Expression of Most Retrotransposons in Human Blood Correlates with Biological Aging

Yi-Ting Tsai[1†], Nogayhan Seymen[1†], Ian R. Thompson[1], Xinchen Zou[2], Warisha Mumtaz[1], Sila Gerlevic[1], Ghulam J. Mufti[1] and Mohammad M. Karimi[1*]

[1] Comprehensive Cancer Centre, School of Cancer & Pharmaceutical Sciences, Faculty of Life Sciences & Medicine, King's College London, London SE5 8AF, UK.
[2] MRC LMS, Imperial College London, London W12 0NN, UK.

[†] These authors contributed equally to this work and share first authorship.
* Address correspondence to mohammad.karimi@kcl.ac.uk

Abstract

Retrotransposons (RTEs) have been postulated to reactivate with age and contribute to aging through activated innate immune response and inflammation. Here, we systematically analyzed the relationship between RTEs expression and aging using published transcriptomic and methylomic datasets of human blood. Despite no observed correlation between RTEs activity and chronological age, most RTE classes and families except short interspersed nuclear elements (SINEs) correlate with age-associated gene signature scores. Strikingly, we found that the expression of SINEs is linked to upregulated DNA repair pathways in multiple cohorts. DNA hypomethylation with aging was observed across RTE classes and associated with increased RTEs expression in most RTE classes and families except SINEs. Additionally, our single-cell transcriptomic analysis suggests a role for plasma cells in aging mediated by RTEs. Altogether, our multi-omics analysis of large human cohorts highlights the role of RTEs in biological aging and suggests possible mechanisms and cell populations for future investigations.

Introduction

Transposable elements (TEs) are genetic elements that can move within the genome and are categorized into DNA transposons and retrotransposons (RTEs), which depend on cDNA intermediates to function. RTEs include three classes: long terminal repeat (LTR) retrotransposons, which integrate into the genome like the retroviruses, and long and short interspersed nuclear elements, known as LINEs and SINEs, respectively. LINEs and SINEs employ target-primed reverse transcription for genome integration, with SINEs depending on the proteins encoded by LINEs[1]. Although most of the RTE sequences are dormant in the host genome, they continue to play vital roles in human evolution and physiology[2,3].

RTEs are silenced through heterochromatinization and DNA methylation in the early developmental stages as part of the host defense mechanism[3,4]. However, increased RTE activity that occurs with aging can lead to genome instability and activation of DNA damage pathways[4-7]. Additionally, accumulation of cytoplasmic RTE cDNAs detected in aging organisms and senescent cells can activate type I interferon (IFN-I) response and inflammation[3,4,8-10], that contributes to inflammaging[11]. Furthermore, chromatin remodeling and cellular senescence also contribute partially to the RTE reactivation[10]. Cellular senescence is a non-proliferative state elicited by stress factors such as DNA damage and is closely intertwined with inflammation and aging[12]. Through the expression of senescence associated secretory phenotype (SASP), senescent cells promote inflammation and senescence in other cells, including immune cells, leading to a compromised immune system and a vicious cycle of inflammation[13]. Understanding how RTEs expression as a factor that might affect the SASP, cellular senescence, and inflammaging could offer important insights into strategies against aging and age-related diseases (ARDs).

Recent studies have documented the relationship between RTEs activation and age-related events, but comprehensive and large-scale studies are lacking, mainly due to the paucity of RNA-seq data for large non-cancerous human cohorts. Most of the existing transcriptomics datasets are microarray based, where the computational methods were not sufficiently developed to analyze repetitive elements. By overlapping the Illumina Microarray expression and methylation probes with RepeatMasker[14], we were able to identify enough probes to calculate the expression of RTEs classes and families. Building on this methodology, we explored how RTEs expression contributes to biological aging using publicly available transcriptomics microarray data derived from human blood samples. More specifically, we first investigated if RTEs expression was correlated with chronological age, and then we analyzed the relationship between RTE and age-related events including cellular senescence, inflammation, and IFN-I response with published gene signatures. Using microarray methylomic data, we also investigated the DNA methylation level of RTEs

66 in blood samples of multiple non-cancerous human cohorts and examined the relation
67 between DNA methylation and RTEs expression and aging. Furthermore, with annotated
68 single-cell transcriptomic data of the peripheral blood mononuclear cells (PBMCs) of 21
69 healthy human samples from a recent study of aging immunity[15], we identified cells that
70 could be implicated in the process of RTEs reactivation and aging. Lastly, the single-cell
71 transcriptomic data of supercentenarians[16] was analyzed to examine the interplay between
72 RTEs expression and aging as a factor of aging and longevity (**Fig. 1**).
73
74

Results

Analyzing DNA methylation and expression levels of RTEs using microarray data

We collected three published microarray datasets from large-scale human studies including the Multi-Ethnic Study of Atherosclerosis[17] (MESA, aged 44 to 83, n=1202), the Grady Trauma Project[18,19] (GTP, aged 15 to 77, n=359), and Genetics, Osteoarthritis and Progression[20] (GARP, aged 43 to 79, n=139), which are from peripheral monocytes of healthy human, whole blood of healthy humans, and PBMCs from primary osteoarthrosis patients, respectively (**Supplementary Table 1** and **Supplementary Information**). The studies were conducted using either Illumina HumanHT-12 V3 or Illumina HumanHT-12 V4 expression microarray kits. After comparing the probes of the two microarray versions, we adopted the more comprehensive probe list of V4, which contains the full intersection of the two lists. To generate the RTEs expression, we mapped the probe locations to RepeatMasker to extract a list of intergenic probes that cover the RTE regions. We included three main RTE classes: (1) the LINE class, which encompasses the L1 and L2 families; (2) the SINE class, with Alu and MIR as the two main families; and (3) the LTR class which comprises the ERV1, ERVl, ERVL-MaLR, and ERVK families. Most of the RTE-covering probes available on Illumina HumanHT-12 V4 are present in MESA and GARP, while fewer are available in GTP (**Supplementary Fig. 1a**, **Supplementary Tables 2-4**).

Four methylation datasets were analyzed, including MESA, Swedish Adoption/Twin Study of Aging[21] (SATSA, aged 48 to 98, n=1072), Brisbane Systems Genetics Study[22] (BSGS, aged 10 to 75, n=862), and Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates[23] (GMPWAR, aged 19 to 101, n=656). SATSA, BSGS, and GMPWAR are genome-wide DNA methylation studies of whole blood samples from healthy individuals generated with Illumina Infinium 450k array (**Supplementary Table 1** and **Supplementary Information**). The DNA methylation probes from the Illumina Infinium 450k array kit were aligned to RepeatMasker to identify the probes overlapping the RTE regions (**Supplementary Table 5**). More than 90% of the RTE-covering probes are present in MESA and GMPWAR, while fewer, but more than half, are available in SATA and BSGS datasets (**Supplementary Fig. 1b**).

The chronological age is not linked with RTEs expression.

Firstly, we examined the relationship between RTEs expression and chronological age in the MESA, GARP, and GTP cohorts. The ages of the individuals enrolled in these studies range from 15 to 83 years, with the mean being 70.2 (MESA), 60.1 (GAPR), and 42.5 (GTP) (**Supplementary Table 1**). Strikingly, no significant relationship (P value < 0.05) is found

between chronological age and expressions of the three RTE classes across the datasets (**Fig. 2a**). Similarly, apart from a weak correlation observed between a few RTE families and chronological age, we could not identify any strong relation between chronological age and expression of RTE families across the three cohorts (**Supplementary Fig.2a**).

A similar correlation analysis was carried out in a scRNA-seq data of the PBMCs of 21 non-obese healthy men annotated into 25 cell types by Mogilenko et al.[15] (**Supplementary Table 6** and **online methods**). When comparing the RTEs expressions of the young (aged 25 to 29, n=11) group with the old (aged 62 to 70, n=10) group, we could not find any significant difference or trend in any cell types between the two groups (**Supplementary Figs. 3-5**). Overall, RTEs expression did not correlate with chronological aging across our microarray and scRNA-seq datasets.

RTEs expression positively correlates with age-associated gene signature scores except for SINEs.

In mammals, aging is characterized by an intricate network of inflammation, IFN-I signaling, and cellular senescence[3]. To measure the level of biological aging from the transcriptomic data, we focused on six gene sets retrieved from widely cited studies: IFN-I[24], inflammatory cytokines[25], inflammatory chemokines[26], inflammaging[27], SASP[28], and cellular senescence[29] (**Supplementary Table 7**). The scores of these six age-associated gene sets were calculated for each individual in the MESA, GTP, and GARP cohorts by applying the *singscore* package in Bioconductor[30] (**online methods**). Except for GTP, in which 12 genes were missing from the gene sets in total (**Supplementary Table 8**), MESA and GARP microarray datasets had the expression of all the genes in the gene sets.

Then, we analyzed the relation between RTEs expression and the six age-associated gene signature scores. Positive correlations were observed between the expressions of LINEs and LTRs and age-associated signature scores across three datasets, while inverse correlation was present between the expression of SINEs and these scores in the PBMC samples from the GARP cohort (**Fig. 2a and b**). The discrepant pattern of LTRs and LINEs versus SINEs in PBMC is more pronounced in the correlation analysis between the expression of RTE families and the age-associated gene sets (**Fig. 2c and d**). Interestingly, the expressive patterns of RTE families are divided into two groups that also reflect their positive or inverse correlations with age-associated signature scores in the GARP cohort. In the GARP cohort, the strongest positive correlation with age-associated scores was observed in L1, ERVK, and ERVL. By contrast, the expressions of Alu and MIR of SINE, L2, and ERV1 were inversely correlated with age-associated signature scores in the GARP cohort (**Fig. 2c and d**). This pattern is also observed in PBMCs in the MESA monocyte

144 samples for L1, ERVL, and ERVK, but not other RTE families (**Supplementary Fig.2a and**
145 **c**). The whole blood GTP cohort showed divergent result to the other two cohorts perhaps
146 because of the absence of certain genes in the six age-associated gene sets
147 (**Supplementary Fig.2a-b** and **Supplementary Table 8**).
148 To study the association between RTEs expression and age-associated gene signatures,
149 we also split the samples in each cohort based on quartiles of RTEs expression into high,
150 low, and medium (top 25%, bottom 25%, and middle 50%) groups and compared the age-
151 associated gene signature scores across these three groups. Across the datasets, there is
152 an overall trend of increased scores of age-associated gene signatures in high versus low
153 LINE and LTR expression groups (**Supplementary Fig. 6**). In fact, amongst all gene sets,
154 we found the most significant increase for SASP and senescence and to a lesser extent for
155 inflammaging signatures. We observed higher SASP, inflammatory cytokine, senescence,
156 and inflammaging in the healthy monocytes (MESA) in high SINE- and Alu-expressing
157 groups (**Supplementary Fig. 7**). However, in the PBMCs samples of the GARP cohorts,
158 groups of high SINE and Alu expressions demonstrate lower expressions of SASP,
159 inflammatory chemokine, senescence, and inflammaging, that reflects the inverse correlation
160 we observed before (**Fig. 2c**). There are no specific trends for SINE expression in the whole
161 blood samples from the GTP cohort (**Supplementary Fig. 7**). Taken together, SINEs display
162 different pattern than LINEs and LTRs in terms of the association between their expressions
163 and age-related gene signatures. This prompted us to conduct a gene set variation analysis
164 (GSVA)[31] to identify the pathways that are regulated in the high versus low RTE expression
165 groups in the three cohorts (**online methods**).

166 **Elevated SINE expression is linked with the upregulation of the DNA repair pathways,**
167 **while elevated LINE and LTR expressions are associated with the inflammatory**
168 **response.**

169 We conducted GSVA for high versus low expression groups of RTE classes and families
170 to identify the pathways that might be affected by RTEs expression. We specifically focused
171 on the pathways (gene sets) related to the inflammatory and DNA repair responses retrieved
172 from the Molecular Signatures Database (MSigDB)[32] with “inflammatory” and “dna_repair” as
173 search keywords.
174 We found that a significant fraction of gene sets related to DNA repair were
175 downregulated in the group of monocyte samples (MESA cohort) highly expressing LINE
176 and LTR classes but upregulated in the monocyte and PBMC samples (MESA and GARP
177 cohorts, respectively) highly expressing SINEs (**Fig. 3a**). At the family level, there is a unique
178 increase in DNA repair activity in the monocytes of the cohort with high Alu expression. In

179 the PBMC samples of the GARP dataset, DNA repair pathways are upregulated in groups
180 highly expressing L2, ERV1, Alu, and MIR, while most of the other RTE classes demonstrate
181 decreased DNA repair activity. In contrast, the inflammatory response is upregulated in the
182 monocyte and PBMC samples of groups expressing LINE and LTR classes and families at
183 high levels, and such upregulation is less common in SINE (**Fig. 3a**). These observations
184 provide insight into the perpetuating contrast between SINE versus LINE and LTR observed
185 in our analysis, as SINE might contribute to aging via genome instability instead of
186 inflammation that is associated with LINE and LTR activities. Overall, our observations are
187 consistent across MESA and GARP cohorts but less so in the GTP cohort (**Fig. 3b**), which
188 might be attributed to less RTE-covering probes and the missing genes from the age-related
189 gene signatures. (**Supplementary Fig. 1a, Supplementary Table 8**).

190 **DNA methylation levels of RTEs inversely correlated with the chronological age and**
191 **the RTEs expression except for SINE expression.**

192 The demethylation of RTEs is suggested to contribute to aging[3,5,7]. To examine the effect of
193 RTEs demethylation on chronological aging, we analyzed DNA methylation data from
194 multiple cohorts including monocytes from the MESA cohort and the whole blood data from
195 GMPWAR, BSGS, SATSA. Our analysis showed an inverse correlation between the
196 methylation level of RTEs and chronological age across all RTE classes and families in the
197 whole blood and monocytes (**Fig. 4a-b** and **Supplementary Fig. 8**). This observed trend is
198 in line with previous reports of demethylation of RTE with age in gene-poor regions[5,21,23,33].
199 To elucidate the relationship between the expression and methylation level of RTEs, we
200 divided the MESA samples, with matched RTEs DNA methylation and expression, into three
201 groups based on the level of expression of RTEs in different classes and families: low (1st
202 quartile), medium (2nd and 3rd quartile), and high (4th quartile) expression groups of monocyte
203 samples from the MESA cohort. The DNA methylation levels of LINEs and LTRs significantly
204 decrease in groups of high LINE and LTR expressions, respectively, but SINEs and the
205 satellite repeats (as negative control) do not display the same pattern (**Fig. 4c**). LINE
206 families, MIR, and LTR families except ERVK display lower levels of methylation in higher
207 expression groups, while such pattern is absent in Alu, CR1, and ERVI (**Supplementary Fig.**
208 **9**). Although the DNA methylation level of SINEs positively correlates with the methylation
209 levels of LINEs and LTRs, the expression of SINEs is not significantly correlated with the
210 expression levels of other RTE classes (**Fig. 4d**). Overall, we found a distinct pattern for
211 SINEs versus LTRs and LINEs by exploring a relationship between DNA methylation and
212 expression levels of RTEs in monocytes, consistent with our other findings showing different
213 patterns for SINEs versus LTRs and LINEs in PBMCs (see above). This result also suggests

214 a dispensable role of DNA methylation for SINE silencing in monocytes, which is in line with
215 previous studies showing SINEs are not derepressed by deletion of DNA methyltransferases
216 or treatment with DNA demethylating agents and are primarily regulated by histone
217 modifications[34].

218 **Elevated level of RTEs in plasma cells of healthy PBMC samples is associated with**
219 **the high SASP and inflammaging gene signature scores.**

220 As described above, we did not identify any association between RTEs expression and
221 chronological aging for any human cohorts we analyzed. However, the positive correlation
222 between LINE and LTR expressions with age-associated gene signature scores in the
223 PBMC samples of the GARP cohort prompted us to investigate PBMC samples in more
224 detail using PBMC scRNAseq data of 21 samples annotated into 25 cell types by Mogilenko
225 et al.[15](**online methods** and **Supplementary Table 6**). The scRNA-seq samples were
226 divided based on low and high age-associated gene signature scores to compare against
227 the expression of RTE classes for each cell type (**online methods**). Among the 25
228 annotated cell types, plasma cells consistently demonstrate significantly elevated RTEs
229 expression in cells of high SASP expression that is not observed in the other cell types (**Fig.**
230 **5a** and **Supplementary Figs. 10-12)**. Increased LINE and SINE expressions were also
231 observed in the plasma cells of the samples with high inflammatory chemokines and
232 inflammaging gene signature scores (**Fig. 5b and c**). Although not statistically significant,
233 elevated RTEs expression was also seen in the groups with high inflammatory cytokines and
234 senescence but not IFN-I scores (**Fig. 5d-f**). Overall, this finding indicates that plasma cells
235 might play an important role in the process in which reactivation of RTEs regulates aging.

236 **Downregulation of RTEs expression in supercentenarians versus normal aged cases**
237 **despite high age-associated signature scores in NK and T cells of supercentenarians**

238 Supercentenarians who have reached 110 years of age are an excellent resource for
239 investigating healthy aging. We hypothesized that supercentenarians should express RTEs
240 less than normal-aged people because RTEs expression, particularly LINE and LTR
241 expressions, can activate the inflammatory pathways (**Fig. 3**), which would be generally
242 harmful to healthy aging. To understand whether there is any RTEs expression change in
243 supercentenarians versus normal-aged people and identify the relationship to age-
244 associated gene signatures, we compared the scRNA-seq data of the PBMCs of seven
245 supercentenarians (over 110 years) to five non-centenarian (50 to 80 years) controls[16]. We
246 collected the annotated cells of these twelve cases as a Seurat object for which the cells
247 were annotated into ten different cell types. Compared with the control group, the

248 supercentenarians demonstrate decreased RTEs expression in the natural killer (NK) cells,
249 B-cells, T-cells, monocytes, and dendritic cells (**Fig. 5g** and **Supplementary Figs. 13**). Both
250 T cell types show decreased LINE expression, while the noncytotoxic cluster, TC1, shows
251 slightly decreased LTR expression, and the expanding cytotoxic T cells, TC2, show
252 decreased SINE expression. On the other hand, the cytotoxic T cells also show an increased
253 level of inflammatory cytokines and inflammaging gene signature scores. In addition, the NK
254 cells show higher senescence and SASP scores in the supercentenarians while
255 demonstrating decreased LINE and SINE expressions (**Fig. 5g** and **Supplementary Figs.**
256 **14-15**). Overall, this result reveals inflammatory activity in cytotoxic T cells and increased
257 senescence in the NK cells that is not associated with the activation of RTEs in
258 supercentenarians. By contrast, the expression of RTEs is significantly reduced in most of
259 the immune cell types in supercentenarians.
260
261 **Discussion**
262
263 The role of RTEs in aging has been studied extensively but not in a systematic manner
264 with large, non-cancerous human cohorts across RTE categories before this study. Current
265 understanding of the relationship between RTEs and aging are largely conducted *in vitro* or
266 in model organisms[7,9,33,35] that might not apply to human due to intraspecies differences[36].
267 Moreover, the link of RTEs to aging in non-cancerous human cohorts has not been studied
268 systematically in large cohorts.
269 We combined bulk and single-cell transcriptomic data to examine the relationship between
270 RTEs expression versus chronological and biological aging, demonstrating that
271 chronological aging is not significantly linked with RTEs reactivation. In this process, we
272 discovered that LTR and LINE expressions are positively correlated with the age-related
273 gene signature scores. However, in contrast, the SINE expression demonstrates an inverse
274 correlation with age-related scores in the PBMCs of osteoarthritis patients and no significant
275 relationship in the whole blood of healthy individuals. We suspect the absence of a positive
276 correlation between age-associated scores and SINE expression might be linked to the
277 same discrepancy observed in the GSVA analysis, in which DNA repair pathways are
278 upregulated in the groups of cases with high SINE and Alu expressions. This result suggests
279 that reactivation of SINE, particularly Alu, might lead to DNA damage, a hallmark of aging.
280 Although SINEs compose up to 13% of the human genome, and Alu elements make up the
281 largest family of human mobile elements, Alu is not studied as extensively as L1 from the
282 LINE class[1,37]. Therefore, this novel discovery requires further study to confirm the
283 connection and elucidate the mechanism by which DNA repair might be linked to the
284 derepression of SINE and Alu expressions. In contrast, the groups of cases with high LINE

285 and LTR expressions have upregulation of the pathways related to the inflammatory
286 response, which is in line with the previous findings from human and model organisms[3,35].
287 The mechanisms of aging and age-related diseases (ARDs) converge on inflammaging[38].
288 Previous studies have shown RTE's pathogenetic role in ARDs including neurodegenerative
289 diseases through elevation of DNA damage and genome instability[39-41]. SINE expression
290 inversely correlated with the age-related signature scores in the PBMCs of osteoarthritis
291 patients, while a positive correlation is observed in healthy monocytes. Moreover, a
292 discrepant upregulation of DNA repair pathways is seen in the osteoarthritis cohorts
293 expressing higher level L2, ERV1, Alu, and MIR. With osteoarthritis being a typical ARD, our
294 analysis provides insight into the relationship between aging and RTEs in an ARD state in
295 comparison with that of healthy individuals.
296 The relationship between DNA methylation and expression of RTEs with age was also
297 investigated in our study. The methylation levels of all three RTE classes decrease with
298 chronological age in our analysis, and the RTEs expression is in inverse correlation with the
299 methylation level in LINE and LTR but not in SINE. DNA methylation at the CpG sites is an
300 important mechanism for RTE repression; through the recruitment of DNA and histone
301 methyltransferases, RTE sequences are repressed via DNA methylation and
302 heterochromatinization[42]. In the aging process, surveillance of transcriptional regulators such
303 as SIRT6 wanes, leading to hypomethylation and reduced heterochromatin repression[4]. L1
304 has been identified to be repressed through DNA methylation[43]. However, SINE has been
305 suggested to be predominantly regulated by histone modification, and loss of DNA
306 methylation has little effect on the transcription of SINEs[34], as reflected in our analysis.
307 Although expression microarrays have been extensively employed to study the expression
308 of genes, they were not systematically used for studying the expression of RTEs. Here, we
309 select the probes that cover the RTE regions to extract RTEs expression. The probe design
310 of microarrays does not accommodate RTEs specifically, owing to its repetitive and
311 interspersed nature[44]. To circumvent this issue, our analysis was conducted on the class and
312 family levels but not subfamily levels of RTEs to retain enough probes (**Supplementary Fig.**
313 **1**) providing sufficient power for calculating the expression scores. We also supported our
314 findings with scRNAseq, particularly when investigating the association between RTEs
315 expression and chronological age, to circumvent the potential constraints posed by using
316 microarray data. However, we are aware of the limitations imposed by using microarray in
317 this study, particularly the low number of intergenic probes in the expression microarray
318 data. Our study can be enriched with the advent of large RNA-seq cohorts for aging studies
319 in the future.
320 In summary, our analysis provides an overview of the relationship between the
321 chronological and biological aging versus RTEs expression and methylation using multi-

omics data from large human cohorts. Our analysis of the transcriptomics data showed no relation between RTEs expression and chronological aging. In contrast, the expressions of LINE and LTR are positively correlated with the age-related signature scores, showing a strong correlation with biological aging. The high expression of SINEs, on the other hand, is associated with an upregulation of DNA repair pathways in the PBMC cohort. Our exploration of multiple cell types based on a published PBMC scRNA-seq cohort unveiled plasma cell as the only cell type for which the expression of RTEs correlates with age-related signature scores. This matches a recent finding showing a late senescent-like phenotype marked by accumulation of TEs in plasma cells during pre-malignant stages[45]. Lastly, we explored a supercentenarian PBMC scRNA-seq cohort as a model for healthy aging and identified that RTEs expression is generally decreased in supercentenarians. However, the age-related signatures scores, particularly senescence and SASP scores, increased in the NK cells in supercentenarians versus normal aged group, highlighting immunosenescence as an important factor even in healthy aging.

References

1 Bourque, G. *et al.* Ten things you should know about transposable elements. *Genome Biology* **19**, 199-199 (2018). https://doi.org:10.1186/s13059-018-1577-z

2 Gorbunova, V., Boeke, J. D., Helfand, S. L. & Sedivy, J. M. Sleeping dogs of the genome. *Science* **346**, 1187-1188 (2014). https://doi.org:10.1126/science.aaa3177

3 Gorbunova, V. *et al.* The role of retrotransposable elements in ageing and age-associated diseases. *Nature* **596**, 43-53 (2021). https://doi.org:10.1038/s41586-021-03542-y

4 Van Meter, M. *et al.* SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat Commun* **5**, 5011 (2014). https://doi.org:10.1038/ncomms6011

5 Tsurumi, A. & Li, W. X. Global heterochromatin loss: a unifying theory of aging? *Epigenetics* **7**, 680-688 (2012). https://doi.org:10.4161/epi.20540

6 De Cecco, M. *et al.* Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues. *Aging (Albany NY)* **5**, 867-883 (2013). https://doi.org:10.18632/aging.100621

7 Wood, J. G. *et al.* Chromatin-modifying genetic interventions suppress age-associated transposable element activation and extend life span in Drosophila. *Proc Natl Acad Sci U S A* **113**, 11277-11282 (2016). https://doi.org:10.1073/pnas.1604621113

8 De Cecco, M. *et al.* L1 drives IFN in senescent cells and promotes age-associated inflammation. *Nature* **566**, 73-78 (2019). https://doi.org:10.1038/s41586-018-0784-9

9 Simon, M. *et al.* LINE1 Derepression in Aged Wild-Type and SIRT6-Deficient Mice Drives Inflammation. *Cell Metab* **29**, 871-885 e875 (2019). https://doi.org:10.1016/j.cmet.2019.02.014

10 De Cecco, M. *et al.* Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* **12**, 247-256 (2013). https://doi.org:10.1111/acel.12047

11 Franceschi, C. *et al.* Inflamm-aging. An evolutionary perspective on immunosenescence. *Annals of the New York Academy of Sciences* **908**, 244-254 (2000). https://doi.org:10.1111/j.1749-6632.2000.tb06651.x

12 Gorgoulis, V. *et al.* Cellular Senescence: Defining a Path Forward. *Cell* **179**, 813-827 (2019). https://doi.org:10.1016/j.cell.2019.10.005

13 Coppe, J. P., Desprez, P. Y., Krtolica, A. & Campisi, J. The senescence-associated secretory phenotype: the dark side of tumor suppression. *Annu Rev Pathol* **5**, 99-118 (2010). https://doi.org:10.1146/annurev-pathol-121808-102144

14 Smit, A., Hubley, R, Green, P. . *RepeatMasker Open-4.0.*, <http://www.repeatmasker.org> (2013-2015).

15 Mogilenko, D. A. *et al.* Comprehensive Profiling of an Aging Immune System Reveals Clonal GZMK(+) CD8(+) T Cells as Conserved Hallmark of Inflammaging. *Immunity* **54**, 99-115 e112 (2021). https://doi.org:10.1016/j.immuni.2020.11.005

16 Hashimoto, K. *et al.* Single-cell transcriptomics reveals expansion of cytotoxic CD4 T cells in supercentenarians. *Proc Natl Acad Sci U S A* **116**, 24242-24251 (2019). https://doi.org:10.1073/pnas.1907883116

17 Bild, D. E. *et al.* Multi-Ethnic Study of Atherosclerosis: objectives and design. *Am J Epidemiol* **156**, 871-881 (2002). https://doi.org:10.1093/aje/kwf113

18 Binder, E. B. *et al.* Association of FKBP5 polymorphisms and childhood abuse with risk of posttraumatic stress disorder symptoms in adults. *JAMA* **299**, 1291-1305 (2008). https://doi.org:10.1001/jama.299.11.1291

19 Gillespie, C. F. *et al.* Trauma exposure and stress-related disorders in inner city primary care patients. *Gen Hosp Psychiatry* **31**, 505-514 (2009). https://doi.org:10.1016/j.genhosppsych.2009.05.003

20 Riyazi, N. *et al.* Evidence for familial aggregation of hand, hip, and spine but not knee osteoarthritis in siblings with multiple joint involvement: the GARP study. *Ann Rheum Dis* **64**, 438-443 (2005). https://doi.org:10.1136/ard.2004.024661

21 Wang, Y. *et al.* Epigenetic influences on aging: a longitudinal genome-wide methylation study in old Swedish twins. *Epigenetics* **13**, 975-987 (2018). https://doi.org:10.1080/15592294.2018.1526028

22 Powell, J. E. *et al.* The Brisbane Systems Genetics Study: genetical genomics meets complex trait genetics. *PLoS One* **7**, e35430 (2012). https://doi.org:10.1371/journal.pone.0035430

23 Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol Cell* **49**, 359-367 (2013). https://doi.org:10.1016/j.molcel.2012.10.016

24 Ivashkiv, L. B. & Donlin, L. T. Regulation of type I interferon responses. *Nat Rev Immunol* **14**, 36-49 (2014). https://doi.org:10.1038/nri3581

25 Mogilenko, D. A., Shchukina, I. & Artyomov, M. N. Immune ageing at single-cell resolution. *Nat Rev Immunol* **22**, 484-498 (2022). https://doi.org:10.1038/s41577-021-00646-4

26 Danaher, P. *et al.* Gene expression markers of Tumor Infiltrating Leukocytes. *J Immunother Cancer* **5**, 18 (2017). https://doi.org:10.1186/s40425-017-0215-8

27 Franceschi, C. & Campisi, J. Chronic inflammation (inflammaging) and its potential contribution to age-associated diseases. *J Gerontol A Biol Sci Med Sci* **69 Suppl 1**, S4-9 (2014). https://doi.org:10.1093/gerona/glu057

28 Hudgins, A. D. *et al.* Age- and Tissue-Specific Expression of Senescence Biomarkers in Mice. *Front Genet* **9**, 59 (2018). https://doi.org:10.3389/fgene.2018.00059

29 Troiani, M. *et al.* Single-cell transcriptomics identifies Mcl-1 as a target for senolytic therapy in cancer. *Nature Communications* **13** (2022). https://doi.org:10.1038/s41467-022-29824-1

30 Foroutan, M. *et al.* Single sample scoring of molecular phenotypes. *BMC Bioinformatics* **19**, 404 (2018). https://doi.org:10.1186/s12859-018-2435-4

31 Hanzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013). https://doi.org:10.1186/1471-2105-14-7

32 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005). https://doi.org:10.1073/pnas.0506580102

33 Sturm, A. *et al.* Downregulation of transposable elements extends lifespan in Caenorhabditis elegans. *Nat Commun* **14**, 5278 (2023). https://doi.org:10.1038/s41467-023-40957-9

34 Varshney, D. *et al.* SINE transcription by RNA polymerase III is suppressed by histone methylation but not by DNA methylation. *Nat Commun* **6**, 6569 (2015). https://doi.org:10.1038/ncomms7569

35 LaRocca, T. J., Cavalier, A. N. & Wahl, D. Repetitive elements as a transcriptomic marker of aging: Evidence in multiple datasets and models. *Aging Cell* **19**, e13167 (2020). https://doi.org:10.1111/acel.13167

36 Venuto, D. & Bourque, G. Identifying co-opted transposable elements using comparative epigenomics. *Dev Growth Differ* **60**, 53-62 (2018). https://doi.org:10.1111/dgd.12423

37 Fukuda, K. The role of transposable elements in human evolution and methods for their functional analysis: current status and future perspectives. *Genes Genet Syst* (2023). https://doi.org:10.1266/ggs.23-00140

38 Franceschi, C., Garagnani, P., Parini, P., Giuliani, C. & Santoro, A. Inflammaging: a new immune–metabolic viewpoint for age-related diseases. *Nature Reviews Endocrinology* **14**, 576-590 (2018). https://doi.org:10.1038/s41574-018-0059-4

39 Kaneko, H. *et al.* DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature* **471**, 325-330 (2011). https://doi.org:10.1038/nature09830

40 Andrenacci, D., Cavaliere, V. & Lattanzi, G. The role of transposable elements activity in aging and their possible involvement in laminopathic diseases. *Ageing Res Rev* **57**, 100995 (2020). https://doi.org:10.1016/j.arr.2019.100995

41 Fukuda, S. *et al.* Alu complementary DNA is enriched in atrophic macular degeneration and triggers retinal pigmented epithelium toxicity via cytosolic innate immunity. *Sci Adv* **7**, eabj3658 (2021). https://doi.org:10.1126/sciadv.abj3658

42 Deniz, O., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA modifications. *Nat Rev Genet* **20**, 417-431 (2019). https://doi.org:10.1038/s41576-019-0106-6

43 Woodcock, D. M., Lawler, C. B., Linsenmeyer, M. E., Doherty, J. P. & Warren, W. D. Asymmetric methylation in the hypermethylated CpG promoter region of the human L1 retrotransposon. *J Biol Chem* **272**, 7810-7816 (1997). https://doi.org:10.1074/jbc.272.12.7810

44 Lanciano, S. & Cristofari, G. Measuring and interpreting transposable element expression. *Nat Rev Genet* **21**, 721-736 (2020). https://doi.org:10.1038/s41576-020-0251-y

45 Borges, G. A. *et al.* Abstract PR012: Pre-malignant plasma cells exhibit a senescence-like phenotype and accumulation of transposable elements. *Cancer Research* **83**, PR012-PR012 (2023). https://doi.org:10.1158/1538-7445.Agca22-pr012

46 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015). https://doi.org:10.1093/nar/gkv007

47 He, J. *et al.* Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. *Nature Communications* **12**, 1-14 (2021). https://doi.org:10.1038/s41467-021-21808-x

48 Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* **33**, 495-502 (2015). https://doi.org:10.1038/nbt.3192

49 King's College London. (2022). King's Computational Research, Engineering and Technology Environment (CREATE). Retrieved Nov 28, 2023, from https://doi.org/10.18742/rnvf-m076.

Acknowledgment

This work was supported by the Matt Wilson Scholarship from the Faculty of Life Sciences & Medicine, King's College London. This study was supported with funding from Bristol Myers Squibb (BMS) company during this project. The scRNA-seq raw and processed data of healthy aging populations and the supercentenarians were provided by Maxim N. Artyomov from Washington University School of Medicine in St. Louis and Piero Carninci from RIKEN Center for Integrative Medical Sciences, Yokohama, respectively. We thank Dr D. Mager for critical reading of the manuscript.

Code availability. Scripts and data used in this study are available on Github (https://github.com/Karimi-Lab/TE_aging_manuscript).

Methods

Illumina HT12-v4 probes. The probe lists of Illumina Human HT-12 V3 (29431 probes) and V4 (33963) share 29311 probes in common, which is 99.6% of the list of V3. Therefore, we proceeded with HT-12 V4 probes throughout the analysis to cover all studies that used either V3 or V4. To identify the Illumina probes covering RTE regions, we first filtered out the Illumina Human HT-12 v4 probes covering intergenic regions, obtaining a list of 924 unique probes. We then overlapped these probe locations with RepeatMasker[14] regions to acquire the probes covering RTE regions. In total, we could find 232 unique probes covering RTE regions.

Generating gene signature scores using *singscore*. For each of the curated gene sets, instead of looking at individual gene expressions, we used *singscore* (version 1.20.0)[30], a method that scores gene signatures in single samples using rank-based statistics on their gene expression profiles, to calculate the gene set enrichment scores. We first compiled the microarray expression matrix for average expression values from RPM values using *limma*[46] package in R, then we used the *rankGenes* function from the *singscore* package to rank each gene sample-wise. Eventually, the *multiScore* function was used to calculate signature scores for all six gene sets at once.

Statistical analyses. All statistical analyses were performed in R (version 4.3.0). Pearson correlation test was used to determine the r and *P* values in correlation analyses. For differential expression analysis, the Wilcoxon test is used to determine the significance. The threshold to determine significance is set at P value < 0.05.

Gene Set Variation Analysis (GSVA). GSVA was performed using the *GSVA* package (version 1.48.3)[31] in R to compare the groups of low and high RTE expression groups. The gene lists were retrieved from the Molecular Signatures Database (MSigDB)[32] by searching for “inflammatory” and “dna_repair” as keywords. We further removed pathways that ended with “_DN” (as for downregulated) or “_UP” (as for upregulated) to reduce repeat and confusion. Three repetitive gene sets were removed from our analysis (GOBP_POSITIVE_REGULATION_OF_ACUTE_INFLAMMATORY_RESPONSE, GOBP_POSITIVE_REGULATION_OF_CYTOKINE_PRODUCTION_INVOLVED_IN_INFLAMMATORY_RESPONSE, GOBP_POSITIVE_REGULATION_OF_ACUTE_INFLAMMATORY_RESPONSE_TO_ANTIGENIC_STIMULUS). The filtering process generated 50 gene sets for inflammatory response and 21 gene sets for DNA repair. To detect any alterations in gene expression, the log fold

change threshold was kept low at |logFC| > 0.1, and a significance threshold of $P < 0.05$ was set.

ScRNA-seq analysis workflow for RTEs. Two single-cell transcriptomic datasets of healthy human cohorts were adopted in this study (**Supplementary Table 6**). To obtain the RTEs expression in single-cell sequencing (scRNA-seq) data, the scRNA-seq bam files were processed through scTE[47] pipeline. More specifically, for each cell in each sample, the read counts for different classes and families of RTEs were generated using the scTE method. Subsequently, in each sample, we calculated the cumulative read counts for each RTE class per annotated cell type by adding the number of reads belonging to all the cells of each cell type. This resulted in pseudo-bulk read counts of RTE classes for different cell types in each sample.

In parallel, we also calculated the pseudo-bulk read counts for all genes per cell type in each sample using the genic read counts embedded in Seurat[48] objects collected for the two scRNA-seq datasets that we employed in our study (**Supplementary Table 6**). Concatenating pseudo-bulk RTE and gene read counts, we performed Reads Per Million (RPM) normalization per cell type per sample. For each cell type, the denominator for RPM normalization was the cumulative counts of genes and RTE classes for that cell type.

To calculate the age-associated gene signature scores, we converted the scRNA-seq read counts to bulk RNA-seq read counts by summing up the number of reads of all cells of the same type for each gene. This resulted in a bulk RNA-seq gene count matrix from which we calculated each sample's RPM values per gene. Next, the RPM matrix was provided to *singscore* to calculate gene signature scores. Lastly, we divided the samples into high vs low gene signature scores by using the median as the cut-off.

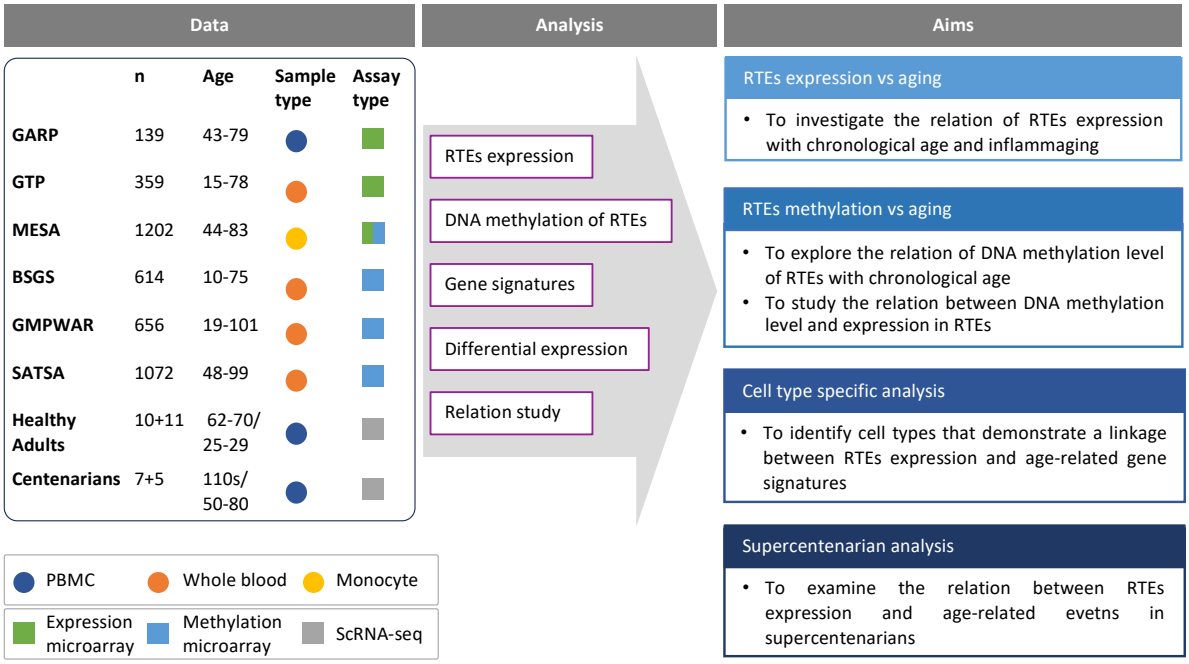


Figure 1. Conceptual framework and the study design. We collected published datasets of human blood samples for gene expression, DNA methylation, and single-cell transcriptomic data. The analysis aimed to study the relation between the expression and DNA methylation of RTEs versus chronological and biological aging in large human cohorts. The single-cell transcriptomic datasets were employed for cell type-specific analysis of RTEs in PBMC to identify the relation between RTEs expression and ageing events for annotated cell types within old versus young PBMC samples.

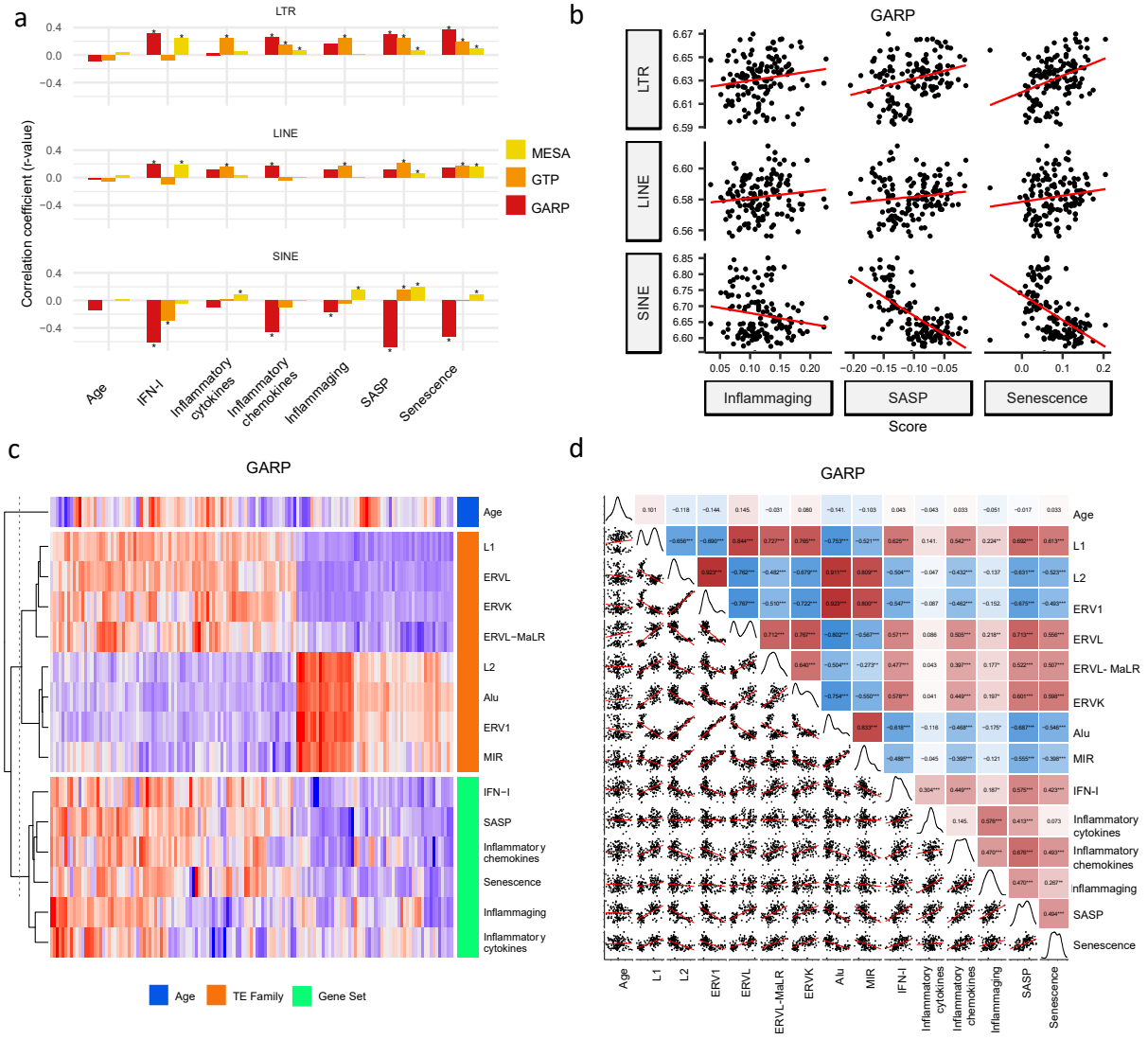
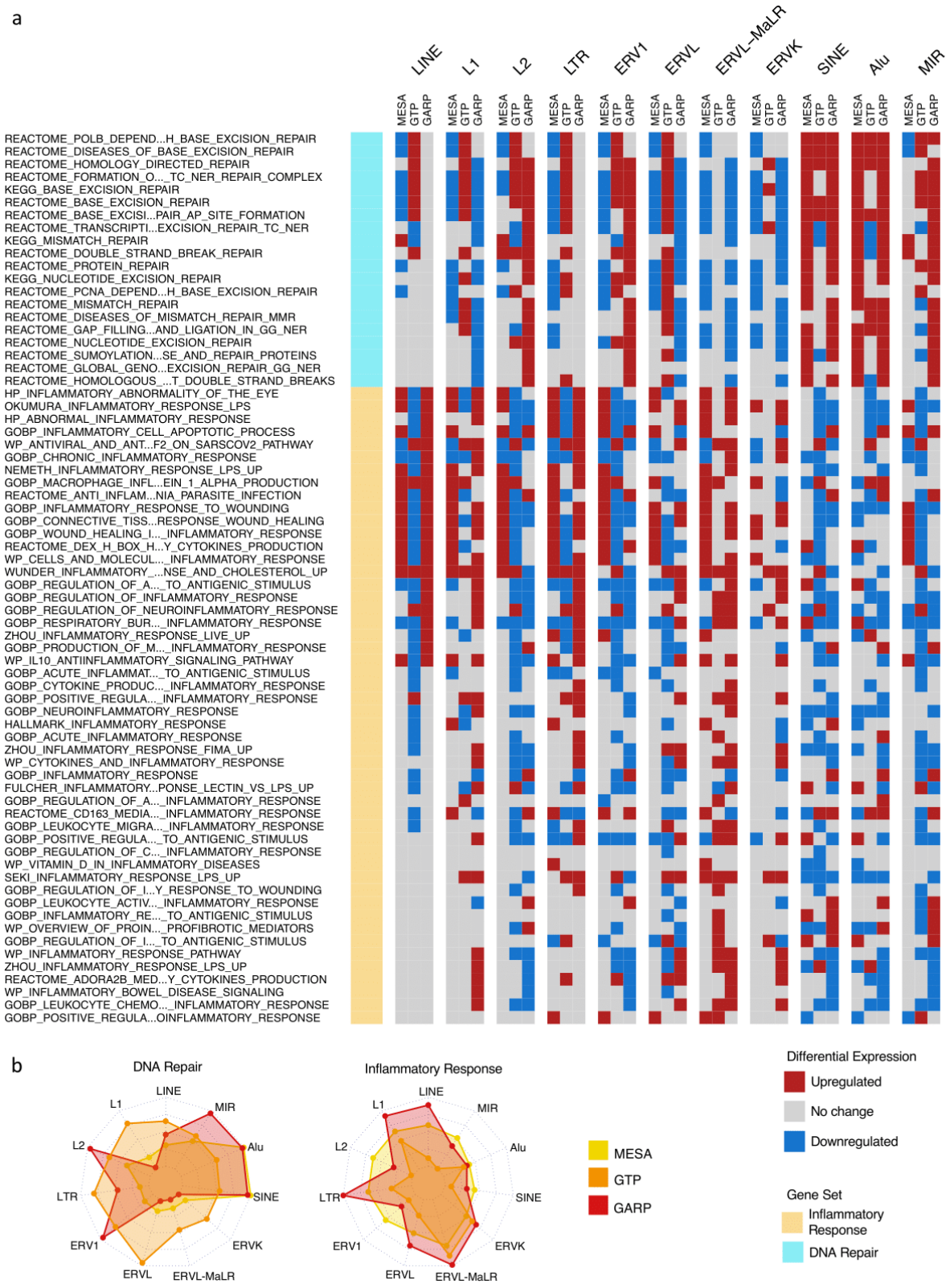


Figure 2. Correlation analysis between RTEs expression, chronological age, and age-associated gene signature scores. a, No correlation between RTEs expression and chronological age versus positive correlations between age-associated gene signature scores and LINE and LTR expressions. Pair-wise correlation coefficients were calculated between the expression of different RTE classes (LTR, LINE & SINE) and chronological age and six age-associated gene signature scores in monocytes (MESA), PBMCs (GARP), and the whole blood (GTP). **b**, Scatter plots displaying a positive correlation between LINE and LTR expressions and inflammaging, SASP, and senescence gene signature scores in PBMCs. **c**, Different families of RTEs were divided into two major groups based on their correlation and inverse correlation with age-associated signature scores in PBMC samples. **d**, Correlation matrix depicting all pair-wise combinations to identify the correlation between chronological age, RTE family expressions, and six age-associated signature scores in PBMCs. ** $P \leq 0.01$, *** $P \leq 0.001$, Pearson's correlation. MESA, n=1202; GARP, n=139, GTP, n=359.

589


a
LINE
L1
L2
LTR
ERV1
ERVL
ERVL-MaLR
ERVK
SINE
Alu
MIR
MESA
GTP
GARP
REACTOME_POLB_DEPEND...H_BASE_EXCISION_REPAIR
REACTOME_DISEASES_OF_BASE_EXCISION_REPAIR
REACTOME_HOMOLOGY_DIRECTED_REPAIR
REACTOME_FORMATION_O..._TC_NER_REPAIR_COMPLEX
KEGG_BASE_EXCISION_REPAIR
REACTOME_BASE_EXCISION_REPAIR
REACTOME_BASE_EXCISI...PAIR_AP_SITE_FORMATION
REACTOME_TRANSCRIPTI...EXCISION_REPAIR_TC_NER
KEGG_MISMATCH_REPAIR
REACTOME_DOUBLE_STRAND_BREAK_REPAIR
REACTOME_PROTEIN_REPAIR
KEGG_NUCLEOTIDE_EXCISION_REPAIR
REACTOME_PCNA_DEPEND...H_BASE_EXCISION_REPAIR
REACTOME_MISMATCH_REPAIR
REACTOME_DISEASES_OF_MISMATCH_REPAIR_MMR
REACTOME_GAP_FILLING...AND_LIGATION_IN_GG_NER
REACTOME_NUCLEOTIDE_EXCISION_REPAIR
REACTOME_SUMOYLATION...SE_AND_REPAIR_PROTEINS
REACTOME_GLOBAL_GENO...EXCISION_REPAIR_GG_NER
REACTOME_HOMOLOGOUS_...T_DOUBLE_STRAND_BREAKS
HP_INFLAMMATORY_ABNORMALITY_OF_THE_EYE
OKUMURA_INFLAMMATORY_RESPONSE_LPS
HP_ABNORMAL_INFLAMMATORY_RESPONSE
GOBP_INFLAMMATORY_CELL_APOPTOTIC_PROCESS
WP_ANTIVIRAL_AND_ANT...F2_ON_SARSCOV2_PATHWAY
GOBP_CHRONIC_INFLAMMATORY_RESPONSE
NEMETH_INFLAMMATORY_RESPONSE_LPS_UP
GOBP_MACROPHAGE_INFL...EIN_1_ALPHA_PRODUCTION
REACTOME_ANTI_INFLAM...NIA_PARASITE_INFECTION
GOBP_INFLAMMATORY_RESPONSE_TO_WOUNDING
GOBP_CONNECTIVE_TISS...RESPONSE_WOUND_HEALING
GOBP_WOUND_HEALING_I..._INFLAMMATORY_RESPONSE
REACTOME_DEX_H_BOX_H...Y_CYTOKINES_PRODUCTION
WP_CELLS_AND_MOLECUL..._INFLAMMATORY_RESPONSE
WUNDER_INFLAMMATORY_...NSE_AND_CHOLESTEROL_UP
GOBP_REGULATION_OF_A..._TO_ANTIGENIC_STIMULUS
GOBP_REGULATION_OF_INFLAMMATORY_RESPONSE
GOBP_REGULATION_OF_NEUROINFLAMMATORY_RESPONSE
GOBP_RESPIRATORY_BUR..._INFLAMMATORY_RESPONSE
ZHOU_INFLAMMATORY_RESPONSE_LIVE_UP
GOBP_PRODUCTION_OF_M..._INFLAMMATORY_RESPONSE
WP_IL10_ANTIINFLAMMATORY_SIGNALING_PATHWAY
GOBP_ACUTE_INFLAMMAT..._TO_ANTIGENIC_STIMULUS
GOBP_CYTOKINE_PRODUC..._INFLAMMATORY_RESPONSE
GOBP_POSITIVE_REGULA..._INFLAMMATORY_RESPONSE
GOBP_NEUROINFLAMMATORY_RESPONSE
HALLMARK_INFLAMMATORY_RESPONSE
GOBP_ACUTE_INFLAMMATORY_RESPONSE
ZHOU_INFLAMMATORY_RESPONSE_FIMA_UP
WP_CYTOKINES_AND_INFLAMMATORY_RESPONSE
GOBP_INFLAMMATORY_RESPONSE
FULCHER_INFLAMMATORY...PONSE_LECTIN_VS_LPS_UP
GOBP_REGULATION_OF_A..._INFLAMMATORY_RESPONSE
REACTOME_CD163_MEDIA..._INFLAMMATORY_RESPONSE
GOBP_LEUKOCYTE_MIGRA..._INFLAMMATORY_RESPONSE
GOBP_POSITIVE_REGULA..._TO_ANTIGENIC_STIMULUS
GOBP_REGULATION_OF_C..._INFLAMMATORY_RESPONSE
WP_VITAMIN_D_IN_INFLAMMATORY_DISEASES
SEKI_INFLAMMATORY_RESPONSE_LPS_UP
GOBP_REGULATION_OF_I...Y_RESPONSE_TO_WOUNDING
GOBP_LEUKOCYTE_ACTIV..._INFLAMMATORY_RESPONSE
GOBP_INFLAMMATORY_RE..._TO_ANTIGENIC_STIMULUS
WP_OVERVIEW_OF_PROIN..._PROFIBROTIC_MEDIATORS
GOBP_REGULATION_OF_I..._TO_ANTIGENIC_STIMULUS
WP_INFLAMMATORY_RESPONSE_PATHWAY
ZHOU_INFLAMMATORY_RESPONSE_LPS_UP
REACTOME_ADORA2B_MED...Y_CYTOKINES_PRODUCTION
WP_INFLAMMATORY_BOWEL_DISEASE_SIGNALING
GOBP_LEUKOCYTE_CHEMO..._INFLAMMATORY_RESPONSE
GOBP_POSITIVE_REGULA...OINFLAMMATORY_RESPONSE
b
DNA Repair
Inflammatory Response
MESA
GTP
GARP
Differential Expression
Upregulated
No change
Downregulated
Gene Set
Inflammatory Response
DNA Repair

590

591

Figure 3. Upregulation of DNA repair vs inflammatory responses for samples with high expression of SINE vs LTR and LINE in the MESA and GARP cohorts. a, GSVA demonstrates increased activity of DNA repair pathways in the group of samples with high vs low SINE expression in the MESA and GARP cohorts. In contrast, the inflammatory response is upregulated in the sample groups highly expressing LINE and LTR classes and families in the MESA and GARP cohorts. The samples in each cohort were divided into low (1st quartile), medium (2nd and 3rd quartile), and high (4th quartile) expression groups based on the expression of RTE classes or families. GSVA was applied on high vs low groups for each class and family of RTEs. The threshold for differential expression is set at $|logFC| > 0.1$ and $P < 0.05$ (**online methods**). **b**, The Radar plot shows the difference between the number of upregulated versus downregulated gene sets related to DNA repair and inflammatory response in each cohort. While high expression of SINE and Alu is significantly associated with high number of up-regulated DNA-repair gene sets, LINE and L1 expressions are highly related to high number of activated gene sets related to inflammatory response in the MESA and GARP cohorts. This result is not highly supported by the GTP cohort, more likely due to the low number of probes in this cohort.

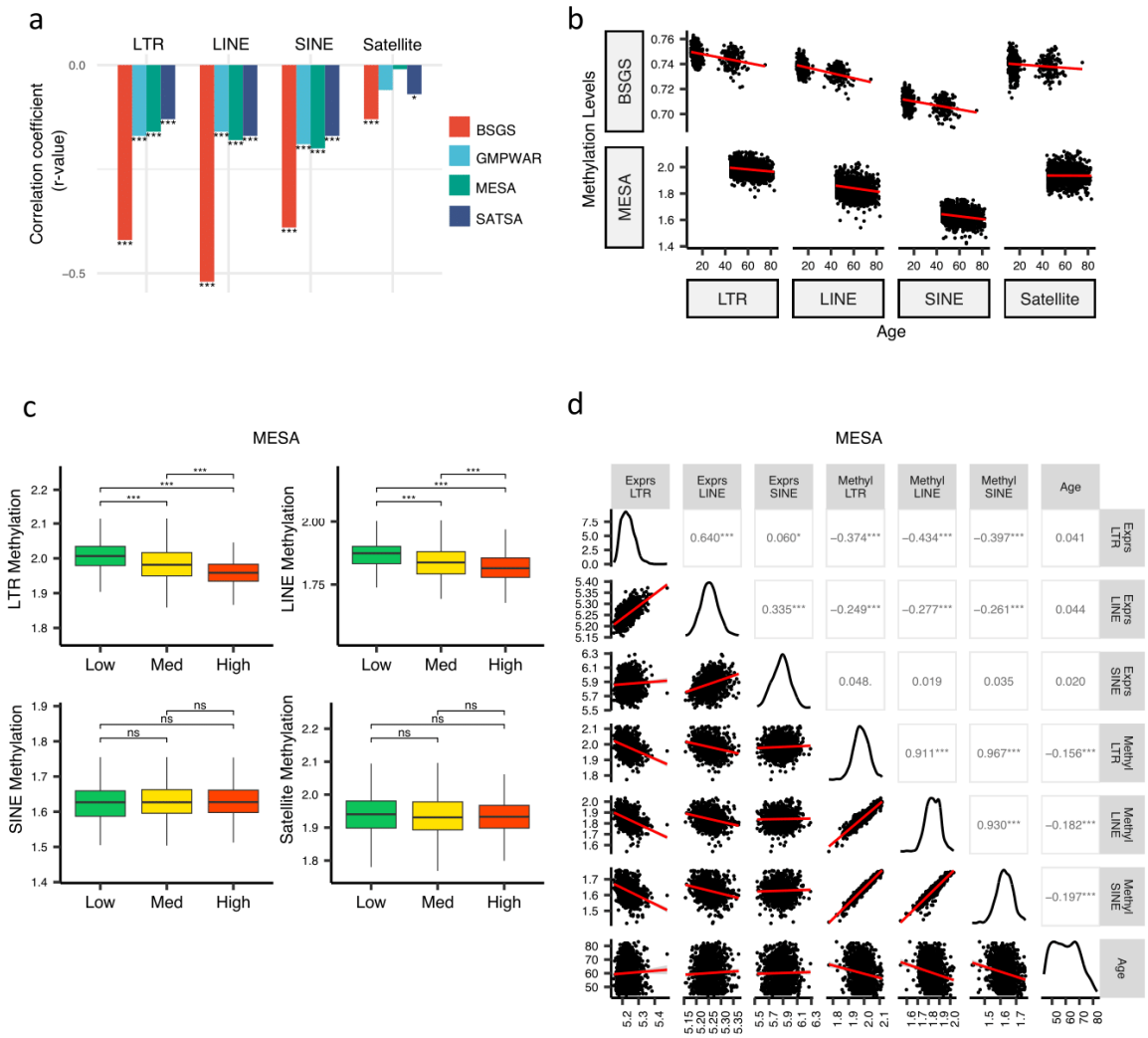


Figure 4. Inverse correlation of DNA methylation levels of RTEs with the chronological age and the RTEs expression except for SINE expression. a, b, Methylation levels of RTE classes inversely correlate with chronological age in monocyte (MESA) and whole blood (BSGS, SATSA, and GMPWAR) samples. Satellite DNA is included as a control group. **c**, Methylation levels versus low (1st quartile), medium (2nd and 3rd quartile), and high (4th quartile) expressions of RTE classes in monocytes (MESA). Wilcoxon test; ns: not significant. **d**, Correlation matrix for RTE expressions and methylation levels, and chronological age. * $P \leq 0.05$, **** $P \leq 0.0001$, Pearson's correlation. MESA, n=1202; BSGS, n=614; GMPWAR, n=656; SATSA, n=1072.

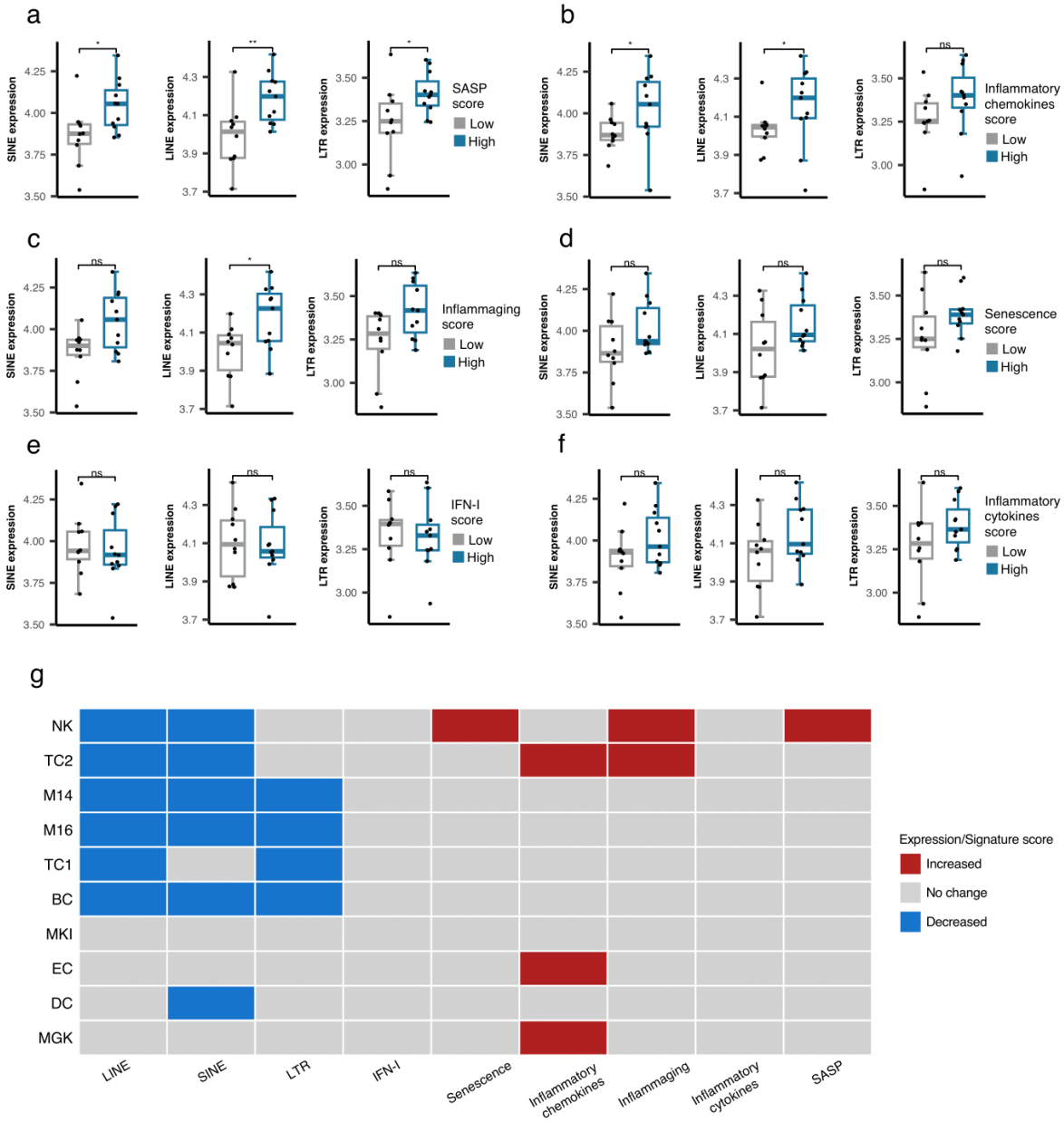
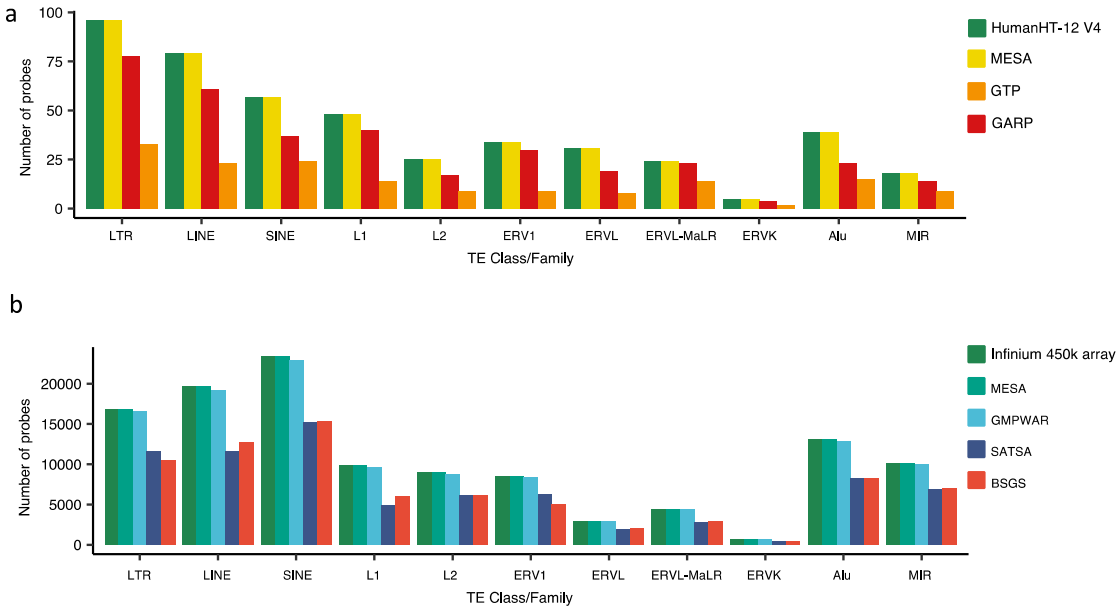
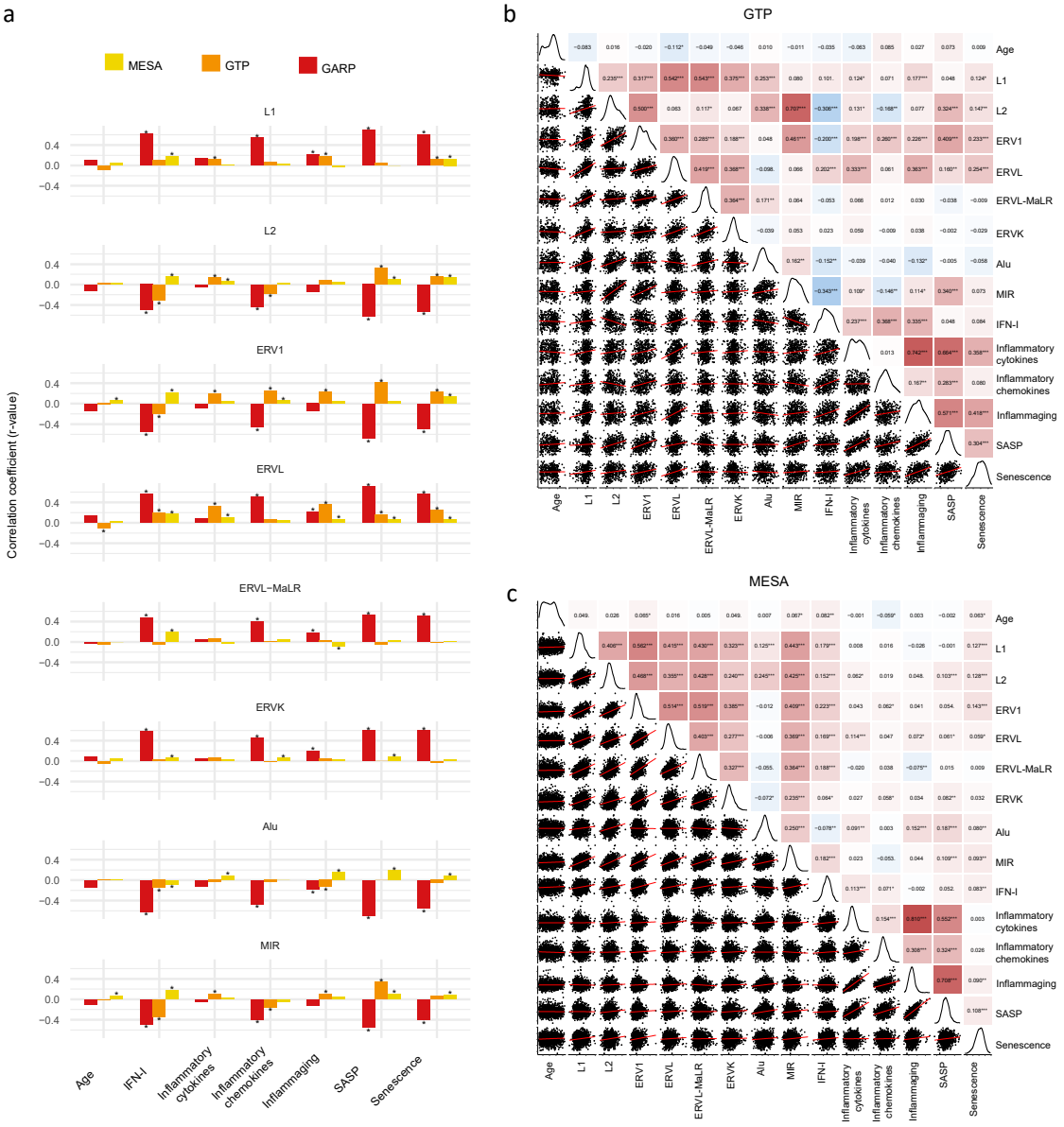


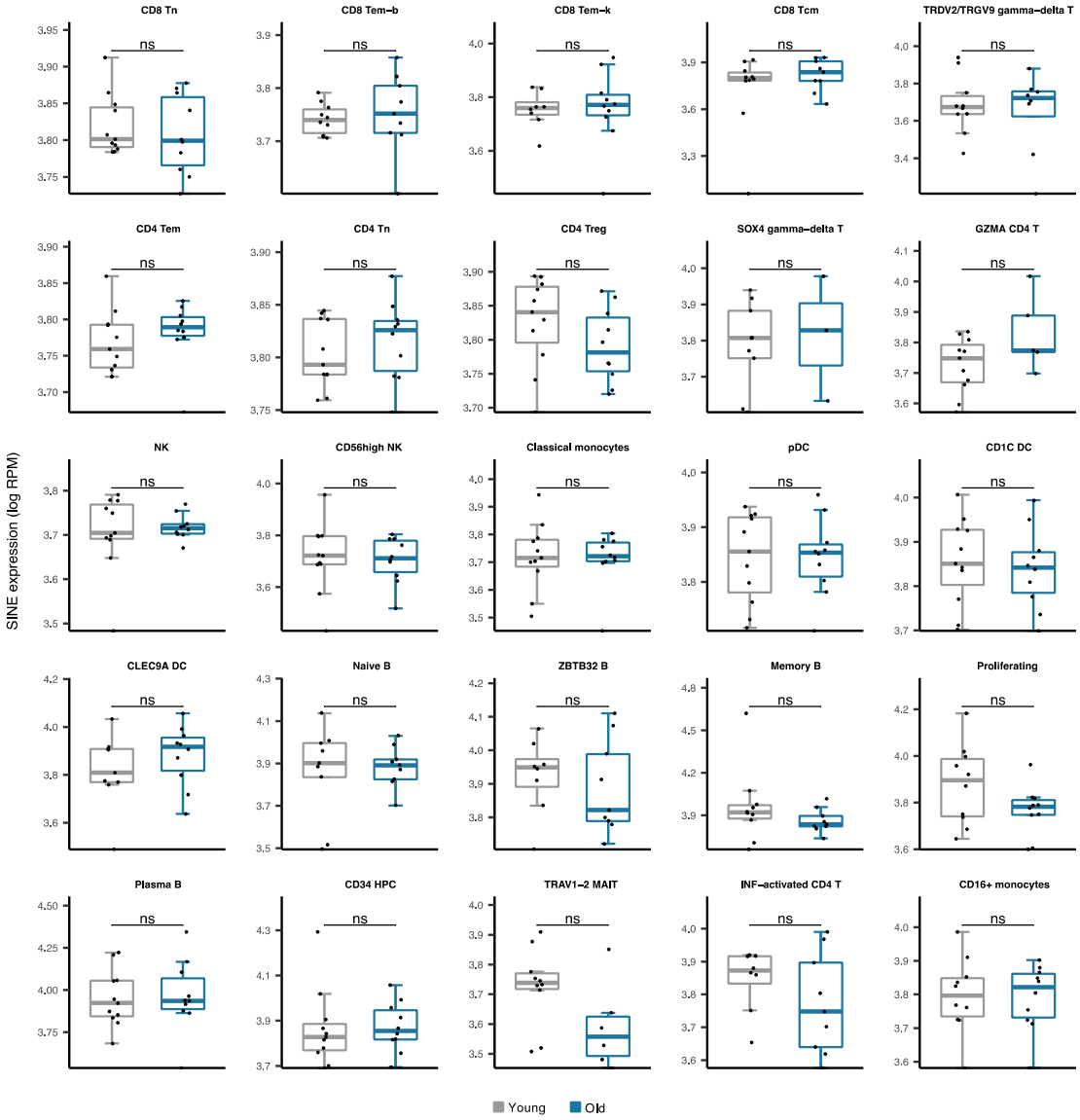
Figure 5. Cell type-specific analysis of RTEs expression vs age-associated gene signature scores in two PBMC scRNA-seq cohorts. a-f, Unique increased expression of RTEs in the Plasma B cells of the samples with high age-associated gene signature scores indicates the potential role of plasma B cells in ageing. Wilcoxon test. n = 21. **g,** Decreased expressions of RTE classes and increased age-associated gene signature scores in multiple annotated cell types obtained from the PBMCs of supercentenarians compared to ordinary elderlies as control. Supercentenarians, n=7; control, n=5, age 50s to 80s. Wilcoxon test was applied to identify the significant changes. NK, Natural killer cell; BC, B-cell; TC1, T-cell 1; TC2, T-cell 2; M14, CD14+ monocyte; M16, CD16+ monocyte; EC, Erythrocytes; MKI, MKI67+ proliferating cell; DC, Dendritic cell; MGK, Megakaryocyte.

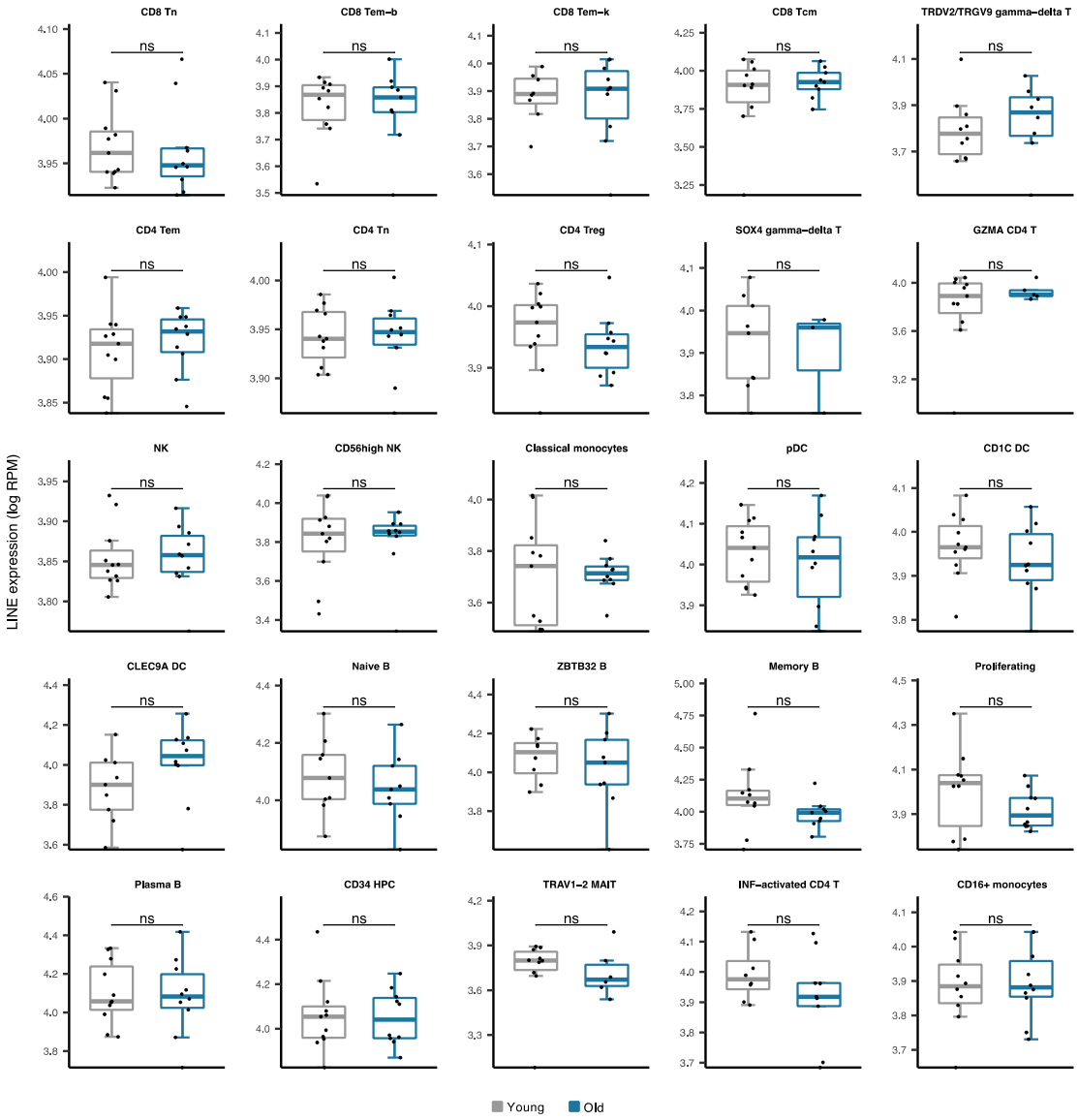


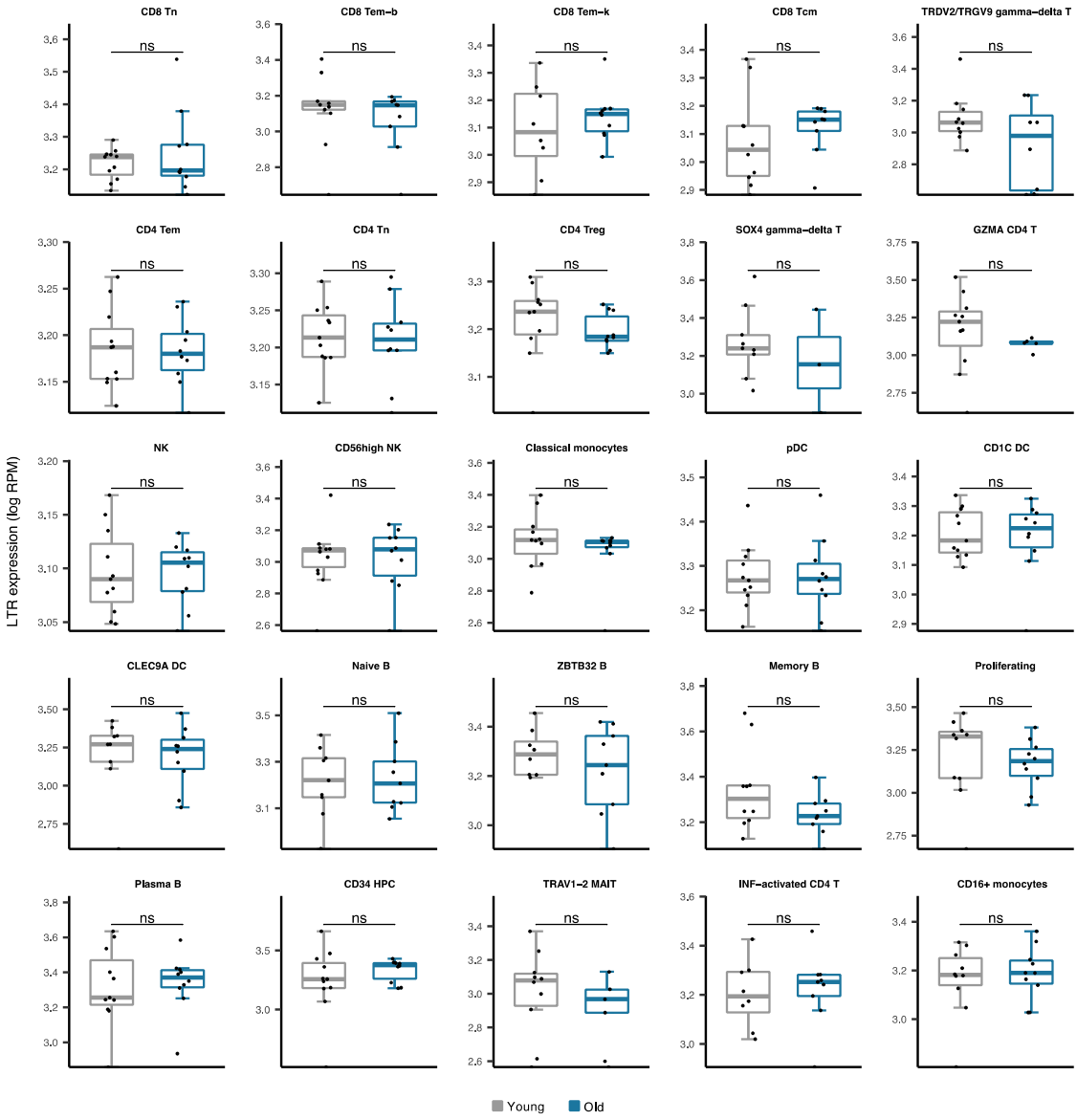
657
658
659
660
661
662
663

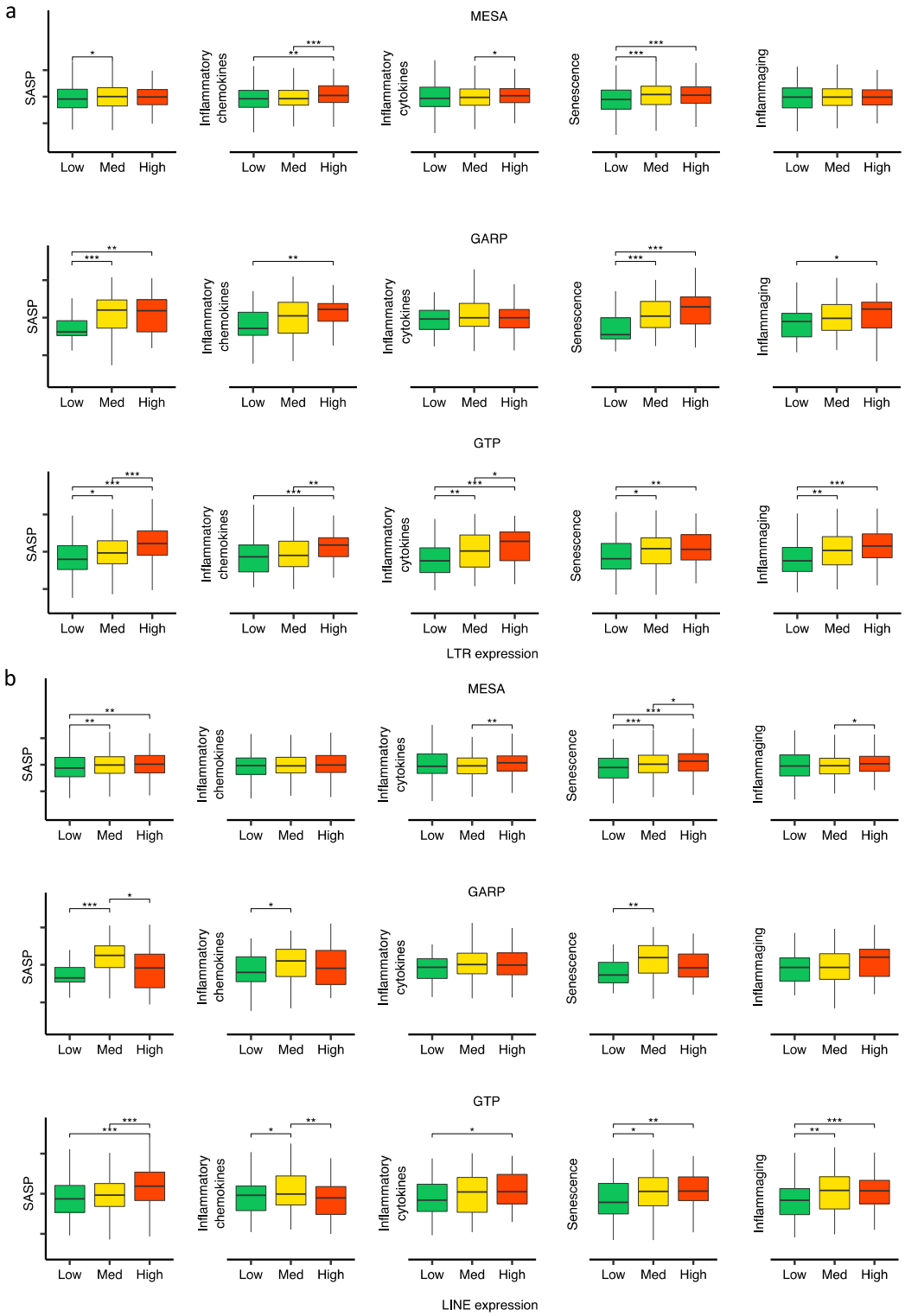


Supplementary Figure 2. Correlation analysis between age-associated gene signatures and RTE family expressions in human cohorts. a, Weak correlation between a few RTE families and chronological age and strong positive correlation between L1, ERVL, ERVK, and MaLR with age-associated signature scores in PBMC samples (GARP cohort). **b, c**, Correlation matrix depicting all pair-wise combinations to identify the correlation between chronological age, RTE family expressions, and six age-associated gene expressions in monocytes (MESA) and the whole blood (GTP). * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, Pearson's correlation.

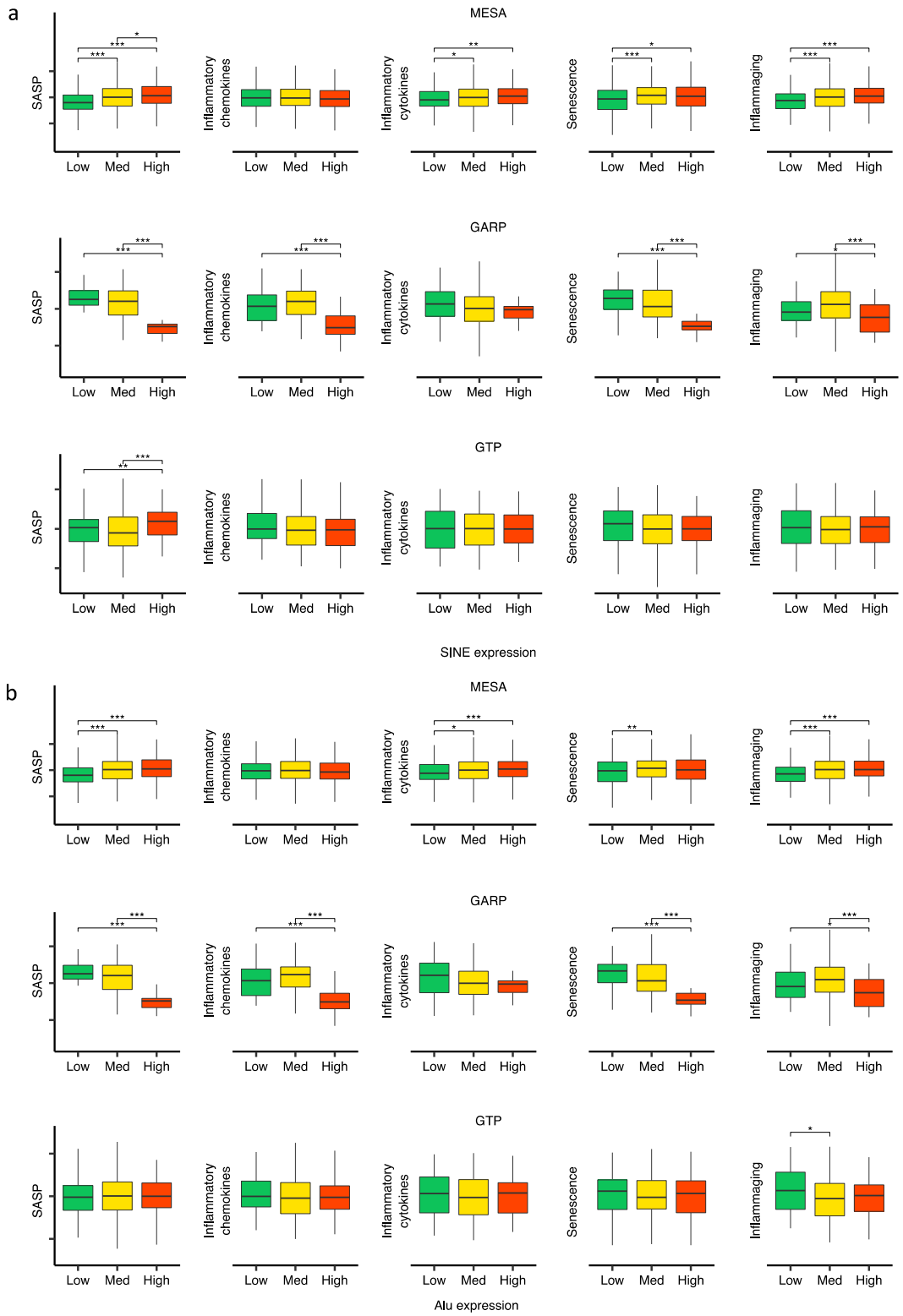

CD8 Tn
CD8 Tem−b
CD8 Tem−k
CD8 Tcm
TRDV2/TRGV9 gamma−delta T
CD4 Tem
CD4 Tn
CD4 Treg
SOX4 gamma−delta T
GZMA CD4 T
NK
CD56high NK
Classical monocytes
pDC
CD1C DC
CLEC9A DC
Naive B
ZBTB32 B
Memory B
Proliferating
Plasma B
CD34 HPC
TRAV1−2 MAIT
INF−activated CD4 T
CD16+ monocytes
ns
SINE expression (log RPM)
Young
Old


CD8 Tn
CD8 Tem−b
CD8 Tem−k
CD8 Tcm
TRDV2/TRGV9 gamma−delta T
CD4 Tem
CD4 Tn
CD4 Treg
SOX4 gamma−delta T
GZMA CD4 T
NK
CD56high NK
Classical monocytes
pDC
CD1C DC
CLEC9A DC
Naive B
ZBTB32 B
Memory B
Proliferating
Plasma B
CD34 HPC
TRAV1−2 MAIT
INF−activated CD4 T
CD16+ monocytes
ns
LINE expression (log RPM)
Young
Old

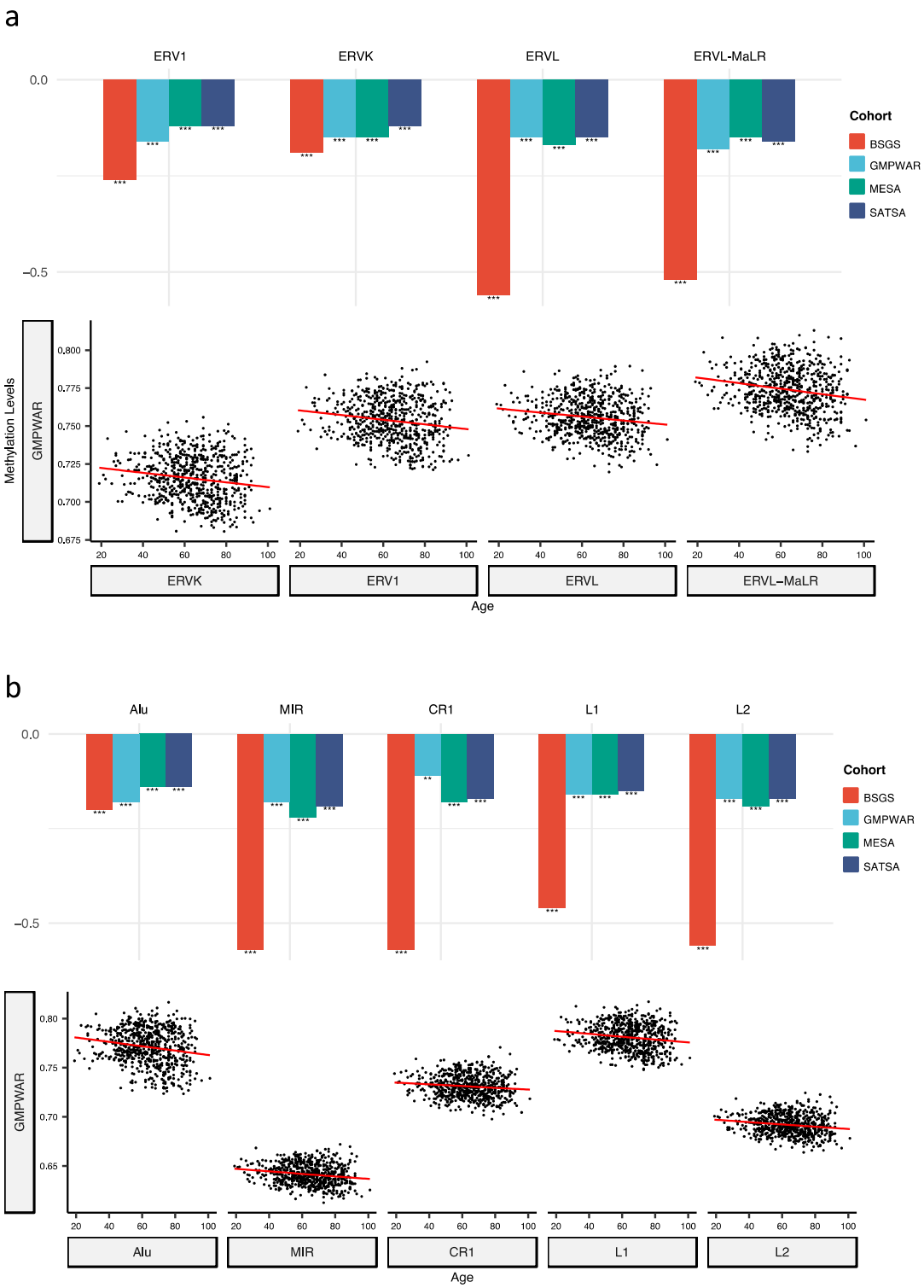

CD8 Tn
CD8 Tem−b
CD8 Tem−k
CD8 Tcm
TRDV2/TRGV9 gamma−delta T
CD4 Tem
CD4 Tn
CD4 Treg
SOX4 gamma−delta T
GZMA CD4 T
NK
CD56high NK
Classical monocytes
pDC
CD1C DC
CLEC9A DC
Naive B
ZBTB32 B
Memory B
Proliferating
Plasma B
CD34 HPC
TRAV1−2 MAIT
INF−activated CD4 T
CD16+ monocytes
ns
LTR expression (log RPM)
Young
Old



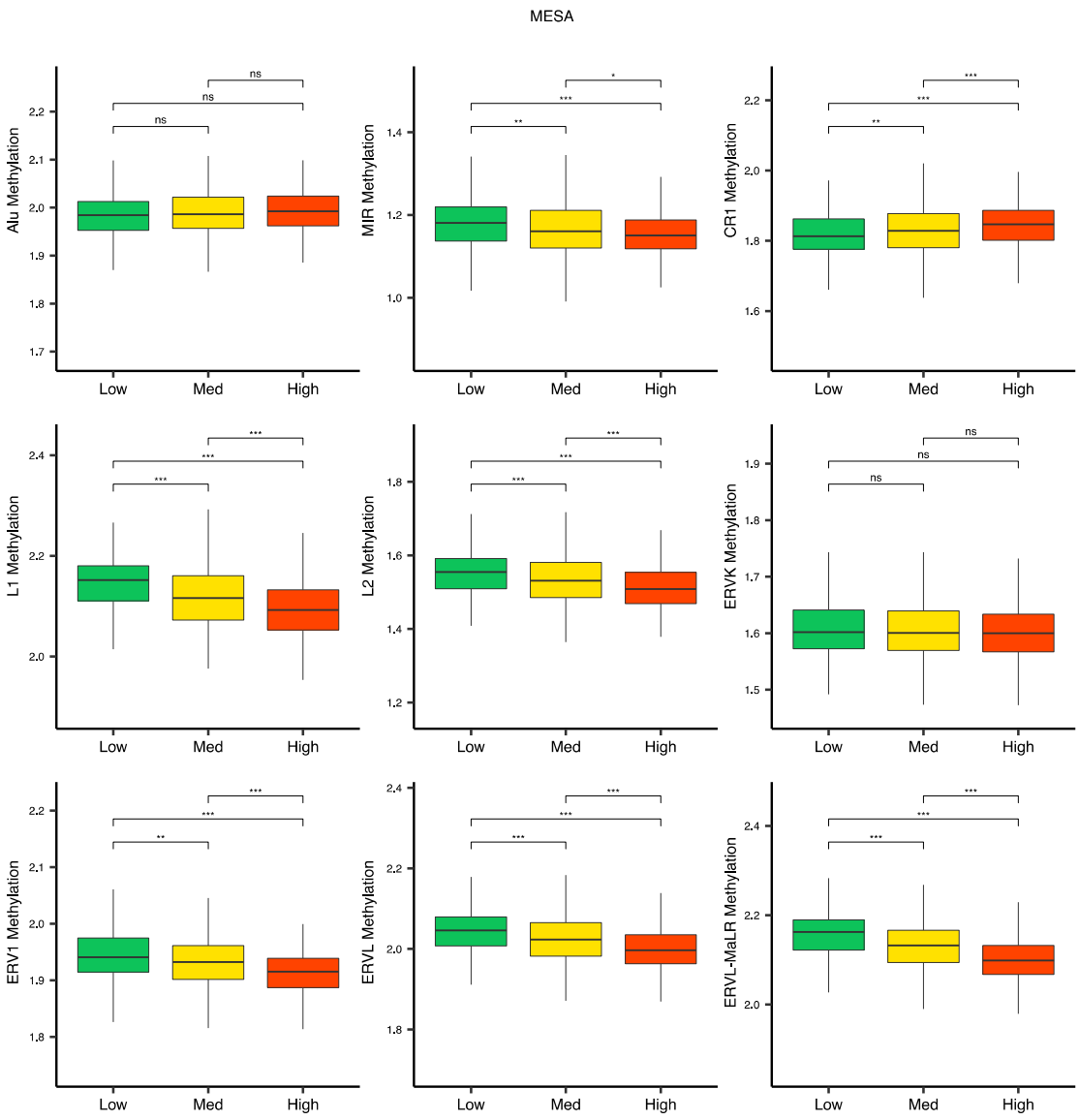
Supplementary Figure 6. Increasing trend of age-associated gene signature scores in high vs low LTR and LINE expression groups in the three human cohorts. a,b The samples in each cohort were divided into (low (1st quartile), medium (2nd and 3rd quartile), and high (4th quartile) LTR and LINE expression groups, respectively. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, Wilcoxon test.



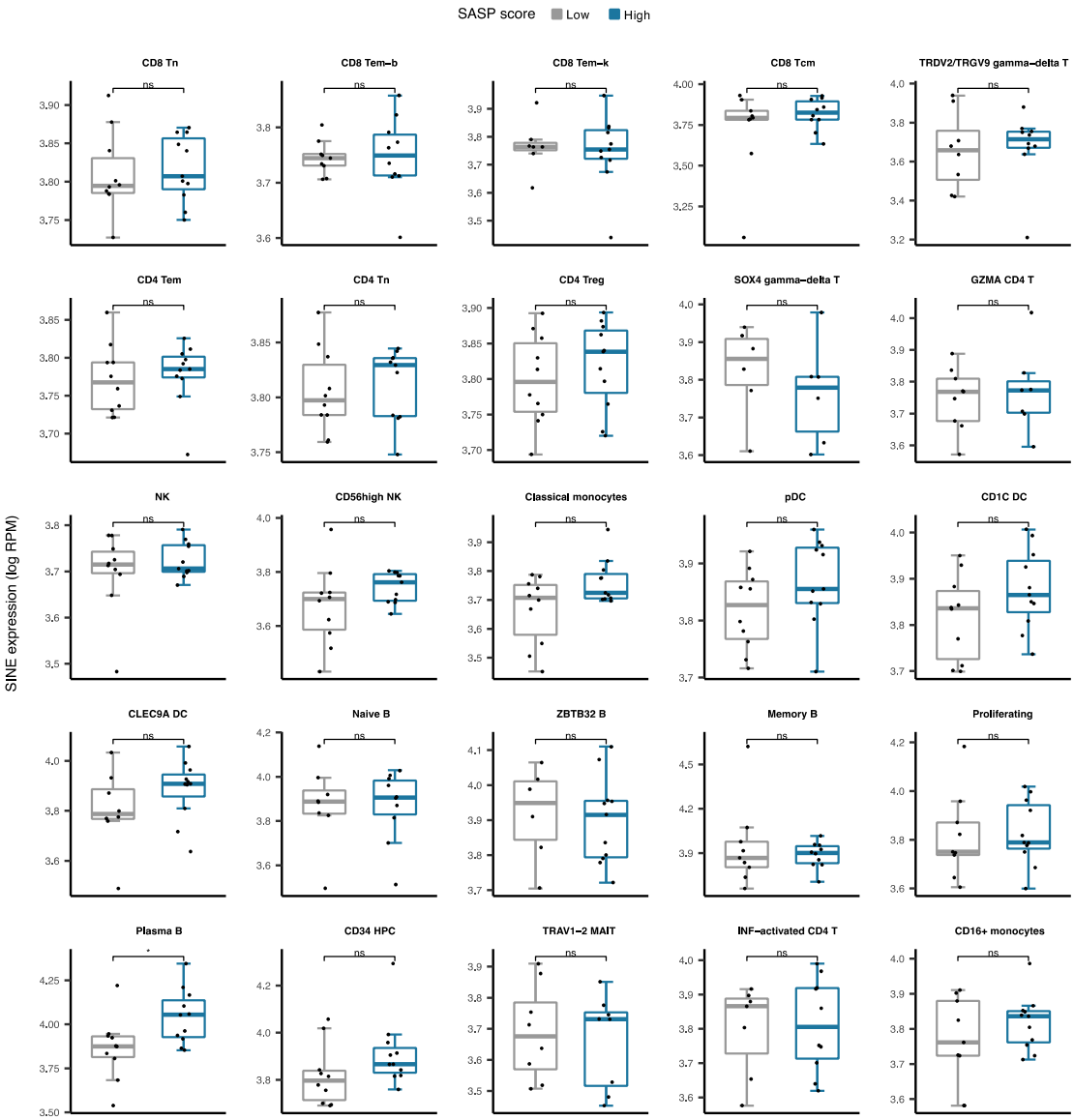
Supplementary Figure 7. Different patterns of age-associated signature scores in MESA vs GARP vs GTP for high vs low SINE and Alu expression groups in the three human cohorts. a, b, The samples in each cohort were divided into low (1st quartile), medium (2nd and 3rd quartile), and high (4th quartile) SINE and Alu expression groups, respectively. * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$, Wilcoxon test.

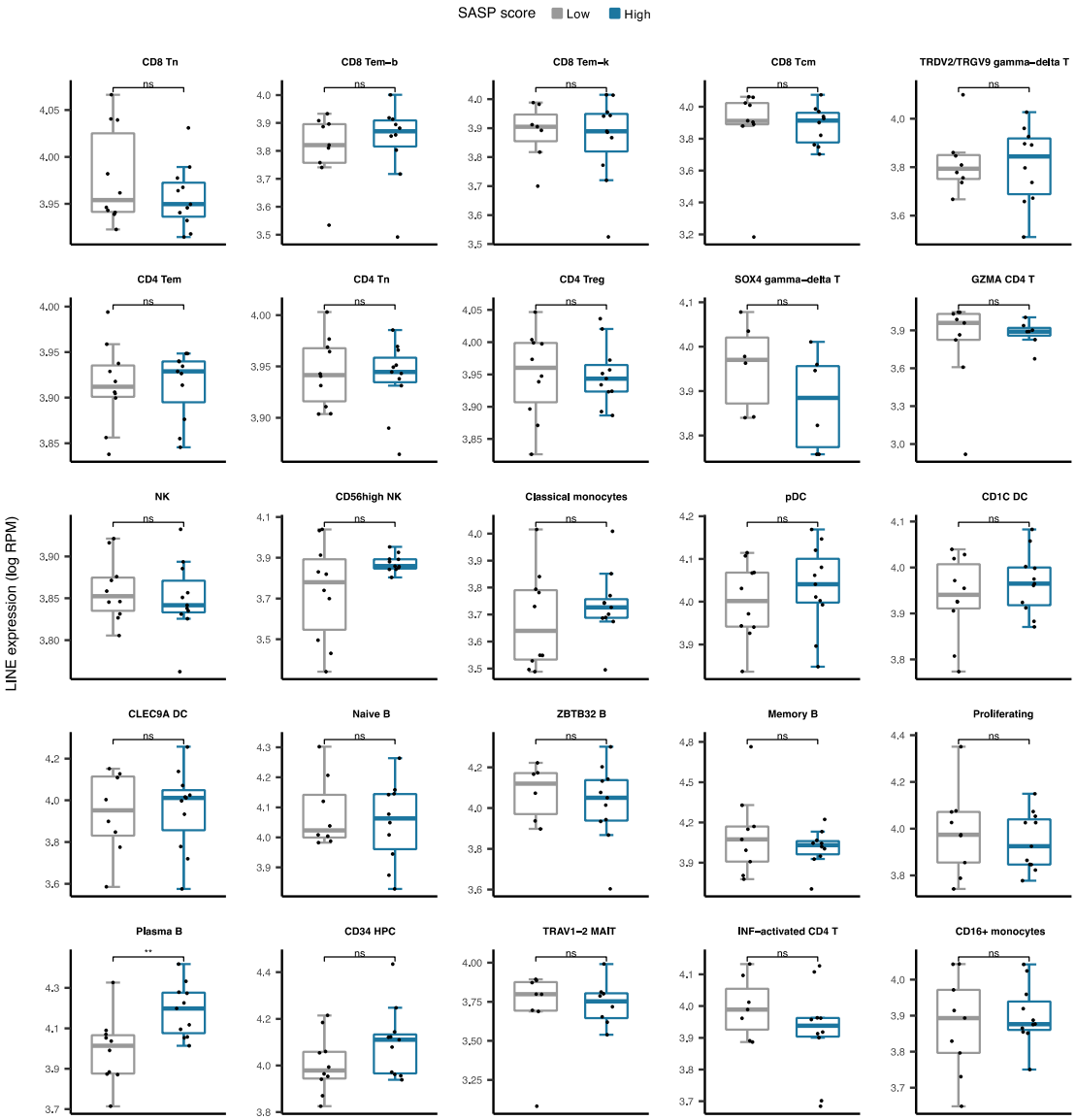


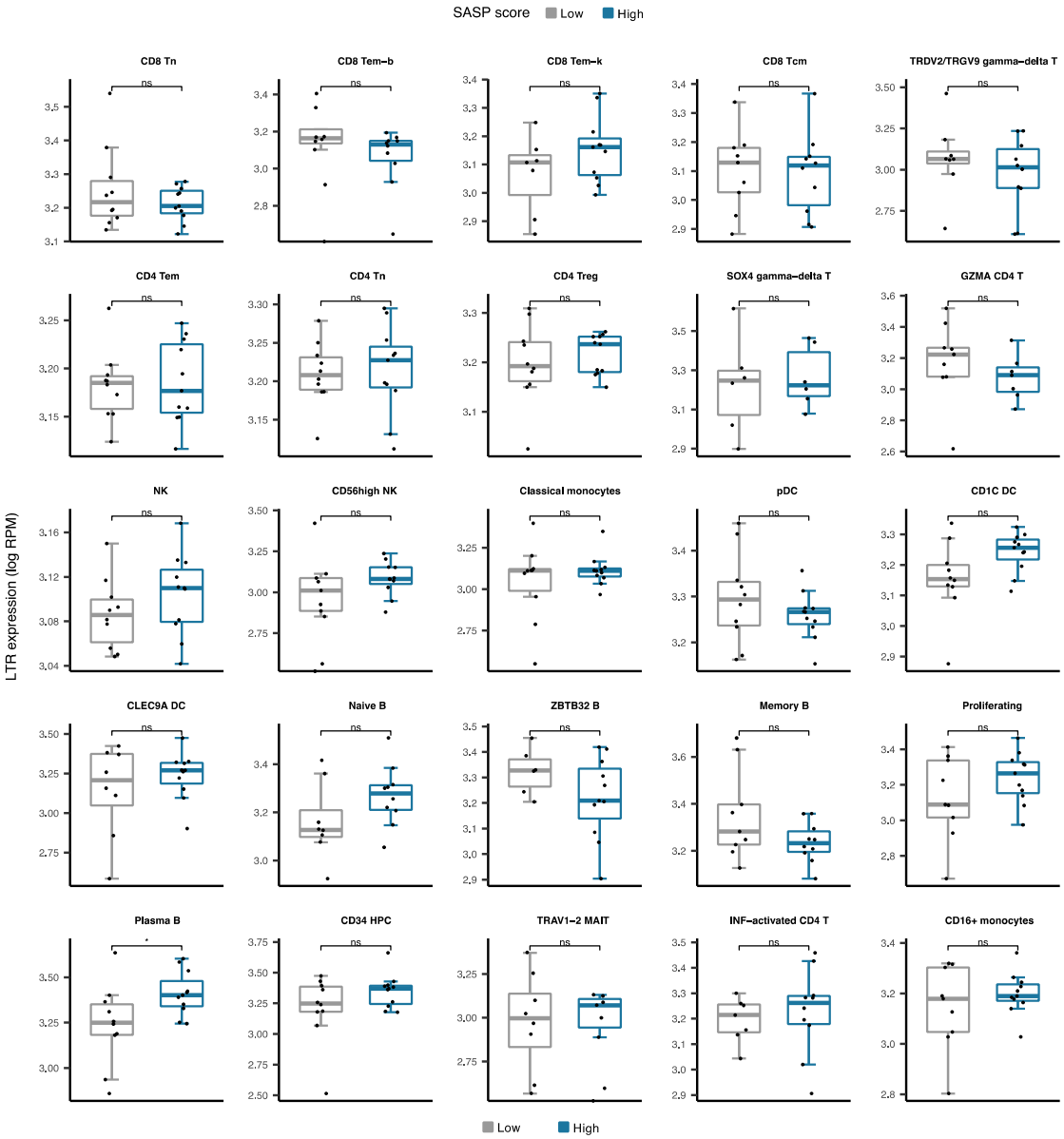
Supplementary Figure 8. Correlation of methylation levels of RTE families with chronological age in monocyte (MESA) and whole blood (BSGS, SATSA, and GMPWAR) samples. a, b, Methylation levels of LTR and LINE/SINE families negatively correlate with chronological age in monocytes (MESA) and the whole blood (BSGS, SATSA, and GMPWAR). *** $P \leq 0.001$, Wilcoxon test.

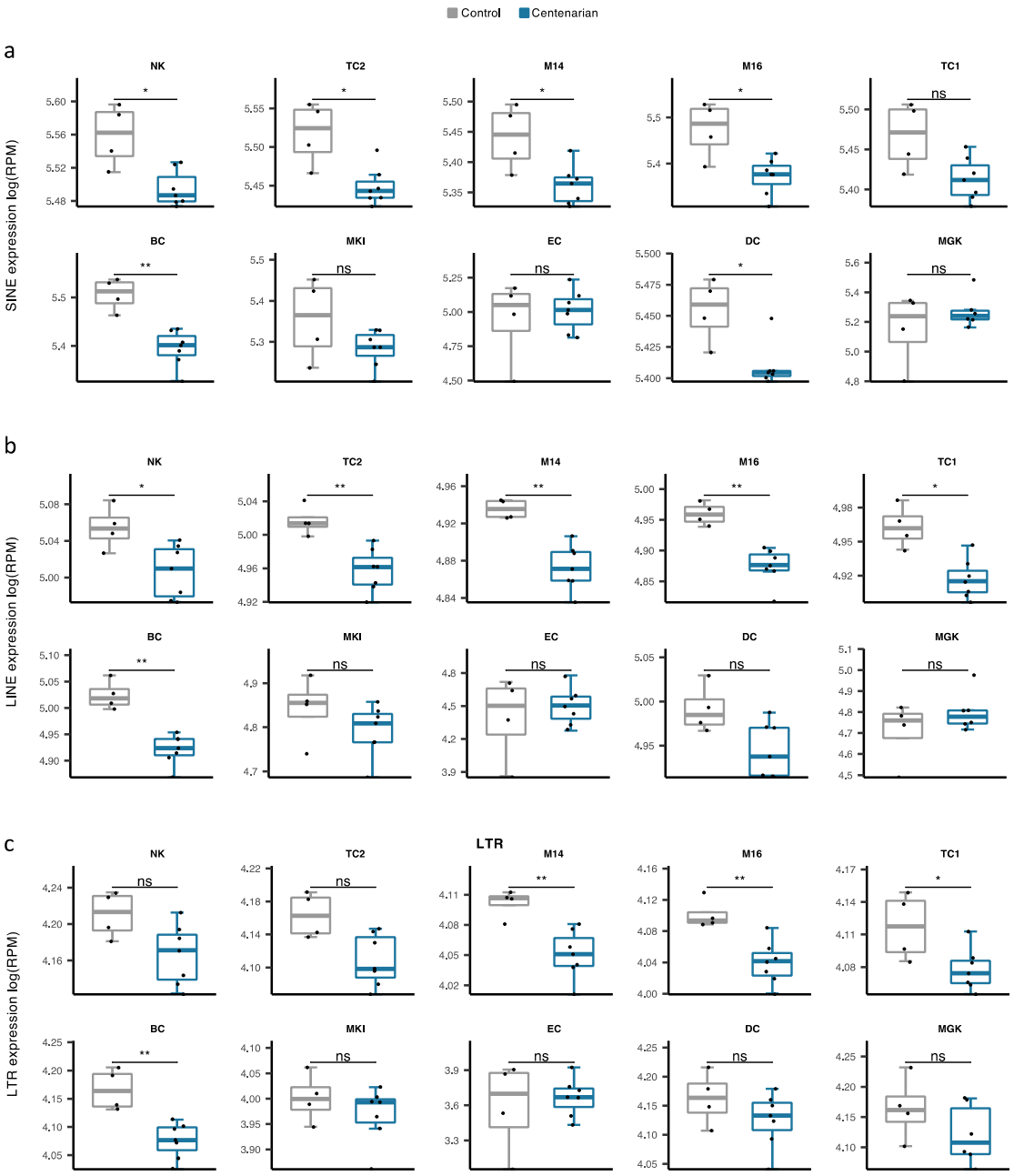


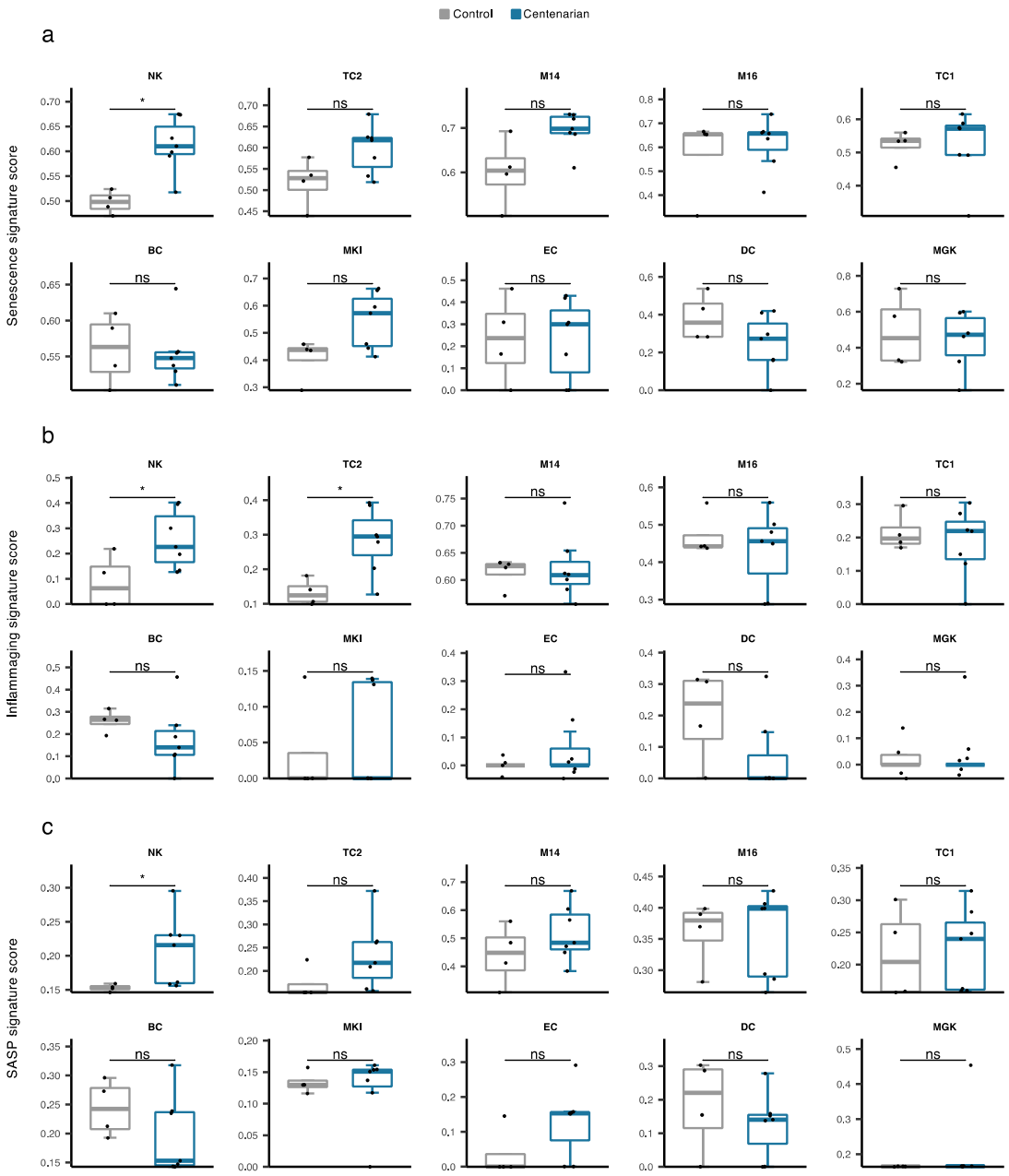
71
71
71
71
71
71
71


SASP score
Low
High
SINE expression (log RPM)
CD8 Tn
CD8 Tem−b
CD8 Tem−k
CD8 Tcm
TRDV2/TRGV9 gamma−delta T
CD4 Tem
CD4 Tn
CD4 Treg
SOX4 gamma−delta T
GZMA CD4 T
NK
CD56high NK
Classical monocytes
pDC
CD1C DC
CLEC9A DC
Naive B
ZBTB32 B
Memory B
Proliferating
Plasma B
CD34 HPC
TRAV1−2 MAIT
INF−activated CD4 T
CD16+ monocytes
ns

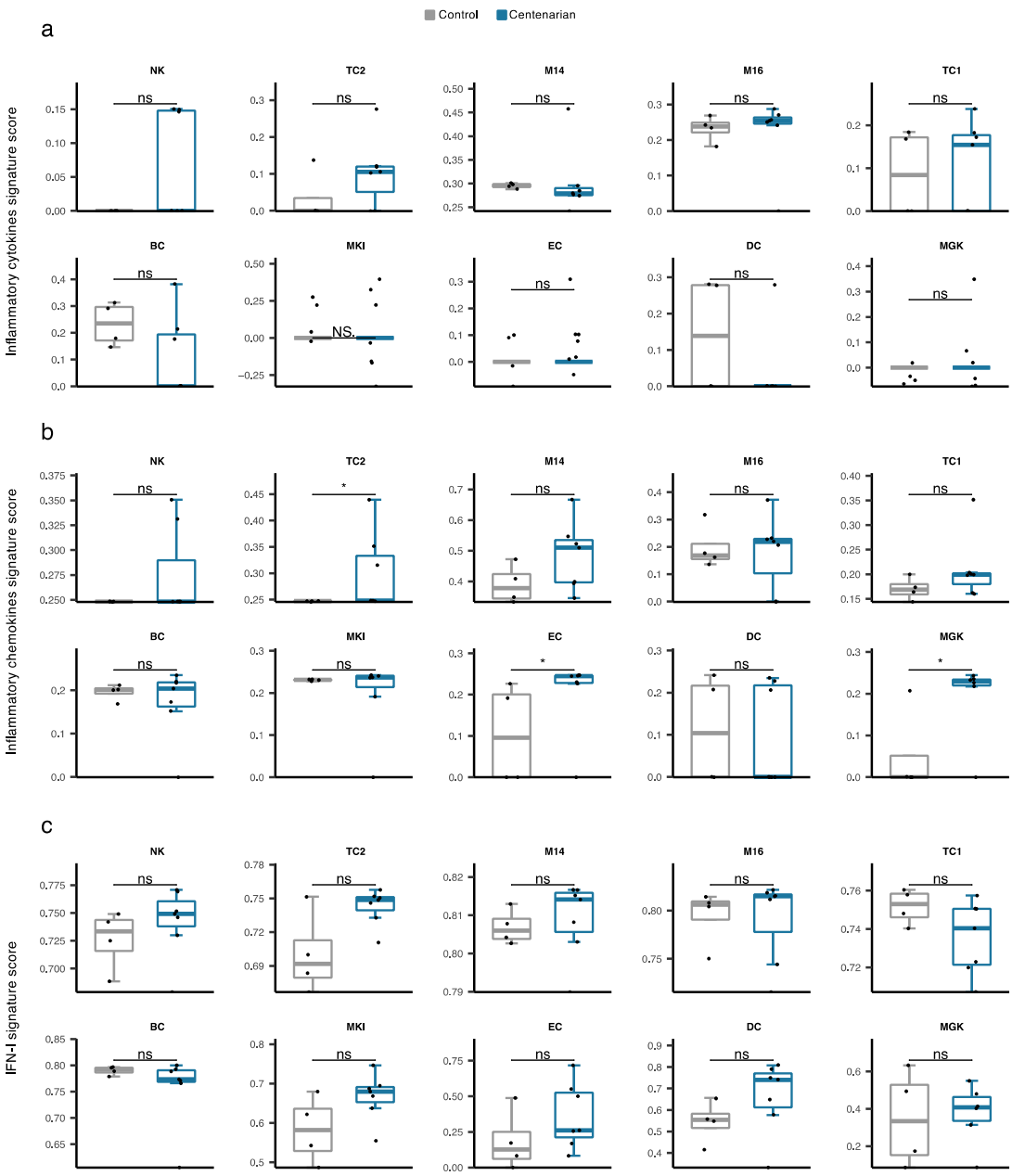

SASP score
Low
High
LINE expression (log RPM)
CD8 Tn
CD8 Tem−b
CD8 Tem−k
CD8 Tcm
TRDV2/TRGV9 gamma−delta T
CD4 Tem
CD4 Tn
CD4 Treg
SOX4 gamma−delta T
GZMA CD4 T
NK
CD56high NK
Classical monocytes
pDC
CD1C DC
CLEC9A DC
Naive B
ZBTB32 B
Memory B
Proliferating
Plasma B
CD34 HPC
TRAV1−2 MAIT
INF−activated CD4 T
CD16+ monocytes
ns
**


SASP score
Low
High
CD8 Tn
CD8 Tem−b
CD8 Tem−k
CD8 Tcm
TRDV2/TRGV9 gamma−delta T
CD4 Tem
CD4 Tn
CD4 Treg
SOX4 gamma−delta T
GZMA CD4 T
NK
CD56high NK
Classical monocytes
pDC
CD1C DC
CLEC9A DC
Naive B
ZBTB32 B
Memory B
Proliferating
Plasma B
CD34 HPC
TRAV1−2 MAIT
INF−activated CD4 T
CD16+ monocytes
LTR expression (log RPM)
ns
Low
High


Control
Centenarian
a
SINE expression log(RPM)
NK
TC2
M14
M16
TC1
BC
MKI
EC
DC
MGK
b
LINE expression log(RPM)
c
LTR expression log(RPM)
LTR
ns



744
745
746
747
748 Wilcoxon test.
749



significant, Wilcoxon test.

757 ## Supplementary Notes

758 ### Descriptions of the study populations included in the analysis

759 **Grady Trauma Project (GTP).** The research concentrated on African American individuals
760 of low socioeconomic backgrounds who had undergone trauma, primarily recruited from
761 Atlanta's Grady Memorial Hospital between 2005 and 2007 (Charles F. Gillespie et al., 2009).
762 Exclusion criteria comprised mental retardation, active psychosis, or inability to provide
763 informed consent. With over 7,000 participants, mostly urban African Americans with low
764 socioeconomic status, the GTP specifically targeted a demographic with a history of
765 significant trauma. The study involved whole blood collection between 8 to 9 a.m. following
766 overnight fasting. Whole genome expression profiles were generated for 398 subjects at the
767 Max-Planck Institute using Illumina HT-12 v3.0 or v4.0 arrays, focusing on 359 subjects of
768 self-reported African American ancestry aged between 16-78 years. The expression data is
769 available under GSE58137 at the GEO public repository. The Institutional Review Boards of
770 Emory University School of Medicine and Grady Memorial Hospital granted approval for all
771 procedures in this study. Funding primarily came from the National Institutes of Mental
772 Health (MH071537).

773 **Genetics, Arthrosis and Progression study (GARP).** The GARP study focused on
774 Dutch Caucasian siblings diagnosed with symptomatic osteoarthritis across multiple joint
775 sites. It comprised 191 sibling pairs (n=382 individuals) of white Dutch ancestry, aged
776 between 40 and 70, diagnosed with primary symptomatic osteoarthritis in multiple joints
777 including hand, spine, knee, or hip (N Riyazi et al., 2005). Whole genome expression profiles
778 were analyzed from 108 participants (68 unrelated families) of the GARP study and 26 age-
779 matched healthy controls. The blood of participants was collected, and mononuclear blood
780 cells were separated prior to RNA isolation. Sample libraries were hybridized onto the
781 microarrays (Illumina Human HT-12_v3_BeadChip's; Illumina). For the methylation analysis,
782 genomic DNA was extracted, and the DNA methylation was assayed at over 450,000 sites
783 on the Illumina Infinium HumanMethylation 450K BeadChips. The data is available at GEO
784 public repository under the accession GSE48556.

785 **Multi-Ethnic Study of Atherosclerosis (MESA).** MESA is a study of subclinical
786 cardiovascular disease and the factors that forecast its progression to clinically evident forms
787 initiated in 2000. This comprehensive investigation involves 6,814 asymptomatic individuals
788 aged 45-84, presenting a diverse demographic: 38% Caucasian, 28% African American, 22
789 percent Hispanic, and 12 percent Asian (primarily of Chinese descent). Enrolling 6,500
790 individuals evenly distributed across genders, the study targets those without clinical
791 cardiovascular disease at the outset, representing four racial/ethnic groups from six US
792 communities (Forsyth County, North Carolina; St. Paul, Minnesota; Chicago, Illinois; New

York, New York; Baltimore, Maryland; Los Angeles County, California). The study protocol has been approved by the institutional review boards of the six field centers: Wake Forest University, Columbia University, Johns Hopkins University, University of Minnesota, Northwestern University, and University of California - Los Angeles. Detailed information about the design of MESA can be found at https://internal.mesa-nhlbi.org, and the data is accessible under GSE56045 at GEO.

Blood samples were gathered to isolate peripheral blood mononuclear cells, from which monocytes were separated using anti-CD14 monoclonal antibody-coated magnetic beads. Flow cytometry analysis consistently demonstrated monocyte samples of over 90% purity across 18 specimens. The DNA and RNA were extracted simultaneously. RNA with integrity scores above 9.0 was chosen for global expression microarrays. Genome-wide expression analysis employed the Illumina HumanHT-12 v4 Expression BeadChip and Illumina Bead Array Reader, following the Illumina expression protocol. Additionally, the Illumina HumanMethylation450 BeadChip and HiScan reader were utilized for epigenome-wide methylation analysis, assigning individual samples to the BeadChips and chip positions using the same sampling scheme as for the expression BeadChips. Details for sample preparation, data processing and quality control were processed as demonstrated in the design paper (Diane Bild et al., 2002). The expression and methylation data are available at GEO public repository under the accession GSE56045 and GSE56046, respectively.

Swedish Adoption/Twin Study of Aging (SATSA). SATSA is a comprehensive exploration of aging variations using twins raised in different environments influenced by genetic and environmental factors. The study draws its data from the Swedish Twin Registry, a population-based national register encompassing twins born from 1886 to 2000. Additionally, the comprehensive assessment across biological, psychological, and social domains allows for the examination of patterns, predicting the onset of age-related health issues. Started in 1984, the study employed rigorous questionnaires every three years until 2010, engaging over 2,000 twins. In parallel, a subset of 861 individuals participated in eight in-person testing (IPT) sessions, encompassing health, cognition, and memory assessments. Notably, IPT9 and IPT10 introduced daily tracking of memory, emotions, and social interactions, providing valuable insights into short-term fluctuations and early indicators of declining health.

A total of 1,122 blood samples were collected at five time points spanning from 1992 to 2012. Across these five longitudinal waves, participant counts with 1 to 5 measurements were 99, 86, 90, 80, and 30, respectively. Phenotype data, encompassing chronological age, sex, and zygosity, were gathered through comprehensive questionnaires and physical testing at each sampling wave. Following the manufacturer's protocol optimized for Illumina's Infinium 450K assay, 200 ng of DNA was prepared for each sample. The bisulfite-converted DNA samples

830 were then hybridized to the Infinium HumanMethylation450 BeadChips using Illumina's
831 Infinium HD protocol. This process allowed the measurement of DNA methylation levels for
832 485,512 CpGs in each sample. Following the quality control procedures described by Peters
833 et al., 1,011 samples from 385 twins were retained after processing and quality control on
834 the methylation data (Marjolein Peters et al., 2015). The data is available at ArrayExpress
835 public repository under the accession E-MTAB-7309.

836 **Brisbane Systems Genetics Study (BSGS).** Participants BSGS were enrolled in the
837 Brisbane Twin Nevus (BTN) and MAPS (Nevus and cognition studies) projects. The study
838 was approved by the Queensland Institute for Medical Research-Human Research Ethics
839 Committee. Over the span of 16 years, adolescent MZ and DZ twins, alongside their siblings
840 and parents, were recruited for an ongoing investigation into genetic and environmental
841 influences on pigmented nevi and the associated risk of skin cancer and cognition. The
842 study comprises individuals of mainly Anglo-Celtic descent with northern European origins
843 (Joseph Powell et al., 2012).

844 Whole blood samples were collected for total RNA extraction, following the sample
845 preparation process detailed by Powell et al. Gene expression profiles were generated by
846 hybridizing 750 ng of cDNA to the Illumina HumanHT-12 v4.0. This process yielded gene
847 expression data for 962 individuals from 314 families. In parallel, DNA methylation analysis
848 was conducted on 614 individuals from 117 European descent families recruited through
849 BSGS5. Bisulfite converted DNA samples were hybridized to 12-sample Illumina
850 HumanMethylation450 BeadChips using the Infinium HD Methylation protocol and Tecan
851 robotics, assessing methylation status across 485,577 CpG sites, covering 99% of RefSeq
852 genes. Post-methylation probe filtration by Powell et al. resulted in 417,069 probes. Both
853 expression and methylation data are accessible in the GEO public repository under the
854 accessions GSE33321 and GSE56105.

855 **Genome-wide Methylation Profiles Reveal Quantitative Views of Human Aging Rates**
856 **(GMPWAR).** This study aims to construct an aging model that gauges one's methylome age
857 acceleration. Approved by the University of California, San Diego, the University of Southern
858 California, and West China Hospital, the research involved 426 Caucasian and 230 Hispanic
859 individuals, spanning ages 19 to 101, providing extensive methylome-wide profiles. Whole
860 blood samples were collected, and genomic DNA underwent analysis using the Illumina
861 Infinium HumanMethylation450 BeadChip assay, assessing methylation states across
862 485,577 CpG markers. Stringent quality controls by Hannum et al. ensured the exclusion of
863 unreliable markers and samples (Gregory Hannum et al., 2013). Access to the complete
864 methylation profiles is available at the GEO repository (GSE40279).

865 **ScRNA-seq of PMBCs of healthy aging populations.** Approval for all human studies
866 was granted by the Washington University in St. Louis School of Medicine Institutional

867 Review Board (IRB- 201804084). The study enrolled healthy Caucasian individuals, both
868 males and females, who were non-obese with a BMI below 30, between 2018 and 2019. The
869 research involved specific age and BMI criteria: young males aged 25-29 years (BMI 21.2-
870 28.9 kg/m2, n = 14), and older non-frail males aged 62-70 years (BMI 17.8-29.8 kg/m2, n =
871 14). Exclusion criteria comprised individuals with a history of cancer, inflammatory
872 conditions, bloodborne diseases, smokers, or recent illness (Denis Mogilenko et al., 2021).
873 Blood samples (~100 ml) were obtained via venous puncture in the morning (7–10 a.m.)
874 after an overnight fast. Single-cell suspensions were isolated and underwent droplet-based
875 massively parallel single-cell RNA sequencing using the Chromium Single Cell 5 Reagent Kit
876 according to the manufacturer's instructions (10x Genomics). Processing involved sample
877 demultiplexing, barcode handling, and cell counting using the Cell Ranger Single-Cell
878 Software Suite (10x Genomics). Subsequent analysis utilized the Seurat R package (Butler
879 et al., 2018), which identified 25 clusters representing primary immune cell populations.
880 Notably, one sample (D18) was excluded from analysis due to its outlier status, identified by
881 distinct clustering in the single-cell RNA sequencing analysis. Raw and processed scATAC-
882 seq data are available at https://www.synapse.org repository (syn22255433)
883 **ScRNA-seq of PBMCs of supercentenarians.** All experiments using human samples in
884 this study were approved by the Keio University School of Medicine Ethics Committee
885 (approval no. 20021020) and the ethical review committee of RIKEN (approval no. H28-6).
886 PBMCs derived from 7 supercentenarians (SC1–SC7) and 5 controls (CT1–CT5, aged in
887 their 50s to 80s) were profiled using droplet-based single-cell RNA sequencing technology
888 (10× Genomics). Fresh whole blood from supercentenarians, their offspring residing with
889 them, and unrelated donors was collected, and PBMCs were isolated from whole blood.
890 Single-cell libraries were prepared from freshly isolated PBMCs by using Chromium Single
891 Cell 3' v2 Reagent Kits. The analysis pipelines in Cell Ranger version 2.1.0 and Cell Ranger
892 R Kit (version 2.0.0) were used for sequencing data processing (Kosuke Hashimoto et al.,
893 2019). Raw UMI counts and normalized expression values for scRNA-Seq are publicly
894 available at http://gerg.gsc.riken.jp/SC2018/.
895
896
897
898
899