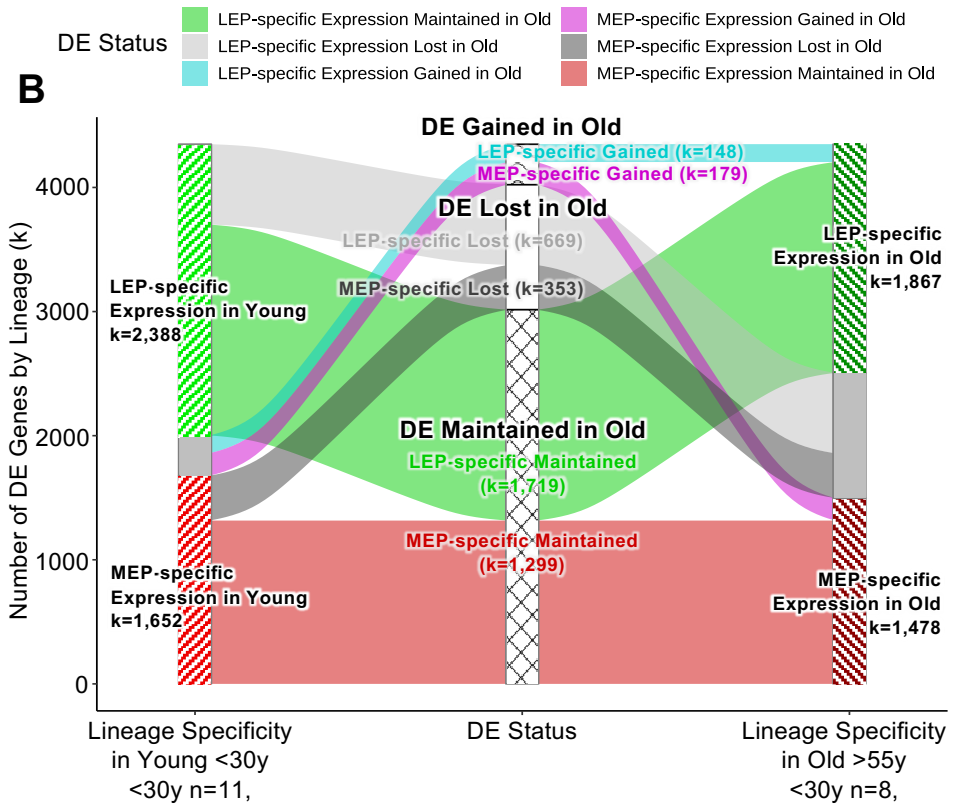
Xu, J., Chen, Y., & Olopade, O. I. (2010). MYC and Breast Cancer. *Genes Cancer*, *1*(6), 629-640. https://doi.org/10.1177/1947601910378691

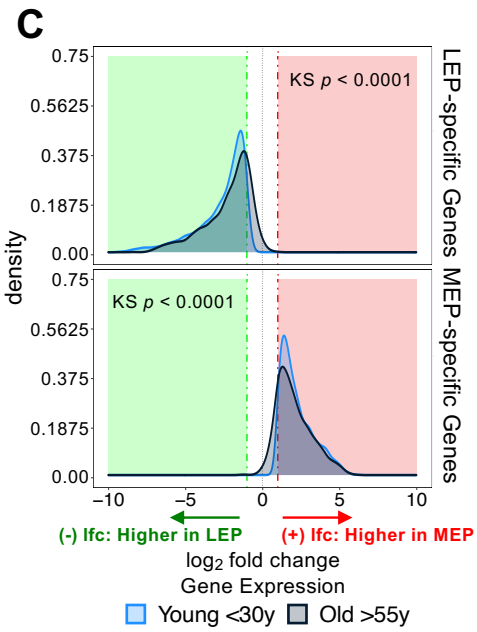
Yau, C., Fedele, V., Roydasgupta, R., Fridlyand, J., Hubbard, A., Gray, J. W., Chew, K., Dairkee, S. H., Moore, D. H., Schittulli, F., Tommasi, S., Paradiso, A., Albertson, D. G., & Benz, C. C. (2007). Aging impacts transcriptomes but not genomes of hormone-dependent breast cancers. *Breast Cancer Res*, *9*(5), R59. https://doi.org/10.1186/bcr1765

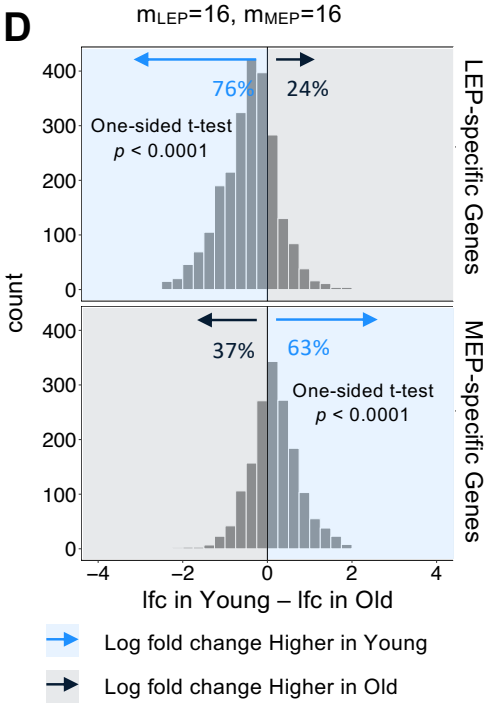
Zirbes, A., Joseph, J., Lopez, J. C., Sayaman, R. W., Basam, M., Seewaldt, V. L., & LaBarge, M. A. (2021). Changes in Immune Cell Types with Age in Breast are Consistent with a Decline in Immune Surveillance and Increased Immunosuppression. *J Mammary Gland Biol Neoplasia*. https://doi.org/10.1007/s10911-021-09495-2

Figure 1


A
NUCLEI
MEP (K14)
LEP (K19)
MERGE


B
DE Status
LEP-specific Expression Maintained in Old
LEP-specific Expression Lost in Old
LEP-specific Expression Gained in Old
MEP-specific Expression Gained in Old
MEP-specific Expression Lost in Old
MEP-specific Expression Maintained in Old
DE Gained in Old
LEP-specific Gained (k=148)
MEP-specific Gained (k=179)
DE Lost in Old
LEP-specific Lost (k=669)
MEP-specific Lost (k=353)
DE Maintained in Old
LEP-specific Maintained (k=1,719)
MEP-specific Maintained (k=1,299)
LEP-specific Expression in Young k=2,388
MEP-specific Expression in Young k=1,652
LEP-specific Expression in Old k=1,867
MEP-specific Expression in Old k=1,478
Number of DE Genes by Lineage (k)
0
1000
2000
3000
4000
Lineage Specificity in Young <30y
<30y n=11,
mLEP=16, mMEP=16
DE Status
Lineage Specificity in Old >55y
<30y n=8,
mLEP=11, mMEP=11


C
KS p < 0.0001
KS p < 0.0001
LEP-specific Genes
MEP-specific Genes
density
0.75
0.5625
0.375
0.1875
0.00
−10
−5
0
5
10
(-) lfc: Higher in LEP
(+) lfc: Higher in MEP
log2 fold change
Gene Expression
Young <30y
Old >55y


D
76%
24%
One-sided t-test
p < 0.0001
37%
63%
One-sided t-test
p < 0.0001
LEP-specific Genes
MEP-specific Genes
count
0
100
200
300
400
−4
−2
0
2
4
lfc in Young – lfc in Old
Log fold change Higher in Young
Log fold change Higher in Old

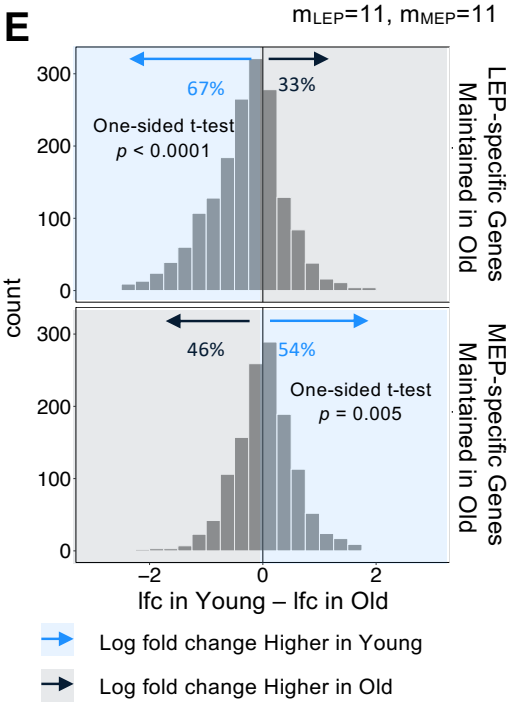
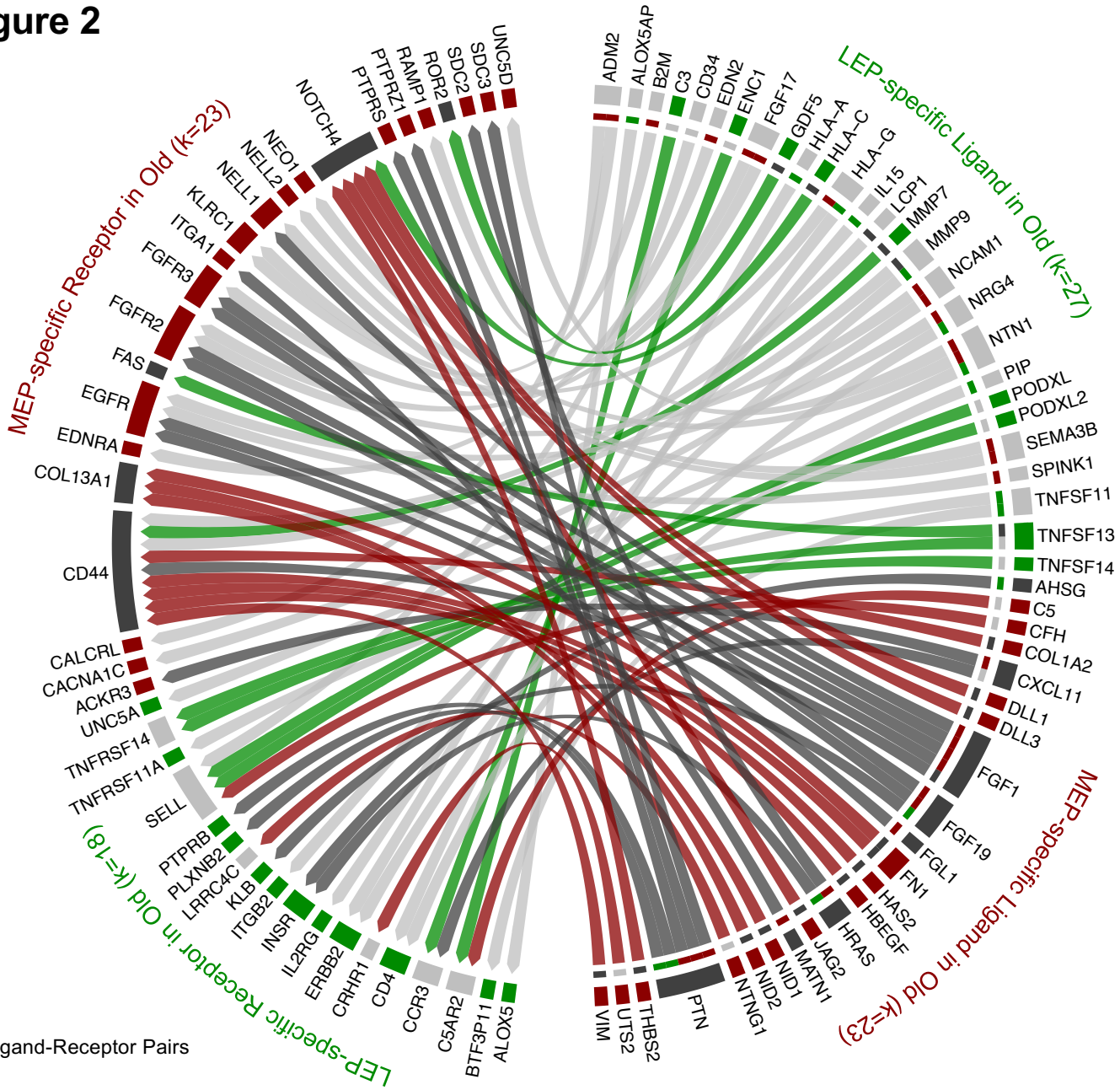

E
67%
33%
One-sided t-test
p < 0.0001
46%
54%
One-sided t-test
p = 0.005
LEP-specific Genes Maintained in Old
MEP-specific Genes Maintained in Old
count
0
100
200
300
−2
0
2
lfc in Young – lfc in Old
Log fold change Higher in Young
Log fold change Higher in Old

Figure 1. Genome-wide loss of lineage-specific expression with age. (**A**) Immunofluorescence staining of normal breast tissue showing the mammary epithelium with an apical LEPs (K19) surrounded by basal MEPs (K14). (**B**) DE LEP-specific and MEP-specific genes (adj. $p<0.001$, lfc≥1) in younger <30y (left) and older >55y (right) women. Strata plot shows changes in lineage-specific DE with age, showing the number of LEP- and MEP-specific genes gained (cyan and magenta), lost (light and dark gray), and maintained (green and red) in older women. Number of subjects (n) and sample replicates (m) in each DE analysis annotated; number of DE genes (k) in each age group and DE status indicated. (**C**) Distribution of lfc in expression between LEPs and MEPs in younger and older subjects for either DE LEP-specific (top panel) or MEP-specific (bottom panel) genes. KS *p*-values for equality of distributions of lfc between younger and older women annotated. (**D-E**) Histogram of pairwise differences in lfc in expression between LEPs and MEPs in younger vs. older women for (**D**) all genes with lineage-specific expression in younger women or (**E**) only genes that maintain lineage-specific expression in older women. Genes with LEP-specific and MEP-specific expression are shown in the top and bottom panels respectively. The percent of genes with higher lfc in younger women (light blue) or higher lfc in older women (blue gray) are indicated. One-sided t-test p-values annotated.

Figure 2

A



74 Ligand-Receptor Pairs

B

Ligand-associated KEGG Pathways

LEP

KEGG Pathway	FDR	Genes
Herpes simplex infection	1.08E-05	C3,HLA-A,HLA-C,HLA-G,IL15,TNFSF14
Cell adhesion molecules (CAMs)	4.39E-05	CD34,HLA-A,HLA-C,HLA-G,NCAM1
Antigen processing and presentation	5.09E-05	B2M,HLA-A,HLA-C,HLA-G
Cytokine-cytokine receptor interaction	0.0003	GDF5,IL15,TNFSF11,TNFSF13,TNFSF14
Type I diabetes mellitus	0.0003	HLA-A,HLA-C,HLA-G
Allograft rejection	0.0003	HLA-A,HLA-C,HLA-G
Graft-versus-host disease	0.0003	HLA-A,HLA-C,HLA-G
Phagosome	0.00038	C3,HLA-A,HLA-C,HLA-G
Autoimmune thyroid disease	0.00038	HLA-A,HLA-C,HLA-G
Viral myocarditis	0.00045	HLA-A,HLA-C,HLA-G
Kaposi's sarcoma-associated herpesvirus infection	0.00067	C3,HLA-A,HLA-C,HLA-G
Viral carcinogenesis	0.00067	C3,HLA-A,HLA-C,HLA-G
Rheumatoid arthritis	0.0011	IL15,TNFSF11,TNFSF13
HTLV-I infection	0.0017	HLA-A,HLA-C,HLA-G,IL15
Natural killer cell mediated cytotoxicity	0.0029	HLA-A,HLA-C,HLA-G
Cellular senescence	0.0052	HLA-A,HLA-C,HLA-G
Intestinal immune network for IgA production	0.0066	IL15,TNFSF13
Epstein-Barr virus infection	0.0084	HLA-A,HLA-C,HLA-G

MEP

KEGG Pathway	FDR	Genes
Breast cancer	1.71E-06	DLL1,DLL3,FGF1,FGF19,HRAS,JAG2
Endocrine resistance	4.51E-06	DLL1,DLL3,HBEGF,HRAS,JAG2
Notch signaling pathway	0.00072	DLL1,DLL3,JAG2
Pathways in cancer	0.00072	DLL1,DLL3,FGF1,FGF19,HRAS,JAG2
PI3K-Akt signaling pathway	0.00088	COL1A2,FGF1,FGF19,HRAS,THBS2
Melanoma	0.0014	FGF1,FGF19,HRAS
Th1 and Th2 cell differentiation	0.0021	DLL1,DLL3,JAG2
Gastric cancer	0.008	FGF1,FGF19,HRAS

C

Receptor-associated KEGG Pathways

LEP

KEGG Pathway	FDR	Genes
Cell adhesion molecules (CAMs)	0.0005	CD4,ITGB2,LRRC4C,SELL
Adherens junction	0.0014	ERBB2,INSR,PTPRB
Cytokine-cytokine receptor interaction	0.0019	CCR3,IL2RG,TNFRSF11A,TNFRSF14
Axon guidance	0.0092	LRRC4C,PLXNB2,UNC5A
Primary immunodeficiency	0.0092	CD4,IL2RG

MEP

KEGG Pathway	FDR	Genes
MAPK signaling pathway	0.0012	CACNA1C,EGFR,FAS,FGFR2,FGFR3
Vascular smooth muscle contraction	0.0012	CACNA1C,CALCRL,EDNRA,RAMP1
Pathways in cancer	0.0012	EDNRA,EGFR,FAS,FGFR2,FGFR3,NOTCH4
MicroRNAs in cancer	0.0012	CD44,EGFR,FGFR3,NOTCH4
Regulation of actin cytoskeleton	0.0013	EGFR,FGFR2,FGFR3,ITGA1
Proteoglycans in cancer	0.0013	CD44,EGFR,FAS,SDC2
Central carbon metabolism in cancer	0.0013	EGFR,FGFR2,FGFR3
EGFR tyrosine kinase inhibitor resistance	0.0014	EGFR,FGFR2,FGFR3
Human papillomavirus infection	0.0053	EGFR,FAS,ITGA1,NOTCH4
Cell adhesion molecules (CAMs)	0.0058	NEO1,SDC2,SDC3
PI3K-Akt signaling pathway	0.0062	EGFR,FGFR2,FGFR3,ITGA1
Graft-versus-host disease	0.0074	FAS,KLRC1
Calcium signaling pathway	0.0087	CACNA1C,EDNRA,EGFR
Bladder cancer	0.0087	EGFR,FGFR3

Figure 2. Loss of lineage fidelity with age leads to disrupted lineage-specific signaling. (**A**) Interactome map of DE lineage-specific ligand-receptor pairs (LRPs) (adj. $p<0.001$, lfc≥1) that show loss of lineage-specific expression of either ligands and/or their cognate receptors in older LEPs (light gray) or MEPs (dark gray). LRPs are connected by chord diagrams from the cell type expressing the ligand (L) to the cell type expressing the cognate receptor (R). Number of LRPs, and genes (k) in each category annotated. (**B-C**) Network functional enrichment of top KEGG pathways (FDR $p<0.001$) associated with loss of lineage-specific DE of (**b**) ligands and/or (**c**) cognate receptors in LEPs and MEPs in older women.

Figure 3

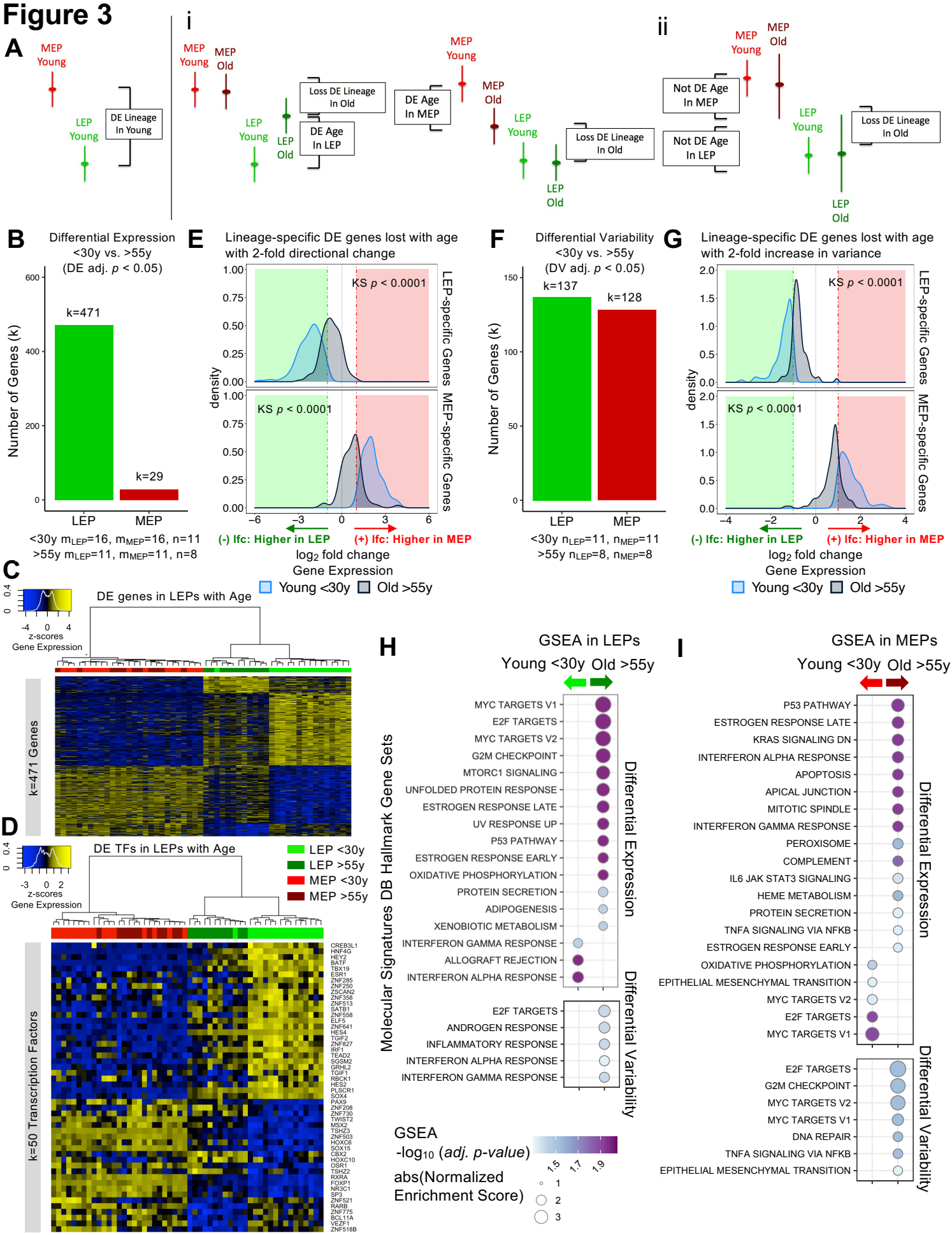
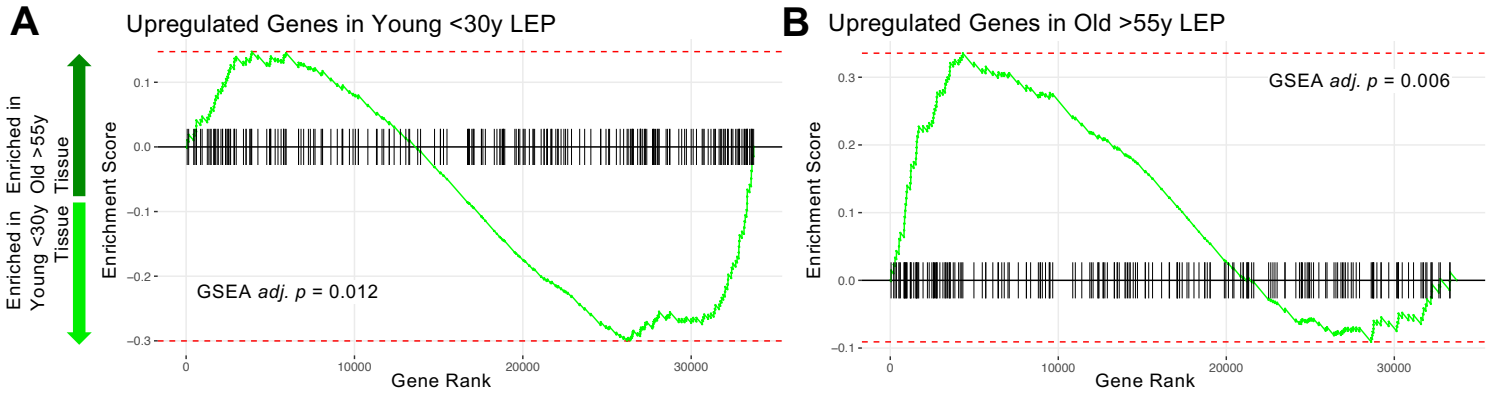
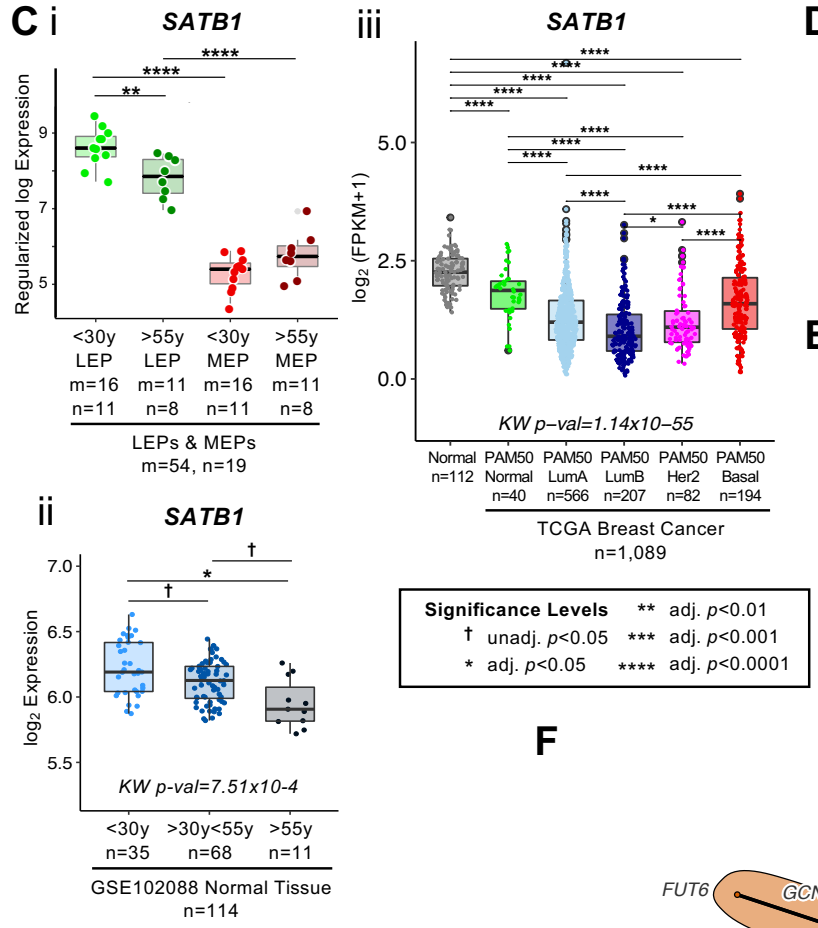
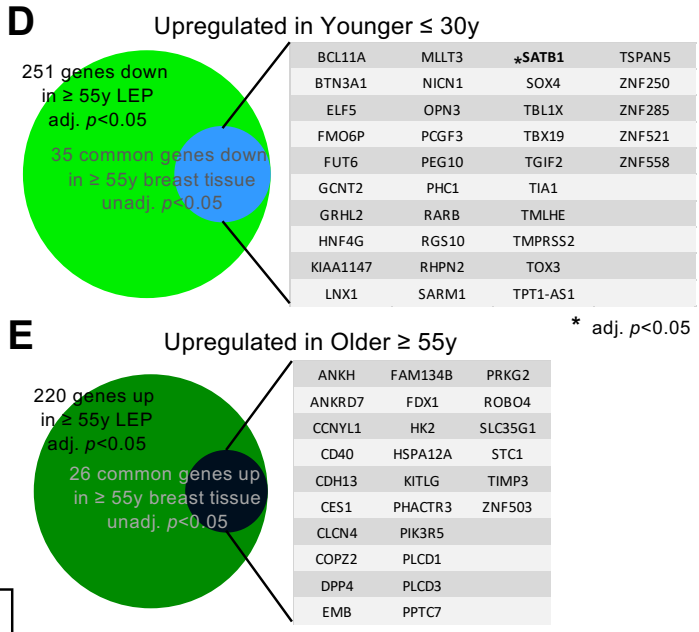

A
MEP Young
LEP Young
DE Lineage In Young
i
MEP Young
MEP Old
LEP Young
LEP Old
Loss DE Lineage In Old
DE Age In LEP
DE Age In MEP
MEP Young
MEP Old
LEP Young
LEP Old
Loss DE Lineage In Old
ii
MEP Young
MEP Old
Not DE Age In MEP
Not DE Age In LEP
LEP Young
LEP Old
Loss DE Lineage In Old
B
Differential Expression
<30y vs. >55y
(DE adj. p < 0.05)
Number of Genes (k)
k=471
k=29
LEP
MEP
<30y mLEP=16, mMEP=16, n=11
>55y mLEP=11, mMEP=11, n=8
E
Lineage-specific DE genes lost with age
with 2-fold directional change
KS p < 0.0001
KS p < 0.0001
LEP-specific Genes
MEP-specific Genes
density
(-) lfc: Higher in LEP
(+) lfc: Higher in MEP
log2 fold change
Gene Expression
Young <30y
Old >55y
F
Differential Variability
<30y vs. >55y
(DV adj. p < 0.05)
Number of Genes (k)
k=137
k=128
LEP
MEP
<30y nLEP=11, nMEP=11
>55y nLEP=8, nMEP=8
G
Lineage-specific DE genes lost with age
with 2-fold increase in variance
KS p < 0.0001
KS p < 0.0001
LEP-specific Genes
MEP-specific Genes
density
(-) lfc: Higher in LEP
(+) lfc: Higher in MEP
log2 fold change
Gene Expression
Young <30y
Old >55y
C
DE genes in LEPs with Age
z-scores
Gene Expression
k=471 Genes
D
DE TFs in LEPs with Age
LEP <30y
LEP >55y
MEP <30y
MEP >55y
z-scores
Gene Expression
k=50 Transcription Factors
CREB3L1
HNF4G
HEY2
BATF
TBX19
ESR1
ZNF285
ZNF250
ZSCAN2
ZNF358
ZNF513
SATB1
ZNF558
ELF5
ZNF641
HES4
TGIF2
ZNF827
IRF1
TEAD2
SGSM2
GRHL2
TGIF1
RBCK1
HES2
PLSCR1
SOX4
PAX9
ZNF208
ZNF730
TWIST2
MSX2
TSHZ3
ZNF503
HOXC6
SOX15
CBX2
HOXC10
OSR1
TSHZ2
RXRA
FOXP1
NR3C1
SP3
ZNF521
RARB
ZNF775
BCL11A
VEZF1
ZNF518B
H
GSEA in LEPs
Young <30y
Old >55y
Molecular Signatures DB Hallmark Gene Sets
MYC TARGETS V1
E2F TARGETS
MYC TARGETS V2
G2M CHECKPOINT
MTORC1 SIGNALING
UNFOLDED PROTEIN RESPONSE
ESTROGEN RESPONSE LATE
UV RESPONSE UP
P53 PATHWAY
ESTROGEN RESPONSE EARLY
OXIDATIVE PHOSPHORYLATION
PROTEIN SECRETION
ADIPOGENESIS
XENOBIOTIC METABOLISM
INTERFERON GAMMA RESPONSE
ALLOGRAFT REJECTION
INTERFERON ALPHA RESPONSE
Differential Expression
E2F TARGETS
ANDROGEN RESPONSE
INFLAMMATORY RESPONSE
INTERFERON ALPHA RESPONSE
INTERFERON GAMMA RESPONSE
Differential Variability
GSEA
-log10 (adj. p-value)
1.5
1.7
1.9
abs(Normalized
Enrichment Score)
1
2
3
I
GSEA in MEPs
Young <30y
Old >55y
P53 PATHWAY
ESTROGEN RESPONSE LATE
KRAS SIGNALING DN
INTERFERON ALPHA RESPONSE
APOPTOSIS
APICAL JUNCTION
MITOTIC SPINDLE
INTERFERON GAMMA RESPONSE
PEROXISOME
COMPLEMENT
IL6 JAK STAT3 SIGNALING
HEME METABOLISM
PROTEIN SECRETION
TNFA SIGNALING VIA NFKB
ESTROGEN RESPONSE EARLY
OXIDATIVE PHOSPHORYLATION
EPITHELIAL MESENCHYMAL TRANSITION
MYC TARGETS V2
E2F TARGETS
MYC TARGETS V1
Differential Expression
E2F TARGETS
G2M CHECKPOINT
MYC TARGETS V2
MYC TARGETS V1
DNA REPAIR
TNFA SIGNALING VIA NFKB
EPITHELIAL MESENCHYMAL TRANSITION
Differential Variability

Figure 3. The luminal lineage is a hotspot for age-dependent directional changes. (**A**) Models of loss of lineage fidelity illustrate hypothesized mechanisms leading to loss of lineage fidelity: (i) Age-dependent DE shifts in gene expression in LEPs and/or MEPs of older relative to younger cells; or (ii) An increase in gene expression variance in older LEPs and/or MEPs that lead to loss of detection of lineage-specific DE between LEPs and MEPs with age. (**B**) Number of DE genes (adj. $p<0.05$) between younger and older LEPs or MEPs. Number of subjects (n) and sample replicates (m) in DE analysis annotated. (**C-D**) Hierarchical clustering of all LEP and MEP samples based on sample-level expression of age-dependent (**C**) DE genes in LEPs (adj. $p < 0.05$) and (**D**) DE transcription factors in LEPs (adj. $p < 0.05$). Number of DE genes (k) indicated. Gene expression scaled regularized log (rlog) values are represented in the heatmap; clustering performed using Euclidean distances and Ward agglomerative method. (**E**) Distribution of lfc in expression between LEPs and MEPs in younger and older women for LEP-specific (top panel) or MEP-specific (bottom panel) genes that are lost with age (DE adj. $p<0.001$, lfc ≥1) and that have at least a 2-fold age-dependent directional change in the older cohort. KS *p*-values annotated. (**F**) Number of DV genes (adj. $p<0.05$) between younger and older LEPs or MEPs. Number of subjects (n) in DV analysis annotated. (**G**) Distribution of lfc in expression between LEPs and MEPs in younger and older women for LEP-specific (top panel) or MEP-specific (bottom panel) genes that are lost with age (DE adj. $p <0.001$, lfc≥1) and that have at least a 2-fold age-dependent increase in variance in the older cohort. KS *p*-values annotated. (**H-I**) MSigDB Hallmark gene sets identified by GSEA to be enriched (adj. $p<0.05$) in younger and older (**H**) LEPs and (**I**) MEPs based on age-dependent DE (top) or DV (bottom).

Figure 4


A
Upregulated Genes in Young <30y LEP
Enriched in Young <30y Tissue
Enriched in Old >55y Tissue
Enrichment Score
GSEA adj. p = 0.012
Gene Rank
B
Upregulated Genes in Old >55y LEP
GSEA adj. p = 0.006
Enrichment Score
Gene Rank


C i
SATB1
Regularized log Expression
<30y LEP m=16 n=11
>55y LEP m=11 n=8
<30y MEP m=16 n=11
>55y MEP m=11 n=8
LEPs & MEPs m=54, n19
ii
SATB1
log2 Expression
KW p-val=7.51x10-4
<30y n=35
>30y<55y n=68
>55y n=11
GSE102088 Normal Tissue n=114
iii
SATB1
log2 (FPKM+1)
KW p-val=1.14x10-55
Normal n=112
PAM50 Normal n=40
PAM50 LumA n=566
PAM50 LumB n=207
PAM50 Her2 n=82
PAM50 Basal n=194
TCGA Breast Cancer n=1,089
Significance Levels
† unadj. p<0.05
* adj. p<0.05
** adj. p<0.01
*** adj. p<0.001
**** adj. p<0.0001


D
Upregulated in Younger ≤ 30y
251 genes down in ≥ 55y LEP adj. p<0.05
35 common genes down in ≥ 55y breast tissue unadj. p<0.05
BCL11A MLLT3 *SATB1 TSPAN5
BTN3A1 NICN1 SOX4 ZNF250
ELF5 OPN3 TBL1X ZNF285
FMO6P PCGF3 TBX19 ZNF521
FUT6 PEG10 TGIF2 ZNF558
GCNT2 PHC1 TIA1
GRHL2 RARB TMLHE
HNF4G RGS10 TMPRSS2
KIAA1147 RHPN2 TOX3
LNX1 SARM1 TPT1-AS1
* adj. p<0.05
E
Upregulated in Older ≥ 55y
220 genes up in ≥ 55y LEP adj. p<0.05
26 common genes up in ≥ 55y breast tissue unadj. p<0.05
ANKH FAM134B PRKG2
ANKRD7 FDX1 ROBO4
CCNYL1 HK2 SLC35G1
CD40 HSPA12A STC1
CDH13 KITLG TIMP3
CES1 PHACTR3 ZNF503
CLCN4 PIK3R5
COPZ2 PLCD1
DPP4 PLCD3
EMB PPTC7

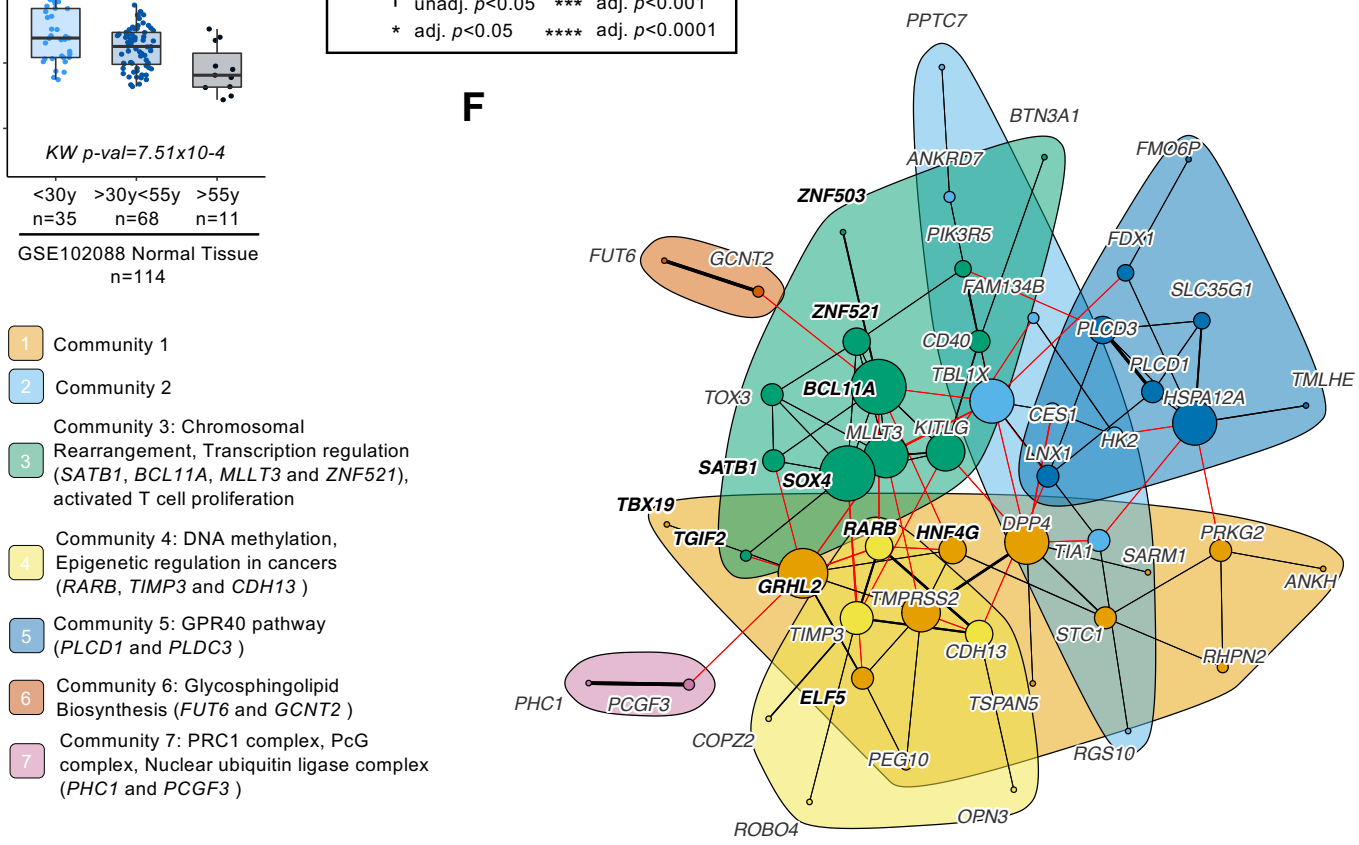

F
PPTC7
BTN3A1
ANKRD7
FMO6P
ZNF503
PIK3R5
FDX1
FUT6
GCNT2
FAM134B
SLC35G1
ZNF521
PLCD3
CD40
PLCD1
TBL1X
TMLHE
TOX3
BCL11A
HSPA12A
CES1
MLLT3
KITLG
HK2
LNX1
SATB1
SOX4
TBX19
DPP4
PRKG2
TGIF2
RARB
HNF4G
TIA1
SARM1
ANKH
GRHL2
TMPRSS2
TIMP3
STC1
CDH13
RHPN2
PHC1
PCGF3
ELF5
TSPAN5
COPZ2
PEG10
RGS10
ROBO4
OPN3
1 Community 1
2 Community 2
3 Community 3: Chromosomal Rearrangement, Transcription regulation (SATB1, BCL11A, MLLT3 and ZNF521), activated T cell proliferation
4 Community 4: DNA methylation, Epigenetic regulation in cancers (RARB, TIMP3 and CDH13)
5 Community 5: GPR40 pathway (PLCD1 and PLDC3)
6 Community 6: Glycosphingolipid Biosynthesis (FUT6 and GCNT2)
7 Community 7: PRC1 complex, PcG complex, Nuclear ubiquitin ligase complex (PHC1 and PCGF3)

Figure 4. Age-dependent directional changes in the luminal lineage are indicators of aging breast tissue. (**A-B**) GSEA enrichments plots from age-dependent DE analysis of bulk tissue showing gene ranks based on DE test statistics and gene set enrichment scores. Age-dependent enrichment of two gene sets composed of (**A**) differentially upregulated genes in younger <30y LEPs and (**B**) differentially upregulated genes in older >55y LEPs in bulk tissue are shown. Negative enrichment scores indicate upregulation of specified gene set in tissue from younger <30y women, while positive enrichment scores indicated upregulation in tissue from older >55y women. GSEA enrichment BH adj. *p*-values are annotated. (**C**) Boxplots of *SATB1* gene expression: (i) subject-level rlog values in LEPs and MEPs of younger and older women; (ii) $\log_2$ values in normal breast tissue (GSE102088); and (iii) $\log_2$ FPKM values in the TCGA breast cancer cohort by PAM50 subtype in cancers and in matched normal tissue. Age-dependent DE adj. *p*-values in normal breast tissue and LEPs and lineage-specific DE adj. *p*-values in LEPs and MEPs are indicated (i-ii). Pair-wise Wilcoxon *p*-values between groups, and KW *p*-value across matched normal and breast cancer subtypes in TCGA annotated (iii). Number of subjects (n) and sample replicates (m) in each analysis annotated. (**D-E**) Venn diagram of genes with age-dependent DE in LEPs (adj. $p<0.05$) and at least nominal DE (unadj. $p<0.05$) in normal primary breast tissue. Genes commonly (**D**) upregulated and (**E**) downregulated in LEPs and bulk tissue with age are listed. (**F**) PPI network of common age-dependent DE genes in LEPs (adj. $p<0.05$) and bulk tissue (unadj. $p<0.05$) with TFs annotated in bold. Seven gene communities identified; corresponding network functional enrichment (FDR $p<0.05$) of selected processes annotated.

Figure 5

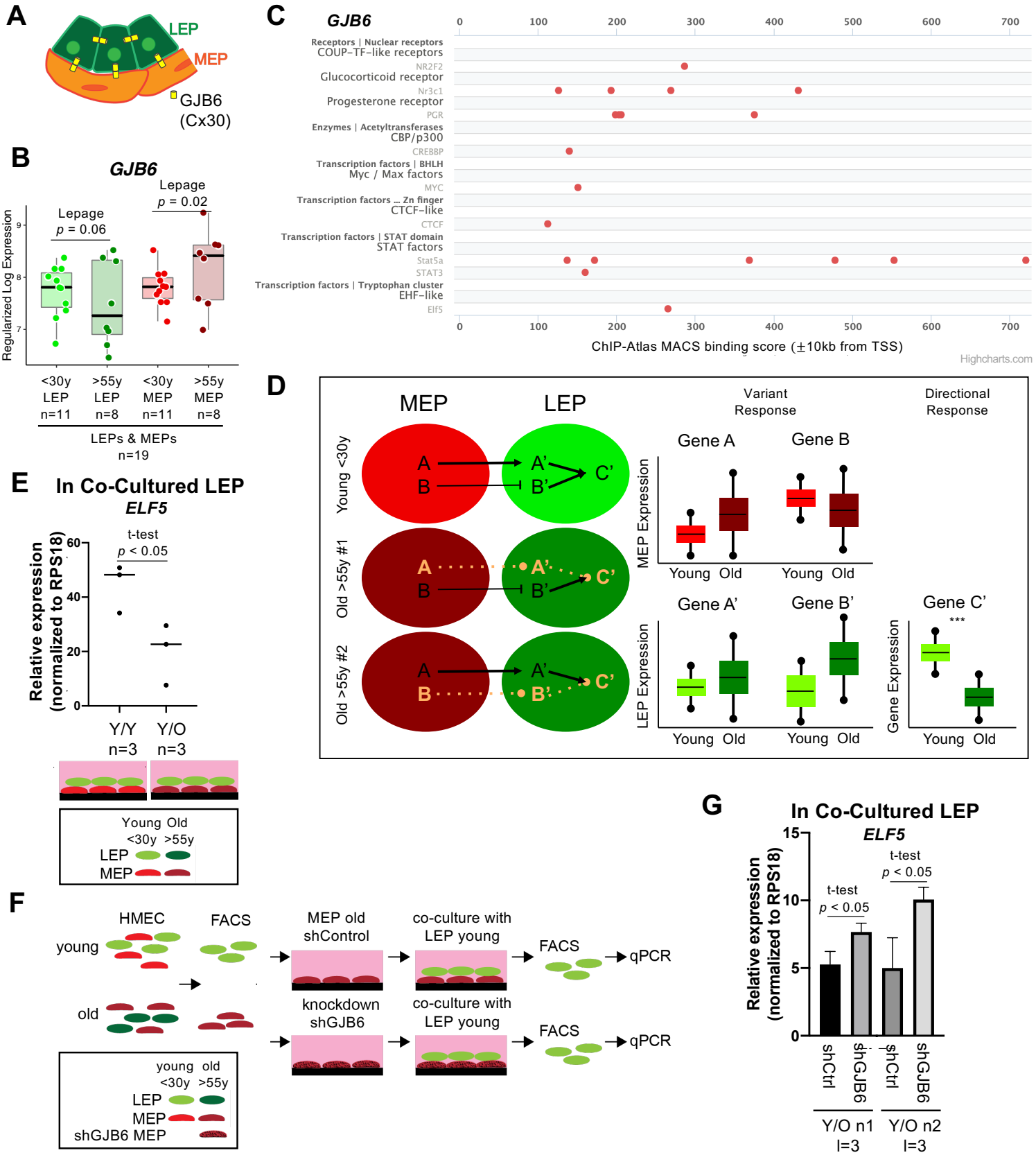
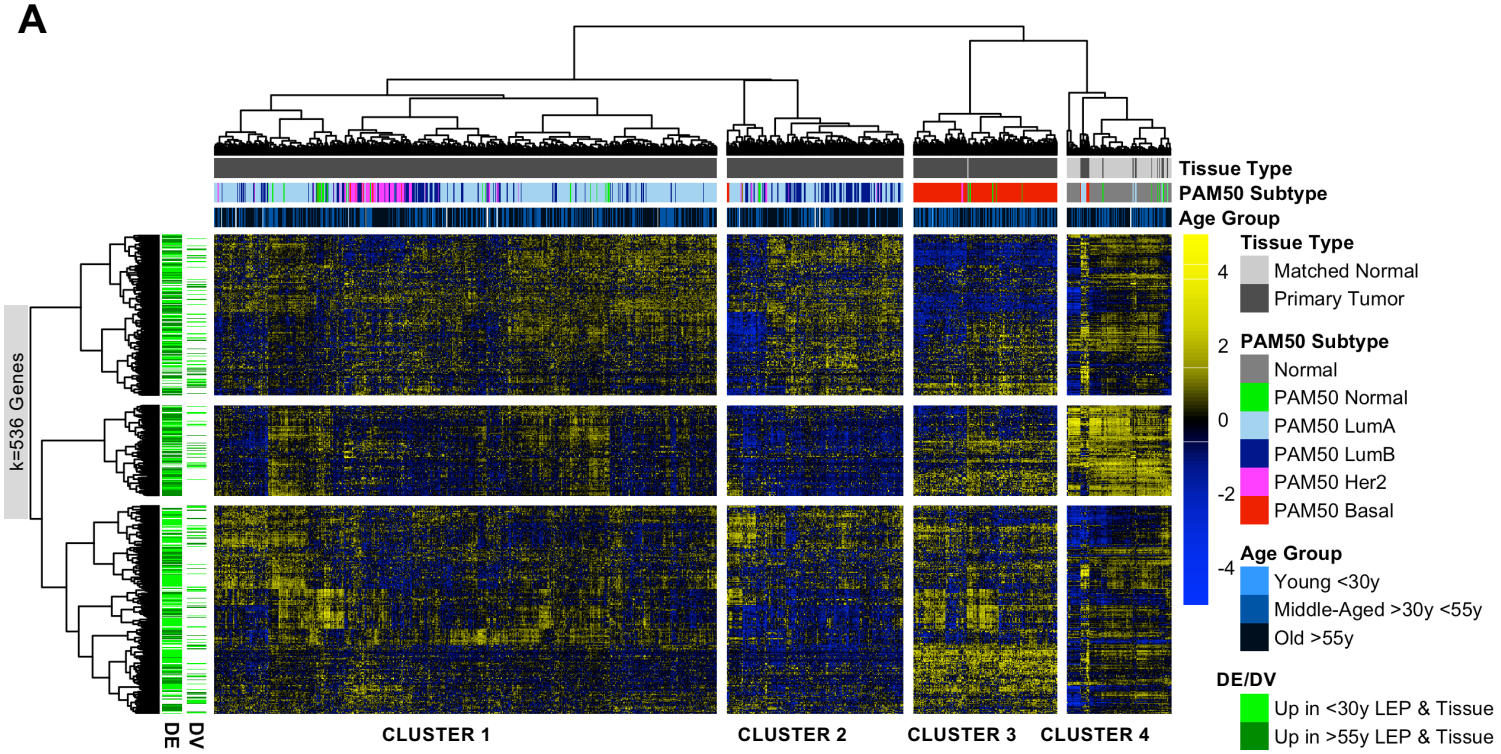

A
LEP
MEP
GJB6
(Cx30)
B
GJB6
Lepage
p = 0.06
Lepage
p = 0.02
Regularized Log Expression
<30y LEP n=11
>55y LEP n=8
<30y MEP n=11
>55y MEP n=8
LEPs & MEPs
n=19
C
GJB6
Receptors | Nuclear receptors
COUP-TF-like receptors
NR2F2
Glucocorticoid receptor
Nr3c1
Progesterone receptor
PGR
Enzymes | Acetyltransferases
CBP/p300
CREBBP
Transcription factors | BHLH
Myc / Max factors
MYC
Transcription factors ... Zn finger
CTCF-like
CTCF
Transcription factors | STAT domain
STAT factors
Stat5a
STAT3
Transcription factors | Tryptophan cluster
EHF-like
Elf5
0 100 200 300 400 500 600 700
ChIP-Atlas MACS binding score (±10kb from TSS)
Highcharts.com
D
MEP
LEP
Variant Response
Directional Response
Young <30y
Old >55y #1
Old >55y #2
A
B
A'
B'
C'
Gene A
Gene B
MEP Expression
Young Old
Gene A'
Gene B'
Gene C'
LEP Expression
Gene Expression
***
E
In Co-Cultured LEP
ELF5
t-test
p < 0.05
Relative expression (normalized to RPS18)
Y/Y n=3
Y/O n=3
Young Old
<30y >55y
LEP
MEP
F
HMEC
FACS
young
old
MEP old shControl
co-culture with LEP young
FACS
qPCR
knockdown shGJB6
co-culture with LEP young
FACS
qPCR
young old
<30y >55y
LEP
MEP
shGJB6 MEP
G
In Co-Cultured LEP
ELF5
t-test
p < 0.05
t-test
p < 0.05
Relative expression (normalized to RPS18)
shCtrl
shGJB6
shCtrl
shGJB6
Y/O n1 I=3
Y/O n2 I=3

Figure 5. *GJB6* is a mediator of the non-cell autonomous mechanism of aging in breast. (**A**) Schematic of gap junction protein, Connexin-30 (*GJB6*) homotypic channel formation between LEPs and between LEP and MEP. (**B**) Boxplot of *GJB6* subject-level rlog expression values in LEPs and MEPs in younger and older women. Lepage test *p*-values are indicated. Number of subjects (n) in analysis annotated. (**C**) *GJB6* ChIP-seq (Cistromics) binding signal within +/-10 from the TSS in mammary gland (SPP Ominer database). (**D**) Schematic illustrating integration of directional and variant responses in older epithelial cells. Different genes are dysregulated in LEPs and MEPs of older individuals leading to an increase in variance in expression across aged cells. Through cell-cell signaling, variant responses in MEPs (gene A or gene B) lead to variant responses in LEPs (gene A' or gene B'). Where these variant responses integrate and affect common downstream genes in LEPs (gene C') lead to detectable age-dependent directional changes (***) that are seen as stereotyped responses in the lineage. (**E**) Relative expression of *ELF5* in younger LEPs co-cultured with either younger (Y/Y) or older (Y/O) MEPs. Two-tailed t-test *p*-value indicated. (**F**) Schematic of co-culture methodology with HMEC cells from younger and older women enriched by FACS for LEPs and MEPs; *GJB6* knock-down in older MEP feeder layer by shRNA; younger LEPs are co-cultured on top of the older MEP feeder layer for 10 days; LEPs separated from MEPs by FACS; and LEP expression levels measured by qPCR. (**G**) Relative expression of *ELF5* in younger LEPs co-cultured with either shControl or shGJB6 older MEPs. Two-tailed t-test *p*-value indicated. Number of subjects (n) and technical replicates (l) annotated.

Figure 6

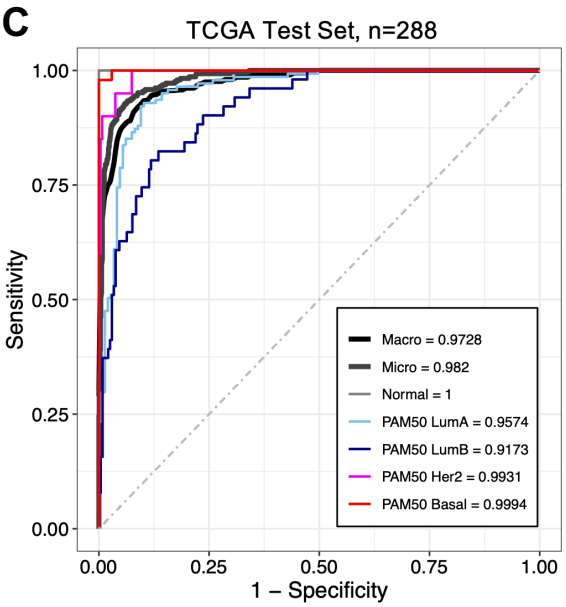
A



B

	TRAIN 75% TCGA	TEST 25% TCGA	INDPT TEST GTEx/GSE81540
Normal	84	28	180
PAM50 LumA	425	141	1,709
PAM50 LumB	156	51	767
PAM50 Her2	62	20	348
PAM50 Basal	146	48	360
TOTAL	**873**	**288**	**3,364**

C



D

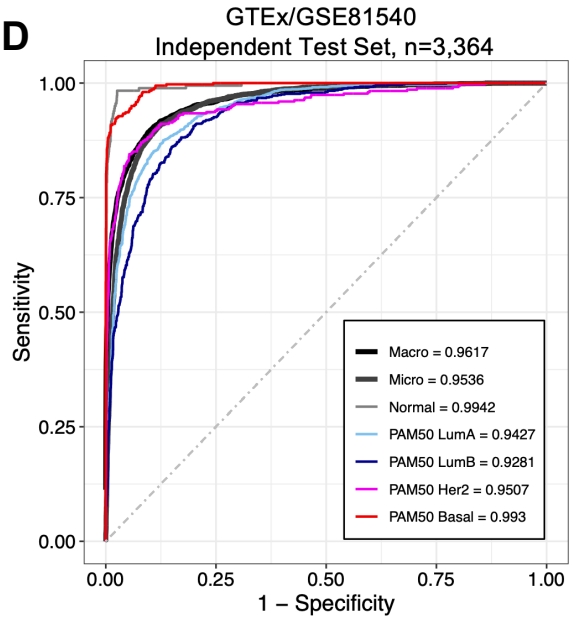


Figure 6. Age-dependent dysregulation in LEPs shape predictors of normal breast tissue and PAM50 subtypes. (**A**) Unsupervised hierarchical clustering of TCGA samples from matched normal and primary tumor tissue based on expression of age-dependent DE and DV genes identified in LEPs (adj. $p < 0.05$). PAM50 intrinsic subtypes and patient age at diagnosis are annotated. Gene expression scaled $\log_2$ FPKM values are represented in the heatmap; clustering performed using Euclidean distances and Ward agglomerative method. (Note: extreme outlier values are set to either the minimum or maximum value of the scale bar). Number of age-dependent DE and DV genes (k) included in analysis annotated. (**B**) Number of individuals in each ML class in the training, test and independent test sets. (**C-D**) Multi-class classification model performance in predicting normal tissue and breast cancer breast cancer subtypes in the (**C**) TCGA test set and (**D**) GTEx/GSE81540 independent test set. Macro AUC, micro AUC, and AUC of each group vs. rest are shown.

Figure 7

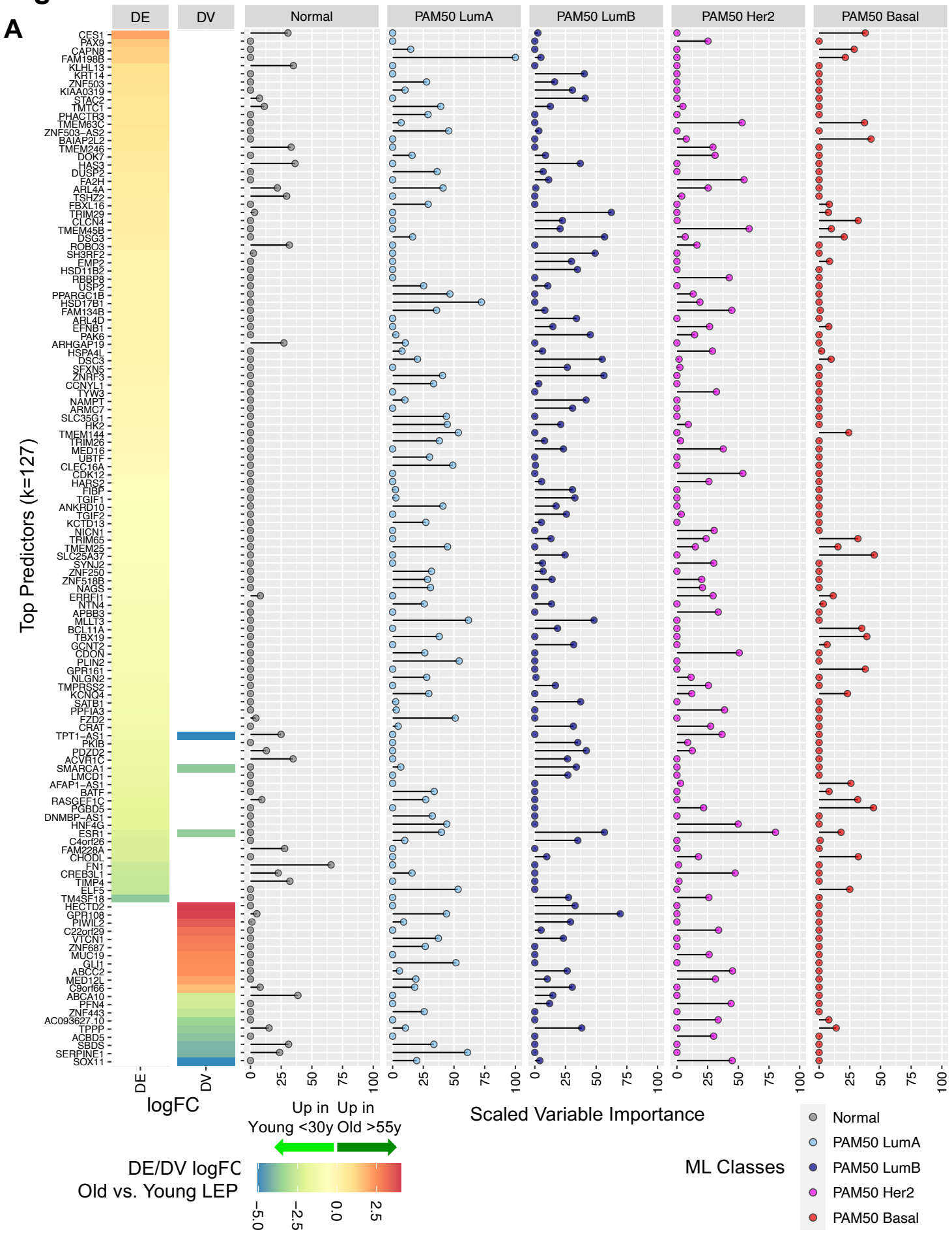

A
DE
DV
Normal
PAM50 LumA
PAM50 LumB
PAM50 Her2
PAM50 Basal
Top Predictors (k=127)
CES1
PAX9
CAPN8
FAM198B
KLHL13
KRT14
ZNF503
KIAA0319
STAC2
TMTC1
PHACTR3
TMEM63C
ZNF503−AS2
BAIAP2L2
TMEM246
DOK7
HAS3
DUSP2
FA2H
ARL4A
TSHZ2
FBXL16
TRIM29
CLCN4
TMEM45B
DSG3
ROBO3
SH3RF2
EMP2
HSD11B2
RBBP8
USP2
PPARGC1B
HSD17B1
FAM134B
ARL4D
EFNB1
PAK6
ARHGAP19
HSPA4L
DSC3
SFXN5
ZNRF3
CCNYL1
TYW3
NAMPT
ARMC7
SLC35G1
HK2
TMEM144
TRIM26
MED16
UBTF
CLEC16A
CDK12
HARS2
FIBP
TGIF1
ANKRD10
TGIF2
KCTD13
NICN1
TRIM65
TMEM25
SLC25A37
SYNJ2
ZNF250
ZNF518B
NAGS
ERRFI1
NTN4
APBB3
MLLT3
BCL11A
TBX19
GCNT2
CDON
PLIN2
GPR161
NLGN2
TMPRSS2
KCNQ4
SATB1
PPFIA3
FZD2
CRAT
TPT1−AS1
PKIB
PDZD2
ACVR1C
SMARCA1
LMCD1
AFAP1−AS1
BATF
RASGEF1C
PGBD5
DNMBP−AS1
HNF4G
ESR1
C4orf26
FAM228A
CHODL
FN1
CREB3L1
TIMP4
ELF5
TM4SF18
HECTD2
GPR108
PIWIL2
C22orf29
VTCN1
ZNF687
MUC19
GLI1
ABCC2
MED12L
C9orf66
ABCA10
PFN4
ZNF443
AC093627.10
TPPP
ACBD5
SBDS
SERPINE1
SOX11
0
25
50
75
100
logFC
Scaled Variable Importance
Up in Young <30y
Up in Old >55y
DE/DV logFC
Old vs. Young LEP
−5.0
−2.5
0.0
2.5
ML Classes
Normal
PAM50 LumA
PAM50 LumB
PAM50 Her2
PAM50 Basal

Figure 7. Expression of top aging-associated predictors of normal breast tissue and PAM50 subtypes in TCGA. (**A**) Gene predictors with scaled variable importance ≥ 25% in prediction of at least one class: normal breast tissue, PAM50 LumA, PAM50 LumB, PAM50 Her2 or PAM50 Basal, in TCGA. Rank ordered heatmap DE and DV lfc in LEPs with (+) lfc higher in older and (-) lfc higher in younger LEPs (left); scaled variable importance of each gene in each TCGA class (right). Number of gene predictors (k) annotated.

Supplementary Files

File name: Sayaman_Miyano_etal_bioRxiv2022_Supplementary_Methods

File format: .pdf

Title of data: Supplementary Methods

Description of data: A full description of the **Method Details** and **Quantification and Statistical Analysis** section of the **Methods.**

File name: Sayaman_Miyano_etal_bioRxiv2022_Supplementary_Table1

File format: .pdf

Title of data: Figure 1—table supplement 1. RNA-sequencing sample list.

Description of data: Metadata for RNA-sequencing samples used in this study.

Supplementary Tables related to Figures 1-7 will be made available upon publication.