Ageing affects DNA methylation drift and transcriptional cell-to-cell variability in muscle stem cells

Authors: Irene Hernando-Herraez[1*], Brendan Evano[2,3,4*], Thomas Stubbs[1*], Pierre-Henri Commere[5], Stephen Clark[1], Simon Andrews[1], Shahragim Tajbakhsh[2,3 #], Wolf Reik[1 #]

Affiliations:

[1] Epigenetics Programme, Babraham Institute, Cambridge CB22 3AT, United Kingdom

[2] Stem Cells & Development, Department of Developmental & Stem Cell Biology, Institut Pasteur, 25 rue du Dr. Roux, 75015, Paris, France.

[3] CNRS UMR 3738, Institut Pasteur, Paris 75015, France.

[4] Current address: CNRS UMR 3664, Nuclear Dynamics, Institut Curie, Pavillon Pasteur, 26 rue d'Ulm 75005 Paris, France.

[5] Cytometry and Biomarkers, Center for Technological Resources and Research, Institut Pasteur, 28 rue du Dr. Roux, 75015, Paris, France.

[*] Equal contributions.

[#] Equal contributions.

Correspondence to Irene.Herraez@babraham.ac.uk, shahragim.tajbakhsh@pasteur.fr and wolf.reik@babraham.ac.uk

Abstract:

Age-related tissue alterations have been associated with a decline in stem cell number and function[1]. Although increased cell-to-cell variability in transcription or epigenetic marks has been proposed to be a major hallmark of ageing[2–5], little is known about the molecular diversity of stem cells during ageing. Here, by combined single-cell transcriptome and DNA methylome profiling in mouse muscle stem cells, we show a striking global increase of uncoordinated transcriptional heterogeneity together with context-dependent alterations of DNA methylation with age. Importantly, promoters with increased methylation heterogeneity are associated with increased transcriptional heterogeneity of the genes they drive. Notably, old cells that change the most with age reveal alterations in the transcription of genes regulating cell-niche interactions. These results indicate that epigenetic drift, by accumulation of stochastic DNA methylation changes in promoters, is a substantial driver of the degradation of coherent transcriptional networks with consequent stem cell functional decline during ageing.

Epigenetic alterations have been proposed to be a major cause of age-related decline in tissue function[6]. Changes in DNA methylation are well correlated with ageing and methylation of specific loci has been used as age biomarker in a large number of tissues[6,7]. However, age-related methylation changes are poorly correlated with transcriptional variation, presumably because the changes are generally small and may not occur homogeneously in all cells[7], a phenomenon also known as epigenetic drift. Although epigenetic drift has long been hypothesised to be an important hallmark of ageing[8], this proposal has been challenging to test because of technical constraints. However, powerful combined single cell methods[9,10] are now available, and epigenetic changes during ageing together with their functional consequences can now be read out in single cells[11].

Degenerative changes in tissue-specific stem cells have been proposed to be a major cause of age-related decline in tissue function[12]. While several reports indicate a loss of clonal diversity during early life stages[13–15] little is known about how cell-to-cell variability at the molecular level is involved in stem cell ageing. Here, we performed parallel single-cell DNA methylation and transcriptome sequencing (scM&T-seq) on the same cell[10] to investigate how ageing affects transcriptional and epigenetic heterogeneity of tissue-specific stem cells, using mouse muscle stem cells as a model. Muscle satellite (stem) cells express the transcription factor *Pax7*[16] and are largely quiescent in adult muscles. They activate upon injury to differentiate and fuse to form new fibers, or self-renew to reconstitute the stem cell pool[16]. Age-associated muscle defects have been attributed to a decrease in stem cell number together with impaired regenerative potential[17]. In addition, clonal lineage-tracing of mouse satellite cells showed that population diversity is unaltered during homeostatic ageing[18].

Satellite cells with high expression of *Pax7* were shown to be in a deep quiescent state[19,20]. To investigate the molecular effects of ageing in a defined population that is less poised to enter the cell cycle, we isolated single satellite cells by fluorescence-activated cell sorting

(FACS) from young (2 months) and old (24 months) *Tg:Pax7-nGFP* mice[21] and selected those with high levels of GFP, to which we applied scM&T-seq (Fig. 1A).

After quality control and filtering, a total of 377 transcriptomes were analysed. Young and old cells from different individuals clustered together, respectively, indicating no global differences with age and absence of sequencing-related batch effects (Fig. 1B). Furthermore, we did not observe significant differences in the levels of *Pax7*, the myogenic factors *Myod* and *Myf5* and the cell cycle inhibitor *Cdkn1b,* nor of senescent markers such as *Cdkn2a*, suggesting that some molecular signatures are conserved between the analysed cell populations (Fig. 1C). Nevertheless, 940 genes were differentially expressed between young and old individuals (SCDE, FDR $P < 0.05$, Table S1). *Spry1*, which is a key factor for maintaining quiescence[22], and the cell cycle regulators *Ccnd1*, *Btg1* and *Gas1* were down-regulated, while ageing markers such as the chemokine genes *Ccl11* and *Ccl19* were up-regulated[20] (Fig. 1C). Furthermore, we uncovered genes not previously reported to change in expression with age, such as the early activation markers *Fosb* and *Egr1*[23] and the metalloproteinase *Mmp2* (Fig. 1C).

To investigate if ageing affects transcriptional heterogeneity of the stem cell pool, we calculated pairwise correlation coefficients between cells within each individual (see Methods) and observed that old individuals showed consistently lower correlation (1.3 mean-fold decrease, Mann-Whitney-Wilcoxon test; $P < 2.2e\text{-}16$, Fig. 1D), indicating a remarkably lower degree of similarity between cells and no obvious population substructure. We also computed an expression-level normalised measure of gene expression heterogeneity (named distance to the median)[24], which proved to be higher in old individuals (Mann-Whitney-Wilcoxon test; $P < 2.2e\text{-}16$, Fig. 1E) revealing a striking global increase of uncoordinated transcriptional variability with age. Strikingly, the proportion of cells expressing a given gene (frequency of gene expression) was reduced with age (Mann-Whitney-Wilcoxon test; $P < 2.2e\text{-}16$, Fig. 1F),

even in genes that did not significantly change mean expression levels (SCDE, FDR $P > 0.05$, Fig. 1F). Importantly, we observed that this was independent of gene expression levels and not restricted to lowly expressed genes suggesting that this global feature is unrelated to technical effects (Fig. 1G).

Genes that displayed increased expression variability with age (expression frequency difference > 15%) include several collagen genes (*Col4a2*, *Col5a3*, *Col4a1*) and other extracellular matrix-related genes such as *Dag1*, *Sparc*, *Cdh15* or *Itgb1* (Fig. 2A). Interestingly, satellite cells without *Itgb1* (β1-integrin) cannot maintain quiescence and its experimental activation improves ageing-related decline in muscle regeneration[25]. Similarly, reduction of N-cadherin and M-cadherin (*Cdh15*) leads to a break of quiescence of satellite cells[26]. Notably, none of the above-mentioned genes were shown to change in expression level during the isolation procedure of satellite cells[27].

The observed increase in transcriptional variability with age could reflect the presence of cell subpopulations or be a purely stochastic process. Despite not observing clear substructure (Fig. 1B and Fig. 1D right), we further investigated the origin of this variability by ranking old cells based on their transcriptome-wide similarity to young cells, and performed correlation analyses to identify the genes driving this ranking. Gene ontology analysis indicated that old cells that differed the most from young cells were enriched in processes such as translation and peptide biosynthesis (Fig. 2B top), while old cells that were most similar to young ones were enriched in extracellular matrix-related functions (Fig. 2B bottom). For example, *Fos* and *Mmp2* were preferentially expressed in the most different old cells, while extracellular markers such as *Dag1*, *Itgb1*, *Cdh15* or *Bgn* were expressed in the most similar ones (Fig. 2C). These results indicate that cells that have accumulated more differences with age are likely to have impaired cell-niche interactions and are more prone to exit quiescence (Fig. 2D).

For the analysis of DNA methylation patterns, we limited potential biases due to uneven sequencing depth between cells or different number of cells per individual by randomly subsampling 1 million reads from each cell and 35 cells per individual (140 cells in total, 2 million CpG sites on average per cell). Global mean DNA methylation levels were around 50%, as previously reported for muscle stem cells[28] (Fig. S1C). As with the transcriptomes, we did not observe clear subpopulations in any of the methylome samples (Fig. S2). Overall, CpG islands, promoters and enhancers were hypomethylated; exons, myoblast enhancers (marked by H3K27ac) and shores (flanking region of the CpG islands) were around 30% methylated, while repeats and bodies of active genes (marked by H3K36me3) were highly methylated (Fig. 3A). We found that DNA methylation levels increased slightly with age, as reported for human muscle stem cells[22], mostly in repeat elements and H3K36me3 regions (Fig. 3B, 3C and 3D).

Identical average methylation levels for a given genomic region may reflect different scenarios, from uniform populations to completely random heterogeneous patterns (Fig. 3E). Since we did not observe substructure in our data (Fig. S2) and as stochastic epigenetic drift has been suggested to be a major hallmark of ageing[8], we computed a score to measure levels of stochastic intrapopulation heterogeneity (Fig. S3, Methods). As expected, our initial measure of heterogeneity depended on average methylation levels (Fig. 3F). Hence, we developed an independent measure of heterogeneity by calculating the distance between the observed heterogeneity for each genomic region and a rolling median (Fig. 3F, Methods). Interestingly, this analysis showed that different genomic contexts displayed different levels of methylation heterogeneity between cells, for example CpG islands were more heterogeneous than enhancers (Fig. 3G).

Global levels of methylation heterogeneity were similar between ages (Fig. S4); we next computed localised Z-score comparisons between young and old to examine changes in specific genomic elements. Notably, methylation of LINE-1 elements became more

131 homogeneous with age whereas regions marked by H3K27me3 became more heterogeneous
132 (Fig. 4A). Specifically, LINE-1 elements also experienced the highest increase in absolute
133 DNA methylation levels, both of which may reflect a coordinated mechanism to prevent
134 deleterious somatic retrotranspositions during ageing. Most of the H3K27me3 regions were
135 associated with genes that are repressed but poised for rapid activation[29]. We hypothesize that
136 this increase in methylation heterogeneity may contribute to an impaired transcriptional
137 response upon activation.

138 Interestingly, we observed a negative correlation between changes in methylation levels
139 and changes in methylation heterogeneity (Promoters: Pearson's coefficient= -0.35, $P < 2.2e-16$,
140 Fig. 4B). Regions becoming more homogeneous showed an increase in methylation,
141 suggesting that *de novo* methylation enzymes (Dnmt3a,b) are recruited to specific sites and add
142 methylation in a coordinated manner between cells. In contrast, regions becoming more
143 heterogeneous showed a decrease in their methylation levels. Despite the low proliferative
144 history of these cells, this pattern could reflect errors in DNA methylation maintenance during
145 DNA replication, or an active demethylation mechanism via TET enzymes (Fig. S5).

146 Epigenetic changes may contribute to the age-associated pattern of transcriptional
147 heterogeneity. To explore this possibility, we analysed the association between promoter DNA
148 methylation and gene expression. We calculated a correlation coefficient for each cell and
149 confirmed the expected negative correlation for methylation and transcription (Fig. 4C).
150 Interestingly, old cells that were most transcriptionally different from young cells showed
151 lower levels of correlation (Mann-Whitney-Wilcoxon test; $P < 0.05$, Fig. 4C). Furthermore, we
152 calculated changes in transcriptional variability between young and old cells (see Methods)
153 and observed that promoters with increased methylation heterogeneity tended to have increased
154 transcriptional heterogeneity (Mann-Whitney-Wilcoxon test; $P < 0.001$) (Fig. 4D). It appears
155 therefore that deterioration of transcriptional coherence during ageing is associated with

156 increased promoter methylation heterogeneity and with decreased connectivity between the
157 epigenome and the transcriptome.

158 In summary, we report transcriptional and epigenetic signatures associated with ageing
159 in a deeply quiescent population of muscle stem cells. Previous studies have investigated
160 transcriptional heterogeneity changes with age in mixed cell populations[4] which are affected
161 by differences in cellular composition, such as an increase in senescent cells[4]. In contrast, our
162 study is focused on a specific population of cells in which known stemness, activation and
163 senescent markers were not affected by ageing. Even in this restricted population, we observe
164 a global increase of uncoordinated transcriptional variability with age, indicating an intrinsic
165 mechanism of cellular ageing. Interestingly, mouse muscle stem cells were shown to maintain
166 clonal diversity during homeostatic ageing by lineage-tracing[18], however, our study uncovers
167 a dramatic underlying molecular heterogeneity in these stem cells that extends beyond
168 maintenance of clonal homogeneity. We also observe that cells that have acquired more
169 differences with age showed alterations in multiple extracellular matrix related genes
170 potentially affecting cell-niche interactions.

171 Elevated transcriptional variability with age has been reported in several studies[2–4],
172 however the underlying causes remain largely unknown. The accumulation of somatic
173 mutations only partially accounts for the increased cell-to-cell transcriptional variability[4],
174 suggesting that epigenetic mechanisms might be a contributing factor[5]. In this study, by
175 applying for the first time a combined single cell method for DNA methylation and the
176 transcriptome, we show that epigenetic drift, or the uncoordinated accumulation of methylation
177 changes in promoters, contributes to the increased transcriptional variability with age (Fig. 4E).
178 Due to the deep quiescent state of the homeostatic cells chosen for study, our data highlight the
179 possibility that the observed epigenetic patterns could be independent of extensive cell
180 proliferation. We propose that this variability is detrimental due to uncoordinated transcription,

181 thereby affecting the ability of stem cells to maintain quiescence or activate coherently upon
182 injury. Future studies of different stem cell populations integrating multiple layers of molecular
183 information will be highly informative for a more complete understanding of the underlying
184 molecular mechanisms of ageing and age-related diseases.

Methods

Mice

Animals were handled according to national and European Community guidelines, and an ethics committee of the Institut Pasteur (CETEA) in France approved protocols. Young (2 months-old) and old (24 months-old) *Tg:Pax7-nGFP*[21] mice were used in this study.

Isolation of satellite cells

Mice were sacrificed by cervical dislocation. *Tibialis anterior* muscles were dissected and placed into cold DMEM (ThermoFisher, 31966). Muscles were then chopped and put into a 15 ml Falcon tube containing 10 ml of DMEM, 0.08% collagenase D (Sigma, 11 088 882 001), 0.1% trypsin (ThermoFisher, 15090), 10 μg/ml DNaseI (Sigma, 11284932) at 37°C under gentle agitation for 25 min. Digests were allowed to stand for 5 min at room temperature and the supernatants were collected on 5 ml of foetal bovine serum (FBS; Gibco) on ice. The digestion was repeated 3 times until complete digestion of the muscle. The supernatants were filtered through a 70-μm cell strainer (Miltenyi, 130-098-462). Cells were spun for 15 min at 515g at 4°C and the pellets were resuspended in 1 ml freezing medium (10% DMSO (Sigma, D2438) in foetal calf serum (FCS, Invitrogen)) for long term storage in liquid nitrogen.

Before isolation by FACS, samples were thawed in 50 ml of cold DMEM, spun for 15 min at 515g at 4°C. Pellets were resuspended in 300 μl of DMEM 2% FCS 1 μg/mL propidium iodide (Calbiochem, 537060) and filtered through a 40-μm cell strainer (BD Falcon, 352235). Viable cells were isolated based on size, granulosity and GFP expression level (top 10% $nGFP^{Hi}$ cells, Fig. S6) using a MoFlo Astrios cell sorter (Beckmann Coulter).

Single cells were collected in 2.5 μL cold RLT Plus buffer (Qiagen, 1053393) containing 1U/μL RNAse inhibitor (Ambion, AM2694) in 96 well-plates (LoBind Eppendorf, 0030129504), flash-frozen on dry ice and stored at -80°C.

210

211 Library preparation and data alignment

212 We prepared scM&T-seq libraries[10] by isolating mRNA on magnetic beads and separating from
213 the single-cell lysate as described[30] prior to reverse transcription and amplification using
214 Smartseq2[31] but with 25 PCR cycles. We then processed the lysate containing genomic DNA
215 according to the published single-cell bisulfite sequencing protocol[32]. Single-cell RNA-seq
216 libraries were aligned using HiSat2 with options --sp 1000,1000 --no-mixed --no-discordant[33].
217 Single-cell bisulfite libraries were processed using Bismark[34] as described[10]. Mapped RNA-seq
218 data were quantitated using the RNA-seq quantitation pipeline in Seqmonk software
219 (www.bioinformatics.babraham.ac.uk/projects/seqmonk/).

220

221 Quality control RNA-seq

222 Cells expressing fewer than 1,000 genes or less than 10^5 mapped reads allocated to nuclear
223 genes were removed in quality control (Fig. S7). These cells were also verified to have less
224 than 10% of mapped on mitochondrial genes. Out of the 768 cells that were captured across
225 the experiment, 377 passed our quality and filtering criteria (Table S2).

226

227 Data analysis RNA-seq

228 Gene expression levels were estimated in terms of reads per million of mapped reads to the
229 transcriptome. A score of variability per gene (named distance to the median) was calculated
230 by fitting the squared coefficient of variation as a function of the mean normalized counts and
231 then calculating the distance to a rolling average (window size=100) (Fig. S8)[24]. We included
232 only genes with an average normalized read count of at least 10. The top 1000 most variable
233 genes of the entire data set were used to perform principal component analyses (as $\log_2$-
234 transformed and median-cantered values) (Fig. 1B, Table S3). Single cell differential

235 expression (SCDE) was used to calculate differential expression analysis between young and
236 old cells (Table S1)[35].

237 Cell-to-cell correlation analyses were performed using the top 500 most variable genes
238 within each individual and using Spearman's correlation as the measure of similarity between
239 cells (Fig. 1D). Distance to the median of the top 500 most variable genes within each
240 individual was computed for Fig. 1E, similar results are observed when restricting the analysis
241 to genes that are expressed in all the individuals (average normalized read count of at least 10)
242 and different numbers of genes (Fig. S9).

243 An average young reference transcriptome was computed by calculating the mean of
244 log transformed expression values for each gene across cells from young individuals. We then
245 performed Spearman's correlation analyses to assess the similarity between each cell from old
246 samples and the young transcriptome. Spearman's correlation analyses were then also used to
247 find gene expression patterns associated with this genome-wide similarity score. Genes
248 expressed in fewer than five cells were excluded from the analysis. The top 200 correlated and
249 anticorrelated genes (Table S4) were used for GO enrichment analysis[36] .

250

251 DNA-methylome

252 We discarded cells that had less than 1 million paired-end alignments or less than 500,000 CpG
253 sites covered (Fig. S1). To avoid biases that might occur due to different sequencing depths or
254 number of cells between individuals, we down-sampled the data to 1 million reads for each cell
255 and randomly selected 35 cells from each individual (2 young and 2 old). ChIP-seq datasets
256 for H3K4me3, H3K27me3, H3K36me3 in satellite cells and H3K27ac in myoblast were
257 obtained from existing studies[23,29]. Bowtie2 and MACS2 were used for mapping and peak
258 calling respectively.

259

DNA methylation heterogeneity

We developed a heterogeneity score based on Hamming distances and Shannon entropy between cell pairs from the same sample. This value captures the properties we desire: i) ability to detect cell-to-cell stochastic heterogeneity ii) not affected by population substructure iii) not biased by missing values. Precisely, let *r* be a matrix with methylation values of cells for a particular gene, each row corresponding to a cell and each column corresponding to a CpG site, and *w* be the weight corresponding to the number of covered CpGs within each pairs of cells. For each pair of cells (*c*), we then computed the Hamming distance (*D*) and the Shannon entropy score of the pairs (*S*) considering sites with coverage in both cells. Then weighted heterogeneity score of the regions is:

$$H(r) = \frac{\sum_{c=1}^{n} w_c \times D_c}{\sum_{c=1}^{n} w_c} \times \frac{\sum_{c=1}^{n} w_c \times S_c}{\sum_{c=1}^{n} w_c}$$

Here D_c is the normalised Hamming distance of a given a pair of cells, which measures the number of bits that are different in two binary sets:

$$D_c = \sum_{i=1}^{k} |x_i - x_j|$$

S_c is the joint Shannon Entropy between a pair of cells which measures the complexity of the pattern:

$$S_c = -\sum_{i=1}^{k} p_i \cdot \log_2(p_i)$$

Here *p* is the frequency of pairs of methylation values.

We validated our approach by applying the method in simulated data with increasing levels of methylation heterogeneity (Fig. S3). We also observed that our algorithm is highly robust to missing data (Fig. S3).

We applied this method across multiple genomic regions for each individual

independently and then computed the average of young and old samples. Pairwise comparisons with fewer than 4 CpG sites were not considered in the analysis. Furthermore, to avoid misinterpretations because of poor coverage depth we excluded regions with: i) less than 20CpG sites, ii) less than an average of 2 CpG sites covered per cell, iii) less than 100 cell-to-cell pairwise comparisons. We also excluded regions with high coverage differences between ages (more than an average of 10 CpG sites or more than 200 cell-to-cell pairwise comparisons). A total of 63,823 genomic regions were used in the analysis (average window size= 2,267 bp).

Coverage-weighted cell methylation values were used to calculate the mean methylation levels of each region. A normalised measure of DNA methylation heterogeneity was calculated for each region (from young or old samples) by fitting the score of heterogeneity as a function of the mean methylation levels and then calculating the distance to a rolling median of 1,000 observations (Fig. 3F). Regions with less than 0.05 or more than 0.9 mean methylation levels were excluded from the analysis.

Differences between young and old DNA methylation heterogeneity values were Z-score normalised using a sliding window of 100 observations ordered by the mean value of young and old (Fig. S10 and Table S5). Same approach was used to calculate differences between young and old transcriptional heterogeneity (mean distance to the median) (Fig S9 and Table S5).

Data availability

Sequencing data have been deposited in GEO with the accession: GSE121364

Software

Custom software is available upon request.

References

1. López-Otín, C., Blasco, M. A., Partridge, L., Serrano, M. & Kroemer, G. The Hallmarks of Aging. *Cell* **153,** 1194–1217 (2013).
2. Bahar, R. *et al*. Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441,** 1011–1014 (2006).
3. Martinez-Jimenez, C. P. *et al*. Aging increases cell-to-cell transcriptional variability upon immune stimulation. *Science*. **355,** 1433–1436 (2017).
4. Enge, M. *et al*. Single-cell analysis of human pancreas reveals transcriptional signatures of aging and somatic mutation patterns. *Cell* **171,** 321–330 (2017).
5. Cheung, P. *et al*. Single-cell chromatin modification profiling reveals increased epigenetic variations with aging. *Cell* **173,** 1385–1397 (2018).
6. Benayoun, B. A., Pollina, E. A. & Brunet, A. Epigenetic regulation of ageing: Linking environmental inputs to genomic stability. *Nat. Rev. Mol. Cell Biol*. **16,** 593–610 (2015).
7. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet*. **19,** 371–384 (2018).
8. Veitia, R. A., Govindaraju, D. R., Bottani, S. & Birchler, J. A. Aging: somatic mutations, epigenetic drift and gene dosage imbalance. *Trends Cell Biol*. **27,** 299–310 (2017).
9. Gravina, S., Dong, X., Yu, B. & Vijg, J. Single-cell genome-wide bisulfite sequencing uncovers extensive heterogeneity in the mouse liver methylome. *Genome Biol*. **17,** 150–158 (2016).
10. Angermueller, C. *et al*. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. *Nat. Methods* **13,** 229–232 (2016).

11. Kelsey, G., Stegle, O. & Reik, W. Single-cell epigenomics: Recording the past and predicting the future. *Science*. **358,** 69–75 (2017).

12. Oh, J., Lee, Y. D. & Wagers, A. J. Stem cell aging: mechanisms, regulators and therapeutic opportunities. *Nat. Med*. **20,** 870–880 (2014).

13. Snippert, H. J. *et al*. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. *Cell* **143,** 134–144 (2010).

14. Nguyen, P. D. *et al*. Muscle stem cells undergo extensive clonal drift during tissue growth via Meox1-mediated induction of G2 cell-cycle arrest. *Cell Stem Cell* **21,** 107–119 (2017).

15. Klein, A. M., Nakagawa, T., Ichikawa, R., Yoshida, S. & Simons, B. D. Mouse germ line stem cells undergo rapid and stochastic turnover. *Cell Stem Cell* **7,** 214–224 (2010).

16. Brack, A. S. & Rando, T. A. Tissue-specific stem cells: lessons from the skeletal muscle satellite cell. *Cell Stem Cell* **10,** 504–514 (2012).

17. Brack, A. S. & Muñoz-Cánoves, P. The ins and outs of muscle stem cell aging. *Skelet. Muscle* **6,** 1–9 (2016).

18. Tierney, M. T., Stec, M. J., Rulands, S., Simons, B. D. & Sacco, A. Muscle stem cells exhibit distinct clonal dynamics in response to tissue repair and homeostatic aging. *Cell Stem Cell* **22,** 119–127 (2018).

19. Rocheteau, P., Gayraud-Morel, B., Siegl-Cachedenier, I., Blasco, M. A. & Tajbakhsh, S. A subpopulation of adult skeletal muscle stem cells retains all template DNA strands after cell division. *Cell* **148,** 112–125 (2012).

20. Chakkalakal, J. V, Jones, K. M., Basson, M. A. & Brack, A. S. The aged niche disrupts muscle stem cell quiescence. *Nature* **490,** 355–360 (2012).

21. Sambasivan, R. *et al*. Distinct regulatory cascades govern extraocular and pharyngeal

arch muscle progenitor cell fates. *Dev. Cell* **16,** 810–821 (2009).

22. Bigot, A. *et al*. Age-associated methylation suppresses SPRY1, leading to a failure of re-quiescence and loss of the reserve stem cell pool in elderly muscle. *Cell Rep*. **13,** 1172–1182 (2015).

23. Machado, L. *et al*. In situ fixation redefines quiescence and early activation of skeletal muscle stem cells. *Cell Rep*. **21,** 1982–1993 (2017).

24. Brennecke, P. *et al*. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10,** 1093–1095 (2013).

25. Rozo, M., Li, L. & Fan, C. M. Targeting β1-integrin signaling enhances regeneration in aged and dystrophic muscle in mice. *Nat. Med*. **22,** 889–896 (2016).

26. Goel, A. J., Rieder, M. K., Arnold, H. H., Radice, G. L. & Krauss, R. S. Niche cadherins control the quiescence-to-activation transition in muscle stem cells. *Cell Rep*. **21,** 2236–2250 (2017).

27. Brink, S. C. van den *et al*. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nat. Methods* **14,** 935–936 (2017).

28. Davegårdh, C. *et al*. Abnormal epigenetic changes during differentiation of human skeletal muscle stem cells from obese subjects. *BMC Med*. **15,** 39–66 (2017).

29. Liu, L. *et al*. Chromatin modifications as determinants of muscle stem cell quiescence and chronological aging. *Cell Rep*. **4,** 189–204 (2013).

30. Macaulay, I. C. *et al*. G&T-seq: Parallel sequencing of single-cell genomes and transcriptomes. *Nat. Methods* **12,** 519–522 (2015).

31. Picelli, S. *et al*. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc*. **9,** 171–181 (2014).

32. Clark, S. J. *et al*. Genome-wide base-resolution mapping of DNA methylation in single cells using single-cell bisulfite sequencing (scBS-seq). *Nat. Protoc*. **12,** 534–547

(2017).

33. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **12,** 357–360 (2015).

34. Krueger, F. & Andrews, S. R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27,** 1571–1572 (2011).

35. Kharchenko, P. V., Silberstein, L. & Scadden, D. T. Bayesian approach to single-cell differential expression analysis. *Nat. Methods* **11,** 740–742 (2014).

36. Eden, E., Navon, R., Steinfeld, I., Lipson, D. & Yakhini, Z. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* **10,** 48–55 (2009).

Acknowledgments: We would like to thank the Flow Cytometry Platform of the Center for Technological Resources and Research (Institut Pasteur) and the Wellcome Trust Sanger Institute sequencing facility for assistance with Illumina sequencing. This project was supported by grants from Institut Pasteur, Agence Nationale de la Recherche (Laboratoire d'Excellence Revive, Investissement d'Avenir; ANR-10-LABX-73), Association Française contre les Myopathies (21857), CNRS, the European Research Council (Advanced Research Grant 332893), and the Biotechnology and Biological Sciences Research Council (BBSRC, CBBS/E/B/000C0425).

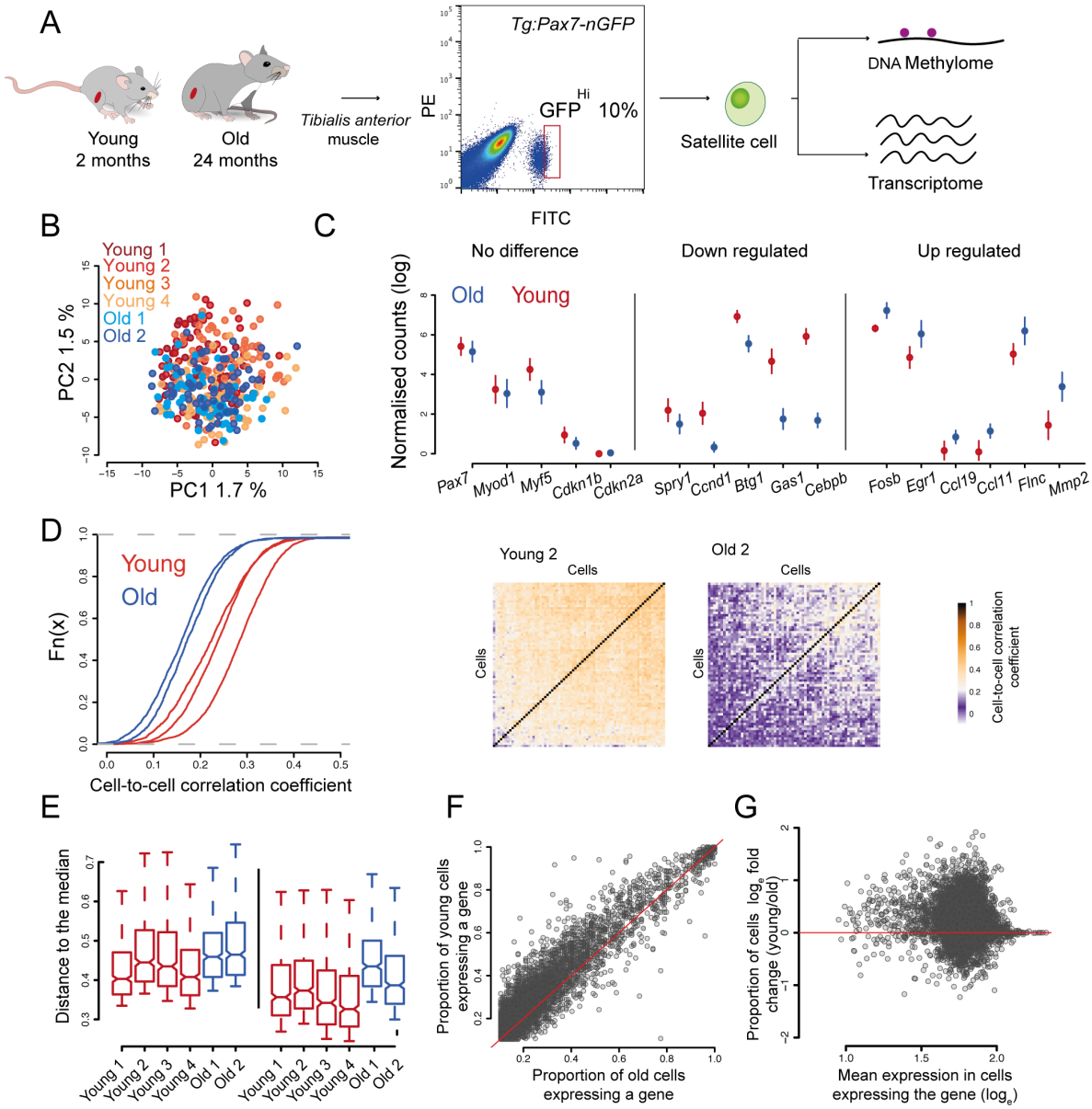
Author contributions: I.H.H., B.E., T.S., S.T. and W.R. proposed the concept and designed the experiments. B.E and P.H.C. performed FACS; T.S. and S.C. performed library preparation and sequencing. I.H.H. developed the analysis methodologies and analysed the experiments with advice from SA. I.H.H., B.E., S.T. and W.R. wrote the paper. All authors read and agreed on the manuscript.

Competing interests: W.R. is a consultant and shareholder of Cambridge Epigenetix. T.S. is CEO of Chronomics. All other authors declare no competing financial interests.

Materials & Correspondence: Correspondence and material requests should be addressed to shahragim.tajbakhsh@pasteur.fr and wolf.reik@babraham.ac.uk.

413 **Figures:**

Fig. 1



414

Fig. 1. Aged satellite cells have increased cell-to-cell transcriptional variability

(A) Experimental scheme. Single cells were isolated from *Tg:Pax7-nGFP* young and old mice and subjected to parallel single-cell methylation and RNA sequencing.

(B) PCA of a total of 377 cells from young (n=4) and old (n=2) individuals.

(C) Selected markers and differentially expressed genes between young and old cells (mean ± standard error).

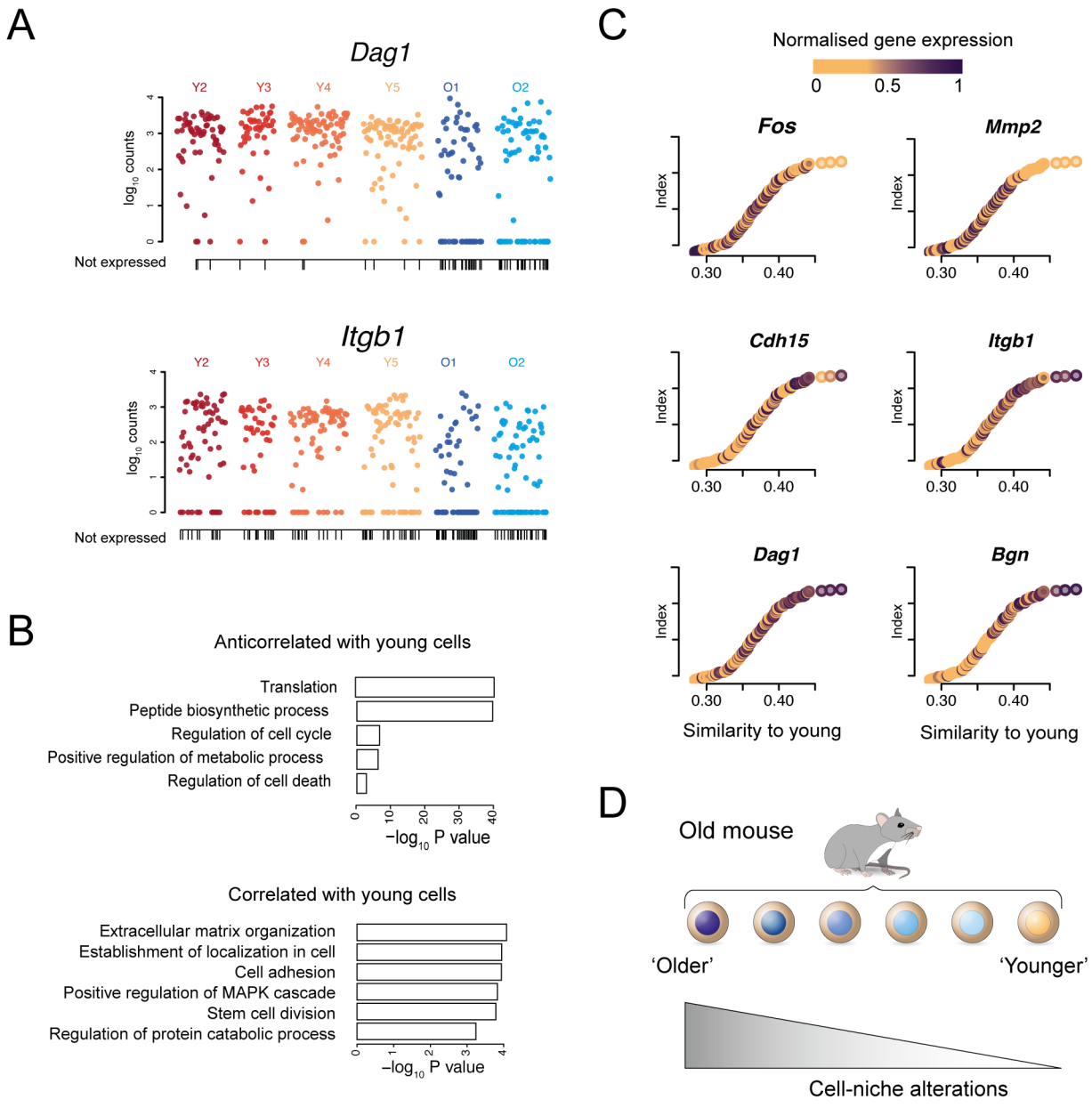
(D) Cumulative distribution of cell-to-cell Spearman correlation values per individual (left) showing that transcriptional heterogeneity dramatically increases with age. Heatmap showing cell-to-cell Spearman correlation values from a young and an old mouse (right).

(E) Distance to the median of the top 500 most variable genes among all genes (left) and of the top 500 most variable genes among the 5,127 common genes expressed in the six individuals (right).

(F) Frequency of gene expression in young and old cells.

(G) Independence between frequency of gene expression differences and gene expression level.

Fig. 2



429

Fig. 2. Variability within aged satellite cells and cell-niche interactions

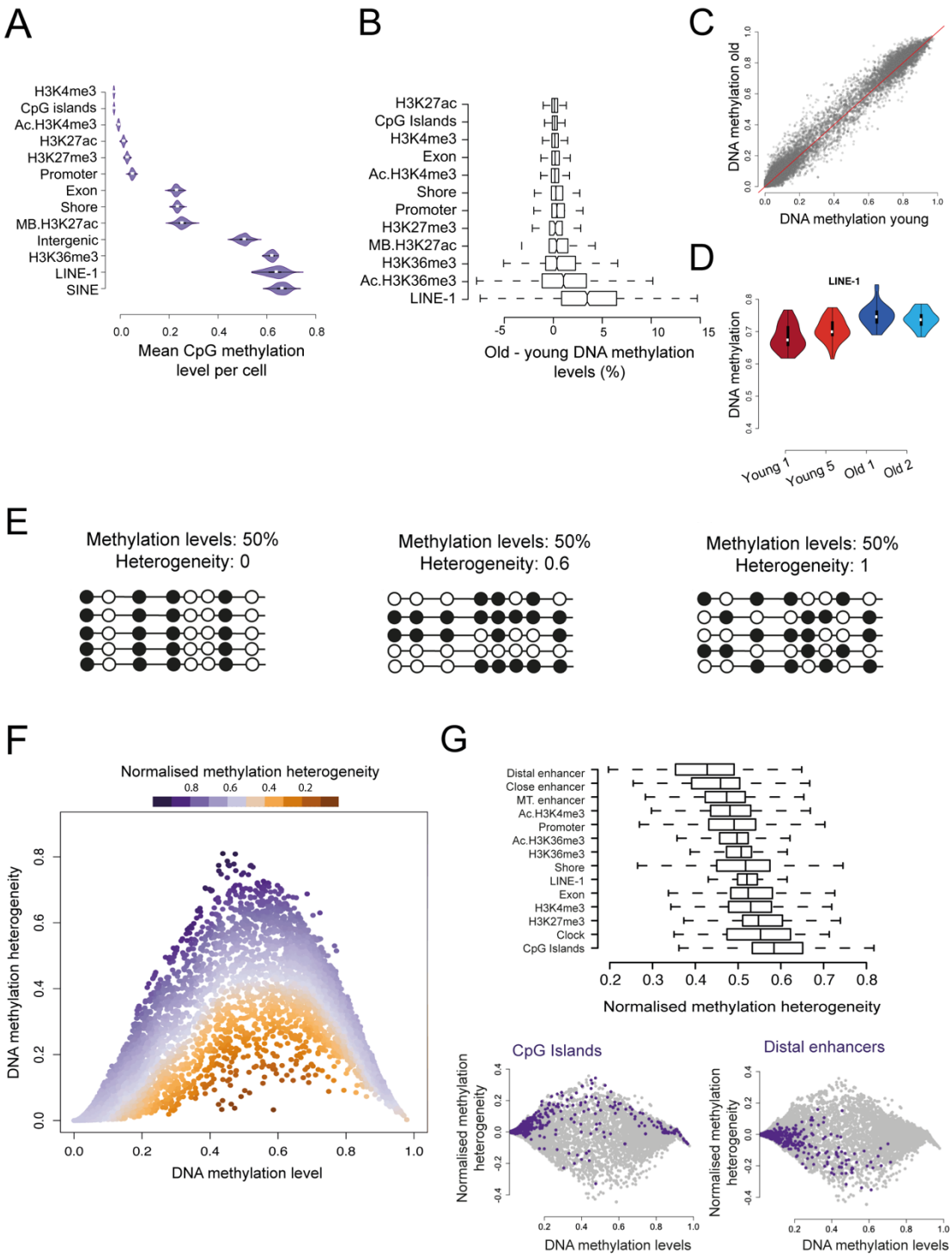
(A) *Dag1* and *Itgb1* expression in young and old cells. Each dot represents a cell. Vertical lines on the x-axis indicate cells that do not express the gene.

(B) P-values of the GO terms associated with the top 200 anticorrelated (top) and correlated (bottom) genes with the similarity score to young cells.

(C) Similarity between old and young cells. Each dot represents an old cell; on the x-axis cells are ordered according to their similarity (Spearman correlation coefficient) to young cells (young mean expression). Colours indicate the normalized levels of expression of selected genes correlated with the x-axis.

(D) Old cells diverging from young cells are likely to have impaired cell-niche interactions.

Fig. 3



440

Fig. 3. Changes in methylation levels and methylation heterogeneity

(A) Levels of DNA methylation per cell across different genomic regions (Chip-seq data from 2-months-old mice [23], Ac: activated satellite cells [23], MB: myoblast [29]).

(B) Mean methylation difference between old and young cells across different genomic elements.

(C) Genome-wide mean methylation values in old and young cells. Each dot represents a genomic region.

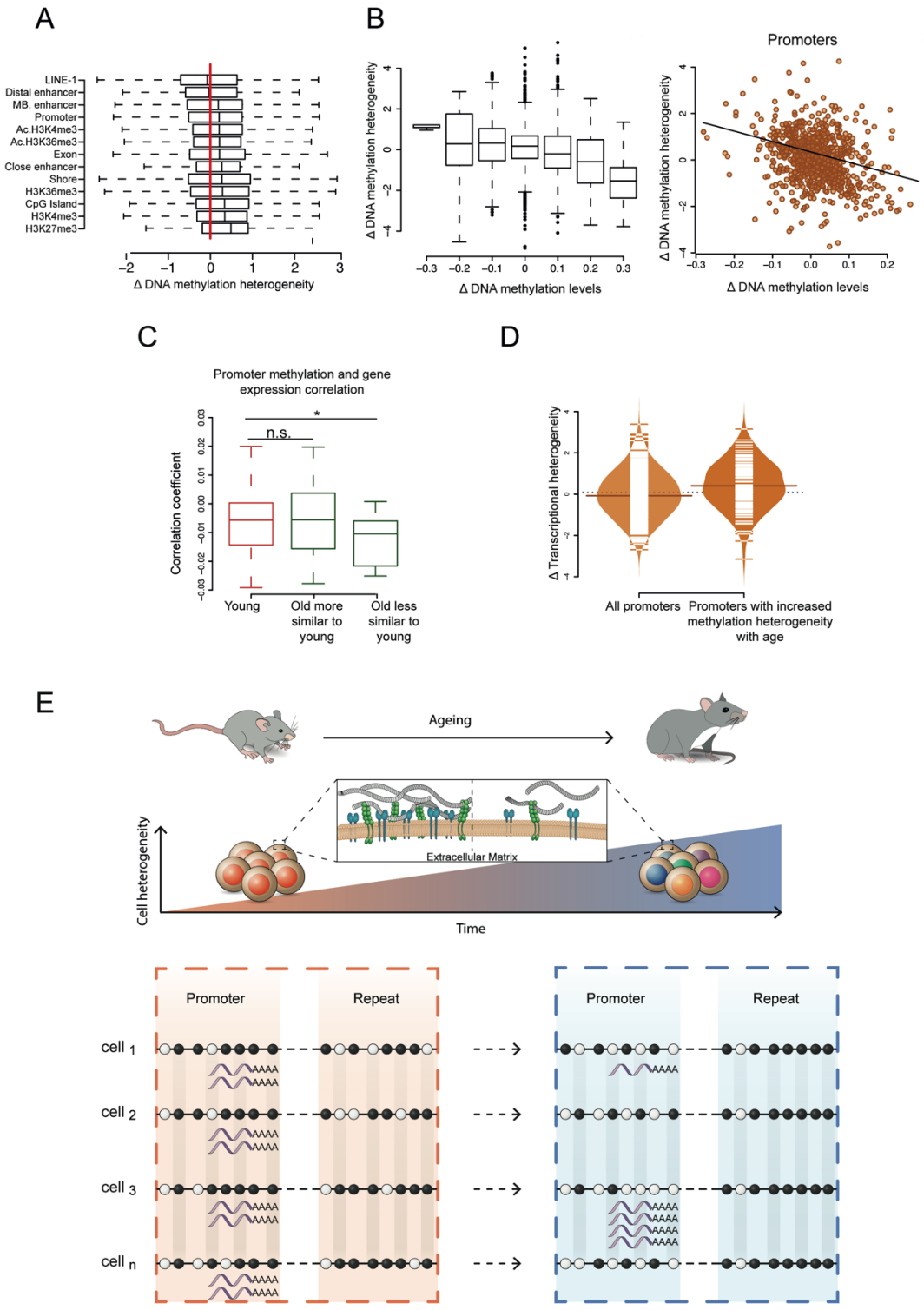
(D) Levels of DNA methylation per cell and individual across Line L1 elements.

(E) Examples of different distributions of DNA methylation heterogeneity at loci with similar average methylation. Empty circles represent unmethylated CpG sites and filled circles methylated CpG sites.

(F) DNA methylation levels and DNA methylation heterogeneity. Each dot represents a genomic region from young or old cells. Colour scale represents the methylation-level normalised measure of DNA methylation heterogeneity.

(G) Boxplot showing the normalised DNA methylation heterogeneity across different genomic elements in young cells (top). Normalised methylation heterogeneity and methylation levels across all the different genomic elements (grey) and across CpG Islands (purple) or enhancer regions (purple) in young cells (bottom).

Fig. 4



459

Fig. 4. Changes in cell-to-cell methylation heterogeneity during ageing

(A) Normalised methylation heterogeneity changes with age (Δ methylation heterogeneity: old–young) across different genomic features (Ac: activated satellite cells [23], MB: myoblast [29]).

(B) Genome-wide normalised methylation heterogeneity difference with ages (Δ methylation heterogeneity: old–young) binned by 0.1 methylation level differences (left). Changes in promoter methylation heterogeneity (y-axis) and methylation levels (x-axis) with age (right).

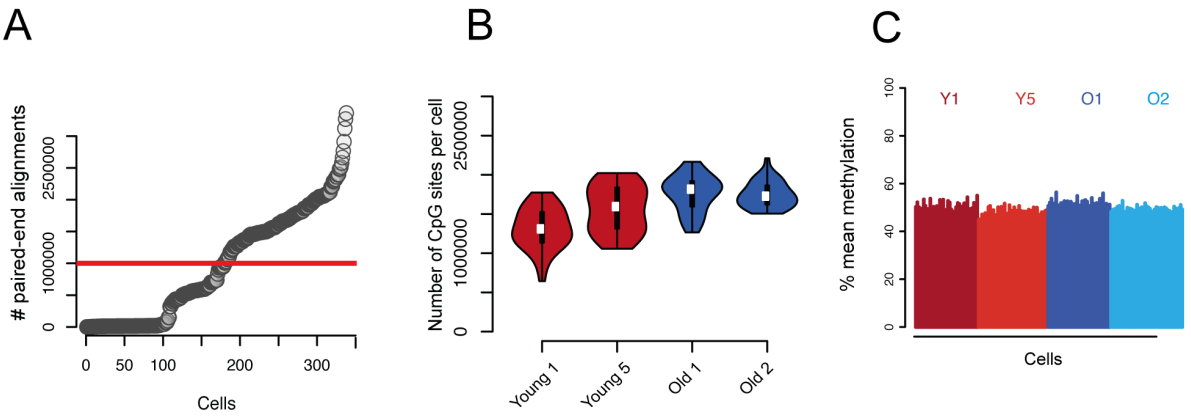
(C) Distribution of Pearson's correlation coefficients between promoter DNA methylation and gene expression (one association test per cell, number of cells: young = 64, old more similar to young = 30, old less similar to young = 20, * $P < 0.05$).

(D) Increase of transcriptional heterogeneity with age across all promoters (n=394) and promoters with increased DNA methylation heterogeneity (Δ methylation heterogeneity > 0.3, n=113) ($P < 0.001$).

(E) Global increase of transcriptional cell-to-cell variability with age with enhanced heterogeneity in the multiple extracellular matrix related genes (top). Relationship between transcriptional and DNA methylation heterogeneity in aged satellite cells (bottom). Empty circles represent unmethylated CpG sites and filled circles methylated CpG sites. Repeat elements become more homogeneous with age by increasing their methylation levels in a coordinated manner. In contrast, promoter regions become more heterogeneous by randomly loosing DNA methylation and this is coupled with an increase of transcriptional variability of the genes they drive.

480 **Extended data:**

Fig. S1



481

482 **Fig. S1. Quality control of single-cell DNA methylation data.**

483 (A) Number of pair-end alignments per cell. Cells below the threshold were excluded from the

484 study.

485 (B) Number of CpG sites per cell and individual.

486 (C) Mean methylation per cell showing no global differences between ages.

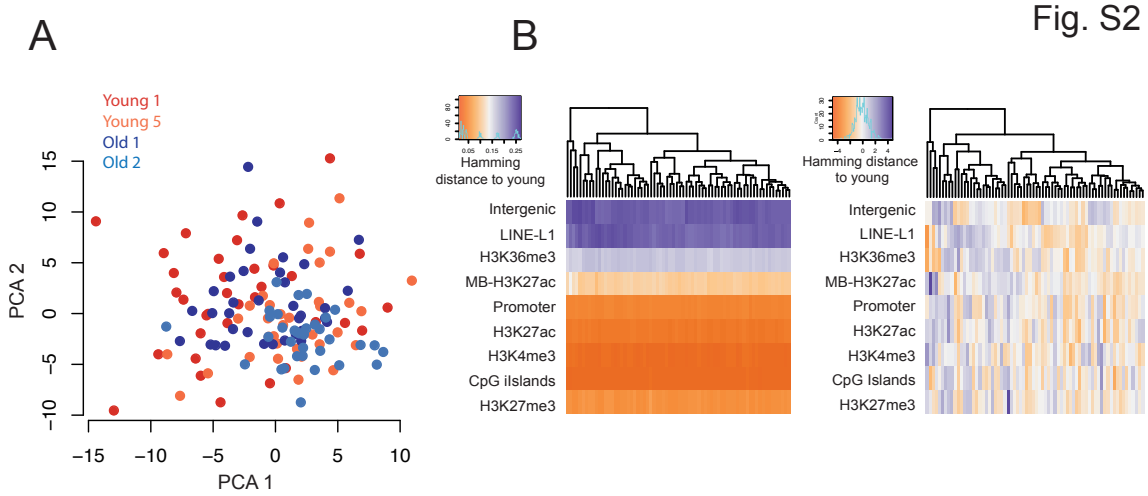
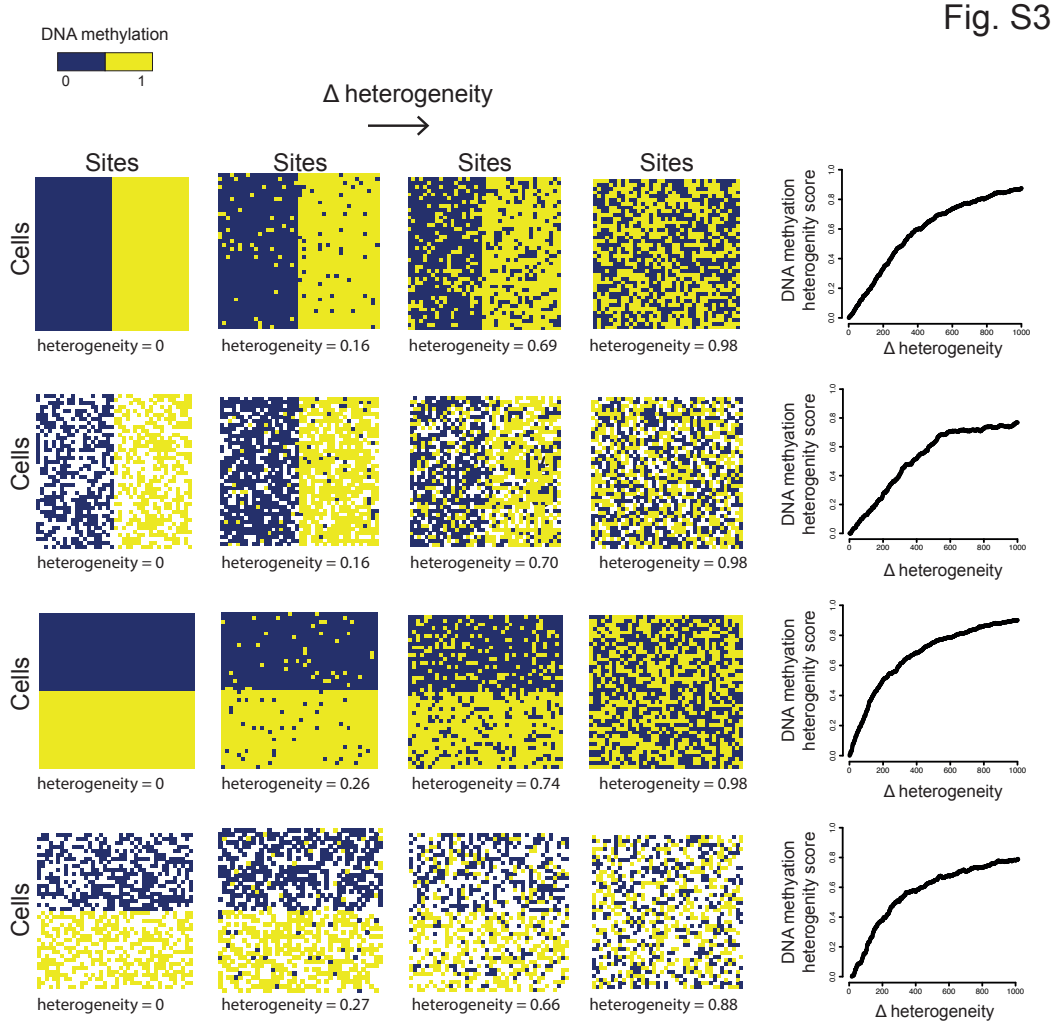


Fig. S2. Cell clustering based on DNA methylation data

(A) PCA on gene body methylation showing no clear differences between ages.

(B) Heatmap showing Hamming distances between the average methylation from young cells and individual old cells (columns) across different genomic context (rows) (left). Same measure normalised by genomic context (right) showing no cellular substructure.

Fig. S3



493

494 **Fig. S3. DNA methylation heterogeneity on simulated data.**

495 Four cases with different population substructure and missing values tested with simulated data

496 of increasing heterogeneity. Missing values are represented in white.

Fig. S4

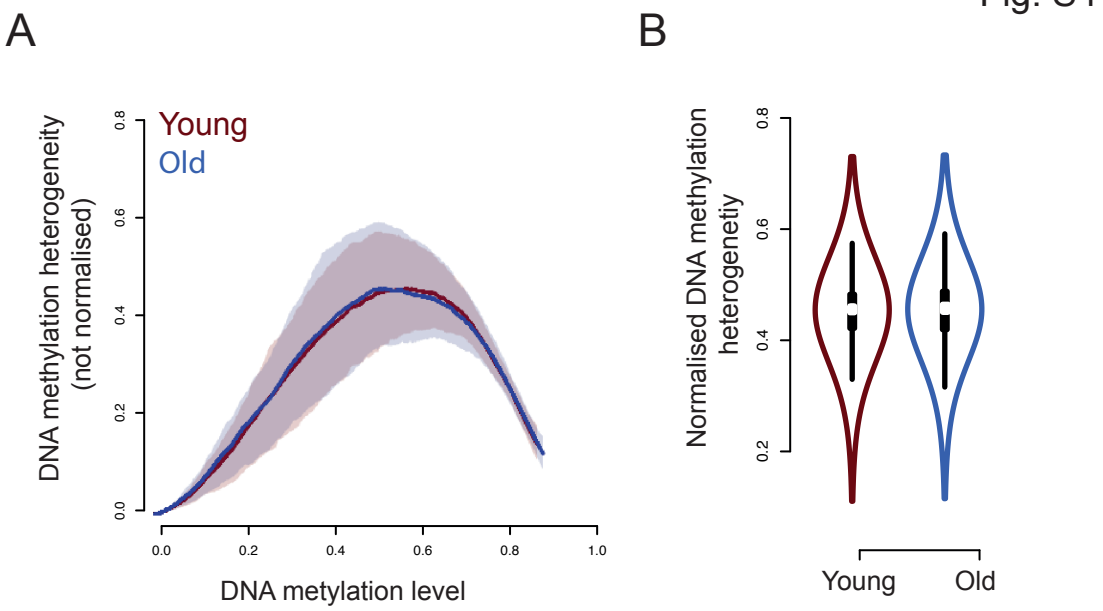


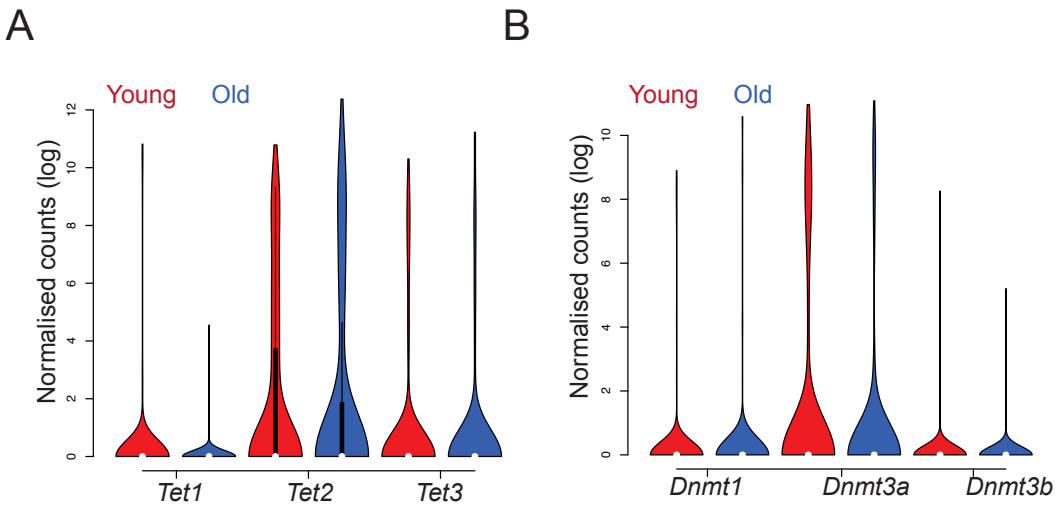
Fig. S4. Global levels of DNA methylation heterogeneity between ages.

(A) DNA methylation levels and methylation heterogeneity in young and old cells.

(B) Normalised DNA methylation heterogeneity in young and old cells.

501

Fig. S5



502

503 **Fig. S5. Expression levels of the DNA methylation enzymes.**

504 (A) Expression levels of the enzymes for active demethylation in young and old samples.

505 (B) Expression levels of the DNA methylation enzymes in young and old samples.

Fig. S6

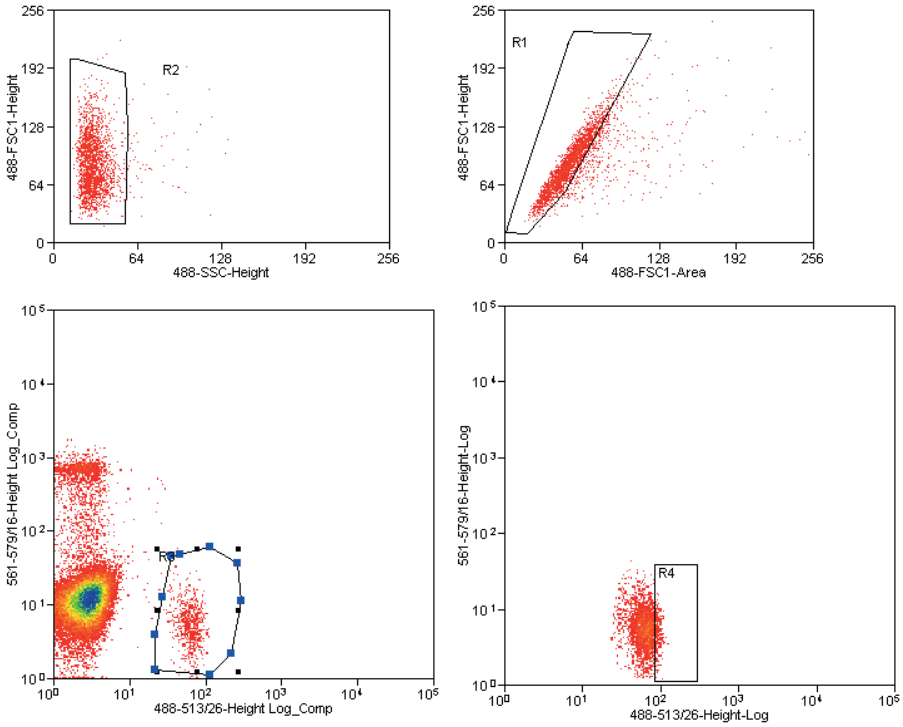
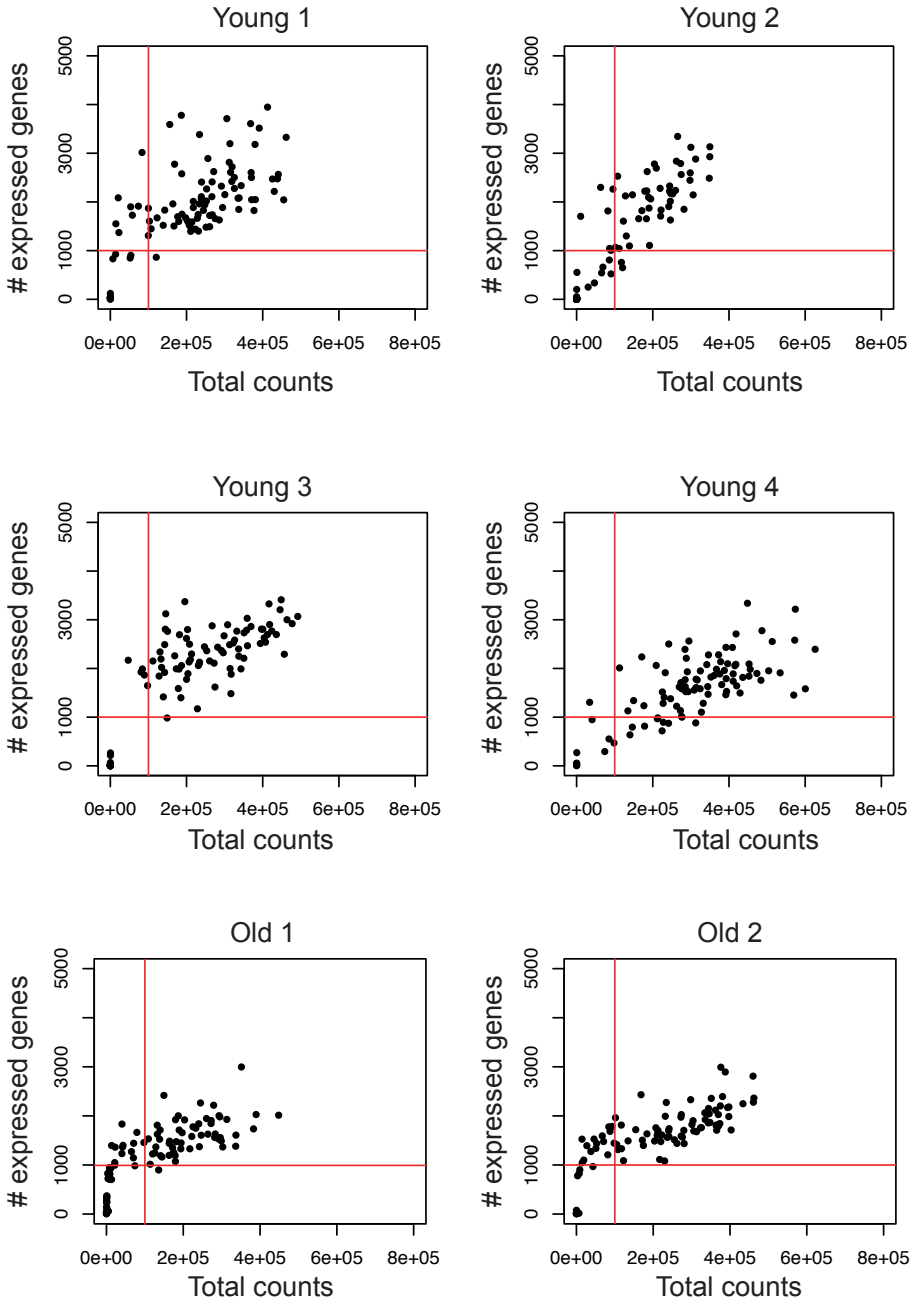


Fig. S6. Isolation of single satellite cells by FACS.

Satellite cells were isolated by FACS by gating first on size and granulosity (R2 gate), excluding doublets (R1 gate) and gating on the GFP^{+}/PI^{-} population (R3 gate). Pax7-$nGFP^{Hi}$ cells (top 10% highest nGFP-expressing cells, R4 gate) were sorted as single cells.

Fig. S7



511

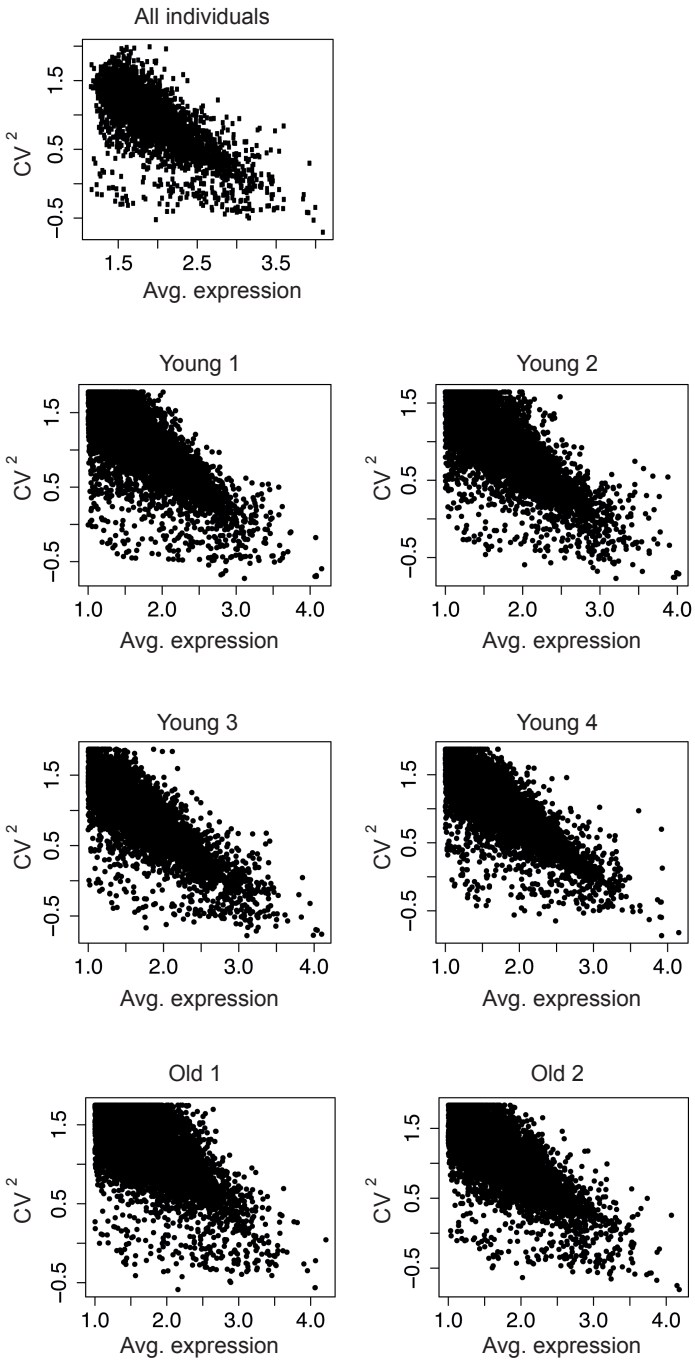
512 **Fig. S7. Quality control of single-cell RNA-seq data.**

513 Plot representing number of genes and total expression counts expressed in each cell per

514 individual. Cells above highlighted threshold (1000 genes, 10^5 counts) were included in the

515 study.

Fig. S8



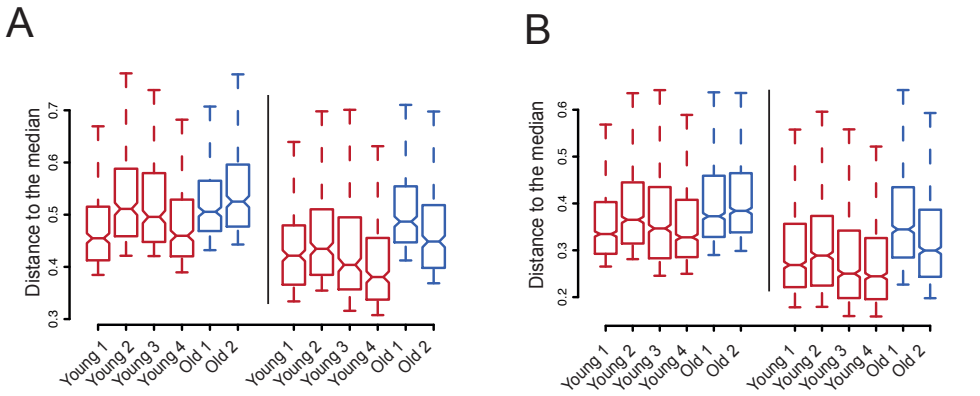
516

517 **Fig. S8. Transcriptional variability.**

518 Gene variability: squared coefficients of variation are plotted against the means of normalized

519 read counts for gene using data from all individuals (top) or each individual separately.

Fig. S9



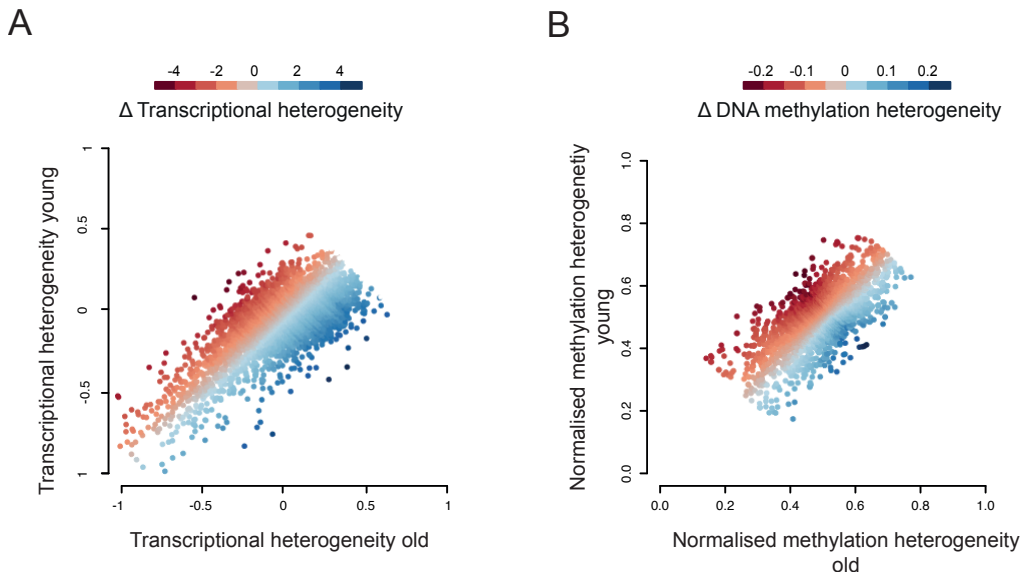
520

521 **Fig. S9. Transcriptional variability: distance to the median**

522 Distance to the median of the top 300 (A) and 1000 (B) most variable genes among all genes

523 (left) and among the 5,127 common genes expressed in the six individuals (right).

Fig. S10



524

525 **Fig. S10. Changes in transcriptional and DNA methylation heterogeneity with age.**

526 (A) Differences in transcriptional heterogeneity measures where Z-score normalised using a
527 sliding window of 100 observations (color code). Transcriptional heterogeneity represents the
528 mean distance to the median for every gene from young (y-axis) and old (x-axis) individuals.
529 (B) Differences in DNA methylation heterogeneity measures where Z-score normalised using
530 a sliding window of 100 observations (color code). DNA methylation heterogeneity represents
531 the normalised measure of methylation heterogeneity from young (y-axis) and old (x-axis)
532 individuals.

Supplementary tables:

Table S1. Differentially expressed genes between cells from young and old mice.

Table S2. scM&T quality control.

		RNA		DNA	
Sequencing label	**Individual**	**Total cells**	**After QC**	**Total cells**	**After QC**
Y2	**Young 1**	96	0	78	35
Y8	**Young 2**	96	75	86	35
Y5	**Young 3**	96	44	NA	NA
Y7	**Young 4**	96	74	72	35
Y4	**Young 5**	96	60	NA	NA
O1	**Old 1**	96	56	80	35
O5	**Old 2**	96	68	90	35
O8	**Old 3**	96	5	8	0

Table S3. Top 1000 most variable genes across the entire data set.

Table S4. Top 200 genes correlated and anticorrelated with the similarity score to the young reference transcriptome.

Table S5. Increase of transcriptional and methylation heterogeneity with age in promoter regions (Δ: Old-young).