Aging leads to DNA methylation alterations associated with loss of lineage fidelity and breast cancer in mammary luminal epithelial cells

Parijat Senapati[1], Masaru Miyano[2], Rosalyn W. Sayaman[2,4], Mudaser Basam[1,2], Candi Trac[1], Amy Leung[1], Mark A. LaBarge[2,3], Dustin E. Schones*[1,3]

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

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

[3]Irell & Manella Graduate School of Biological Sciences, City of Hope, Duarte, CA, USA

[4]Department of Laboratory Medicine, Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, San Francisco, CA, USA

*Corresponding Author:

Dustin E. Schones, Ph.D.

Department of Diabetes Complications and Metabolism

City of Hope

Duarte, CA, 91010

Email: dschones@coh.org

Phone: 626-218-1319

Abstract

Age is a significant risk factor for breast cancer and luminal epithelial cells (LEps) are a key cell lineage implicated in age-related luminal breast cancers. The aging process is characterized by alterations to the epigenome. However, the extent of age-associated DNA methylation alterations in LEps and its functional consequences have remained unclear. We report here that aging leads to distinct methylation changes in LEps that may increase susceptibility to breast cancer. Regulatory elements display consistent methylation changes at lineage-specific TF binding sites consistent with loss of lineage fidelity. CpG islands and transposable elements (TEs) showed increased methylation disorder with stochastic methylation gain and loss respectively that were also detected in luminal breast cancer. Each of these classes of methylation changes impact the regulation of genes associated with luminal breast cancer. Altogether, our results indicate that aging leads to DNA methylation changes that could determine breast cancer susceptibility.

Keywords: aging, DNA methylation, human mammary epithelial cells, luminal epithelial cells, breast cancer, transposable elements,

Introduction

Over 75% of breast cancer diagnoses are in women over the age of 50, making age one of the most significant risk factors for breast cancer (Siegel, Miller et al. 2021). This is especially true for luminal subtypes of breast cancer which account for 80% of age-related breast cancers (Jenkins, Deal et al. 2014). The mammary epithelium is composed of two principal lineages, namely luminal epithelial cells (LEps) and myoepithelial cells (MEps). Due to high transcriptional similarity LEps are the chief suspect cell of origin for luminal breast cancer (Prat and Perou 2009, Fu, Nolan et al. 2020). As most breast cancers are sporadic with no known inherited mutation or family history, a better understanding of how age-dependent molecular changes contribute to breast cancer risk is critical for the design of better therapies and prevention strategies.

Aging in the breast tissue is associated with significant physiological changes, including increased fat content and a decrease in connective tissue (Benz 2008), resulting in an altered microenvironment of the mammary epithelium. There is a complex interplay between the various epithelial cell types and microenvironment that leads to changes in molecular properties and the proportion of the different cell populations with age (Garbe, Pepin et al. 2012, Pelissier, Garbe et al. 2014, Miyano, Sayaman et al. 2017). Human mammary epithelial cells (HMECs) from older post-menopausal women, when transformed *in vitro*, exhibit gene expression patterns characteristic of luminal breast cancer subtypes. In contrast, transformed HMECs from younger pre-menopausal women show more properties of basal breast cancers (Lee, Garbe et al. 2015). Striking age-dependent changes occur in LEps, which gain expression of characteristic MEp genes (Garbe, Pepin et al. 2012, Pelissier Vatter, Schapiro et al. 2018, Sayaman, Miyano et al. 2021) and signaling molecules consistent with those of immortalized luminal cells (Pelissier Vatter, Schapiro et al. 2018). Aging increases susceptibility to cancers, and changes within the luminal lineage merit significant attention as it is the most likely lineage of origin for age-related breast cancers.

DNA methylation is a relatively stable component of the epigenome that gets established in early development and is maintained through successive cycles of cell division to preserve cell identity (Bird 2002). Studies on other model systems and tissue types, however, have shown that DNA methylation patterns can be altered with age (Maegawa, Hinkal et al. 2010, Rakyan, Down et al. 2010, Heyn, Li et al. 2012, Horvath 2013) and DNA methylation levels of specific CpG sites show a significant correlation with chronological age (Horvath 2013). This 'epigenetic clock' can be used to predict the biological age of diverse human tissues (Horvath 2013). However, this pan-tissue epigenetic clock is poorly calibrated in breast tissue, even for women without cancer (Horvath 2013). It has been speculated that the mammary gland has an aging progression unique from other tissues due to physiological factors such as changes in hormonal status with age. Previous studies have investigated DNA methylation changes in breast tissue with age, but these were either performed in bulk mammary tissue or interrogated only a subset of CpGs in the genome (Hofstatter, Horvath et al. 2018, Castle, Lin et al. 2020). A complete understanding of age-associated DNA methylation changes in human mammary LEps and how it relates to age-related breast cancer has been lacking.

To begin to address this, we performed whole-genome bisulfite sequencing (WGBS) in non-senescent primary human mammary luminal epithelial cells from pre-menopausal (young) and post-menopausal (old) women. LEps from older women showed DNA methylation changes at regulatory elements driven by altered expression and activity of lineage-specific transcription factors such as ELF5 and TP63. MEp lineage-specific regulatory elements lose methylation, whereas luminal-specific regions gain methylation. Additionally, PRC2 (polycomb repressive complex 2) regulated genes gain DNA methylation at their promoters with age, consistent with these genes having methylated promoters in luminal breast cancer (Holm, Hegardt et al. 2010). We also observed increased methylation disorder with age, resulting in stochastic methylation gain at CGIs and loss at transposable elements (TEs). TEs that lose methylation with age were

also hypomethylated in breast cancers, leading to their aberrant activity. Overall, our results suggest that aging mammary luminal epithelial cells undergo DNA methylation alterations at regulatory elements, CpG islands, and evolutionarily recent TEs that could contribute to key events relevant to luminal breast cancer pathogenesis.

Results

Aging leads to altered DNA methylation patterns in mammary luminal epithelial cells.

To investigate the genome-wide changes in DNA methylation associated with aging in the mammary tissue, we performed whole-genome bisulfite sequencing (WGBS-seq) on human mammary LEps from five younger (<30y) and five older (>55y) women (**Figure S1A, Table S1**). The age groups were chosen to represent pre-menopausal and post-menopausal status. LEps from these primary cultures were shown to maintain gene expression and DNA methylation profiles that mirror those in tissue (Miyano, Sayaman et al. 2017, Sayaman, Miyano et al. 2021). Moreover, these are non-senescent cells and therefore are a valuable resource that can be used to interrogate aging-dependent changes in the breast tissue – distinct from those arising from cellular senescence.

We sequenced all ten WGBS libraries to ~30X genome coverage, generating ~ 1B 150-bp paired-end reads (~86 Gb uniquely mapped sequence per sample) (**Table S2**) and obtained a representation of about 95% of individual CpG sites in our samples (see Methods and Tables S2, S3). We compared our DNA methylation profiles with previously published WGBS studies on distinct human mammary epithelial cell types from reduction mammoplasty (RM tissue) of a young (19y old) donor (Pellacani, Bilenky et al. 2016) and found that the cultured LEps resemble the ER-negative (ER^-) LEp population (**Figure S1B**). Moreover, young and old LEPs clustered separately indicating differences in methylomes (**Figure S1B**) which were not apparent on a global scale (**Figure S1C, D**). Pairwise comparisons of methylomes also showed higher number

of differentially methylated regions (DMRs) between young and old pairs (**Figure S1E**). Comparing the methylomes of old versus young LEps, we identified 18,860 differentially methylated regions (DMRs) (see Methods), with 7,207 showing methylation loss and 11,653 showing methylation gain in old LEps (**Figure 1A**). Most DMRs were 100-500bp in length (**Figure S1F**). The methylation changes at DMRs were confirmed with heatmaps showing methylation levels across samples (**Figure 1A**), as well as kernel smoothed methylation plots of representative DMRs (**Figure 1B**). DNA methylation changes were observed at all common genomic features including in introns (53%), gene promoters (10%), CpG islands (3%) and intergenic regions (45%) (**Figure S1G-H**).

We assigned DMRs to genes based on overlap with either promoters or gene body (exonic or intronic). Gene set enrichment analysis (GSEA) using genes with methylation loss in old LEps revealed gene sets expressed in mammary stem cells, epithelial-to-mesenchymal transition, apical junction, and genes downregulated in luminal breast cancer (**Figure 1C**). For example, the *FOXP1* gene locus shows methylation loss at several sites at the promoter and gene body (**Figure 1D**). FOXP1 is expressed in mammary stem cells residing in the basal compartment and is essential for differentiation into various mammary epithelial lineages (Fu, Pal et al. 2018). Genes with promoter methylation gain were enriched for Polycomb target gene sets, luminal specific genes upregulated in luminal breast cancer (**Figure 1E**). For example, *ALDH1A3*, an ER^- luminal lineage restricted gene (Nguyen, Pervolarakis et al. 2018), gains methylation at several sites including the promoter in old LEps (**Figure 1F**). These results indicate that aging induces DNA methylation changes at genes specific to mammary epithelial lineages and luminal subtype breast cancers.

Aging-associated DNA methylation changes occur at lineage-specific transcription factor binding sites

To investigate whether DNA methylation changes were associated with transcription factor (TF) binding signatures, we performed motif enrichment analysis using position weight matrices (PWMs) from the JASPAR database (Fornes, Castro-Mondragon et al. 2020). DMRs that lose methylation are enriched for binding sites of TFs such as EGR1, TP63, TP73, TEAD1, TEAD4, and RUNX1 (**Figure 2A left panel**). Several of these TFs, including TEAD, TP63, TP73, NFI, and EGR1, have been previously shown to be enriched at active regulatory elements in basal/myoepithelial cells (Pellacani, Bilenky et al. 2016). To investigate whether these TFs change expression with age, we checked their expression using RNA-sequencing data from LEps and MEps of young and old individuals (Sayaman, Miyano et al. 2021). *TP63* and *NFIA* showed increased expression in old compared to young LEps (**Figure 2B and S2A**). Although not all TFs showed an increase in expression with age, we did note that *TP63, NFIA (NFI family TF; binds to NFI half-sites), EGR1, and TP73* have relatively higher expression in MEps as compared to LEps (**Figure 2A right panel and 2E**). The expression patterns of TFs and the hypomethylation of their targets are concordant with the known roles of these TFs. For example, young LEps do not express *TP63*, which is a MEp lineage-specific transcription factor (Miyano, Sayaman et al. 2017). Furthermore, TEAD1 and TEAD4 are DNA binding partners for YAP (Yes-associated protein), a transcriptional co-activator regulated by the Hippo signaling pathway. YAP is sequestered in the cytoplasm through activity of kinases downstream of the Hippo pathway (Badouel, Garg et al. 2009). Old LEps exhibit increased expression of *TP63* as well as increased nuclear YAP localization (Pelissier, Garbe et al. 2014, Miyano, Sayaman et al. 2017), and thereby increased YAP activity and TEAD binding. YAP is also known to bind TP73 and RUNX2 (Badouel, Garg et al. 2009). These results suggest that increased expression/activity of myoepithelial-specific TFs in old LEps could lead to increased binding and loss of DNA methylation in these regions (**Figure 2F**).

Conversely, DMRs with methylation gain are enriched for binding sites of the ETS family of transcription factors EHF, ELF3, ELF5, ELF4, ETV1, ETS1, ETV4, GABPA, and NFI half and full sites (**Figure 2C left panel**). ELF5 is a luminal lineage-specific transcription factor that exhibits decreased expression with age (**Figure 2D**) (Miyano, Sayaman et al. 2017). Besides *ELF5*, transcription factors *EHF* and *ELF3* show relatively higher expression in LEps as compared to MEps (**Figure 2C right panel and 2E**). These results further suggest that decreased expression of luminal lineage-specific TFs may lead to a decrease in binding and protection of binding sites from DNA methyltransferases (DNMTs). This, in turn, may result in the methylation and repression of these luminal lineage-specific regions (**Figure 2F**). We also checked for the known DNA methylation binding sensitivities (Yin, Morgunova et al. 2017) for enriched TF binding sites. TFs with binding sites enriched at methylation loss DMRs either do not have a CpG or are unaffected by DNA methylation (**Table S4**). Presumably they recruit other factors such as TETs to bring about DNA methylation changes. However, TFs with binding sites enriched at methylation gain DMRs are sensitive to DNA methylation (**Table S4**). Generally, sites of methylation loss were enriched for myoepithelial-specific TF binding motifs (**Figure 2E**), whereas sites of methylation gain were enriched for luminal lineage-specific TF binding motifs (**Figure 2E**). For example, SOX4 sites are enriched in DMRs with methylation gain, and it is a luminal-specific TF (**Figure 2E**). SOX15, on the other hand, is a myoepithelial-specific TF and its binding sites are enriched at DMRs with methylation loss (**Figure 2E**). Expression of *SOX4* decreases, whereas *SOX15* increases with age in old LEps (**Figure S2B-C**), indicating loss of expression of luminal lineage-specific TFs and increased expression of myoepithelial-specific TFs. Collectively, these results suggest that aging-associated DNA methylation changes in LEps are orchestrated by lineage-specific transcription factors.

Myoepithelial lineage-specific regulatory elements lose methylation, whereas luminal-specific regulatory elements gain methylation with age.

Given that aging-associated DNA methylation changes occurred at sites containing lineage-specific TF binding sites, we examined the functional consequences of these changes by measuring chromatin accessibility using ATAC-seq in young and old LEps (**Table S1**). We found that DMRs with methylation loss showed corresponding increased accessibility in the older LEps (**Figure 3A-B**). On the other hand, DMRs with methylation gain showed a decrease in chromatin accessibility (**Figure 3B-C**).-These results indicate that functional regulatory elements undergo DNA methylation changes with age.

To further investigate the functional relevance of the age-dependent DNA methylation changes in the breast tissue, we used published histone modification ChIP-seq datasets in mammary epithelial lineages to predict chromatin states at DMRs (Pellacani, Bilenky et al. 2016). To discern whether DMRs reflect changes in cell properties, we predicted chromatin states in ER^+ luminal, ER^- luminal, and basal (MEp) cells from RM tissue of a young (19y old) donor (Pellacani, Bilenky et al. 2016). We classified DMRs into seven chromatin states in each cell type based on the extent of overlap with each state (see Methods; **Figure S3A-C**). DMRs with methylation loss in old LEps overlap primarily with MEp-specific promoter and enhancer regions, accounting for about 80% of all DMRs (**Figure 3E**). As these sites also contain binding sites for TFs enriched at MEp-specific active enhancers (Pellacani, Bilenky et al. 2016), these results indicate that MEp-specific regulatory elements lose methylation in old LEps. Active enhancer sites are hypomethylated (Lister, Pelizzola et al. 2009, Stadler, Murr et al. 2011) and methylation loss at MEp-specific regulatory elements might indicate an epigenetic shift towards MEp gene expression patterns in LEps with age. Indeed, we do observe an increase in expression of MEp genes in old luminal cells (**Figure 1H**) (Sayaman, Miyano et al. 2021). These data are in line

with previous studies suggesting that aging luminal cells acquire basal-like gene expression signatures (Garbe, Pepin et al. 2012, Pelissier, Garbe et al. 2014).

Conversely, DMRs that gain methylation overlap with putative promoter and enhancer regions in ER^- luminal cells, accounting for about 42% of all DMRs (**Figure 3F**). As these DMRs also contain binding sites for TFs enriched at luminal lineage-specific active enhancers (Pellacani, Bilenky et al. 2016), these results indicate that luminal-specific regulatory regions gain DNA methylation in aging LEps. Together, these results suggest that aging LEps acquire permissive chromatin states at MEp-specific regions and repress ER^- luminal-specific chromatin regions through changes in DNA methylation.

Polycomb repressive complex 2 target genes also gain DNA methylation with age.

Age-dependent DNA methylation gain at Polycomb repressive complex 2 (PRC2) target gene promoters has been observed in other tissues and model systems (Maegawa, Hinkal et al. 2010, Teschendorff, Menon et al. 2010, Beerman, Bock et al. 2013, Sun, Luo et al. 2014, Jung and Pfeifer 2015). We previously found that genes with methylation gain at their promoters were enriched in PRC2 target genes identified in ESCs (**Figure 1E**). Moreover, about 15% of the DMRs with methylation gain overlapped with Polycomb repressed regions marked by H3K27me3 (**Figure 3F and S3A-C**). Since the PRC2 target gene sets were based on datasets derived from embryonic stem cells, we used H3K27me3 ChIP-seq data from ER^-, ER^+ luminal cells, and basal cells (Pellacani, Bilenky et al. 2016) to determine if they were indeed PRC2 targets in mammary epithelial lineages. We found that 39-46% of promoters that gain methylation with age are marked by H3K27me3 across the three cell lineages (**Figure 4A**). In contrast, only 14-27% of promoters with methylation loss are enriched for H3K27me3 (**Figure 4B**). Additionally, promoters with DNA methylation gain have conserved H3K27me3 occupancy in all three epithelial lineages (**Figure 4C**) as compared to promoters with DNA methylation loss

that show H3K27me3 only in the luminal lineage (**Figure 4D**). These results indicate that promoters with age-dependent DNA methylation gain are indeed *bona fide* PRC2 target genes in mammary epithelial cells. These results are consistent with PRC2 protecting unmethylated CpG-rich regions and that erosion of the PRC2 targeting with aging leads to increased methylation (Jung and Pfeifer 2015).

Aging leads to stochastic methylation loss at transposable elements and methylation gain at CpG islands.

Aging is associated with increased transcriptional variability, cell-to-cell heterogeneity, and DNA methylation variability (Bahar, Hartmann et al. 2006, Martinez-Jimenez, Eling et al. 2017, Cheung, Vallania et al. 2018, Hernando-Herraez, Evano et al. 2019). To examine whether there is an increase in methylation heterogeneity in older individuals, we calculated methylation entropy, a measure of methylation heterogeneity between adjacent CpGs, for 150 bp non-overlapping windows across the genome (Jenkinson, Pujadas et al. 2017). There are two modes by which a given region could increase methylation entropy; a) a completely unmethylated region could gain methylation stochastically or b) a completely methylated region could lose methylation (**Figure 5A-B**). We identified regions that increase methylation entropy in older individuals and found a net decrease in methylation levels at these regions (**Figure S4A-D**), consistent with the fact that most of the genome is methylated in normal cells. Of these, unmethylated regions with increase in methylation entropy were mainly at CpG islands (69.2%; 1017 of 1469). Conversely, methylated regions with increase in entropy were mainly at repeat regions (70.3%; 76575 of 108877). We then focused our analysis on all unmethylated CpG islands (CGIs) with increased entropy and found that they show a net gain in methylation in old individuals (**Figure 5C**). Given that one of the major roles of DNA methylation in mammalian genomes is silencing repetitive elements, we examined the repeat elements with increase in methylation entropy. We found that

they were enriched at primate specific Alu elements. Alu elements showed a net decrease in methylation with increase in methylation entropy (**Figure 5D**). These results indicate that increased DNA methylation heterogeneity with age leads to stochastic methylation gain at CpG islands and methylation loss at repeats.

To determine whether stochastic methylation gain and loss accompanied by increased disorder has any consequences in breast cancer, we used a published WGBS dataset of primary tumors containing ER+ (luminal) and ER- (basal) samples (Brinkman, Nik-Zainal et al. 2019). We found that at CGIs with increased entropy, luminal breast tumors showed an even greater increase in methylation and entropy (**Figure 5E**). In contrast, basal tumors did not show an increase in CpG island methylation or entropy (**Figure 5F**), indicating that this behavior is specific to the molecular subtype of breast cancer. We identified the subset of CGIs that were hypermethylated in at least 20% (>=6) of the breast tumors. These CGIs were enriched at PRC2 target genes (**Figure S4E**) which have been shown to be hypermethylated in luminal breast cancer (Holm, Hegardt et al. 2010) and other types of cancer. We also found gene sets that show differential expression between luminal and basal breast cancer to be enriched at these CGIs (**Figure S4E, Figure 5G**). Of note, PRC2 target genes are repressed in normal tissues through H3K27me3 and although they show DNA methylation gain at their CGI promoters, expression changes are not expected. Altogether, these results indicate that increased DNA methylation heterogeneity with age leads to stochastic methylation gain at CpG islands which is exacerbated in luminal breast cancer.

Aging leads to increased DNA methylation variability between old individuals at TEs.

We next tested whether in addition to increased methylation entropy, aging would lead to increased interindividual variability. We indeed do observe increased variability of CpG

methylation in older individuals (**Figure S5A**). Since some amount of DNA methylation variability is expected in human populations and differences at individual CpG sites may not have functional consequences, we first identified regions of variable methylation across all 10 individuals using the Systematic Interindividual Variation (SIV) approach (Gunasekara, Scott et al. 2019) (see Methods). We identified 23,921 regions that were variably methylated only in old samples and 25,585 regions that were specific to the young samples (**Figure S5B**). Regions variably methylated in old samples mainly showed methylation loss (17,780 of 23,921; 74.3%). The variable regions usually were 150bp-450bp in length with some spanning 2-2.5kbs (**Figure S5C and S5D**). We next examined whether variable regions contained TEs and found that 68.7% of regions variable in young as compared to 67.59% of control non-variable regions. 71.4% regions variable in old overlapped with TEs as compared to 67.65% of control non-variable regions (**Figure S5E**). Among the TEs that are variably methylated in old, about 76.9% lose methylation and 23.13% gain methylation (**Figure S5F**). We further confirmed that variable methylation at TEs was not due to mappability issues (**Figure S5G**).

We further investigated the characteristics of TEs that are variably methylated in old. The evolutionarily recent elements (<100 million years) (Burns and Boeke 2012) tended towards DNA methylation loss, whereas the older TE elements (>100 million years) tended to gain methylation in old cells (**Figure 6A-D)**. These results indicate that evolutionarily recent TEs that are likely to be regulated by DNA methylation can lose methylation with age in a stochastic manner, contributing to increased interindividual variation.

We further inspected the specific subfamilies enriched at the variably methylated regions. While many evolutionarily recent Alu elements such as AluY and AluS showed stochastic loss of methylation, these elements are highly abundant in the genome (**Figure 6E**). Other TE

subfamilies show enriched at regions with variable methylation loss AluSq2, L1PA(2-6), HERVH, AluSg4, AluYc, L1P1, L1P2, and MER11A elements (**Figure 6E-F**).

Transposable elements with age-dependent methylation loss are activated in breast cancer.

TEs have latent regulatory potential and, when not repressed, can function as regulatory elements. Given our observations that evolutionarily recent, CpG dense transposons lose methylation stochastically with age, we examined whether breast tumors show loss of epigenetic suppression at these elements. Loss of TE suppression can lead to aberrant transcription of TE-derived transcripts and alternative isoforms due to splicing with neighboring genes in cancer cells (Jang, Shah et al. 2019). To investigate the potential of this happening in breast cancer, we examined published RNA-seq data from high-grade Ductal carcinoma in situ (DCIS), an early form of breast cancer (Abba, Gong et al. 2015). We used the SQuIRE pipeline (Yang, Ardeljan et al. 2019) to measure gene and TE expression changes in DCIS as compared to normal tissues. We found 554 age-hypomethylated TEs that showed higher expression in DCIS. Of these, 468 were variably hypomethylated with age and 86 were within DMRs with methylation loss. HERVH elements and L1PA2 elements were most common among the TEs that showed aberrant expression. A representative example of a HERVH element that loses methylation in aging LEps and breast tumors is shown (**Figure 7A**). Aberrant transcription from this HERVH element was observed only in DCIS samples (**Figure 7A**). We also found examples of exonization of TEs, where a MIRb element was included within the exon of a non-coding RNA DRAIC expressed in DCIS (**Figure S6A)**. While aberrant expression of hypomethylated TEs in DCIS was clear, we did not observe any breast cancer subtype specificity in the expression of these TEs.

We next examined whether TEs with methylation loss function as regulatory elements in breast cancer. To this end we utilized ATAC-seq data from The Cancer Genome Atlas (TCGA) breast cancer cohort (TCGA-BRCA) and published ChIP-seq datasets from breast cancer cell lines (Franco, Nagari et al. 2018). We identified 1197 TCGA-BRCA ATAC-seq peaks that overlap TEs with methylation loss in aging LEps (**Figure 7B**). Most of the TEs had open chromatin sites across luminal tumors especially in lobular carcinomas (**Figure 7B**). Some TEs were accessible in basal subtype tumors as well. These results indicate that TEs hypomethylated with age could function as regulatory elements in breast cancer.

To determine whether loss of methylation at TEs can lead to changes in expression of genes important in breast cancer, we mapped the 554 age-hypomethylated TEs to genes within ±500kb and checked correlation of expression of TE with each gene mapped within this window in DCIS samples (Abba, Gong et al. 2015). We identified 266 genes which were significantly correlated with the expression of a nearby TE. A representative example of an L1PA4 element with significant expression correlation (correlation: 0.8, *p*-val: $7.29e^{-9}$) with nearby *ZNF92* gene is shown (**Figure S7A**). We find that *ZNF92* is also highly expressed in luminal breast tumors (Luminal A and Luminal B) as compared to basal tumors (**Figure 7C**). These results indicate that TEs hypomethylated with age could potentially regulate the expression of genes important for luminal breast cancer and thereby support transformation.

Discussion

Previous studies analyzing aging-associated DNA methylation changes in human mammary luminal epithelial cells have either used array-based technologies (Miyano, Sayaman et al. 2017, Hofstatter, Horvath et al. 2018, Castle, Lin et al. 2020, Sayaman, Miyano et al. 2021) or were performed using composite tissues that represented a mixture of cells (Hofstatter, Horvath et al. 2018, Castle, Lin et al. 2020). Array probes are mainly enriched at genic regions

and intergenic elements including repeats and distal regulatory elements are under-represented (Bell, Lowe et al. 2019). It has furthermore been shown that different cell lineages have unique responses to aging (Miyano, Sayaman et al. 2017, Kimmel, Penland et al. 2019), making pure cell populations important for interpretation. Our use of WGBS on sorted normal human mammary epithelial cells addresses both shortcomings.

We find that aging leads to DNA methylation changes at lineage-specific gene regulatory elements in mammary luminal epithelial cells, in-line with previous studies that show loss of lineage fidelity in luminal cells with age (Miyano, Sayaman et al. 2017). Specifically, luminal cells show DNA methylation loss at myoepithelial-specific regulatory elements containing MEp-specific TF binding sites. Moreover, luminal-specific regulatory regions containing LEp-specific TF binding sites are repressed by DNA methylation with age. This loss of lineage-fidelity has been shown to occur with age and in mammary tissue that is susceptible to breast cancer (Sayaman, Miyano et al. 2021, Shalabi, Miyano et al. 2021).

Previous DNA methylome studies in human cells suggested a global loss of methylation with age (Heyn, Li et al. 2012, Vandiver, Irizarry et al. 2015, Jenkinson, Pujadas et al. 2017). These studies were performed using $CD4^+$ T cells from a neonate and a centenarian (Heyn, Li et al. 2012), sorted $CD4^+$ T cells (Jenkinson, Pujadas et al. 2017) and skin keratinocytes (Vandiver, Irizarry et al. 2015) from young (<25 y) and old (>70 y) individuals. The global methylation loss with age was accompanied by a global increase in entropy or methylation discordance with age (Jenkinson, Pujadas et al. 2017). We did not observe a global methylation loss or a global increase in entropy in mammary luminal epithelial cells with age. This could be due to higher fidelity of DNA methylation maintenance in luminal cells or due to a smaller separation between the age of young (<30 years) and old (> 55 years) individuals in our study. This observation is supported by DNA methylome studies in aging mouse livers, where no global loss of methylation was observed, although specific changes at regulatory elements were

detected (Cole, Robertson et al. 2017). Whereas no global methylation loss was detected, we did, however, find methylation changes at regulatory elements and variable methylation changes at evolutionarily recent transposons.

Global loss of methylation and focal increase in methylation, termed ‘epigenetic drift’, has been reported to occur with age and in cancer (Issa 2014). Focal increase in methylation has been reported to occur mainly at CpG islands. We find CpG islands showing both gain and loss in methylation reflecting changes in lineage-specific gene expression. However, we do observe increased methylation disorder/entropy at CpG islands that lead to methylation gain with age that is enhanced further in luminal breast cancer. These CpG islands were predominantly PRC2 targets that are known to gain methylation in other cancer subtypes (Easwaran, Johnstone et al. 2012). PRC2 targets were also found among genes that showed methylation gain at promoters. Loss/repositioning of PRC2 has been proposed to cause DNA methylation gain at its targets. Presumably, CpG islands with increased methylation entropy are in the process of becoming hypermethylated in luminal breast cancer. Hypermethylation of PRC2 targets seems to be specific to the luminal subtype of breast cancer (Holm, Hegardt et al. 2010), although the reasons for this is not currently clear.

DNA methylation is an important mechanism for silencing of transposable elements. Reactivation of transposable elements has been shown in cancer cells, coincident with a global loss of DNA methylation (Burns 2017, Ishak and De Carvalho 2020). While previous research comparing DNA methylation levels between a neonate and a centenarian reported that 12% of DMRs that lost methylation with age were located on Alu elements (Heyn, Li et al. 2012), changes in epigenetic silencing at TEs with aging has not been systematically investigated. Increased expression of retrotransposons and active retrotransposition have been reported in aging mouse tissues (De Cecco, Criscione et al. 2013, Benayoun, Pollina et al. 2019) and increased LINE-1 expression has been reported in senescent human fibroblasts *in vitro* (De

Cecco, Ito et al. 2019). However, these studies used bulk quantitative strategies such as RT-qPCR that are not ideal due to a high sequence between multiple active and inactive copies. Our approach allowed us to investigate the impact of aging on DNA methylation at TEs using non-senescent mammary epithelial cells. Our results demonstrate that evolutionarily recent CpG-dense transposons are more likely to be silenced by DNA methylation and show increased variability and methylation loss with age.

We further show that the methylation loss observed at TEs with age becomes exacerbated in frank luminal breast cancers. The derepressed TEs have the potential to lead to aberrant transcription or function as regulatory elements in breast cancer cells. We demonstrate that hypomethylated TEs could regulate genes associated with luminal subtype breast cancers. Retrotransposon activation has been widely observed in nearly every human cancer, including breast cancer (Rodic, Sharma et al. 2014, Rooney, Shukla et al. 2015), and even in premalignant lesions (Ishak and De Carvalho 2020). This suggests that early events in tumorigenesis may be driven by retrotransposon activity. We show here, for the first time, that loss of transposon silencing is evident in aging luminal epithelial cells that could predispose towards breast cancer. Future studies are needed to characterize the role of these TEs in the early events of mammary cancer initiation. Altogether, our results showing DNA methylation alterations at regulatory elements, CpG islands and TEs with aging provide further support to the observation that epigenetic changes with age could determine susceptibility towards breast cancer.

Materials and Methods

Isolation of mammary luminal epithelial cells

LEps were isolated from 4th passage (p4) cultures of pre-stasis finite-lifespan HMECs from normal reduction mammoplasty (RM) tissues (Garbe, Pepin et al. 2012). Primary non-immortalized HMECs were generated and maintained, as described previously (Labarge, Garbe

et al. 2013). HMEC strains were established in the laboratory of Dr. Martha Stampfer (Lawrence Berkeley National Laboratory) who has all required IRB approvals to distribute these samples. Cells were grown in M87A medium with cholera toxin and oxytocin at 0.5 ng/ml and 0.1nM, respectively. 4th passage HMECs were stained with anti-human CD133-PE (Biolegend, clone 7) and anti-human CD271 (Biolegend, clone ME20.4) by following standard flow cytometry protocol. LEps and MEps were enriched using Aria III (Becton Dickinson) flow sorter as CD133^{+}/CD271^{-} and CD133^{-}/CD271^{+} fractions, respectively. The different HMEC strains used are described in Table S1.

WGBS

We examined the DNA methylome of primary LEps from younger (< 30y) premenopausal and older (> 55y) postmenopausal women. Genomic DNA from FACS enriched LEps were extracted using Quick-DNA Microprep kit (Zymo Research). Whole-genome bisulfite sequencing (WGBS-seq) was performed at Admera Health Biopharma Services. Genomic DNA was bisulfite converted using the Zymo EZ DNA methylation kit (Zymo Research) per manufacturer's protocol. Library preparation was then performed using Accel-NGS® Methyl-Seq DNA Library kit (Swift Biosciences, Michigan, USA) per manufacturer's recommendations. Equimolar pooling of libraries was performed based on QC values. Samples were sequenced on an Illumina® NovaSeq S4 (Illumina, California, USA) with a read length configuration of 150bp paired-end. We obtained about 38X coverage per genome to an average of 414 million paired-end reads (Table S2).

Sequencing reads were hard trimmed 15 bp from both ends using Trim Galore (https://github.com/FelixKrueger/TrimGalore) (version 0.4.5) to remove unwanted methylation bias arising from the library preparation. Adapter contamination was also removed using Trim Galore using --illumina option. Trimmed reads were aligned to the human reference genome (hg19) using Bismark (version 0.19.1) (Krueger and Andrews 2011) with the --bowtie1 option

(Langmead, Trapnell et al. 2009). Bismark (deduplicate_bismark) to remove PCR duplicates. Methpipe (Song, Decato et al. 2013) (version 3.4.3) was used to determine the methylation status of individual CpG sites. The hg19 reference genome contains about 56.4 million CpG loci (hg19) or 28.2 million CpG dyads. Methylation levels were calculated for individual CpG dyads and are henceforth referred to as CpG sites.

Power Analysis

We used previously generated RRBS data from LEps derived from tissue of young (n=6) and old (n=4) women for determining sample size for WGBS. We used MethylSeqDesign (Liu, Lin et al. 2021) to determine sample size needed for different discovery rates. We found that a sample size of 5 per group allows 74.82% power for detecting differences in methylation (FDR < 0.05). A post-hoc analysis of our WGBS data using n=3 as pilot data on chromosome 1 also showed that we have 75% power with n=5 per age group.

DMR identification

DMRfinder (Gaspar and Hart 2017) was used to identify DMRs. Briefly, CpG sites with at least 4X coverage in at least 6 out of 10 samples were determined. Genomic regions with at least 4 CpG sites within a distance of 100 bp were identified. The maximum length of regions was set to 500 bp. Of these regions, ones that had less than 20 reads covering them were discarded. The final set of regions was then used to calculate differentially methylated regions with greater than 20% change (*P*-value < 0.05) between pooled young and old samples.

Characterization of DMRs

Overlaps with genomic features were calculated using bedtools intersect (Quinlan and Hall 2010) with masked CpG islands from the UCSC genome browser (Kent, Sugnet et al. 2002) and repeats from the RepeatMasker annotation (Smit 2013-2015). DMR overlaps with CpG islands, and repeat elements were determined using bedtools intersect such that the overlapping region contains at least 1 CpG site.

Kernel smoothed plots were made using the BSSeq package with default smoothing parameters and CpG sites with greater than 3X coverage (Hansen, Langmead et al. 2012). Heatmaps were made using the pheatmap package (Kolde 2018).

Identification of variably methylated regions

Variably methylated regions were identified using the Systematic Interindividual Variation (SIV) approach (Gunasekara, Scott et al. 2019). Briefly, the human genome was divided into 100 bp non-overlapping windows containing at least 1 CpG site. About 14.4 million such windows were obtained. The coverage and methylation values were calculated for each sample across these 100bp windows using Methpipe. Windows were then filtered by coverage as follows to obtain informative bins: Windows with CpG <= 2 that had a coverage of at least 10 reads and windows with CpG > 2 that had a coverage of at least 10 reads at half the number of CpG sites. We obtained about 12.9 million bins that had sufficient coverage. Next, we calculated the Pearson correlation co-efficient for each bin with its adjacent bin. Adjacent bins that were within 100 bp of each other and a correlation >= 0.71 were merged into correlated blocks of methylation. The average methylation for each sample was calculated for the correlated blocks. Correlated blocks which had an interindividual range (IR) of >= 20% and at least 4 CpG sites were defined as variable regions. Old specific variable regions were defined as regions that had an IR of >= 20% across the five old samples and IR_{old} -IR_{young} >= 20%. Similarly, young specific variable regions were defined as regions that had an IR of >= 20% across the five young samples and IR_{young} - IR_{old} >= 20%. These regions were further filtered to retain those that have coverage of at least 3 CpGs in all 10 samples. To categorize variably methylated regions into variable loss or variable gain, the highest (old_{max}) and the lowest (old_{min}) methylation values among old samples were compared to the methylation average in young (avg_y). The absolute values of differences in methylation between avg_y and $old_{max,}$ old_{min} were compared. If the difference in methylation

between the avg_y and old_{min} were higher than that between avg_y and old_{max}, that region was assigned as a variable region with methylation loss

Chromatin state predictions

ChIP-seq data for mammary epithelial populations were downloaded using the International Human Epigenome Consortium data portal (https://epigenomesportal.ca/ihec/). Bigwig files containing ChIP-seq signal for six different histone modifications in three cell types (sample IDs: CEMT0154, CEMT0155, and CEMT0156) were used (**Figure S4A-C**). ChIP-seq signals from the bigwig files were binned into 200 bp windows. These binned signal files were then binarized using BinarizeSignal in ChromHMM (version 1.20) (Ernst and Kellis 2017). The binarized files were used to run ChromHMM LearnModel with 16 states. Functionally similar chromatin states were grouped together to seven distinct chromatin states (**Figure S4C**). DMR overlaps with chromatin states were performed using bedtools intersect. In cases where a DMR overlapped more than one chromatin state, DMR was assigned to the state with a higher base overlap.

ATAC-seq and analysis

50,000 live FACS sorted young (*n* = 4) and old (*n* = 4) LEps were used for Tn5 transposition in duplicates using the Omni-ATAC method (Corces, Trevino et al. 2017). Libraries were sequenced to a depth of 40 million paired reads each on an Illumina HiSeq 2500 system. Raw sequencing reads were trimmed using Trim Galore (https://github.com/FelixKrueger/TrimGalore) (version 0.4.5) to remove adapters. Trimmed reads were aligned to the human reference genome (hg19) using bowtie (Langmead, Trapnell et al. 2009) with the --best -k 1 -X 2000 --mm --chunkmbs 1024 parameters. Unmapped reads, mates as well as low quality reads and secondary alignments were removed using samtools. PCR duplicates were identified using Picard and removed using samtools. The reads were shifted by 4 bases on the sense strand and 5 bases on the anti-sense strand to account for Tn5 bias. Reads from technical replicates were pooled to generate pseudoreplicates. Peaks were

called using MACS2 version 2.1.1.20160309 using -p 0.01 --nomodel --shift -37 --extsize 73 --SPMR --keep-dup all --call-summits as parameters. Peak calls from technical replicates and pseudo-replicates were compared using IDR to get the most conservative set of peaks for each sample.

RNA-seq analysis

Total RNA was isolated using Quick-RNA Microprep kit (Zymo Research) from FACS sorted young (*n* = 10) and old LEps (*n* = 8). Similarly, total RNA was isolated from young (*n* = 6) MEps. PolyA-enriched RNA was isolated and used for library preparation using the TruSeq RNA Library Prep kit (Illumina). About 50 million single-end reads of length 51 bp were generated for all samples. Raw sequences were aligned to the hg19 reference genome using HISAT2 (version 2.1.0) (Kim, Langmead et al. 2015) with the --dta parameter. Stringtie (version 1.3.4) (Pertea, Pertea et al. 2015) was used with default parameters to assemble transcripts using the Gencode v19 transcript annotation. Assembled transcripts from all libraries were further merged using --merge option in Stringtie. Merged transcript abundances were measured using bedtools coverage.

For the old LEp versus young LEp comparison, edgeR (v3.26.4) (Robinson, McCarthy et al. 2010) was used to conduct differential gene expression analysis. Briefly, normalization factors were calculated using the trimmed mean of M values method (TMM) to eliminate composition biases between libraries. Genewise biological dispersion was estimated using edgeR's empirical Bayes approach. Then, we employed a generalized linear model (GLM) approach to account for the unnormalized distribution of gene counts, as well as additional factors, such as batch effects. Differentially regulated genes were identified with an absolute log2 fold change $\geq$ 0.5 and FDR < 0.05. Gene expression boxplots were generated using batch corrected RPKM values. Batch correction of RPKM values was done using limma package (Ritchie, Phipson et al. 2015). For

LEp versus MEp comparison, DESeq2 (Love, Huber et al. 2014) was used to normalize counts and identify differentially expressed genes (absolute log2 fold change ≥ 0.5 and p-value < 0.05). RNA-seq data from normal and DCIS samples were obtained from GEO (Accession number: GSE69240 (Abba, Gong et al. 2015)). Fastq reads were quality and adapter trimmed using Trim Galore with the --paired option. Quality trimmed reads were aligned to the hg19 genome using STAR (version 2.6.0.a) (Dobin, Davis et al. 2013) with --sjdbOverhang 100 --winAnchorMultimapNmax 100 --outFilterMultimapNmax 100 --outFilterMismatchNmax 3 --alignEndsType EndToEnd options. The SQuIRE pipeline (Yang, Ardeljan et al. 2019) was used to measure gene and TE expression changes in DCIS as compared to normal tissues.

Enrichment of DMRs and variably methylated regions at TE subfamilies

The extent of overlap of DMRs and variable regions at repat subfamilies (RepeatMasker annotation) was calculated using bedtools intersect. Randomly shuffled control regions of the same length were generated 1000 times, and the expected number of overlaps calculated using bedtools intersect. Enrichment values were calculated by dividing the number of DMR overlaps (observed) with the average number of random overlaps (expected). *P* values were calculated by dividing the number of extreme values (at either tail of distribution) from random overlaps by 1000. Significantly enriched repeat subfamilies had $p < 0.05$, enrichment > 1.5-fold, and > 5 copies overlapping DMRs or variable regions.

Additional data analysis and external data

Mappability of 100bp bisulfite converted kmers for the hg19 genome generated using Bismap (Karimzadeh, Ernst et al. 2018) were downloaded from the UCSC Genome Browser (Kent, Sugnet et al. 2002). Bedgraph files were used to obtain the mappability score of individual CpGs within TEs. Motif scanning was performed using fimo (Grant, Bailey et al. 2011) with JASPAR 2020 database with default parameters (Fornes, Castro-Mondragon et al. 2020). Motif enrichment analysis was performed using findMotifsGenome.pl in homer (Heinz, Benner et al.

2010). For each set of genomic regions, background sequences with matched GC content were selected. *P*-values for motif enrichment was calculated using the cumulative binomial distributions. The evolutionary age of TEs was calculated as described previously using the Jukes-Cantor correction (Bourque, Leong et al. 2008). Normalized methylation entropy and mean methylation were calculated on 150bp non-overlapping windows across the genome using informME (Jenkinson, Abante et al. 2018). Normalized counts for TCGA-BRCA cancer specific ATAC-seq peaks were downloaded from the Genomic Data Commons (GDC) Portal. The clinical information and PAM50 subtypes for these samples were accessed using TCGAbiolinks (Colaprico, Silva et al. 2016). Primary breast tumor WGBS data generated as part of the International Cancer Genome Consortium (Brinkman, Nik-Zainal et al. 2019) and was obtained from European Genome-phenome Archiver after DACO approval. Figures were made in R (version 3.5.1) using the ggplot2 package (Wickham 2016) and assembled in Adobe Illustrator.

Availability of data and materials

All raw and processed sequencing data files have been deposited at NCBI's Gene Expression Omnibus (GEO) under accession GSE153696.

Funding

This work was supported by the National Institutes of Health, grants R01DK112041, R01CA220693 (D.E.S.); K01DK104993 (A.L.); Department of Defense/Army Breast Cancer Era of Hope Scholar Award BC141351, and Expansion Award BC181737, Conrad N. Hilton Foundation, and City of Hope Center for Cancer and Aging (M.A.L.); the National Cancer Institute (NCI) Cancer Metabolism Training Program Postdoctoral Fellowship T32CA221709 (R.W.S.) and City of Hope Program in Molecular and Cellular Biology pilot award (D.E.S and M.A.L.). The research reported in this publication included work performed in the Integrative

Genomics and Analytical Cytometry Cores supported by the National Cancer Institute of the National Institutes of Health under award 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 funding bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Authors' contributions

PS, MAL and DES designed the study. PS, MM and CT performed experiments. PS, RWS, MB and AL analyzed and interpreted the data. PS, MAL and DES prepared the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank other members of the Schones laboratory for helpful discussions and comments, and the continued involvement of patient advocates Susan Samson and Sandy Preto. 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 ENCODE Consortium: https://www.encodeproject.org/.

Conflict of Interest Statement

The authors declare that they have no potential conflicts of interest.

References

Abba, M. C., T. Gong, Y. Lu, J. Lee, Y. Zhong, E. Lacunza, M. Butti, Y. Takata, S. Gaddis, J. Shen, M. R. Estecio, A. A. Sahin and C. M. Aldaz (2015). "A Molecular Portrait of High-Grade Ductal Carcinoma In Situ." Cancer Res **75**(18): 3980-3990.

Badouel, C., A. Garg and H. McNeill (2009). "Herding Hippos: regulating growth in flies and man." Curr Opin Cell Biol **21**(6): 837-843.

Bahar, R., C. H. Hartmann, K. A. Rodriguez, A. D. Denny, R. A. Busuttil, M. E. Dolle, R. B. Calder, G. B. Chisholm, B. H. Pollock, C. A. Klein and J. Vijg (2006). "Increased cell-to-cell variation in gene expression in ageing mouse heart." Nature **441**(7096): 1011-1014.

Beerman, I., C. Bock, B. S. Garrison, Z. D. Smith, H. Gu, A. Meissner and D. J. Rossi (2013). "Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging." Cell Stem Cell **12**(4): 413-425.

Bell, C. G., R. Lowe, P. D. Adams, A. A. Baccarelli, S. Beck, J. T. Bell, B. C. Christensen, V. N. Gladyshev, B. T. Heijmans, S. Horvath, T. Ideker, J. J. Issa, K. T. Kelsey, R. E. Marioni, W. Reik, C. L. Relton, L. C. Schalkwyk, A. E. Teschendorff, W. Wagner, K. Zhang and V. K. Rakyan (2019). "DNA methylation aging clocks: challenges and recommendations." Genome Biol **20**(1): 249.

Benayoun, B. A., E. A. Pollina, P. P. Singh, S. Mahmoudi, I. Harel, K. M. Casey, B. W. Dulken, A. Kundaje and A. Brunet (2019). "Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses." Genome Res **29**(4): 697-709.

Benz, C. C. (2008). "Impact of aging on the biology of breast cancer." Crit Rev Oncol Hematol **66**(1): 65-74.

Bird, A. (2002). "DNA methylation patterns and epigenetic memory." Genes Dev **16**(1): 6-21.

Bourque, G., B. Leong, V. B. Vega, X. Chen, Y. L. Lee, K. G. Srinivasan, J. L. Chew, Y. Ruan, C. L. Wei, H. H. Ng and E. T. Liu (2008). "Evolution of the mammalian transcription factor binding repertoire via transposable elements." Genome Res **18**(11): 1752-1762.

Brinkman, A. B., S. Nik-Zainal, F. Simmer, F. G. Rodriguez-Gonzalez, M. Smid, L. B. Alexandrov, A. Butler, S. Martin, H. Davies, D. Glodzik, X. Zou, M. Ramakrishna, J. Staaf, M. Ringner, A. Sieuwerts, A. Ferrari, S. Morganella, T. Fleischer, V. Kristensen, M. Gut, M. J. van de Vijver, A. L. Borresen-Dale, A. L. Richardson, G. Thomas, I. G. Gut, J. W. M. Martens, J. A. Foekens, M. R. Stratton and H. G. Stunnenberg (2019). "Partially methylated domains are hypervariable in breast cancer and fuel widespread CpG island hypermethylation." Nat Commun **10**(1): 1749.

Burns, K. H. (2017). "Transposable elements in cancer." Nat Rev Cancer **17**(7): 415-424.

Burns, K. H. and J. D. Boeke (2012). "Human transposon tectonics." Cell **149**(4): 740-752.

Castle, J. R., N. Lin, J. Liu, A. M. V. Storniolo, A. Shendre, L. Hou, S. Horvath, Y. Liu, C. Wang and C. He (2020). "Estimating breast tissue-specific DNA methylation age using next-generation sequencing data." Clin Epigenetics **12**(1): 45.

Cheung, P., F. Vallania, H. C. Warsinske, M. Donato, S. Schaffert, S. E. Chang, M. Dvorak, C. L. Dekker, M. M. Davis, P. J. Utz, P. Khatri and A. J. Kuo (2018). "Single-Cell Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging." Cell **173**(6): 1385-1397 e1314.

Colaprico, A., T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi and H. Noushmehr (2016). "TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data." Nucleic Acids Res **44**(8): e71.

Cole, J. J., N. A. Robertson, M. I. Rather, J. P. Thomson, T. McBryan, D. Sproul, T. Wang, C. Brock, W. Clark, T. Ideker, R. R. Meehan, R. A. Miller, H. M. Brown-Borg and P. D. Adams (2017). "Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions." Genome Biol **18**(1): 58.

Corces, M. R., A. E. Trevino, E. G. Hamilton, P. G. Greenside, N. A. Sinnott-Armstrong, S. Vesuna, A. T. Satpathy, A. J. Rubin, K. S. Montine, B. Wu, A. Kathiria, S. W. Cho, M. R. Mumbach, A. C. Carter, M. Kasowski, L. A. Orloff, V. I. Risca, A. Kundaje, P. A. Khavari, T. J. Montine, W. J. Greenleaf and H. Y. Chang (2017). "An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues." Nat Methods **14**(10): 959-962.

De Cecco, M., S. W. Criscione, A. L. Peterson, N. Neretti, J. M. Sedivy and J. A. Kreiling (2013). "Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues." Aging (Albany NY) **5**(12): 867-883.

De Cecco, M., T. Ito, A. P. Petrashen, A. E. Elias, N. J. Skvir, S. W. Criscione, A. Caligiana, G. Brocculi, E. M. Adney, J. D. Boeke, O. Le, C. Beausejour, J. Ambati, K. Ambati, M. Simon, A. Seluanov, V. Gorbunova, P. E. Slagboom, S. L. Helfand, N. Neretti and J. M. Sedivy (2019). "L1 drives IFN in senescent cells and promotes age-associated inflammation." Nature **566**(7742): 73-78.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner." Bioinformatics **29**(1): 15-21.
Easwaran, H., S. E. Johnstone, L. Van Neste, J. Ohm, T. Mosbruger, Q. Wang, M. J. Aryee, P. Joyce, N. Ahuja, D. Weisenberger, E. Collisson, J. Zhu, S. Yegnasubramanian, W. Matsui and S. B. Baylin (2012). "A DNA hypermethylation module for the stem/progenitor cell signature of cancer." Genome Res **22**(5): 837-849.
Ernst, J. and M. Kellis (2017). "Chromatin-state discovery and genome annotation with ChromHMM." Nat Protoc **12**(12): 2478-2492.
Fornes, O., J. A. Castro-Mondragon, A. Khan, R. van der Lee, X. Zhang, P. A. Richmond, B. P. Modi, S. Correard, M. Gheorghe, D. Baranasic, W. Santana-Garcia, G. Tan, J. Cheneby, B. Ballester, F. Parcy, A. Sandelin, B. Lenhard, W. W. Wasserman and A. Mathelier (2020). "JASPAR 2020: update of the open-access database of transcription factor binding profiles." Nucleic Acids Res **48**(D1): D87-D92.
Franco, H. L., A. Nagari, V. S. Malladi, W. Li, Y. Xi, D. Richardson, K. L. Allton, K. Tanaka, J. Li, S. Murakami, K. Keyomarsi, M. T. Bedford, X. Shi, W. Li, M. C. Barton, S. Y. R. Dent and W. L. Kraus (2018). "Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis." Genome Res **28**(2): 159-170.
Fu, N. Y., E. Nolan, G. J. Lindeman and J. E. Visvader (2020). "Stem Cells and the Differentiation Hierarchy in Mammary Gland Development." Physiol Rev **100**(2): 489-523.
Fu, N. Y., B. Pal, Y. Chen, F. C. Jackling, M. Milevskiy, F. Vaillant, B. D. Capaldo, F. Guo, K. H. Liu, A. C. Rios, N. Lim, A. J. Kueh, D. M. Virshup, M. J. Herold, H. O. Tucker, G. K. Smyth, G. J. Lindeman and J. E. Visvader (2018). "Foxp1 Is Indispensable for Ductal Morphogenesis and Controls the Exit of Mammary Stem Cells from Quiescence." Dev Cell **47**(5): 629-644 e628.
Garbe, J. C., F. Pepin, F. A. Pelissier, K. Sputova, A. J. Fridriksdottir, D. E. Guo, R. Villadsen, M. Park, O. W. Petersen, A. D. Borowsky, M. R. Stampfer and M. A. Labarge (2012). "Accumulation of multipotent progenitors with a basal differentiation bias during aging of human mammary epithelia." Cancer Res **72**(14): 3687-3701.
Gaspar, J. M. and R. P. Hart (2017). "DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data." BMC Bioinformatics **18**(1): 528.
Grant, C. E., T. L. Bailey and W. S. Noble (2011). "FIMO: scanning for occurrences of a given motif." Bioinformatics **27**(7): 1017-1018.
Gunasekara, C. J., C. A. Scott, E. Laritsky, M. S. Baker, H. MacKay, J. D. Duryea, N. J. Kessler, G. Hellenthal, A. C. Wood, K. R. Hodges, M. Gandhi, A. B. Hair, M. J. Silver, S. E. Moore, A. M. Prentice, Y. Li, R. Chen, C. Coarfa and R. A. Waterland (2019). "A genomic atlas of systemic interindividual epigenetic variation in humans." Genome Biol **20**(1): 105.
Hansen, K. D., B. Langmead and R. A. Irizarry (2012). "BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions." Genome Biol **13**(10): R83.
Heinz, S., C. Benner, N. Spann, E. Bertolino, Y. C. Lin, P. Laslo, J. X. Cheng, C. Murre, H. Singh and C. K. Glass (2010). "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities." Mol Cell **38**(4): 576-589.
Hernando-Herraez, I., B. Evano, T. Stubbs, P. H. Commere, M. Jan Bonder, S. Clark, S. Andrews, S. Tajbakhsh and W. Reik (2019). "Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells." Nat Commun **10**(1): 4361.
Heyn, H., N. Li, H. J. Ferreira, S. Moran, D. G. Pisano, A. Gomez, J. Diez, J. V. Sanchez-Mut, F. Setien, F. J. Carmona, A. A. Puca, S. Sayols, M. A. Pujana, J. Serra-Musach, I. Iglesias-Platas, F. Formiga, A. F. Fernandez, M. F. Fraga, S. C. Heath, A. Valencia, I. G. Gut, J. Wang and M. Esteller (2012). "Distinct DNA methylomes of newborns and centenarians." Proc Natl Acad Sci U S A **109**(26): 10522-10527.
Hofstatter, E. W., S. Horvath, D. Dalela, P. Gupta, A. B. Chagpar, V. B. Wali, V. Bossuyt, A. M. Storniolo, C. Hatzis, G. Patwardhan, M. K. Von Wahlde, M. Butler, L. Epstein, K. Stavris, T. Sturrock,

A. Au, S. Kwei and L. Pusztai (2018). "Increased epigenetic age in normal breast tissue from luminal breast cancer patients." Clin Epigenetics **10**(1): 112.
Holm, K., C. Hegardt, J. Staaf, J. Vallon-Christersson, G. Jonsson, H. Olsson, A. Borg and M. Ringner (2010). "Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns." Breast Cancer Res **12**(3): R36.
Horvath, S. (2013). "DNA methylation age of human tissues and cell types." Genome Biol **14**(10): R115.
Ishak, C. A. and D. D. De Carvalho (2020). "Reactivation of Endogenous Retroelements in Cancer Development and Therapy." Annual Review of Cancer Biology **4**.
Issa, J. P. (2014). "Aging and epigenetic drift: a vicious cycle." J Clin Invest **124**(1): 24-29.
Jang, H. S., N. M. Shah, A. Y. Du, Z. Z. Dailey, E. C. Pehrsson, P. M. Godoy, D. Zhang, D. Li, X. Xing, S. Kim, D. O'Donnell, J. I. Gordon and T. Wang (2019). "Transposable elements drive widespread expression of oncogenes in human cancers." Nat Genet **51**(4): 611-617.
Jenkins, E. O., A. M. Deal, C. K. Anders, A. Prat, C. M. Perou, L. A. Carey and H. B. Muss (2014). "Age-specific changes in intrinsic breast cancer subtypes: a focus on older women." Oncologist **19**(10): 1076-1083.
Jenkinson, G., J. Abante, A. P. Feinberg and J. Goutsias (2018). "An information-theoretic approach to the modeling and analysis of whole-genome bisulfite sequencing data." BMC Bioinformatics **19**(1): 87.
Jenkinson, G., E. Pujadas, J. Goutsias and A. P. Feinberg (2017). "Potential energy landscapes identify the information-theoretic nature of the epigenome." Nat Genet **49**(5): 719-729.
Jung, M. and G. P. Pfeifer (2015). "Aging and DNA methylation." BMC Biol **13**: 7.
Karimzadeh, M., C. Ernst, A. Kundaje and M. M. Hoffman (2018). "Umap and Bismap: quantifying genome and methylome mappability." Nucleic Acids Res **46**(20): e120.
Kent, W. J., C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler and D. Haussler (2002). "The human genome browser at UCSC." Genome Res **12**(6): 996-1006.
Kim, D., B. Langmead and S. L. Salzberg (2015). "HISAT: a fast spliced aligner with low memory requirements." Nat Methods **12**(4): 357-360.
Kimmel, J. C., L. Penland, N. D. Rubinstein, D. G. Hendrickson, D. R. Kelley and A. Z. Rosenthal (2019). "Murine single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging." Genome Res **29**(12): 2088-2103.
Kolde, R. (2018). pheatmap: Pretty Heatmaps. R package version 1.0.10.
Krueger, F. and S. R. Andrews (2011). "Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications." Bioinformatics **27**(11): 1571-1572.
Labarge, M. A., J. C. Garbe and M. R. Stampfer (2013). "Processing of human reduction mammoplasty and mastectomy tissues for cell culture." J Vis Exp(71).
Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.
Lee, J. K., J. C. Garbe, L. Vrba, M. Miyano, B. W. Futscher, M. R. Stampfer and M. A. LaBarge (2015). "Age and the means of bypassing stasis influence the intrinsic subtype of immortalized human mammary epithelial cells." Front Cell Dev Biol **3**: 13.
Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren and J. R. Ecker (2009). "Human DNA methylomes at base resolution show widespread epigenomic differences." Nature **462**(7271): 315-322.
Liu, P., C. W. Lin, Y. Park and G. Tseng (2021). "MethylSeqDesign: a framework for Methyl-Seq genome-wide power calculation and study design issues." Biostatistics **22**(1): 35-50.
Love, M. I., W. Huber and S. Anders (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biol **15**(12): 550.

Maegawa, S., G. Hinkal, H. S. Kim, L. Shen, L. Zhang, J. Zhang, N. Zhang, S. Liang, L. A. Donehower and J. P. Issa (2010). "Widespread and tissue specific age-related DNA methylation changes in mice." Genome Res **20**(3): 332-340.

Martinez-Jimenez, C. P., N. Eling, H. C. Chen, C. A. Vallejos, A. A. Kolodziejczyk, F. Connor, L. Stojic, T. F. Rayner, M. J. T. Stubbington, S. A. Teichmann, M. de la Roche, J. C. Marioni and D. T. Odom (2017). "Aging increases cell-to-cell transcriptional variability upon immune stimulation." Science **355**(6332): 1433-1436.

Miyano, M., R. W. Sayaman, M. H. Stoiber, C. H. Lin, M. R. Stampfer, J. B. Brown and M. A. LaBarge (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.

Nguyen, Q. H., N. Pervolarakis, K. Blake, D. Ma, R. T. Davis, N. James, A. T. Phung, E. Willey, R. Kumar, E. Jabart, I. Driver, J. Rock, A. Goga, S. A. Khan, D. A. Lawson, Z. Werb and K. Kessenbrock (2018). "Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity." Nat Commun **9**(1): 2028.

Pelissier, F. A., J. C. Garbe, B. Ananthanarayanan, M. Miyano, C. Lin, T. Jokela, S. Kumar, M. R. Stampfer, J. B. Lorens and M. A. LaBarge (2014). "Age-related dysfunction in mechanotransduction impairs differentiation of human mammary epithelial progenitors." Cell Rep **7**(6): 1926-1939.

Pelissier Vatter, F. A., D. Schapiro, H. Chang, A. D. Borowsky, J. K. Lee, B. Parvin, M. R. Stampfer, M. A. LaBarge, B. Bodenmiller and J. B. Lorens (2018). "High-Dimensional Phenotyping Identifies Age-Emergent Cells in Human Mammary Epithelia." Cell Rep **23**(4): 1205-1219.

Pellacani, D., M. Bilenky, N. Kannan, A. Heravi-Moussavi, D. Knapp, S. Gakkhar, M. Moksa, A. Carles, R. Moore, A. J. Mungall, M. A. Marra, S. J. M. Jones, S. Aparicio, M. Hirst and C. J. Eaves (2016). "Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks." Cell Rep **17**(8): 2060-2074.

Pertea, M., G. M. Pertea, C. M. Antonescu, T. C. Chang, J. T. Mendell and S. L. Salzberg (2015). "StringTie enables improved reconstruction of a transcriptome from RNA-seq reads." Nat Biotechnol **33**(3): 290-295.

Prat, A. and C. M. Perou (2009). "Mammary development meets cancer genomics." Nat Med **15**(8): 842-844.

Quinlan, A. R. and I. M. Hall (2010). "BEDTools: a flexible suite of utilities for comparing genomic features." Bioinformatics **26**(6): 841-842.

Rakyan, V. K., T. A. Down, S. Maslau, T. Andrew, T. P. Yang, H. Beyan, P. Whittaker, O. T. McCann, S. Finer, A. M. Valdes, R. D. Leslie, P. Deloukas and T. D. Spector (2010). "Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains." Genome Res **20**(4): 434-439.

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

Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics **26**(1): 139-140.

Rodic, N., R. Sharma, R. Sharma, J. Zampella, L. Dai, M. S. Taylor, R. H. Hruban, C. A. Iacobuzio-Donahue, A. Maitra, M. S. Torbenson, M. Goggins, M. Shih Ie, A. S. Duffield, E. A. Montgomery, E. Gabrielson, G. J. Netto, T. L. Lotan, A. M. De Marzo, W. Westra, Z. A. Binder, B. A. Orr, G. L. Gallia, C. G. Eberhart, J. D. Boeke, C. R. Harris and K. H. Burns (2014). "Long interspersed element-1 protein expression is a hallmark of many human cancers." Am J Pathol **184**(5): 1280-1286.

Rooney, M. S., S. A. Shukla, C. J. Wu, G. Getz and N. Hacohen (2015). "Molecular and genetic properties of tumors associated with local immune cytolytic activity." Cell **160**(1-2): 48-61.

Sayaman, R. W., M. Miyano, P. Senapati, S. Shalabi, A. Zirbes, M. E. Todhunter, V. Seewaldt, S. L. Neuhausen, M. R. Stampfer, D. E. Schones and M. A. LaBarge (2021). "Epigenetic changes with age primes mammary luminal epithelia for cancer initiation." bioRxiv: 2021.2002.2012.430777.
Shalabi, S. F., M. Miyano, R. W. Sayaman, J. C. Lopez, T. A. Jokela, M. E. Todhunter, S. Hinz, J. C. Garbe, M. R. Stampfer and K. Kessenbrock (2021). "Evidence for accelerated aging in mammary epithelia of women carrying germline BRCA1 or BRCA2 mutations." Nature Aging **1**(9): 838-849.
Siegel, R. L., K. D. Miller, H. E. Fuchs and A. Jemal (2021). "Cancer Statistics, 2021." CA Cancer J Clin **71**(1): 7-33.
Smit, A., Hubley, R & Green, P. (2013-2015). "RepeatMasker Open-4.0. http://www.repeatmasker.org.".
Song, Q., B. Decato, E. E. Hong, M. Zhou, F. Fang, J. Qu, T. Garvin, M. Kessler, J. Zhou and A. D. Smith (2013). "A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics." PLoS One **8**(12): e81148.
Stadler, M. B., R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Scholer, E. van Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari and D. Schubeler (2011). "DNA-binding factors shape the mouse methylome at distal regulatory regions." Nature **480**(7378): 490-495.
Sun, D., M. Luo, M. Jeong, B. Rodriguez, Z. Xia, R. Hannah, H. Wang, T. Le, K. F. Faull, R. Chen, H. Gu, C. Bock, A. Meissner, B. Gottgens, G. J. Darlington, W. Li and M. A. Goodell (2014). "Epigenomic profiling of young and aged HSCs reveals concerted changes during aging that reinforce self-renewal." Cell Stem Cell **14**(5): 673-688.
Teschendorff, A. E., U. Menon, A. Gentry-Maharaj, S. J. Ramus, D. J. Weisenberger, H. Shen, M. Campan, H. Noushmehr, C. G. Bell, A. P. Maxwell, D. A. Savage, E. Mueller-Holzner, C. Marth, G. Kocjan, S. A. Gayther, A. Jones, S. Beck, W. Wagner, P. W. Laird, I. J. Jacobs and M. Widschwendter (2010). "Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer." Genome Res **20**(4): 440-446.
Vandiver, A. R., R. A. Irizarry, K. D. Hansen, L. A. Garza, A. Runarsson, X. Li, A. L. Chien, T. S. Wang, S. G. Leung, S. Kang and A. P. Feinberg (2015). "Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin." Genome Biol **16**: 80.
Wickham, H. (2016). ggplot2: elegant graphics for data analysis, Springer.
Yang, W. R., D. Ardeljan, C. N. Pacyna, L. M. Payer and K. H. Burns (2019). "SQuIRE reveals locus-specific regulation of interspersed repeat expression." Nucleic Acids Res **47**(5): e27.
Yin, Y., E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P. K. Das, T. Kivioja, K. Dave, F. Zhong, K. R. Nitta, M. Taipale, A. Popov, P. A. Ginno, S. Domcke, J. Yan, D. Schubeler, C. Vinson and J. Taipale (2017). "Impact of cytosine methylation on DNA binding specificities of human transcription factors." Science **356**(6337).

Figure Legends

Figure 1: Mammary luminal epithelial cells from older individuals show DNA methylation changes at distinct genomic loci. (A) Methylation levels of identified DMRs across all samples. Samples are in rows, and DMRs are in columns. The color intensity represents the Z-score scaled from the mean. Values ranging from higher to lower methylation are shown as purple to green. (B) Kernel smoothed line plots of representative DMRs±500bp with methylation gain (left panel) and methylation loss (right panel). Blue

and red lines respectively represent young and old samples. DMRs are highlighted in pink and black ticks mark each CpG position. (C) MSigDB gene sets enriched in genes with age dependent DNA methylation loss at promoters. (D) Genome browser tracks showing DMRs with methylation loss in old LEps at the *FOXP1* gene locus. (E) MSigDB gene sets enriched in genes with age dependent DNA methylation gain at promoters. Promoters were defined as TSS+/-5kb. Gene ratio denotes the proportion of genes represented of the total genes in gene set. (F) Genome browser tracks showing methylation gain at the CpG island promoter of *ALDH1A3* gene.

Figure 2: Differentially methylated regions show enrichment of lineage-specific transcription factor binding sites. (A) Top ten TF binding sites enriched in the DMRs that lose methylation with age are shown. Dots in the left panel show -log10(p-value) for the enrichment of the transcription factor binding site. Dots in the right panel show expression levels (RPKM) of the respective transcription factor in young myoepithelial (MEp) and luminal (LEp) cells. (B) Boxplot showing expression levels of myoepithelial-specific TF TP63 in young and old LEps (n=8, *p*-value < 0.01; two-tailed unpaired t test with Welch's correction). (C) Top ten TF binding sites enriched in the DMRs that gain methylation with age are shown. Dots in the left panel show -log10(p-value) for the enrichment of the TF binding site. Dots in the right panel show expression levels (RPKM) of the respective transcription factor in young myoepithelial (MEp) and luminal (LEp) cells. (D) Boxplot showing expression levels of luminal lineage-specific TF ELF5 in young and old LEps (n=8, *p*-value < 0.05; two-tailed Mann Whitney test). (E) Volcano plot showing expression levels of all TFs expressed in LEPs and MEPs (RPKM > 0.1). TFs with motifs enriched in DMRs with methylation loss and gain are highlighted in blue and red respectively. *P*-value < 0.05. (F) Schematic showing the role of myoepithelial-specific transcription factors in causing loss of methylation. Conversely, loss of expression of luminal-specific transcription factors leads to methylation of unbound sites and methylation gain.

Figure 3: Differentially methylated regions are enriched at lineage specific regulatory elements. (A) Genome browser tracks showing a representative example of a DMR with methylation loss

accompanied with increase in chromatin accessibility in old LEPs. DMRs are highlighted in green. Bottom tracks show location of MEp-specific TF binding sites. (B) Volcano plot showing log2 fold change (log2fc) of ATAC-seq peaks overlapping DMRs with methylation loss in old versus young LEPs. Positive log2 fold change indicates increase in accessibility. (C) Genome browser tracks showing a representative example of a DMR with methylation gain accompanied with decrease in chromatin accessibility in old LEPs. DMRs are highlighted in green. Bottom tracks show location of LEp-specific TF binding sites. (D) Volcano plot showing log2 fold change (log2fc) of ATAC-seq peaks overlapping DMRs with methylation gain in old versus young LEPs. Negative log2 fold change indicates decrease in accessibility. (E) Percentage number of DMRs with methylation loss and (F) methylation gain that overlap with chromHMM features of ER^-, ER^+ luminal cells, and basal cells (rows) is shown. The color key indicates the various chromatin states identified from ChIP-seq datasets.

Figure 4: (A) Percentage number of promoters with DNA methylation gain that overlap H3K27me3 peaks (filled bar) in ER^-, ER^+ luminal cells, and basal cells is shown. (B) Percentage number of promoters with DNA methylation gain that overlap H3K27me3 peaks (filled bar) in indicated mammary cell types. (C) Aggregate plot of H3K27me3 signal around the TSS of genes with age-dependent DNA methylation gain and marked with H3K27me3 in all three cell types (union set). (D) Aggregate plot of H3K27me3 signal around the TSS of genes with age-dependent DNA methylation loss and marked with H3K27me3 in all three cell types (union set).

Figure 5: (A) Schematic depicting changes in entropy associated with stochastic methylation gain at CpG islands. (B) Schematic depicting changes in entropy associated with stochastic methylation loss at fully methylated regions such as repeats/Alu elements. (C) CpG island methylation and entropy levels at unmethylated CpG islands that show increase in entropy. (D) Alu methylation and entropy levels at fully methylated Alu elements that show increase in entropy. (E) Aggregate profiles of CpG island methylation and entropy levels at unmethylated CpG islands that show increase in entropy with age. Luminal breast cancer samples (red) show an increase in DNA methylation as well as entropy at these CpG islands.

(F) Aggregate profiles of CpG island methylation and entropy levels at unmethylated CpG islands that show increase in entropy with age. Basal breast cancer samples (pink) do not show an increase in DNA methylation or entropy at these CpG islands. (G) Expression of *FOXC1* in five subtypes of breast tumors from the TCGA BRCA collection. Luminal A (LumA) versus Basal p-value < 2.2x10-16; using two-tailed Wilcoxon signed-rank test. Luminal B (LumB) versus Basal (p-value < 2.2x10-16; using two-tailed Wilcoxon signed-rank test).

Figure 6: Variably methylated regions with methylation loss are enriched at primate-specific transposable elements (A-D) Evolutionary age of (A) all TE elements (B) SINE, (C) LINE, and (D) LTR/ERV elements that are variably methylated with methylation loss or gain. Evolutionary ages of only the significantly enriched TE subfamilies are plotted. (E) TE subfamilies enriched in regions with variable methylation loss. Left-right; panels show evolutionary age, CpG density, number of elements from each subfamily that lose methylation and the observed/expected ratio for each TE subfamily. Expected number of overlaps in shuffled controls was used to calculate the ratio. In the evolutionary age and CpG density panels, brown points indicate the median. (F) Genome browser tracks showing a representative example of a HERVH-int element showing variable methylation loss in old samples. Bottom tracks are from ATAC-seq data showing gain of accessibility at the variably methylated region in the corresponding sample.

Figure 7: Transposable elements with age-dependent methylation loss are activated in breast cancer. (A) A representative HERVH element showing variable loss of methylation in old samples and in breast cancer (BC) samples. BC WGBS tracks are shown in the dense mode to save space. Methylation loss is evident from grey bars instead of black (fully methylated). PAM50 subtypes and estrogen receptor (ER) expression status of the tumors are shown for the BC WGBS tracks. Bottom tracks are RNA-seq tracks from DCIS samples showing cryptic transcription of the HERVH element. All RNA-seq tracks are scaled to the same value and PAM50 subtype is indicated on the right. (B) Heatmap showing signals across 74 TCGA-BRCA samples for ATAC-seq peaks that overlap TEs hypomethylated with age. (C) Expression of *ZNF92* in five subtypes of breast tumors from the TCGA BRCA collection (p-

value Luminal A (LumA) versus Basal and Luminal B (LumB) versus Basal (p-value < 2.2x10-16; using two-tailed Wilcoxon signed-rank test).

Figures

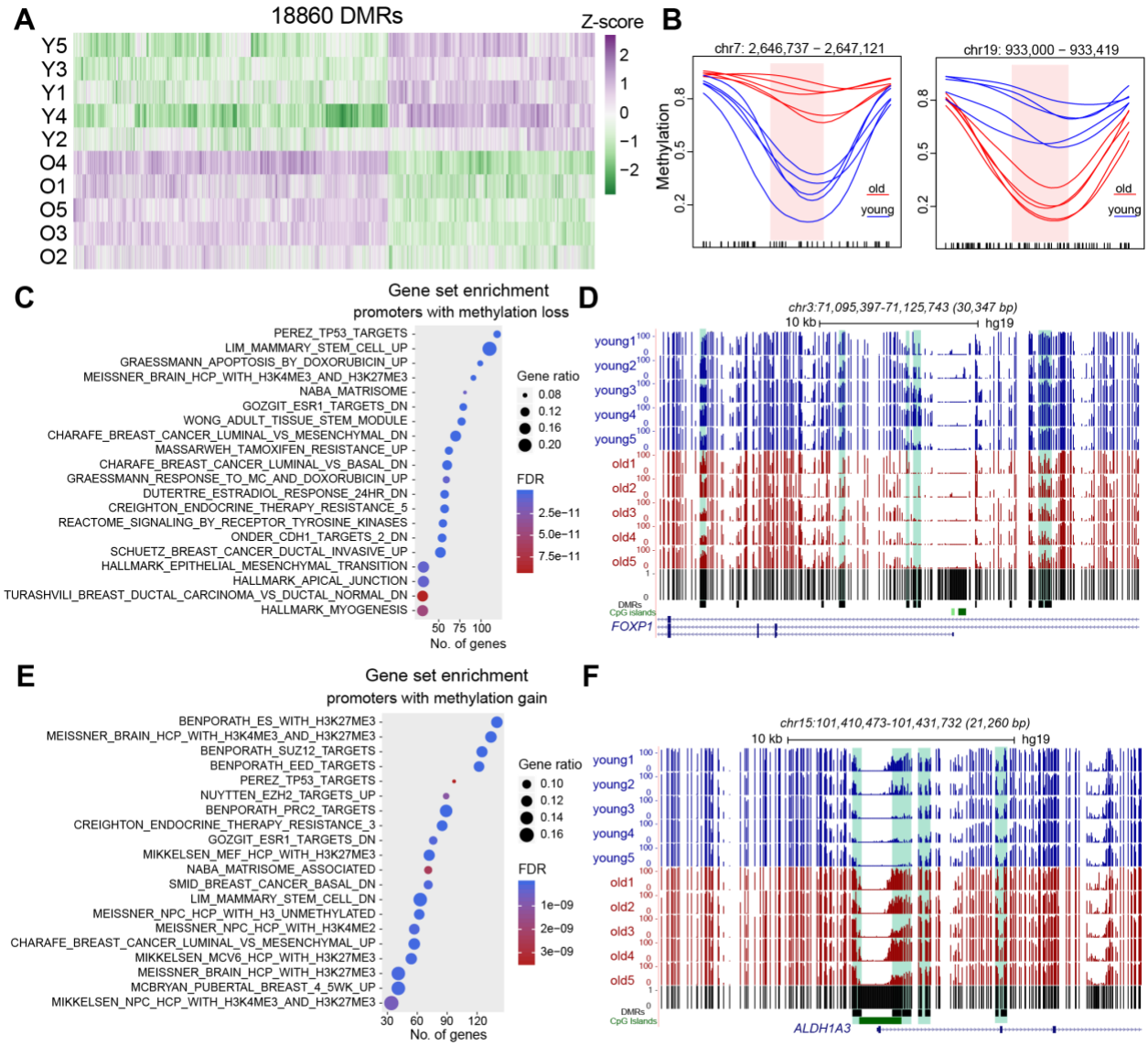

Figure 1

Figure 1: Mammary luminal epithelial cells from older individuals show DNA methylation changes at distinct genomic loci. (A) Methylation levels of identified DMRs across all samples. Samples are in rows, and DMRs are in columns. The color intensity represents the Z-score scaled from the mean. Values ranging from higher to lower methylation are shown as purple to green. (B) Kernel smoothed line plots of representative DMRs±500bp with methylation gain (left panel) and methylation loss (right panel). Blue and red lines respectively represent young and old samples. DMRs are highlighted in pink and black ticks mark each CpG position. (C) MSigDB gene sets enriched in genes with age dependent DNA methylation loss at promoters. (D) Genome browser tracks showing DMRs with methylation loss in old LEps at the *FOXP1* gene locus. (E) MSigDB gene sets enriched in genes with age dependent DNA methylation gain at promoters. Promoters were defined as TSS+/-5kb. Gene ratio denotes the proportion of genes represented of the total genes in gene set. (F) Genome browser tracks showing methylation gain at the CpG island promoter of *ALDH1A3* gene.

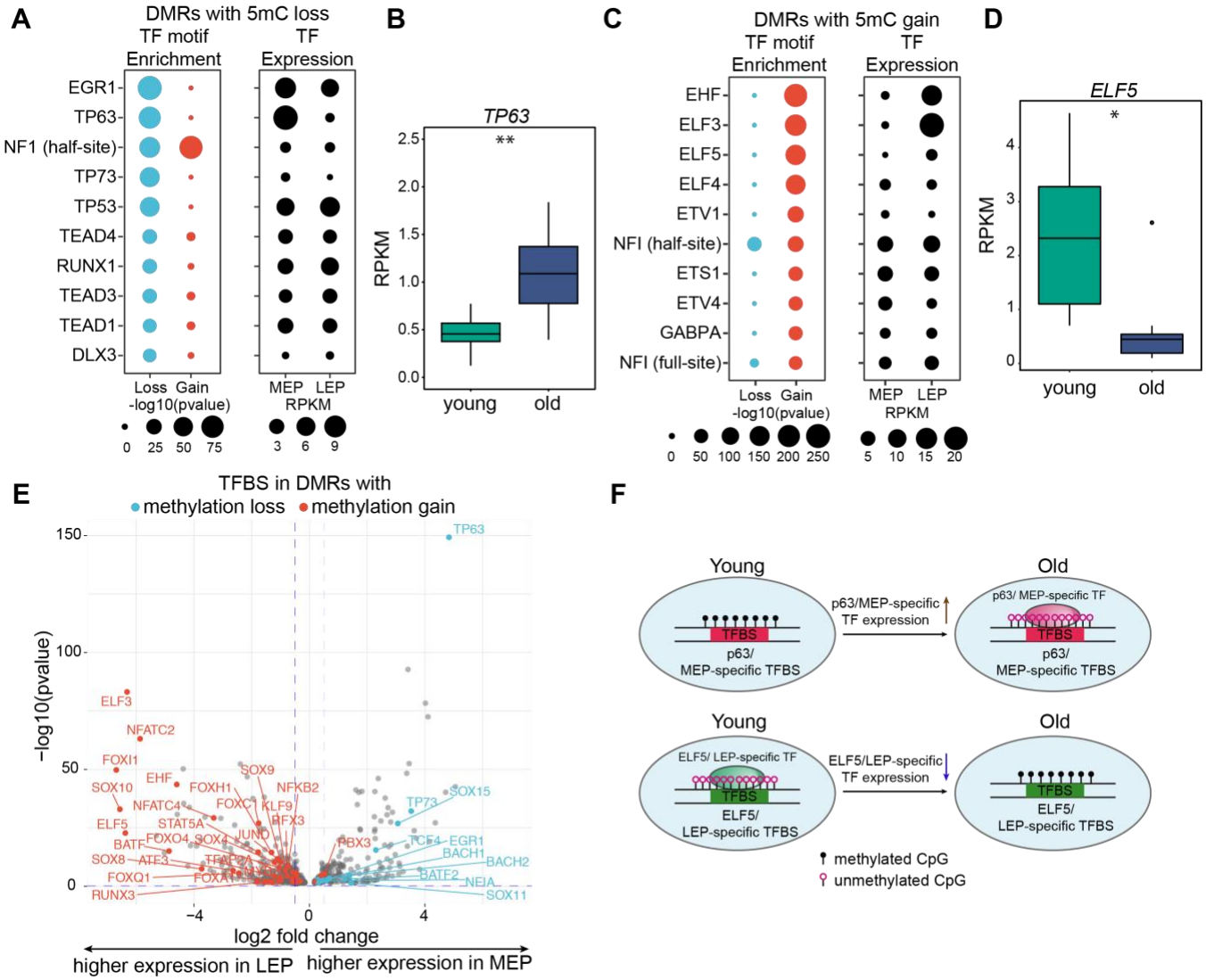

Figure 2

Figure 2: Differentially methylated regions show enrichment of lineage-specific transcription factor binding sites. (A) Top ten TF binding sites enriched in the DMRs that lose methylation with age are shown. Dots in the left panel show -log10(p-value) for the enrichment of the transcription factor binding site. Dots in the right panel show expression levels (RPKM) of the respective transcription factor in young myoepithelial (MEp) and luminal (LEp) cells. (B) Boxplot showing expression levels of myoepithelial-specific TF TP63 in young and old LEps (n=8, *p*-value < 0.01; two-tailed unpaired t test with Welch's correction). (C) Top ten TF binding sites enriched in the DMRs that gain methylation with age are shown. Dots in the left panel show -log10(p-value) for the enrichment of the TF binding site. Dots in the right panel show expression levels (RPKM) of the respective transcription factor in young myoepithelial (MEp) and luminal (LEp) cells. (D) Boxplot showing expression levels of luminal lineage-specific TF ELF5 in young and old LEps (n=8, *p*-value < 0.05; two-tailed Mann Whitney test). (E) Volcano plot showing expression levels of all TFs expressed in LEPs and MEPs (RPKM > 0.1). TFs with motifs enriched in DMRs with methylation loss and gain are highlighted in blue and red respectively. *P*-value < 0.05. (F) Schematic showing the role of myoepithelial-specific transcription factors in causing loss of methylation. Conversely, loss of expression of luminal-specific transcription factors leads to methylation of unbound sites and methylation gain.

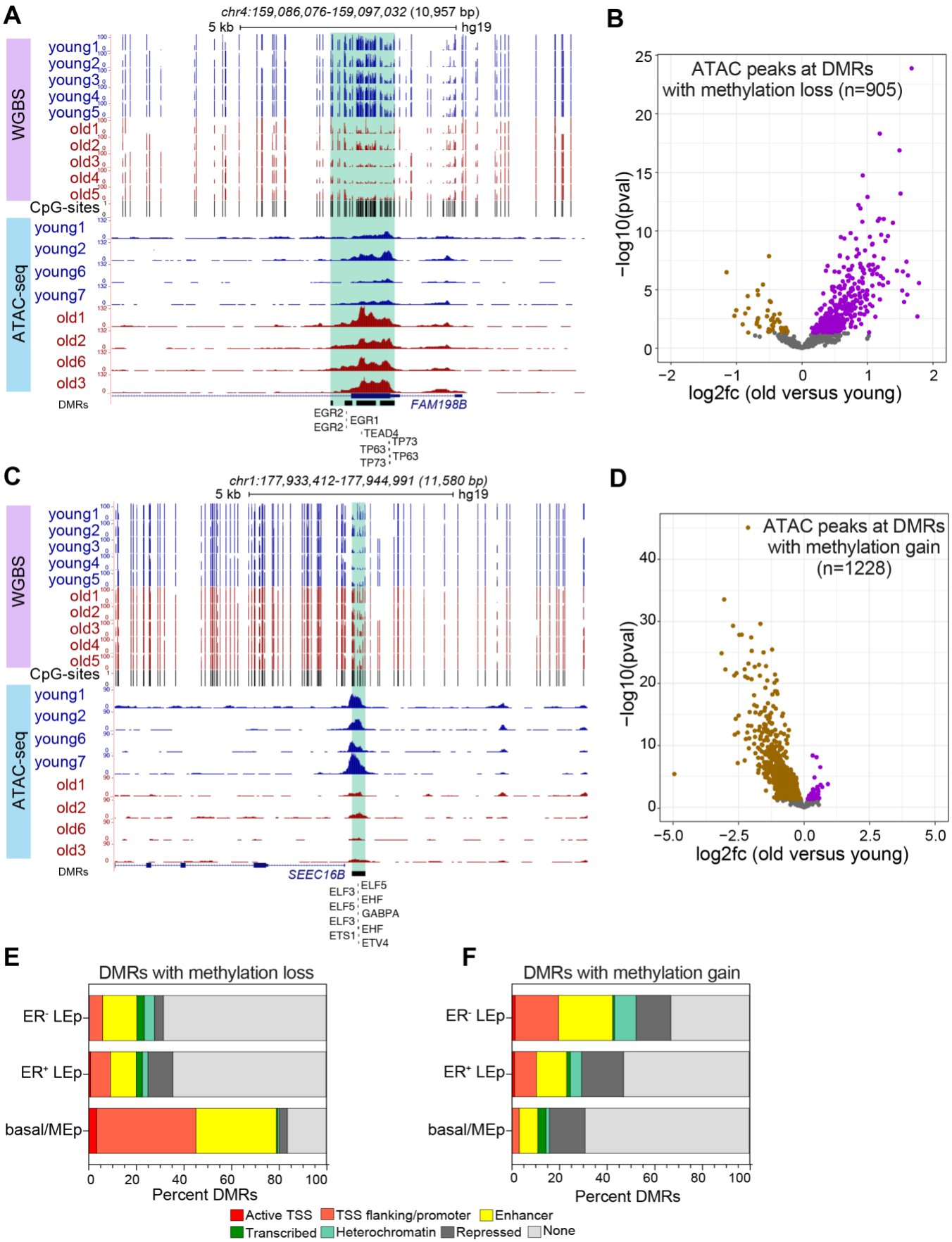
A
chr4:159,086,076-159,097,032 (10,957 bp)
5 kb
hg19
WGBS
young1
young2
young3
young4
young5
old1
old2
old3
old4
old5
CpG-sites
ATAC-seq
young1
young2
young6
young7
old1
old2
old6
old3
DMRs
FAM198B
EGR2
EGR2
EGR1
TEAD4
TP63
TP73
TP73
TP63
B
ATAC peaks at DMRs
with methylation loss (n=905)
−log10(pval)
log2fc (old versus young)
C
chr1:177,933,412-177,944,991 (11,580 bp)
5 kb
hg19
WGBS
young1
young2
young3
young4
young5
old1
old2
old3
old4
old5
CpG-sites
ATAC-seq
young1
young2
young6
young7
old1
old2
old6
old3
DMRs
SEEC16B
ELF3
ELF5
ELF3
ETS1
ELF5
EHF
GABPA
EHF
ETV4
D
ATAC peaks at DMRs
with methylation gain
(n=1228)
−log10(pval)
log2fc (old versus young)
E
DMRs with methylation loss
ER⁻ LEp
ER⁺ LEp
basal/MEp
Percent DMRs
F
DMRs with methylation gain
ER⁻ LEp
ER⁺ LEp
basal/MEp
Percent DMRs
Active TSS
TSS flanking/promoter
Enhancer
Transcribed
Heterochromatin
Repressed
None

Figure 3

Figure 3: Differentially methylated regions are enriched at lineage specific regulatory elements. (A) Genome browser tracks showing a representative example of a DMR with methylation loss accompanied with increase in chromatin accessibility in old LEPs. DMRs are highlighted in green. Bottom tracks show location of MEp-specific TF binding sites. (B) Volcano plot showing log2 fold change (log2fc) of ATAC-seq peaks overlapping DMRs with methylation loss in old versus young LEPs. Positive log2 fold change indicates increase in accessibility. (C) Genome browser tracks showing a representative example of a DMR with methylation gain accompanied with decrease in chromatin accessibility in old LEPs. DMRs are highlighted in green. Bottom tracks show location of LEp-specific TF binding sites. (D) Volcano plot showing log2 fold change (log2fc) of ATAC-seq peaks overlapping DMRs with methylation gain in old versus young LEPs. Negative log2 fold change indicates decrease in accessibility. (E) Percentage number of DMRs with methylation loss and (F) methylation gain that overlap with chromHMM features of ER^-, ER^+ luminal cells, and basal cells (rows) is shown. The color key indicates the various chromatin states identified from ChIP-seq datasets.

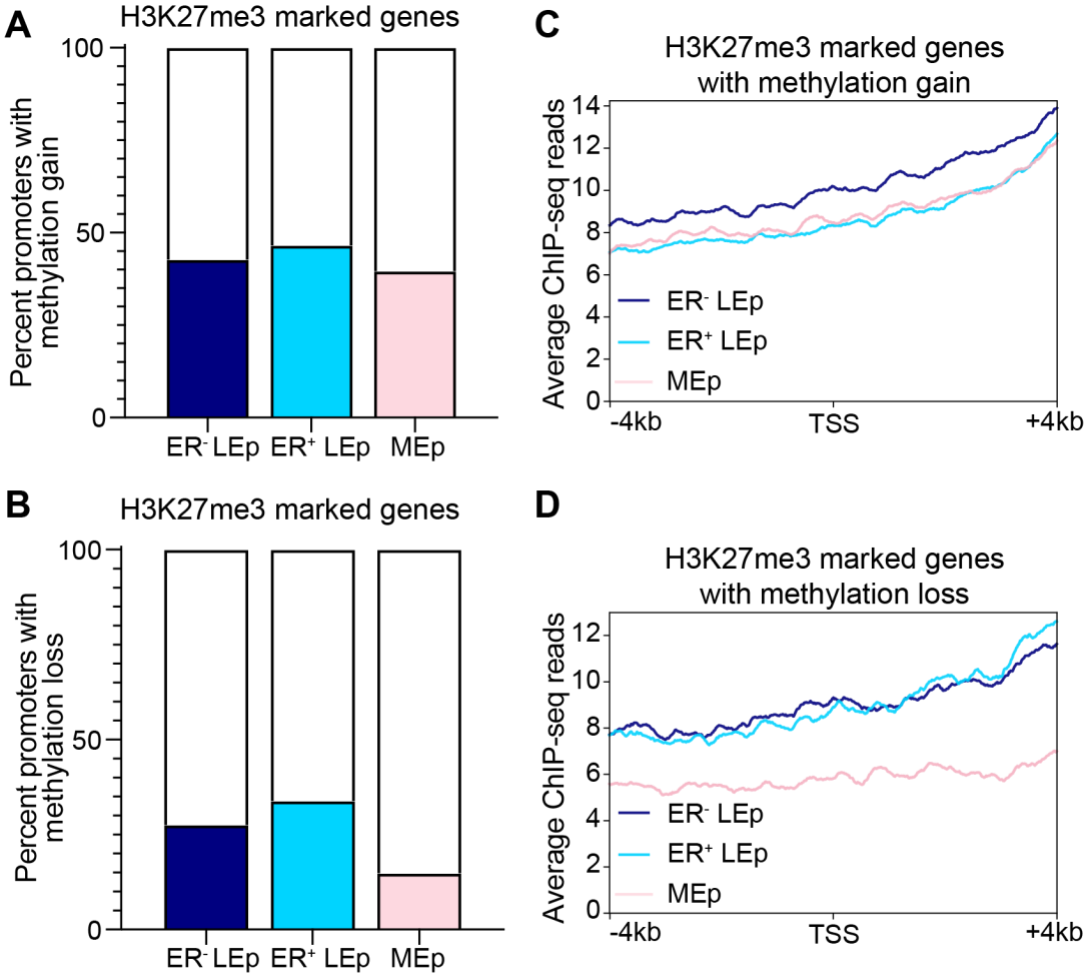

Figure 4

Figure 4: (A) Percentage number of promoters with DNA methylation gain that overlap H3K27me3 peaks (filled bar) in ER^-, ER^+ luminal cells, and basal cells is shown. (B) Percentage number of promoters with DNA methylation gain that overlap H3K27me3 peaks (filled bar) in indicated mammary cell types. (C) Aggregate plot of H3K27me3 signal around the TSS of genes with age-dependent DNA methylation gain and marked with H3K27me3 in all three cell types (union set). (D) Aggregate plot of H3K27me3 signal

around the TSS of genes with age-dependent DNA methylation loss and marked with H3K27me3 in all three cell types (union set).

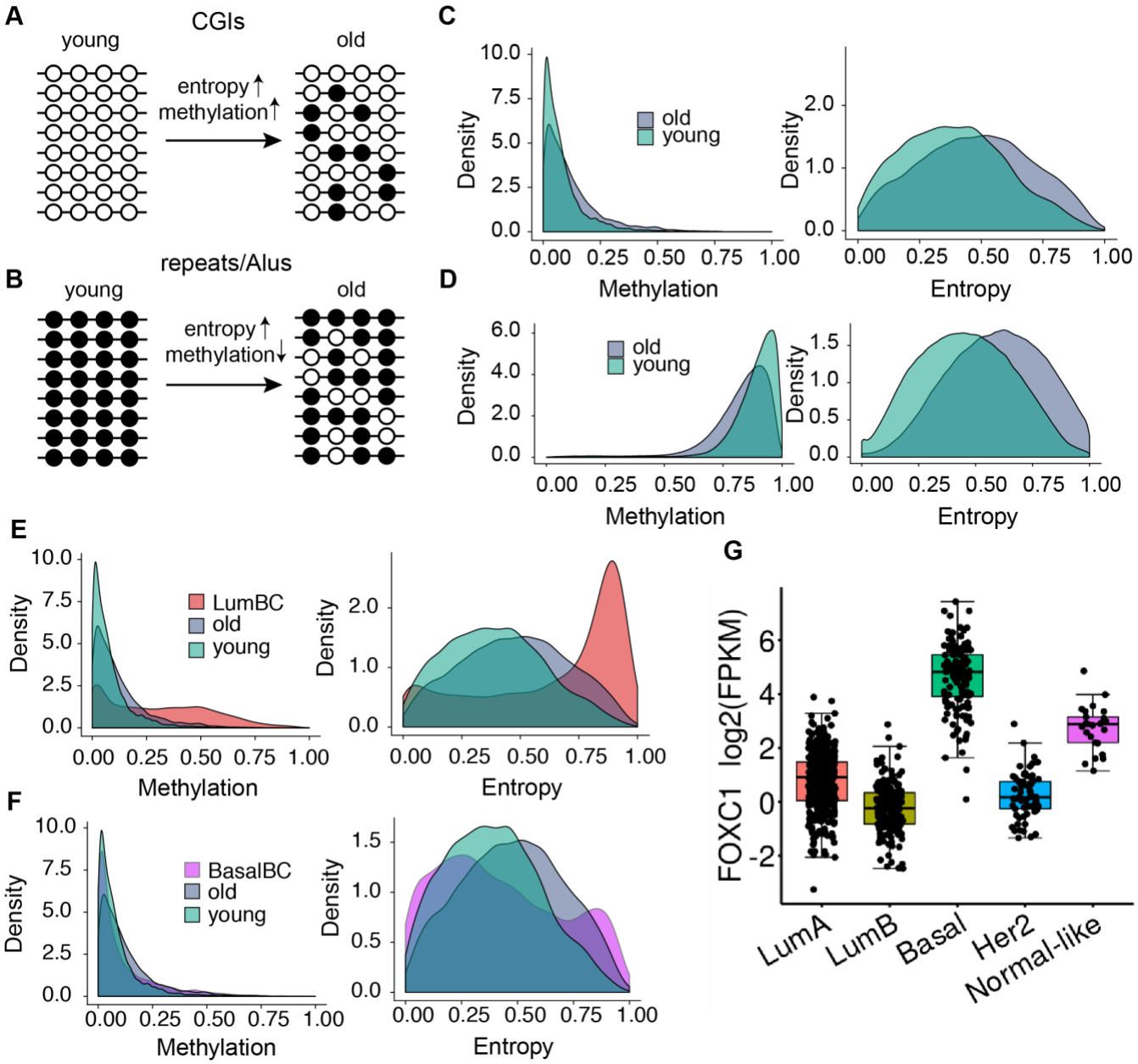

Figure 5: (A) Schematic depicting changes in entropy associated with stochastic methylation gain at CpG islands. (B) Schematic depicting changes in entropy associated with stochastic methylation loss at fully methylated regions such as repeats/Alu elements. (C) CpG island methylation and entropy levels at unmethylated CpG islands that show increase in entropy. (D) Alu methylation and entropy levels at fully methylated Alu elements that show increase in entropy. (E) Aggregate profiles of CpG island methylation and entropy levels at unmethylated CpG islands that show increase in entropy with age. Luminal breast cancer samples (red) show an increase in DNA methylation as well as entropy at these CpG islands. (F) Aggregate profiles of CpG island methylation and entropy levels at unmethylated CpG islands that show increase in entropy with age. Basal breast cancer samples (pink) do not show an increase in DNA methylation or entropy at these CpG islands. (G) Expression of *FOXC1* in five subtypes of breast tumors from the TCGA BRCA collection. Luminal A (LumA) versus Basal p-value < 2.2x10-16; using two-tailed

Wilcoxon signed-rank test. Luminal B (LumB) versus Basal (p-value < 2.2x10-16; using two-tailed Wilcoxon signed-rank test).

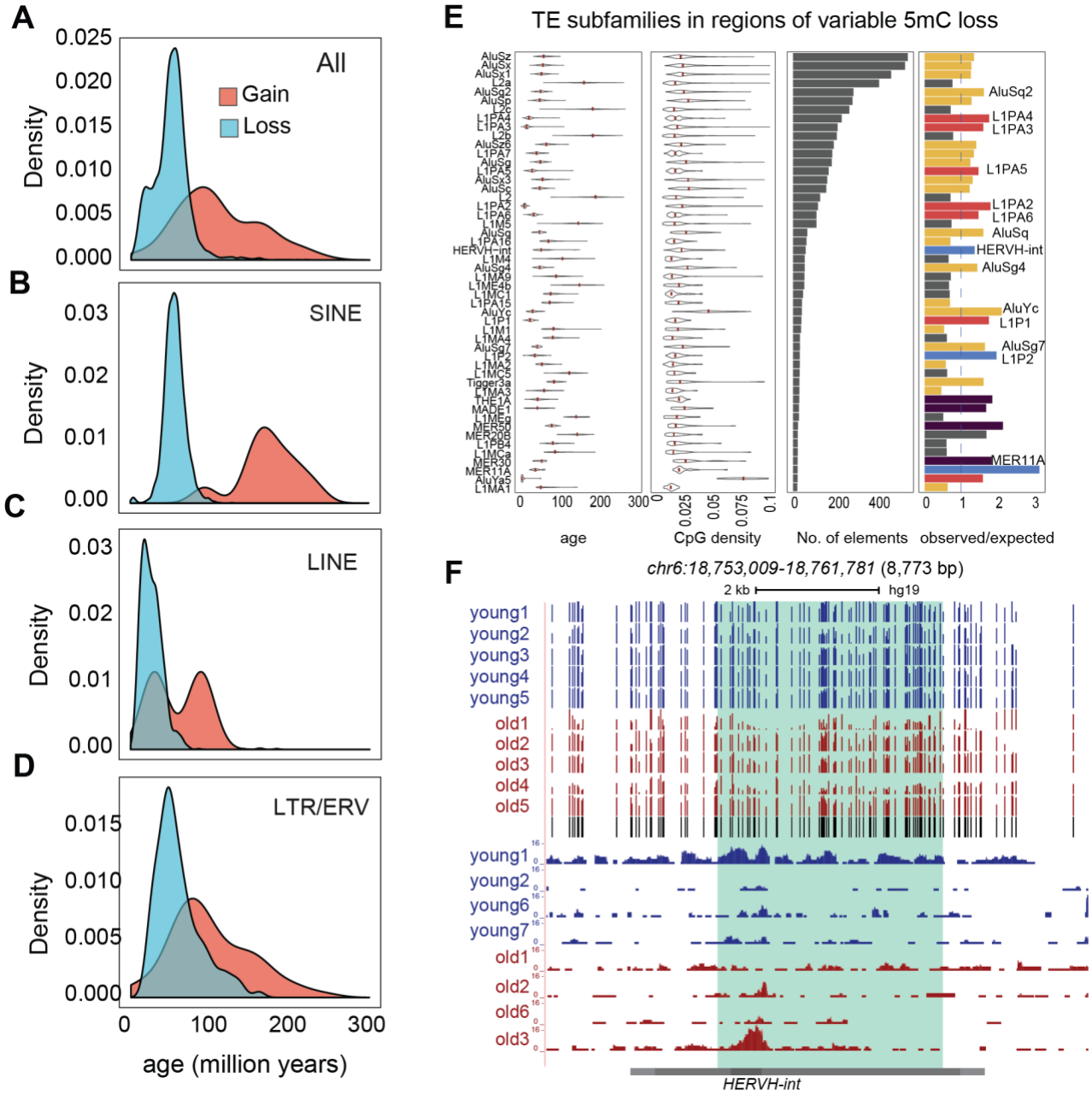

Figure 6: Variably methylated regions with methylation loss are enriched at primate-specific transposable elements (A-D) Evolutionary age of (A) all TE elements (B) SINE, (C) LINE, and (D) LTR/ERV elements that are variably methylated with methylation loss or gain. Evolutionary ages of only the significantly enriched TE subfamilies are plotted. (E) TE subfamilies enriched in regions with variable methylation loss. Left-right; panels show evolutionary age, CpG density, number of elements from each subfamily that lose methylation and the observed/expected ratio for each TE subfamily. Expected number of overlaps in shuffled controls was used to calculate the ratio. In the evolutionary age and CpG density panels, brown points indicate the median. Colors in the observed/expected graph highlight primate-specific or younger elements. (F) Genome browser tracks showing a representative example of a

HERVH-int element showing variable methylation loss in old samples. Bottom tracks are from ATAC-seq data showing gain of accessibility at the variably methylated region in the corresponding sample.

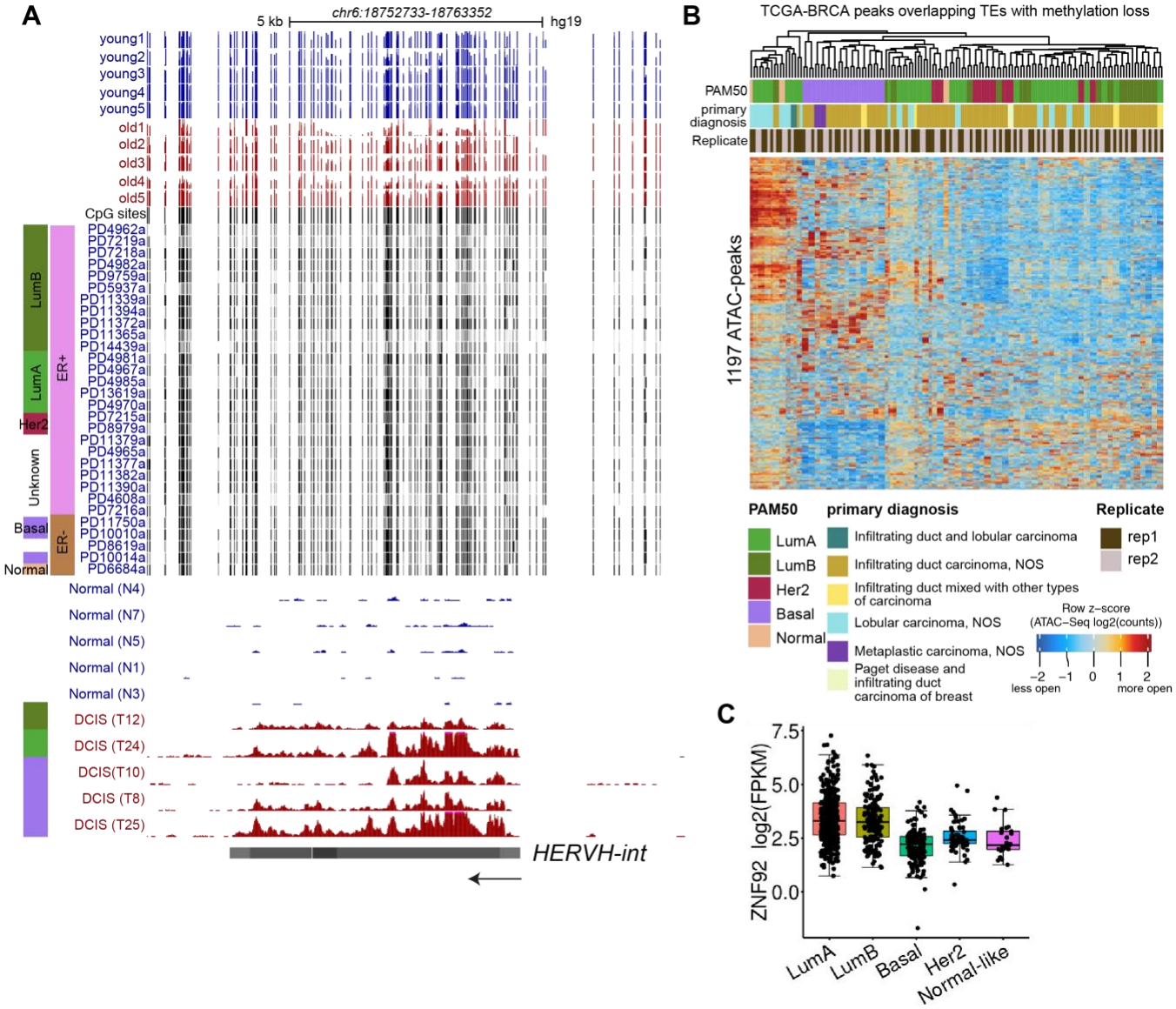

Figure 7: Transposable elements with age-dependent methylation loss are activated in breast cancer. (A) A representative HERVH element showing variable loss of methylation in old samples and in breast cancer (BC) samples. BC WGBS tracks are shown in the dense mode to save space. Methylation loss is evident from grey bars instead of black (fully methylated). PAM50 subtypes and estrogen receptor (ER) expression status of the tumors are shown for the BC WGBS tracks. Bottom tracks are RNA-seq tracks from DCIS samples showing cryptic transcription of the HERVH element. All RNA-seq tracks are scaled to the same value and PAM50 subtype is indicated on the right. (B) Heatmap showing signals across 74 TCGA-BRCA samples for ATAC-seq peaks that overlap TEs hypomethylated with age. (C) Expression of *ZNF92* in five subtypes of breast tumors from the TCGA BRCA collection (p-value Luminal A (LumA) versus Basal and Luminal B (LumB) versus Basal (p-value < 2.2x10-16; using two-tailed Wilcoxon signed-rank test).