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

DNA methylation in mammalian genomes is critical for repression of transposable elements (TEs) and regulation of transcription factor (TF) binding. Methylation loss in cancer cells results in widespread TE activation and altered TF binding that subsequently impact transcriptional networks. While the aging process is similarly characterized by changes to the methylome, the extent and its functional consequences have remained unclear. We report here that aging leads to distinct methylation changes in luminal epithelial cells, a key cell lineage implicated in age-related breast cancers. Regulatory elements and evolutionarily old TEs display consistent methylation changes at lineage-specific TF binding sites. Evolutionarily younger TEs, however, exhibit stochastic methylation loss that corresponds with increased methylation entropy. Each of these classes of methylation change impact the regulation of genes associated with luminal breast cancer. Altogether, our results indicate that aging leads to distinct DNA methylation changes that can rewire gene networks important for breast cancer.

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

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 [1,2]. 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 [3]. It is now well-established, however, that DNA methylation patterns are altered with age [4-7]. DNA methylome comparisons of human $CD4^+$ T cells from a newborn and a centenarian suggested an overall decrease in methylation levels with age [7]. Moreover, DNA methylation levels of specific CpG sites show a significant correlation with age across diverse tissues [5]. This set of CpG dinucleotides, known as an 'epigenetic clock', can predict the chronological age of cells based on DNA methylation patterns [5]. It is clear that DNA methylation changes to specific CpG sites occur with aging, whereas the causes of these alterations and their impact on disease risk and progression are not well understood.

Over 75% of breast cancer diagnoses are in women over the age of 50 [8], and 80% of those age-related cancers are of the luminal subtype [9]. Aging in breast tissue is accompanied by increases in estrogen receptor expression [10], and significant physiological changes such as an increase in fat content and a decrease in connective tissue [11] that alter the microenvironment of the mammary epithelium. Previous studies have suggested a complex interplay between the various epithelial cell types and microenvironment that leads to changes in molecular properties and the proportion of cell populations [12,13]. Luminal epithelial cells (LEps) and myoepithelial cells (MEps) are the two principal lineages in breast epithelia. Striking age-dependent changes in abundance and phenotype occur among LEps [12,14-16], which are the chief suspect cell of origin for luminal cancer subtypes [17]. Furthermore, aging-associated tissue microenvironment changes contribute to epigenetic and transcriptional changes in the mammary epithelium [18].

76 *In vitro*, transformed human mammary epithelial cells (HMECs) from older
77 postmenopausal women (> 55 y) exhibit gene expression patterns characteristic of luminal
78 breast cancer subtypes. In contrast, transformed HMECs from younger premenopausal women
79 (< 30 y) show more properties of basal breast cancers [19]. This suggests that cells from older
80 women may harbor age-related epigenetic states that predispose them to develop luminal-
81 subtype breast cancers. Intriguingly, LEps from older women express signaling molecules
82 consistent with those of immortalized luminal cells, suggesting that age-associated changes can
83 elevate cancer risk [14]. Epigenetic clock studies in breast tissue have indicated that women with
84 luminal breast cancers exhibit increased epigenetic age even in the adjacent normal breast
85 tissue [20]. Although earlier studies have investigated the role of epigenetic changes in breast
86 tissue with age and its potential role in elevating breast cancer risk [20,21], these studies were
87 either performed in bulk mammary tissue or interrogated a small subset of CpG sites in the
88 genome. Therefore, a complete understanding of age-associated DNA methylation changes in
89 specific human mammary cell types is still lacking.

90 Repetitive elements derived from transposable elements (TEs) constitute ~50% of the
91 human genome and are classified based on their method of transposition [22]. While some
92 elements, such as Alu, LINE, and SVA elements, have maintained retrotransposition
93 competence [23], most elements have accumulated enough mutations to lose their ability to
94 propagate [24]. They can still, however, carry latent regulatory potential [24]. Cancer cells with
95 genome-wide loss of DNA methylation frequently exhibit activation of transposable elements
96 [25,26]. Increased expression of retrotransposons and active retrotransposition have been
97 observed in aging mouse tissues [27,28], and LINE-1 expression is increased during fibroblast
98 replicative senescence, a form of cellular aging [29]. We, therefore, reasoned that aging would
99 lead to the loss of epigenetic silencing of transposons.

100 To evaluate the impact of aging on TEs we profiled global DNA methylation levels using
101 WGBS in non-senescent human mammary luminal epithelial cells from young and old women.
102 We found that the chronological age of human mammary luminal epithelial cells was associated
103 with DNA methylation changes at regulatory elements and loss of methylation at evolutionarily
104 young transposable elements. Methylation changes at regulatory elements are driven by altered
105 expression and activity of lineage-specific transcription factors such as ELF5 and TP63.
106 Recently evolved TEs mainly lose methylation in a stochastic manner and exhibit increased
107 methylation disorder, presumably due to loss of fidelity of maintenance methylation [30]. TEs that
108 lose methylation with age were also hypomethylated in breast cancers. Loss of suppression of
109 TEs leads to their cryptic expression in pre-breast cancer tissues (DCIS). Furthermore, they
110 function as regulatory elements in breast cancer influencing increased expression of genes
111 associated with luminal-subtype breast cancers. Overall, our results suggest that CpG-dense
112 TEs lose silencing with age and can function as regulatory elements that lead to expression of
113 key deleterious genes during breast cancer development.
114

115 **Results**

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

117 To investigate the genome-wide changes in DNA methylation associated with aging in
118 the mammary tissue, we performed whole-genome bisulfite sequencing (WGBS-seq) on human
119 mammary LEps from five younger (<30y) and five older (>55y) women (**Figure S1A**). We
120 isolated LEps from 4th passage (p4) cultures of pre-stasis finite-lifespan HMECs from normal
121 reduction mammoplasty (RM) tissues [12]. These pre-stasis LEps from primary cultures maintain
122 gene expression and DNA methylation profiles that mirror those in tissue [16,18]. Moreover, these
123 are non-senescent cells and therefore are a valuable resource that can be used to interrogate

124 aging-dependent changes in the breast tissue – distinct from those arising from cellular
125 senescence.

126 We sequenced all ten WGBS libraries to ~30X genome coverage, generating ~ 1B 150-
127 bp paired-end reads (~86 Gb uniquely mapped sequence per sample) (Table S1) and obtained
128 a representation of about 95% of individual CpG sites in our samples (see Methods and Tables
129 S1, S2). To obtain a global view of aging-dependent changes to the methylome, we compared
130 the WGBS profiles of young and old individuals. We did not observe any changes in the global
131 methylation levels across the genome with age (**Figure S1B, C**). However, principal component
132 analysis (PCA) of the WGBS data separated the old and the young samples on the first principal
133 component axes (16.9% cumulative variance; **Figure S1D**) indicating that there are differences
134 in the methylomes that are not apparent on a genome-scale. Pairwise comparison of CpG
135 methylation of all ten samples (45 combinations) at various genomic features showed that
136 methylation at CpG islands, promoters and exons are highly correlated among samples (**Figure**
137 **S1E**). However, methylation levels at repeat regions, especially at SVA elements and SINE
138 elements were the least correlated, indicating high methylation variability at these sites (**Figure**
139 **S1E**). As mapping to repetitive regions can be challenging, we assessed mappability and
140 coverage at repetitive regions. We found no systematic biases, suggesting these results are not
141 due to technical artifacts from mapping (**Fig S1F-G**).

142 To identify genome-wide age-associated changes to DNA methylation, we found all
143 differentially methylated regions (DMRs) (see Methods) in old LEps as compared to the young
144 LEps. We identified 18,860 age dependent DMRs, with 7,207 showing methylation loss and
145 11,653 showing methylation gain in old LEps (**Figure 1A**). Most DMRs were 100-500bp in length
146 (**Figure S1H**). The methylation changes at DMRs were confirmed with heatmaps showing
147 methylation levels across samples (**Figure 1A**), as well as kernel smoothed methylation plots of
148 representative DMRs (**Figure 1B**). DNA methylation changes were observed at all common

149 genomic features including in introns (53%), gene promoters (10%), CpG islands (3%) and
150 intergenic regions (45%) (**Figure S1I-J**). These results suggest that specific loci, including repeat
151 elements and CpG islands undergo methylation changes in LEps in response to aging. We
152 identified genes with methylation changes at promoter regions and performed gene set
153 enrichment analysis (GSEA) to identify pathways associated with aging in LEps. Genes with
154 methylation loss in old LEps were enriched in gene sets expressed in mammary stem cells,
155 epithelial-to-mesenchymal transition, apical junction and genes downregulated in luminal breast
156 cancer (**Figure S2A**). Genes that gained methylation at promoters were enriched in Polycomb
157 target genes in embryonic stem cells and luminal genes that are downregulated in mammary
158 stem cells and basal breast cancer but upregulated in luminal breast cancer (**Figure S2B**). Age-
159 dependent increase in DNA methylation of Polycomb (PRC2) target gene promoters has been
160 observed in other tissues and model systems [4,31-34]. Moreover, EZH2, a component of the PRC2
161 complex is important for maintaining the mammary luminal progenitor population [35,36]. These
162 results indicate aging induces DNA methylation changes at genes specific to mammary epithelial
163 lineages and luminal subtype breast cancers.

164 Transposable elements contribute transcription factor binding sites [36] and can function as
165 regulatory elements even in normal tissue [37,38]. We identified the repeat elements that overlap
166 DMRs and determined if any TE subfamilies were enriched at these regions. We found that LTR
167 subfamilies, HERVH, MER52A, MER52D and LTR27E were enriched at DMRs with methylation
168 loss in old LEps (**Figure 1C**). HERVH elements have been shown to function as enhancers,
169 regulate long non-coding RNAs, and are important for the maintenance of pluripotency [39].
170 MER52A and MER52D elements function as enhancers in other cell types [40,41] indicating these
171 could function as enhancers in old LEps. Centromeric gamma satellite repeat (GSAT) and
172 human satellite 4 (HSAT4) were also enriched at regions of methylation loss (**Figure 1C**). DNA
173 methylation loss at these elements was accompanied by an increase in chromatin accessibility

174 in old LEps (**Figure 1D, S2E and F**). Decondensation of pericentromeric heterochromatin is a
175 characteristic feature of senescent cells [42] and centromere dysfunction has been reported with
176 age [43]; however, methylation changes in centromeric DNA in aging cells has not been studied.
177 DMRs with methylation gain showed enrichment of LTR subfamilies LTR10C, HERVE, and
178 MER11D (**Figure S2C**), which were less methylated than expected in the younger cells,
179 suggesting they have been co-opted in LEps (**Figure S2D**). MER11D elements have previously
180 been shown to be hypomethylated in breast tissue and function as tissue-specific enhancers [37].
181

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

184 To investigate whether DNA methylation changes were associated with transcription
185 factor (TF) binding signatures, we performed motif enrichment analysis using position weight
186 matrices (PWMs) from the JASPAR database [44]. DMRs that lose methylation are enriched for
187 binding sites of TFs such as EGR1, TP63, TP73, TEAD1, TEAD4, and RUNX1 (**Figure 2A left**
188 **panel**). TEs with methylation loss also showed enrichment of similar transcription factor binding
189 sites, including TEAD1-4, TP63, TP73, and EGR1 (**Figure 2B left panel**). NFI half binding sites
190 were enriched in regions with loss as well as gain of methylation. NFI half-sites can be bound
191 by NFI family TFs NFIA, NFIC, and NFIX. TEAD, TP63, TP73, NFI, and EGR1 sites have been
192 previously shown to be enriched in active regulatory elements in basal/myoepithelial cells [45]. To
193 investigate whether these TFs change expression with age, we checked their expression using
194 RNA-sequencing data in young versus old LEps, and in LEps versus MEps [16]. *TP63* and *NFIA*
195 showed increased expression in old compared to young LEps (**Figure 2C and S3A**). Although
196 not all TFs showed an increase in expression with age, we did note that *TP63, NFIA, EGR1, and*
197 *TP73* have relatively higher expression in MEps as compared to LEps (**Figure 2A-B right panel**
198 **and S3D**). The expression patterns of TFs and the hypomethylation of their targets are

199 concordant with the known roles of these TFs. For example, young LEps do not express *TP63*,
200 which is a MEp lineage-specific transcription factor [18]. Furthermore, TEAD1 and TEAD4 are DNA
201 binding partners for YAP (Yes-associated protein), a transcriptional co-activator regulated by the
202 Hippo signaling pathway. YAP is sequestered in the cytoplasm through activity of kinases
203 downstream of the Hippo pathway [46]. Old LEps exhibit increased expression of *TP63* as well as
204 increased nuclear YAP localization [13,18], and thereby increased YAP activity and TEAD binding.
205 YAP is also known to bind TP73 and RUNX2 [46]. These results suggest that increased
206 expression/activity of myoepithelial-specific TFs in old LEps could lead to increased binding and
207 loss of DNA methylation in these regions (**Figure S3E**).

208 Conversely, DMRs with methylation gain are enriched for binding sites of the ETS family
209 of transcription factors EHF, ELF3, ELF5, ELF4, ETV1, ETS1, ETV4, GABPA, and NFI half and
210 full sites (**Figure 2D left panel**). TEs that gain methylation also showed enrichment of similar TF
211 binding sites in addition to the SOX family of TFs (**Figure 2E, left panel**). ELF5 is a luminal
212 lineage-specific transcription factor that exhibits decreased expression with age (**Figure 2F**) [18].
213 Besides *ELF5*, transcription factors *EHF* and *ELF3* show relatively higher expression in LEps as
214 compared to MEps (**Figure 2D and E right panel and 3D**). Expression of *SOX4* decreases,
215 whereas *SOX15* increases with age in LEps (**Figure S3B-C**), indicating loss of expression of
216 luminal lineage-specific TFs and increased expression of myoepithelial-specific TFs (**Figure**
217 **S3D**). These results further suggest that decreased expression of luminal lineage-specific TFs
218 may lead to a decrease in binding and protection of binding sites from DNA methyltransferases
219 (DNMTs). This, in turn, may result in the methylation and repression of these luminal lineage-
220 specific regions (**Figure S3E**). Generally, sites of methylation loss were enriched for
221 myoepithelial-specific TF binding motifs (**Figure 2G**), whereas sites of methylation gain were
222 enriched for luminal lineage-specific TF binding motifs (**Figure 2G**). Collectively, these results

223 suggest that aging-associated DNA methylation changes in LEps are orchestrated by lineage-
224 specific transcription factors.

225 Next, we determined which TE subfamilies contribute TF binding sites to the differentially
226 methylated regions (DMRs). Among the TEs that lose methylation, HERVH elements showed
227 the highest enrichment for TEAD and RUNX TF binding sites (**Figure 2H**). TP63 binding sites
228 were enriched at AluSx, AluSp, L2, and L3 elements. TP63 (p63) belongs to the p53 family of
229 TFs that includes TP53 (p53) and TP73 (p73). p53 binding sites occur in Alu elements and other
230 retrotransposons through deamination of a methylated CpG site [47,48]. p53 can repress
231 transposons through direct binding at these sites [49]. Since p63 and p73 share similar binding
232 sites as p53, it is possible that in mammary epithelial cells, these sites are repurposed as
233 regulatory elements through p63 binding. ETS family TF binding sites that gain methylation were
234 enriched in MIRb, MIR subfamilies of SINE elements and L2, L2a-c LINE subfamilies (**Figure**
235 **S3E**). SOX family TF binding sites were contributed by AluY, LTR10 and MER11D elements
236 (**Figure S3E**). Together, these results suggest that TEs have been co-opted by lineage-specific
237 TF binding sites and contribute to aging-dependent DNA methylation changes in LEps.

238

239 **Luminal lineage-specific regulatory elements gain methylation, whereas myoepithelial**
240 **regulatory elements lose methylation with age.**

241 Given that aging-associated DNA methylation changes occurred at sites containing
242 lineage-specific TF binding sites, we examined the functional relevance of these DMR regions
243 in the breast tissue. We predicted chromatin states at the DMRs using published histone
244 modification ChIP-seq datasets performed in mammary epithelial lineages [45]. To discern whether
245 DMRs reflect changes in cell properties, we predicted chromatin states in mature luminal, luminal
246 progenitor, and basal (MEp) cells from RM tissue of a young (19y old) donor [45]. We classified
247 DMRs into seven chromatin states in each cell type based on the extent of overlap with each

248 state (see Methods; **Figure S4A-C**). DMRs with methylation loss in old LEps overlap primarily
249 with MEp-specific promoter and enhancer regions, accounting for about 80% of all DMRs
250 (**Figure S4D**) and 74% of DMRs containing TEs (**Figure S4E**). As these sites also contain
251 binding sites for TFs enriched at MEp-specific active enhancers [45], these results indicate that
252 MEp-specific regulatory elements lose methylation in old LEps. Active enhancer sites are
253 hypomethylated [50,51] and methylation loss at MEp-specific regulatory elements might indicate an
254 epigenetic shift towards MEp gene expression patterns in LEps with age. Indeed, we do observe
255 an increase in expression of MEp genes in old luminal cells (**Figure S2A**) [16]. These data are in
256 line with previous studies suggesting that aging luminal cells acquire basal-like gene expression
257 signatures [12,13]. As regulatory elements such as promoters and enhancers are characterized by
258 accessible chromatin, we compared sites of chromatin accessibility using ATAC-seq in young
259 and old LEps to study changes in chromatin accessibility at DMRs. We found that open
260 chromatin peaks overlapping DMRs that lose methylation in the older LEps showed
261 corresponding increased accessibility in older LEps (**Figure S4F, J**). This result indicates loss
262 of methylation is associated with an increase in chromatin accessibility, consistent with these
263 sites containing regulatory elements [52].

264 Conversely, DMRs that gain methylation and overlap with putative promoter and
265 enhancer regions in luminal progenitor cells account for about 42% of all DMRs (**Figure S4G**)
266 and 43% of DMRs containing TEs (**Figure S4H**). Of note, our strategy of FACS enrichment of
267 luminal cells using PROM1 (CD133) positive and NGFR (CD271) negative cells (**Figure S1A**)
268 includes luminal progenitor cells [53]. As these DMR sites also contain binding sites for TFs
269 enriched at luminal lineage-specific active enhancers, these results indicate that luminal
270 progenitor-specific regulatory regions gain DNA methylation in aging LEps. Furthermore, open
271 chromatin peaks overlapping DMRs with methylation gain showed a decrease in chromatin
272 accessibility, suggesting repression of these regulatory elements (**Figure S4I, K**). Together,

273 these results suggest that aging LEps acquire permissive chromatin states at MEp-specific
274 regions and repress luminal progenitor-specific chromatin regions through changes in DNA
275 methylation. TEs with lineage-specific TF binding sites contribute to aging-dependent DNA
276 methylation changes at regulatory regions.
277
278 **Aging leads to methylation loss at young transposable elements.**
279 Aging is associated with increased transcriptional variability, cell-to-cell heterogeneity,
280 and DNA methylation variability [54-57]. Given that DNA methylation is an important mechanism for
281 silencing of evolutionarily young TEs [58], we asked whether increased variability in DNA
282 methylation with age leads to loss of methylation at young TEs. Since some amount of DNA
283 methylation variability is expected in human populations, we first identified regions of variable
284 methylation across all 10 individuals using the Systematic Interindividual Variation (SIV)
285 approach [59] (see Methods). We identified 23,921 regions that were variably methylated only in
286 old samples and 25,585 regions that were specific to the young samples (**Figure S5A**). Regions
287 variably methylated in old samples mainly showed methylation loss (17,780 of 23,921; 74.3%)
288 with some regions gaining methylation (6,141 of 23,921; 25.67%). The variable regions usually
289 were 150bp-450bp in length with some spanning 2-2.5kbs (**Figure S5B and S5C**). We next
290 examined whether variable regions contained TEs and found that 68.7% of regions variable in
291 young and 71.4% regions variable in old overlapped with TEs (**Figure S5D**). Among the TEs
292 that are variably methylated in old, about 76.9% lose methylation and 23.13% gain methylation
293 (**Figure S5E**).
294
295 We further investigated the characteristics of TEs that are variably methylated in old. The
296 evolutionarily younger elements (<100 million years) [60] tended towards DNA methylation loss,
297 whereas the older TE elements (>100 million years) tended to gain methylation in old cells

298 (**Figure S6A and Figure 3A)**. As some TE subfamilies, especially Alu elements, are much more
299 abundant than others, we determined the specific TE subfamilies that are enriched at these
300 regions using random controls. SINE elements were most abundant at variably methylated
301 regions and young SINEs showed variable methylation loss whereas older SINE elements
302 showed variable methylation gain in old LEps (**Figure 3B**). LINE elements with variable
303 methylation loss were younger than those that gained methylation (**Figure 3C**). Young LTR/ERV
304 elements showed methylation loss and gain (**Figure 3D**), indicating possible exaptation of these
305 elements in LEps. These results indicate that evolutionarily young TEs that are likely to be
306 regulated by DNA methylation can lose methylation with age and that this process could be
307 variable among individuals.

308 We further inspected the specific subfamilies enriched at the variably methylated regions.
309 While young Alu elements such as AluY and AluS are highly abundant in the genome, a few
310 AluS subfamilies were slightly enriched among the variably methylated regions (**Figure 3E**). The
311 TE subfamilies enriched at these regions were AluSq2, L1PA(2-6), HERVH, AluSg4, AluYc,
312 L1P1, L1P2, and MER11A elements (**Figure 3E**). HERVH elements are particularly interesting
313 as they were enriched at DMRs as well as variable regions of methylation loss (**Figure 1F and**
314 **3E-F**).

315

316 **Aging leads to stochastic methylation loss at transposable elements associated with**
317 **heterochromatin regions.**

318 We next examined the possible mechanisms of increased methylation variability at TEs
319 with age. Errors in maintenance methylation during replication has been proposed to be involved
320 in aging-dependent stochastic methylation changes [61]. Methylation loss in cancer cells occurs
321 at late replicating regions associated with heterochromatin due to inefficient maintenance
322 methylation by DNMT1 [62]. We therefore tested whether variably methylated regions are

323 associated with replication timing and heterochromatin. Given that replication time is known to
324 be conserved in normal and cancer cell lines [63], we utilized Repli-Seq data from MCF7 cells
325 available through the ENCODE project. We found that a higher proportion of variably methylated
326 regions in older LEps were at late replicating regions compared to variably methylated regions
327 in younger samples (**Figure 4A-B**). Similarly, more variably methylated TEs in older samples
328 were at late replicating regions as compared to those in younger samples (**Figure 4C-D**). We
329 also used Hi-C data from MCF10A cells to determine the A and B-type compartments in
330 mammary epithelial cells [64] that are associated with euchromatin and heterochromatin
331 respectively. A higher proportion of variably methylated regions in older samples (54.3%) were
332 in B-type compartments as compared to 47.7% in young (**Figure 4E, top panel**). Variably
333 methylated TEs in old also had a higher proportion (55.1%) in B-type compartments as
334 compared to 49.5% in young (**Figure 4E, bottom panel**). Among the variably methylated TEs
335 with methylation loss, we identified several TE subfamilies that had a higher proportion (> 55.1
336 %) of variably methylated copies in the B-compartment (**Figure S6B**) indicating their enrichment
337 at heterochromatin regions.

338 Since impaired DNA methylation maintenance at heterochromatin regions could lead to
339 discordant methylation patterns among adjacent CpGs, we examined whether there is an
340 increase in intra-sample methylation heterogeneity at variably methylated TEs in older
341 individuals. For this, we calculated methylation entropy, a measure of methylation heterogeneity
342 between adjacent CpGs, for 150 bp non-overlapping windows across the genome [65]. Variably
343 methylated regions identified in old individuals had a higher methylation entropy in B-type
344 compartments (**Figure 4F**). Similarly, variably methylated TEs in old individuals showed higher
345 methylation entropy in B-type compartments (**Figure 4G**). Variably methylated TEs in old
346 showed an increase in methylation heterogeneity in old samples as compared to young (**Figure**
347 **S6C and D**). We further examined whether increase in entropy was associated with methylation

348 loss at the variably methylated TEs enriched at B-type compartments. We found that indeed
349 variably methylated HERVH and MER52A loci showed methylation loss accompanied by
350 increase in entropy (**Figure 4H and I**). Together, these results indicate that evolutionarily young
351 CpG dense TEs exhibit stochastic loss of methylation with age.
352

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

355 TEs have latent regulatory potential and, when not repressed, can function as regulatory
356 elements. Given our observations that evolutionarily young, CpG dense transposons lose
357 methylation stochastically with age, we examined whether breast tumors show loss of epigenetic
358 suppression at these elements. We utilized a published dataset of 30 breast cancer methylomes
359 [66] to examine whether TEs that lose methylation with age are hypomethylated in breast cancer
360 and whether lack of methylation at these elements is also variable in breast cancer. Variable
361 DNA methylation loss at heterochromatic regions leads to formation of partially methylated
362 domains (PMDs) in breast cancer that are hyper-variable among individuals [66]. We therefore
363 reasoned that TEs at heterochromatin regions that exhibit variable methylation loss would be
364 enriched at PMDs in breast cancer.

365 For each copy in a TE subfamily, we examined the number of breast tumors in which it
366 overlapped with a PMD. The number of breast tumors (of 30) that contained a variably
367 methylated TE at a PMD was denoted as PMD frequency. There was a large variation in the
368 PMD frequency at variably methylated TEs (**Figure 5A**). However, MER52A, L1P1, L1PA3,
369 L1PA4 and HERVH elements were more likely to be at PMDs in multiple breast cancer samples
370 (**Figure 5A**). Loss of TE suppression leads to cryptic transcription and alternative isoforms due
371 to splicing with neighboring genes in cancer cells [67]. In order to investigate the potential of this
372 happening in breast cancer, we examined published RNA-seq data from high-grade Ductal

373 carcinoma in situ (DCIS) [68]. DCIS is considered to be a form of pre-breast cancer. We used the
374 SQuIRE pipeline [69] to measure gene and TE expression changes in DCIS as compared to
375 normal tissues. We found 197 age-hypomethylated TEs that showed higher expression in DCIS.
376 HERVH elements and L1PA2 elements were most common among the TEs that showed cryptic
377 transcription. A representative example of a HERVH element that loses methylation in aging
378 LEps and breast tumors is shown (**Figure 5B**). Cryptic transcription from this HERVH element
379 was observed only in DCIS samples (**Figure 5B**). We also found examples of exonization of
380 TEs; a MIRb into a non-coding RNA LOC1558357 expressed in DCIS (**Figure S7A)**. While
381 aberrant expression of age hypomethylated TEs in DCIS was clear, we did not observe any
382 subtype specificity in the expression of these TEs.

383 Next, we determined whether these TEs with methylation loss function as regulatory
384 elements in breast cancer. To this end we utilized ATAC-seq data from The Cancer Genome
385 Atlas (TCGA) breast cancer cohort (TCGA-BRCA) and published ChIP-seq datasets from breast
386 cancer cell lines [70]. We identified 1197 TCGA-BRCA ATAC-seq peaks that overlap TEs with
387 methylation loss in aging LEps (**Figure 5C**). The L1PA2 LINE subfamily was highly represented
388 in the ATAC-seq dataset. Most of the TEs had open chromatin sites across luminal tumors
389 especially in lobular carcinomas (**Figure 5C**). Some TEs were accessible in basal subtype
390 tumors as well. These results indicate that TEs hypomethylated with age could function as
391 regulatory elements in breast cancer. To further verify these results, we identified chromatin
392 states at hypomethylated L1PA2 elements from ChIP-seq data available from breast cancer cell
393 lines [70]. L1PA2 elements exist in a bivalent promoter state in luminal (3 of 4) and HER2-positive
394 (2 of 3) cell lines (**Figure 5D**), with the presence of both H3K4me3 and H3K9me3 marks (**Figure
395 5D and Figure 7B**). Since, H3K4me3 and H3K9me3 marks are mutually exclusive, they might
396 be occupying the same sites in different cells. These results indicate that some cancer cells that

397 lose repression at TEs from DNA methylation and H3K9me3 can utilize these elements as
398 regulatory regions to rewire transcriptomes.

399 To determine whether loss of methylation at TEs can lead to changes in expression of
400 genes important in breast cancer, we mapped variably methylated TEs to genes within 50kb and
401 checked their expression in DCIS samples [68] and across TCGA datasets. A representative
402 example of a variably methylated HERVH element upstream of *CYB5A* gene is shown (**Figure**
403 **5E**). We found increased expression of *CYB5A* gene in DCIS samples (**Figure 5F**) and is highly
404 expressed in luminal breast tumors (Luminal A and Luminal B) as compared to basal tumors
405 (**Figure 5G**). These results indicate that TEs hypomethylated with age can regulate the
406 expression of genes important for luminal breast cancer and thereby support transformation.

407

408 **Discussion**

409 We report a comprehensive WGBS analysis of the aging-associated DNA methylation
410 alterations in normal human mammary luminal epithelial cells. Previous studies were done using
411 array-based technologies that cover only a fraction of the CpG sites in the genome [16,18,20,21].
412 Repeat elements are under-represented among the arrayed probes [71], making interpretation of
413 DNA methylation at repetitive regions difficult. Moreover, earlier DNA methylation studies were
414 performed using composite tissues that represented a mixture of cells [20,21]. We and others
415 showed that different cell lineages have unique responses to aging, and this is dependent on
416 cell-cell interactions and microenvironment changes [18,72]. For example, aging leads to cell
417 composition changes in blood cells where older individuals show a skew towards myeloid
418 lineages [73]. Changes in cell composition can confound the interpretation of DNA methylation
419 levels in bulk tissues. Thus, it is imperative to study age-dependent molecular changes in purified
420 cell populations.

421 Previous DNA methylome studies in human cells have suggested a global loss of
422 methylation with age [7,65,74]. These studies were performed using CD4$^+$ T cells from a neonate
423 and a centenarian [7], and sorted CD4$^+$ T cells [65] and skin keratinocytes [74] from young (<25 y) and
424 old (>70 y) individuals. The global methylation loss with age was accompanied by a global
425 increase in entropy or methylation discordance with age [65]. We did not observe a global
426 methylation loss or a global increase in entropy in mammary luminal epithelial cells with age.
427 This could be due to higher fidelity of DNA methylation maintenance in luminal cells or due to a
428 smaller separation between the age of young (<30 years) and old (> 55 years) individuals in our
429 study. This observation is supported by DNA methylome studies in aging mouse livers, where
430 no global loss of methylation was observed, although specific changes at regulatory elements
431 were detected [75]. Whereas no global methylation loss was detected, we did, however, find
432 methylation changes at regulatory elements and variable methylation changes at young
433 transposons.

434 DNA methylation is an important mechanism for silencing of transposable elements.
435 Reactivation of transposable elements has been shown in cancer cells with global loss of DNA
436 methylation [25,26]. While previous research comparing DNA methylation levels between a neonate
437 and a centenarian reported that 12% of DMRs that lost methylation with age were located on
438 Alu elements [7], changes in epigenetic silencing at TEs with aging has not previously been
439 systematically investigated. Increased expression of retrotransposons and active
440 retrotransposition have been reported in aging mouse tissues using bulk quantitative strategies
441 [27,28], and, increased LINE-1 expression has been reported in senescent human fibroblasts *in*
442 *vitro* [29]. However, the impacts of altered DNA methylation patterning and TE derepression on
443 gene regulation in these systems was not investigated. We examined the impact of aging on
444 DNA methylation at TEs using non-senescent mammary epithelial cells. We show that the
445 evolutionary age of TEs is important for determining the response to aging. As older TEs can be

446 exapted and function as tissue-specific regulatory elements [24], aging-associated changes can
447 be attributed to cell-type-specific transcription factor activity. However, young CpG-dense
448 transposons are more likely to be silenced by DNA methylation and presumably lose methylation
449 by inefficient maintenance methylation by DNMT1. We indeed find that transposons with variable
450 methylation loss occur in late replicating regions associated with heterochromatin. This
451 observation is relevant in the context of proliferative capacity of breast epithelial cells. LEps have
452 significantly shorter telomere lengths as compared to isogenic MEps indicating that they are
453 among the most proliferative cells in breast epithelia [76,77]. Cancer cells are highly proliferative
454 and lose methylation at heterochromatin loci by a similar mechanism [62]. Taken together, it is
455 tempting to speculate that the LEp lineage is particularly susceptible to DNA methylation loss at
456 young TEs in addition to the loss of lineage fidelity in the aging breast tissue.

457 We further show that methylation loss at TEs with age is recapitulated in luminal breast
458 cancer. The derepressed TEs could lead to cryptic transcription or function as regulatory
459 elements in breast cancer cells. We also showed that TEs that lose silencing with age are
460 present as bivalent promoter loci in breast cancer cell lines, indicating loss of repression in
461 heterogenous populations of cells as opposed to being tightly regulated and completely silent in
462 normal cells. In addition, we showed that hypomethylated TEs could potentially regulate genes
463 associated with the luminal subtype of breast cancer. Altogether, these observations provide
464 further support to the observation that epigenetic changes with age could determine
465 susceptibility towards breast cancer. Retrotransposon activation has been widely observed in
466 nearly every human cancer, including breast cancer [78,79], and even in premalignant lesions [26].
467 This suggests that early events in tumorigenesis maybe driven by retrotransposon activity. We
468 show here, for the first time, that loss of transposon silencing is evident in aging luminal epithelial
469 cells that could predispose towards breast cancer. Future studies are needed to study the role
470 of these TEs in the early events of mammary cancer initiation.

471 **Materials and Methods**

472 **Isolation of mammary luminal epithelial cells**

473 Primary non-immortalized HMECs were generated and maintained, as described previously [80].
474 HMEC strains were established in the laboratory of Dr. Martha Stampfer (Lawrence Berkeley
475 National Laboratory) who has all required IRB approvals to distribute these samples. Cells were
476 grown in M87A medium with cholera toxin and oxytocin at 0.5 ng/ml and 0.1nM, respectively. 4th
477 passage HMECs were stained with anti-human CD133-PE (Biolegend, clone 7) and anti-human
478 CD271 (Biolegend, clone ME20.4) by following standard flow cytometry protocol. LEps and
479 MEps were enriched using Aria III (Becton Dickinson) flow sorter as CD133^{+}/CD271^{-} and CD133^{-}
480 /CD271^{+} fractions, respectively. The different HMEC strains used are described in Table S3.

481 **WGBS**

482 We examined the DNA methylome of primary LEps from younger (< 30y) premenopausal and
483 older (> 55y) postmenopausal women. Genomic DNA from FACS enriched LEps were extracted
484 using Quick-DNA Microprep kit (Zymo Research). Whole-genome bisulfite sequencing (WGBS-
485 seq) was performed at Admera Health Biopharma Services. Genomic DNA was bisulfite
486 converted using the Zymo EZ DNA methylation kit (Zymo Research). Library preparation was
487 done using TruSeq DNA Methylation Kit with 75ng of bisulfite converted DNA as per
488 manufacturer's recommendations. 150bp paired-end sequencing was performed on a single
489 lane of Illumina Hiseq X Ten. We obtained about 38X coverage per genome to an average of
490 414 million paired-end reads (Table S1).

491 Sequencing reads were hard trimmed 8 bp from both ends using Trim Galore
492 (https://github.com/FelixKrueger/TrimGalore) (version 0.4.5) to remove unwanted methylation
493 bias arising from the library preparation. Adapter contamination was also removed using Trim
494 Galore using --illumina option. Trimmed reads were aligned to the human reference genome
495 (hg19) using Bismark (version 0.19.1) [81] with the --bowtie1 option [82]. Bismark

496 (deduplicate_bismark) to remove PCR duplicates. Methpipe [83] (version 3.4.3) was used to
497 determine the methylation status of individual CpG sites. The hg19 reference genome contains
498 about 56.4 million CpG loci (hg19) or 28.2 million CpG dyads. Methylation levels were calculated
499 for individual CpG dyads and are henceforth referred to as CpG sites.

500 **DMR identification**

501 DMRfinder [84] was used to identify DMRs. Briefly, CpG sites with at least 4X coverage in at least
502 6 out of 10 samples were determined. Genomic regions with at least 4 CpG sites within a
503 distance of 100 bp were identified. The maximum length of regions was set to 500 bp. Of these
504 regions, ones that had less than 20 reads covering them were discarded. The final set of regions
505 was then used to calculate differentially methylated regions with greater than 20% change (*P*-
506 value < 0.05) between pooled young and old samples.

507 **Characterization of DMRs**

508 Overlaps with genomic features were calculated using bedtools intersect [85] with masked CpG
509 islands from the UCSC genome browser [86] and repeats from the RepeatMasker annotation [87].
510 DMR overlaps with CpG islands, and repeat elements were determined using bedtools intersect
511 such that the overlapping region contains at least 1 CpG site.

512 Kernel smoothed plots were made using the BSSeq package with default smoothing parameters
513 and CpG sites with greater than 3X coverage [88]. Heatmaps were made using the pheatmap
514 package [89].

515 **Identification of variably methylated regions**

516 Variably methylated regions were identified using the Systematic Interindividual Variation (SIV)
517 approach [59]. Briefly, the human genome was divided into 100 bp non-overlapping windows
518 containing at least 1 CpG site. About 14.4 million such windows were obtained. The coverage
519 and methylation values were calculated for each sample across these 100bp windows using
520 Methpipe. Windows were then filtered by coverage as follows to obtain informative bins:

521 Windows with CpG <= 2 that had a coverage of at least 10 reads and windows with CpG > 2
522 that had a coverage of at least 10 reads at half the number of CpG sites. We obtained about
523 12.9 million bins that had sufficient coverage. Next, we calculated the Pearson correlation co-
524 efficient for each bin with its adjacent bin. Adjacent bins that were within 100 bp of each other
525 and a correlation >= 0.71 were merged into correlated blocks of methylation. The average
526 methylation for each sample was calculated for the correlated blocks. Correlated blocks which
527 had an interindividual range (IR) of >= 20% and at least 4 CpG sites were defined as variable
528 regions. Old specific variable regions were defined as regions that had an IR of >= 20% across
529 the five old samples and IR_{old} -IR_{young} >= 20%. Similarly, young specific variable regions were
530 defined as regions that had an IR of >= 20% across the five young samples and IR_{young} -IR_{old} >=
531 20%. These regions were further filtered to retain those that have coverage of at least 3 CpGs
532 in all 10 samples. To categorize variably methylated regions into variable loss or variable gain,
533 the highest (old_{max}) and the lowest (old_{min}) methylation values among old samples were
534 compared to the methylation average in young (avg_y). The absolute values of differences in
535 methylation between avg_y and $old_{max,}$ old_{min} were compared. If the difference in methylation
536 between the avg_y and old_{min} were higher than that between avg_y and old_{max}, that region was
537 assigned as a variable region with methylation loss

538 **Chromatin state predictions**

539 ChIP-seq data for mammary epithelial populations were downloaded using the International
540 Human Epigenome Consortium data portal (https://epigenomesportal.ca/ihec/). Bigwig files
541 containing ChIP-seq signal for six different histone modifications in three cell types (sample IDs:
542 CEMT0154, CEMT0155, and CEMT0156) were used (**Figure S4A-C**). ChIP-seq signals from
543 the bigwig files were binned into 200 bp windows. These binned signal files were then binarized
544 using BinarizeSignal in ChromHMM (version 1.20) [90]. The binarized files were used to run
545 ChromHMM LearnModel with 16 states. Functionally similar chromatin states were grouped

546 together to seven distinct chromatin states (**Figure S4C**). ChIP-seq data for breast cancer cell
547 lines were downloaded from GEO (Accession number: GSE85158 [70]). Aligned bed files were
548 used to generate binarized files using BinarizeBed with default options. The binarized data were
549 used to run ChromHMM LearnModel with 18 states (**Figure S7B**). Functionally similar chromatin
550 states were grouped together to nine distinct chromatin states. DMR overlaps with chromatin
551 states were performed using bedtools intersect. In cases where a DMR overlapped more than
552 one chromatin state, DMR was assigned to the state with a higher base overlap.

553 **ATAC-seq and analysis**

554 50,000 live FACS sorted young (*n* = 4) and old (*n* = 4) LEps were used for Tn5 transposition in
555 duplicates using the Omni-ATAC method [91]. Libraries were sequenced to a depth of 40 million
556 paired reads each on an Illumina HiSeq 2500 system. Raw sequencing reads were trimmed
557 using Trim Galore (https://github.com/FelixKrueger/TrimGalore) (version 0.4.5) to remove
558 adapters. Trimmed reads were aligned to the human reference genome (hg19) using bowtie [82]
559 with the --best -k 1 -X 2000 --mm --chunkmbs 1024 parameters. Unmapped reads, mates as
560 well as low quality reads and secondary alignments were removed using samtools. PCR
561 duplicates were identified using Picard and removed using samtools. The reads were shifted by
562 4 bases on the sense strand and 5 bases on the anti-sense strand to account for Tn5 bias.
563 Reads from technical replicates were pooled to generate pseudoreplicates. Peaks were called
564 using MACS2 version 2.1.1.20160309 using -p 0.01 --nomodel --shift -37 --extsize 73 --SPMR -
565 -keep-dup all --call-summits as parameters. Peak calls from technical replicates and pseudo-
566 replicates were compared using IDR to get the most conservative set of peaks for each sample.

567 **RNA-seq analysis**

568 Total RNA was isolated using Quick-RNA Microprep kit (Zymo Research) from FACS sorted
569 young (*n* = 10) and old LEps (*n* = 8). Similarly, total RNA was isolated from young (*n* = 6) MEps.
570 PolyA-enriched RNA was isolated and used for library preparation using the TruSeq RNA Library

571 Prep kit (Illumina). About 50 million single-end reads of length 51 bp were generated for all
572 samples. Raw sequences were aligned to the hg19 reference genome using HISAT2 (version
573 2.1.0) [92] with the --dta parameter. Stringtie (version 1.3.4) [93] was used with default parameters
574 to assemble transcripts using the Gencode v19 transcript annotation. Assembled transcripts
575 from all libraries were further merged using --merge option in Stringtie. Merged transcript
576 abundances were measured using bedtools coverage.
577 For the old LEp versus young LEp comparison, edgeR (v3.26.4) [94] was used to conduct
578 differential gene expression analysis. Briefly, normalization factors were calculated using the
579 trimmed mean of M values method (TMM) to eliminate composition biases between libraries.
580 Genewise biological dispersion was estimated using edgeR's empirical Bayes approach. Then,
581 we employed a generalized linear model (GLM) approach to account for the unnormalized
582 distribution of gene counts, as well as additional factors, such as batch effects. Differentially
583 regulated genes were identified with an absolute log2 fold change $\geq$ 0.5 and FDR < 0.05. Gene
584 expression boxplots were generated using batch corrected RPKM values. Batch correction of
585 RPKM values was done using limma package [95]. For LEp versus MEp comparison, DESeq2 [96]
586 was used to normalize counts and identify differentially expressed genes (absolute log2 fold
587 change $\geq$ 0.5 and p-value < 0.05).
588 RNA-seq data from normal and DCIS samples were obtained from GEO (Accession number:
589 GSE69240 [68]). Fastq reads were quality and adapter trimmed using Trim Galore with the --paired
590 option. Quality trimmed reads were aligned to the hg19 genome using STAR (version 2.6.0.a) [97]
591 with --sjdbOverhang 100 --winAnchorMultimapNmax 100 --outFilterMultimapNmax 100 --
592 outFilterMismatchNmax 3 --alignEndsType EndToEnd options. The SQuIRE pipeline [69] was
593 used to measure gene and TE expression changes in DCIS as compared to normal tissues.
594 **Enrichment of DMRs and variably methylated regions at TE subfamilies**

595 The extent of overlap of DMRs and variable regions at repat subfamilies (RepeatMasker
596 annotation) was calculated using bedtools intersect. Randomly shuffled control regions of the
597 same length were generated 1000 times, and the expected number of overlaps calculated using
598 bedtools intersect. Enrichment values were calculated by dividing the number of DMR overlaps
599 (observed) with the average number of random overlaps (expected). *P* values were calculated
600 by dividing the number of extreme values (at either tail of distribution) from random overlaps by
601 1000. Significantly enriched repeat subfamilies had $p < 0.05$, enrichment > 1.5-fold, and > 5
602 copies overlapping DMRs or variable regions.

603 **Additional data analysis and external data**

604 Mappability of 100bp bisulfite converted kmers for the hg19 genome generated using Bismap [98]
605 were downloaded from the UCSC Genome Browser [86]. Begraph files were used to obtain the
606 mappability score of individual CpGs within TEs. Motif scanning was performed using fimo [99]
607 with JASPAR 2020 database with default parameters [44]. Motif enrichment analysis was
608 performed using findMotifsGenome.pl in homer [100]. For each set of genomic regions,
609 background sequences with matched GC content were selected. *P*-values for motif enrichment
610 was calculated using the cumulative binomial distributions. The evolutionary age of TEs was
611 calculated as described previously using the Jukes-Cantor correction [101]. Normalized
612 methylation entropy and mean methylation were calculated on 150bp non-overlapping windows
613 across the genome using informME [102]. Replication time for MC7 was obtained from ENCODE
614 data (GSE34399). A-type and B-type compartment classification for MCF10A cells were
615 downloaded from GSE66733 [64]. Normalized counts for TCGA-BRCA cancer specific ATAC-seq
616 peaks were downloaded from the Genomic Data Commons (GDC) Portal. The clinical
617 information and PAM50 subtypes for these samples were accessed using TCGAbiolinks [103].
618 Primary breast tumor WGBS data generated as part of the International Cancer Genome
619 Consortium [66] and was obtained from European Genome-phenome Archiver after DACO

620 approval. Figures were made in R (version 3.5.1) using the ggplot2 package [104] and assembled
621 in Adobe Illustrator.
622
623 **Availability of data and materials**
624 All raw and processed sequencing data files have been deposited at NCBI's Gene Expression
625 Omnibus (GEO) under accession GSE153696.
626
627 **Funding**
628 This work was supported by the National Institutes of Health, grants R01DK112041,
629 R01CA220693 (D.E.S.); K01DK104993 (A.L.); Department of Defense/Army Breast Cancer Era
630 of Hope Scholar Award BC141351, and Expansion Award BC181737, Conrad N. Hilton
631 Foundation, and City of Hope Center for Cancer and Aging (M.A.L.); the National Cancer
632 Institute (NCI) Cancer Metabolism Training Program Postdoctoral Fellowship T32CA221709
633 (R.W.S.) and City of Hope Program in Molecular and Cellular Biology pilot award (D.E.S and
634 M.A.L.). The research reported in this publication included work performed in the Integrative
635 Genomics and Analytical Cytometry Cores supported by the National Cancer Institute of the
636 National Institutes of Health under award number P30CA033572. The content is solely the
637 responsibility of the authors and does not necessarily represent the official views of the National
638 Institutes of Health. The funding bodies had no role in the design of the study and collection,
639 analysis, and interpretation of data and in writing the manuscript.
640
641 **Authors' contributions**
642 PS, MAL and DES designed the study. PS, MM and CT performed experiments. PS, RWS, MB
643 and AL analyzed and interpreted the data. PS, MAL and DES prepared the manuscript. All
644 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.

670 **References**

671 1 Lopez-Otin, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The hallmarks of aging.
672 *Cell* **153**, 1194-1217, doi:10.1016/j.cell.2013.05.039 (2013).
673 2 Booth, L. N. & Brunet, A. The Aging Epigenome. *Mol Cell* **62**, 728-744,
674 doi:10.1016/j.molcel.2016.05.013 (2016).
675 3 Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21,
676 doi:10.1101/gad.947102 (2002).
677 4 Maegawa, S. *et al.* Widespread and tissue specific age-related DNA methylation changes in
678 mice. *Genome Res* **20**, 332-340, doi:10.1101/gr.096826.109 (2010).
679 5 Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* **14**, R115,
680 doi:10.1186/gb-2013-14-10-r115 (2013).
681 6 Rakyan, V. K. *et al.* Human aging-associated DNA hypermethylation occurs preferentially at
682 bivalent chromatin domains. *Genome Res* **20**, 434-439, doi:10.1101/gr.103101.109 (2010).
683 7 Heyn, H. *et al.* Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S*
684 *A* **109**, 10522-10527, doi:10.1073/pnas.1120658109 (2012).
685 8 Jemal, A. *et al.* Cancer statistics, 2007. *CA Cancer J Clin* **57**, 43-66,
686 doi:10.3322/canjclin.57.1.43 (2007).
687 9 Jenkins, E. O. *et al.* Age-specific changes in intrinsic breast cancer subtypes: a focus on older
688 women. *Oncologist* **19**, 1076-1083, doi:10.1634/theoncologist.2014-0184 (2014).
689 10 Quong, J. *et al.* Age-dependent changes in breast cancer hormone receptors and oxidant stress
690 markers. *Breast Cancer Res Treat* **76**, 221-236, doi:10.1023/a:1020886801674 (2002).
691 11 Benz, C. C. Impact of aging on the biology of breast cancer. *Crit Rev Oncol Hematol* **66**, 65-74,
692 doi:10.1016/j.critrevonc.2007.09.001 (2008).
693 12 Garbe, J. C. *et al.* Accumulation of multipotent progenitors with a basal differentiation bias
694 during aging of human mammary epithelia. *Cancer Res* **72**, 3687-3701, doi:10.1158/0008-
695 5472.CAN-12-0157 (2012).
696 13 Pelissier, F. A. *et al.* Age-related dysfunction in mechanotransduction impairs differentiation of
697 human mammary epithelial progenitors. *Cell Rep* **7**, 1926-1939,
698 doi:10.1016/j.celrep.2014.05.021 (2014).
699 14 Pelissier Vatter, F. A. *et al.* High-Dimensional Phenotyping Identifies Age-Emergent Cells in
700 Human Mammary Epithelia. *Cell Rep* **23**, 1205-1219, doi:10.1016/j.celrep.2018.03.114 (2018).
701 15 Dos Santos, C. O., Dolzhenko, E., Hodges, E., Smith, A. D. & Hannon, G. J. An epigenetic
702 memory of pregnancy in the mouse mammary gland. *Cell Rep* **11**, 1102-1109,
703 doi:10.1016/j.celrep.2015.04.015 (2015).
704 16 Sayaman, R. W. *et al.* Epigenetic changes with age primes mammary luminal epithelia for cancer
705 initiation. *bioRxiv*, 2021.2002.2012.430777, doi:10.1101/2021.02.12.430777 (2021).
706 17 Prat, A. & Perou, C. M. Mammary development meets cancer genomics. *Nat Med* **15**, 842-844,
707 doi:10.1038/nm0809-842 (2009).
708 18 Miyano, M. *et al.* Age-related gene expression in luminal epithelial cells is driven by a
709 microenvironment made from myoepithelial cells. *Aging (Albany NY)* **9**, 2026-2051,
710 doi:10.18632/aging.101298 (2017).
711 19 Lee, J. K. *et al.* Age and the means of bypassing stasis influence the intrinsic subtype of
712 immortalized human mammary epithelial cells. *Front Cell Dev Biol* **3**, 13,
713 doi:10.3389/fcell.2015.00013 (2015).
714 20 Hofstatter, E. W. *et al.* Increased epigenetic age in normal breast tissue from luminal breast
715 cancer patients. *Clin Epigenetics* **10**, 112, doi:10.1186/s13148-018-0534-8 (2018).
716 21 Castle, J. R. *et al.* Estimating breast tissue-specific DNA methylation age using next-generation
717 sequencing data. *Clin Epigenetics* **12**, 45, doi:10.1186/s13148-020-00834-4 (2020).

718 22 Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nat*
719 *Rev Genet* **10**, 691-703, doi:10.1038/nrg2640 (2009).
720 23 Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biol* **19**,
721 199, doi:10.1186/s13059-018-1577-z (2018).
722 24 Sundaram, V. *et al.* Widespread contribution of transposable elements to the innovation of gene
723 regulatory networks. *Genome Res* **24**, 1963-1976, doi:10.1101/gr.168872.113 (2014).
724 25 Burns, K. H. Transposable elements in cancer. *Nat Rev Cancer* **17**, 415-424,
725 doi:10.1038/nrc.2017.35 (2017).
726 26 Ishak, C. A. & De Carvalho, D. D. Reactivation of Endogenous Retroelements in Cancer
727 Development and Therapy. *Annual Review of Cancer Biology* **4**.
728 27 De Cecco, M. *et al.* Transposable elements become active and mobile in the genomes of aging
729 mammalian somatic tissues. *Aging (Albany NY)* **5**, 867-883, doi:10.18632/aging.100621 (2013).
730 28 Benayoun, B. A. *et al.* Remodeling of epigenome and transcriptome landscapes with aging in
731 mice reveals widespread induction of inflammatory responses. *Genome Res* **29**, 697-709,
732 doi:10.1101/gr.240093.118 (2019).
733 29 De Cecco, M. *et al.* L1 drives IFN in senescent cells and promotes age-associated inflammation.
734 *Nature* **566**, 73-78, doi:10.1038/s41586-018-0784-9 (2019).
735 30 Ciccarone, F. *et al.* Age-dependent expression of DNMT1 and DNMT3B in PBMCs from a large
736 European population enrolled in the MARK-AGE study. *Aging Cell* **15**, 755-765,
737 doi:10.1111/acel.12485 (2016).
738 31 Beerman, I. *et al.* Proliferation-dependent alterations of the DNA methylation landscape underlie
739 hematopoietic stem cell aging. *Cell Stem Cell* **12**, 413-425, doi:10.1016/j.stem.2013.01.017
740 (2013).
741 32 Sun, D. *et al.* Epigenomic profiling of young and aged HSCs reveals concerted changes during
742 aging that reinforce self-renewal. *Cell Stem Cell* **14**, 673-688, doi:10.1016/j.stem.2014.03.002
743 (2014).
744 33 Teschendorff, A. E. *et al.* Age-dependent DNA methylation of genes that are suppressed in stem
745 cells is a hallmark of cancer. *Genome Res* **20**, 440-446, doi:10.1101/gr.103606.109 (2010).
746 34 Jung, M. & Pfeifer, G. P. Aging and DNA methylation. *BMC Biol* **13**, 7, doi:10.1186/s12915-
747 015-0118-4 (2015).
748 35 Pal, B. *et al.* Global changes in the mammary epigenome are induced by hormonal cues and
749 coordinated by Ezh2. *Cell Rep* **3**, 411-426, doi:10.1016/j.celrep.2012.12.020 (2013).
750 36 Michalak, E. M. *et al.* Polycomb group gene Ezh2 regulates mammary gland morphogenesis and
751 maintains the luminal progenitor pool. *Stem Cells* **31**, 1910-1920, doi:10.1002/stem.1437 (2013).
752 37 Xie, M. *et al.* DNA hypomethylation within specific transposable element families associates
753 with tissue-specific enhancer landscape. *Nat Genet* **45**, 836-841, doi:10.1038/ng.2649 (2013).
754 38 Pehrsson, E. C., Choudhary, M. N. K., Sundaram, V. & Wang, T. The epigenomic landscape of
755 transposable elements across normal human development and anatomy. *Nat Commun* **10**, 5640,
756 doi:10.1038/s41467-019-13555-x (2019).
757 39 Lu, X. *et al.* The retrovirus HERVH is a long noncoding RNA required for human embryonic
758 stem cell identity. *Nat Struct Mol Biol* **21**, 423-425, doi:10.1038/nsmb.2799 (2014).
759 40 Zhang, B. *et al.* Comparative DNA methylome analysis of endometrial carcinoma reveals
760 complex and distinct deregulation of cancer promoters and enhancers. *BMC Genomics* **15**, 868,
761 doi:10.1186/1471-2164-15-868 (2014).
762 41 Leung, D. *et al.* Integrative analysis of haplotype-resolved epigenomes across human tissues.
763 *Nature* **518**, 350-354, doi:10.1038/nature14217 (2015).
764 42 Swanson, E. C., Manning, B., Zhang, H. & Lawrence, J. B. Higher-order unfolding of satellite
765 heterochromatin is a consistent and early event in cell senescence. *J Cell Biol* **203**, 929-942,
766 doi:10.1083/jcb.201306073 (2013).

43 Nakagome, Y., Abe, T., Misawa, S., Takeshita, T. & Iinuma, K. The "loss" of centromeres from chromosomes of aged women. *Am J Hum Genet* **36**, 398-404 (1984).
44 Fornes, O. *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* **48**, D87-D92, doi:10.1093/nar/gkz1001 (2020).
45 Pellacani, D. *et al.* Analysis of Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and Associated Transcription Factor Networks. *Cell Rep* **17**, 2060-2074, doi:10.1016/j.celrep.2016.10.058 (2016).
46 Badouel, C., Garg, A. & McNeill, H. Herding Hippos: regulating growth in flies and man. *Curr Opin Cell Biol* **21**, 837-843, doi:10.1016/j.ceb.2009.09.010 (2009).
47 Zemojtel, T., Kielbasa, S. M., Arndt, P. F., Chung, H. R. & Vingron, M. Methylation and deamination of CpGs generate p53-binding sites on a genomic scale. *Trends Genet* **25**, 63-66, doi:10.1016/j.tig.2008.11.005 (2009).
48 Wang, T. *et al.* Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci U S A* **104**, 18613-18618, doi:10.1073/pnas.0703637104 (2007).
49 Tiwari, B. *et al.* p53 directly represses human LINE1 transposons. *Genes Dev* **34**, 1439-1451, doi:10.1101/gad.343186.120 (2020).
50 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-495, doi:10.1038/nature10716 (2011).
51 Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322, doi:10.1038/nature08514 (2009).
52 de la Rica, L. *et al.* TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol* **17**, 234, doi:10.1186/s13059-016-1096-8 (2016).
53 Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15**, 907-913, doi:10.1038/nm.2000 (2009).
54 Hernando-Herraez, I. *et al.* Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem cells. *Nat Commun* **10**, 4361, doi:10.1038/s41467-019-12293-4 (2019).
55 Bahar, R. *et al.* Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**, 1011-1014, doi:10.1038/nature04844 (2006).
56 Cheung, P. *et al.* Single-Cell Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging. *Cell* **173**, 1385-1397 e1314, doi:10.1016/j.cell.2018.03.079 (2018).
57 Martinez-Jimenez, C. P. *et al.* Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science* **355**, 1433-1436, doi:10.1126/science.aah4115 (2017).
58 Ohtani, H., Liu, M., Zhou, W., Liang, G. & Jones, P. A. Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses. *Genome Res* **28**, 1147-1157, doi:10.1101/gr.234229.118 (2018).
59 Gunasekara, C. J. *et al.* A genomic atlas of systemic interindividual epigenetic variation in humans. *Genome Biol* **20**, 105, doi:10.1186/s13059-019-1708-1 (2019).
60 Burns, K. H. & Boeke, J. D. Human transposon tectonics. *Cell* **149**, 740-752, doi:10.1016/j.cell.2012.04.019 (2012).
61 Issa, J. P. Aging and epigenetic drift: a vicious cycle. *J Clin Invest* **124**, 24-29, doi:10.1172/JCI69735 (2014).
62 Zhou, W. *et al.* DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat Genet* **50**, 591-602, doi:10.1038/s41588-018-0073-4 (2018).
63 Du, Q. *et al.* Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nat Commun* **10**, 416, doi:10.1038/s41467-019-08302-1 (2019).

64 Barutcu, A. R. *et al.* Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells. *Genome Biol* **16**, 214, doi:10.1186/s13059-015-0768-0 (2015).
65 Jenkinson, G., Pujadas, E., Goutsias, J. & Feinberg, A. P. Potential energy landscapes identify the information-theoretic nature of the epigenome. *Nat Genet* **49**, 719-729, doi:10.1038/ng.3811 (2017).
66 Brinkman, A. B. *et al.* Partially methylated domains are hypervariable in breast cancer and fuel widespread CpG island hypermethylation. *Nat Commun* **10**, 1749, doi:10.1038/s41467-019-09828-0 (2019).
67 Jang, H. S. *et al.* Transposable elements drive widespread expression of oncogenes in human cancers. *Nat Genet* **51**, 611-617, doi:10.1038/s41588-019-0373-3 (2019).
68 Abba, M. C. *et al.* A Molecular Portrait of High-Grade Ductal Carcinoma In Situ. *Cancer Res* **75**, 3980-3990, doi:10.1158/0008-5472.CAN-15-0506 (2015).
69 Yang, W. R., Ardeljan, D., Pacyna, C. N., Payer, L. M. & Burns, K. H. SQuIRE reveals locus-specific regulation of interspersed repeat expression. *Nucleic Acids Res* **47**, e27, doi:10.1093/nar/gky1301 (2019).
70 Franco, H. L. *et al.* Enhancer transcription reveals subtype-specific gene expression programs controlling breast cancer pathogenesis. *Genome Res* **28**, 159-170, doi:10.1101/gr.226019.117 (2018).
71 Bell, C. G. *et al.* DNA methylation aging clocks: challenges and recommendations. *Genome Biol* **20**, 249, doi:10.1186/s13059-019-1824-y (2019).
72 Kimmel, J. C. *et al.* Murine single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging. *Genome Res* **29**, 2088-2103, doi:10.1101/gr.253880.119 (2019).
73 Rimmele, P. *et al.* Aging-like phenotype and defective lineage specification in SIRT1-deleted hematopoietic stem and progenitor cells. *Stem Cell Reports* **3**, 44-59, doi:10.1016/j.stemcr.2014.04.015 (2014).
74 Vandiver, A. R. *et al.* Age and sun exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. *Genome Biol* **16**, 80, doi:10.1186/s13059-015-0644-y (2015).
75 Cole, J. J. *et al.* Diverse interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at critical gene regulatory regions. *Genome Biol* **18**, 58, doi:10.1186/s13059-017-1185-3 (2017).
76 Meeker, A. K. *et al.* Telomere shortening occurs in subsets of normal breast epithelium as well as in situ and invasive carcinoma. *Am J Pathol* **164**, 925-935, doi:10.1016/S0002-9440(10)63180-X (2004).
77 Sputova, K. *et al.* Aging phenotypes in cultured normal human mammary epithelial cells are correlated with decreased telomerase activity independent of telomere length. *Genome Integr* **4**, 4, doi:10.1186/2041-9414-4-4 (2013).
78 Rodic, N. *et al.* Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am J Pathol* **184**, 1280-1286, doi:10.1016/j.ajpath.2014.01.007 (2014).
79 Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48-61, doi:10.1016/j.cell.2014.12.033 (2015).
80 Labarge, M. A., Garbe, J. C. & Stampfer, M. R. Processing of human reduction mammoplasty and mastectomy tissues for cell culture. *J Vis Exp*, doi:10.3791/50011 (2013).
81 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572, doi:10.1093/bioinformatics/btr167 (2011).

82 Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).
83 Song, Q. *et al.* A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. *PLoS One* **8**, e81148, doi:10.1371/journal.pone.0081148 (2013).
84 Gaspar, J. M. & Hart, R. P. DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. *BMC Bioinformatics* **18**, 528, doi:10.1186/s12859-017-1909-0 (2017).
85 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
86 Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006, doi:10.1101/gr.229102 (2002).
87 Smit, A., Hubley, R & Green, P. . RepeatMasker Open-4.0. http://www.repeatmasker.org. (2013-2015).
88 Hansen, K. D., Langmead, B. & Irizarry, R. A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* **13**, R83, doi:10.1186/gb-2012-13-10-r83 (2012).
89 Kolde, R. (2018).
90 Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12**, 2478-2492, doi:10.1038/nprot.2017.124 (2017).
91 Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat Methods* **14**, 959-962, doi:10.1038/nmeth.4396 (2017).
92 Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods* **12**, 357-360, doi:10.1038/nmeth.3317 (2015).
93 Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**, 290-295, doi:10.1038/nbt.3122 (2015).
94 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2010).
95 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47, doi:10.1093/nar/gkv007 (2015).
96 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
97 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
98 Karimzadeh, M., Ernst, C., Kundaje, A. & Hoffman, M. M. Umap and Bismap: quantifying genome and methylome mappability. *Nucleic Acids Res* **46**, e120, doi:10.1093/nar/gky677 (2018).
99 Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018, doi:10.1093/bioinformatics/btr064 (2011).
100 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
101 Bourque, G. *et al.* Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* **18**, 1752-1762, doi:10.1101/gr.080663.108 (2008).
102 Jenkinson, G., Abante, J., Feinberg, A. P. & Goutsias, J. An information-theoretic approach to the modeling and analysis of whole-genome bisulfite sequencing data. *BMC Bioinformatics* **19**, 87, doi:10.1186/s12859-018-2086-5 (2018).
103 Colaprico, A. *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res* **44**, e71, doi:10.1093/nar/gkv1507 (2016).

104 Wickham, H. *ggplot2: elegant graphics for data analysis*. (Springer, 2016).

Figure Legends

Figure 1: Luminal cells from old 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) TE subfamilies enriched in DMRs with methylation loss. Left panel shows number of elements from each subfamily that lose methylation. Right panel shows the observed/expected ratio for each TE subfamily. Expected number of overlaps in a background set of regions with matched length and CpG density was used to calculate the ratio. (D) Genome browser tracks showing a representative example of a HERVH-int element losing methylation in old samples. Bottom tracks are from ATAC-seq data showing gain of accessibility at the DMR.

Figure 2: Transposable elements contribute to aging induced DNA methylation changes at regulatory regions. (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) Same as A, but showing motifs enriched in DMRs overlapping transposable elements. Only the portion overlapping a TE was used for motif enrichment analysis. (C) 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). (D) 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. (E) Same as D, but showing motifs enriched in DMRs overlapping transposable elements. Only the portion

overlapping TEs was used for motif enrichment analysis. (F) 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). (G) 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. (H) Enrichment of TF binding site in each TE subfamily with methylation loss. Color scale is –log10(p-value) of TF motif enrichment.

Figure 3: Variably methylated regions with methylation loss are enriched at young 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 DMRs with 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 4: Stochastic methylation loss at transposons associated with heterochromatin regions. **(**A, C) Density plots showing replication time of (A) all variably methylated regions and (C) variably methylated TEs in young and old using MCF7 Repli-Seq data. Vertical dashed lines indicate late (weighted average < 17) or early (weighted average < 73) cutoffs. (B) Proportion of variably methylated regions in young and old samples that are at late or early replicating region. (D) Same as B but showing proportion of variably methylated TEs. For B and D, asterisks indicate enrichment (p-value < 0.0001, Fisher's exact test) (E) Percent of variable regions in old and young cells (top panel) that are in A or B type compartments in mammary epithelial cells. Bottom panel shows percent of variable TEs in old and young cells that are in A or B type compartments. Published MCF10A HiC data was used for A and B-

type prediction. Asterisks indicate enrichment (p-value < 0.0001, Fisher's exact test) (F) Plots showing methylation entropy, a measure of methylation heterogeneity at variably methylated regions and (G) variably methylated TEs in A and B-type compartments. F and G. *p*-values were calculated using Kolmogorov-Smirnov (KS) test for equality of distributions. (H) Scatterplots of DNA methylation and entropy levels of variably methylated HERVH and (I) MER52A elements that lose methylation. Marginal density plots are shown to display the density of points on the X and Y axes.

Figure 5: Transposable elements with age-dependent methylation loss are activated in breast cancer. (A) Frequency of overlap with partially methylated domains at variably methylated TEs among 30 primary breast cancer samples. (B) 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. (C) Heatmap showing signals across 74 TCGA-BRCA samples for ATAC-seq peaks that overlap TEs hypomethylated with age. (D) Heatmaps showing chromatin states identified at hypomethylated L1PA2 elements in normal cells, transformed mammary cell lines and breast cancer cell lines. Each column shows a hypomethylated L1PA2 element in aging LEps. (E) Genome browser tracks showing methylation loss at a HERVH element near the *CYB5A* gene promoter. (F) Expression of *CYB5A* in normal and ductal carcinoma in situ (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 (p-value < 2.2x10-16; using two-tailed Wilcoxon signed-rank test).

Figures

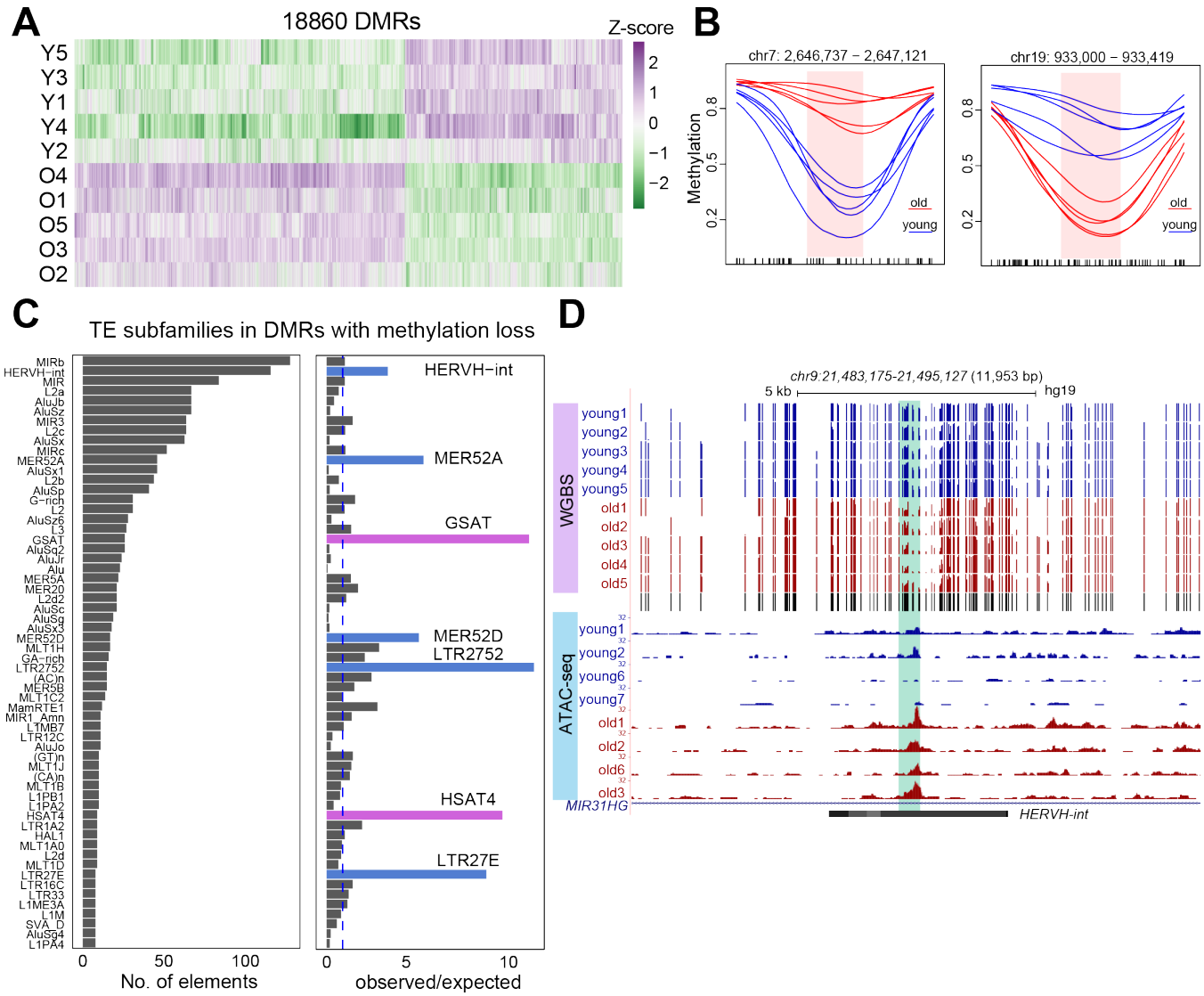

Figure 1: Luminal cells from old individuals show DNA methylation changes at distinct genomic loci. (A) Methylation levels of identified DMRs across all samples. Samples are in rows, and DMRs are in columns. The color intensity represents the Z-score scaled from the mean. Values ranging from higher to 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) TE subfamilies enriched in DMRs with methylation loss. Left panel shows number of elements from each subfamily that lose methylation. Right panel shows the observed/expected ratio for each TE subfamily. Expected number of overlaps in a background set of regions with matched length and CpG density was used to calculate the ratio. (D) Genome browser tracks showing a representative example of a HERVH-int element losing methylation in old samples. Bottom tracks are from ATAC-seq data showing gain of accessibility at the DMR.

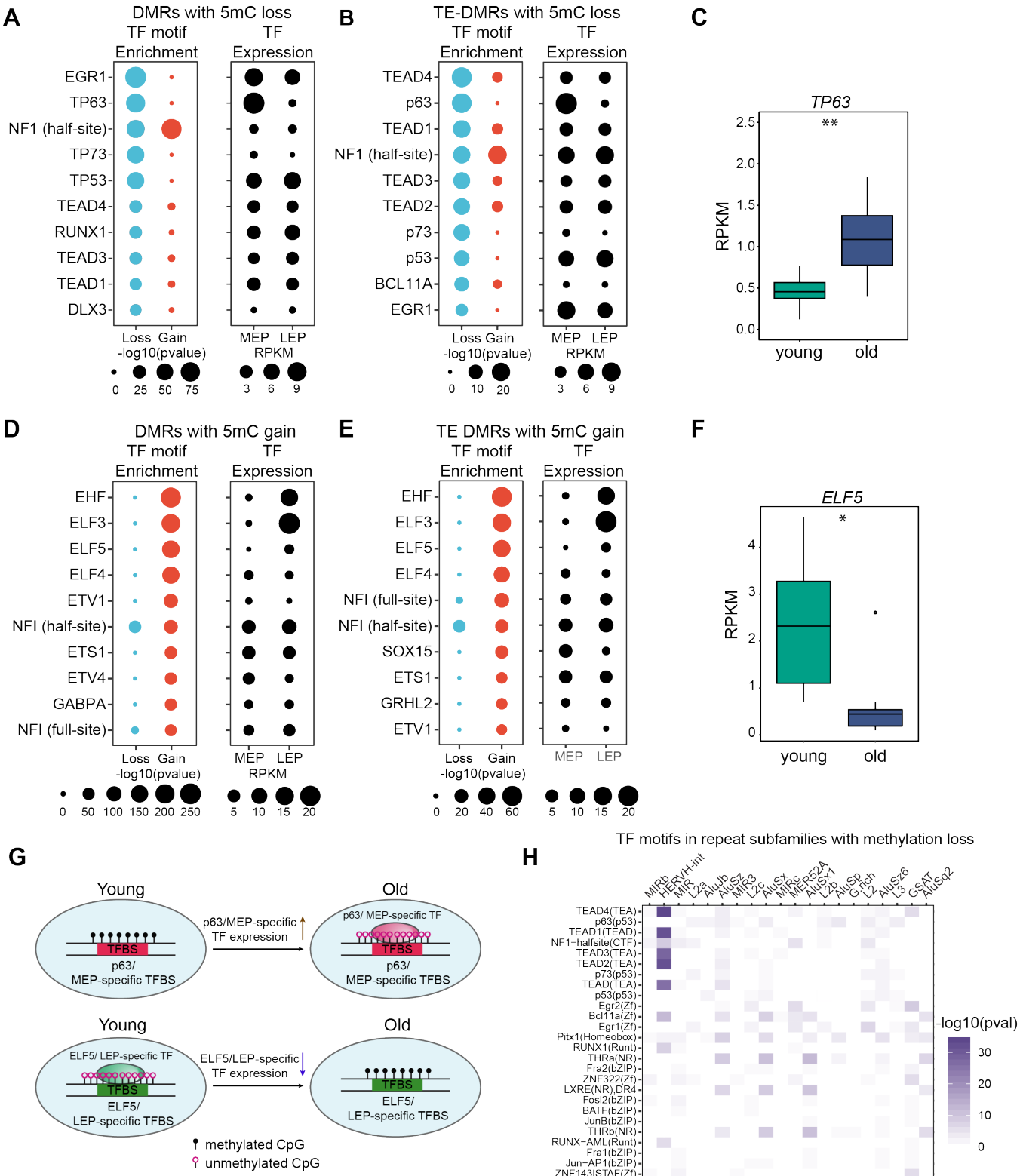
A
DMRs with 5mC loss
TF motif Enrichment
TF Expression
EGR1
TP63
NF1 (half-site)
TP73
TP53
TEAD4
RUNX1
TEAD3
TEAD1
DLX3
Loss Gain
-log10(pvalue)
0 25 50 75
MEP LEP
RPKM
3 6 9
B
TE-DMRs with 5mC loss
TF motif Enrichment
TF Expression
TEAD4
p63
TEAD1
NF1 (half-site)
TEAD3
TEAD2
p73
p53
BCL11A
EGR1
Loss Gain
-log10(pvalue)
0 10 20
MEP LEP
RPKM
3 6 9
C
TP63
**
RPKM
2.5
2.0
1.5
1.0
0.5
0.0
young
old
D
DMRs with 5mC gain
TF motif Enrichment
TF Expression
EHF
ELF3
ELF5
ELF4
ETV1
NFI (half-site)
ETS1
ETV4
GABPA
NFI (full-site)
Loss Gain
-log10(pvalue)
0 50 100 150 200 250
MEP LEP
RPKM
5 10 15 20
E
TE DMRs with 5mC gain
TF motif Enrichment
TF Expression
EHF
ELF3
ELF5
ELF4
NFI (full-site)
NFI (half-site)
SOX15
ETS1
GRHL2
ETV1
Loss Gain
-log10(pvalue)
0 20 40 60
MEP LEP
5 10 15 20
F
ELF5
*
RPKM
4
3
2
1
0
young
old
G
Young
TFBS
p63/
MEP-specific TFBS
p63/MEP-specific
TF expression
Old
p63/ MEP-specific TF
TFBS
p63/
MEP-specific TFBS
Young
ELF5/ LEP-specific TF
TFBS
ELF5/
LEP-specific TFBS
ELF5/LEP-specific
TF expression
Old
TFBS
ELF5/
LEP-specific TFBS
methylated CpG
unmethylated CpG
H
TF motifs in repeat subfamilies with methylation loss
MIRb
HERVH-int
MIR
L2a
AluJb
AluSz
MIR3
L2c
AluSx
MIRc
MER52A
AluSx1
L2b
AluSp
G1ich
L2
AluSz6
L3
GSAT
AluSq2
TEAD4(TEA)
p63(p53)
TEAD1(TEAD)
NF1−halfsite(CTF)
TEAD3(TEA)
TEAD2(TEA)
p73(p53)
TEAD(TEA)
p53(p53)
Egr2(Zf)
Bcl11a(Zf)
Egr1(Zf)
Pitx1(Homeobox)
RUNX1(Runt)
THRa(NR)
Fra2(bZIP)
ZNF322(Zf)
LXRE(NR),DR4
Fosl2(bZIP)
BATF(bZIP)
JunB(bZIP)
THRb(NR)
RUNX−AML(Runt)
Fra1(bZIP)
Jun−AP1(bZIP)
ZNF143|STAF(Zf)
Atf3(bZIP)
GLI3(Zf)
Mef2a(MADS)
AP−1(bZIP)
Dlx3(Homeobox)
NF1(CTF)
VDR(NR),DR3
Nkx3.1(Homeobox)
CEBP(bZIP)
ZSCAN22(Zf)
Sox4(HMG)
EAR2(NR)
-log10(pval)
30
20
10
0

Figure 2: Transposable elements contribute to aging induced DNA methylation changes at regulatory regions. (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) Same as A, but showing motifs enriched in DMRs overlapping transposable elements. Only the portion overlapping a TE was used for motif enrichment analysis. (C) 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). (D) 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. (E) Same as D, but showing motifs enriched in DMRs overlapping transposable elements. Only the portion overlapping TEs was used for motif enrichment analysis. (F) 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). (G) 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. (H) Enrichment of TF binding site in each TE subfamily with methylation loss. Color scale is –log10(p-value) of TF motif enrichment.

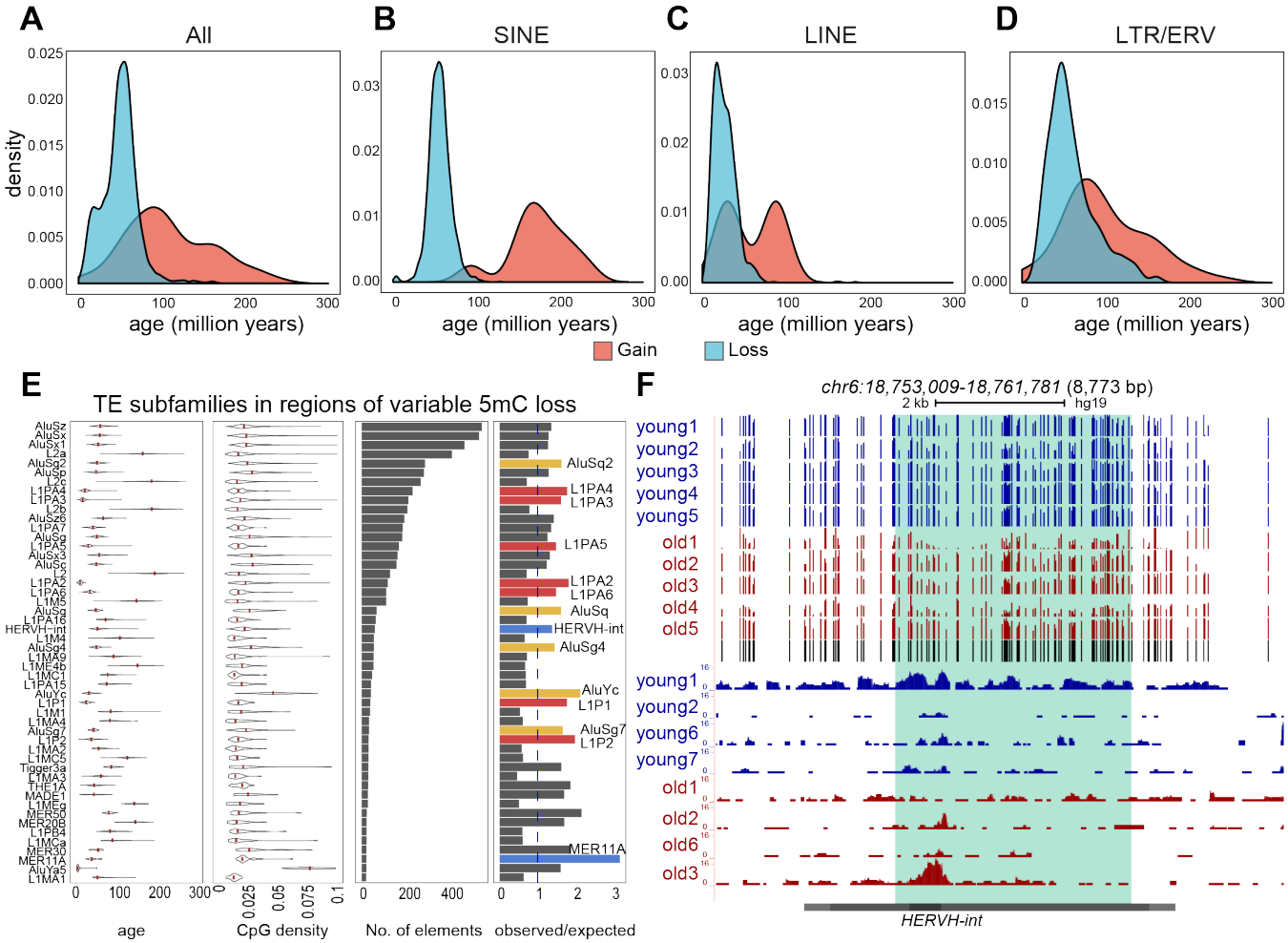

Figure 3: Variably methylated regions with methylation loss are enriched at young transposable elements (A-D) Evolutionary age of (A) all TE elements (B) SINE, (C) LINE, and (D) LTR/ERV elements that are variably methylated with methylation loss or gain. Evolutionary ages of only the significantly enriched TE subfamilies are plotted. (E) TE subfamilies enriched in DMRs with 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.

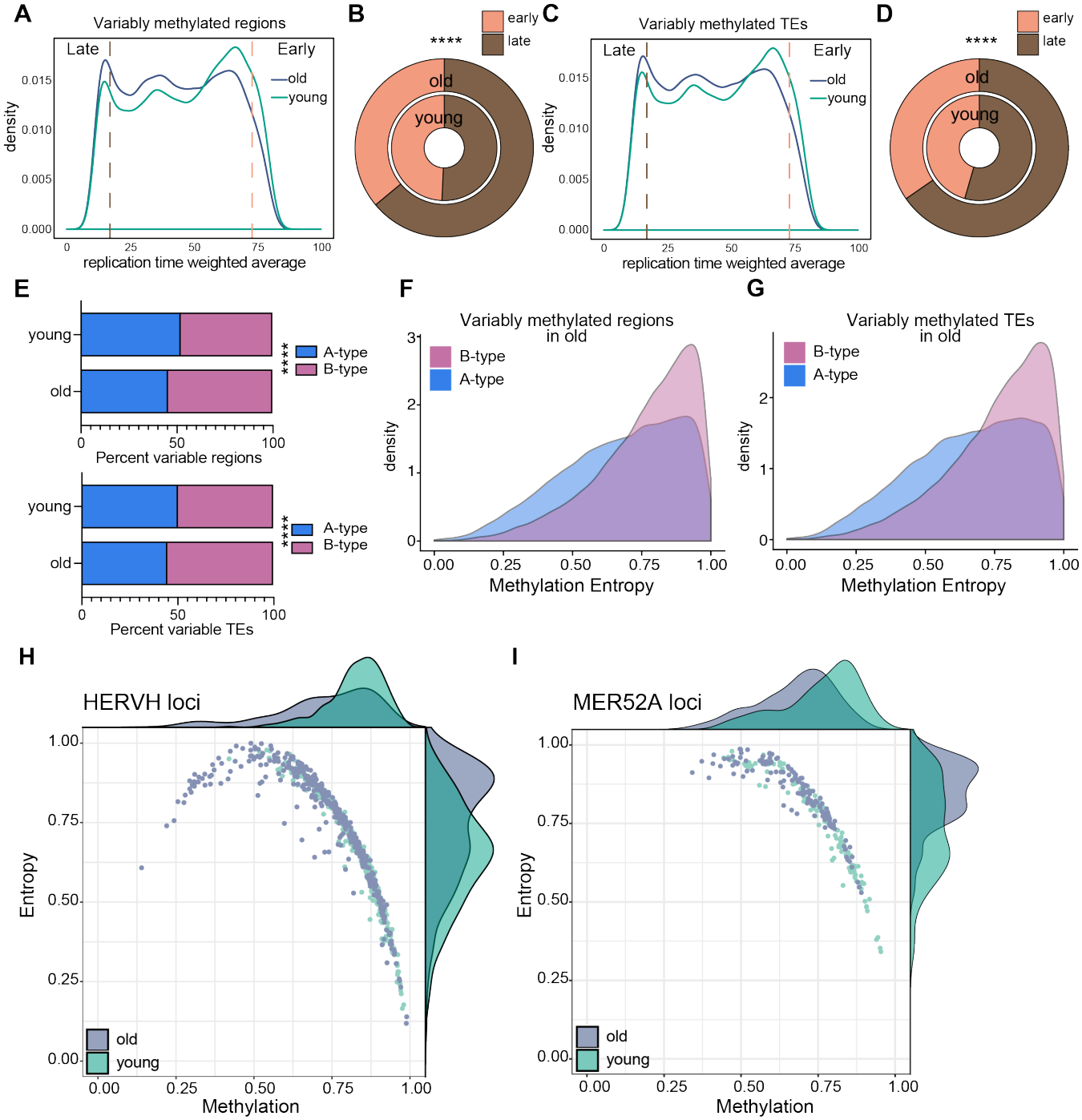

Figure 4: Stochastic methylation loss at transposons associated with heterochromatin regions. (A, C) Density plots showing replication time of (A) all variably methylated regions and (C) variably methylated TEs in young and old using MCF7 Repli-Seq data. Vertical dashed lines indicate late (weighted average < 17) or early (weighted average < 73) cutoffs. (B) Proportion of variably methylated regions in young and old samples that are at late or early replicating region. (D) Same as B but showing proportion of variably methylated TEs. For B and D, asterisks indicate enrichment (p-value < 0.0001, Fisher's exact test) (E) Percent of variable regions in old and young cells (top panel) that are in A or B

type compartments in mammary epithelial cells. Bottom panel shows percent of variable TEs in old and young cells that are in A or B type compartments. Published MCF10A HiC data was used for A and B-type prediction. Asterisks indicate enrichment (p-value < 0.0001, Fisher's exact test) (F) Plots showing methylation entropy, a measure of methylation heterogeneity at variably methylated regions and (G) variably methylated TEs in A and B-type compartments. F and G. *p*-values were calculated using Kolmogorov-Smirnov (KS) test for equality of distributions. (H) Scatterplots of DNA methylation and entropy levels of variably methylated HERVH and (I) MER52A elements that lose methylation. Marginal density plots are shown to display the density of points on the X and Y axes.

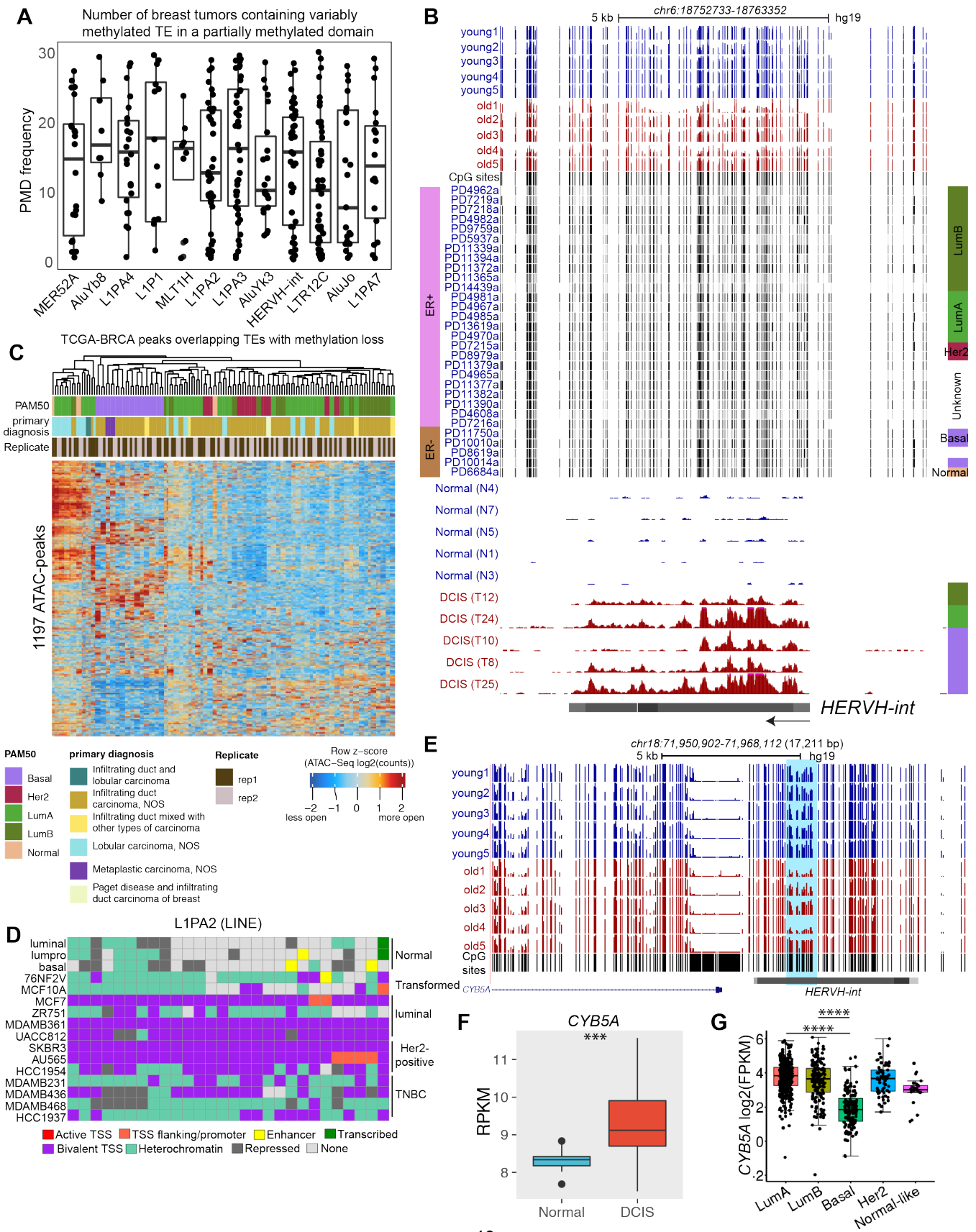
A
Number of breast tumors containing variably methylated TE in a partially methylated domain
PMD frequency
MER52A
AluYb8
L1PA4
L1P1
MLT1H
L1PA2
L1PA3
AluYk3
HERVH-int
LTR12C
AluJo
L1PA7
B
chr6:18752733-18763352
5 kb
hg19
young1
young2
young3
young4
young5
old1
old2
old3
old4
old5
CpG sites
ER+
ER-
LumB
LumA
Her2
Unknown
Basal
Normal
Normal (N4)
Normal (N7)
Normal (N5)
Normal (N1)
Normal (N3)
DCIS (T12)
DCIS (T24)
DCIS(T10)
DCIS (T8)
DCIS (T25)
HERVH-int
C
TCGA-BRCA peaks overlapping TEs with methylation loss
PAM50
primary diagnosis
Replicate
1197 ATAC-peaks
PAM50
Basal
Her2
LumA
LumB
Normal
primary diagnosis
Infiltrating duct and lobular carcinoma
Infiltrating duct carcinoma, NOS
Infiltrating duct mixed with other types of carcinoma
Lobular carcinoma, NOS
Metaplastic carcinoma, NOS
Paget disease and infiltrating duct carcinoma of breast
Replicate
rep1
rep2
Row z-score (ATAC-Seq log2(counts))
−2 −1 0 1 2
less open
more open
E
chr18:71,950,902-71,968,112 (17,211 bp)
5 kb
hg19
young1
young2
young3
young4
young5
old1
old2
old3
old4
old5
CpG sites
CYB5A
HERVH-int
D
L1PA2 (LINE)
luminal
lumpro
basal
76NF2V
MCF10A
MCF7
ZR751
MDAMB361
UACC812
SKBR3
AU565
HCC1954
MDAMB231
MDAMB436
MDAMB468
HCC1937
Normal
Transformed
luminal
Her2-positive
TNBC
Active TSS
TSS flanking/promoter
Enhancer
Transcribed
Bivalent TSS
Heterochromatin
Repressed
None
F
CYB5A
***
RPKM
Normal
DCIS
G
****
****
CYB5A log2(FPKM)
LumA
LumB
Basal
Her2
Normal-like

Figure 5: Transposable elements with age-dependent methylation loss are activated in breast cancer. (A) Frequency of overlap with partially methylated domains at variably methylated TEs among 30 primary breast cancer samples. (B) 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. (C) Heatmap showing signals across 74 TCGA-BRCA samples for ATAC-seq peaks that overlap TEs hypomethylated with age. (D) Heatmaps showing chromatin states identified at hypomethylated L1PA2 elements in normal cells, transformed mammary cell lines and breast cancer cell lines. Each column shows a hypomethylated L1PA2 element in aging LEps. (E) Genome browser tracks showing methylation loss at a HERVH element near the *CYB5A* gene promoter. (F) Expression of *CYB5A* in normal and ductal carcinoma in situ (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 (p-value < 2.2x10-16; using two-tailed Wilcoxon signed-rank test).