Aging leads to stochastic loss of silencing at transposons in mammary luminal epithelial cells

Parijat Senapati[1], Masaru Miyano[2], Mudaser Basam[1,2], Rosalyn W. Sayaman[2], 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

*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

Running title: Stochastic loss of transposon silencing during aging.

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

Abstract

The aging process is characterized by an accumulation of cellular damage, including alterations to the epigenome. While it is clear that DNA methylation patterns change with age, the mechanisms behind these epigenomic alterations – and their functional consequences – have remained unclear. One of the primary roles of DNA methylation in mammalian genomes is in the repression of retrotransposons. Many of these repetitive elements contain regulatory sequences, and loss of repression can lead to aberrant transcriptional regulation. Aging-associated DNA methylation loss at transposable elements (TEs), therefore, has the potential to rewire gene expression networks to be favorable for pathologies, including age-related diseases such as cancer. However, the extent of loss of epigenetic silencing at transposable elements during aging and the functional consequences of this have remained unclear. Luminal epithelial cells in the breast tissue are a key cell lineage implicated in age-related luminal breast cancers. We report here that aging leads to distinct DNA methylation patterns at tissue-specific regulatory elements and TEs in luminal epithelial cells. DNA methylation changes at regulatory elements are driven by altered activity of lineage-specific transcription factors such as ELF5 and TP63. However, transposable elements, especially recently evolved TEs, exhibit stochastic loss of methylation and increased methylation entropy with age. TEs that lose methylation with age regulate known breast cancer genes and are associated with elevated breast cancer risk. Altogether, our results indicate that aging leads to DNA methylation loss at evolutionarily young TEs that can impact gene networks important for age-related breast cancers.

Background

Aging is associated with a progressive decline of tissue and organ functions and is a significant risk factor for many diseases, including cancer. At the genomic level, alterations to the epigenome may contribute to the development of age-associated diseases (Lopez-Otin et al. 2013; Booth and Brunet 2016). DNA methylation is a relatively stable component of the epigenome that gets established in early development and is maintained through successive cell division cycles to preserve cell identity (Bird 2002). However, DNA methylation patterns are altered with age (Maegawa et al. 2010; Rakyan et al. 2010; Heyn et al. 2012; Horvath 2013). DNA methylome comparison of human $CD4^+$ T cells from newborns and centenarians suggested an overall decrease in methylation levels with age (Heyn et al. 2012). Moreover, DNA methylation levels of specific CpG sites show a significant correlation with age across diverse tissues (Horvath 2013). These sets of CpG sites known as 'epigenetic clocks' can accurately predict the chronological age of cells based on DNA methylation (Horvath 2013). While it is clear that distinct DNA methylation changes to specific CpG sites occur with aging, the causes of these alterations and their impact on disease are still unclear.

Over 75% of breast cancers are diagnosed in women over the age of 50 (Jemal et al. 2007). Aging in the breast tissue is accompanied by increases in estrogen receptor expression (Quong et al. 2002), and significant physiological changes such as an increase in fat content and a decrease in connective tissue (Benz 2008) that alter the microenvironment of the mammary epithelium. Previous studies suggest a complex interplay between the various epithelial cell types and microenvironment that leads to changes in molecular properties and the proportion of cell populations (Garbe et al. 2012; Pelissier et al. 2014). Luminal epithelial cells (LEPs) and myoepithelial cells (MEPs) are the two principal lineages in breast epithelia, and the most striking age and microenvironment-related changes in abundance and phenotype occur

among LEPs (Garbe et al. 2012; Dos Santos et al. 2015; Pelissier Vatter et al. 2018). Furthermore, aging-associated tissue microenvironment changes contribute to the observed age-related epigenetic and transcriptional changes in the mammary epithelium (Miyano et al. 2017).

In vitro transformed human mammary epithelial cells (HMECs) from older women (> 55 y) exhibit gene expression patterns characteristic of luminal breast cancer subtypes. In contrast, transformed HMECs from younger women (< 30 y) show more properties of basal breast cancers (Lee et al. 2015). This is consistent with the fact that 80% of age-related breast cancers are of the hormone-receptor expressing luminal subtypes (Jenkins et al. 2014). This suggests that cells from older women may harbor age-related epigenetic states that make them predisposed to develop luminal breast cancers. Intriguingly, the putative breast cancer cells of origin in older women express signaling molecules consistent with those of immortalized luminal cells, suggesting that age-associated changes can elevate cancer risk (Pelissier Vatter et al. 2018). Epigenetic clock studies in breast tissue have suggested that women with luminal breast cancers exhibit increased epigenetic age even in the adjacent normal breast tissue (Hofstatter et al. 2018). Although earlier studies suggest that aging can lead to epigenetic changes in breast tissue, thereby elevating breast cancer risk, a complete understanding of age-associated DNA methylation changes in specific human mammary cell types is lacking. Most previous studies were performed either in bulk mammary tissue or interrogated a small subset of CpG sites in the genome (Hofstatter et al. 2018; Castle et al. 2020).

Repetitive elements derived from transposable elements constitute roughly about ~50% of the human genome and are classified based on their method of transposition (Cordaux and Batzer 2009). Although most elements have accumulated enough mutations and lost their ability to propagate, they can still carry latent regulatory potential (Sundaram et al. 2014). Some

elements, such as Alu, LINE, and SVA elements, have maintained retrotransposition competence (Bourque et al. 2018). Cancer cells with global losses of DNA methylation frequently exhibit activation of transposable elements (Ishak and De Carvalho ; Burns 2017). Interestingly, increased expression of retrotransposons and active retrotransposition have been observed in aging mouse tissues (De Cecco et al. 2013; Benayoun et al. 2019). Moreover, LINE-1 expression is increased during cell senescence, a cell state synonymous with aging (De Cecco et al. 2019). We therefore reasoned that aging would lead to the loss of epigenetic silencing of transposons in human mammary luminal epithelial cells.

To evaluate the impact of aging on transposable elements we profiled global DNA methylation levels using WGBS and RRBS in non-senescent human mammary luminal epithelial cells from young and old individuals. We report here that aging in human mammary luminal epithelial cells leads to DNA methylation changes at regulatory elements and loss of methylation at CpG-dense transposable elements. Methylation changes at regulatory elements are driven by altered expression and activity of lineage dependent transcription factors such as ELF5 and TP63. Recently evolved transposable elements mainly lose methylation in a stochastic manner and exhibit increased methylation disorder, presumably due to loss of fidelity of maintenance methylation. Aging dependent methylation loss at TEs leads to increased expression of luminal breast cancer genes in old luminal cells as well as early breast cancer tissues. Overall, our results suggest that CpG-dense TEs lose silencing with age and can function as regulatory elements supporting the expression of key genes during breast cancer development.

Results

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

To investigate the genome-wide changes in DNA methylation associated with healthy aging, we performed whole-genome bisulfite sequencing (WGBS-seq) on human mammary LEPs from two young and two old individuals (**Figure 1A**). We isolated LEPs from primary cultures of pre-stasis HMECs from discarded reduction mammoplasty (RM) tissue (Garbe et al. 2012). Tissues were obtained from women with differing chronological ages but without breast cancer. These pre-stasis LEPs from primary cultures maintain gene expression and DNA methylation profiles that mirror those in tissue (Miyano et al. 2017). 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 *in vitro* cellular senescence.

We obtained high coverage (~9.4×) WGBS data across individual CpG sites and a representation of about 90% of individual CpG sites in our samples (see Methods and Tables S1, S2). Covered CpG sites had a slightly higher representation at CpG islands and lower representation at repeat elements, as expected, given the difficulty of mapping to repetitive regions of the genome (**Figure S1A**). Across repeat classes, SINE element CpGs had a higher representation, whereas LINE CpGs were under-represented (**Figure S1B**), consistent with the known mappability problems at young LINE elements (De Cecco et al. 2013). Comparing the methylomes of young and old individuals, we did not observe any change in the global methylation levels across the genome (**Figure S1C, D**). However, principal component analysis (PCA) of the WGBS data separated the old and the young samples on the first and second principal component axes (35.95% and 33.1% of the respective cumulative variance; **Figure S1E**) indicating that there are local differences in the methylomes.

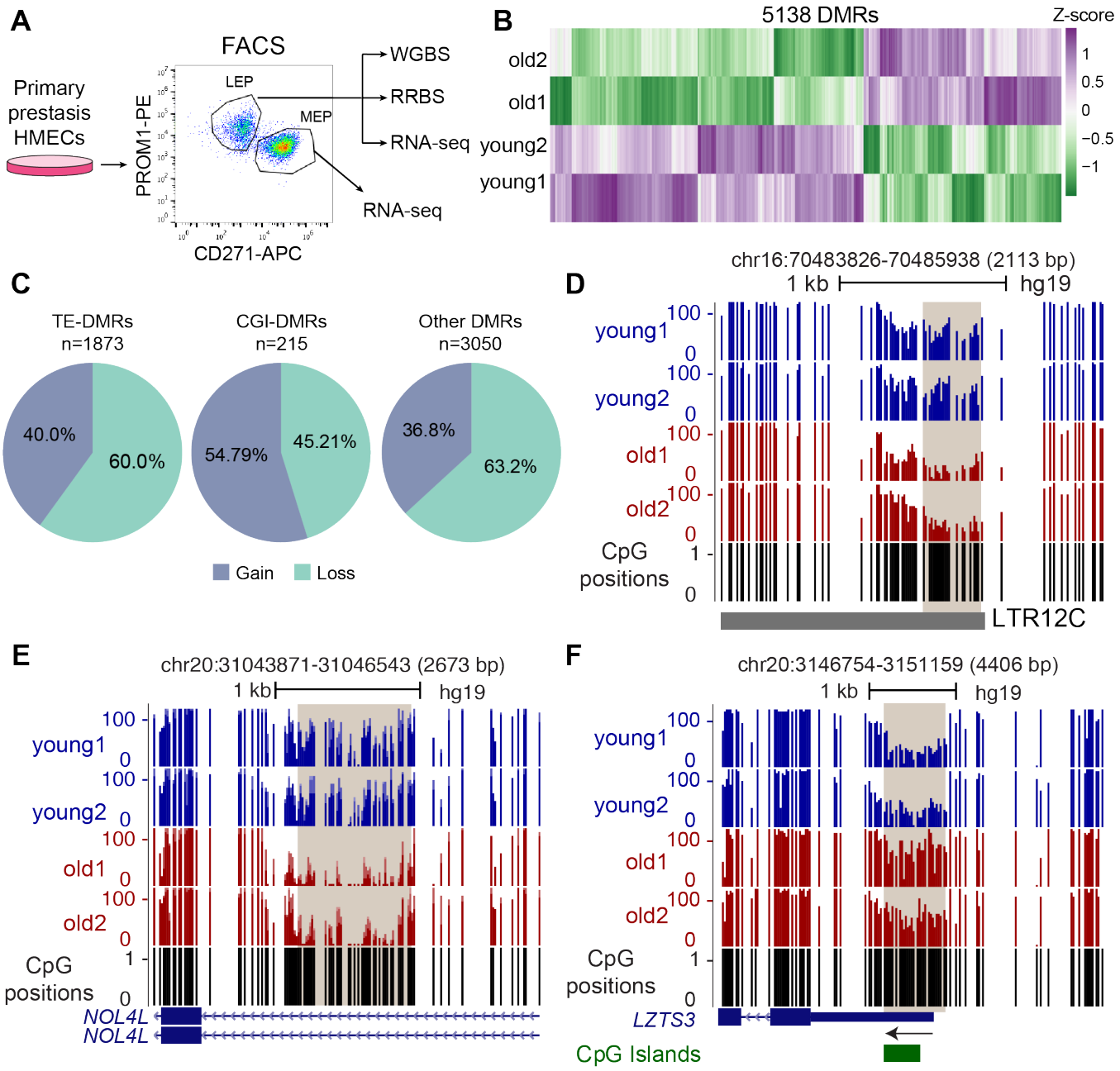

Figure 1

Figure 1: Luminal cells from old individuals show DNA methylation changes at distinct genomic loci. (A) Strategy for enrichment of the luminal and myoepithelial cell populations and design of study. (B) 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 methylation level across samples. Values ranging from higher to lower methylation are shown as purple to green. (C) Percentage DMRs gaining and losing methylation with age are shown for DMRs overlapping transposable elements (TEs), CpG islands (CGIs), and other DMRs that do not overlap either TEs or CGIs. (D) Genome browser tracks showing a representative example of an LTR12C element losing methylation in old samples. (E)

A representative example of an intronic DMR (Other DMRs) at the *NOL4L* locus showing decreased methylation in old LEPs. (F) A representative example of a DMR overlapping a CpG island at the *LZTS3* locus showing increased methylation in old LEPs. DMRs are highlighted in brown.

To determine age-associated changes to DNA methylation, we identified CpGs differentially methylated between the pooled young and old samples. About 900K sites were differentially methylated (n = 967,534, P-value < 0.05) with ~500K sites (n = 539,928) showing decreased methylation (> 25% methylation change) in old LEPs (**Figure S1F**). Next, we identified differentially methylated regions (DMRs) (see Methods) in old LEPs as compared to the young LEPs. We identified a total of 5,138 DMRs (>25% methylation change and P-value < 0.05) with 3,149 DMRs losing methylation and 1,989 gaining methylation in old LEPs (**Figure S1G**). The validity of identified DMRs was confirmed with heatmaps showing methylation levels across samples (**Figure 1B**) as well as kernel smoothed methylation plots of representative DMRs (**Figure S1H**).

We further characterized the locus of all DMRs by overlapping with common genomic features (**Figure S1I**). ~48% of DMRs are intragenic and overlap introns (**Figure S1I**). DMRs overlapping TEs (~36% of all DMRs, **Figure 1C, D**) and DMRs overlapping non-CGI and non-repeat regions (59% of all DMRs, **Figure 1C, E**) both lose and gain DNA methylation, with more DMRs losing DNA methylation. In contrast, more DMRs overlapping CpG islands gain methylation (**Figure 1C, F**). These results suggest that specific loci, including repeat elements and CpG islands, show a differential response to aging in LEPs.

Aging-specific DNA methylation alterations occur at cell type specific regulatory regions.

As the PCA indicated that there were methylation differences between the two old individuals (**Figure S1E)**, we examined whether this could be due to increased heterogeneity

associated with aging (Bahar et al. 2006; Martinez-Jimenez et al. 2017; Cheung et al. 2018; Hernando-Herraez et al. 2019). We first focused on identifying aging-specific signatures in LEPs. To do that, we classified the DMRs as 'shared' or 'variable' based on whether they showed the same behavior in the two old individuals. Shared DMRs are those that showed the most consistent changes in the old LEPs compared to young LEPs (*P*-value < 0.05, methylation difference > 25%) (**Figure S2A-B**). Variable DMRs are those that showed a change in only one of the two old LEPs when compared to the young LEPs (**Figure S2C-D**). Presumably, the shared DMRs are aging-specific signatures that relate to the response of LEPs to the aging breast tissue microenvironment. We characterized the genomic location of these DMRs and found that a higher number of shared DMRs are in genic regions that include promoters, exons, and introns; as well as on CpG islands (**Figure S2E**). Variable DMRs are over-represented at transposable elements, especially at SINE elements (**Figure S2E**). Moreover, a majority of the shared and variable DMRs were located away from transcription start sites (**Figure S2F**) indicating these could be distal regulatory regions.

To understand the functional relevance of these distal DMR sites in the breast tissue, we compared them with chromatin states predicted from published histone modification ChIP-seq datasets performed in mammary epithelial lineages (Pellacani et al. 2016). To discern whether DMRs reflect changes in cell properties, we predicted chromatin states in mature luminal, luminal progenitor, and basal (MEP) cells from RM tissue of a young (19y old) donor (Pellacani 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 basal/MEP-specific promoter and enhancer regions, accounting for about 68% of shared (**Figure S3D, F and H**) and 40% of variable DMRs (**Figure S3E and G**). Since active enhancer sites are hypomethylated (Lister et al. 2009; Stadler et al. 2011), these results

indicate that basal/MEP-specific regulatory elements lose methylation in old LEPs and function as active enhancers. These data are in-line with previous studies suggesting that aging luminal cells acquire basal-like gene expression signatures (Garbe et al. 2012; Pelissier et al. 2014).

Conversely, DMRs that gain methylation overlap with promoter and enhancer regions in luminal progenitor cells accounting for about 40% of shared (**Figure S3I, K and M**) and 32% of the variable DMRs (**Figure S3J and L**). Of note, our strategy of FACS enrichment of luminal cells using PROM1 (CD133) positive and NGFR (CD271) negative cells (**Figure 1A**) includes luminal progenitor cells (Lim et al. 2009). These results indicate that luminal progenitor-specific regulatory regions gain DNA methylation in aging luminal cells. Together, these results suggest that aging LEPs acquire permissive states at basal-specific regions and repress luminal progenitor specific chromatin regions through changes in DNA methylation.

Lineage-specific transcription factors cause aging-associated DNA methylation changes.

Given that DNA methylation changes at DMRs correlated with cell-type-specific chromatin states, we examined the underlying DNA sequence at DMRs for signatures of TF binding. We performed motif enrichment analysis using TF position weight matrices (PWMs) from the JASPAR database (Fornes et al. 2020). DMRs that lose methylation are enriched for binding sites of TFs such as NFIA, EGR1, TEAD1, TEAD4, TP63, TP73, and SOX15 (**Figure 2A left panel: shared DMRs and S4A left panel: variable DMRs**). We determined the expression of these TFs in LEPs versus MEPs and in young and old LEPs using RNA-sequencing data. *TP63*, *NFIA,* and *SOX15* showed increased expression in old LEPs compared to young LEPs (**Figure 2B-D**). In addition, *TP63, NFIA, EGR1, TP73*, and *SOX15* all have relatively higher expression in MEPs as compared to LEPs (**Figure 2A right panel and 2I**). The expression patterns of TFs and the hypomethylation of their targets are concordant with known roles of these TFs. For

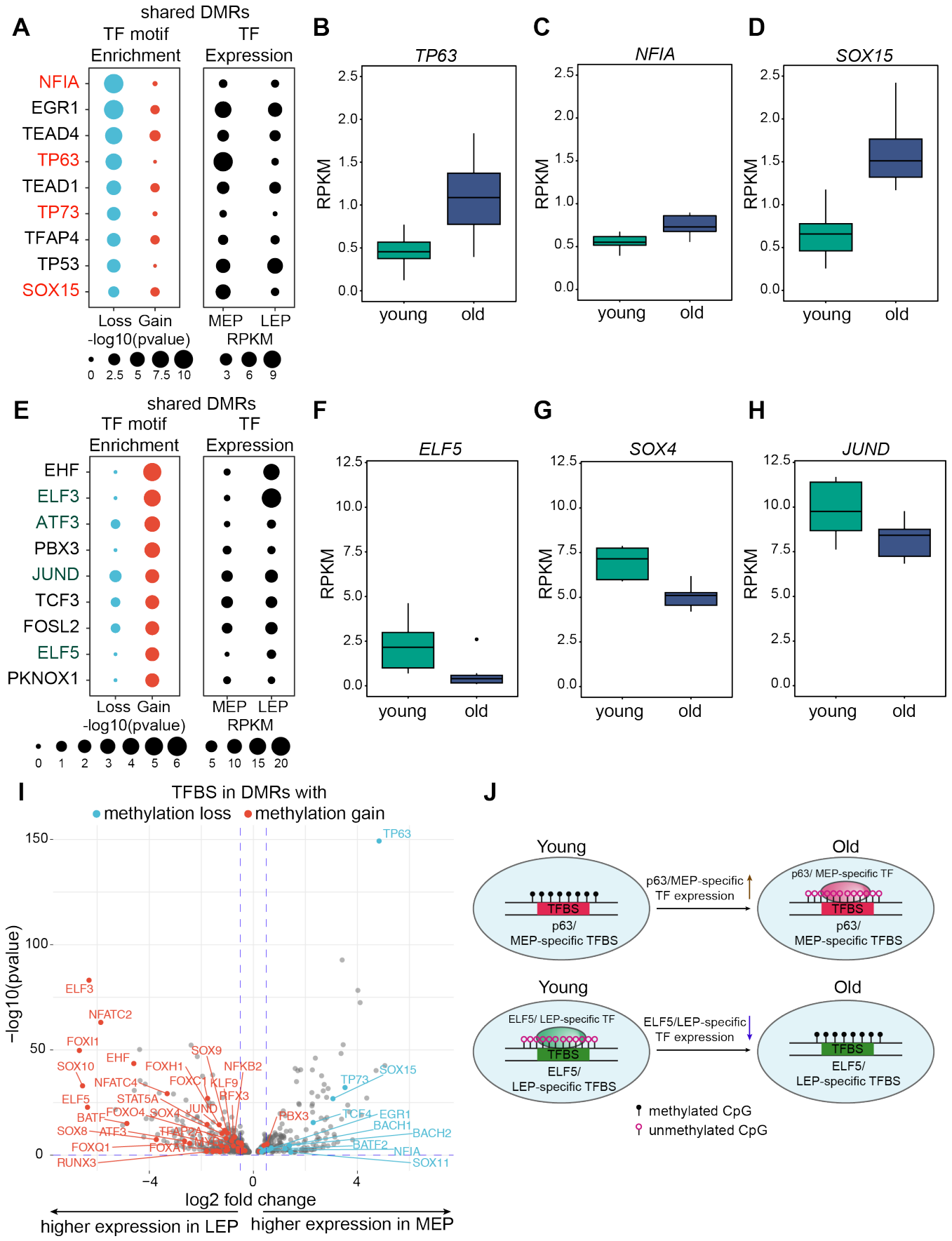
A
shared DMRs
TF motif Enrichment
TF Expression
NFIA
EGR1
TEAD4
TP63
TEAD1
TP73
TFAP4
TP53
SOX15
Loss Gain
MEP LEP
-log10(pvalue)
RPKM
B
TP63
C
NFIA
D
SOX15
young
old
RPKM
E
EHF
ELF3
ATF3
PBX3
JUND
TCF3
FOSL2
ELF5
PKNOX1
F
ELF5
G
SOX4
H
JUND
I
TFBS in DMRs with
methylation loss
methylation gain
−log10(pvalue)
log2 fold change
higher expression in LEP
higher expression in MEP
J
Young
Old
p63/MEP-specific TF expression
p63/ MEP-specific TF
TFBS
p63/ MEP-specific TFBS
ELF5/ LEP-specific TF
ELF5/LEP-specific TF expression
ELF5/ LEP-specific TFBS
methylated CpG
unmethylated CpG

Figure 2

Figure 2: Differentially methylated regions show enrichment of lineage-specific transcription factors. (A) Top ten TF binding sites enriched in the shared DMRs that lose or (E) gain methylation 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-D) Boxplots showing expression levels of transcription factors as indicated in young and old LEPs (n=8, *p*-value < 0.01; two-tailed unpaired t test with Welch's correction). (F-H) Boxplots showing expression levels of luminal lineage-specific transcription factors in young and old LEPs (n=8, *p*-value < 0.05; two-tailed Mann Whitney test). (I) Volcano plot showing expression levels of all TFs expressed in LEPs and MEPs (RPKM > 0.1). TFs with motifs enriched in shared DMRs with methylation loss and gain are highlighted in blue and red respectively. *P*-value < 0.05. (J) 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 unbound sites being methylated, leading to methylation gain.

example, young LEPs do not express *TP63*, which is a myoepithelial lineage-specific transcription factor (Miyano et al. 2017). Furthermore, TEAD1 and TEAD4 are DNA binding partners for YAP, a component of the Hippo signaling pathway. Old LEPs exhibit increased expression of *TP63* as well as increased nuclear YAP localization (Pelissier et al. 2014; Miyano et al. 2017), and thereby increased YAP activity and TEAD binding. 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 2J**).

Conversely, DMRs with methylation gain are enriched for binding sites of transcription factors EHF, ELF3, ELF5, JUND, FOSL2, and ATF3 (**Figure 2E left panel: shared DMRs and S4B left panel: variable DMRs**). ELF5 is a luminal lineage-specific transcription factor that decreases expression with age (**Figure 2F**) (Miyano et al. 2017). Besides *ELF5*, transcription factors *ELF3, SOX4,* and *JUND* show relatively lower expression in MEPs as compared to LEPs (**Figure 2E right panel and 2I**). Expression of *SOX4* and *JUND* also decreases with age in LEPs, indicating a loss of expression of these lineage-specific TFs (**Figure 2F-H**). These results

suggest that decreased expression of luminal lineage-specific TFs may lead to a decrease in binding and protection of binding sites from DNMTs. This might result in DNA methylation and repression of these luminal lineage-specific regions (**Figure 2J**). Generally, sites of methylation loss were enriched for myoepithelial specific TF binding motifs (**Figure 2I: shared DMRs and S4C: variable DMRs**) whereas sites of methylation gain were enriched for luminal lineage-specific TF binding motifs (**Figure 2I and S4C**). Collectively, these results suggest that aging-associated DNA methylation changes in luminal epithelial cells are orchestrated at least in part by lineage-specific transcription factors.

Aging leads to hypomethylation at young transposable elements.

Given that DNA methylation is an important mechanism for suppression of transposable elements (TEs), we asked whether loss of DNA methylation with age occurs at TEs. We observed that TEs were indeed present in DMRs, and the TE content was significantly different than expected for random regions of the same length (**Figure S5A and B**). In general, shared DMRs had a lower TE content than variable DMRs (**Figure S5A and B**), indicating that age-dependent methylation differences at TEs are mainly stochastic.

We further investigated the characteristics of TEs within differentially methylated regions. We observed that evolutionarily younger elements (<100 million years) tend towards DNA methylation loss, whereas the older TE elements (>100 million years) tend to gain methylation (**Figure 3A: variable DMRs and S5C: shared DMRs**). Separating by TE class, SINE elements showed a distribution similar to all TEs, indicating their relative abundance in the DMRs (**Figure 3B: variable DMRs and S5D: shared DMRs**). Young LINE elements lose methylation variably in the old LEPs (**Figure 3C**). However, older LINE elements showed both loss and gain of methylation but tended more towards methylation gain (**Figure 3C: variable DMRs** and **Figure**

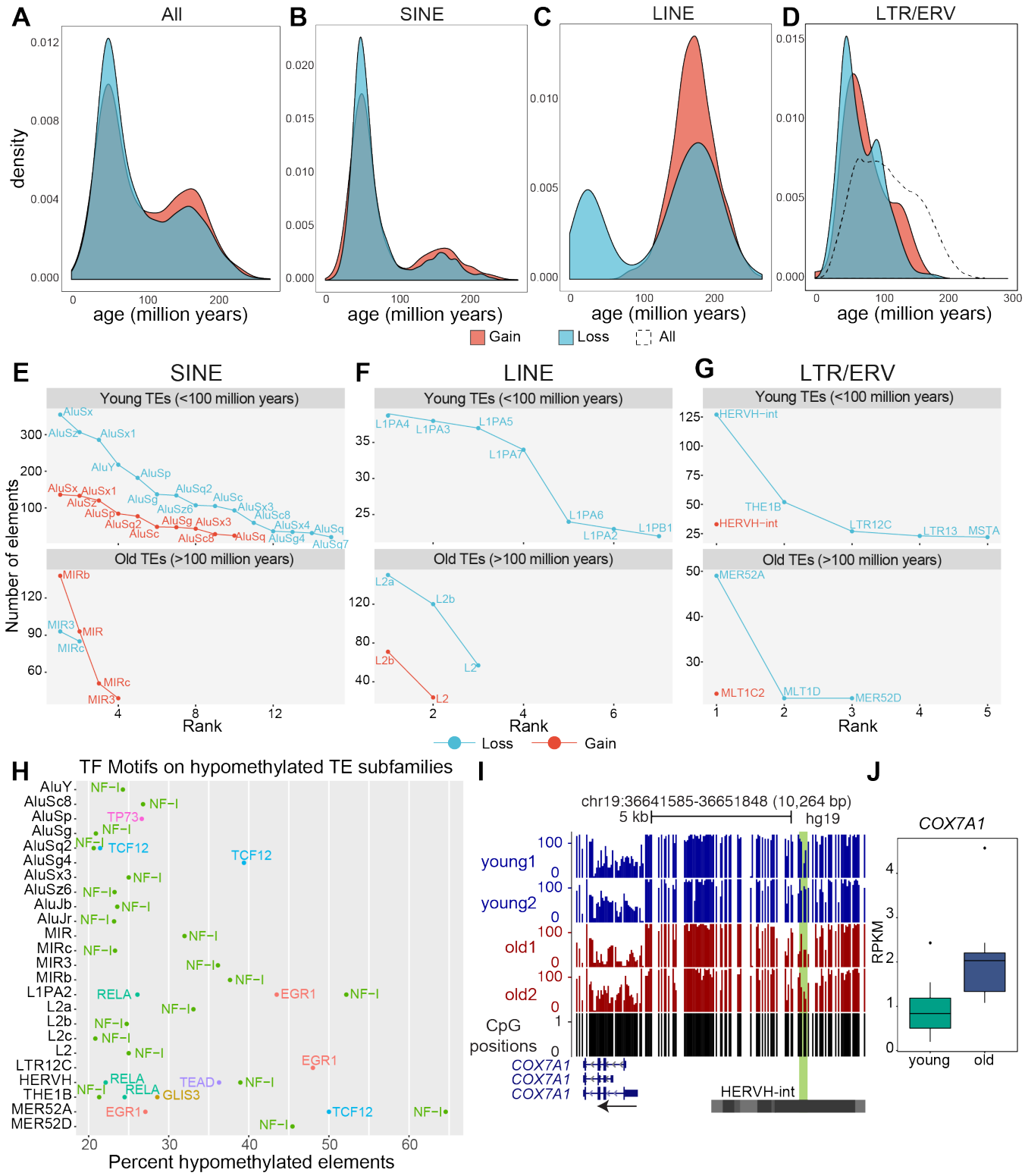

Figure 3

Figure 3: Aging leads to loss of methylation at young transposable elements. (A) Density plots of evolutionary age of transposable elements at variable DMRs with methylation loss (Loss) or gain (Gain). (B) Evolutionary age of SINE, (C) LINE, and (D) LTR/ERV elements at variable DMRs with methylation

loss or gain. (E) Ranked plots of SINE (F) LINE and (G) LTR/ERV subfamilies enriched in variable DMRs. E-G. Plots show the number of elements that lose or gain methylation within young and old TE subfamilies. (H) Percentage number of variable hypomethylated DMRs at specific TE subfamilies containing transcription factor binding sites as indicated. (I) Genome browser tracks showing hypomethylation of a HERVH element near the *COX7A1* gene. (J) Expression of *COX7A1* gene in young and old LEPs, n=8, *p*-value <0.01; two-tailed Mann Whitney test.

S5E: shared DMRs). Interestingly, young LTR/ERV elements showed methylation loss and gain (**Figure 3D: variable DMRs and S5F: shared DMRs**), indicating possible exaptation of these elements in LEPs.

We further inspected the specific subfamilies enriched at variable and shared DMRs with methylation loss and gain. As some TE subfamilies, especially Alu elements, are much more abundant than others, we calculated the expected number of overlaps for each TE subfamily using random genomic regions. To be conservative, TE subfamilies that showed an observed/expected ratio closer to 1 (0.75 - 1.25) were discarded (**Figure S5G-I**). Among the variable DMRs, younger Alu subfamilies were highly abundant in regions with methylation loss, whereas older SINE subfamilies MIRb, and MIR were enriched at regions with methylation gain (**Figure 3E**). Young LINE subfamilies L1PA2-7, and L1PB1 lost methylation, whereas older LINE elements L2a, L2b, and L2 showed methylation loss as well as gain (**Figure 3F**). These results indicate that older LINE and SINE elements are possibly exapted in these cells and change methylation similar to regulatory regions. However, LTR/ERV elements mainly showed methylation loss irrespective of evolutionary age. (**Figure 3G**). LTR/ERV subfamilies HERVH and MER52A were particularly interesting as they were highly enriched at shared as well as variable DMRs (**Figure S5I-J**).

Next, we determined whether the hypomethylated TEs harbor binding sites of transcription factors that could explain the methylation loss at these sites. We performed motif

scanning using TF PWMs from the JASPAR database (Fornes et al. 2020) on the hypomethylated regions of TEs. We found that NF-I binding sites are prevalent across all hypomethylated TE subfamilies (**Figure 3H**). Hypomethylated AluSp elements showed an abundance of TP73 sites, whereas hypomethylated HERVH elements showed a prevalence of TEAD motifs and NF-κB (REL, RELA) motifs (**Figure 3H and S5K**). DMRs in MER52A elements showed the presence of EGR1, TCF12, and NF-I motifs (**Figure 3H**). Binding of these transcription factors that are active in old LEPs (**Figure 2A and I**) could result in DNA methylation loss at these sites through the recruitment of Ten-eleven translocation (TET) enzymes. These demethylated TEs could, in turn, function as regulatory elements and influence the expression of nearby genes. For example, a HERVH element located upstream (~5kb) from the *COX7A1* gene loses methylation with age (**Figure 3I**). *COX7A1* mRNA expression increases with age, indicating that the upstream HERVH element could function as a regulatory element for this gene (**Figure 3J**). Interestingly, *COX7A1* is highly upregulated in the Luminal A subtype of breast cancer (**Figure S5L**). Altogether, these results suggest that evolutionarily young TEs lose methylation with age and could regulate nearby gene expression. However, we note that TF binding sites explained only about 20-70% of the hypomethylation events at TEs (**Figure 3H**).

Aging leads to increased methylation heterogeneity at young CpG dense transposons.

Aging is associated with increased transcriptional variability and cell-to-cell heterogeneity (Bahar et al. 2006; Martinez-Jimenez et al. 2017; Cheung et al. 2018; Hernando-Herraez et al. 2019). As we observed increased variation in CpG methylation between the two old individuals, we examined whether there is an increase in methylation heterogeneity with age in older individuals. For this, we calculated methylation entropy, a measure of methylation heterogeneity between adjacent CpGs, for 150 bp non-overlapping windows across the genome (Jenkinson et

al. 2017). We did not observe any global changes in methylation entropy with increased age across the genome (**Figure S6A-B**).

Since methylation changes could occur through changes in methylation heterogeneity, we asked whether the DMRs show changes in methylation entropy. Indeed, we observed that regions with methylation loss showed a net increase in entropy (**Figure 4A-B and S6C-D)**, whereas regions with methylation gain showed a decrease in entropy (**Figure 4C-D and S6E-F**). These results indicate that concerted methylation changes are brought about by DNA methyltransferases (DNMTs) at regions that gain methylation. Since the variable DMRs represent regions with differing methylation status in the two old individuals, we reasoned that loss of maintenance by DNMT1 could lead to increases in methylation heterogeneity accompanied by methylation loss. A similar loss of methylation due to inefficient maintenance methylation by DNMT1 has been reported at late replicating regions associated with heterochromatin (Zhou et al. 2018). Given that TEs are more likely to be associated with heterochromatin regions and variable DMRs have a higher TE content, we examined whether variable DMRs overlapping TEs exhibit an increase in methylation entropy. Results showed that TE containing DMRs with methylation loss have increased methylation entropy, suggesting a stochastic loss of methylation (**Figure 4E-F**). In contrast, CpG islands that lose methylation do not show net changes in entropy (**Figure 4G-H**). These results indicate that in contrast to regulatory regions enriched with TF binding sites, methylation loss at TEs are brought about by stochastic events indicated by increased disorderness of methylation states.

We further noted that CpG dense transposons such as L1PA2, LTR12C, and AluY show variable methylation loss in the old individuals (**Figure 3E-G**). We investigated whether the loss of methylation at these elements is stochastic accompanied with increased methylation entropy.

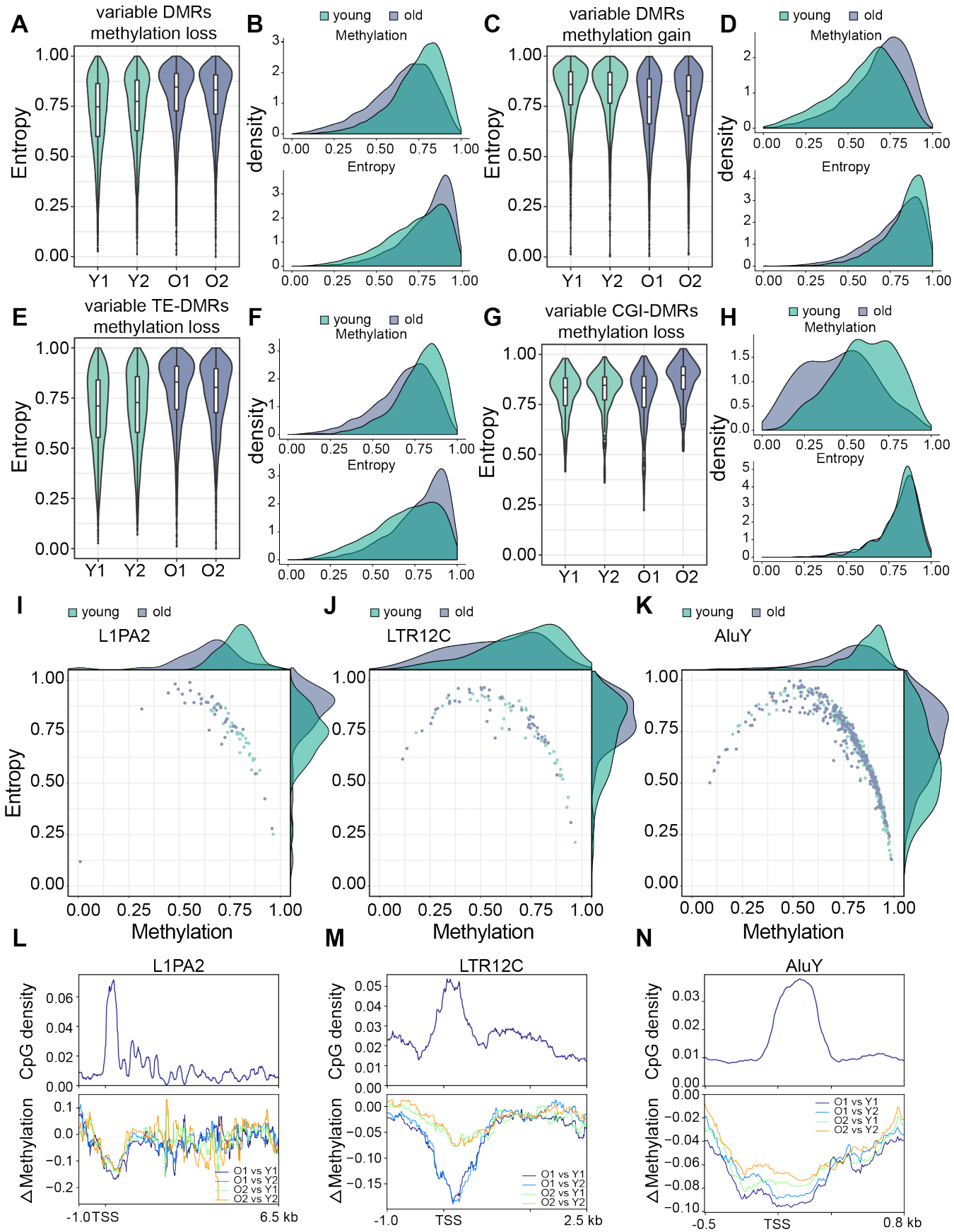
A
variable DMRs
methylation loss
Entropy
Y1
Y2
O1
O2
B
young
old
Methylation
Entropy
density
C
variable DMRs
methylation gain
D
E
variable TE-DMRs
methylation loss
F
G
variable CGI-DMRs
methylation loss
H
I
L1PA2
J
LTR12C
K
AluY
Methylation
L
M
N
CpG density
ΔMethylation
O1 vs Y1
O1 vs Y2
O2 vs Y1
O2 vs Y2
-1.0TSS
6.5 kb
-1.0
TSS
2.5 kb
-0.5
0.8 kb

Figure 4

Figure 4: Aging increases methylation heterogeneity at young CpG dense transposons. (A) Normalized methylation entropy (Entropy) levels at variable DMRs with methylation loss in young and old individuals. (B) Density plots showing DNA methylation (top panel) and entropy levels (bottom panel) at variable DMRs with methylation loss (C) As A, but showing variable DMRs with methylation gain. (D) As B, but showing variable DMRs with methylation gain. (E) Entropy levels at variable DMRs overlapping TEs and (G) variable DMRs overlapping CGIs with methylation loss (F) Density plots showing DNA methylation (top) and entropy levels (bottom) at variable DMRs overlapping TEs and (H) variable DMRs overlapping CGIs with methylation loss. (I) Scatterplots of DNA methylation and entropy levels in L1PA2, (J) LTR12C, and (K) AluY elements that lose methylation. Marginal density plots are shown to display the density of points on the X and Y axes. (L) Aggregate CpG density and methylation change (Δ Methylation) at hypomethylated L1PA2, (M) LTR12C, and (N) AluY elements. Only full-length elements were used to generate aggregate plots. (Y1: young1, Y2: young2, O1: old1, O2: old2)

Indeed, we found an increased methylation entropy associated with methylation loss at L1PA2, LTR12C, and AluY elements in the old individuals (**Figure 4I-K**). Moreover, the methylation loss occurred in the CpG dense region of the element (**Figure 4L-N**). For example, the 5'UTR of L1PA2 elements contain a CpG island-like dense region, and the methylation loss was focused on this region (**Figure 4L**). These results indicate that young CpG dense transposons exhibit stochastic loss of methylation with age.

Age-dependent methylation loss at transposable elements is associated with luminal breast cancer risk.

Given that CpG dense transposons lose methylation with age, we determined whether this might lead to increased breast cancer risk. To this end, we examined DNA methylation profiles generated from reduced representation bisulfite sequencing (RRBS) with additional LEPs from young (n = 6) and old (n = 4) individuals and individuals with a higher-than-average risk for breast cancer (n = 7) (Table S3). We reasoned that since RRBS enriches CpG dense regions of the genome, we would be able to examine methylation levels of CpG dense

transposons in more number of individuals. Indeed, we obtained a higher representation of CpG sites from the CpG-dense SVA class of transposons and CpG islands (**Figure S6G: upper panel**). RRBS was able to capture more than 50% of copies for the CpG dense LTR12 and SVA subfamilies and younger LINE subfamilies (**Figure S6G: lower panel**). We obtained a higher or equivalent coverage at CpG dense regions of SVA_D and LTR12C elements as compared to WGBS (**Figure S6H**). Next, we identified TE containing DMRs with methylation loss in old individuals. We used this set to ask whether high-risk LEPs also show methylation loss at these TEs. PCA and hierarchical clustering analysis on the hypomethylated TE-DMRs showed that, indeed, high-risk LEPs cluster together with the old LEPs (**Figure 5A-B**). This indicates that age-dependent methylation loss at TEs is recapitulated in LEPs from high-risk individuals.

TEs have latent regulatory potential and, when derepressed, can function as regulatory elements. Given that age is a risk factor for luminal breast cancer, we asked whether TEs hypomethylated with age play any role in luminal breast cancer. To examine the potential function of the underlying DNA, we identified chromatin states at these TEs from published breast cancer cell line ChIP-seq datasets (Franco et al. 2018). The results revealed that hypomethylated CpG-dense TEs exist in a bivalent promoter state with H3K4me3 and H3K27me3 marks (**Figure 5C**). Hypomethylated L1HS and L1PA2 elements exist in a bivalent promoter state in luminal and HER2-positive cell lines (**Figure 5C**), whereas bivalent state at SVA_D loci was observed only in HER2-positive cell lines (**Figure 5C**). These results indicate that CpG-denseTEs hypomethylated with age exhibit a poised promoter state in luminal breast cancer cell lines and may function as regulatory elements.

To further support this idea, we analyzed published RNA-seq data from normal and ductal carcinoma in situ (DCIS) samples (Abba et al. 2015). DCIS is considered to be the earliest form of breast cancer. We identified genes upregulated in DCIS samples compared to normal and

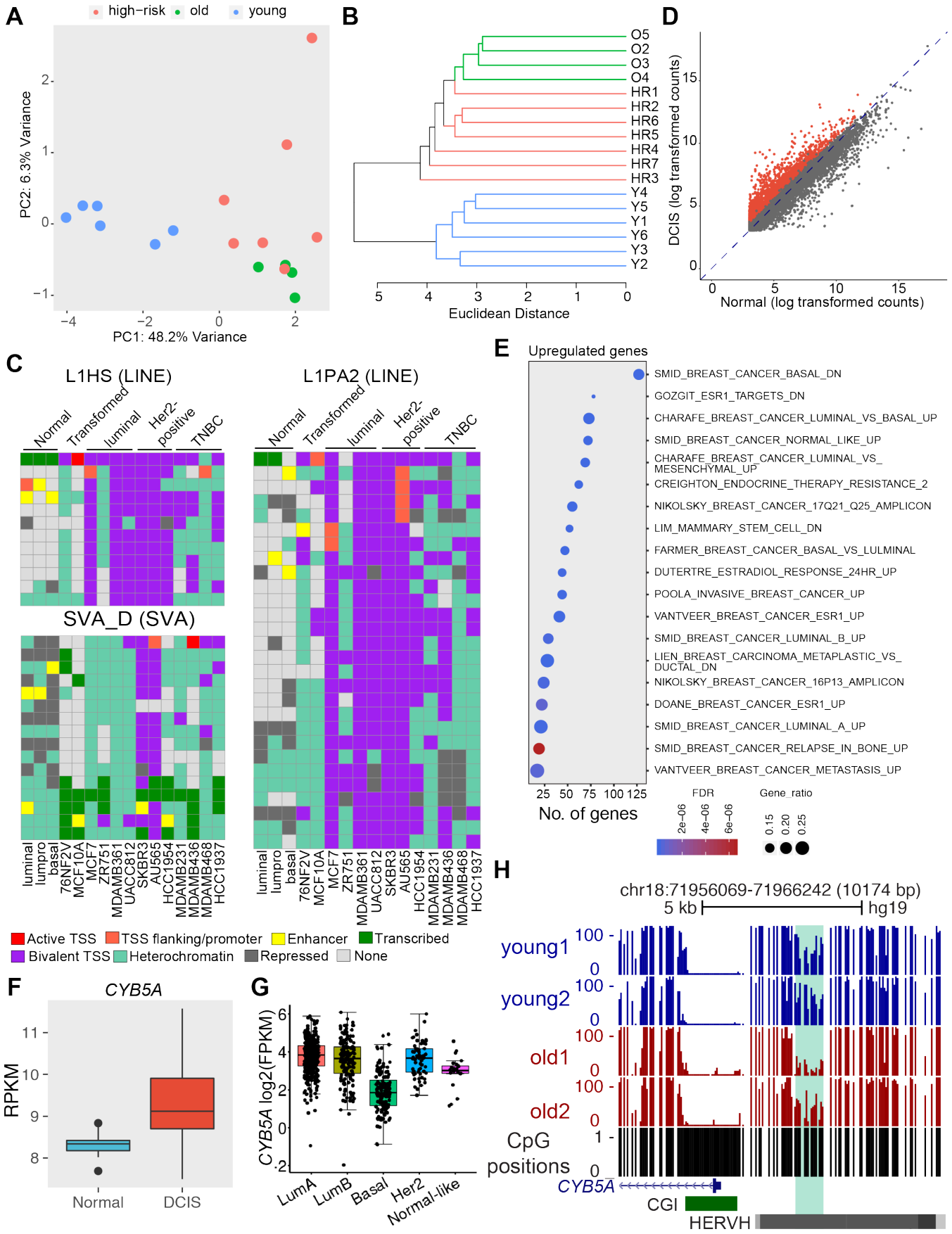
A
high−risk
old
young
PC2: 6.3% Variance
PC1: 48.2% Variance
B
O5
O2
O3
O4
HR1
HR2
HR6
HR5
HR4
HR7
HR3
Y4
Y5
Y1
Y6
Y3
Y2
Euclidean Distance
D
DCIS (log transformed counts)
Normal (log transformed counts)
C
L1HS (LINE)
L1PA2 (LINE)
SVA_D (SVA)
Normal
Transformed
luminal
Her2-positive
TNBC
luminal
lumpro
basal
76NF2V
MCF10A
MCF7
ZR751
MDAMB361
UACC812
SKBR3
AU565
HCC1954
MDAMB231
MDAMB436
MDAMB468
HCC1937
Active TSS
TSS flanking/promoter
Enhancer
Transcribed
Bivalent TSS
Heterochromatin
Repressed
None
E
Upregulated genes
SMID_BREAST_CANCER_BASAL_DN
GOZGIT_ESR1_TARGETS_DN
CHARAFE_BREAST_CANCER_LUMINAL_VS_BASAL_UP
SMID_BREAST_CANCER_NORMAL_LIKE_UP
CHARAFE_BREAST_CANCER_LUMINAL_VS_MESENCHYMAL_UP
CREIGHTON_ENDOCRINE_THERAPY_RESISTANCE_2
NIKOLSKY_BREAST_CANCER_17Q21_Q25_AMPLICON
LIM_MAMMARY_STEM_CELL_DN
FARMER_BREAST_CANCER_BASAL_VS_LULMINAL
DUTERTRE_ESTRADIOL_RESPONSE_24HR_UP
POOLA_INVASIVE_BREAST_CANCER_UP
VANTVEER_BREAST_CANCER_ESR1_UP
SMID_BREAST_CANCER_LUMINAL_B_UP
LIEN_BREAST_CARCINOMA_METAPLASTIC_VS_DUCTAL_DN
NIKOLSKY_BREAST_CANCER_16P13_AMPLICON
DOANE_BREAST_CANCER_ESR1_UP
SMID_BREAST_CANCER_LUMINAL_A_UP
SMID_BREAST_CANCER_RELAPSE_IN_BONE_UP
VANTVEER_BREAST_CANCER_METASTASIS_UP
No. of genes
FDR
Gene_ratio
F
CYB5A
RPKM
Normal
DCIS
G
CYB5A log2(FPKM)
LumA
LumB
Basal
Her2
Normal-like
H
chr18:71956069-71966242 (10174 bp)
5 kb
hg19
young1
young2
old1
old2
CpG positions
CYB5A
CGI
HERVH

Figure 5

Figure 5: Age linked methylation loss at transposable elements is associated with luminal breast cancer risk. (A) Principal component analysis and (B) hierarchical clustering of young, old, and high-risk samples using methylation levels at hypomethylated DMRs overlapping TEs. (C) Heatmaps showing chromatin states identified at hypomethylated L1HS, L1PA2 and SVA_D elements in normal cells, transformed mammary cell lines and breast cancer cell lines. Each row shows a hypomethylated TE identified from WGBS and RRBS data. (D) Upregulated genes in DCIS samples (red) within 50kb of a hypomethylated TE. (E) MSigDb gene sets enriched in genes highlighted in red in D. (F) Expression of *CYB5A* in Normal and DCIS samples (p-value < 0.001; two-tailed unpaired t test with Welch's correction) and (G) in five subtypes of breast tumors from the TCGA BRCA collection (p-value Luminal A (LumA) versus Basal and Luminal B (LumB) versus Basal < 2.2×10^{-16}; using two-tailed Wilcoxon signed-rank test). (H) Genome browser tracks showing hypomethylation of a HERVH element near the *CYB5A* gene promoter.

present within 50kb of a hypomethylated TE identified from the WGBS and RRBS data (**Figure 5D**). We overlapped this set of genes with the curated gene sets available in the MSigDB database (Liberzon et al. 2011). We found that most of these genes are involved in luminal breast cancer etiology and distinguish luminal breast cancers from basal breast cancers (**Figure 5E**). For example, the *CYB5A* gene increases expression in DCIS samples (**Figure 5F**) and is highly expressed in luminal breast tumors (Luminal A and Luminal B) as compared to basal tumors (**Figure 5G**). Genome browser tracks of methylation levels at the *CYB5A* locus show decreased methylation at an upstream HERVH element with age (**Figure 5H**). These results indicate that TEs hypomethylated with age can regulate the expression of genes important for luminal breast cancer and thereby support transformation.

Discussion

We report here a comprehensive analysis of the aging-associated DNA methylation alterations in primary human mammary luminal epithelial cells. This is the first study to use a WGBS approach to examine the impact of age on the DNA methylome of primary human

mammary epithelial cells. Previous studies were done using array-based technologies that cover only a fraction of the CpG sites in the genome (Miyano et al. 2017; Hofstatter et al. 2018; Castle et al. 2020). Importantly, repeat elements are under-represented in these array probes (Bell et al. 2019), making interpretation of DNA methylation at repetitive regions difficult. Moreover, earlier DNA methylation studies have been performed using composite tissues with a mixture of cells (Hofstatter et al. 2018; Castle et al. 2020). We and others have shown that different cell lineages have unique responses to aging, and this is dependent on cell-cell interactions and microenvironment changes (Miyano et al. 2017; Kimmel et al. 2019). Furthermore, aging-related changes in cell composition have been shown in blood cells where older individuals show a skew towards myeloid lineages (Rimmele et al. 2014). Changes in cell composition can confound the interpretation of DNA methylation levels in bulk tissues. Thus, it is imperative to study age-dependent molecular changes in purified cell populations.

Previous DNA methylome studies in human cells have suggested a global loss of methylation with age (Heyn et al. 2012; Vandiver et al. 2015; Jenkinson et al. 2017). These studies were performed using $CD4^+$ T cells from a neonate and centenarian (Heyn et al. 2012), sorted $CD4^+$ T cells (Jenkinson et al. 2017) and skin keratinocytes (Vandiver et al. 2015) from young (<25 y) and old (>70 y) individuals. Furthermore, the global methylation loss with age was accompanied by a global increase in entropy or methylation discordance with age (Jenkinson 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 et al. 2017). Whereas no

global methylation loss was detected, we did, however, find methylation increase at CpG islands. Global loss of methylation and focal increase in methylation at CpG islands have been reported to occur as part of a phenomenon known as 'epigenetic drift' with an increase in age (Christensen et al. 2009).

DNA methylation is an important mechanism for silencing of transposable elements. Reactivation of transposable elements has been shown in cancer cells with global loss of DNA methylation (Ishak and De Carvalho ; Burns 2017). While previous research comparing DNA methylation levels between neonates and centenarians reported that 12% of DMRs that lost methylation with age were located on Alu elements (Heyn et al. 2012), changes in epigenetic silencing at TEs with aging has not previously been systematically investigated. Increased expression of retrotransposons and active retrotransposition have been reported in aging mouse tissues (De Cecco et al. 2013; Benayoun et al. 2019). Furthermore, increased LINE-1 expression has been reported in senescent fibroblasts (De Cecco et al. 2019). However, the role of DNA methylation patterns in these systems was not examined. We examined the impact of aging on DNA methylation at TEs using non-senescent mammary epithelial cells. We show that the evolutionary age of TEs is important for determining the response to aging. As older TEs can be exapted and function as tissue-specific regulatory elements (Sundaram et al. 2014), aging-associated changes can be attributed to cell-type-specific transcription factor activity. This observation may help explain why cancers tend to be tissue-specific. However, young CpG-dense transposons are more likely to be silenced by DNA methylation and presumably lose methylation by inefficient maintenance methylation by DNMT1. Heterochromatin loci that occur in late replicating regions lose methylation by a similar mechanism in cancer cells (Zhou et al. 2018).

We further show that methylation loss at TEs with age is recapitulated in luminal cells from high-risk individuals. Moreover, we show that these TEs are present as bivalent promoter loci in luminal breast cancer cell lines indicating their poised regulatory potential as opposed to being completely silent in normal cells. In addition, we showed that hypomethylated TEs could potentially regulate genes associated with the luminal subtype of breast cancer. Altogether, these observations provide further support to the observation that epigenetic changes with age could determine risk for age-related breast cancer. We also provide evidence for the potential role of TE derepression in luminal breast cancer. Future studies are needed to study the role of these TEs in the early events of mammary tumor initiation.

Methods

Isolation of mammary luminal epithelial cells

Primary HMECs were generated and maintained, as described previously (Labarge 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. Cells 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 with Aria III (Becton Dickinson) as $CD133^{+}/CD271^{-}$ and $CD133^{-}/CD271^{+}$ fractions, respectively.

WGBS

Quick-DNA Microprep kit (Zymo Research) was used to extract DNA. 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). Library

preparation was done using TruSeq DNA Methylation Kit with 75ng of bisulfite converted DNA as per manufacturer's recommendations. 150bp paired-end sequencing was performed on a single lane of Illumina Hiseq X Ten. We obtained about 38X coverage per genome to an average of 414 million paired-end reads (Table S1).

Sequencing reads were hard trimmed 8 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 et al. 2009). Bismark (deduplicate_bismark) to remove PCR duplicates. Methpipe (Song 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. We achieved sufficient coverage of 91.3% and 90.4% of the CpG sites in young and old samples, respectively, with a mean coverage of 9.4 methylation calls per site.

DMR identification

For identifying differentially methylated CpG sites, Methpipe (Song et al. 2013) was used (**Figure S1F**), Only CpG sites with coverage > 3× in pooled samples were considered. CpG sites with greater than 25% change in methylation levels (*P*-value < 0.05) were considered as differentially methylated.

DMRfinder (Gaspar and Hart 2017) was used to identify DMRs. Briefly, CpG sites with at least 4X coverage in all four 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 25% change (*P*-value < 0.05) between pooled young and old samples (**Figure S1F, G**). For variable DMRs, each old sample was compared with the pooled young samples to determine methylation changes. Regions with greater than 25% change (*P*-value < 0.05) were identified for both old individuals. A union set of regions with loss and gain of methylation were classified as variable DMRs.

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 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. In case a DMR overlaps both a CpG island and a repeat element, the DMR was assigned to the element contributing higher number of CpG sites.

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

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 S3A-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 S3C**). ChIP-seq data for breast cancer cell lines were downloaded from GEO (Accession number: GSE85158 (Franco et al. 2018)). Aligned bed files were used to generate binarized files using BinarizeBed with default options. The binarized data were used to run ChromHMM LearnModel with 18 states. Functionally similar chromatin states were grouped together to nine distinct chromatin states. 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.

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 et al. 2015) with the --dta parameter. Stringtie (version 1.3.4) (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 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 ≥ 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 et al. 2015). For LEP versus MEP comparison, DESeq2 (Love 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 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 et al. 2013) with --sjdbOverhang 100 --winAnchorMultimapNmax 100 --outFilterMultimapNmax 100 --outFilterMismatchNmax 3 --alignEndsType EndToEnd options. DESeq2 package (Love et al. 2014) was used to normalize counts and identify differentially expressed genes (absolute log2 fold change ≥ 0.5 and p-value < 0.05).

DMR overlaps with transposable elements

The extent of overlap of DMRs and random regions with TEs (RepeatMasker annotation) was calculated using Genomic Association Tester (GAT) (Heger et al. 2013). Random regions with the same size and GC content distribution as DMRs were selected from ungapped regions of the hg19 genome. The number of DMRs overlapping TEs and the number of nucleotide bases overlapping TEs was calculated. Simulations with random regions were performed 10,000 times. 95% confidence intervals for the number of overlaps with random regions (expected) were plotted with the observed number of overlaps (**Figures S5A and S5B**). Empirical *P*-values are reported for 10000 simulations. Multiple testing correction was performed using the Benjamini-Hochberg method.

RRBS

Quick-DNA Microprep kit (Zymo Research) was used to extract DNA and subjected to a streamlined RRBS protocol modified from (Meissner et al. 2008). Briefly, 250 ng of genomic DNA was digested with *Msp*I at 37°C overnight followed by end-filling, A-tailing and ligation with methylated universal adaptors. Insert fragments over 40 bp were size selected and bisulfite converted using EZ DNA Methylation kit (Zymo Research, Cat D5001). After library amplification, 160-500bp fragments were purified using 6% of TBE polyacrylamide gels and sequenced on an Illumina HiSeq2500 with paired end mode of 101cycles (Table S3). Sequencing reads were adapter trimmed as well as clipped 4 bp on each end using Trim Galore (version 0.4.5) using –illumina --rrbs --trim1 options. 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 et al. 2009). Methpipe (Song et al. 2013) (version 3.4.3) was used to determine the methylation status of individual CpG sites. DMRs were identified using a similar approach as WGBS.

Additional data analysis

Motif scanning was performed using fimo (Grant et al. 2011) with JASPAR 2020 database with default parameters (Fornes et al. 2020). Motif enrichment analysis was performed using findMotifsGenome.pl in homer (Heinz 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 et al. 2008). Normalized methylation entropy and mean methylation were calculated on 150bp non-overlapping windows across the genome using informME (Jenkinson et al. 2018). Figures were made in R (version 3.51) using the ggplot2 package (Wickham 2016) and assembled in Adobe Illustrator.

Data Access

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

Acknowledgements

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. We thank other members of the Schones laboratory for helpful discussions and comments.

Authors' contributions

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

Disclosure Declaration

The authors declare that there are no conflicts of interests.

References

Abba MC, Gong T, Lu Y, Lee J, Zhong Y, Lacunza E, Butti M, Takata Y, Gaddis S, Shen J et al. 2015. A Molecular Portrait of High-Grade Ductal Carcinoma In Situ. *Cancer Res* **75**: 3980-3990.

Bahar R, Hartmann CH, Rodriguez KA, Denny AD, Busuttil RA, Dolle ME, Calder RB, Chisholm GB, Pollock BH, Klein CA et al. 2006. Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**: 1011-1014.

Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, Christensen BC, Gladyshev VN, Heijmans BT, Horvath S et al. 2019. DNA methylation aging clocks: challenges and recommendations. *Genome Biol* **20**: 249.

Benayoun BA, Pollina EA, Singh PP, Mahmoudi S, Harel I, Casey KM, Dulken BW, Kundaje A, Brunet A. 2019. Remodeling of epigenome and transcriptome landscapes with aging in mice reveals widespread induction of inflammatory responses. *Genome Res* **29**: 697-709.

Benz CC. 2008. Impact of aging on the biology of breast cancer. *Crit Rev Oncol Hematol* **66**: 65-74.

Bird A. 2002. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**: 6-21.

Booth LN, Brunet A. 2016. The Aging Epigenome. *Mol Cell* **62**: 728-744.

Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvak Z, Levin HL, Macfarlan TS et al. 2018. Ten things you should know about transposable elements. *Genome Biol* **19**: 199.

Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL, Ng HH et al. 2008. Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18**: 1752-1762.

Burns KH. 2017. Transposable elements in cancer. *Nat Rev Cancer* **17**: 415-424.

Castle JR, Lin N, Liu J, Storniolo AMV, Shendre A, Hou L, Horvath S, Liu Y, Wang C, He C. 2020. Estimating breast tissue-specific DNA methylation age using next-generation sequencing data. *Clin Epigenetics* **12**: 45.

Cheung P, Vallania F, Warsinske HC, Donato M, Schaffert S, Chang SE, Dvorak M, Dekker CL, Davis MM, Utz PJ et al. 2018. Single-Cell Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging. *Cell* **173**: 1385-1397 e1314.

Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH, Karagas MR, Padbury JF, Bueno R et al. 2009. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. *PLoS Genet* **5**: e1000602.

Cole JJ, Robertson NA, Rather MI, Thomson JP, McBryan T, Sproul D, Wang T, Brock C, Clark W, Ideker T et al. 2017. Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol* **18**: 58.

Cordaux R, Batzer MA. 2009. The impact of retrotransposons on human genome evolution. *Nat Rev Genet* **10**: 691-703.

De Cecco M, Criscione SW, Peterson AL, Neretti N, Sedivy JM, Kreiling JA. 2013. Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging (Albany NY)* **5**: 867-883.

De Cecco M, Ito T, Petrashen AP, Elias AE, Skvir NJ, Criscione SW, Caligiana A, Brocculi G, Adney EM, Boeke JD et al. 2019. L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**: 73-78.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21.

Dos Santos CO, Dolzhenko E, Hodges E, Smith AD, Hannon GJ. 2015. An epigenetic memory of pregnancy in the mouse mammary gland. *Cell Rep* **11**: 1102-1109.

Ernst J, Kellis M. 2017. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12**: 2478-2492.
Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP, Correard S, Gheorghe M, Baranasic D et al. 2020. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**: D87-D92.
Franco HL, Nagari A, Malladi VS, Li W, Xi Y, Richardson D, Allton KL, Tanaka K, Li J, Murakami S et al. 2018. Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis. *Genome Res* **28**: 159-170.
Garbe JC, Pepin F, Pelissier FA, Sputova K, Fridriksdottir AJ, Guo DE, Villadsen R, Park M, Petersen OW, Borowsky AD et al. 2012. Accumulation of multipotent progenitors with a basal differentiation bias during aging of human mammary epithelia. *Cancer Res* **72**: 3687-3701.
Gaspar JM, Hart RP. 2017. DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. *BMC Bioinformatics* **18**: 528.
Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017-1018.
Hansen KD, Langmead B, Irizarry RA. 2012. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* **13**: R83.
Heger A, Webber C, Goodson M, Ponting CP, Lunter G. 2013. GAT: a simulation framework for testing the association of genomic intervals. *Bioinformatics* **29**: 2046-2048.
Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**: 576-589.
Hernando-Herraez I, Evano B, Stubbs T, Commere PH, Jan Bonder M, Clark S, Andrews S, Tajbakhsh S, Reik W. 2019. Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nat Commun* **10**: 4361.
Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F, Carmona FJ et al. 2012. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A* **109**: 10522-10527.
Hofstatter EW, Horvath S, Dalela D, Gupta P, Chagpar AB, Wali VB, Bossuyt V, Storniolo AM, Hatzis C, Patwardhan G et al. 2018. Increased epigenetic age in normal breast tissue from luminal breast cancer patients. *Clin Epigenetics* **10**: 112.
Horvath S. 2013. DNA methylation age of human tissues and cell types. *Genome Biol* **14**: R115.
Ishak CA, De Carvalho DD. Reactivation of Endogenous Retroelements in Cancer Development and Therapy. *Annual Review of Cancer Biology* **4**.
Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ. 2007. Cancer statistics, 2007. *CA Cancer J Clin* **57**: 43-66.
Jenkins EO, Deal AM, Anders CK, Prat A, Perou CM, Carey LA, Muss HB. 2014. Age-specific changes in intrinsic breast cancer subtypes: a focus on older women. *Oncologist* **19**: 1076-1083.
Jenkinson G, Abante J, Feinberg AP, Goutsias J. 2018. An information-theoretic approach to the modeling and analysis of whole-genome bisulfite sequencing data. *BMC Bioinformatics* **19**: 87.
Jenkinson G, Pujadas E, Goutsias J, Feinberg AP. 2017. Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nat Genet* **49**: 719-729.
Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12**: 996-1006.
Kim D, Langmead B, Salzberg SL. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**: 357-360.

Kimmel JC, Penland L, Rubinstein ND, Hendrickson DG, Kelley DR, Rosenthal AZ. 2019. Murine single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging. *Genome Res* **29**: 2088-2103.

Kolde R. 2018. pheatmap: Pretty Heatmaps. R package version 1.0.10.

Krueger F, Andrews SR. 2011. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**: 1571-1572.

Labarge MA, Garbe JC, Stampfer MR. 2013. Processing of human reduction mammoplasty and mastectomy tissues for cell culture. *J Vis Exp* doi:10.3791/50011.

Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Lee JK, Garbe JC, Vrba L, Miyano M, Futscher BW, Stampfer MR, LaBarge MA. 2015. Age and the means of bypassing stasis influence the intrinsic subtype of immortalized human mammary epithelial cells. *Front Cell Dev Biol* **3**: 13.

Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**: 1739-1740.

Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A et al. 2009. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15**: 907-913.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**: 315-322.

Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G. 2013. The hallmarks of aging. *Cell* **153**: 1194-1217.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550.

Maegawa S, Hinkal G, Kim HS, Shen L, Zhang L, Zhang J, Zhang N, Liang S, Donehower LA, Issa JP. 2010. Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res* **20**: 332-340.

Martinez-Jimenez CP, Eling N, Chen HC, Vallejos CA, Kolodziejczyk AA, Connor F, Stojic L, Rayner TF, Stubbington MJT, Teichmann SA et al. 2017. Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science* **355**: 1433-1436.

Miyano M, Sayaman RW, Stoiber MH, Lin CH, Stampfer MR, Brown JB, LaBarge MA. 2017. Age-related gene expression in luminal epithelial cells is driven by a microenvironment made from myoepithelial cells. *Aging (Albany NY)* **9**: 2026-2051.

Pelissier FA, Garbe JC, Ananthanarayanan B, Miyano M, Lin C, Jokela T, Kumar S, Stampfer MR, Lorens JB, LaBarge MA. 2014. Age-related dysfunction in mechanotransduction impairs differentiation of human mammary epithelial progenitors. *Cell Rep* **7**: 1926-1939.

Pelissier Vatter FA, Schapiro D, Chang H, Borowsky AD, Lee JK, Parvin B, Stampfer MR, LaBarge MA, Bodenmiller B, Lorens JB. 2018. High-Dimensional Phenotyping Identifies Age-Emergent Cells in Human Mammary Epithelia. *Cell Rep* **23**: 1205-1219.

Pellacani D, Bilenky M, Kannan N, Heravi-Moussavi A, Knapp D, Gakkhar S, Moksa M, Carles A, Moore R, Mungall AJ et al. 2016. Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks. *Cell Rep* **17**: 2060-2074.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**: 290-295.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.

Quong J, Eppenberger-Castori S, Moore D, 3rd, Scott GK, Birrer MJ, Kueng W, Eppenberger U, Benz CC. 2002. Age-dependent changes in breast cancer hormone receptors and oxidant stress markers. *Breast Cancer Res Treat* **76**: 221-236.

Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT, Finer S, Valdes AM et al. 2010. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res* **20**: 434-439.

Rimmele P, Bigarella CL, Liang R, Izac B, Dieguez-Gonzalez R, Barbet G, Donovan M, Brugnara C, Blander JM, Sinclair DA et al. 2014. Aging-like phenotype and defective lineage specification in SIRT1-deleted hematopoietic stem and progenitor cells. *Stem Cell Reports* **3**: 44-59.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**: e47.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Smit A, Hubley, R & Green, P. . 2013-2015. RepeatMasker Open-4.0. http://www.repeatmasker.org.

Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, Garvin T, Kessler M, Zhou J, Smith AD. 2013. A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* **8**: e81148.

Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**: 490-495.

Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T. 2014. Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* **24**: 1963-1976.

Vandiver AR, Irizarry RA, Hansen KD, Garza LA, Runarsson A, Li X, Chien AL, Wang TS, Leung SG, Kang S et al. 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.

Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, Laird PW, Berman BP. 2018. DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* **50**: 591-602.