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

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

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

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

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

*Corresponding Author:

Dustin E. Schones, Ph.D.

Department of Diabetes Complications and Metabolism

City of Hope

Duarte, CA, 91010

Email: dschones@coh.org

Phone: 626-218-1319

Abstract

Background: The aging process is characterized by an accumulation of cellular damage, including alterations to the epigenome. While it is clear, there is a general lack of DNA methylation fidelity with aging, the mechanisms behind epigenomic alterations – and the functional consequences of these changes – have remained unclear. One of the primary roles of DNA methylation in mammalian genomes is in the repression of retrotransposons. Many of these repetitive elements contain regulatory sequences, and loss of repression can lead to aberrant transcriptional regulation. Aging-associated DNA methylation loss at transposable elements (TEs), therefore, has the potential to rewire gene expression networks to be favorable for pathologies, including age-related diseases such as cancer. However, the extent of loss of epigenetic silencing at transposable elements during aging and the functional consequences of this have remained unclear.

Results: We profiled DNA methylation across the genome for human mammary luminal epithelial cells (LEPs), a key cell type associated with age-related luminal breast cancer at different ages. Our analysis reveals distinct DNA methylation patterns in old LEPs. Luminal lineage-specific regulatory elements gain methylation, while myoepithelial lineage-specific regulatory elements lose methylation. These methylation changes are partly driven by altered activity of lineage-specific transcription factors. Analysis of DNA methylation differences at transposable elements between young and old LEPs reveal that TEs can lose DNA methylation. Evolutionarily young transposons lose DNA methylation in a stochastic manner, presumably due to loss of fidelity of maintenance methylation. Moreover, aging dependent methylation loss at TEs correlates with elevated breast cancer risk.

Conclusions: The results indicate that aging leads to widespread alterations in the DNA methylation landscape of the genome, especially at young TEs. These alterations have the potential to rewire gene networks that may be important for age-associated breast cancers.

Keywords: transposable elements, aging, DNA methylation, human mammary 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]. 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 [2]. However, DNA methylation patterns are altered with age [3-6]. DNA methylome comparison of human CD4+ T cells from newborns and centenarians suggested an overall decrease in methylation levels with age [6]. Moreover, DNA methylation levels of specific CpG sites show a significant correlation with age across diverse tissues [4]. These sets of CpG sites known as 'epigenetic clocks' can accurately predict the chronological age of cells based on DNA methylation [4]. While it is clear that distinct DNA methylation changes to specific CpG sites occur with aging, the causes of these alterations and their impact on disease are still unclear.

Over 75% of breast cancers are diagnosed in women over the age of 50 [7]. Aging in the breast tissue is accompanied by increases in estrogen receptor expression [8], and significant physiological changes such as an increase in fat content and a decrease in connective tissue [9] that alter the microenvironment of the mammary epithelium. Previous studies suggest a complex interplay between the various epithelial cell types and microenvironment that leads to changes in molecular properties and the proportion of cell populations [10, 11]. Luminal epithelial cells (LEPs) and myoepithelial cells (MEPs) are the two principal lineages in breast epithelia, and the most striking age-related changes in abundance and phenotype occur among LEPs [10, 12]. Furthermore, aging-associated tissue microenvironment changes contribute to the observed age-related epigenetic and transcriptional changes in the mammary epithelium [13].

95 *In vitro* transformed human mammary epithelial cells (HMECs) from older women (> 55
96 y) exhibit gene expression patterns characteristic of luminal breast cancer subtypes. In contrast,
97 transformed HMECs from younger women (< 30 y) show more properties of basal breast cancers
98 [14]. This is consistent with the fact that 80% of age-related breast cancers are of the hormone-
99 receptor expressing luminal subtypes [15]. This suggests that cells from older women may
100 harbor age-related epigenetic states that make them predisposed to develop luminal breast
101 cancers. Intriguingly, the putative breast cancer cells of origin in older women express signaling
102 molecules consistent with those of immortalized luminal cells, suggesting that age-associated
103 changes can elevate cancer risk [12]. Epigenetic clock studies in breast tissue have suggested
104 that women with luminal breast cancers exhibit increased epigenetic age even in the adjacent
105 normal breast tissue [16]. Although earlier studies suggest that aging can lead to epigenetic
106 changes in breast tissue, thereby elevating breast cancer risk, a complete understanding of age-
107 associated DNA methylation changes in specific mammary cell types is lacking. Most previous
108 studies were performed either in bulk mammary tissue or interrogated a small subset of CpG
109 sites in the genome [16, 17].

110 Repetitive elements derived from transposable elements constitute roughly about ~50%
111 of the human genome and are classified based on their method of transposition [18]. Although
112 most elements have accumulated enough mutations and lost their ability to copy themselves,
113 they carry latent regulatory potential [19]. Some elements such as Alu, LINE and SVA elements
114 are still retrotransposition competent [20]. Cancer cells with global losses of DNA methylation
115 frequently exhibit activation of transposable elements [21, 22]. Interestingly, increased
116 expression of retrotransposons and active retrotransposition have been observed in aging
117 mouse tissues [23]. Moreover, LINE-1 expression is increased during cell senescence, a cell

118 state synonymous with aging [24]. We, therefore, reasoned that aging would lead to the loss of
119 epigenetic silencing of transposons in mammary luminal epithelial cells.

120 Here, we studied the genome-wide DNA methylation changes that occur with age in
121 healthy non-senescent luminal mammary epithelial cells. We show that aging leads to
122 methylation loss at specific transposable elements and methylation gain at CpG islands.
123 Methylation changes are partly driven by altered activity of lineage dependent transcription
124 factors. Young transposable elements mainly lose methylation in a stochastic manner,
125 presumably due to loss of fidelity of maintenance methylation. Finally, we show that aging
126 dependent methylation loss at transposable elements correlates with elevated breast cancer
127 risk.

128

129 **Results**

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

131 To investigate the genome-wide changes in DNA methylation associated with healthy
132 aging, we performed whole-genome bisulfite sequencing (WGBS-seq) on human mammary
133 LEPs from two young and two old individuals (**Figure 1A**). We isolated LEPs from primary
134 cultures of pre-stasis HMECs from discarded reduction mammoplasty (RM) tissue [10]. Tissues
135 were obtained from women with differing chronological ages but without breast cancer. These
136 pre-stasis LEPs from primary cultures maintain gene expression and DNA methylation profiles
137 that mirror those in tissue [13]. Moreover, these are non-senescent cells and therefore are a
138 valuable resource that can be used to interrogate aging dependent changes in the breast tissue
139 – distinct from those arising from *in vitro* cellular senescence.

140 We obtained high coverage (~9.4×) WGBS data across individual CpG sites and a
141 representation of about 90% of individual CpG sites in our samples (see Methods and Tables

142 S1, S2). Covered CpG sites had a slightly higher representation at CpG islands and lower
143 representation at repeat elements, as expected, given the difficulty of mapping to repetitive
144 regions of the genome (**Figure S1A**). Across repeat classes, SINE element CpGs had a higher
145 representation, whereas LINE CpGs were under-represented (**Figure S1B**), consistent with the
146 known mappability problems at young LINE elements [23]. Comparing the methylomes of young
147 and old individuals, we did not observe any change in the global methylation levels across the
148 genome (**Figure S1C, D**). However, principal component analysis (PCA) of the WGBS data
149 separated the old and the young samples on the first and second principal component axes
150 (35.95% and 33.1% of the respective cumulative variance; **Figure S1E**) indicating that there are
151 local differences in the methylomes.

152 To identify age-associated changes to DNA methylation, we identified CpGs differentially
153 methylated between the pooled young and old samples. Of the differentially methylated CpG
154 sites (n = 967,534, P-value < 0.05), a higher number of sites (n = 539,928) showed a decrease
155 in methylation (> 25% methylation change) in old LEPs (**Figure S1F**). To identify loci with many
156 CpG sites changing methylation, we identified differentially methylated regions (DMRs) (see
157 Methods) between the pooled young and old LEPs. We identified a total of 5,138 DMRs (>25%
158 methylation change and P-value < 0.05) with 3,149 DMRs losing methylation and 1,989 gaining
159 methylation in old LEPs (**Figure S1G**). The validity of identified DMRs was confirmed with
160 heatmaps showing methylation levels across samples (**Figure 1B**) as well as kernel smoothed
161 methylation plots of representative DMRs (**Figure S1H**).

162 We further characterized the locus of all DMRs by overlapping with common genomic
163 features (**Figure S1I**). ~48% of DMRs are intragenic and overlap introns (**Figure S1I**). DMRs
164 overlapping TEs (~36% of all DMRs, **Figure 1C, D**) and DMRs overlapping non-CGI and non-
165 repeat regions (59% of all DMRs, **Figure 1C, E**) both lose and gain DNA methylation, with more

166 DMRs losing DNA methylation. In contrast, more DMRs overlapping CpG islands gain
167 methylation (**Figure 1C, F**). These results suggest that specific loci, including repeat elements
168 and CpG islands, show a differential response to aging in LEPs.
169
170 **Aging leads to decreased methylation at myoepithelial-specific regulatory regions.**
171 As the principal component analysis indicated that there were methylation differences
172 between the two old individuals (**Figure S1E)**, we examined whether this could be due to
173 increased cellular heterogeneity associated with aging [25-28]. We first focused on finding aging-
174 specific signatures in LEPs. To do that, we classified the DMRs as 'shared DMRs' or 'variable
175 DMRs' based on whether they showed the same behavior in the two old individuals. Shared
176 DMRs are those that showed the most consistent changes in the old LEPs compared to young
177 LEPs (P-value < 0.05, methylation difference > 25%) (**Figure 2A-B**). Variable DMRs are those
178 that showed a change in only one of the two old LEPs when compared to the young LEPs
179 (**Figure 2C-D**). Presumably, the shared DMRs are aging-specific signatures that relate to the
180 response of LEPs to the aging breast tissue microenvironment. We further characterized the
181 genomic location of these DMRs and found that 1) shared DMRs have a higher representation
182 in genic regions, including promoters, exons, and introns, as well as on CpG islands (**Figure**
183 **2E**), and 2) variable DMRs have a higher representation of regions overlapping with
184 transposable elements, especially at SINE elements (**Figure 2E**). Moreover, although the shared
185 DMRs were enriched at genic regions, a majority of the shared and variable DMRs were located
186 away from transcription start sites (**Figure 2F**).
187 To understand the functional relevance of these distal DMR sites in the breast tissue, we
188 overlapped them with chromatin states predicted from published chromatin modification ChIP-
189 seq datasets performed in mammary epithelial lineages [29]. Our strategy of FACS enrichment

190 of luminal cells using PROM1 (CD133) positive and NGFR (CD271) negative cells includes
191 luminal progenitor cells [30]. Furthermore, previous studies have suggested that aging luminal
192 cells acquire basal-like characteristics [10, 11]. To discern whether DMRs reflect changes in cell
193 properties, we compared DMRs with chromatin states predicted in mature luminal, luminal
194 progenitor, and basal (MEP) cells from RM tissue of a young (19y old) donor [29]. We classified
195 DMRs into seven chromatin states in each cell type based on the extent of overlap with each
196 state (see Methods; **Figure S2A-C**). We observed that DMRs that lose methylation in LEPs
197 overlap primarily with basal/MEP-specific promoter and enhancer regions, accounting for about
198 68% of shared (**Figure 2G-H**) and 40% of variable DMRs (**Figure S2D-E**). For example, a DMR
199 with methylation loss at an alternative *ABR* isoform promoter (**Figure S2H**) is identified as a
200 promoter state only in basal cells. Since active enhancer sites are hypomethylated [31, 32],
201 these results indicate that basal/MEP-specific regulatory elements that lose DNA methylation
202 with age may be active enhancers in older LEPs. These data are in-line with previous studies
203 suggesting that aging luminal cells acquire basal-like characteristics [10, 11].
204 Conversely, DMRs that gain methylation overlap with promoter and enhancer regions in
205 luminal progenitor cells accounting for about 40% of shared (**Figure 2I-J**) and 32% of the
206 variable DMRs (**Figure S2F-G**). An example DMR with methylation gain at *LBX2-AS1* promoter
207 (**Figure S2I**) is identified as a promoter state in luminal progenitor and luminal cells. These
208 results indicate that luminal progenitor-specific regulatory regions gain DNA methylation in aging
209 luminal cells. Together, these results suggest that aging LEPs acquire permissive states at
210 basal-specific regions and repress luminal progenitor specific chromatin regions through
211 changes in DNA methylation.
212
213 **Lineage-specific transcription factors cause aging-associated DNA methylation changes.**

214 Given that DNA methylation changes at DMRs correlated with cell-type-specific chromatin
215 states, we examined the underlying DNA sequence at DMRs for signatures of TF binding. We
216 performed motif enrichment analysis using TF position weight matrices (PWMs) from the
217 JASPAR database [33]. Intriguingly, DMRs that lose methylation are enriched for binding sites
218 of TFs such as NFIA, EGR1, TEAD1, TEAD4, TP63, TP73, and SOX15 (**Figure 3A left panel:**
219 **shared DMRs and S3A left panel: variable DMRs**). We determined the expression of these
220 TFs in LEPs versus MEPs and in old LEPs using RNA-sequencing data. *TP63*, *NFIA,* and
221 *SOX15* showed increased expression in old LEPs compared to young LEPs (**Figure 3B-D**). In
222 addition, *TP63, NFIA, EGR1, TP73*, and *SOX15* all have relatively higher expression in MEPs
223 as compared to LEPs (**Figure 3A right panel and 3I**). The expression patterns of TFs and the
224 hypomethylation of their targets are concordant with known roles of these TFs and the changes
225 previously reported for old LEPs, which gain characteristics of MEPs. For example, young LEPs
226 do not express *TP63*, which is a myoepithelial lineage-specific transcription factor [13]. Similarly,
227 TEAD1 and TEAD4 are DNA binding partners for YAP, a component of the Hippo signaling
228 pathway. Old LEPs have been shown to exhibit increased nuclear YAP localization [11],
229 indicating increased YAP activity and TEAD binding. These results suggest that increased
230 expression of myoepithelial-specific TFs in old LEPs could lead to increased binding and loss of
231 DNA methylation in these regions (**Figure 3J**).

232 Conversely, DMRs with methylation gain showed enrichment for binding sites of
233 transcription factors EHF, ELF3, ELF5, JUND, FOSL2, and ATF3 (**Figure 3E right panel:**
234 **shared DMRs and S3B right panel: variable DMRs**). ELF5 is a luminal lineage-specific
235 transcription factor that shows a decrease in expression with age (**Figure 3F**) [13]. Besides
236 *ELF5*, we found that transcription factors *ELF3, SOX4* and *JUND* show relatively lower
237 expression in MEPs as compared to LEPs (**Figure 3E right panel and 3I**). Expression of *SOX4*

238 and *JUND* also decreases with age in LEPs, indicating a loss of expression of these lineage-
239 specific TFs (**Figure 3F-H**). These results suggest that decreased expression of luminal lineage-
240 specific TFs may lead to a decrease in binding and protection of binding sites on DNA from
241 DNMTs. This might result in DNA methylation and repression of these luminal lineage-specific
242 regions (**Figure 3J**). We further note that altered TF activity due to post-translational
243 modifications could occur without any changes to mRNA expression. Given our observation that
244 the differentially methylated regions show enrichment of binding sites for lineage-specific TFs
245 such as TP63 and ELF5, we further examined whether the other TFs whose binding sites are
246 enriched on DMRs also show expression in a lineage-specific manner. We found that both in the
247 case of shared and variable DMRs, sites of methylation loss were enriched for binding sites of
248 myoepithelial specific TFs (**Figure 3I: shared DMRs and S3C: variable DMRs**). Sites of
249 methylation gain were enriched for binding sites of luminal lineage-specific TFs (**Figure 3I and**
250 **S3C**). Collectively, these results suggest that aging-associated changes in luminal epithelial cells
251 are orchestrated at least in part by lineage-specific transcription factors.
252
253 **Aging leads to hypomethylation at young transposable elements.**
254 Given that DNA methylation is an important mechanism for suppression of transposable
255 elements (TEs), we asked whether loss of DNA methylation with age occurs at TEs. We
256 observed that TEs were indeed present in DMRs, and the TE content was significantly different
257 than expected for random regions of the same length (**Figure S4A and B**). In general, shared
258 DMRs had a lower TE content than variable DMRs (**Figure S4A and B**), indicating that age-
259 dependent methylation differences at TEs are mainly stochastic.
260 We further investigated the characteristics of TEs within differentially methylated regions.
261 We found that the evolutionarily younger elements (<100 million years) tend towards DNA

262 methylation loss, whereas the older TE elements (>100 million years) tend to gain methylation
263 (**Figure 4A: variable DMRs and S4C: shared DMRs**). Separating by TE class, SINE elements
264 showed a distribution similar to all TEs, indicating their relative abundance in the DMRs (**Figure**
265 **4B: variable DMRs and S4D: shared DMRs**). Young LINE elements lose methylation variably
266 in the old LEPs (**Figure 4C**). However, older LINE elements in shared and variable DMRs
267 showed both loss and gain of methylation but tended more towards methylation gain (**Figure**
268 **4C: variable DMRs** and **Figure S4E: shared DMRs**). Interestingly, young LTR/ERV elements
269 were present in both shared and variable DMRs with methylation loss and gain (**Figure 4D:**
270 **variable DMRs and S4F: shared DMRs**), indicating possible exaptation of these elements in
271 LEPs.

272 We further inspected the specific subfamilies enriched at variable and shared DMRs with
273 methylation loss and gain. As some TE subfamilies, especially Alu elements, are much more
274 abundant than others, we calculated the expected number of overlaps for each TE subfamily
275 using random genomic regions. To be conservative, TE subfamilies that showed an
276 observed/expected ratio closer to 1 (0.75 - 1.25) were discarded (**Figure S4G-I**). Among the
277 variable DMRs, younger Alu subfamilies were highly abundant in regions with methylation loss
278 whereas older SINE subfamilies MIRb, and MIR were highly enriched at regions with methylation
279 gain (**Figure 4E**). Young LINE subfamilies L1PA2-7, and L1PB1 lost methylation, whereas older
280 LINE elements L2a, L2b, and L2 showed methylation loss as well as gain (**Figure 4F**). Young
281 LTR/ERV subfamilies LTR13, LTR12C, THE1B, and MSTA, were observed at regions of
282 methylation loss, whereas, HERVH elements were enriched at regions of methylation loss as
283 well as methylation gain. Older LTR subfamilies MER52A, MER52D, and MLT1D also showed
284 methylation loss (**Figure 4G**). Furthermore, shared DMRs showed enrichment of two LTR
285 subfamilies HERVH and MER52A that lost methylation (**Figure S4J**). Next, we determined

286 whether the hypomethylated TEs harbor binding sites of transcription factors that could explain
287 the methylation loss at these sites. We performed motif scanning using TF PWMs from the
288 JASPAR database [33] on the hypomethylated regions of TEs. NF-I binding sites are prevalent
289 across all hypomethylated TE subfamilies (**Figure 4H**). Hypomethylated AluSp elements
290 showed an abundance of TP73 sites, whereas hypomethylated HERVH elements showed a
291 prevalence of TEAD motifs and NF-κB (REL, RELA) motifs (**Figure 4H and S4K**). DMRs in
292 MER52A elements showed the presence of EGR1, TCF12, and NF-I motifs (Figure 4H). Binding
293 of these transcription factors that are active in old LEPs (**Figure 3I**) could result in DNA
294 methylation loss at these sites through the recruitment of ten eleven translocation (TET)
295 enzymes. These demethylated TEs could, in turn, function as regulatory elements and influence
296 the expression of nearby genes. An example locus of a HERVH element located upstream
297 (~5kb) from the *COX7A1* gene is shown that loses methylation with age (**Figure 4I**). *COX7A1*
298 mRNA expression increases with age, indicating that the upstream HERVH element could
299 function as a regulatory element for this gene (**Figure 4J**). Interestingly, *COX7A1* is highly
300 upregulated in the Luminal A subtype of breast cancer (**Figure S4L**). However, we note that TF
301 binding sites explained only about 20-70% of the hypomethylation events at TEs (**Figure 4H**).
302 These results indicate that young TEs lose methylation with age, and TF binding events can only
303 partially explain these changes.
304

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

306 Aging is associated with increased transcriptional variability and cell-to-cell heterogeneity
307 [25-28]. As we observed increased variation in CpG methylation between the two old individuals,
308 we examined whether there is an increase in methylation heterogeneity with age in older
309 individuals. For this, we calculated methylation entropy, a measure of methylation heterogeneity

310 for 150 bp non-overlapping windows across the genome [34]. We did not find any global changes
311 in methylation entropy in the old individuals across the genome (**Figure S5A-B**).

312 Since methylation changes could occur through changes in methylation heterogeneity,
313 we asked whether the DMRs show changes in methylation entropy. Indeed, we observed that
314 regions with methylation loss showed a net increase in entropy (**Figure 5A-B and S5C-D)**,
315 whereas regions with methylation gain showed a decrease in entropy (**Figure 5C-D and S5E-**
316 **F**). These results indicate that concerted methylation changes are brought about by DNA
317 methyltransferases (DNMTs) at regions that gain methylation. Since the variable DMRs
318 represent regions with differing methylation status in the two old individuals, we reasoned that
319 loss of maintenance by DNMT1 could lead to increases in methylation heterogeneity
320 accompanied by methylation loss. Similar loss of methylation due to inefficient maintenance
321 methylation by DNMT1 has been reported at late replicating regions associated with
322 heterochromatin [35]. Given that TEs are more likely to be associated with these regions and
323 variable DMRs have a higher TE content, we investigated whether variable DMRs overlapping
324 TEs exhibit an increase in methylation entropy. Results showed that TE containing DMRs with
325 methylation loss have increased methylation entropy, suggesting a stochastic loss of methylation
326 (**Figure 5E-F**). In contrast, CpG islands with methylation loss do not show net changes in entropy
327 (**Figure 5G-H**).

328 We further noted that CpG dense transposons such as L1PA2, LTR12C, and AluY show
329 variable methylation loss in the old individuals (**Figure 4E-G**). We investigated whether the loss
330 of methylation at these elements is stochastic. Indeed, we found an increase in methylation
331 entropy associated with methylation loss at L1PA2, LTR12C, and AluY elements in the old
332 individuals (**Figure 5I-K**). Moreover, the methylation loss occurred in the CpG dense region of
333 the element (**Figure 5L-N**). For example, the 5'UTR of L1PA2 elements contain a CpG island-

like dense region, and the methylation loss was focused at this region (**Figure 5L-N**). These results indicate that young CpG dense transposons exhibit stochastic loss of methylation with age.

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

Given that CpG dense transposons lose methylation with age, we determined whether this might lead to increased breast cancer risk. To this end, we examined DNA methylation profiles generated from reduced representation bisulfite sequencing (RRBS) approach with additional LEPs from young (n = 6) and old (n = 4) individuals and individuals with a higher-than-average risk for breast cancer (n = 7) (Table S3). We reasoned that since RRBS enriches CpG dense regions of the genome, we would be able to examine methylation of CpG dense transposons in more individuals. Indeed, we obtained a higher representation of CpG sites from CpG-dense SVA class of transposons and CpG islands (**Figure S5G: upper panel**). RRBS was able to capture more than 50% of copies for the CpG dense LTR12 and SVA subfamilies and younger LINE subfamilies (**Figure S5G: lower panel**). We obtained a higher or equivalent coverage at CpG dense regions of SVA_D and LTR12C elements as compared to WGBS (**Figure S5H**). Next, we identified TE containing DMRs with methylation loss in the old individuals. We used this set to ask whether high-risk LEPs also show methylation loss at these TEs. PCA and hierarchical clustering analysis on the hypomethylated TE-DMRs showed that indeed high-risk LEPs cluster together with the old LEPs (**Figure 6A-B**). This indicates that age-dependent methylation loss at TEs are recapitulated in LEPs from high-risk individuals. TEs have latent regulatory potential and, when derepressed, can function as regulatory elements. Given that age is a risk factor for luminal breast cancer, we asked whether TEs hypomethylated with

358 age play any role in luminal breast cancer. To examine the potential function of the underlying
359 DNA, we identified chromatin states at these TEs from published breast cancer cell line ChIP-
360 seq datasets [36]. The results revealed that hypomethylated CpG-dense TEs exist in a bivalent
361 promoter state with H3K4me3 and H3K27me3 marks (**Figure 6C**). Hypomethylated L1HS and
362 L1PA2 elements exist in a bivalent promoter state in luminal and HER2-positive cell lines (**Figure**
363 **6C**), whereas bivalent state at SVA_D loci was observed only in HER2-positive cell lines (**Figure**
364 **6C**). These results indicate that these TEs exhibit a poised promoter state in luminal breast
365 cancer cell lines and may function as regulatory elements. To further support this idea, we
366 analyzed published RNA-seq data from normal and ductal carcinoma in situ (DCIS) samples
367 [37]. We identified genes upregulated in DCIS samples compared to normal and present within
368 50kb of a hypomethylated TE identified from the WGBS and RRBS data (**Figure 6D**). We
369 overlapped this set of genes with the curated gene sets available in the MSigDB database [38].
370 We found that most of these genes were involved in luminal breast cancer etiology and
371 distinguish luminal breast cancers from basal breast cancers (**Figure 6E**). For example, the
372 *CYB5A* gene increases expression in DCIS samples (**Figure 6F**) and is highly expressed in
373 luminal breast tumors (Luminal A and Luminal B) as compared to basal tumors (**Figure 6G**).
374 Genome browser tracks of methylation levels at the *CYB5A* locus show decreased methylation
375 at an upstream HERVH element (**Figure 6H**). These results indicate that TEs hypomethylated
376 with age can regulate the expression of genes important for luminal breast cancer. Thus, age
377 hypomethylated TEs can predict breast cancer risk.
378

379 **Discussion**

380 Here we report a comprehensive analysis of the aging-associated DNA methylation
381 alterations in primary human mammary luminal epithelial cells. This is the first study to use a

382 WGBS approach to examine the impact of age on the DNA methylome of primary mammary
383 epithelial cells. Previous studies were done using array-based technologies that cover only a
384 fraction of the CpG sites in the genome [13, 16, 17]. Importantly, repeat elements are under-
385 represented in these array probes [39], making interpretation of DNA methylation at repetitive
386 regions difficult. Moreover, earlier DNA methylation studies have been performed using
387 composite tissues with a mixture of cells [16, 17]. We and others have shown that different cell
388 lineages have unique responses to aging, and this is dependent on cell-cell interactions and
389 microenvironment changes [13, 40]. Furthermore, aging-related changes in cell composition
390 have been shown in blood cells where older individuals show a skew towards myeloid lineages
391 [41]. Changes in cell composition can confound the interpretation of DNA methylation levels in
392 bulk tissues. Thus, it is imperative to study age-dependent molecular changes in purified cell
393 populations.

394 Previous DNA methylome studies in human cells have suggested a global loss of
395 methylation with age [6, 34, 42]. These studies were performed using $CD4^+$ T cells from a
396 neonate and centenarian [6], sorted $CD4^+$ T cells [34] and skin keratinocytes [42] from young
397 (<25 y) and old (>70 y) individuals. Furthermore, the global methylation loss with age was
398 accompanied by a global increase in entropy or methylation discordance with age [34]. We did
399 not observe a global methylation loss or a global increase in entropy in mammary luminal
400 epithelial cells with age. This could be due to a higher fidelity of DNA methylation maintenance
401 in luminal cells or due to a smaller separation between the age of young (<30 years) and old (>
402 55 years) individuals in our study. This observation is supported by DNA methylome studies in
403 aging mouse livers, where no global loss of methylation was observed, although specific
404 changes at regulatory elements were detected [43]. Whereas no global methylation loss was
405 detected, we did, however, find methylation increase at CpG islands. Global loss of methylation

406 and focal increase in methylation at CpG islands have been reported to occur as part of a
407 phenomenon known as ‘epigenetic drift’ with an increase in age [44].

408 DNA methylation is an important mechanism for silencing of transposable elements.
409 Reactivation of transposable elements has been shown in cancer cells with global loss of DNA
410 methylation [21, 22]. While previous research comparing DNA methylation levels between a
411 neonate and centenarian DNA reported that 12% of DMRs that lost methylation with age were
412 located on Alu elements [6], changes in epigenetic silencing at TEs with aging has not previously
413 been systematically investigated. Increased expression of retrotransposons and active
414 retrotransposition have been reported in aging mouse tissues [23]. Furthermore, increased
415 LINE-1 expression has been reported in senescent fibroblasts [24]. However, the role of DNA
416 methylation patterns in these systems was not examined. We examined the impact of aging on
417 DNA methylation at TEs using non-senescent mammary epithelial cells. We show that the
418 evolutionary age of TEs is important for determining the response to aging. As older TEs are
419 exapted and function as tissue-specific regulatory elements [19], aging-associated changes can
420 be attributed to cell-type-specific transcription factor activity. This observation may help explain
421 why cancers tend to be tissue-specific. However, young CpG-dense transposons are more likely
422 to be silenced by DNA methylation and presumably lose methylation by inefficient maintenance
423 methylation by DNMT1. Heterochromatin loci that occur in late replicating regions lose
424 methylation by a similar mechanism in cancer cells [35].

425 We further show that methylation loss at TEs with age is recapitulated in luminal cells
426 from high-risk individuals. Moreover, we show that these TEs are present as bivalent promoter
427 loci in luminal breast cancer cell lines indicating their poised regulatory potential as opposed to
428 being completely silent in normal cells. In addition, we showed that hypomethylated TEs could
429 potentially regulate genes associated with the luminal subtype of breast cancer. Altogether,

430 these observations provide further support to the observation that epigenetic changes with age
431 could determine risk for age-related breast cancer. We also provide evidence for the potential
432 role of TE derepression in luminal breast cancer. Future studies would be needed to study the
433 role of these TEs in the early events of mammary tumor initiation.
434

435 **Methods**

436 **Isolation of mammary luminal epithelial cells**

437 Primary HMECs were generated and maintained, as described previously [45]. Cells were grown
438 in M87A medium with cholera toxin and oxytocin at 0.5 ng/ml and 0.1nM, respectively. Cells
439 were stained with anti-human CD133-PE (Biolegend, clone 7) and anti-human CD271
440 (Biolegend, clone ME20.4) by following standard flow cytometry protocol. LEPs and MEPs were
441 enriched with Aria III (Becton Dickinson) as $CD133^{+}/CD271^{-}$ and $CD133^{-}/CD271^{+}$ fractions,
442 respectively.

443 **WGBS**

444 Quick-DNA Microprep kit (Zymo Research) was used to extract DNA. Whole-genome bisulfite
445 sequencing (WGBS-seq) was performed at Admera Health Biopharma Services. Genomic DNA
446 was bisulfite converted using the Zymo EZ DNA methylation kit (Zymo Research). Library
447 preparation was done using TruSeq DNA Methylation Kit with 75ng of bisulfite converted DNA
448 as per manufacturer's recommendations. 150bp paired-end sequencing was performed on a
449 single lane of Illumina Hiseq X Ten. We obtained about 38X coverage per genome to an average
450 of 414 million paired end reads (Table S1).

451 Sequencing reads were hard trimmed 8 bp from both ends using Trim Galore
452 (https://github.com/FelixKrueger/TrimGalore) (version 0.4.5) to remove unwanted methylation
453 bias arising from the library preparation. Adapter contamination was also removed using Trim

454 Galore using --illumina option. Trimmed reads were aligned to the human reference genome
455 (hg19) using Bismark (version 0.19.1) [46] with the --bowtie1 option [47]. Bismark
456 (deduplicate_bismark) to remove PCR duplicates. Methpipe [48] (version 3.4.3) was used to
457 determine the methylation status of individual CpG sites. The hg19 reference genome contains
458 about 56.4 million CpG loci (hg19) or 28.2 million CpG dyads. Methylation levels were calculated
459 for individual CpG dyads and are henceforth referred to as CpG sites. We achieved sufficient
460 coverage of 91.3% and 90.4% of the CpG sites in young and old samples, respectively, with a
461 mean coverage of 9.4 methylation calls per site.

462 **DMR identification**

463 For identifying differentially methylated CpG sites, Methpipe [48] was used (Figure 1B), Only
464 CpG sites with coverage > 3× in pooled samples were considered. CpG sites with greater than
465 25% change in methylation levels (*P*-value < 0.05) were considered as differentially methylated.
466 DMRfinder [49] was used to identify DMRs. Briefly, CpG sites with at least 4X coverage in all
467 four samples were determined. Genomic regions with at least 4 CpG sites within a distance of
468 100 bp were identified. The maximum length of regions was set to 500 bp. Of these regions,
469 ones that had less than 20 reads covering them were discarded. The final set of regions was
470 then used to calculate differentially methylated regions with greater than 25% change (*P*-value
471 < 0.05) between pooled young and old samples (Figure 1C, D). For variable DMRs, each old
472 sample was compared with the pooled young samples to determine methylation changes.
473 Regions with greater than 25% change (*P*-value < 0.05) were identified for both old individuals.
474 A union set of regions with loss and gain of methylation were classified as variable DMRs.

475 **Characterization of DMRs**

476 Overlaps with genomic features were calculated using bedtools intersect [50] with masked CpG
477 islands from the UCSC genome browser [51] and repeats from the RepeatMasker annotation

478 [52]. DMR overlaps with CpG islands, and repeat elements were determined using bedtools
479 intersect such that the overlapping region contains at least 1 CpG site. In case a DMR overlaps
480 both a CpG island and a repeat element, the DMR was assigned to the element contributing
481 higher number of CpG sites.
482 Kernel smoothed plots were made using the BSSeq package with default smoothing parameters
483 and CpG sites with greater than 3X coverage [53]. Heatmaps were made using the pheatmap
484 package [54].

485 **Chromatin state predictions**

486 ChIP-seq data for mammary epithelial populations were downloaded using the International
487 Human Epigenome Consortium data portal (https://epigenomesportal.ca/ihec/). Bigwig files
488 containing ChIP-seq signal for six different histone modifications in three cell types (sample IDs:
489 CEMT0154, CEMT0155, and CEMT0156) were used (**Figure S2A-C**). ChIP-seq signals from
490 the bigwig files were binned into 200 bp windows. These binned signal files were then binarized
491 using BinarizeSignal in ChromHMM (version 1.20) [55]. The binarized files were used to run
492 ChromHMM LearnModel with 16 states. Functionally similar chromatin states were grouped
493 together to seven distinct chromatin states (**Figure S2C**). ChIP-seq data for breast cancer cell
494 lines were downloaded from GEO (Accession number: GSE85158 [36]). Aligned bed files were
495 used to generate binarized files using BinarizeBed with default options. The binarized data were
496 used to run ChromHMM LearnModel with 18 states. Functionally similar chromatin states were
497 grouped together to nine distinct chromatin states. DMR overlaps with chromatin states were
498 performed using bedtools intersect. In cases where a DMR overlapped more than one chromatin
499 state, DMR was assigned to the state with higher base overlap.

500 **RNA-seq analysis**

501 Total RNA was isolated using Quick-RNA Microprep kit (Zymo Research) from FACS sorted
502 young (n = 10) and old LEPs (n = 8). Similarly, total RNA was isolated from young (n = 6) MEPs.
503 PolyA-enriched RNA was isolated and used for library preparation using the TruSeq RNA Library
504 Prep kit (Illumina). About 50 million single-end reads of length 51 bp were generated for all
505 samples. Raw sequences were aligned to the hg19 reference genome using HISAT2 (version
506 2.1.0) [56] with the --dta parameter. Stringtie (version 1.3.4) [57] was used with default
507 parameters to assemble transcripts using the Gencode v19 transcript annotation. Assembled
508 transcripts from all libraries were further merged using --merge option in Stringtie. Merged
509 transcript abundances were measured using bedtools coverage.
510 For the old LEP versus young LEP comparison, edgeR (v3.26.4) [58] was used to conduct
511 differential gene expression analysis. Briefly, normalization factors were calculated using the
512 trimmed mean of M values method (TMM) to eliminate composition biases between libraries.
513 Genewise biological dispersion was estimated using edgeR's empirical Bayes approach. Then,
514 we employed a generalized linear model (GLM) approach to account for the unnormalized
515 distribution of gene counts, as well as additional factors, such as batch effects. Differentially
516 regulated genes were identified with an absolute log2 fold change $\geq$ 0.5 and FDR < 0.05. Gene
517 expression boxplots were generated using batch corrected RPKM values. Batch correction of
518 RPKM values was done using limma package [59]. For LEP versus MEP comparison, DESeq2
519 [60] was used to normalize counts and identify differentially expressed genes (absolute log2 fold
520 change $\geq$ 0.5 and p-value < 0.05).
521 RNA-seq data from normal and DCIS samples were obtained from GEO (Accession number:
522 GSE69240 [37]). Fastq reads were quality and adapter trimmed using Trim Galore with the --
523 paired option. Quality trimmed reads were aligned to the hg19 genome using STAR (version
524 2.6.0.a) [61] with --sjdbOverhang 100 --winAnchorMultimapNmax 100 --outFilterMultimapNmax

525 100 --outFilterMismatchNmax 3 --alignEndsType EndToEnd options. DESeq2 package [60] was
526 used to normalize counts and identify differentially expressed genes (absolute log2 fold change
527 ≥ 0.5 and p-value < 0.05).

528 **DMR overlaps with transposable elements**

529 The extent of overlap of DMRs and random regions with TEs (RepeatMasker annotation) was
530 calculated using Genomic Association Tester (GAT) [62]. Random regions with the same size
531 and GC content distribution as DMRs were selected from ungapped regions of hg19 genome.
532 Number of DMRs overlapping TEs and the number of nucleotide bases overlapping TEs was
533 calculated. Simulations with random regions was calculated 10,000 times. 95% confidence
534 intervals for the number of overlaps with random regions (expected) were plotted with the
535 observed number of overlaps (Figures S4A and S4B). Empirical *P*-values are reported for 10000
536 simulations. Multiple testing correction was performed using Benjamini-Hochberg method.

537 **RRBS**

538 Quick-DNA Microprep kit (Zymo Research) was used to extract DNA and bisulfite converted
539 DNA was used to prepare libraries for RRBS (Table S3). 100bp paired-end sequencing was
540 performed on Illumina Hiseq 2500 (Table S3). Sequencing reads were adapter trimmed as well
541 as clipped 4 bp on each end using Trim Galore (version 0.4.5) using –illumina --rrbs --trim1
542 options. Trimmed reads were aligned to the human reference genome (hg19) using Bismark
543 (version 0.19.1) [46] with the --bowtie1 option [47]. Methpipe [48] (version 3.4.3) was used to
544 determine the methylation status of individual CpG sites. DMRs were identified using a similar
545 approach as WGBS.

546 **Additional data analysis**

547 Motif scanning was performed using fimo [63] with JASPAR 2020 database [33]. Motif
548 enrichment analysis was performed using homer [64]. The evolutionary age of TEs was

549 calculated as described previously using the Jukes-Cantor correction [65]. Normalized
550 methylation entropy and mean methylation were calculated on 150bp non-overlapping windows
551 across the genome using informME [66]. Figures were made in R (version 3.51) using the
552 ggplot2 package [67] and assembled in Adobe Illustrator.
553

554 **Declarations**

555 **Consent for publication**

556 Not applicable

557 **Availability of data and materials**

558 All raw and processed sequencing data files have been deposited at NCBI's Gene Expression
559 Omnibus (GEO) under accession GSE153309.

560 **Competing interests**

561 The authors declare that they have no competing interests.

562 **Funding**

563 This work was supported by the National Institutes of Health, grants R01DK112041,
564 R01CA220693 (D.E.S.), and Department of Defense/Army Breast Cancer Era of Hope Scholar
565 Award BC141351, and Expansion Award BC181737, Conrad N. Hilton Foundation, and City of
566 Hope Center for Cancer and Aging (M.A.L.), and the National Cancer Institute (NCI) Cancer
567 Metabolism Training Program Postdoctoral Fellowship T32CA221709 (R.W.S.). City of Hope
568 Program in Molecular and Cellular Biology pilot award (D.E.S and M.A.L.). The research
569 reported in this publication included work performed in the Integrative Genomics and Analytical
570 Cytometry Cores supported by the National Cancer Institute of the National Institutes of Health
571 under award number P30CA033572. The content is solely the responsibility of the authors and
572 does not necessarily represent the official views of the National Institutes of Health. The funding

bodies had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

597 **References**

598 1. Lopez-Otin C, Blasco MA, Partridge L, Serrano M, Kroemer G: **The hallmarks of aging.** *Cell*
599 2013, **153:**1194-1217.
600 2. Bird A: **DNA methylation patterns and epigenetic memory.** *Genes Dev* 2002, **16:**6-21.
601 3. Maegawa S, Hinkal G, Kim HS, Shen L, Zhang L, Zhang J, Zhang N, Liang S, Donehower LA,
602 Issa JP: **Widespread and tissue specific age-related DNA methylation changes in mice.**
603 *Genome Res* 2010, **20:**332-340.
604 4. Horvath S: **DNA methylation age of human tissues and cell types.** *Genome Biol* 2013,
605 **14:**R115.
606 5. Rakyan VK, Down TA, Maslau S, Andrew T, Yang TP, Beyan H, Whittaker P, McCann OT,
607 Finer S, Valdes AM, et al: **Human aging-associated DNA hypermethylation occurs**
608 **preferentially at bivalent chromatin domains.** *Genome Res* 2010, **20:**434-439.
609 6. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, Diez J, Sanchez-Mut JV, Setien F,
610 Carmona FJ, et al: **Distinct DNA methylomes of newborns and centenarians.** *Proc Natl Acad*
611 *Sci U S A* 2012, **109:**10522-10527.
612 7. Jemal A, Siegel R, Ward E, Murray T, Xu J, Thun MJ: **Cancer statistics, 2007.** *CA Cancer J*
613 *Clin* 2007, **57:**43-66.
614 8. Quong J, Eppenberger-Castori S, Moore D, 3rd, Scott GK, Birrer MJ, Kueng W, Eppenberger U,
615 Benz CC: **Age-dependent changes in breast cancer hormone receptors and oxidant stress**
616 **markers.** *Breast Cancer Res Treat* 2002, **76:**221-236.
617 9. Benz CC: **Impact of aging on the biology of breast cancer.** *Crit Rev Oncol Hematol* 2008,
618 **66:**65-74.
619 10. Garbe JC, Pepin F, Pelissier FA, Sputova K, Fridriksdottir AJ, Guo DE, Villadsen R, Park M,
620 Petersen OW, Borowsky AD, et al: **Accumulation of multipotent progenitors with a basal**
621 **differentiation bias during aging of human mammary epithelia.** *Cancer Res* 2012, **72:**3687-
622 3701.
623 11. Pelissier FA, Garbe JC, Ananthanarayanan B, Miyano M, Lin C, Jokela T, Kumar S, Stampfer
624 MR, Lorens JB, LaBarge MA: **Age-related dysfunction in mechanotransduction impairs**
625 **differentiation of human mammary epithelial progenitors.** *Cell Rep* 2014, **7:**1926-1939.
626 12. Pelissier Vatter FA, Schapiro D, Chang H, Borowsky AD, Lee JK, Parvin B, Stampfer MR,
627 LaBarge MA, Bodenmiller B, Lorens JB: **High-Dimensional Phenotyping Identifies Age-**
628 **Emergent Cells in Human Mammary Epithelia.** *Cell Rep* 2018, **23:**1205-1219.
629 13. Miyano M, Sayaman RW, Stoiber MH, Lin CH, Stampfer MR, Brown JB, LaBarge MA: **Age-**
630 **related gene expression in luminal epithelial cells is driven by a microenvironment made**
631 **from myoepithelial cells.** *Aging (Albany NY)* 2017, **9:**2026-2051.
632 14. Lee JK, Garbe JC, Vrba L, Miyano M, Futscher BW, Stampfer MR, LaBarge MA: **Age and the**
633 **means of bypassing stasis influence the intrinsic subtype of immortalized human mammary**
634 **epithelial cells.** *Front Cell Dev Biol* 2015, **3:**13.
635 15. Jenkins EO, Deal AM, Anders CK, Prat A, Perou CM, Carey LA, Muss HB: **Age-specific**
636 **changes in intrinsic breast cancer subtypes: a focus on older women.** *Oncologist* 2014,
637 **19:**1076-1083.
638 16. Hofstatter EW, Horvath S, Dalela D, Gupta P, Chagpar AB, Wali VB, Bossuyt V, Storniolo AM,
639 Hatzis C, Patwardhan G, et al: **Increased epigenetic age in normal breast tissue from luminal**
640 **breast cancer patients.** *Clin Epigenetics* 2018, **10:**112.
641 17. Castle JR, Lin N, Liu J, Storniolo AMV, Shendre A, Hou L, Horvath S, Liu Y, Wang C, He C:
642 **Estimating breast tissue-specific DNA methylation age using next-generation sequencing**
643 **data.** *Clin Epigenetics* 2020, **12:**45.

644 18. Cordaux R, Batzer MA: **The impact of retrotransposons on human genome evolution.** *Nat*
645 *Rev Genet* 2009, **10:**691-703.
646 19. Sundaram V, Cheng Y, Ma Z, Li D, Xing X, Edge P, Snyder MP, Wang T: **Widespread**
647 **contribution of transposable elements to the innovation of gene regulatory networks.**
648 *Genome Res* 2014, **24:**1963-1976.
649 20. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvak
650 Z, Levin HL, Macfarlan TS, et al: **Ten things you should know about transposable elements.**
651 *Genome Biol* 2018, **19:**199.
652 21. Burns KH: **Transposable elements in cancer.** *Nat Rev Cancer* 2017, **17:**415-424.
653 22. Ishak CA, De Carvalho DD: **Reactivation of Endogenous Retroelements in Cancer**
654 **Development and Therapy.** *Annual Review of Cancer Biology*, **4**.
655 23. De Cecco M, Criscione SW, Peterson AL, Neretti N, Sedivy JM, Kreiling JA: **Transposable**
656 **elements become active and mobile in the genomes of aging mammalian somatic tissues.**
657 *Aging (Albany NY)* 2013, **5:**867-883.
658 24. De Cecco M, Ito T, Petrashen AP, Elias AE, Skvir NJ, Criscione SW, Caligiana A, Brocculi G,
659 Adney EM, Boeke JD, et al: **L1 drives IFN in senescent cells and promotes age-associated**
660 **inflammation.** *Nature* 2019, **566:**73-78.
661 25. Hernando-Herraez I, Evano B, Stubbs T, Commere PH, Jan Bonder M, Clark S, Andrews S,
662 Tajbakhsh S, Reik W: **Ageing affects DNA methylation drift and transcriptional cell-to-cell**
663 **variability in mouse muscle stem cells.** *Nat Commun* 2019, **10:**4361.
664 26. Bahar R, Hartmann CH, Rodriguez KA, Denny AD, Busuttil RA, Dolle ME, Calder RB,
665 Chisholm GB, Pollock BH, Klein CA, Vijg J: **Increased cell-to-cell variation in gene**
666 **expression in ageing mouse heart.** *Nature* 2006, **441:**1011-1014.
667 27. Cheung P, Vallania F, Warsinske HC, Donato M, Schaffert S, Chang SE, Dvorak M, Dekker CL,
668 Davis MM, Utz PJ, et al: **Single-Cell Chromatin Modification Profiling Reveals Increased**
669 **Epigenetic Variations with Aging.** *Cell* 2018, **173:**1385-1397 e1314.
670 28. Martinez-Jimenez CP, Eling N, Chen HC, Vallejos CA, Kolodziejczyk AA, Connor F, Stojic L,
671 Rayner TF, Stubbington MJT, Teichmann SA, et al: **Aging increases cell-to-cell**
672 **transcriptional variability upon immune stimulation.** *Science* 2017, **355:**1433-1436.
673 29. Pellacani D, Bilenky M, Kannan N, Heravi-Moussavi A, Knapp D, Gakkhar S, Moksa M, Carles
674 A, Moore R, Mungall AJ, et al: **Analysis of Normal Human Mammary Epigenomes Reveals**
675 **Cell-Specific Active Enhancer States and Associated Transcription Factor Networks.** *Cell*
676 *Rep* 2016, **17:**2060-2074.
677 30. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T,
678 Partanen A, et al: **Aberrant luminal progenitors as the candidate target population for basal**
679 **tumor development in BRCA1 mutation carriers.** *Nat Med* 2009, **15:**907-913.
680 31. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C,
681 Oakeley EJ, Gaidatzis D, et al: **DNA-binding factors shape the mouse methylome at distal**
682 **regulatory regions.** *Nature* 2011, **480:**490-495.
683 32. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z,
684 Ngo QM, et al: **Human DNA methylomes at base resolution show widespread epigenomic**
685 **differences.** *Nature* 2009, **462:**315-322.
686 33. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, Modi BP,
687 Correard S, Gheorghe M, Baranasic D, et al: **JASPAR 2020: update of the open-access**
688 **database of transcription factor binding profiles.** *Nucleic Acids Res* 2020, **48:**D87-D92.
689 34. Jenkinson G, Pujadas E, Goutsias J, Feinberg AP: **Potential energy landscapes identify the**
690 **information-theoretic nature of the epigenome.** *Nat Genet* 2017, **49:**719-729.

691 35. Zhou W, Dinh HQ, Ramjan Z, Weisenberger DJ, Nicolet CM, Shen H, Laird PW, Berman BP:
692 **DNA methylation loss in late-replicating domains is linked to mitotic cell division.** *Nat*
693 *Genet* 2018, **50:**591-602.
694 36. Franco HL, Nagari A, Malladi VS, Li W, Xi Y, Richardson D, Allton KL, Tanaka K, Li J,
695 Murakami S, et al: **Enhancer transcription reveals subtype-specific gene expression**
696 **programs controlling breast cancer pathogenesis.** *Genome Res* 2018, **28:**159-170.
697 37. Abba MC, Gong T, Lu Y, Lee J, Zhong Y, Lacunza E, Butti M, Takata Y, Gaddis S, Shen J, et
698 al: **A Molecular Portrait of High-Grade Ductal Carcinoma In Situ.** *Cancer Res* 2015,
699 **75:**3980-3990.
700 38. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP:
701 **Molecular signatures database (MSigDB) 3.0.** *Bioinformatics* 2011, **27:**1739-1740.
702 39. Bell CG, Lowe R, Adams PD, Baccarelli AA, Beck S, Bell JT, Christensen BC, Gladyshev VN,
703 Heijmans BT, Horvath S, et al: **DNA methylation aging clocks: challenges and**
704 **recommendations.** *Genome Biol* 2019, **20:**249.
705 40. Kimmel JC, Penland L, Rubinstein ND, Hendrickson DG, Kelley DR, Rosenthal AZ: **Murine**
706 **single-cell RNA-seq reveals cell-identity- and tissue-specific trajectories of aging.** *Genome*
707 *Res* 2019, **29:**2088-2103.
708 41. Rimmele P, Bigarella CL, Liang R, Izac B, Dieguez-Gonzalez R, Barbet G, Donovan M,
709 Brugnara C, Blander JM, Sinclair DA, Ghaffari S: **Aging-like phenotype and defective lineage**
710 **specification in SIRT1-deleted hematopoietic stem and progenitor cells.** *Stem Cell Reports*
711 2014, **3:**44-59.
712 42. Vandiver AR, Irizarry RA, Hansen KD, Garza LA, Runarsson A, Li X, Chien AL, Wang TS,
713 Leung SG, Kang S, Feinberg AP: **Age and sun exposure-related widespread genomic blocks**
714 **of hypomethylation in nonmalignant skin.** *Genome Biol* 2015, **16:**80.
715 43. Cole JJ, Robertson NA, Rather MI, Thomson JP, McBryan T, Sproul D, Wang T, Brock C, Clark
716 W, Ideker T, et al: **Diverse interventions that extend mouse lifespan suppress shared age-**
717 **associated epigenetic changes at critical gene regulatory regions.** *Genome Biol* 2017, **18:**58.
718 44. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, Nelson HH,
719 Karagas MR, Padbury JF, Bueno R, et al: **Aging and environmental exposures alter tissue-**
720 **specific DNA methylation dependent upon CpG island context.** *PLoS Genet* 2009,
721 **5:**e1000602.
722 45. Labarge MA, Garbe JC, Stampfer MR: **Processing of human reduction mammoplasty and**
723 **mastectomy tissues for cell culture.** *J Vis Exp* 2013.
724 46. Krueger F, Andrews SR: **Bismark: a flexible aligner and methylation caller for Bisulfite-Seq**
725 **applications.** *Bioinformatics* 2011, **27:**1571-1572.
726 47. Langmead B, Trapnell C, Pop M, Salzberg SL: **Ultrafast and memory-efficient alignment of**
727 **short DNA sequences to the human genome.** *Genome Biol* 2009, **10:**R25.
728 48. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, Garvin T, Kessler M, Zhou J, Smith AD: **A**
729 **reference methylome database and analysis pipeline to facilitate integrative and**
730 **comparative epigenomics.** *PLoS One* 2013, **8:**e81148.
731 49. Gaspar JM, Hart RP: **DMRfinder: efficiently identifying differentially methylated regions**
732 **from MethylC-seq data.** *BMC Bioinformatics* 2017, **18:**528.
733 50. Quinlan AR, Hall IM: **BEDTools: a flexible suite of utilities for comparing genomic features.**
734 *Bioinformatics* 2010, **26:**841-842.
735 51. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D: **The human**
736 **genome browser at UCSC.** *Genome Res* 2002, **12:**996-1006.
737 52. Smit A, Hubley, R & Green, P. : **RepeatMasker Open-4.0. http://www.repeatmasker.org.**
738 2013-2015.

739 53. Hansen KD, Langmead B, Irizarry RA: **BSmooth: from whole genome bisulfite sequencing**
740 **reads to differentially methylated regions.** *Genome Biol* 2012, **13:**R83.
741 54. Kolde R: **pheatmap: Pretty Heatmaps. R package version 1.0.10.** 2018.
742 55. Ernst J, Kellis M: **Chromatin-state discovery and genome annotation with ChromHMM.** *Nat*
743 *Protoc* 2017, **12:**2478-2492.
744 56. Kim D, Langmead B, Salzberg SL: **HISAT: a fast spliced aligner with low memory**
745 **requirements.** *Nat Methods* 2015, **12:**357-360.
746 57. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL: **StringTie enables**
747 **improved reconstruction of a transcriptome from RNA-seq reads.** *Nat Biotechnol* 2015,
748 **33:**290-295.
749 58. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential**
750 **expression analysis of digital gene expression data.** *Bioinformatics* 2010, **26:**139-140.
751 59. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK: **limma powers differential**
752 **expression analyses for RNA-sequencing and microarray studies.** *Nucleic Acids Res* 2015,
753 **43:**e47.
754 60. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-**
755 **seq data with DESeq2.** *Genome Biol* 2014, **15:**550.
756 61. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras
757 TR: **STAR: ultrafast universal RNA-seq aligner.** *Bioinformatics* 2013, **29:**15-21.
758 62. Heger A, Webber C, Goodson M, Ponting CP, Lunter G: **GAT: a simulation framework for**
759 **testing the association of genomic intervals.** *Bioinformatics* 2013, **29:**2046-2048.
760 63. Grant CE, Bailey TL, Noble WS: **FIMO: scanning for occurrences of a given motif.**
761 *Bioinformatics* 2011, **27:**1017-1018.
762 64. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass
763 CK: **Simple combinations of lineage-determining transcription factors prime cis-regulatory**
764 **elements required for macrophage and B cell identities.** *Mol Cell* 2010, **38:**576-589.
765 65. Bourque G, Leong B, Vega VB, Chen X, Lee YL, Srinivasan KG, Chew JL, Ruan Y, Wei CL,
766 Ng HH, Liu ET: **Evolution of the mammalian transcription factor binding repertoire via**
767 **transposable elements.** *Genome Res* 2008, **18:**1752-1762.
768 66. Jenkinson G, Abante J, Feinberg AP, Goutsias J: **An information-theoretic approach to the**
769 **modeling and analysis of whole-genome bisulfite sequencing data.** *BMC Bioinformatics*
770 2018, **19:**87.
771 67. Wickham H: *ggplot2: elegant graphics for data analysis.* Springer; 2016.
772

773

774

775

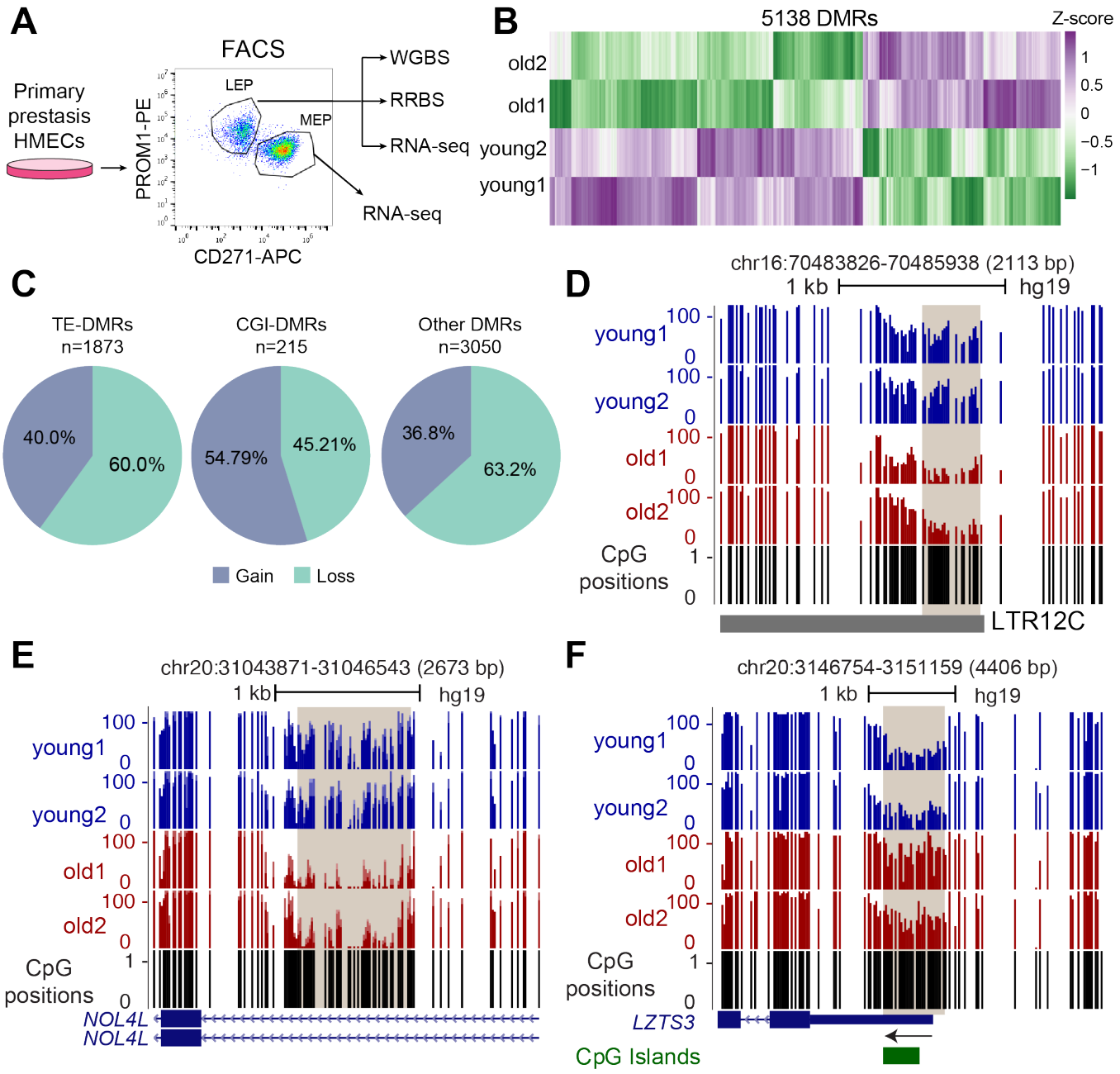
776

777

778

779

780 **Figures**



781 **Figure 1**

782 **Figure 1: Luminal cells from old individuals show DNA methylation changes at distinct genomic**
783 **loci.** (A) Schematic (FACS sorting of the luminal population). (B) Methylation levels of identified DMRs
784 across all samples. Samples are in rows, and DMRs are in columns. The color intensity represents the
785 Z-score scaled from the mean methylation level across samples. Values ranging from higher to lower
786 methylation are shown as purple to green. (C) Percentage DMRs gaining and losing methylation with age

787 are shown for DMRs overlapping transposable elements (TEs), CpG islands (CGIs), and other DMRs
788 that do not overlap either TEs or CGIs. (D) Genome browser tracks showing a representative example
789 of an LTR12C element losing methylation in old samples. (E) A representative example of an intronic
790 DMR (Other DMRs) at the *NOL4L* locus showing decreased methylation in old LEPs. (F) A representative
791 example of a DMR overlapping a CpG island at the *LZTS3* locus showing increased methylation in old
792 LEPs. DMRs are highlighted in brown.
793

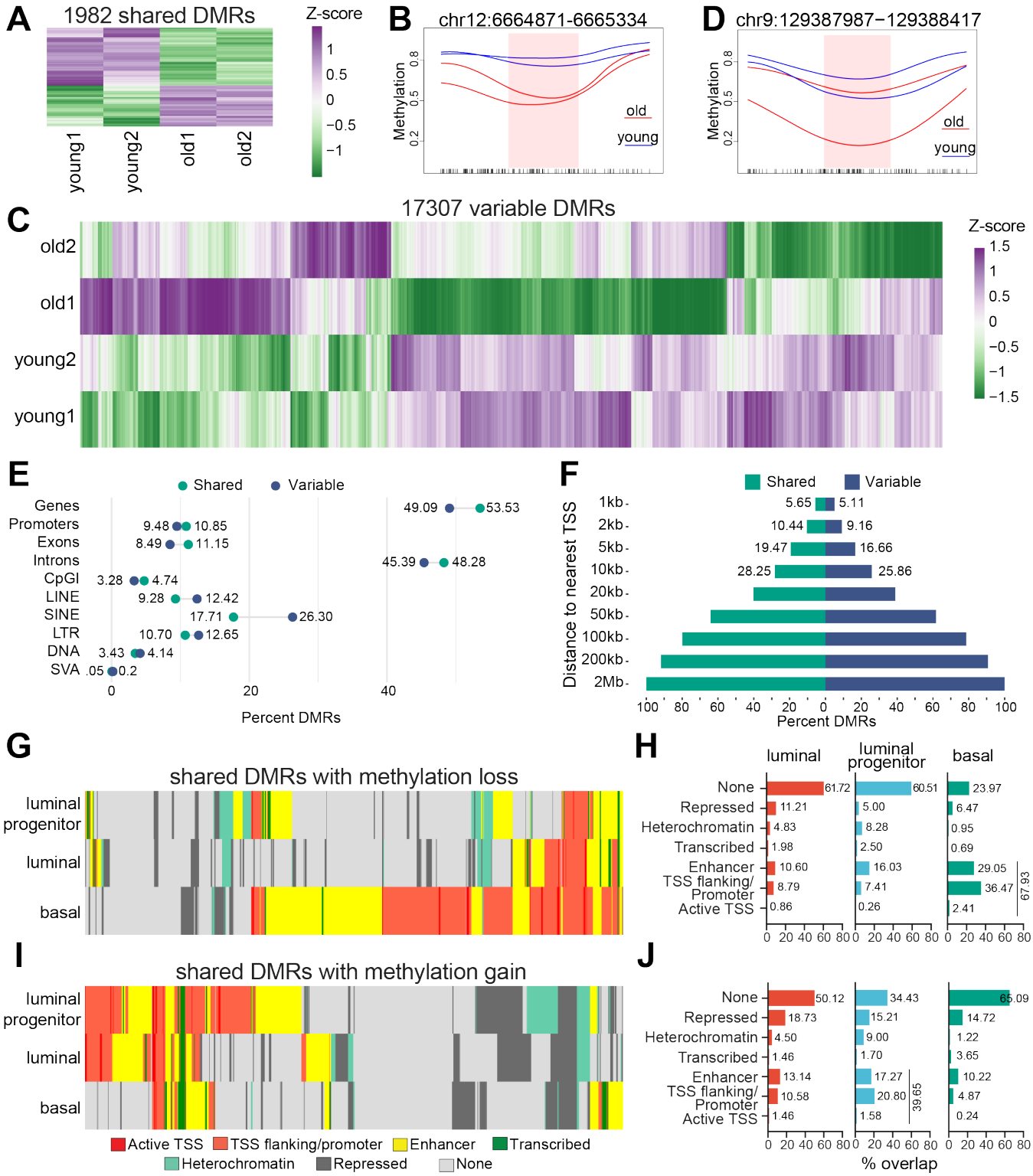


Figure 2

Figure 2: Differentially methylated regions are enriched at distinct chromatin features of mammary epithelial lineages. (A) Methylation levels of shared DMRs across all samples. Samples are

797 in columns, and DMRs are in rows. (C) Methylation levels of variable DMRs across all samples. The color
798 intensity represents the Z-score scaled from the mean methylation level across samples. Values ranging
799 from higher to lower methylation are shown as purple to green. (B) Kernel smoothed line plots of
800 representative shared and (D) variable DMRs±500bp. Blue and red lines respectively represent young
801 and old samples. DMRs are highlighted in pink and black ticks mark each CpG position. (E) Percentage
802 DMRs that overlap various gene features, CGIs, and major repeat classes. Percentage overlaps are
803 shown for both shared and variable DMRs. (F) The percentage number of shared or variable DMRs
804 present within 1kb-2Mb of the nearest TSS is shown. (G) Overlap of shared DMRs that lose or (I) gain
805 methylation with chromHMM features of luminal progenitors, luminal cells, and basal cells (rows) is
806 shown. The color key indicates the various chromatin states identified from ChIP-seq datasets. (H)
807 Percentage number of shared DMRs with methylation loss or (J) gain that overlap each chromatin feature
808 as indicated in luminal progenitors, luminal cells, and basal cells.
809
810

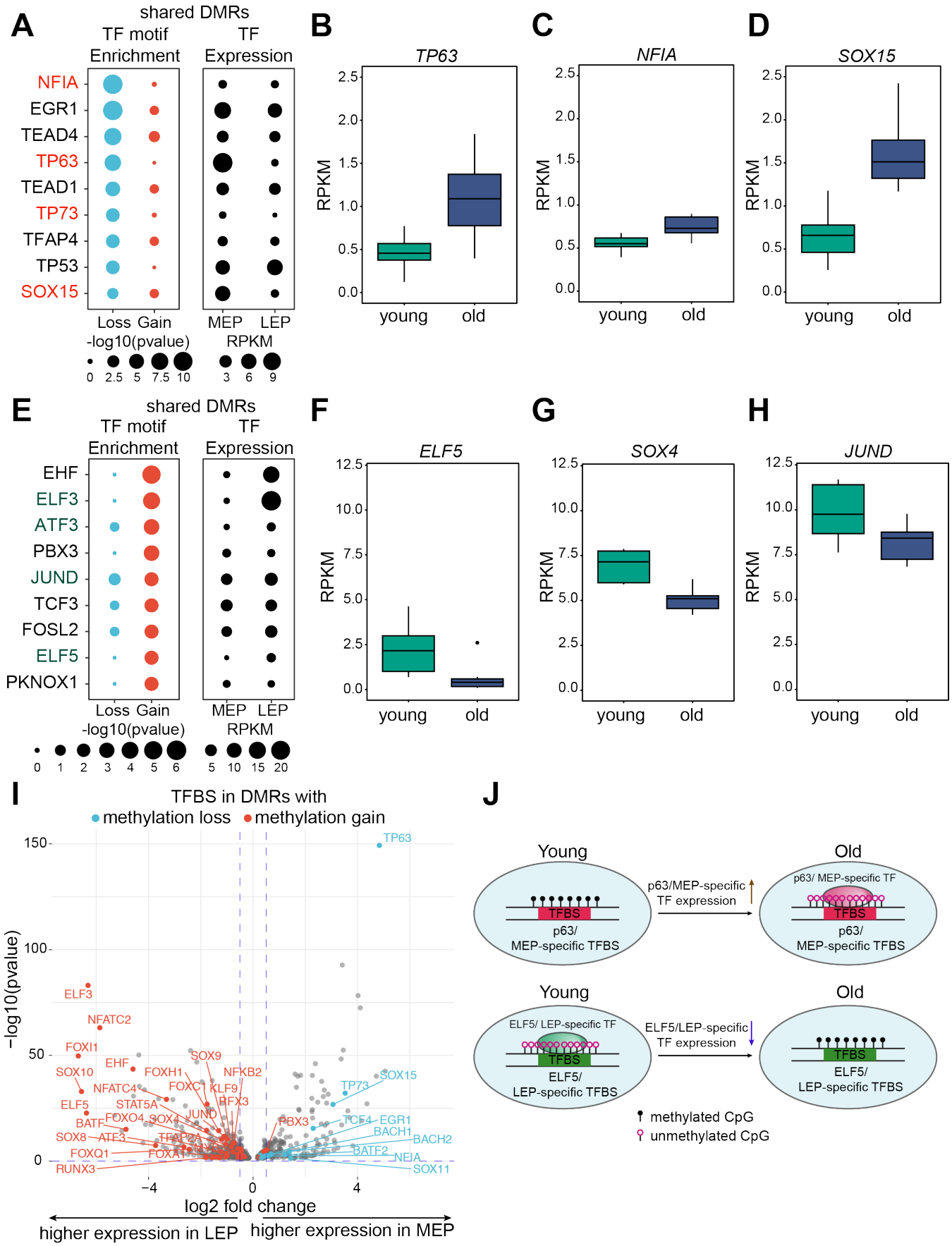

A
shared DMRs
TF motif Enrichment
TF Expression
NFIA
EGR1
TEAD4
TP63
TEAD1
TP73
TFAP4
TP53
SOX15
Loss Gain
MEP LEP
-log10(pvalue)
RPKM
B
TP63
C
NFIA
D
SOX15
young
old
E
EHF
ELF3
ATF3
PBX3
JUND
TCF3
FOSL2
ELF5
PKNOX1
F
ELF5
G
SOX4
H
JUND
I
TFBS in DMRs with
methylation loss
methylation gain
−log10(pvalue)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

Figure 3

811

812 **Figure 3: Differentially methylated regions show enrichment of lineage-specific transcription**
813 **factors.** (A) Top ten TF binding sites enriched in the shared DMRs that lose or (E) gain methylation are
814 shown. Dots in the left panel show -log10(p-value) for the enrichment of the transcription factor binding
815 site. Dots in the right panel show expression levels (RPKM) of the respective transcription factor in young
816 myoepithelial (MEP) and luminal (LEP) cells. (B-D) Boxplots showing expression levels of transcription
817 factors as indicated in young and old LEPs (n=8, *p*-value < 0.01). (F-H) Boxplots showing expression
818 levels of luminal lineage-specific transcription factors in young and old LEPs (n=8, *p*-value < 0.05). (I)
819 Volcano plot showing expression levels of all TFs expressed in LEPs and MEPs (RPKM > 0.1). TFs with
820 motifs enriched in shared DMRs with methylation loss and gain are highlighted in blue and red
821 respectively. *P*-value < 0.05. (J) Schematic showing the role of myoepithelial-specific transcription factors
822 in causing loss of methylation. Conversely, loss of expression of luminal-specific transcription factors
823 leads to unbound sites being methylated, leading to a gain in methylation.
824
825

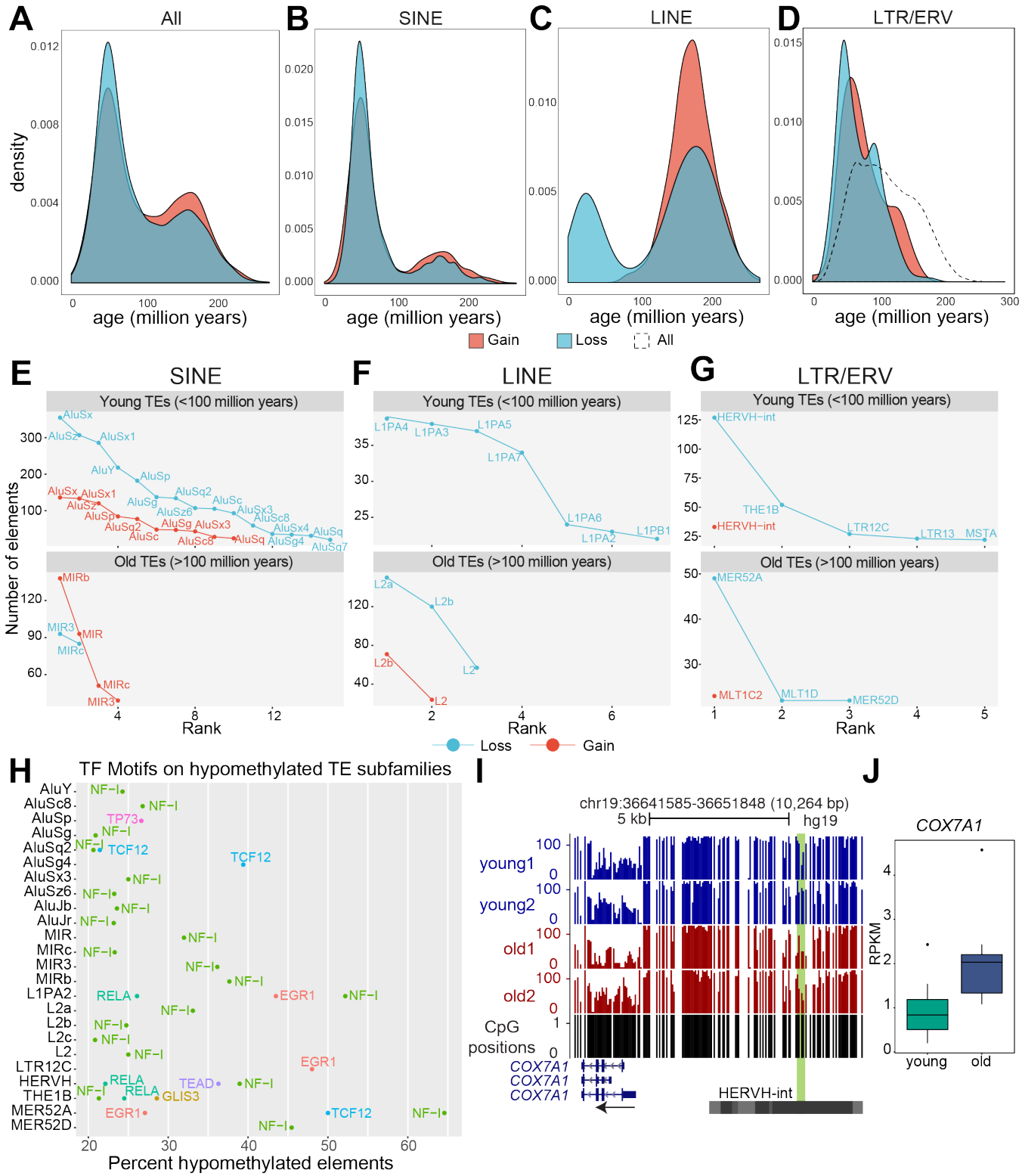


Figure 4

Figure 4: Aging leads to loss of methylation at young transposable elements. (A) Density plots of evolutionary age of transposable elements at variable DMRs with methylation loss (Loss) or gain (Gain).

829 (B) Evolutionary age of SINE, (C) LINE, and (D) LTR/ERV elements at variable DMRs with methylation
830 loss or gain. (E) Ranked plots of SINE (F) LINE and (G) LTR/ERV subfamilies enriched in variable DMRs.
831 E-G. Plots show the number of elements that lose or gain methylation within young and old TE
832 subfamilies. (H) Percentage number of variable hypomethylated DMRs at specific TE subfamilies
833 containing transcription factor binding sites as indicated. (I) Genome browser tracks showing
834 hypomethylation of a HERVH element near the *COX7A1* gene. (J) Expression of *COX7A1* gene in young
835 and old LEPs, n=8, *p*-value <0.05.
836
837

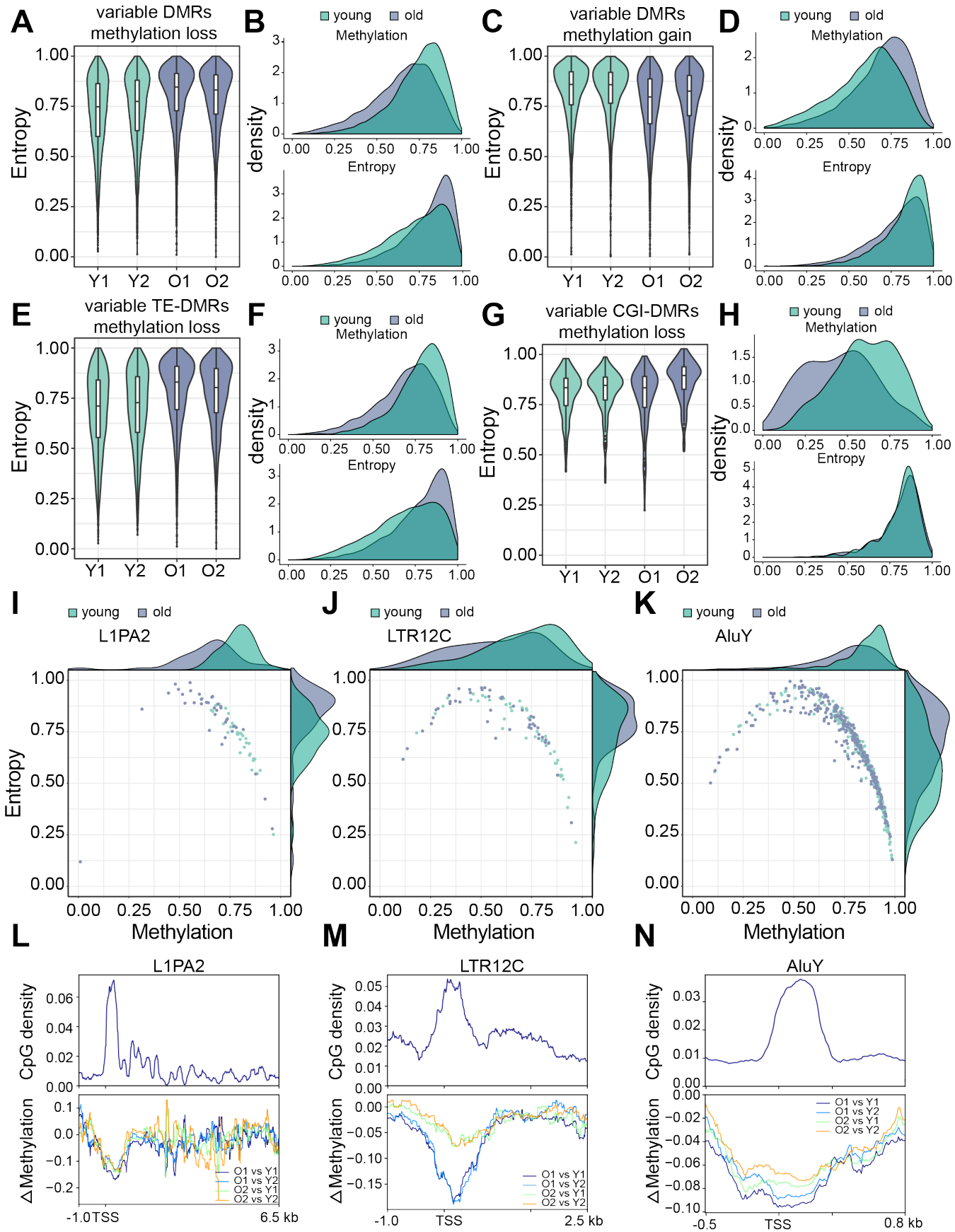

A
variable DMRs
methylation loss
Entropy
Y1
Y2
O1
O2
B
young
old
Methylation
Entropy
density
C
variable DMRs
methylation gain
D
E
variable TE-DMRs
methylation loss
F
G
variable CGI-DMRs
methylation loss
H
I
L1PA2
J
LTR12C
K
AluY
Methylation
L
M
N
CpG density
ΔMethylation
O1 vs Y1
O1 vs Y2
O2 vs Y1
O2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

Figure 5

839 **Figure 5: Aging increases methylation heterogeneity at young CpG dense transposons.** (A)
840 Normalized methylation entropy (Entropy) levels at variable DMRs with methylation loss in young and old
841 individuals. (B) Density plots showing DNA methylation (top panel) and entropy levels (bottom panel) at
842 variable DMRs with methylation loss (C) As A, but showing variable DMRs with methylation gain. (D) As
843 B, but showing variable DMRs with methylation gain. (E) Entropy levels at variable DMRs overlapping
844 TEs and (G) variable DMRs overlapping CGIs with methylation loss (F) Density plots showing DNA
845 methylation (top) and entropy levels (bottom) at variable DMRs overlapping TEs and (H) variable DMRs
846 overlapping CGIs with methylation loss. (I) Scatterplots of DNA methylation and entropy levels across
847 L1PA2, (J) LTR12C, and (K) AluY elements that lose methylation. Marginal density plots are shown to
848 display the density of points on the X and Y axes. (L) Aggregate CpG density and methylation change (Δ
849 Methylation) at hypomethylated L1PA2, (M) LTR12C, and (N) AluY elements. Only full-length elements
850 were used to generate aggregate plots. (Y1: young1, Y2: young2, O1: old1, O2: old2)
851
852
853

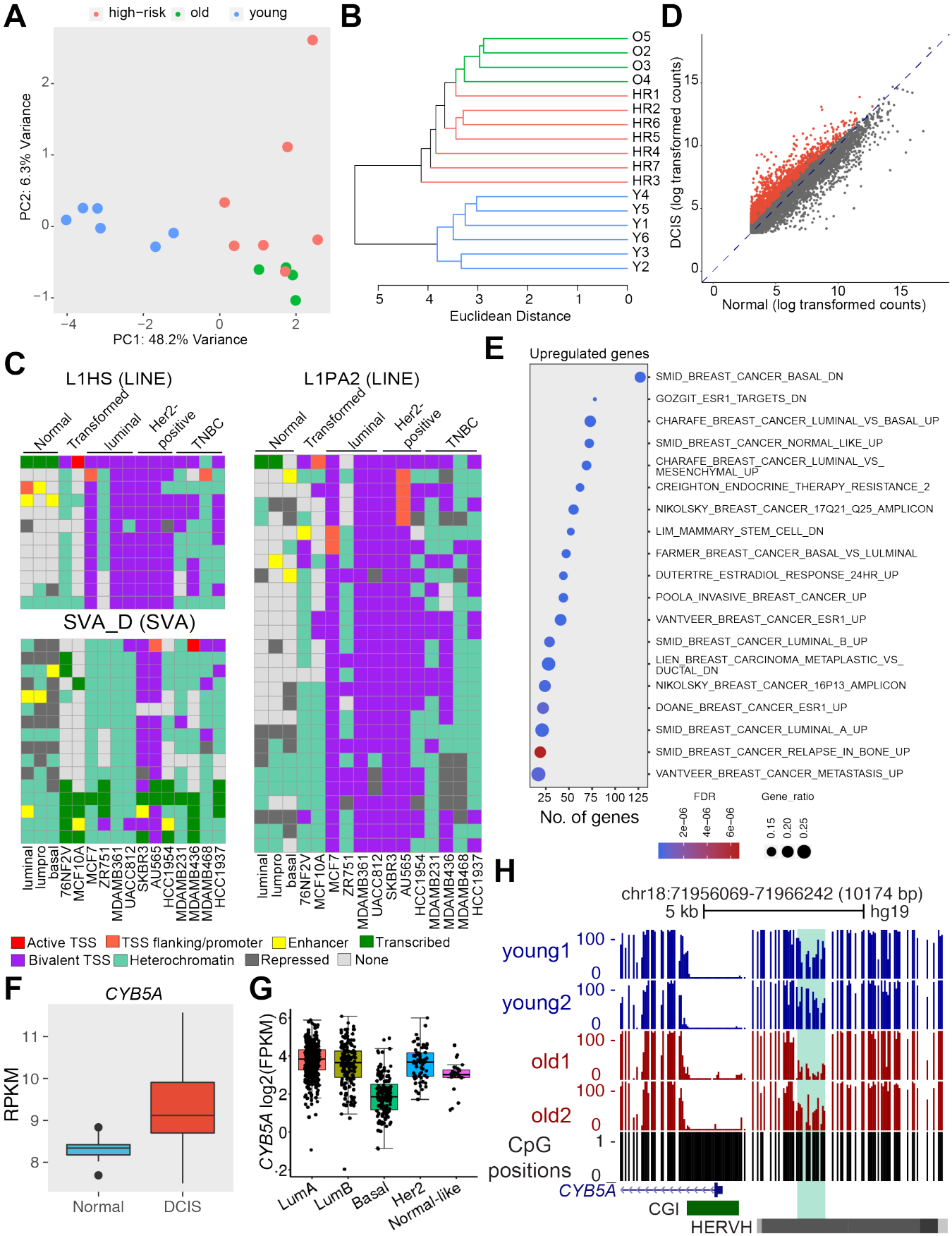

A
high−risk
old
young
PC2: 6.3% Variance
PC1: 48.2% Variance
B
O5
O2
O3
O4
HR1
HR2
HR6
HR5
HR4
HR7
HR3
Y4
Y5
Y1
Y6
Y3
Y2
Euclidean Distance
D
DCIS (log transformed counts)
Normal (log transformed counts)
C
L1HS (LINE)
L1PA2 (LINE)
SVA_D (SVA)
Normal
Transformed
luminal
Her2-positive
TNBC
luminal
lumpro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

Figure 6

855 **Figure 6: Age linked methylation loss at transposable elements is associated with luminal breast**
856 **cancer risk.** (A) Principal component analysis and (B) hierarchical clustering of young, old, and high-risk
857 samples using methylation levels at hypomethylated DMRs overlapping TEs. (C) Heatmaps showing
858 chromatin states identified at hypomethylated L1HS, L1PA2 and SVA_D elements in normal cells,
859 transformed mammary cell lines and breast cancer cell lines. Each row shows a hypomethylated TE
860 identified from WGBS and RRBS data. (D) Upregulated genes in DCIS samples (red) within 50kb of a
861 hypomethylated TE. (E) MSigDb gene sets enriched in genes highlighted in red in D. (F) Expression of
862 *CYB5A* in Normal and DCIS samples (*p*-value < 0.0001) and (G) in five subtypes of breast tumors from
863 the TCGA BRCA collection (*p*-value Luminal A (LumA) versus Basal and Luminal B (LumB) versus Basal
864 $< 2.2 \times 10^{-16}$). (H) Genome browser tracks showing hypomethylation of a HERVH element near the *CYB5A*
865 gene promoter.
866
867