Normal mammary epithelia accumulate DNA methylation changes with age that are consistent with breast cancer methylomes

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

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

Running Title: DNA methylome characteristics of aging mammary luminal cells

Significance: Mammary luminal epithelial cells lose lineage-specific expression with age and accumulate stochastic methylation changes that lead to loss of silencing at transposable elements. These age-dependent events potentially promote breast cancer susceptibility.

Abstract

Luminal epithelial cells (LEps) are a key cell lineage implicated in age-related luminal breast cancers. Alterations to the epigenome are a hallmark of aging cells. However, the extent of age-associated DNA methylation alterations in LEps, and the corresponding functional consequences of these alterations, have remained unclear. We report here that aging leads to distinct methylation changes in LEps. Luminal lineage-specific genes gain promoter methylation, whereas myoepithelial-specific genes lose promoter methylation. Regulatory elements display methylation changes at lineage-specific TF binding sites consistent with the loss of lineage fidelity. CpG islands (CGIs) and transposable elements (TEs) have stochastic methylation gain and loss, respectively. PRC2 target genes that are hypermethylated in luminal breast cancer exhibit stochastic methylation increase with age. TEs with stochastic methylation loss are activated in breast cancer and potentially function as regulatory elements contributing to the loss of lineage fidelity with age. Each of these classes of methylation changes impact the regulation of genes associated with luminal breast cancer. Altogether, our results indicate that aging leads to DNA methylation changes that could determine breast cancer susceptibility.

Introduction

Over 75% of breast cancer diagnoses are in women over the age of 50, making age the most significant risk factor for breast cancer (1). This is especially true for luminal subtypes of all age-related breast cancers (2). The mammary epithelium is composed of two principal lineages, myoepithelial cells (MEps) and luminal epithelial cells (LEps), with the latter being the chief suspect cell of origin for luminal breast cancers (3,4). Most breast cancers are sporadic with no known inherited mutation or family history. A better understanding of how age-dependent molecular changes contribute to breast cancer risk is critical for designing better therapies and prevention strategies.

Aging in the breast tissue is associated with significant physiological changes that lead to changes in molecular properties and the proportion of the different cell populations with age (5-8). Human mammary epithelial cells (HMECs) from older post-menopausal women, when transformed *in vitro*, exhibit gene expression patterns characteristic of luminal breast cancer subtypes. In contrast, transformed HMECs from younger pre-menopausal women show more properties of basal breast cancers (9). Striking age-dependent changes in abundance and phenotype occur in LEps, which gain expression of characteristic MEp genes (6,10,11) and signaling molecules consistent with those of immortalized luminal cells (10), suggesting that age-associated epigenetic changes in luminal cells are directly related to elevated cancer risk.

DNA methylation is a relatively stable component of the epigenome that gets established in early development and is maintained through successive cycles of cell division to preserve cell identity (12). DNA methylation patterns, however, are known to change with age (13-16) and DNA methylation levels of specific CpG sites show a significant correlation with chronological age (14). This ‘epigenetic clock’ can be used to predict the biological age of diverse human tissues (14). However, this pan-tissue epigenetic clock is “poorly calibrated” in breast tissue, even for women without cancer, possibly due to physiological factors such as changes in

76 hormonal status with age (14). Previous studies examining DNA methylation changes in breast
77 tissue with age were either performed in bulk mammary tissue with heterogeneous cell
78 populations or interrogated only a subset of CpGs in the genome (17,18). A complete
79 understanding of age-associated DNA methylation changes in human mammary LEps and how
80 it relates to age-related breast cancer has been lacking.

81 We performed whole-genome bisulfite sequencing (WGBS) and Assay for Transposase-
82 Accessible Chromatin (ATAC-seq) in non-senescent primary human mammary luminal epithelial
83 cells from pre-menopausal (young) and post-menopausal (old) women. We find that LEps from
84 older women display DNA methylation changes at regulatory elements associated with altered
85 expression and activity of lineage-specific transcription factors such as ELF5 and TP63. We
86 used chromatin state predictions from mammary epithelial cell lineages to show that MEp
87 lineage-specific regulatory elements lose methylation, whereas luminal-specific regions gain
88 methylation. Additionally, PRC2 (polycomb repressive complex 2) regulated genes gain DNA
89 methylation at their promoters with age, consistent with these genes having methylated
90 promoters in luminal breast cancer (19). We also observed increased methylation disorder with
91 age, resulting in stochastic methylation gain at CpG islands (CGIs) and loss at evolutionarily
92 recent transposable elements (TEs). Using breast cancer WGBS data, we show that TEs that
93 lose methylation with age were also hypomethylated in breast cancers. In addition, we used
94 ATAC-seq and RNA-seq data from the TCGA-BRCA cohort and RNA-seq done on DCIS (ductal
95 carcinoma in situ) samples to show that TEs with age-dependent methylation loss are activated
96 in breast cancer and can function as regulatory elements for nearby genes. Overall, our results
97 suggest that aging mammary luminal epithelial cells undergo DNA methylation alterations at
98 regulatory elements, CGIs, and evolutionarily recent TEs, similar to methylation changes
99 observed in luminal breast cancer.

100

101 **Results**

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

103 We performed whole-genome bisulfite sequencing (WGBS-seq) on human mammary LEps
104 from younger (<30y) and older (>55y) women (**Figure S1A, Table S1**). The age groups were
105 chosen to represent samples from women who were clearly pre-menopausal or post-
106 menopausal to avoid confounding issues of peri-menopause. We previously demonstrated that
107 LEps from these primary cultures maintain gene expression and DNA methylation profiles that
108 mirror those in tissue (8,11). Moreover, these are non-senescent cells and therefore are a
109 valuable resource that can be used to interrogate aging-dependent changes in the breast tissue
110 – distinct from those arising from cellular senescence.

111 WGBS libraries were sequenced to ~30X genome coverage, generating ~ 1B 150-bp paired-
112 end reads (~86 Gb uniquely mapped sequence per sample) (**Table S2**) to obtain a
113 representation of about 95% of individual CpG sites in our samples (see Methods and **Tables**
114 **S2, S3**). We compared our DNA methylation profiles with previously published WGBS studies
115 on distinct human mammary epithelial cell types from reduction mammoplasty (RM) tissue of a
116 young (19y old) donor (20) and found that the cultured LEps resemble the ER-negative (ER^-)
117 LEp population (**Figure S1B**). Young and old LEPs clustered separately indicating differences
118 in methylomes (**Figure S1B**) that were not apparent when considering total methylation levels
119 (**Figure S1C, D**). Pairwise comparisons of methylomes also showed higher number of
120 differentially methylated regions (DMRs) between young and old pairs (**Figure S1E**). Comparing
121 the methylomes of old versus young LEps, we identified 18,860 differentially methylated regions
122 (DMRs) (see Methods), with 7,207 showing methylation loss and 11,653 showing methylation
123 gain in old LEps. Most DMRs were 100-500bp in length (**Figure S1F**). The methylation changes
124 at DMRs were confirmed with heatmaps showing methylation levels across samples (**Figure**

125 **1A**), as well as kernel smoothed methylation plots of representative DMRs (**Figure 1B**). The
126 majority of DNA methylation changes were observed at non-promoter regions (**Figure S1G-H**).
127 We assigned DMRs to genes based on overlap with either promoters or gene body (exonic
128 or intronic) (**Table S4A-B**). We used the Molecular Signatures Database (MSigDB) to identify
129 the gene sets or pathways over-represented among these genes (**Table S5A-D**). Genes with
130 methylation loss at promoters belong to gene sets expressed in mammary stem cells, cell-cell
131 anchoring junctions epithelial-to-mesenchymal transition, and genes downregulated in luminal
132 breast cancer (**Figure 1C, Table S5A**). Genes with promoter methylation gain were enriched for
133 luminal specific genes upregulated in luminal breast cancer and Polycomb (PRC2) target genes
134 (**Figure 1D, Table S5C**, BENPORATH gene sets). We also noted that genes expressed in
135 mammary stem cell enriched basal compartment showed methylation loss (**Figure S2A**)
136 whereas luminal lineage restricted genes showed methylation gain (**Figure S2B**). As an
137 example, *FOXP1*, a mammary stem cell specific gene essential for differentiation into various
138 mammary epithelial lineages (21), shows methylation loss at several sites its promoter and gene
139 body (**Figure 1E**). *ALDH1A3*, on the other hand, is an ER^- luminal lineage restricted gene (22)
140 which gains methylation at several sites including the promoter in old LEps (**Figure 1F**). DNA
141 methylation changes with age suggest that luminal cells increase stem cell related properties
142 while repressing genes that are specifically related to maintenance of the luminal lineage.
143

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

146 To investigate whether DNA methylation changes were associated with transcription
147 factor (TF) binding signatures, we performed motif enrichment analysis (see Supplemental
148 Methods for details) using position weight matrices (PWMs) from the JASPAR database (23).
149 DMRs that lose methylation are enriched for binding sites of TFs such as EGR1, TP63, TP73,

150 TEAD1, TEAD4, and RUNX1 (**Figure 2A left panel, Table S6A**). Several of these TF binding
151 sites, including TEAD, TP63, TP73, NFI, and EGR1, are enriched at active regulatory elements
152 in basal/myoepithelial cells (20). To investigate whether these TFs change expression with age,
153 we checked their expression using RNA-sequencing data from LEps and MEps of young and
154 old individuals (11). *TP63* and *NFIA (NFI family TF; binds to NFI half-sites)* showed increased
155 expression in old compared to young LEps (**Figure 2B and S3A**). We also checked whether
156 these TFs showed any DNA methylation changes. *EGR1* showed significant methylation loss at
157 its promoter (**Figure S3B**) whereas *NFIA* and *TP63* exhibit methylation loss at specific sites
158 within their gene bodies (**Table S4A**) suggesting altered methylation at regulatory elements.
159 Although not all TFs showed an increase in expression with age, we noted that *TP63, NFIA,*
160 *EGR1, and TP73* have relatively higher expression in MEps as compared to LEps (**Figure 2A**
161 **right panel and 2E**). The expression patterns of TFs and the hypomethylation of their targets
162 are concordant with the known roles of these TFs. For example, young LEps do not express
163 *TP63*, which is a MEp lineage-specific transcription factor (8). Furthermore, TEAD1 and TEAD4
164 are DNA binding partners for YAP (Yes-associated protein), a transcriptional co-activator
165 regulated by the Hippo signaling pathway. YAP, typically a MEp specific protein, is sequestered
166 in the cytoplasm through activity of kinases downstream of the Hippo pathway (24). Old LEps
167 exhibit increased *TP63* expression and nuclear YAP localization (7,8), which leads to increased
168 TEAD binding and YAP activity. YAP is also known to bind p73 (TP73 protein) and RUNX2 (24).
169 These results suggest that increased expression/activity of myoepithelial-specific TFs in old
170 LEps lead to increased binding and loss of DNA methylation in these regions (**Figure 2F**).

171 DMRs with methylation gain are enriched for binding sites of the ETS family of
172 transcription factors EHF, ELF3, ELF5, ELF4, ETV1, ETS1, ETV4, GABPA, and NFI half and full
173 sites (**Figure 2C left panel, Table S6B**). ELF5 is a luminal lineage-specific transcription factor
174 that decreases expression with age (**Figure 2D**) (8). Both *ELF3* and *ELF5* genes also show

175 significant methylation gain at their promoter and gene locus (**Figure S3C-D**). In addition, *ELF5*,
176 *EHF* and *ELF3* exhibit relatively higher expression in LEps as compared to MEps (**Figure 2C**
177 **right panel and 2E**). These results further suggest that decreased expression of luminal
178 lineage-specific TFs may lead to a decrease in binding and protection of binding sites from DNA
179 methyltransferases (DNMTs). This, in turn, may result in the methylation and repression of these
180 luminal lineage-specific regions (**Figure 2F**). We also examined the known DNA methylation
181 binding sensitivities of TFs (25) with enriched binding sites. TFs with binding sites showing
182 methylation loss either do not have a CpG at their respective site or their binding is unaffected
183 by DNA methylation (**Table S7**). Presumably they recruit other factors such as TETs to bring
184 about DNA methylation changes. However, TFs with binding sites showing methylation gain are
185 sensitive to DNA methylation (**Table S7**). Generally, sites of methylation loss were enriched for
186 myoepithelial-specific TF binding motifs (**Figure 2E**), whereas sites of methylation gain were
187 enriched for luminal lineage-specific TF binding motifs (**Figure 2E**). For example, SOX4 sites
188 are enriched in DMRs with methylation gain, and it is a luminal-specific TF (**Figure 2E**). SOX15,
189 on the other hand, is a myoepithelial-specific TF and its binding sites are enriched at DMRs with
190 methylation loss (**Figure 2E**). Expression of *SOX4* decreases, whereas *SOX15* increases with
191 age in old LEps (**Figure S3E-F**), indicating loss of expression of luminal lineage-specific TFs
192 and increased expression of myoepithelial-specific TFs. Collectively, we observe that aging-
193 associated DNA methylation changes in LEps occur at lineage-specific TF binding sites that
194 seem to contribute to loss of cell identity. Enticingly this is a characteristic of cancer cells
195 whereby even the so-called luminal subtype cancer cells exhibit expression of genes and
196 proteins that are characteristic of basal cells.
197
198 **Lineage-specific regulatory regions are hotspots for methylation changes with age.**

199 Given that aging-associated DNA methylation changes occurred at regions containing
200 lineage-specific TF binding sites, we examined whether these binding sites are functional by
201 measuring chromatin accessibility using ATAC-seq in young and old LEps (**Table S1**). Open
202 chromatin region at a TF binding site is a reliable indicator of a regulatory element with bound
203 TFs as not all TF binding sites in a given condition in a cell are bound. DMRs with methylation
204 loss showed corresponding increased accessibility in older LEps (**Figure 3A-B**). On the other
205 hand, DMRs with methylation gain showed a decrease in chromatin accessibility (**Figure 3C-D**).
206 These results indicate that functional regulatory elements undergo DNA methylation changes
207 with age.

208 To further investigate the functional relevance of the age-dependent DNA methylation
209 changes in the breast tissue, we used published histone modification ChIP-seq datasets in
210 mammary epithelial lineages to predict chromatin states at DMRs (20). To discern whether age-
211 dependent DNA methylation changes were occurring at lineage-specific regulatory elements,
212 we predicted chromatin states in ER^+ luminal, ER^- luminal, and basal (MEp) cells from RM tissue
213 of a young (19y old) donor (20). We classified DMRs into seven different chromatin states that
214 are characteristic of i) Active TSS, ii) TSS flanking/promoter, iii) transcribed, iv) Enhancer, v)
215 Heterochromatin, vi) Repressed and vii) None/unclassified regions based on the combinatorial
216 histone modification levels in each cell type (see Supplemental Methods; **Figure S4A-C**). DMRs
217 with methylation loss in old LEps overlap primarily with MEp-specific promoter and enhancer
218 regions (~80% of methylation loss DMRs) (**Figure 3E**). Because these sites also contain binding
219 sites for TFs enriched at MEp-specific active enhancers (20), these results indicate that MEp-
220 specific regulatory elements lose methylation in old LEps. Active enhancer sites are
221 hypomethylated (26,27) and methylation loss at MEp-specific regulatory elements might indicate
222 an epigenetic shift towards MEp gene expression patterns in LEps with age.

223 Conversely, DMRs with methylation gain overlap putative promoter and enhancer regions
224 in ER^- luminal cells, accounting for about 42% of methylation gain DMRs (**Figure 3F**). As these
225 DMRs also contain binding sites for TFs enriched at luminal lineage-specific active enhancers
226 (20), these results indicate that luminal-specific regulatory regions gain DNA methylation in aging
227 LEps. Together, these results suggest that aging LEps acquire permissive chromatin states at
228 MEp-specific regions and repress ER^- luminal-specific chromatin regions through changes in
229 DNA methylation.
230
231 **Polycomb repressive complex 2 target genes gain DNA methylation with age.**
232 The Polycomb repressive complex 2 (PRC2) is a histone methyltransferase complex that
233 trimethylates lysine 27 of histone H3 (H3K27me3) and mediates transcriptional silencing of
234 developmentally important genes. As per current understanding PRC2 is targeted to
235 unmethylated CpG-dense regions, primarily CpG islands (CGIs) that occur at gene promoters
236 (28). Earlier we found that genes with methylation gain at their promoters were enriched in
237 PRC2 target genes identified in embryonic stem cells (ESCs) (**Figure 1D**). Moreover, about 15%
238 of the DMRs with methylation gain overlapped with Polycomb repressed regions marked by
239 H3K27me3 (**Figure 3F and S4A-C**). PRC2 target gene promoters have been described to gain
240 DNA methylation with age in various tissues as well as in mice (13,29-32). They are also known
241 to be hypermethylated in diverse types of cancer, some among them are known tumor
242 suppressor genes (33). Since the MSigDB PRC2 target gene sets are ESC-specific, we used
243 H3K27me3 ChIP-seq data from ER^- and ER^+ luminal cells, and basal cells (20) to determine if
244 genes with promoter methylation gain are PRC2 targets in mammary epithelial lineages. We
245 found that 39-46% of promoters that gain methylation with age are marked by H3K27me3 across
246 the three cell lineages (**Figure S5A**). In contrast, only 14-27% of promoters with methylation loss
247 are enriched for H3K27me3 (**Figure S5B**). Promoters with DNA methylation gain also have

248 conserved H3K27me3 occupancy in all three epithelial lineages (**Figure S5C**) as compared to
249 promoters with DNA methylation loss that show H3K27me3 only in the luminal lineage (**Figure**
250 **S5D**). These results indicate that promoters with age-dependent DNA methylation gain are
251 indeed *bona fide* PRC2 target genes in mammary epithelial cells. These results are consistent
252 with PRC2 protecting unmethylated CpG-rich regions and that erosion of the PRC2 targeting
253 with aging (32) may be fundamentally related to increased methylation at its target genes.
254

255 **Stochastic methylation loss at transposable elements (TE) and methylation gain at CpG**
256 **islands (CGIs) with age.**

257 Aging is associated with increased transcriptional variability, cell-to-cell heterogeneity,
258 and DNA methylation variability (34-37). Because we use a sorted luminal lineage cell
259 population, we can discern DNA methylation variability in LEps eliminating confounding
260 problems such as changes in cell proportions with age. To examine whether there is an increase
261 in methylation heterogeneity in older individuals, we calculated methylation entropy, a measure
262 of methylation heterogeneity between adjacent CpGs, for 150 bp non-overlapping windows
263 across the genome (38). There are two modes by which a given region could increase
264 methylation entropy. An unmethylated region can gain methylation stochastically or a methylated
265 region can lose methylation stochastically (**Figure 4A-B**). Considering these regions, we
266 observed an overall increase in methylation entropy in the old LEps. We identified regions with
267 greater methylation entropy in older individuals and found a net decrease in methylation levels
268 at these regions (**Figure S6A-D**), consistent with the fact that most of the genome is methylated
269 in normal cells. Of these, unmethylated regions with increased methylation entropy were mainly
270 at CGIs (69.2%; 1017 of 1469; **Figure 4C**). Conversely, methylated regions with increased
271 entropy were found mainly at repeat regions (70.3%; 76575 of 108877; **Figure 4D**), which are
272 known to be enriched with TEs. These results indicate that increased DNA methylation

273 heterogeneity with age leads to stochastic methylation gain at CGIs and methylation loss at
274 repeat regions.

275 To examine the potential of stochastic methylation changes to have consequences for
276 breast cancer, we used a published WGBS dataset of 30 primary tumors containing ER+ (luminal
277 breast cancer) and ER- (basal breast cancer) samples (39). At CGIs with increased entropy,
278 luminal breast tumors showed an even greater increase in methylation and entropy than
279 observed in old LEps (**Figure 4E**). In contrast, basal tumors did not show an increase in CGI
280 methylation or entropy (**Figure 4F**), indicating that this behavior is specific to the luminal
281 subtypes of breast cancer. We identified the subset of entropy-gained CGIs that were
282 hypermethylated in at least 20% (n >= 6) of the breast tumors. These CGIs were enriched at
283 PRC2 target genes, which are known to be hypermethylated in luminal breast cancer (19) and
284 other types of cancer (**Figure S6E**). Of note, PRC2 target genes are repressed in normal tissues
285 through H3K27me3, and although they show DNA methylation gain at their CGI promoters,
286 expression changes are not expected. Apart from PRC2 target genes, we also found gene sets
287 that are upregulated in basal breast cancer and downregulated in luminal B breast cancer to be
288 enriched at these CGIs (**Figure S6E**). For example, *FOXC1* gene is expressed in LEps and
289 normal breast tissue but downregulated specifically in luminal and Her2-enriched breast cancers
290 (**Figure 4G**). Taken together, these results indicate that increased DNA methylation
291 heterogeneity with age leads to stochastic methylation gain at CGIs which is further exacerbated
292 in luminal breast cancer.

293

294 **Increased DNA methylation variability between old individuals at TEs.**

295 We next tested whether aging would lead to increased interindividual variability in addition
296 to increased methylation entropy. We indeed do observe increased variability of CpG
297 methylation in older individuals (**Figure S7A**). Since some amount of DNA methylation variability

298 is expected in human populations and differences at individual CpG sites may not have
299 functional consequences, we first identified systemic regions of variable methylation across all
300 10 individuals using the Systemic Interindividual Variation (SIV) approach (40) (see
301 Supplemental Methods). We identified 23,921 regions that were variably methylated only in old
302 samples and 25,585 regions that were specific to the young samples (**Figure S7B**). Regions
303 variably methylated in old samples mainly showed methylation loss compared methylation levels
304 in younger individuals (17,780 of 23,921; 74.3%). The variable regions were ~150p-450bp in
305 length with some spanning 2-2.5kb (**Figure S7C and S7D**). We next examined whether variable
306 regions contained TEs and found that 68.7% of the regions variable in young samples contained
307 TEs as compared to 67.59% of control non-variable regions. 71.4% regions variable in old
308 samples overlapped with TEs as compared to 67.65% of control non-variable regions (**Figure**
309 **S7E**). Among the TEs that are variably methylated in old samples, about 76.9% lose methylation
310 and 23.13% gain methylation (**Figure S7F**). We also confirmed that variable methylation at TEs
311 was not due to mapping issues (**Figure S7G**).
312
313 One of the main functions of DNA methylation is to silence TEs, especially those that are
314 evolutionarily recent and more likely to be capable of activity (41,42). To investigate whether the
315 TEs that are variably methylated in old samples are evolutionarily recent, we calculated the
316 Jukes-Cantor evolutionary age of each TE copy. Among the variably methylated TEs, the
317 evolutionarily recent elements (inserted <100 million years ago; (43)) tended towards DNA
318 methylation loss, whereas the older TE elements (>100 million years) tended to gain methylation
319 in old cells (**Figure 5A-D)**. These results indicate that evolutionarily recent TEs lose methylation
320 with age in a stochastic manner. While evolutionarily recent Alu elements such as AluY and AluS
321 are highly abundant in the genome, a few AluS subfamilies such as AluSz and AluSx were
322 enriched above background for stochastic loss of methylation (**Figure 5E, S7H**). Other

323 evolutionarily recent TE subfamilies were enriched at regions with variable methylation loss as
324 well, including both L1PA (primate-specific LINE-1) and HERVH elements (**Figure 5E-F)**.

325 Many TEs in the human genome have retained latent regulatory potential and can serve
326 as regulatory elements when not suppressed (44-46). We therefore examined the TF motifs
327 enriched at specific TE subfamilies with variable DNA methylation loss in old. We found that Alu
328 subfamilies contain motifs for Homeobox TFs such as SIX4, BAPX1 and nuclear receptor TFs
329 such as RARA which are upregulated in breast cancer (**Figure 5G**). Moreover, the younger
330 L1PA subfamilies L1PA(2-4) have motifs for many TFs; the most relevant among them being
331 binding sites for GATA, FOXA1 TFs and EREs which are bound by Estrogen receptor (ESR1)
332 (**Figure 5G**). GATA3, FOXA1 and ESR1 are highly upregulated in luminal breast cancers (47).
333 These results indicate that stochastic loss of methylation at these TEs can lead to binding of TFs
334 expressed during breast cancer development.

335 To examine whether TEs similarly contribute TF binding sites to DMR regions associated
336 with loss of lineage fidelity, we identified the TEs that overlap DMRs. Although many DMRs
337 overlap TEs, given the high number of copies in the genome only a few TE subfamilies,
338 especially those of the LTR class, were enriched over background. Several LTR subfamilies,
339 including HERVH, MER52A, MER52D and LTR27E were enriched at DMRs with methylation
340 loss in old LEps (**Figure S8A**) while LTR10C, HERVE, and MER11D (**Figure S8B**) were
341 enriched in DMRs with methylation gain. Among the TEs that lose methylation, HERVH elements
342 showed the highest enrichment for TEAD and RUNX TF binding sites (**Figure S8C**), motifs that
343 were also found in non-repetitive DMRs with methylation loss (**Figure 2A**) and known binding
344 partners of the MEp specific YAP. TP63 binding sites, which were also found at DMRs with
345 methylation loss (**Figure 2A**) were enriched at AluSx, AluSp, L2, and L3 elements. ETS family
346 TF binding sites that gain methylation were found within MIRb, MIR subfamilies of SINE
347 elements and L2, L2a-c LINE subfamilies (**Figure S8D**), similar to non-repetitive DMRs with

348 methylation gain (**Figure 2C**). SOX family TF binding sites that also occur in DMRs with
349 methylation gain (**Figure 2E**) were present in AluY, LTR10 and MER11D elements with
350 methylation gain (**Figure S8D**). Most of these DMR-associated TEs are evolutionarily older,
351 indicating that they have been co-opted by lineage-specific TFs and contribute to aging-
352 dependent DNA methylation changes in LEps.
353
354 **Transposable elements that lose methylation with age are activated in breast cancer.**
355 Given our observations that evolutionarily recent transposons lose methylation
356 stochastically with age and contain motifs for TFs overexpressed in breast cancer, we examined
357 whether breast cancers show loss of epigenetic suppression at these elements. Loss of TE
358 suppression can lead to aberrant transcription of TE-derived transcripts and alternative isoforms
359 due to splicing with neighboring genes in cancer cells (48). To investigate the potential of this
360 happening in breast cancer, we examined published RNA-seq data from high-grade ductal
361 carcinoma in situ (DCIS), a form of pre-cancer (49). We used the SQuIRE pipeline (50) to
362 measure gene and TE expression changes in DCIS as compared to normal tissues; 554 age-
363 hypomethylated TEs showed higher expression in DCIS. Of these, 468 were variably
364 hypomethylated with age and 86 were within DMRs with methylation loss. As with the loci with
365 variable methylation loss in older samples, HERVH elements and L1PA2 elements were most
366 common among the TEs that showed aberrant expression. A representative example of a
367 HERVH element that loses methylation in aging LEps and breast tumors is shown (**Figure 6A**).
368 Aberrant transcription from this HERVH element was observed only in DCIS samples (**Figure**
369 **6A**). We also found examples of exonization of TEs, where a MIRb element was included within
370 the exon of a non-coding RNA DRAIC expressed in DCIS (**Figure S9A)**. While aberrant
371 expression of hypomethylated TEs in DCIS was clear, we did not observe breast cancer subtype
372 specificity in the expression of these TEs.

373 We next examined whether TEs with methylation loss function as regulatory elements in
374 breast cancer. To this end we utilized ATAC-seq data from The Cancer Genome Atlas (TCGA)
375 breast cancer cohort (TCGA-BRCA) and ChIP-seq datasets from breast cancer cell lines (51).
376 We identified 1197 TCGA-BRCA ATAC-seq peaks that overlap TEs with methylation loss in
377 aging LEps (**Figure 6B**). Although some TEs were accessible in basal subtype tumors (418 of
378 1197), most of them were accessible across luminal tumors (779 of 1197 peaks) (**Figure 6B**).
379 Some luminal tumors that were diagnosed as lobular carcinomas had particularly high chromatin
380 accessibility across all TEs (**Figure 6B**). These results indicate that TEs hypomethylated with
381 age may function as regulatory elements in breast cancer.

382 To determine whether loss of methylation at TEs can lead to changes in expression of
383 genes that are thought to be important in breast cancer, we mapped the 554 age-
384 hypomethylated TEs to genes within ±500kb and calculated the correlation of TE expression
385 with each gene mapped within this window in DCIS samples (49). We identified 406 genes that
386 were significantly correlated with the expression of a nearby TE (**Table S8**). A representative
387 example of an L1PA4 element with significant expression correlation (correlation: 0.8, *p*-val:
388 $7.29e^{-9}$) with the nearby *ZNF92* gene is shown (**Figure S10A**). We find that *ZNF92* is also highly
389 expressed in luminal subtype breast tumors (Luminal A and Luminal B) as compared to basal
390 tumors (**Figure 6C**). These results suggest that TEs located at deleterious locations in the
391 genome are reawakened through stochastic loss of methylation causing expression of genes
392 that are associated with breast cancers. Overall suggesting that reawakening of TEs is directly
393 related to increased cancer susceptibility with age.

394

395 **Discussion**

396 Previous studies analyzing aging-associated DNA methylation changes in human
397 mammary luminal epithelial cells have either used array-based technologies (8,11,17,18) or

398 were performed using composite tissues that represented a mixture of cells (17,18). Array
399 probes are mainly enriched at genic regions and intergenic elements including repeats and distal
400 regulatory elements are under-represented (52). It has furthermore been shown that different
401 cell lineages have unique responses to aging (8,53), making pure cell populations critical for
402 interpretation. Our use of WGBS on sorted normal human mammary epithelial cells addresses
403 both shortcomings.

404 We find that aging leads to DNA methylation changes at lineage-specific gene regulatory
405 elements in mammary luminal epithelial cells, in-line with previous studies that show loss of
406 lineage fidelity in luminal cells with age (8). Specifically, luminal cells show DNA methylation loss
407 at myoepithelial-specific regulatory elements containing MEp-specific TF binding sites.
408 Moreover, luminal-specific regulatory regions containing LEp-specific TF binding sites are
409 repressed by DNA methylation with age. This loss of lineage-fidelity has been shown to occur
410 with age and in mammary tissue that is susceptible to breast cancer (11,54).

411 Previous DNA methylome studies in human cells suggested a global loss of methylation
412 with age (16,38,55). Those studies were performed using $CD4^+$ T cells from a neonate and a
413 centenarian (16), sorted $CD4^+$ T cells (38) and skin keratinocytes (55) from young (<25 y) and
414 old (>70 y) individuals. The global methylation loss with age was accompanied by a global
415 increase in entropy or methylation discordance (38). We did not observe a global methylation
416 loss or a global increase in entropy in mammary luminal epithelial cells with age. This could be
417 due to higher fidelity of DNA methylation maintenance in luminal cells or due to a lesser degree
418 of separation between the age of young (<30 years) and old (> 55 years) individuals in our study.
419 DNA methylome studies in aging mouse livers also showed no global loss of methylation, but
420 specific changes at regulatory elements were detected (56). Whereas global methylation loss
421 was not detected, we did detect methylation changes at regulatory elements and variable
422 methylation changes at evolutionarily recent transposons.

423 Global loss of methylation and focal increase in methylation, termed 'epigenetic drift', has
424 been reported to occur with age and in cancer (57). The focal increase in methylation has been
425 reported to occur mainly at CpG islands (CGIs). We find CGIs showing both gain and loss in
426 methylation reflecting changes in lineage-specific gene expression. However, we do observe
427 increased methylation disorder/entropy at CGIs that lead to methylation gains with age, which is
428 further enhanced during the transition to luminal breast cancer. These CGIs were predominantly
429 PRC2 targets that are known to gain methylation in other cancer types (58). PRC2 targets were
430 also found among genes that showed methylation gain at promoters. Loss/repositioning of PRC2
431 has been proposed to cause DNA methylation gain at its targets. Presumably, CGIs with
432 increased methylation entropy are in the process of becoming hypermethylated in luminal breast
433 cancer. Hypermethylation of PRC2 targets seems to be specific to the luminal subtype of breast
434 cancer (19), although the reasons for this are not currently clear.

435 DNA methylation is an important mechanism for silencing of transposable elements.
436 Reactivation of transposable elements has been shown in cancer cells, coincident with a global
437 loss of DNA methylation (59,60). Previous research comparing DNA methylation levels of $CD4^+$
438 T cells between a neonate and a centenarian reported that 12% of DMRs that lost methylation
439 with age were located on Alu elements (16), whereas changes in epigenetic silencing at TEs
440 with aging has not been systematically investigated. Increased expression of retrotransposons
441 and active retrotransposition have been reported in aging mouse tissues (61,62) and increased
442 LINE-1 expression has been reported in senescent human fibroblasts *in vitro* (63). However,
443 these studies used bulk quantitative strategies such as RT-qPCR, which are not ideal due to
444 high sequence similarity between multiple active and inactive copies. Our approach allowed us
445 to investigate the impact of aging on DNA methylation at specific TEs using non-senescent
446 mammary epithelial cells. Our results demonstrate that evolutionarily recent (e.g primate and
447 hominid specific) transposons are more likely to be silenced by DNA methylation in young and

448 show increased variability and methylation loss with age. These TEs could be an important
449 component of understanding differential susceptibility to cancer as not all aged individuals
450 develop cancer. Our data also underscore the need for use of human cells or organoids for future
451 studies as these cannot be recapitulated in mice.

452 We further show that the methylation loss observed at TEs with age becomes
453 exacerbated in frank luminal breast cancers. The derepressed TEs can lead to aberrant
454 transcription or function as regulatory elements in breast cancer cells. We demonstrate that
455 hypomethylated TEs could regulate genes associated with luminal subtype breast cancers. TEs
456 also contribute lineage-specific TF binding sites to regulatory elements associated with loss of
457 lineage fidelity with age. Retrotransposon activation has been widely observed in nearly every
458 human cancer, including breast cancer (64,65), and even in premalignant lesions (60). This
459 suggests that early events in tumorigenesis may be driven by retrotransposon activity. We show
460 here, for the first time, that loss of transposon silencing is evident in aging luminal epithelial cells
461 that could predispose towards breast cancer. Future studies are needed to characterize the role
462 of these TEs in the early events of mammary cancer initiation. Altogether, our results showing
463 DNA methylation alterations at regulatory elements, CGIs and TEs with aging provide further
464 evidence in support of the observation that epigenetic changes with age could determine
465 susceptibility towards breast cancer.

466

467 **Materials and Methods**

468 **Isolation of mammary luminal epithelial cells**

469 LEps were isolated from 4th passage (p4) cultures of pre-stasis finite-lifespan HMECs from
470 normal reduction mammoplasty (RM) tissues (6). Primary non-immortalized HMECs were
471 generated and maintained, as described previously (66). HMEC strains were established in the
472 laboratory of Dr. Martha Stampfer (Lawrence Berkeley National Laboratory) who has all required

473 IRB approvals to distribute these samples. Cells were grown in M87A medium with cholera toxin
474 and oxytocin at 0.5 ng/ml and 0.1nM, respectively. 4th passage HMECs were stained with anti-
475 human CD133-PE (Biolegend, clone 7) and anti-human CD271 (Biolegend, clone ME20.4) by
476 following standard flow cytometry protocol. LEps and MEps were enriched using Aria III (Becton
477 Dickinson) flow sorter as CD133^{+}/CD271^{-} and CD133^{-}/CD271^{+} fractions, respectively. The
478 different HMEC strains used are described in Table S1.

479 **WGBS**

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

491 Sequencing reads were hard trimmed 15 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) (67) with the --bowtie1 option (68). Bismark
496 (deduplicate_bismark) to remove PCR duplicates. Methpipe (69) (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 (70) was used to identify DMRs. Briefly, CpG sites with at least 4X coverage in at
502 least 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 **ATAC-seq and analysis**

508 50,000 live FACS sorted young (*n* = 4) and old (*n* = 4) LEps were used for Tn5 transposition in
509 duplicates using the Omni-ATAC method (71). Libraries were sequenced to a depth of 40 million
510 paired reads each on an Illumina HiSeq 2500 system. Raw sequencing reads were trimmed
511 using Trim Galore (https://github.com/FelixKrueger/TrimGalore) (version 0.4.5) to remove
512 adapters. Trimmed reads were aligned to the human reference genome (hg19) using bowtie (68)
513 with the --best -k 1 -X 2000 --mm --chunkmbs 1024 parameters. Unmapped reads, mates as
514 well as low quality reads and secondary alignments were removed using samtools. PCR
515 duplicates were identified using Picard and removed using samtools. The reads were shifted by
516 4 bases on the sense strand and 5 bases on the anti-sense strand to account for Tn5 bias.
517 Reads from technical replicates were pooled to generate pseudoreplicates. Peaks were called
518 using MACS2 version 2.1.1.20160309 using -p 0.01 --nomodel --shift -37 --extsize 73 --SPMR -
519 -keep-dup all --call-summits as parameters. Peak calls from technical replicates and pseudo-
520 replicates were compared using IDR to get the most conservative set of peaks for each sample.

521 **Enrichment of DMRs and variably methylated regions at TE subfamilies**

522 The extent of overlap of DMRs and variable regions at repat subfamilies (RepeatMasker
523 annotation) was calculated using bedtools intersect. Randomly shuffled control regions of the
524 same length were generated 1000 times, and the expected number of overlaps calculated using
525 bedtools intersect. Enrichment values were calculated by dividing the number of DMR overlaps
526 (observed) with the average number of random overlaps (expected). *P* values were calculated
527 by dividing the number of extreme values (at either tail of distribution) from random overlaps by
528 1000. Significantly enriched repeat subfamilies had $p < 0.05$, enrichment > 1.5-fold, and > 5
529 copies overlapping DMRs or variable regions.
530

531 **Availability of data and materials**

532 All raw and processed sequencing data files have been deposited at NCBI's Gene Expression
533 Omnibus (GEO) under accession GSE153696 (WGBS and ATAC-seq) and GSE182338 (RNA-
534 seq).
535

536 **Funding**

537 This work was supported by the National Institutes of Health, grants R01DK112041,
538 R01CA220693 (D.E.S.); K01DK104993 (A.L.); Department of Defense/Army Breast Cancer Era
539 of Hope Scholar Award BC141351, and Expansion Award BC181737, Conrad N. Hilton
540 Foundation, and City of Hope Center for Cancer and Aging (M.A.L.); the National Cancer
541 Institute (NCI) Cancer Metabolism Training Program Postdoctoral Fellowship T32CA221709
542 (R.W.S.) and City of Hope Program in Molecular and Cellular Biology pilot award (D.E.S and
543 M.A.L.). The research reported in this publication included work performed in the Integrative
544 Genomics and Analytical Cytometry Cores supported by the National Cancer Institute of the
545 National Institutes of Health under award number P30CA033572. The content is solely the
546 responsibility of the authors and does not necessarily represent the official views of the National

547 Institutes of Health. The funding bodies had no role in the design of the study, data collection,
548 analysis, and interpretation, decision to publish of data or in writing the manuscript.

549

550 **Authors' contributions**

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

554

555 **Acknowledgements**

556 We thank other members of the Schones laboratory for helpful discussions and comments, and
557 the continued involvement of patient advocates Susan Samson and Sandy Preto. The results
558 shown and referenced here are based in part upon data generated by the TCGA Research
559 Network: https://www.cancer.gov/tcga, and the ENCODE Consortium:
560 https://www.encodeproject.org/.

561

562 **Conflict of Interest Statement**

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

564

565 **References**

566 1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer Statistics, 2021. CA Cancer J Clin **2021**;71:7-
567 33
568 2. Jenkins EO, Deal AM, Anders CK, Prat A, Perou CM, Carey LA*, et al.* Age-specific changes in
569 intrinsic breast cancer subtypes: a focus on older women. Oncologist **2014**;19:1076-83
570 3. Prat A, Perou CM. Mammary development meets cancer genomics. Nat Med **2009**;15:842-4
571 4. Fu NY, Nolan E, Lindeman GJ, Visvader JE. Stem Cells and the Differentiation Hierarchy in
572 Mammary Gland Development. Physiol Rev **2020**;100:489-523
573 5. Benz CC. Impact of aging on the biology of breast cancer. Crit Rev Oncol Hematol **2008**;66:65-
574 74

575 6. Garbe JC, Pepin F, Pelissier FA, Sputova K, Fridriksdottir AJ, Guo DE*, et al.* Accumulation of
576 multipotent progenitors with a basal differentiation bias during aging of human mammary
577 epithelia. Cancer Res **2012**;72:3687-701
578 7. Pelissier FA, Garbe JC, Ananthanarayanan B, Miyano M, Lin C, Jokela T*, et al.* Age-related
579 dysfunction in mechanotransduction impairs differentiation of human mammary epithelial
580 progenitors. Cell Rep **2014**;7:1926-39
581 8. Miyano M, Sayaman RW, Stoiber MH, Lin CH, Stampfer MR, Brown JB*, et al.* Age-related
582 gene expression in luminal epithelial cells is driven by a microenvironment made from
583 myoepithelial cells. Aging (Albany NY) **2017**;9:2026-51
584 9. Lee JK, Garbe JC, Vrba L, Miyano M, Futscher BW, Stampfer MR*, et al.* Age and the means of
585 bypassing stasis influence the intrinsic subtype of immortalized human mammary epithelial cells.
586 Front Cell Dev Biol **2015**;3:13
587 10. Pelissier Vatter FA, Schapiro D, Chang H, Borowsky AD, Lee JK, Parvin B*, et al.* High-
588 Dimensional Phenotyping Identifies Age-Emergent Cells in Human Mammary Epithelia. Cell
589 Rep **2018**;23:1205-19
590 11. Sayaman RW, Miyano M, Senapati P, Shalabi S, Zirbes A, Todhunter ME*, et al.* Epigenetic
591 changes with age primes mammary luminal epithelia for cancer initiation. bioRxiv
592 **2021**:2021.02.12.430777
593 12. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev **2002**;16:6-21
594 13. Maegawa S, Hinkal G, Kim HS, Shen L, Zhang L, Zhang J*, et al.* Widespread and tissue specific
595 age-related DNA methylation changes in mice. Genome Res **2010**;20:332-40
596 14. Horvath S. DNA methylation age of human tissues and cell types. Genome Biol **2013**;14:R115
597 15. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H*, et al.* Human aging-associated
598 DNA hypermethylation occurs preferentially at bivalent chromatin domains. Genome Res
599 **2010**;20:434-9
600 16. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A*, et al.* Distinct DNA methylomes of
601 newborns and centenarians. Proc Natl Acad Sci U S A **2012**;109:10522-7
602 17. Hofstatter EW, Horvath S, Dalela D, Gupta P, Chagpar AB, Wali VB*, et al.* Increased epigenetic
603 age in normal breast tissue from luminal breast cancer patients. Clin Epigenetics **2018**;10:112
604 18. Castle JR, Lin N, Liu J, Storniolo AMV, Shendre A, Hou L*, et al.* Estimating breast tissue-
605 specific DNA methylation age using next-generation sequencing data. Clin Epigenetics
606 **2020**;12:45
607 19. Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jonsson G, Olsson H*, et al.* Molecular
608 subtypes of breast cancer are associated with characteristic DNA methylation patterns. Breast
609 Cancer Res **2010**;12:R36
610 20. Pellacani D, Bilenky M, Kannan N, Heravi-Moussavi A, Knapp D, Gakkhar S*, et al.* Analysis of
611 Normal Human Mammary Epigenomes Reveals Cell-Specific Active Enhancer States and
612 Associated Transcription Factor Networks. Cell Rep **2016**;17:2060-74
613 21. Fu NY, Pal B, Chen Y, Jackling FC, Milevskiy M, Vaillant F*, et al.* Foxp1 Is Indispensable for
614 Ductal Morphogenesis and Controls the Exit of Mammary Stem Cells from Quiescence. Dev
615 Cell **2018**;47:629-44 e8
616 22. Nguyen QH, Pervolarakis N, Blake K, Ma D, Davis RT, James N*, et al.* Profiling human breast
617 epithelial cells using single cell RNA sequencing identifies cell diversity. Nat Commun
618 **2018**;9:2028
619 23. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA*, et al.*
620 JASPAR 2020: update of the open-access database of transcription factor binding profiles.
621 Nucleic Acids Res **2020**;48:D87-D92
622 24. Badouel C, Garg A, McNeill H. Herding Hippos: regulating growth in flies and man. Curr Opin
623 Cell Biol **2009**;21:837-43

624 25. Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, *et al.* Impact of cytosine
625 methylation on DNA binding specificities of human transcription factors. Science **2017**;356
626 26. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, *et al.* DNA-binding factors shape
627 the mouse methylome at distal regulatory regions. Nature **2011**;480:490-5
628 27. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, *et al.* Human DNA
629 methylomes at base resolution show widespread epigenomic differences. Nature **2009**;462:315-
630 22
631 28. Blackledge NP, Klose RJ. The molecular principles of gene regulation by Polycomb repressive
632 complexes. Nat Rev Mol Cell Biol **2021**;22:815-33
633 29. Beerman I, Bock C, Garrison BS, Smith ZD, Gu H, Meissner A, *et al.* Proliferation-dependent
634 alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. Cell Stem
635 Cell **2013**;12:413-25
636 30. Sun D, Luo M, Jeong M, Rodriguez B, Xia Z, Hannah R, *et al.* Epigenomic profiling of young
637 and aged HSCs reveals concerted changes during aging that reinforce self-renewal. Cell Stem
638 Cell **2014**;14:673-88
639 31. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, *et al.*
640 Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of
641 cancer. Genome Res **2010**;20:440-6
642 32. Jung M, Pfeifer GP. Aging and DNA methylation. BMC Biol **2015**;13:7
643 33. Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, *et al.* Epigenetic
644 stem cell signature in cancer. Nat Genet **2007**;39:157-8
645 34. Hernando-Herraez I, Evano B, Stubbs T, Commere PH, Jan Bonder M, Clark S, *et al.* Ageing
646 affects DNA methylation drift and transcriptional cell-to-cell variability in mouse muscle stem
647 cells. Nat Commun **2019**;10:4361
648 35. Bahar R, Hartmann CH, Rodriguez KA, Denny AD, Busuttil RA, Dolle ME, *et al.* Increased
649 cell-to-cell variation in gene expression in ageing mouse heart. Nature **2006**;441:1011-4
650 36. Cheung P, Vallania F, Warsinske HC, Donato M, Schaffert S, Chang SE, *et al.* Single-Cell
651 Chromatin Modification Profiling Reveals Increased Epigenetic Variations with Aging. Cell
652 **2018**;173:1385-97 e14
653 37. Martinez-Jimenez CP, Eling N, Chen HC, Vallejos CA, Kolodziejczyk AA, Connor F, *et al.*
654 Aging increases cell-to-cell transcriptional variability upon immune stimulation. Science
655 **2017**;355:1433-6
656 38. Jenkinson G, Pujadas E, Goutsias J, Feinberg AP. Potential energy landscapes identify the
657 information-theoretic nature of the epigenome. Nat Genet **2017**;49:719-29
658 39. Brinkman AB, Nik-Zainal S, Simmer F, Rodriguez-Gonzalez FG, Smid M, Alexandrov LB, *et*
659 *al.* Partially methylated domains are hypervariable in breast cancer and fuel widespread CpG
660 island hypermethylation. Nat Commun **2019**;10:1749
661 40. Gunasekara CJ, Scott CA, Laritsky E, Baker MS, MacKay H, Duryea JD, *et al.* A genomic atlas
662 of systemic interindividual epigenetic variation in humans. Genome Biol **2019**;20:105
663 41. Edwards JR, Yarychkivska O, Boulard M, Bestor TH. DNA methylation and DNA
664 methyltransferases. Epigenetics Chromatin **2017**;10:23
665 42. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian
666 development and disease. Nat Rev Mol Cell Biol **2019**;20:590-607
667 43. Burns KH, Boeke JD. Human transposon tectonics. Cell **2012**;149:740-52
668 44. Michalak EM, Nacerddine K, Pietersen A, Beuger V, Pawlitzky I, Cornelissen-Steijger P, *et al.*
669 Polycomb group gene Ezh2 regulates mammary gland morphogenesis and maintains the luminal
670 progenitor pool. Stem Cells **2013**;31:1910-20

671 45. Xie M, Hong C, Zhang B, Lowdon RF, Xing X, Li D*, et al.* DNA hypomethylation within
672 specific transposable element families associates with tissue-specific enhancer landscape. Nat
673 Genet **2013**;45:836-41
674 46. Pehrsson EC, Choudhary MNK, Sundaram V, Wang T. The epigenomic landscape of
675 transposable elements across normal human development and anatomy. Nat Commun
676 **2019**;10:5640
677 47. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA*, et al.* Molecular portraits of
678 human breast tumours. Nature **2000**;406:747-52
679 48. Jang HS, Shah NM, Du AY, Dailey ZZ, Pehrsson EC, Godoy PM*, et al.* Transposable elements
680 drive widespread expression of oncogenes in human cancers. Nat Genet **2019**;51:611-7
681 49. Abba MC, Gong T, Lu Y, Lee J, Zhong Y, Lacunza E*, et al.* A Molecular Portrait of High-Grade
682 Ductal Carcinoma In Situ. Cancer Res **2015**;75:3980-90
683 50. Yang WR, Ardeljan D, Pacyna CN, Payer LM, Burns KH. SQuIRE reveals locus-specific
684 regulation of interspersed repeat expression. Nucleic Acids Res **2019**;47:e27
685 51. Franco HL, Nagari A, Malladi VS, Li W, Xi Y, Richardson D*, et al.* Enhancer transcription
686 reveals subtype-specific gene expression programs controlling breast cancer pathogenesis.
687 Genome Res **2018**;28:159-70
688 52. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT*, et al.* DNA methylation aging
689 clocks: challenges and recommendations. Genome Biol **2019**;20:249
690 53. Kimmel JC, Penland L, Rubinstein ND, Hendrickson DG, Kelley DR, Rosenthal AZ. Murine
691 single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging. Genome Res
692 **2019**;29:2088-103
693 54. Shalabi SF, Miyano M, Sayaman RW, Lopez JC, Jokela TA, Todhunter ME*, et al.* Evidence for
694 accelerated aging in mammary epithelia of women carrying germline BRCA1 or BRCA2
695 mutations. Nature Aging **2021**;1:838-49
696 55. Vandiver AR, Irizarry RA, Hansen KD, Garza LA, Runarsson A, Li X*, et al.* Age and sun
697 exposure-related widespread genomic blocks of hypomethylation in nonmalignant skin. Genome
698 Biol **2015**;16:80
699 56. Cole JJ, Robertson NA, Rather MI, Thomson JP, McBryan T, Sproul D*, et al.* Diverse
700 interventions that extend mouse lifespan suppress shared age-associated epigenetic changes at
701 critical gene regulatory regions. Genome Biol **2017**;18:58
702 57. Issa JP. Aging and epigenetic drift: a vicious cycle. J Clin Invest **2014**;124:24-9
703 58. Easwaran H, Johnstone SE, Van Neste L, Ohm J, Mosbruger T, Wang Q*, et al.* A DNA
704 hypermethylation module for the stem/progenitor cell signature of cancer. Genome Res
705 **2012**;22:837-49
706 59. Burns KH. Transposable elements in cancer. Nat Rev Cancer **2017**;17:415-24
707 60. Ishak CA, De Carvalho DD. Reactivation of Endogenous Retroelements in Cancer Development
708 and Therapy. Annual Review of Cancer Biology **2020**;4
709 61. De Cecco M, Criscione SW, Peterson AL, Neretti N, Sedivy JM, Kreiling JA. Transposable
710 elements become active and mobile in the genomes of aging mammalian somatic tissues. Aging
711 (Albany NY) **2013**;5:867-83
712 62. Benayoun BA, Pollina EA, Singh PP, Mahmoudi S, Harel I, Casey KM*, et al.* Remodeling of
713 epigenome and transcriptome landscapes with aging in mice reveals widespread induction of
714 inflammatory responses. Genome Res **2019**;29:697-709
715 63. De Cecco M, Ito T, Petrashen AP, Elias AE, Skvir NJ, Criscione SW*, et al.* L1 drives IFN in
716 senescent cells and promotes age-associated inflammation. Nature **2019**;566:73-8
717 64. Rodic N, Sharma R, Sharma R, Zampella J, Dai L, Taylor MS*, et al.* Long interspersed element-
718 1 protein expression is a hallmark of many human cancers. Am J Pathol **2014**;184:1280-6

65. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell **2015**;160:48-61
66. Labarge MA, Garbe JC, Stampfer MR. Processing of human reduction mammoplasty and mastectomy tissues for cell culture. J Vis Exp **2013**
67. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics **2011**;27:1571-2
68. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol **2009**;10:R25
69. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J*, et al.* A reference methylome database and analysis pipeline to facilitate integrative and comparative epigenomics. PLoS One **2013**;8:e81148
70. Gaspar JM, Hart RP. DMRfinder: efficiently identifying differentially methylated regions from MethylC-seq data. BMC Bioinformatics **2017**;18:528
71. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S*, et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. Nat Methods **2017**;14:959-62

Figures

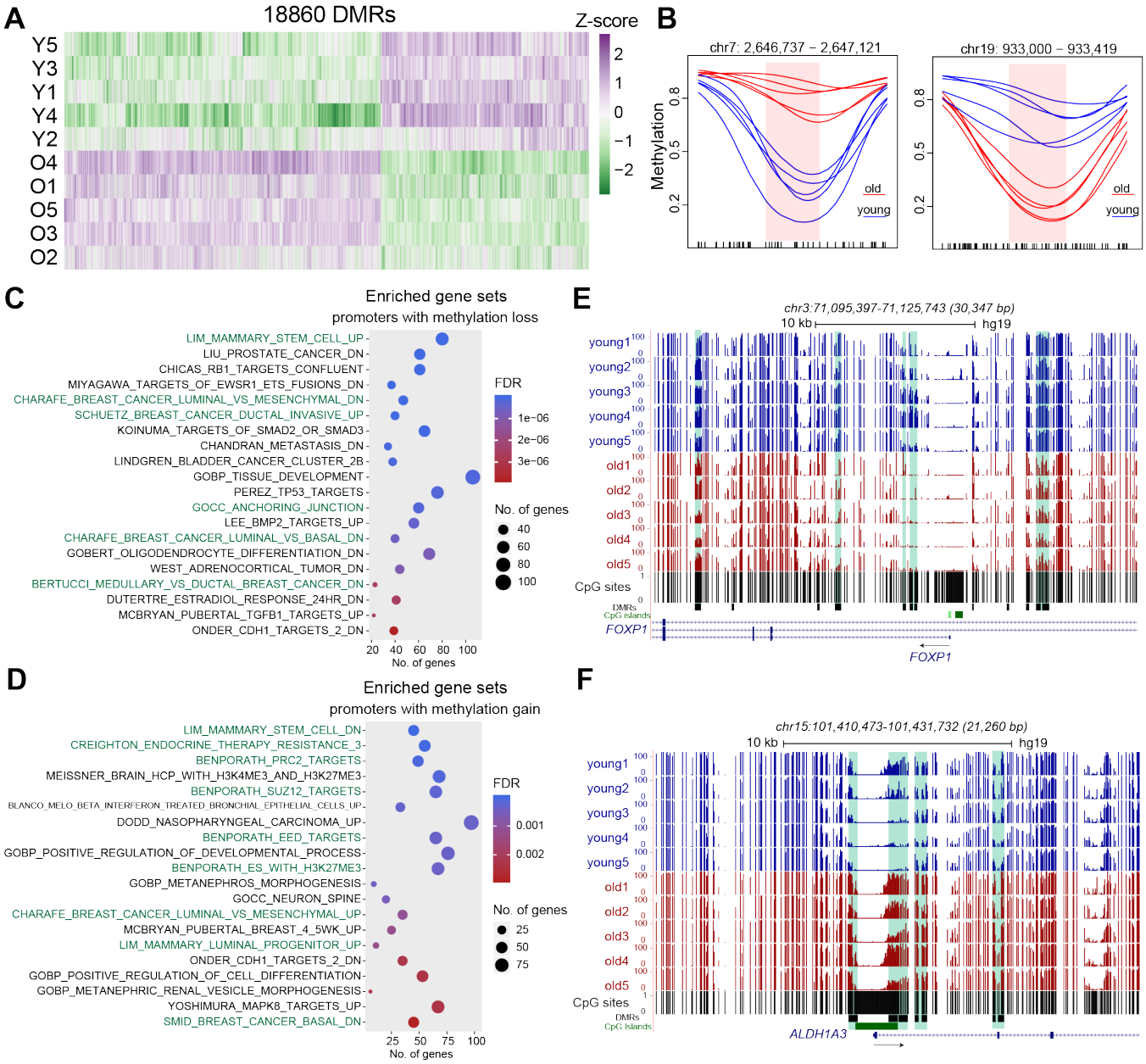


Figure 1: Mammary luminal epithelial cells from older individuals show DNA methylation changes at distinct genomic loci. (A) Methylation levels of identified DMRs across all samples. Samples are in rows, and DMRs are in columns. The color intensity represents the Z-score scaled from the mean. Values ranging from higher to lower methylation are shown as purple to green. (B) Kernel smoothed line plots of representative DMRs±500bp with methylation gain (left panel) and methylation loss (right panel). Blue and red lines respectively represent young and old samples. DMRs are highlighted in pink and black ticks mark each CpG position. (C) MSigDB gene sets enriched in genes with age dependent DNA methylation loss at promoters (TSS+/-2kb). (D) MSigDB gene sets enriched in genes with age dependent DNA methylation gain at promoters (TSS+/-2kb). Relevant gene sets are highlighted in green. C-D. UP: upregulated, DN: downregulated. (E) Genome browser tracks showing DMRs with methylation loss in old LEps at the *FOXP1* gene locus. (F) Genome browser tracks showing methylation gain at the CpG island promoter of *ALDH1A3* gene. E-F. DMR positions are shown as black boxes.

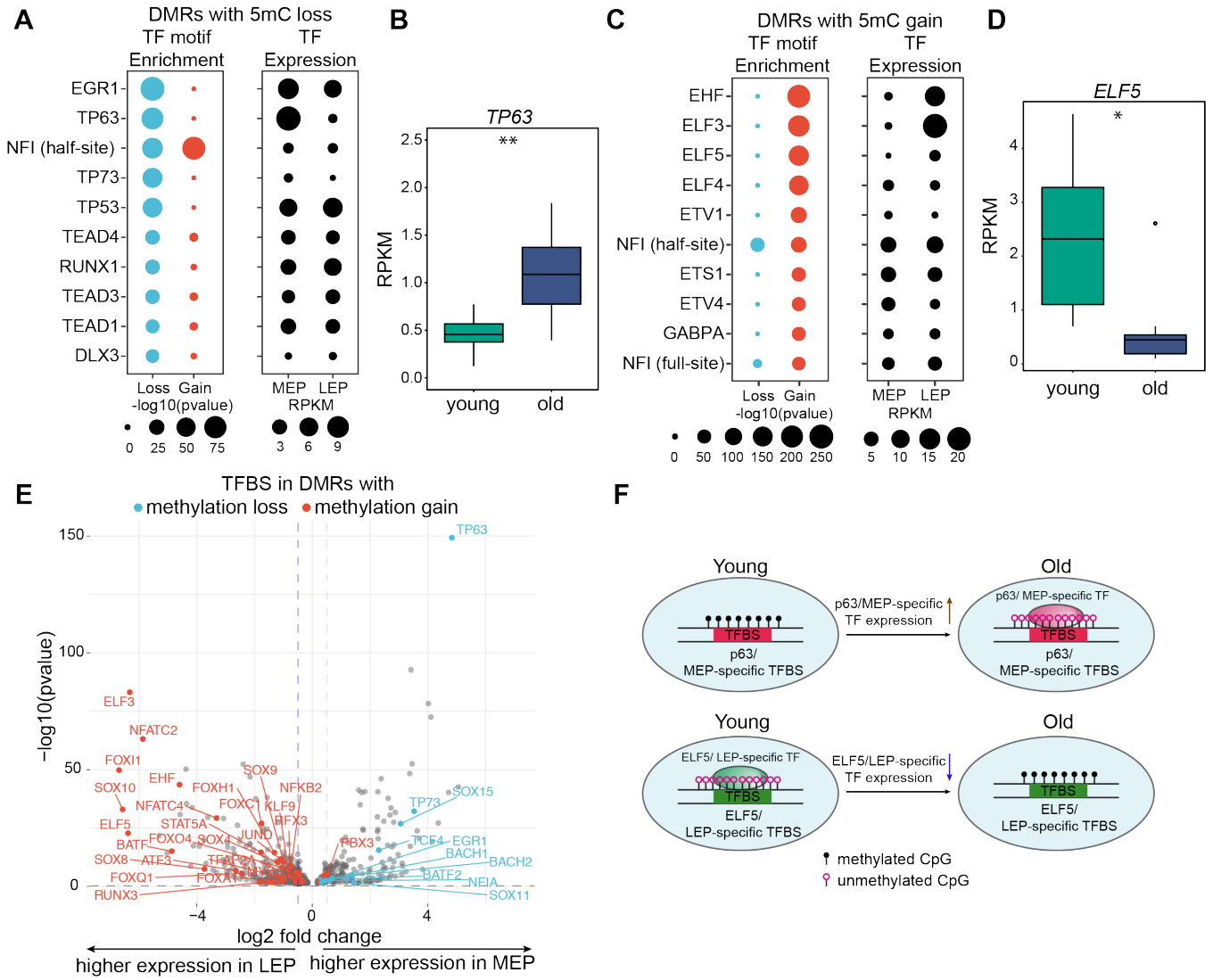


Figure 2: Differentially methylated regions show enrichment of lineage-specific transcription factor binding sites. (A) Top ten TF binding sites enriched in the DMRs that lose methylation with age are shown. Dots in the left panel show -log10(p-value) for the enrichment of the transcription factor binding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

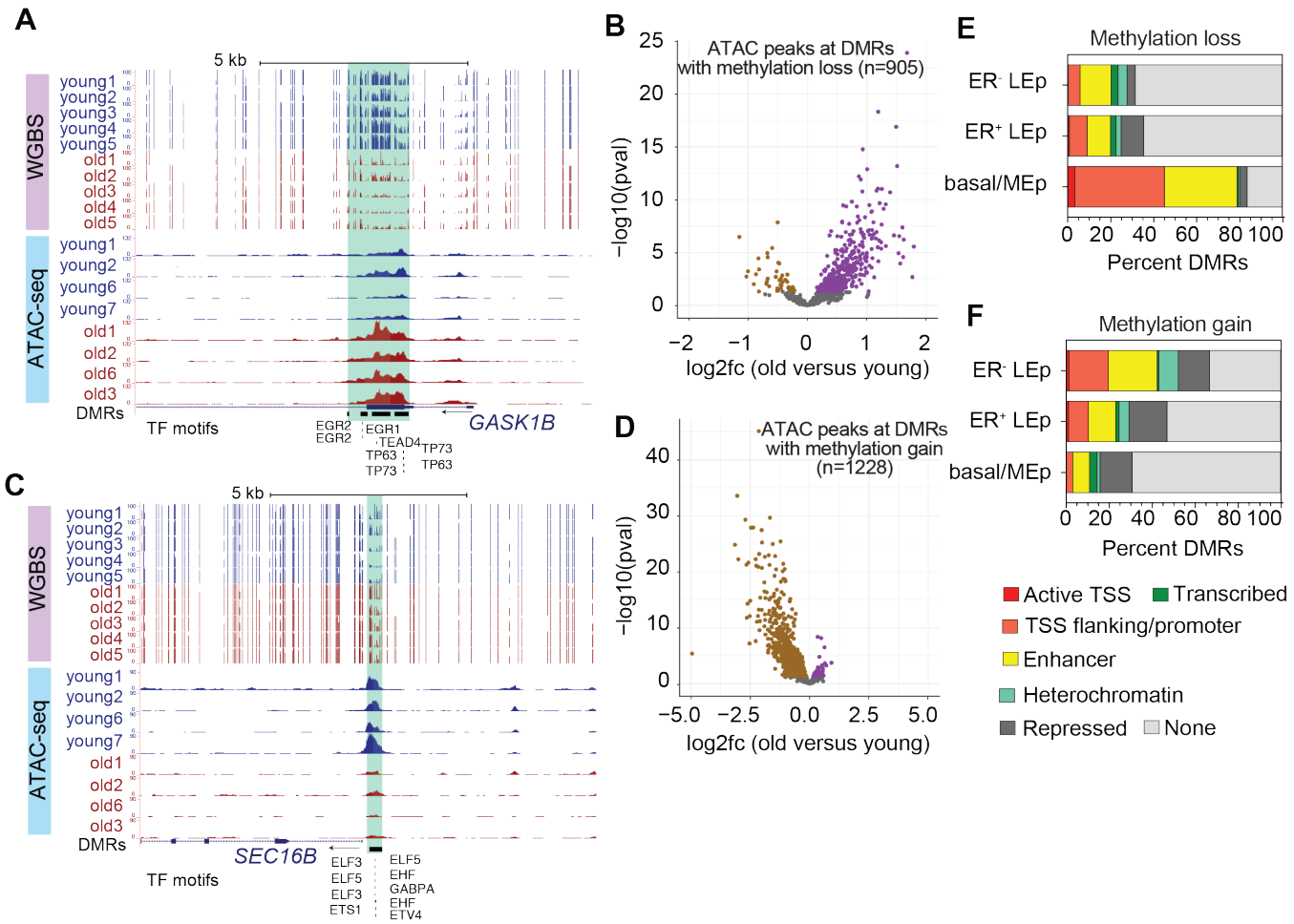


Figure 3: Differentially methylated regions are enriched at lineage-specific regulatory elements. (A) Genome browser tracks showing a representative example of a DMR with methylation loss accompanied with increase in chromatin accessibility in old LEPs. DMRs are highlighted in green. Bottom tracks show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

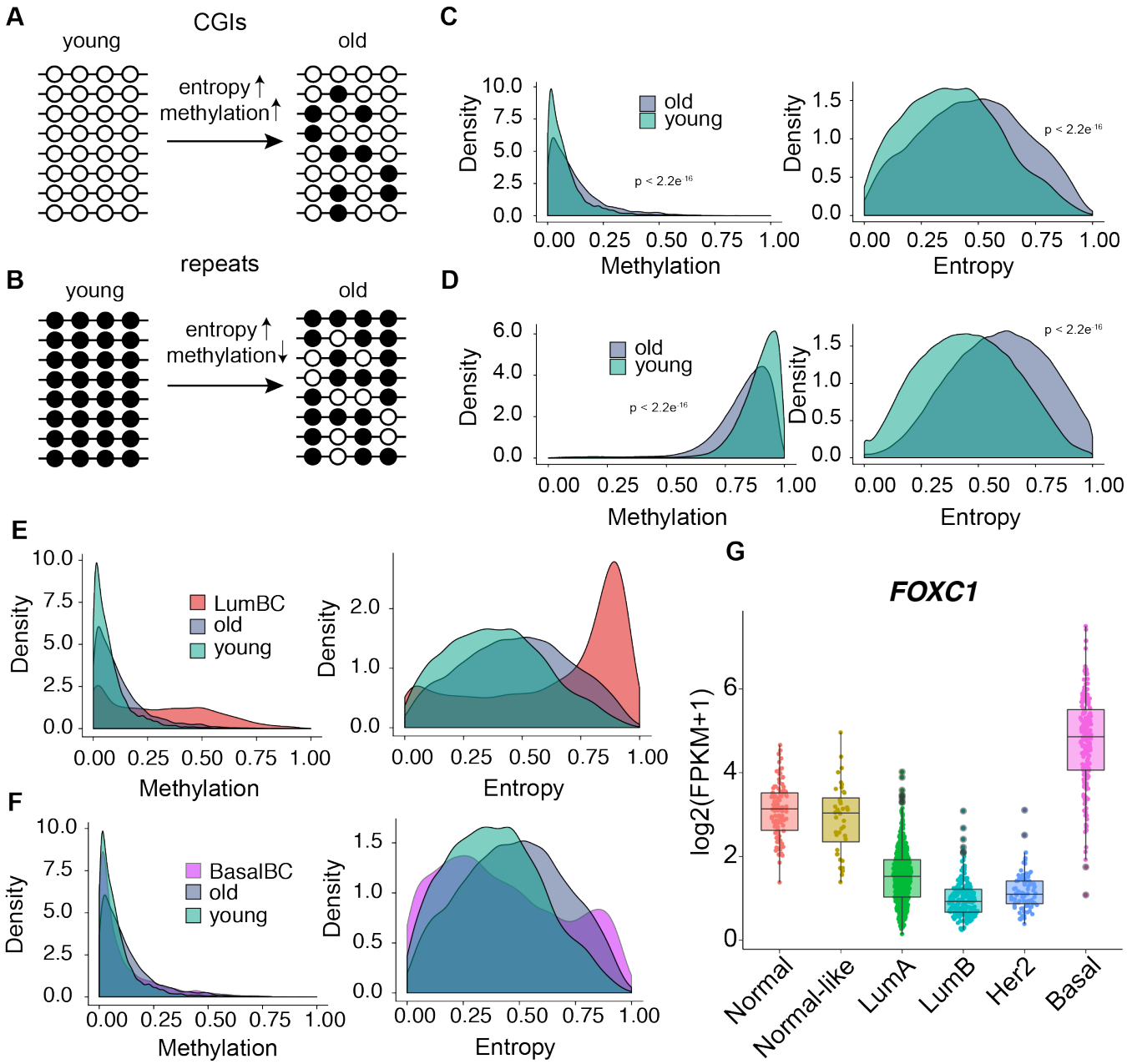
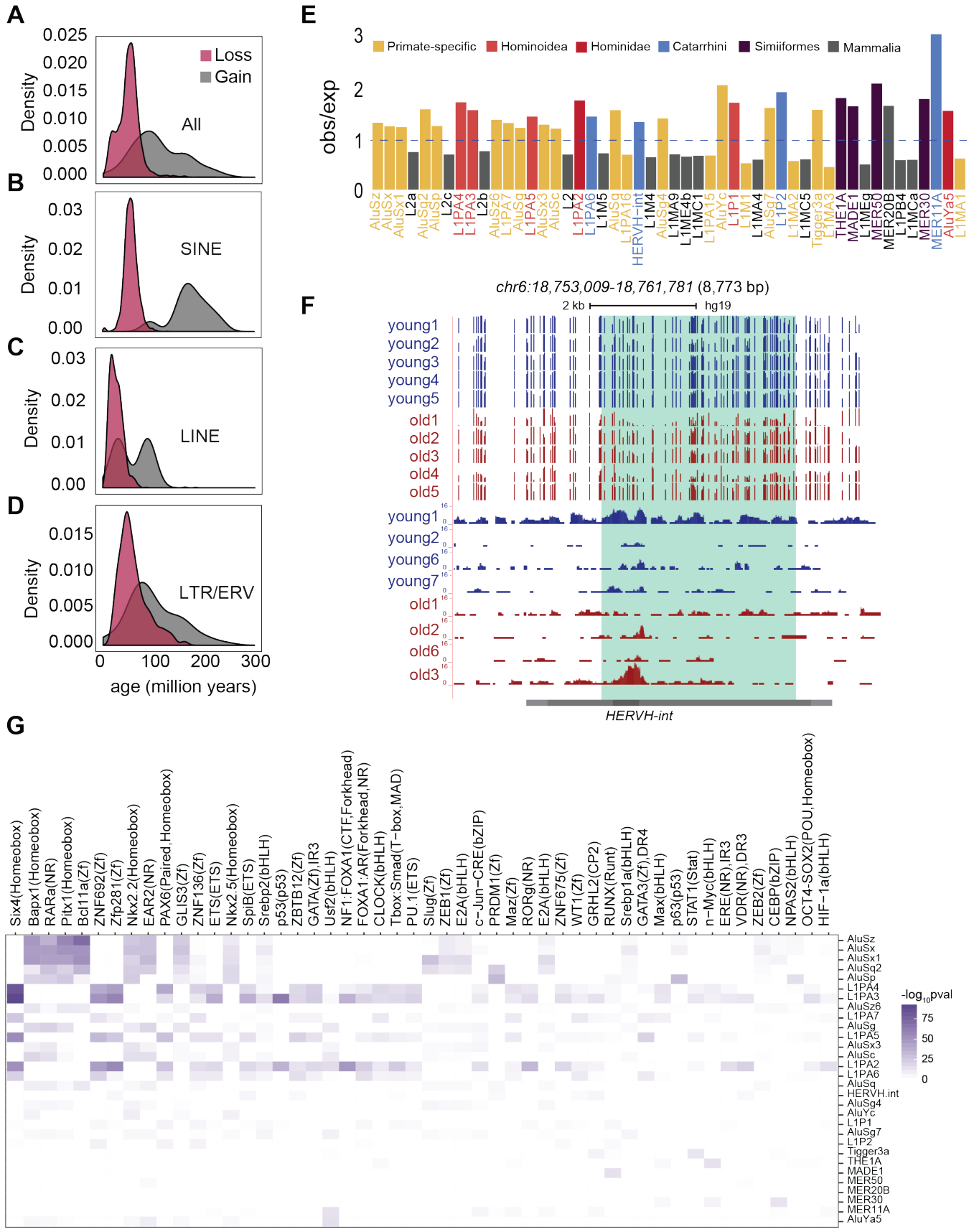


Figure 4: Aging leads to stochastic methylation loss at transposable elements and methylation gain at CpG islands (CGIs). (A) Schematic depicting changes in entropy associated with stochastic methylation gain at CGIs. (B) Schematic depicting changes in entropy associated with stochastic methylation loss at fully methylated regions such as repeat elements. (C) Methylation and entropy levels at unmethylated CGIs that show increase in entropy. (D) Alu methylation and entropy levels at fully methylated Alu elements that show increase in entropy. C-D. Significant differences in distributions were tested using Kolmogorov–Smirnov test. (E) Aggregate profiles of methylation and entropy levels at unmethylated CGIs that show increase in entropy with age. Luminal breast cancer samples (red) show an increase in DNA methylation as well as entropy at these CGIs. (F) Aggregate profiles of methylation and entropy levels at unmethylated CGIs that show increase in entropy with age. Basal breast cancer samples (pink) do not show an increase in DNA methylation or entropy at these CGIs. (G) Expression of *FOXC1* in normal tissues and five PAM50 subtypes of breast tumors from the TCGA BRCA collection. Luminal A (LumA) versus Basal p-value < 2.2×10^{-16}; using two-tailed Wilcoxon signed-rank test. Luminal B (LumB) versus Basal (p-value < 2.2×10^{-16}; using two-tailed Wilcoxon signed-rank test).

825


A
Loss
Gain
All
Density
B
SINE
C
LINE
D
LTR/ERV
age (million years)
E
Primate-specific
Hominoidea
Hominidae
Catarrhini
Simiiformes
Mammalia
obs/exp
F
chr6:18,753,009-18,761,781 (8,773 bp)
2 kb
hg19
young1
young2
young3
young4
young5
old1
old2
old3
old4
old5
young6
young7
old6
HERVH-int
G
-log10pval

826

Figure 5: Variably methylated regions with methylation loss are enriched at primate-specific transposable elements (A-D) Evolutionary age of (A) all TE elements (B) SINE, (C) LINE, and (D) LTR/ERV elements that are variably methylated with methylation loss or gain. Evolutionary ages of only the significantly enriched TE subfamilies are plotted. (E) TE subfamilies enriched in regions with variable methylation loss. Bars show observed/expected ratio for each TE subfamily. Expected number of overlaps in shuffled controls was used to calculate the ratio. Colors in the observed/expected graph highlight primate-specific or younger elements and grey bars indicate older elements. (F) Genome browser tracks showing a representative example of a HERVH-int element showing variable methylation loss in old samples. Bottom tracks are from ATAC-seq data showing gain of accessibility at the variably methylated region in the corresponding sample. (G) TF binding motifs enriched within each TE subfamily with variable methylation loss. Color scale is –log10(p-value) of TF motif enrichment.

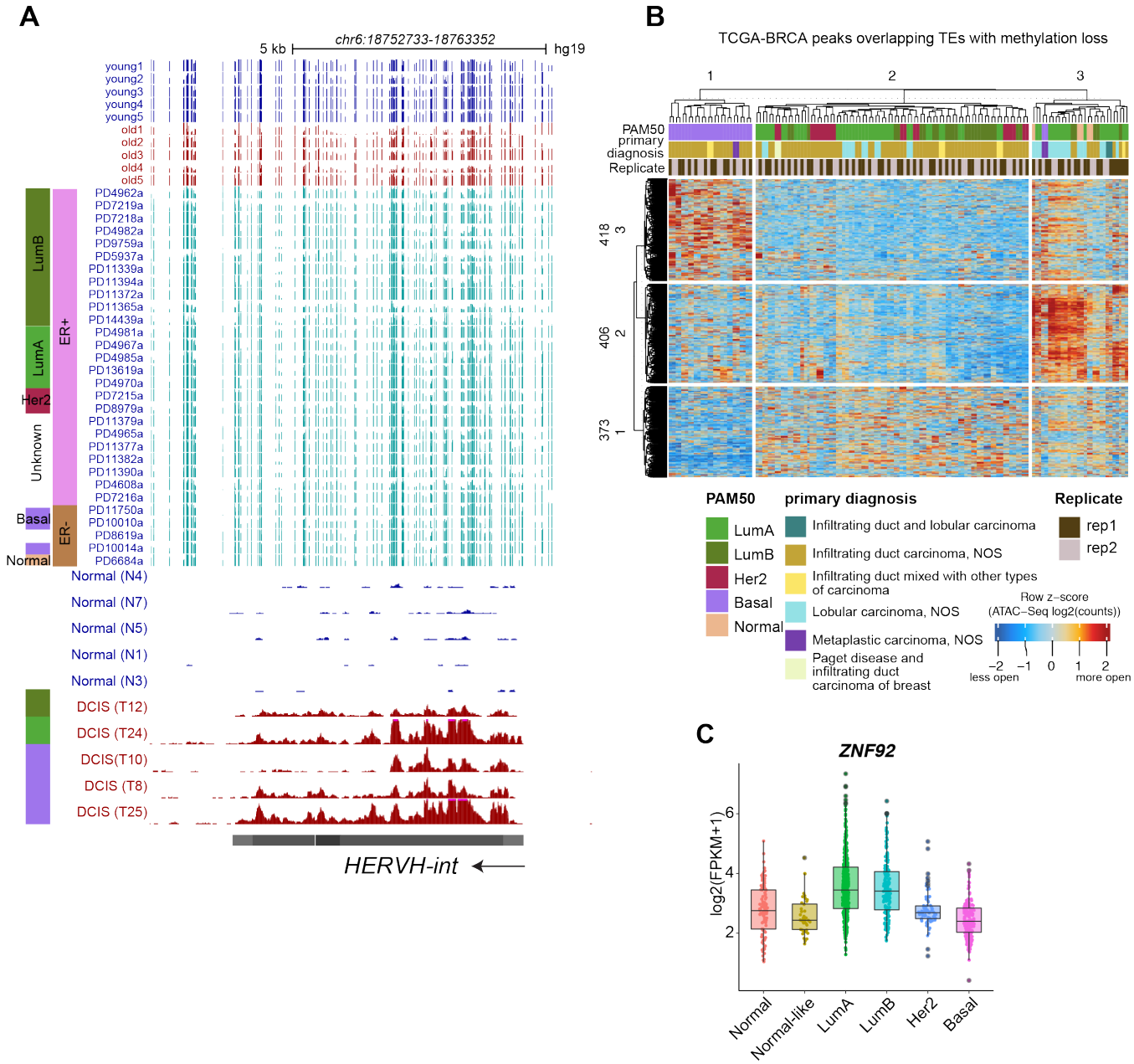


Figure 6: Transposable elements with age-dependent methylation loss are activated in breast cancer. (A) A representative HERVH element showing variable loss of methylation in old samples and

842 in breast cancer (BC) samples. PAM50 subtypes and estrogen receptor (ER) expression status of the
843 tumors are shown for the BC WGBS tracks. Bottom tracks are RNA-seq tracks from DCIS samples
844 showing cryptic transcription of the HERVH element. All RNA-seq tracks are scaled to the same value
845 and PAM50 subtype is indicated on the right. (B) Heatmap showing signals across 74 TCGA-BRCA
846 samples for ATAC-seq peaks that overlap TEs hypomethylated with age. ATAC-seq peaks (rows) and
847 TCGA BRCA samples (columns) are split into 3 clusters each using k-means clustering. (C) Expression
848 of *ZNF92* in normal tissues and five PAM50 subtypes of breast tumors from the TCGA BRCA collection
849 (p-value Luminal A (LumA) versus Basal and Luminal B (LumB) versus Basal (p-value < 2.2×10^{-16}; using
850 two-tailed Wilcoxon signed-rank test).