bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 # Meta-Analysis of the Gene Expression Profiles of Aging Brain Reveals a
2 Consistent Increase in Heterogeneity

3

4 Ulaş Işıldak[1], Mehmet Somel[1], Janet M. Thornton[2], Handan Melike Dönertaş[2*]

5

6 [1]Department of Biological Sciences, Middle East Technical University, 06800, Ankara, Turkey.
7 [2]European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome
8 Campus, Hinxton, Cambridge, CB10 1SD, UK.
9 *correspondence: donertas.melike@gmail.com

10

11 **ORCID**
12 *Ulaş Işıldak - https://orcid.org/0000-0001-6497-6254*
13 *Mehmet Somel - https://orcid.org/0000-0002-3138-1307*
14 *Janet M. Thornton - http://orcid.org/0000-0003-0824-4096*
15 *Handan Melike Dönertaş - http://orcid.org/0000-0002-9788-6535*

16

17

18 ## Abstract
19 In largely non-mitotic tissues such as the brain, cells are prone to a gradual accumulation of
20 stochastic genetic and epigenetic alterations. This may lead to increased gene expression variation
21 between cells and possibly also between individuals over time. Although increased inter-individual
22 heterogeneity in gene expression during brain aging was previously reported, whether this process
23 starts during development or if it is mainly restricted to the aging period has not yet been studied. The
24 regulatory dynamics and functional significance of putative age-related heterogeneity are also
25 unknown. Here we address these issues by a systematic analysis of 19 transcriptome datasets from
26 diverse human brain regions in human covering the whole postnatal lifespan. Among all datasets, we
27 observed a significant increase in inter-individual gene expression heterogeneity during aging (20 to
28 98 years of age) compared to postnatal development (0 to 20 years of age). Increased heterogeneity
29 during aging was consistent among different brain regions at the gene level. Genes showing
30 increased heterogeneity were associated with biological processes that are known to be important for
31 lifespan regulation and neuronal function, including longevity regulating pathway, autophagy, mTOR
32 signaling, axon guidance, and synaptic function. Overall, our results show that increased gene
33 expression heterogeneity during aging is a general effect in the human brain, and may influence
34 aging-related changes in brain functions. We also provide the necessary functions to calculate
35 heterogeneity change with age as an R package, 'hetAge'.

36

37 **Keywords:** aging, development, gene expression, transcriptome, heterogeneity, human, brain

38

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

39 ## Introduction

40 Aging is a complex process characterized by a gradual decline in maintenance and repair
41 mechanisms, accompanied by an increase in genetic and epigenetic mutations, and oxidative
42 damage to nucleic acids, protein and lipids (Gorbunova, Seluanov, Mao, & Hine, 2007; Lu et al.,
43 2004). The human brain experiences dramatic structural and functional changes in the course of
44 aging. These include decline in gray matter and white matter volumes (Sowell, Thompson, & Toga,
45 2004), increase in axonal bouton dynamics (Grillo et al., 2013) and reduced synaptic plasticity, all
46 processes that may be associated with decline in cognitive functions (Dorszewska, 2013). Changes
47 during brain aging are suggested to be a result of stochastic processes, unlike changes associated
48 with postnatal neuronal development that are known to be primarily controlled by adaptive regulatory
49 processes (Polleux, Ince-Dunn, & Ghosh, 2007; Schratt, 2009; Stefani & Slack, 2008). The molecular
50 mechanisms underlying age-related alteration of regulatory processes and eventually leading to
51 aging-related phenotypes, however, are little understood.

52

53 Over the past decade, a number of transcriptome studies focusing on age-related changes in human
54 brain gene expression profiles were published (Kang et al., 2011; Lu et al., 2004; Miller et al., 2014;
55 Somel et al., 2010; Tebbenkamp, Willsey, State, & Šestan, 2014). These studies report aging-related
56 differential expression patterns in many functions, including synaptic functions, energy metabolism,
57 inflammation, stress response, and DNA repair. By analyzing age-related change in gene expression
58 profiles in diverse brain regions, we previously showed that for many genes, gene expression
59 changes occur in opposite directions during postnatal development (pre-20 years of age) and aging
60 (post-20 years of age), which may be associated with aging-related phenotypes in healthy brain aging
61 (Dönertaş et al., 2017). While different brain regions are associated with specific, and often
62 independent, gene expression profiles (Kang et al., 2011; Miller et al., 2014; Tebbenkamp et al.,
63 2014), these studies also show that age-related alteration of gene expression profiles during aging is
64 a widespread effect across different brain regions.

65

66 One of the suggested effects of aging is increased variability between individuals and somatic cells,
67 which has been previously reported by several studies. Some of these studies find an increase in
68 age-related heterogeneity in heart, lung and white blood cells of mice (Angelidis et al., 2019; Bahar et
69 al., 2006; Martinez-Jimenez et al., 2017), *Caenorhabditis elegans* (Herndon et al., 2002), and
70 human twins (Fraga et al., 2005). A study analysing microarray datasets from different tissues of
71 humans and rats also reported an increase in age-related heterogeneity in expression as a general
72 trend (Somel, Khaitovich, Bahn, Pääbo, & Lachmann, 2006), although this study found no significant
73 consistency across datasets, nor any significant enrichment in functional gene groups. That said, the
74 generality of increase in expression heterogeneity remains unresolved. For instance, Viñuela et al.
75 find more decrease than an increase in heterogeneity in human twins (Viñuela et al., 2018) and
76 Ximerakis et al. show the direction of the heterogeneity change depends on cell type in aging mice

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

77 brain (Ximerakis et al., 2018). Using GTEx data covering different brain regions (20 to 70 years of
78 age), Brinkmeyer-Langford et al. identify a set of differentially variable genes between age groups, but
79 they do not observe increased heterogeneity at old age (Brinkmeyer-Langford, Guan, Ji, & Cai, 2016).
80 Meanwhile, another study performing single-cell RNA sequencing of human pancreatic cells, identifies
81 an increase in transcriptional heterogeneity and somatic mutations with age (Enge et al., 2017). A
82 meta-analysis also suggested more shared expression patterns during development than in aging,
83 implying an increase in inter-individual variability (Dönertaş et al., 2017). Likewise, a prefrontal cortex
84 transcriptome analysis we recently conducted revealed a weak increase in age-dependent
85 heterogeneity at the gene, transcriptome and pathway levels, irrespective of the preprocessing
86 methods (Kedlian, Donertas, & Thornton, 2019).
87
88 Whether age-related increase in heterogeneity is a universal phenomenon thus remains contentious.
89 Furthermore, where it can be detected, whether this is a time-dependent process that starts at the
90 beginning of life or whether this increase and its functional consequences are only seen after
91 developmental processes are completed, have not yet been explored. In this study, we retrieved
92 transcriptome data from independent studies covering the whole lifespan, including data from diverse
93 brain regions, and conducted a comprehensive analysis to identify the prevalence of age-related
94 heterogeneity changes in human brain aging compared with those observed during postnatal
95 development. We confirmed that increased age-related heterogeneity is a consistent trend in the
96 human brain transcriptome during aging but not during development, and it is associated with the
97 pathways and biological functions that are related to longevity and neuronal function.
98

99 Results

100 To investigate how heterogeneity in gene expression changes with age, we used 19 published
101 microarray datasets from three independent studies. Datasets included 1,010 samples from 17
102 different brain regions of 298 individuals whose ages ranged from 0 to 98 years (Table S1, Figure
103 S1). In order to analyze the age-related change in gene expression heterogeneity during aging
104 compared to the change in development, we divided datasets into two subsets as development (0 to
105 20 years of age, n = 441) and aging (20 to 98 years of age, n = 569). We used the age of 20 to
106 separate pre-adulthood and adulthood based on commonly used age intervals in earlier studies (see
107 Methods). For the analysis, we focused only on the genes for which we have a measurement across
108 all datasets (n = 11,137).
109

110 Age-related change in gene expression levels

111 To quantify age-related changes in gene expression, we used a linear model between gene
112 expression levels and age (see Methods, Figure S2). We transformed the ages to the fourth root
113 scale before fitting the model as it provides relatively uniform distribution of sample ages across the
114 lifespan (as in Somel et al., 2010), but we also confirmed that different age scales yield quantitatively

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

similar results (see Figure S3). We quantified expression change of each gene in aging and development periods separately and considered regression coefficients from the linear model (β values) as a measure of age-related expression change (Figure S4, Table S2).

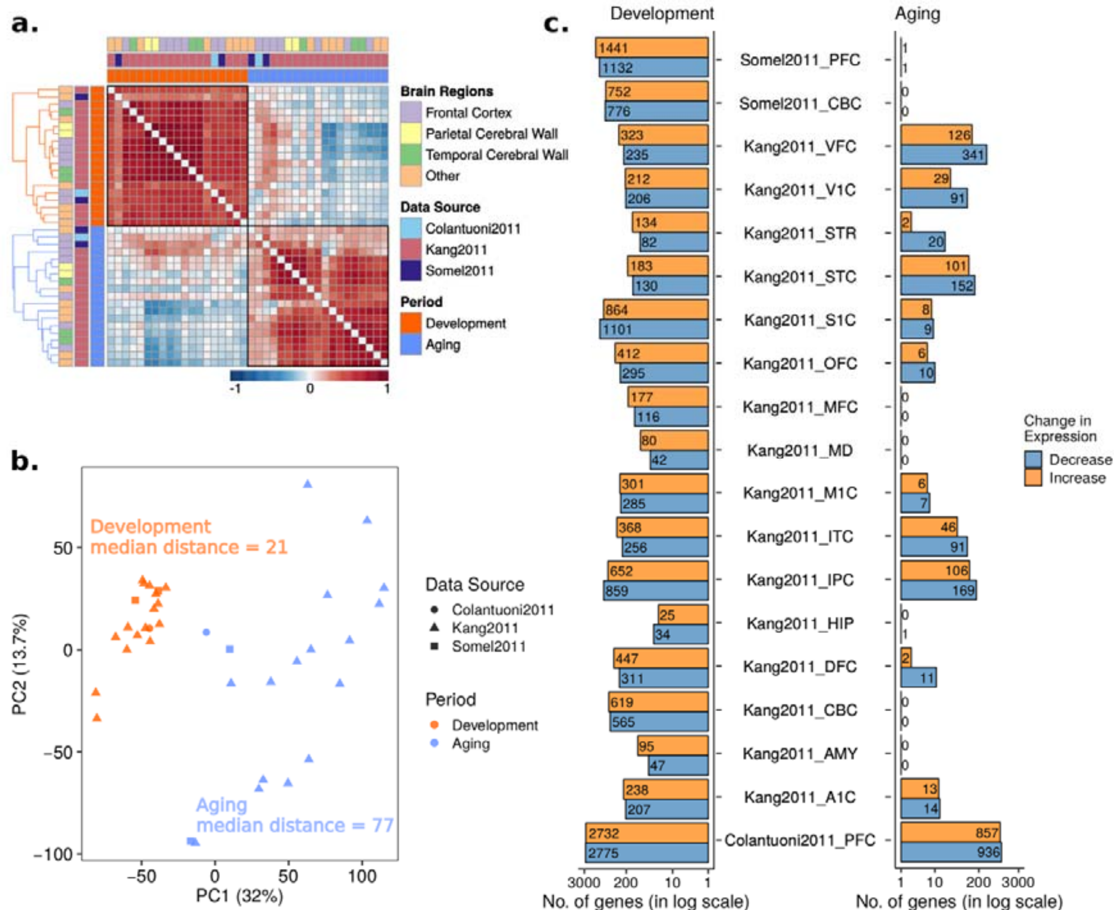


Figure 1. Age-related change in gene expression during postnatal development and aging. (a) Spearman correlations among age-related expression changes (β values) across datasets. The color of the squares indicates if the correlation between the corresponding pair of datasets (across β values of 11,137 common genes) is positive (red) or negative (blue), while darker color specifies a stronger correlation. Diagonal values were removed in order to enhance visuality. Annotation rows and columns indicate data source, brain region and period of each dataset. Hierarchical clustering was performed for each period separately (color of the dendrogram indicates periods) to determine the order of datasets. (b) Principal component analysis (PCA) of age-related expression changes during aging and development. The analysis was performed on age-related expression change values of 11,137 common genes among all 38 datasets. The values of the first principal component on the x-axis and second principal component on the y-axis were drawn, where the values in the parenthesis indicate the variation explained by the corresponding principal component. Median Euclidean pairwise distances among development and aging datasets calculated using PC1 and PC2 were annotated on the figure. Different shapes show different data sources and colors show development (dark orange) and aging (blue) (c) Number of significant (FDR corrected $p < 0.05$) gene expression changes in development (left panel)

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

134 *and aging (right panel). The x-axis shows the number of genes in the log scale. The color of the bars shows the*
135 *direction of change, decrease (steel gray), and increase (orange). The exact number of genes are also displayed*
136 *on the plot.*
137
138 We first analyzed similarity in age-related expression changes across datasets by calculating pairwise
139 Spearman's correlation coefficients among the β values (Figure 1a). Both development (median
140 correlation coefficient = 0.56, permutation test $p < 0.001$, Figure S5a) and aging datasets (median
141 correlation coefficient = 0.43, permutation test $p = 0.003$, Figure S5b) showed moderate correlation
142 with the datasets within the same period. Although the difference between dataset correlations within
143 development and aging datasets was not significant (permutation test $p = 0.1$, Figure S6a), weaker
144 consistency during aging may reflect the stochastic nature of aging, causing increased heterogeneity
145 between aging datasets.
146
147 The principal component analysis (PCA) of age-related expression changes (β) revealed distinct
148 clusters of development and aging datasets (Figure 1b). Moreover, aging datasets were more
149 dispersed than development datasets (median pairwise Euclidean distances between PC1 and PC2
150 were 77 for aging and 21 for development), which may again reflect stochasticity in gene expression
151 change during aging and can indicate more heterogeneity among different brain regions or datasets
152 during aging than in development.
153
154 We next identified genes showing significant age-related expression change (FDR-corrected $p <$
155 0.05), for development and aging datasets separately (Figure 1c). Development datasets showed
156 more significant changes compared to aging (permutation test $p = 0.003$, Figure S6c), which may
157 again indicate higher expression variability among individuals during aging. The direction of change in
158 development was mostly positive (14 datasets with more positive and 5 with more negative), whereas
159 in aging datasets, we observed more genes with a decrease in expression level (13 datasets with
160 more genes decreasing expression and 5 with no significant change, and 1 with an equal number of
161 positive and negative changes).
162

163 **Age-related change in gene expression heterogeneity**

164 To assess age-related change in heterogeneity, we obtained the unexplained variance (residuals)
165 from the linear models used to calculate the change in gene expression level. For each gene in each
166 dataset, we separately calculated Spearman's correlation coefficient (ρ) between the absolute value
167 of residuals and age, irrespective of whether the gene shows a significant change in expression (see
168 Methods, Figure S2). We considered ρ values as a measure of heterogeneity change, where positive
169 values mean an increase in heterogeneity with age (Table S2). We also repeated this approach using
170 loess regression instead of a linear model between expression level and age, and found high
171 correspondence between ρ values based on linear and loess regression models (Figure S7). Still,

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

172 loess regression was more sensitive to the changes in sample sizes and parameters and we therefore
173 continued downstream analyses with the ρ estimates based on the residuals from the linear model.
174
175 We next asked if datasets show similar ρ, *i.e.* age-related changes in heterogeneity, by calculating
176 pairwise Spearman's correlation between pairs of datasets, across shared genes (Figure 2a). Unlike
177 the correlations among expression level changes, ρ values did not show a higher consistency during
178 development. In fact, although the difference is not significant (permutation test $p = 0.2$, Figure S6b),
179 the median value of the correlation coefficients was higher in aging (median correlation coefficient =
180 0.21, permutation test $p = 0.24$, Figure S5c), than in development (median correlation coefficient =
181 0.11, permutation test $p = 0.25$, Figure S5d).
182
183 A principal component analysis (PCA) showed that, like expression change, heterogeneity change
184 with age can also differentiate aging datasets from development (Figure 2b). Similar to the pairwise
185 correlations (Figure 2a), aging datasets clustered more closely than development datasets (median
186 pairwise Euclidean distances between PC1 and PC2 are 41 and 44 for aging and development,
187 respectively). Both observations imply more similar changes in heterogeneity during aging.
188
189 Using the *p*-values from Spearman's correlation between age and the absolute value of residuals for
190 each gene, we then investigated the genes showing a significant change in heterogeneity during
191 aging and development (FDR corrected *p*-value < 0.05). We found almost no significant change in
192 heterogeneity during development, except for the Colantuoni2011 dataset, for which we have high
193 statistical power due to its large sample size. In aging datasets, on the other hand, we observed more
194 genes with significant changes in heterogeneity (permutation test $p = 0.06$, Figure S6d) and the
195 majority of the genes with significant changes in heterogeneity tended to increase in heterogeneity
196 (Figure 2c). However, the genes showing a significant change did not overlap across aging datasets
197 (Figure S8).
198
199 Nevertheless, our analyses indicated relatively more consistent heterogeneity changes among
200 datasets in aging compared to development, implying that heterogeneity change could be a
201 characteristic linked to aging (see Discussion).
202

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

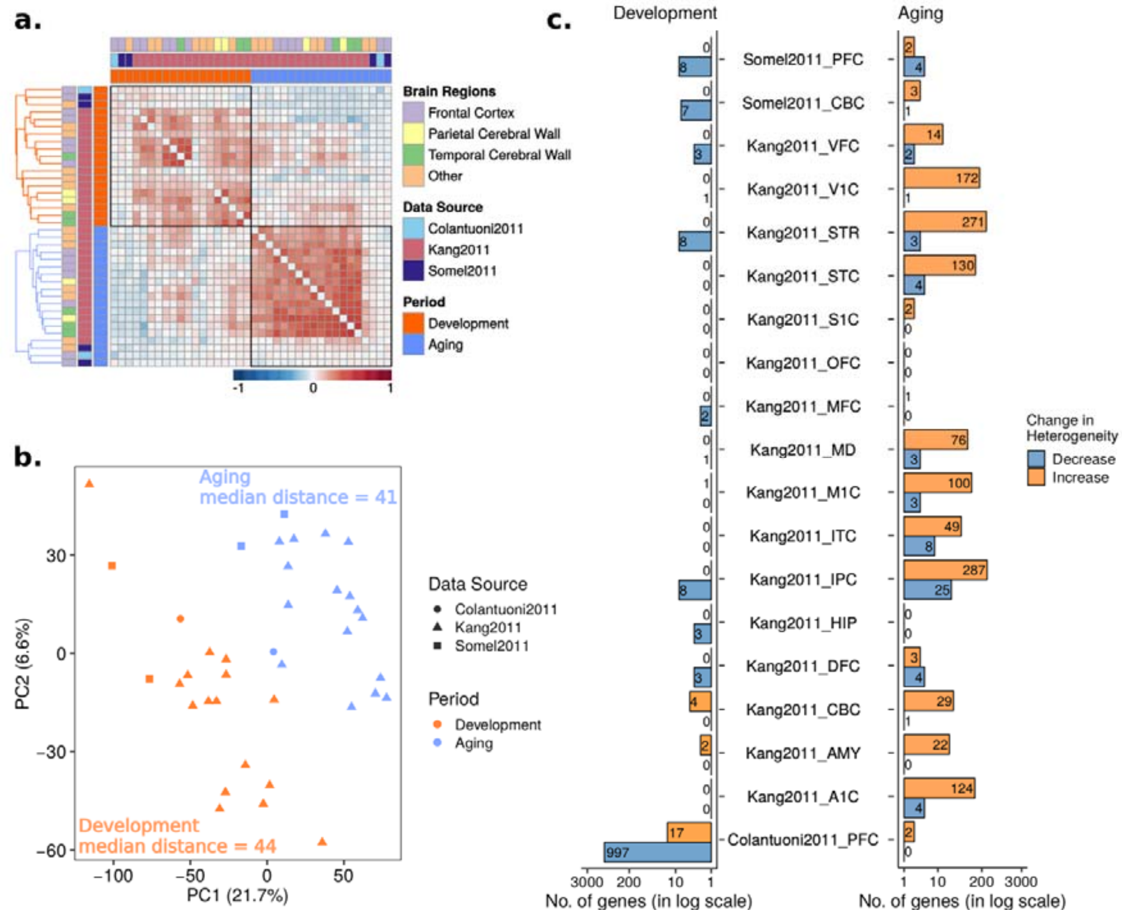


__Figure 2.__ Age-related change in gene expression heterogeneity during development and aging. The procedures are similar to those in Figure 1, except, age-related heterogeneity changes (ρ values) were used instead of expression changes (β values). (a) Spearman correlations among age-related heterogeneity changes (ρ values) across datasets. (b) Principal component analysis (PCA) of heterogeneity change with age. (c) The number of genes showing significant heterogeneity change in aging and development.

Consistent increase in heterogeneity during aging

As our previous analyses suggested age-related changes in heterogeneity can differentiate development from aging and show more similarity during aging, we sought to characterize the genes displaying such changes. Since the significance of the changes is highly dependent on the sample size, instead of focusing on significant genes identified within individual datasets, we leveraged upon the availability of multiple datasets and focused on their shared trends, capturing weak but reproducible trends across multiple datasets (as in Dönertaş et al., 2017). Consequently, we used the level of consistency in age-related heterogeneity change across datasets to sort genes.

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

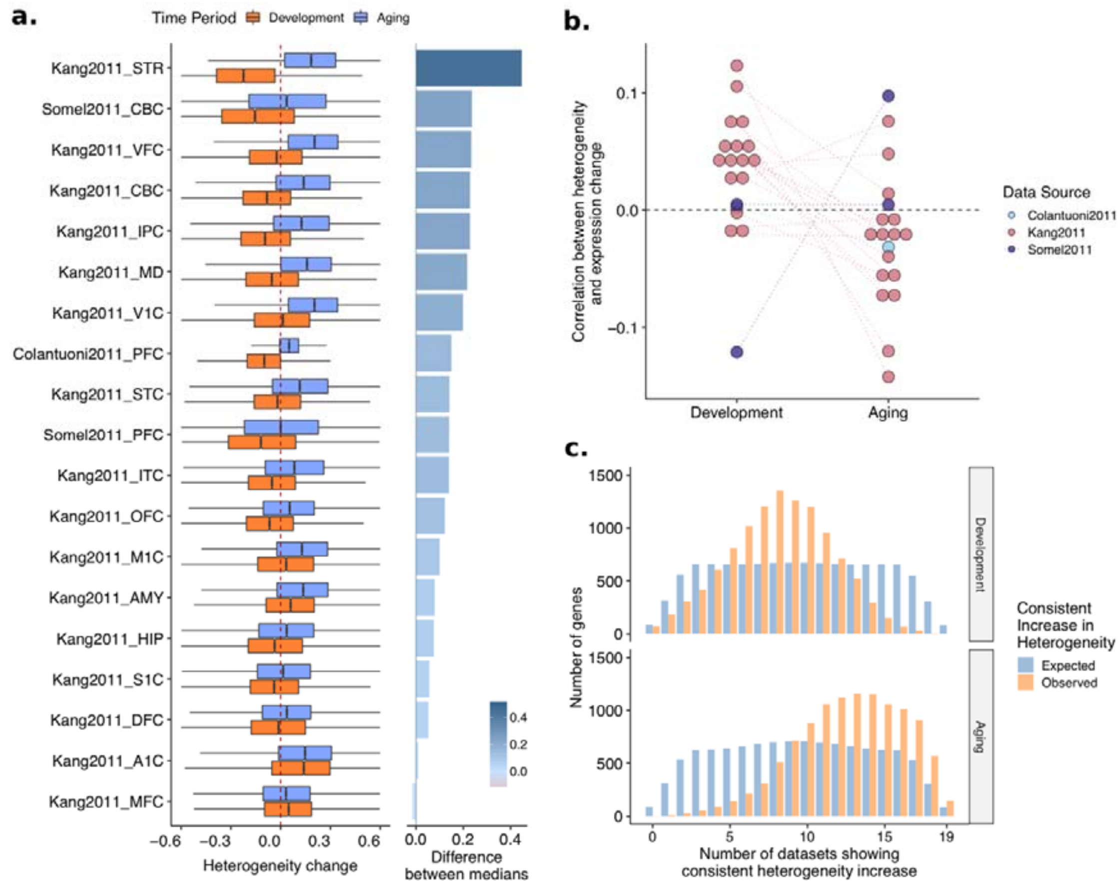


Figure 3. *(a) Boxplots, showing distributions of age-related heterogeneity changes (ρ values) of 11,1137 common genes for each dataset and period separately. The dotted red line (vertical line at x = 0) reflects no change in heterogeneity. The difference between median heterogeneity change in aging and development is given as a bar plot on the right panel. Datasets are ordered by the differences in median heterogeneity changes in aging and development. (b) The relationship between expression and heterogeneity change with age. Spearman correlation analysis was performed between age-related expression changes (β values) and age-related heterogeneity changes (ρ values) of 11,137 common genes, separately for each dataset. The dotted gray line at y = 0 reflects no correlation between expression and heterogeneity. (c) Expected and observed consistency in the heterogeneity change across datasets in development and aging. There is a significant shift toward heterogeneity increase in aging (permutation test $p<10^{-7}$) (lower panel), while there is no significant consistency in either direction in development (upper panel). The expected distribution is constructed using a permutation scheme that accounts for the dependence among datasets and is more stringent than random permutations (see Figure S10 for details).*

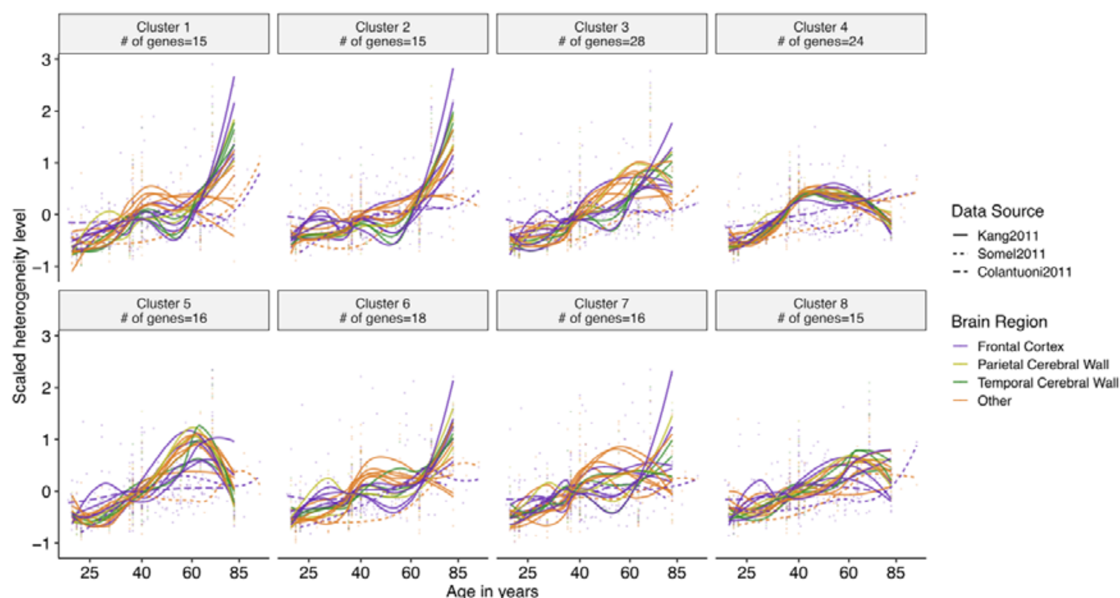
We first examined profiles of age-related heterogeneity change in aging and development. Among aging datasets 18/19 showed more increase than decrease in heterogeneity with age (median $\rho > 0$, *i.e.* higher numbers of genes with increase), while the median heterogeneity change in one dataset was zero. In development, on the other hand, only 5/19 datasets showed more increase in

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

238 heterogeneity, while the remaining 14/19 datasets showed more decrease with age (median $\rho < 0$)
239 (Figure 3a). The age-related change in heterogeneity during aging was significantly higher than during
240 development (permutation test $p < 0.001$, Figure S6e). We also checked if there is a relationship
241 between changes in heterogeneity during development and during aging (*e.g.* if those genes that
242 decrease in heterogeneity tend to increase in heterogeneity during aging) but did not find any
243 significant trend (Figure S9).
244
245 A potential explanation why we see different patterns of heterogeneity change with age in
246 development and aging could be the accompanying changes in the expression levels, as it is
247 challenging to remove dependence between the mean and variance. To address this possibility, we
248 first calculated Spearman's correlation coefficient between the changes in heterogeneity (ρ values)
249 and expression (β values), for each dataset. Overall, all datasets had values close to zero, suggesting
250 the association is not strong. Surprisingly, we saw an opposing profile for development and aging;
251 while the change in heterogeneity and expression were positively correlated in development, they
252 showed a negative correlation in aging (Figure 3b).
253
254 Having observed both a tendency to increase and a higher consistency in heterogeneity change
255 during aging, we asked which genes show consistent increase in heterogeneity across datasets in
256 aging and development. We therefore calculated the number of datasets with an increase in
257 heterogeneity during development and aging for each gene (Figure 3c). To calculate significance and
258 expected consistency, while controlling for dataset dependence, we performed 1,000 random
259 permutations of individuals' ages and re-calculated the heterogeneity changes (see Methods). In
260 development, there was no significant consistency in heterogeneity change in either increase or
261 decrease. During aging, however, there was a significant signal of consistent heterogeneity increase,
262 *i.e.* more genes showed consistent heterogeneity increase across aging datasets than randomly
263 expected (Figure 3c, lower panel). We identified 147 common genes with a significant increase in
264 heterogeneity across all aging datasets (permutation test $p < 0.001$, Table S3). Based on our
265 permutations, we estimated that 84/147 genes could be expected to have consistent increase just by
266 chance, suggesting only ~40% true positives. In development, in contrast, there was no significant
267 consistency in heterogeneity change in either direction (increase or decrease). Nevertheless,
268 comparing the consistency in aging and development, there was an apparent shift towards a
269 consistent increase in aging – even if we cannot confidently report the genes that become significantly
270 more heterogeneous with age across multiple datasets.
271
272 **Heterogeneity Trajectories**
273 We next asked if there are specific patterns of heterogeneity change, *e.g.* increase only after a certain
274 age. We used the genes with a consistent increase in heterogeneity with age during aging ($n = 147$)
275 to explore the trajectories of heterogeneity change (Figure 4). Genes grouped with k-means clustering

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

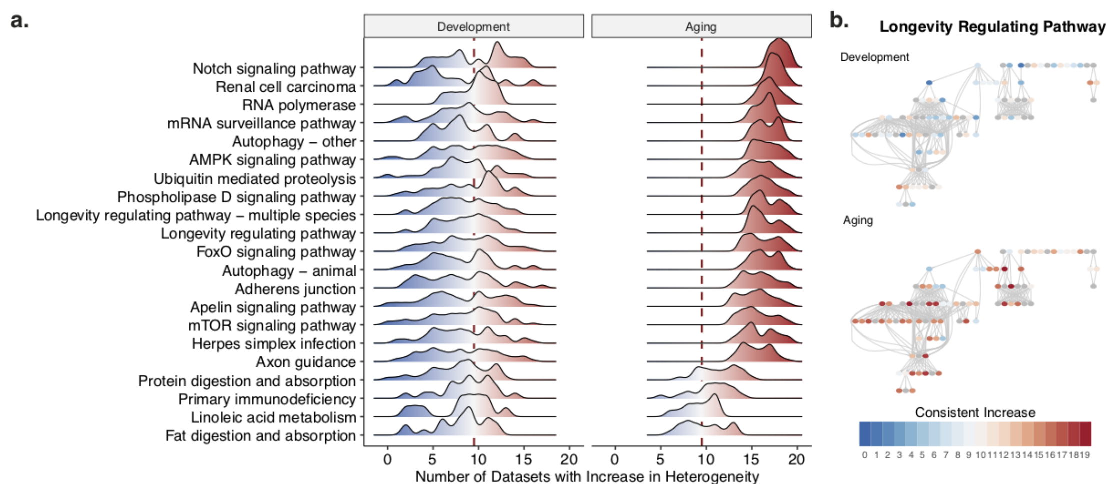
276 revealed three main patterns of heterogeneity increase (Table S3): i) genes in clusters 3 and 7 show
277 noisy but *a steady increase* throughout aging, ii) genes in clusters 4, 5 and 8 show *increase in early*
278 *aging but a later slight decrease*, revealing a reversal (up-down) pattern, and iii) the remaining genes
279 in cluster 1, 2 and 6 *increase in heterogeneity dramatically after the age of 60*. We next asked if these
280 genes have any consistent heterogeneity change pattern in development (Figure S11), but most of
281 the clusters showed no or only weak age-related changes during development. We also analyzed the
282 accompanying changes in mean expression levels for these clusters. Except for cluster 1, which
283 shows a decrease in expression level at around the age of 60 and then shows a dramatic increase, all
284 clusters show a steady scaled mean expression level at around zero, *i.e.* different genes in a cluster
285 show different patterns (Figure S12).
286
287 We further tested the genes showing dramatic heterogeneity increase after the age of 60 (clusters 1,
288 2 and 6) for association with Alzheimer's Disease, as the disease incidence increases after 60 (Evans
289 et al., 1989) as well; however, we found no evidence for such an association (see Figure S13).
290
291



292
293 ***Figure 4.*** *Clusters of genes showing a consistent heterogeneity increase in aging (n = 147). Clustering was*
294 *performed based on patterns of the change in heterogeneity, using the k-means clustering method (see*
295 *Methods). The x- and y-axes show age and heterogeneity levels, respectively. Mean heterogeneity change for*
296 *the genes in each cluster was drawn by spline curves. The colors and line-types of curves specify different brain*
297 *regions and data sources, respectively.*
298
299 **Functional analysis**
300 To examine the functional associations of heterogeneity changes with age, we performed gene set
301 enrichment analysis using KEGG pathways (Kanehisa, Sato, Furumichi, Morishima, & Tanabe, 2019),

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

302 Gene Ontology (GO) categories (Ashburner et al., 2000; The Gene Ontology Consortium, 2019),
303 Disease Ontology (DO) categories (Kibbe et al., 2015), Reactome pathways (Fabregat et al., 2018),
304 transcription factor (TF) targets (TRANSFAC) (Matys et al., 2003), and miRNA targets (MiRTarBase)
305 (Chou et al., 2016). Specifically, we rank-ordered genes based on the number of datasets that show a
306 consistent increase in heterogeneity and asked if the extremes of this distribution are associated with
307 the gene sets that we analyzed. There was no significant enrichment for any of the functional
308 categories and pathways for the consistent changes in development. The significantly enriched KEGG
309 pathways for the genes that become consistently heterogeneous during aging included multiple
310 KEGG pathways known to be relevant for aging, including the longevity regulating pathway,
311 autophagy (Rubinsztein, Mariño, & Kroemer, 2011), mTOR signaling (Johnson, Rabinovitch, &
312 Kaeberlein, 2013) and FoxO signaling (Martins, Lithgow, & Link, 2016) (Figure 5a). Among the
313 pathways with a significant association (listed in Figure 5a), only protein digestion and absorption,
314 primary immunodeficiency, linoleic acid metabolism, and fat digestion and absorption pathways had
315 negative enrichment scores, meaning these pathways were significantly associated with the genes
316 having the least number of datasets showing an increase. However, it is important to note that this
317 does not mean these pathways have a decrease in heterogeneity as the distribution of consistent
318 heterogeneity levels is skewed (Figure 3c, lower panel). We also calculated if the KEGG pathways
319 that we identified are particularly enriched in any of the heterogeneity trajectories we identified.
320 Although we lack the necessary power to test the associations statistically due to small number of
321 genes, we saw that i) group 1, which showed a stable increase in heterogeneity, is associated more
322 with the metabolic pathways and mRNA surveillance pathway, ii) group 2, which showed first an
323 increase and a slight decrease at later ages, is associated with axon guidance, mTOR signaling, and
324 phospholipase D signaling pathways, and iii) group 3, which showed a dramatic increase after age of
325 60, is associated with autophagy, longevity regulating pathway and FoxO signaling pathways. The full
326 list is available as Figure S14.
327



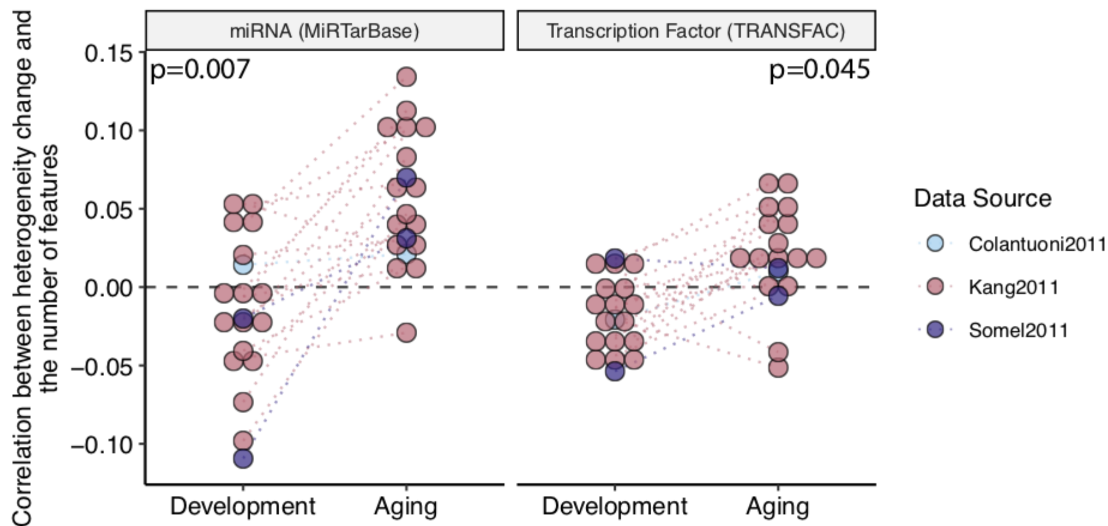
328

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

329 ***Figure 5.*** *Functional analysis of consistent heterogeneity changes. (a) Distribution of consistent heterogeneity*
330 *increase for the significantly enriched KEGG pathways, in development and aging. x- and y-axes show the*
331 *number of datasets with a consistent increase and the density for each significant pathway, respectively. The*
332 *dashed red line shows x = 9.5, which is the middle point for 19 datasets, representing no tendency to increase or*
333 *decrease. Values higher than 9.5, shown with red color, indicate an increase in heterogeneity while values lower*
334 *than 9.5, shown with blue color, indicate a decrease in heterogeneity and the darkness shows the consistency in*
335 *change across datasets. b) The longevity regulating pathway (KEGG Pathway ID: hsa04211), exemplifying the*
336 *distribution of the genes (circles), their heterogeneity across datasets (color – the same color scheme as panel*
337 *(a)), and their relationship in the pathway (edges). More detailed schemes for all significant pathways with the*
338 *gene names are given as SI.*
339
340 The distribution of consistent heterogeneity in development and aging also showed a clear difference.
341 The pathway scheme for the longevity regulating pathway (Figure 5b), colored based on the number
342 of datasets with a consistent increase, shows how particular genes compare between development
343 and aging. The visualizations for all significant pathways, including the gene names are given in the
344 Supplementary Information. Other significantly enriched gene sets, including GO, Reactome, TF and
345 miRNA sets, are included as Tables S4-11. In general, while the consistent heterogeneity changes in
346 development did not show any enrichment (except for miRNAs, see Table S11), we detected a
347 significant enrichment for the genes that become more heterogeneous during aging, with the
348 exception that Disease Ontology terms were not significantly associated with consistent changes in
349 either development or aging. The gene sets included specific categories such as autophagy and
350 synaptic functions as well as broad functional categories such as regulation of transcription and
351 translation processes, cytoskeleton or histone modifications. We also performed GSEA for each
352 dataset separately and confirmed that these pathways show consistent patterns in aging (Figure S15-
353 S19). There were 30 significantly enriched transcription factors, including *EGR* and *FOXO*, and 99
354 miRNAs (see Table S9-10 for the full list). We also asked if the genes that become more
355 heterogenous consistently across datasets are known aging-related genes, using the GenAge Human
356 gene set (Tacutu et al., 2018), but did not find a significant association (Figure S20).
357
358 It has been reported that the total number of distinct regulators of a gene (apart from its specific
359 regulators) is correlated with gene expression noise (Barroso, Puzovic, & Dutheil, 2018). Accordingly,
360 we asked if the total number of transcription factors (TFs) or miRNAs regulating a gene might be
361 related to the heterogeneity change with age (Figure 6). We calculated the correlations between the
362 total number of regulators and the heterogeneity changes and found a mostly positive (18 / 19 for
363 miRNA and 15 / 19 for TFs), and higher correlation between change in heterogeneity and the number
364 of regulators in aging (p = 0.007 for miRNA and p = 0.045 TFs). We further tested the association
365 while controlling for the expression changes in development and aging since regulation of expression
366 changes during development could confound a relationship. However, we found that the pattern is
367 mainly associated with the genes that show a decrease in expression during aging, irrespective of
368 their expression during development (Figure S21).

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

369



370
371 ***Figure 6.*** *Correlation between the change in heterogeneity and number of transcriptional regulators, i.e. miRNA*
372 *and transcription factors. Each point represents a dataset, and the color shows the data source. p-values are*
373 *calculated using a permutation test. The dashed line at y = 0 shows zero correlation.*
374
375 We further tested if genes with a consistent heterogeneity increase in aging are more central in the
376 protein interaction network using STRING database (von Mering et al., 2005). Using multiple cutoffs
377 and repeating the analysis, we observed a higher degree of interactions for the genes with increasing
378 heterogeneity (Figure S22).
379
380 Johnson and Dong et al. previously compiled a list of traits that are age-related and have been
381 sufficiently tested for genome-wide associations (n = 39) (Johnson, Dong, Vijg, & Suh, 2015). Using
382 the genetic associations for GWAS Catalog traits, we tested if there are significantly enriched traits for
383 the consistent changes in heterogeneity during aging (Table S12). Although there was no significant
384 enrichment, all these age-related terms had positive enrichment scores, *i.e.* they all tended to include
385 genes that consistently become more heterogeneous with age during aging.
386
387 Using cell-type specific transcriptome data generated from FACS-sorted cells in mouse brain (Cahoy
388 et al., 2008), we also analyzed if there is an association between genes that become heterogeneous
389 with age and cell-type specific genes, which could be expected if brain cell-type composition
390 progressively varied among individuals with age. Although there was an overlap with
391 oligodendrocytes and myelinated oligodendrocytes, there was no significant enrichment (which could
392 be attributed to low power due to small overlap between aging and cell-type specific expression
393 datasets) (Figure S23).
394

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

395 ## Discussion

396 Aging is characterized by a gradual decrease in the ability to maintain homeostatic processes, which
397 leads to functional decline, age-related diseases, and eventually to death. This age-related
398 deterioration, however, is thought as not a result of expression changes in a few individual genes, but
399 rather as a consequence of an age-related alteration of the whole genome, which could be a result of
400 an accumulation of both epigenetic and genetic errors in a stochastic manner (Enge et al., 2017; Vijg,
401 2004). This stochastic nature of aging impedes the identification of conserved age-related changes in
402 gene expression from a single dataset with a limited number of samples.
403
404 In this study, we examined 19 gene expression datasets compiled from three independent studies to
405 identify the changes in gene expression heterogeneity with age. While all datasets have samples
406 representing the whole lifespan, we separated postnatal development (0 to 20 years of age) and
407 aging (20 to 98 years of age) by the age of 20, as this age is considered to be a turning-point in gene
408 expression trajectories (Dönertaş et al., 2017). We implemented a regression-based method and
409 identified genes showing a consistent change in heterogeneity with age, during development and
410 aging separately. At the single gene level, we did not observe significant age-related heterogeneity
411 change in most of the datasets, possibly due to insufficient statistical power due to small sample sizes
412 and the subtle nature of the phenomenon. We hence took advantage of a meta-analysis approach
413 and focused on consistent signals among datasets, irrespective of their effect sizes and significance.
414 Although this approach fails to capture patterns that are specific to individual brain regions, it identifies
415 genes that would otherwise not pass the significance threshold due to insufficient power.
416 Furthermore, we demonstrated that our method is robust to noise and confounding effects within
417 individual datasets.
418
419 ***Increase in gene expression heterogeneity during aging***

420 By analyzing age-related gene expression changes, we first observed that there are more significant
421 and more similar changes during development than aging. Additionally, genes showing significant
422 change during aging tended to decrease in expression (Figure 1). These results can be explained by
423 the accumulation of stochastic detrimental effects during aging, leading to a decrease in expression
424 levels (Lu et al., 2004). Our initial analysis of gene expression changes suggested a higher
425 heterogeneity between aging datasets.
426
427 We next focused on age-related heterogeneity change between individuals and found a significant
428 increase in age-related heterogeneity during aging, compared to development. Notably, increased
429 heterogeneity is not limited to individual brain regions, but a consistent pattern across different
430 regions during aging. We found that age-related heterogeneity change is more consistent among
431 aging datasets, which may reflect an underlying systemic mechanism. Further, a larger number of

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

432 genes showed more significant heterogeneity changes during aging than in development, and the
433 majority of these genes tended to have more heterogeneous expression.
434
435 It was previously proposed that somatic mutation accumulations (Lodato et al., 2018; Lombard et al.,
436 2005; Lu et al., 2004; Vijg, 2004) and epigenetic regulations (Cheung et al., 2018) might be
437 associated with transcriptome instability. While Enge et al. and Lodato et al. suggested that genome-
438 wide substitutions in single cells are not so common as to influence genome stability and cause
439 transcriptional heterogeneity at the cellular level (Enge et al., 2017; Lodato et al., 2015), epigenetic
440 mechanisms may be relevant. Although we cannot test age-related somatic mutation accumulation
441 and epigenetic regulation in this study, an alternative mechanism might be related to transcriptional
442 regulation, which is considered to be inherently stochastic (Maheshri & O'Shea, 2007). Several
443 studies demonstrated that variation in gene expression is positively correlated with the number of TFs
444 controlling gene's regulation (Barroso et al., 2018). We also found that genes with a higher number of
445 regulators and a decrease in expression during aging become more heterogeneous. Further,
446 significantly enriched TFs include early growth response (*EGF*), known to be regulating the
447 expression of many genes involved in synaptic homeostasis and plasticity, and *FOXO* TFs, which
448 regulate stress resistance, metabolism, cell cycle arrest and apoptosis. Together with these studies,
449 our results support that transcriptional regulation may be associated with age-related heterogeneity
450 increase during aging and may have important functional consequences in brain aging.
451
452 ***Increased heterogeneity is not a result of technical or statistical artifacts***
453 We next confirmed that observed increase in heterogeneity was not a result of low statistical power
454 (Figure S1) or a technical artifact (Figure 3b, S24, S25). Specifically, we tested whether increased
455 heterogeneity during aging can be a result of the mean-variance relationship, but we found no
456 significant effect that can confound our results. In fact, the mean-variance relationship in development
457 and aging showed opposing profiles. We further analyzed this by grouping genes based on their
458 expression in development and aging (Figure S24). The genes that decrease in expression both in
459 development and aging showed the most opposing profiles in terms of the mean-variance
460 relationship, which could suggest that the decrease in development are more coordinated and well-
461 regulated whereas the decrease in aging occurs due to stochastic errors. Another potential
462 confounder is the post-mortem interval (PMI), which is the time between death and sample collection.
463 Since we do not have this data for all datasets we analyzed, we could not account for PMI in our
464 model. However, using the list of genes previously suggested as associated with PMI (Zhu, Wang,
465 Yin, & Yang, 2017), we checked if the consistency among aging datasets could be driven by PMI.
466 Only 2 PMI-associated genes were among the 147 that become consistently heterogeneous, and the
467 distribution also suggested there is no significant relationship (Figure S25). We also confirmed that
468 the increase in heterogeneity is not caused by outlier samples in datasets (Figure S26) or by the
469 confound of sex with age (Figure S27).
470

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

471 ***Microarrays do not bias against identifying age-related heterogeneity change***
472 One important limitation of our study is that we analyze microarray-based data. Since gene
473 expression levels measured by microarray do not reflect an absolute abundance of mRNAs, but
474 rather are relative expression levels, we were only able to examine relative changes in gene
475 expression. A recent study analyzing single-cell RNA sequencing data from the aging *Drosophila*
476 brain identified an age-related decline in total mRNA abundance (Davie et al., 2018). It is also
477 suggested that, in microarray studies, genes with lower expression levels tend to have higher
478 variance (Aris et al., 2004). In this context, whether the change in heterogeneity is a result of the total
479 mRNA decay is an important question. As an attempt to see if the age-related increase in
480 heterogeneity is dependent on the technology used to generate data, we repeated the initial analysis
481 using RNA sequencing data for the human brain, generated by GTEx Consortium (Ardlie et al., 2015)
482 (Figure S28-30). Nine out of thirteen datasets displayed more increase than decrease in
483 heterogeneity during aging, consistent with 18/19 microarray datasets, while the remaining four
484 datasets showed the opposite pattern (BA24, cerebellar hemisphere, cerebellum and substantia
485 nigra). Unlike what we observed for the microarray datasets, the change in expression levels and
486 heterogeneity were strongly positively correlated (Figure S30). Unfortunately, average expression
487 levels and variation levels in RNA sequencing is challenging to disentangle. Thus, the biological
488 relevance of the relationship between the age-related change in expression levels and expression
489 heterogeneity still awaits to be studied through novel experimental and computational approaches.
490 Nevertheless, RNA sequencing analysis also suggests an overall increase in age-related
491 heterogeneity increase.
492
493 Another limitation is related to use of bulk RNA expression datasets, where each value is an average
494 for the tissue. While it is important to note that our results indicate increased heterogeneity between
495 individuals rather than cells, the fact that the brain is composed of different cell types raises the
496 question if increased heterogeneity may be a result of changes in brain cell-type proportions. To
497 explore the association between heterogeneity and cell-type specific genes, we used FACS-sorted
498 cell type specific transcriptome dataset from mouse brain (Cahoy et al., 2008). We only had nine
499 genes that have consistent heterogeneity increase and are specific to one cell-type. Eight out of nine
500 were highly expressed in oligodendrocytes, which is consistent with the results reported in our earlier
501 work (Kedlian et al., 2019). However, we did not observe any significant association between cell-type
502 specific genes and heterogeneity (Figure S23).
503
504 ***Biological processes are associated with increased heterogeneity***
505 Gene set enrichment analysis of the genes with increased heterogeneity with age revealed a set of
506 significantly enriched pathways that are known to modulate aging, including longevity regulating
507 pathway, autophagy, mTOR signaling pathway (Figure 5a). Furthermore, GO terms shared among
508 these genes include some previously identified common pathways in aging and age-related diseases
509 (Figure S16-18). We have also tested if these genes are associated with age-related diseases

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

510 through GWAS, and although not significant, we found a positive association with all age-related traits
511 defined in Johnson and Dong et al.. Overall, these results indicate the effect of heterogeneity on
512 pathways that modulate aging and may reflect the significance of increased heterogeneity in aging.
513 Importantly, we identified genes that are enriched in terms related to neuronal and synaptic functions,
514 such as axon guidance, neuron to neuron synapse, postsynaptic specialization, which may reflect the
515 role of increased heterogeneity in synaptic dysfunction observed in the mammalian brain, which is
516 considered to be a major factor in age-related cognitive decline (Morrison & Baxter, 2012). We also
517 observed genes that become more heterogeneous with age consistently across datasets are more
518 central (*i.e.* have a higher number of interactions) in a protein-protein interaction network (Figure
519 S22). Although this could mean the effect of heterogeneity could be even more critical because it
520 affects hub genes, another explanation is research bias that these genes are studied more than
521 others.
522
523 In summary, by performing a meta-analysis of transcriptome data from diverse brain regions we found
524 a significant increase in gene expression heterogeneity during aging, compared to development.
525 Increased heterogeneity was a consistent pattern among diverse brain regions in aging, while no
526 significant consistency was observed across development datasets. Our results support the view of
527 aging as a result of stochastic molecular alterations, whilst development has a higher degree of gene
528 expression regulation. We also found that genes showing a consistent increase in heterogeneity
529 during aging are involved in pathways important for aging and neuronal function. Therefore, our
530 results demonstrate that increased heterogeneity is one of the characteristics of brain aging and is
531 unlikely to be only driven by the passage of time starting from developmental stages.
532

533 Methods

534

535 Dataset collection

536 *Microarray datasets:* Raw data used in this study were retrieved from the NCBI Gene Expression
537 Omnibus (GEO) from three different sources (Table S1). All three datasets consist of human brain
538 gene expression data generated on microarray platforms. In total, we obtained 1017 samples from
539 298 individuals, spanning the whole lifespan with ages ranging from 0 to 98 years (Figure S1).
540 *RNA sequencing dataset:* We used the transcriptome data generated by the GTEx Consortium (v6p)
541 (Ardlie et al., 2015). We only used the samples with a death circumstance of 1 (violent and fast
542 deaths due to an accident) and 2 (fast death of natural causes) on the Hardy Scale excluding
543 individuals who died of illnesses. As we focus only on the brain, we used all 13 brain tissue data in
544 GTEx. We thus analyzed 623 samples obtained from 99 individuals.
545 *Separating datasets into development and aging datasets:* To differentiate changes in gene
546 expression heterogeneity during aging from those during development, we used the age of 20 to
547 separate pre-adulthood from adulthood. It was shown that the age of 20 corresponds to the first age

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

548 of reproduction in human societies (Walker et al., 2006). Structural changes after the age of 20 in the
549 human brain were previously linked to age-related phenotypes, specifically neuronal shrinkage and a
550 decline in total length of myelinated fibers (Sowell et al., 2004). Earlier studies examining age-related
551 gene expression changes in different brain regions also showed a global change in gene expression
552 patterns after the age of 20 (Colantuoni et al., 2011; Dönertaş et al., 2017; Somel et al., 2010). Thus,
553 consistent with these studies, we separated datasets using the age of 20 into development (0 to 20
554 years of age, $n = 441$) and aging (20 to 98 years of age, $n = 569$).
555

556 **Preprocessing**

557 _Microarray datasets:_ RMA correction (using the 'oligo' library in R) (Carvalho & Irizarry, 2010) and
558 log2 transformation were applied to Somel2011 and Kang2011 datasets. For the Colantuoni2011
559 dataset, as there was no public R package to analyze the raw data, we used the preprocessed data
560 deposited in GEO, which had been loess normalized by the authors. We quantile normalized all
561 datasets using the 'preprocessCore' library in R (Bolstad, 2019). Notably, our analysis focused on
562 consistent patterns across datasets, instead of considering significant changes within individual
563 datasets. Since we don't expect random confounding factors to be shared among datasets, we used
564 quantile normalization to minimize the effects of confounders, and we treated consistent results as
565 potentially a biological signal. We also applied an additional correction procedure for Somel2011
566 datasets, in which there was a batch effect influencing the expression levels, as follows: for each
567 probeset (1) calculate mean expression (M), (2) scale each batch separately (to mean = 0, standard
568 deviation = 1), (3) add M to each value. We excluded outliers given in Table S1, through a visual
569 inspection of the first two principal components for the probeset expression levels (same as in
570 Dönertaş, Fuentealba Valenzuela, Partridge, & Thornton, 2018; Dönertaş et al., 2017). We mapped
571 probeset IDs to Ensembl gene IDs 1) using the Ensembl database, through the 'biomaRt' library
572 (Durinck, Spellman, Birney, & Huber, 2009) in R for the Somel2011 dataset, 2) using the GPL file
573 deposited in GEO for Kang2011, as probeset IDs for this dataset were not complete in Ensembl, and
574 3) using the Entrez gene IDs in the GPL file deposited in GEO for the Colantuoni2011 dataset and
575 converting them into Ensembl gene IDs using the Ensemble database, through the "biomaRt" library
576 in R. Lastly, we scaled expression levels for genes (to mean = 0, standard deviation = 1) using the
577 'scale' function in R. Age values of individuals in each dataset were converted to the fourth root of age
578 (in days) to have a linear relationship between age and expression both in development and aging.
579 However, we repeated the analysis using different age scales and confirmed that the results were
580 quantitatively similar (Figure S3).

581 _RNA sequencing dataset:_ The genes with median RPKM value of 0 were excluded from the dataset.
582 The RPKM values provided in the GTEx data were log2 transformed and quantile-normalized. Similar
583 to the microarray data, we excluded the outliers based on the visual inspection of the first and second
584 principal components (Table S1). In GTEx, ages are given as 10 year intervals. We therefore used
585 the middle point of these age intervals to represent that individual's age.
586

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

587
588 **Age-related expression change**
589 We used linear regression to assess the relationship between age and gene expression. The model
590 used in the analysis is:
591
592 $$(1)\quad Y_i = \beta_{i0} + \beta_{i1} * Age^{1/4} + \varepsilon_i$$
593
594 where Y_i is the scaled log2 expression level for the ith gene, β_{i0} is the intercept, β_{i1} is the slope, and ε_i
595 is the residual. We performed the analysis for each dataset (development and aging datasets
596 separately) and considered the β_1 value as a measure of change in expression. *p*-values obtained
597 from the model were corrected for multiple testing according to Benjamini and Hochberg procedure
598 (Benjamini & Hochberg, 1995) by using 'p.adjust' function in R.
599
600 **Age-related heterogeneity change**
601 In order to quantify the age-related change in gene expression heterogeneity, we calculated
602 Spearman's correlation coefficient (ρ). The correlations were calculated between the absolute values
603 of residuals obtained from equation (1) and the fourth root of individual age. We regarded the
604 absolute values of residuals as a measure of heterogeneity. Therefore, high positive correlation
605 coefficients suggest that heterogeneity increases with age, whereas strong negative correlation
606 implies heterogeneity decreases with age. *p*-values were calculated from the correlation analysis and
607 corrected for multiple testing with the Benjamini and Hochberg method using the 'p.adjust' function in
608 R. To compare heterogeneity changes in aging and development, we employed paired Wilcoxon test
609 ('wilcox.test' in the R 'stats' package) in which we compared median heterogeneity changes in aging
610 and development dataset pairs.
611
612 **Principal Component Analysis**
613 We conducted principal component analysis on both age-related changes in expression (β) and
614 heterogeneity (ρ). We followed a similar procedure for both analyses, in which we used the 'prcomp'
615 function in R. The analysis was performed on a matrix containing β values (for the change in
616 expression level) and ρ values (for the change in heterogeneity), for 11,137 commonly expressed
617 genes for all 38 development and aging datasets. In each dataset, the estimates of expression
618 change (β) or heterogeneity change (ρ) values were scaled for each dataset before calculating
619 principal components.
620
621 **Permutation test**
622 We performed a permutation test, taking into account the non-independence of samples across the
623 Somel2011 and Kang2011 datasets, due to the fact that these datasets include multiple samples from
624 the same individuals for different brain regions. We first randomly permuted ages among individuals,

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

625 not samples, for 1,000 times in each data source, using the 'sample' function in R. Next, we assigned
626 ages of individuals to corresponding samples and calculated age-related expression and
627 heterogeneity change for each dataset, corresponding to different brain regions. For the tests related
628 to the changes in gene expression with age, we used a linear model between gene expression levels
629 and the randomized ages. In contrast, for the tests related to the changes in heterogeneity with age,
630 we measured the correlation between the randomized ages and the absolute value of residuals from
631 the linear model that is between expression levels and non-randomized ages for each gene. In this
632 way, we preserved the relationship between age and expression, and we were able to ensure that our
633 regression model was viable for calculating age-related heterogeneity change. Using expression and
634 heterogeneity change estimates calculated using permuted ages, we tested (a) if the correlation of
635 expression (and heterogeneity) change in aging is higher than in development datasets; (b) if the
636 correlations of expression (and heterogeneity) change in development and in aging datasets are
637 significantly higher than null expectation; (c) if the number of genes showing significant change in
638 expression (and heterogeneity) is significantly higher in aging than in development datasets; (d) if the
639 overall increase in age-related heterogeneity during aging is significantly higher than development; (e)
640 if the observed consistency in heterogeneity increase is significantly different from expected. We also
641 demonstrate that our permutation strategy is more stringent than random permutations in Figure S10,
642 giving the distributions calculated using both dependent permutations and random permutations.
643
644 To test the overall correlation within development or aging datasets for the changes in expression (β)
645 and heterogeneity (ρ), we calculated median correlations among independent three subsets of
646 datasets (one Kang2011, one Somel2011 and the Colantuoni2011 dataset), taking the median value
647 calculated for each possible combination of independent subsets (16 x 2 x 1 = 32 combinations).
648 Using 1,000 permutations of individuals' ages, we generated an expected distribution for the median
649 correlation coefficient for triples and compared these with the observed values, asking how many
650 times we observe a higher value. We used this approach to calculate expected median correlation
651 among development (and aging) datasets, because the number of independent pairwise comparisons
652 are outnumbered by the number of dependent pairwise comparisons, causing low statistical power.
653
654 To further test the significance of the difference between correlations among development and aging
655 datasets, we calculated the median difference in correlations between aging and development
656 datasets for each permutation. We next constructed the null distribution of 1,000 median differences
657 and calculated empirical p-values compering the observed differences with these null distributions.
658 Next, to test the significance of the difference in the number of significantly changing genes between
659 development and aging, we calculated the difference in the number of genes showing significant
660 change between development and aging datasets for each permutation. Empirical p-values were
661 computed according to observed differences. Likewise, to test if the overall increase in age-related
662 heterogeneity during aging is significant compared to development, we computed median differences
663 between median heterogeneity change values of each aging and development dataset, for each

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

664 permutation, followed by an empirical *p*-value calculation to answer if the aging datasets have a
665 higher increase in age-related heterogeneity.
666
667 **Expected heterogeneity consistency**
668 Expected consistency in heterogeneity change was calculated from heterogeneity change values
669 (ρ) measured using permuted ages. For each permutation, we first calculated the total number of
670 genes showing consistent heterogeneity increase for N number of datasets (N = 0, ..., 19). To test if
671 observed consistency significantly differed from the expected, we compared observed consistency
672 values to the distribution of expected numbers, by performing a one-sided test for the consistency in
673 N number of datasets, N = 1, …, 19.
674
675 **Clustering**
676 We used the k-means algorithm ('kmeans' function in R) to cluster genes showing consistent
677 heterogeneity change (n=147) according to their heterogeneity profiles. We first took the subset of the
678 heterogeneity levels (absolute value of the residuals from equation [1]) to include only the genes that
679 show a consistent increase with age and then scaled the heterogeneity levels to the same mean and
680 standard deviation. Since the number of samples in each dataset is different, just running k-means on
681 the combined dataset would not equally represent all datasets. Thus, we first calculated the spline
682 curves for scaled heterogeneity levels for each gene in each dataset (using the 'smooth.spline'
683 function in R, with three degrees of freedom). We interpolated at 11 (the smallest sample size) equally
684 distant age points within each dataset. Then we used the combined interpolated values to run the k-
685 means algorithm with k = 8, a liberal choice, given the total number of genes being 147.
686
687 To test association of the clusters with Alzheimer's Disease, we retrieved overall AD association
688 scores of the 147 consistent genes (*n* = 40) from the Open Targets Platform (Carvalho-Silva et al.,
689 2019).
690
691 **Functional Analysis**
692 We used the "clusterProfiler" package in R to run Gene Set Enrichment Analysis, using Gene
693 Ontology (GO) Biological Process (BP), GO Molecular Function (MF), GO Cellular Compartment
694 (CC), Reactome, Disease Ontology (DO), and KEGG Pathways. We performed GSEA on all gene
695 sets with a size between 5 and 500, and we corrected the resulting *p*-values with the Benjamini and
696 Hochberg correction method. To test if the genes with a consistent increase or decrease in their
697 expression are associated with specific functions, we used the number of datasets with a consistent
698 increase to run GSEA. Since we are running GSEA using number of datasets showing consistency,
699 our data includes many ties, potentially making the ranking ambiguous and non-robust. In order to
700 assess how robust our results are, we ran GSEA 1,000 times on the same data and counted how
701 many times we observed the same set of KEGG pathways as significant (Table S4). The lowest
702 number among the pathways with a significant positive enrichment score was 962 out of 1,000

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

703 (Phospholipase D signaling pathway). Moreover, we repeated the same analysis using the
704 heterogeneity change levels (ρ), instead of using the number of datasets with a consistent change, for
705 each dataset to confirm the gene sets are indeed associated with the increase or decrease in
706 heterogeneity (Figure S15-S19). We visualized the KEGG pathways using 'KEGGgraph' library in R
707 and colored the genes by the number of datasets that show an increase.
708
709 We also performed an enrichment analysis of the transcription factors and miRNA to test if specific
710 TFs or miRNAs regulate the genes that become more heterogeneous consistently. We collected
711 gene-regulator association information using the Harmonizome database (Rouillard et al., 2016),
712 "MiRTarBase microRNA Targets" (12086 genes, 596 miRNAs) and "TRANSFAC Curated
713 Transcription Factor Targets" (13216 genes, 201 TFs) sets. We used the 'fgsea' package in R, which
714 allows GSEA on a custom gene set. We tested the association for each regulator with at least 10 and
715 at most 500 targets. Moreover, we tested if the number of regulators is associated with the change in
716 heterogeneity. We first calculated the correlation between heterogeneity change with age (or the
717 number of datasets with an increase in expression heterogeneity) and the number of TFs or miRNAs
718 regulating that gene, for aging and development separately. We repeated the analysis while
719 accounting for the direction of expression changes in these periods (*i.e.* separating genes into down-
720 down, down-up, up-down, and up-up categories based on their expression in development and aging,
721 Figure S21). To test the difference in the correlations between aging and development, we used
722 1,000 random permutations of the number of TFs. For each permutation, we randomized the number
723 of TFs and calculated the correlation between heterogeneity change (or the number of datasets with
724 an increase in heterogeneity) and the randomized numbers. We then calculated the percentage of
725 datasets where aging has a higher correlation than development. Using the distribution of
726 percentages, we tested if the observed value is expected by chance.
727
728 **Protein-protein interaction network analysis**
729 We downloaded all human protein interaction data from the STRING database (v11) (von Mering et
730 al., 2005). Ensembl Peptide IDs are mapped to Ensembl Gene IDs using the "biomaRt" package in R.
731 Here we aimed to test whether genes showing consistent increase in heterogeneity have a different
732 number of interactors than other genes. For this we calculated the degree distributions for the genes
733 that become consistently more heterogeneous with age and all remaining genes using different
734 cutoffs for interaction confidence scores. In order to calculate the significance of the difference, we i)
735 calculated the number of interactors (degree) for each gene, ii) for 10,000 times, randomly sampled k
736 genes from all interactome data (k = number of genes that become heterogeneous with age across all
737 datasets and have interaction information in STRING database, after filtering for a cutoff), iii)
738 calculated the median of degree for each sample. We then calculated an empirical p-value by asking
739 how many of these 10,000 samples we see a median degree that is equivalent to or higher than our
740 original value. The number of genes and interactions after each cutoff are given in Figure S22.
741

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

742 **Cell-type specificity analysis**

743 Using FACS-sorted cell-type specific transcriptome data from the mouse brain (Cahoy et al., 2008),
744 we checked if there is any overlap between genes that become heterogeneous with age and cell-type
745 specific genes. We downloaded the dataset from the GEO database (GSE9566) and preprocessed it
746 by performing: i) RMA correction using the 'affy' package in R (Gautier, Cope, Bolstad, & Irizarry,
747 2004), ii) log2 transformation, iii) quantile normalization using the 'preprocessCore' package in R
748 (Bolstad, 2019), iv) mapping probeset IDs to first mouse genes, and then human genes. We only
749 included genes that have one to one orthologs in humans, after filtering out probesets that map to
750 multiple genes. We defined cell-type specific genes by calculating the effect size (Cohen's D) for each
751 gene and cell type and identifying genes that have an effect size higher than or equal to 2 as specific
752 to that cell type. At this cutoff, there was no overlap between cell type specific gene lists. To test for
753 association between heterogeneity and cell type specificity, we used the Fisher's exact test using the
754 R 'fisher.test' function.
755

756 **Software**

757 All analysis was performed using R and the code to calculate heterogeneity changes with age is
758 available as an R package 'hetAge', documented at https://mdonertas.github.io/hetAge/. "ggplot2"
759 (Wickham, 2017) and "ggpubr" (Kassambara, 2018) R libraries were used for the visualization.
760

761 **Data availability**

762 Raw data used in this study was downloaded from the GEO database using GSE numbers specified
763 in Table S1. All data generated in this study, *i.e.* changes in expression and heterogeneity with age
764 for each dataset and functional enrichment results are available as Supplementary Tables.
765

766 ## Author Contributions

767 H.M.D. conceived and designed the study with the contributions from M.S., and J.M.T.. U.I. and
768 H.M.D. analyzed the data. U.I. and H.M.D. interpreted the results and wrote the manuscript with the
769 contributions from M.S. and J.M.T. All authors read, revised and approved the final version of this
770 manuscript.
771

772 ## Acknowledgements

773 The authors thank Hamit Izgi, Matias Fuentealba Valenzuela, Dr. Daniel K. Fabian, and Prof Linda
774 Partridge for helpful discussions. H.M.D. is a member of Darwin College, University of Cambridge.
775

776 ## Funding Statement

777 This work is funded by EMBL (H.M.D., J.M.T.) and the Wellcome Trust (098565/Z/12/Z; J.M.T).
778

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

779 **References**

780 Angelidis, I., Simon, L. M., Fernandez, I. E., Strunz, M., Mayr, C. H., Greiffo, F. R., … Schiller, H. B.
781 (2019). An atlas of the aging lung mapped by single cell transcriptomics and deep tissue
782 proteomics. *Nature Communications*, *10*(1), 963. https://doi.org/10.1038/s41467-019-08831-9
783 Ardlie, K. G., Deluca, D. S., Segre, A. V., Sullivan, T. J., Young, T. R., Gelfand, E. T., … Dermitzakis,
784 E. T. (2015). The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation
785 in humans. *Science*, *348*(6235), 648–660. https://doi.org/10.1126/science.1262110
786 Aris, V. M., Cody, M. J., Cheng, J., Dermody, J. J., Soteropoulus, P., Recce, M., & Tolias, P. P.
787 (2004). Noise filtering and nonparametric analysis of microarray data underscores discriminating
788 markers of oral, prostate, lung, ovarian and breast cancer. *BMC Bioinformatics*, *5*.
789 https://doi.org/10.1186/1471-2105-5-185
790 Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., … Sherlock, G. (2000,
791 May). Gene ontology: Tool for the unification of biology. *Nature Genetics*. Nature America Inc.
792 https://doi.org/10.1038/75556
793 Bahar, R., Hartmann, C. H., Rodriguez, K. A., Denny, A. D., Busuttil, R. A., Dollé, M. E. T., … Vijg, J.
794 (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature*,
795 *441*(7096), 1011–1014. https://doi.org/10.1038/nature04844
796 Barroso, G. V., Puzovic, N., & Dutheil, J. Y. (2018). The Evolution of Gene-Specific Transcriptional
797 Noise Is Driven by Selection at the Pathway Level. *Genetics*, *208*(1), 173–189.
798 https://doi.org/10.1534/genetics.117.300467
799 Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful
800 Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*.
801 WileyRoyal Statistical Society. Retrieved from https://www.jstor.org/stable/2346101
802 Bolstad, B. M. (2019). preprocessCore: A collection of pre-processing functions. Retrieved from
803 https://github.com/bmbolstad/preprocessCore
804 Brinkmeyer-Langford, C. L., Guan, J., Ji, G., & Cai, J. J. (2016). Aging Shapes the Population-Mean
805 and -Dispersion of Gene Expression in Human Brains. *Frontiers in Aging Neuroscience*, *8*, 183.
806 https://doi.org/10.3389/fnagi.2016.00183
807 Cahoy, J. D., Emery, B., Kaushal, A., Foo, L. C., Zamanian, J. L., Christopherson, K. S., … Barres, B.
808 A. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new
809 resource for understanding brain development and function. *The Journal of Neuroscience□: The*
810 *Official Journal of the Society for Neuroscience*, *28*(1), 264–278.
811 https://doi.org/10.1523/JNEUROSCI.4178-07.2008
812 Carvalho-Silva, D., Pierleoni, A., Pignatelli, M., Ong, C., Fumis, L., Karamanis, N., … Dunham, I.
813 (2019). Open Targets Platform: new developments and updates two years on. *Nucleic Acids*
814 *Research*, *47*(D1), D1056–D1065. https://doi.org/10.1093/nar/gky1133
815 Carvalho, B. S., & Irizarry, R. A. (2010). A framework for oligonucleotide microarray preprocessing.
816 *Bioinformatics*, *26*(19), 2363–2367. https://doi.org/10.1093/bioinformatics/btq431

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

817 Cheung, P., Vallania, F., Warsinske, H. C., Donato, M., Schaffert, S., Chang, S. E., … Kuo, A. J.
818 (2018). Single-Cell Chromatin Modification Profiling Reveals Increased Epigenetic Variations
819 with Aging. *Cell*, *173*(6), 1385-1397.e14. https://doi.org/10.1016/j.cell.2018.03.079
820 Chou, C.-H., Chang, N.-W., Shrestha, S., Hsu, S.-D., Lin, Y.-L., Lee, W.-H., … Huang, H.-D. (2016).
821 miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database.
822 *Nucleic Acids Research*, *44*(D1), D239–D247. https://doi.org/10.1093/nar/gkv1258
823 Colantuoni, C., Lipska, B. K., Ye, T., Hyde, T. M., Tao, R., Leek, J. T., … Kleinman, J. E. (2011).
824 Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*,
825 *478*(7370), 519–523. https://doi.org/10.1038/nature10524
826 Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, Ł., … Aerts, S. (2018). A
827 Single-Cell Transcriptome Atlas of the Aging Drosophila Brain. *Cell*, *174*(4), 982-998.e20.
828 https://doi.org/10.1016/j.cell.2018.05.057
829 Dönertaş, H. M., Fuentealba Valenzuela, M., Partridge, L., & Thornton, J. M. (2018). Gene
830 expression-based drug repurposing to target aging. *Aging Cell*, *17*(5), e12819.
831 https://doi.org/10.1111/acel.12819
832 Dönertaş, H. M., İzgi, H., Kamacıoğlu, A., He, Z., Khaitovich, P., & Somel, M. (2017). Gene
833 expression reversal toward pre-adult levels in the aging human brain and age-related loss of
834 cellular identity. *Scientific Reports*, *7*(1), 5894. https://doi.org/10.1038/s41598-017-05927-4
835 Dorszewska, J. (2013). Cell biology of normal brain aging: synaptic plasticity–cell death. *Aging*
836 *Clinical and Experimental Research*, *25*(1), 25–34. https://doi.org/10.1007/s40520-013-0004-2
837 Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of
838 genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, *4*(8), 1184–
839 1191. https://doi.org/10.1038/nprot.2009.97
840 Enge, M., Arda, H. E., Mignardi, M., Beausang, J., Bottino, R., Kim, S. K., & Quake, S. R. (2017).
841 Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and
842 Somatic Mutation Patterns. *Cell*, *171*(2), 321-330.e14. https://doi.org/10.1016/j.cell.2017.09.004
843 Evans, D. A., Funkenstein, H. H., Albert, M. S., Scherr, P. A., Cook, N. R., Chown, M. J., … Taylor, J.
844 O. (1989). Prevalence of Alzheimer's disease in a community population of older persons.
845 Higher than previously reported. *JAMA*, *262*(18), 2551–2556. Retrieved from
846 http://www.ncbi.nlm.nih.gov/pubmed/2810583
847 Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., … D'Eustachio, P.
848 (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Research*, *46*(D1), D649–D655.
849 https://doi.org/10.1093/nar/gkx1132
850 Fraga, M. F., Ballestar, E., Paz, M. F., Ropero, S., Setien, F., Ballestar, M. L., … Esteller, M. (2005).
851 Epigenetic differences arise during the lifetime of monozygotic twins. *Proceedings of the*
852 *National Academy of Sciences*, *102*(30), 10604–10609.
853 https://doi.org/10.1073/pnas.0500398102
854 Gautier, L., Cope, L., Bolstad, B. M., & Irizarry, R. A. (2004). affy--analysis of Affymetrix GeneChip
855 data at the probe level. *Bioinformatics*, *20*(3), 307–315.

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

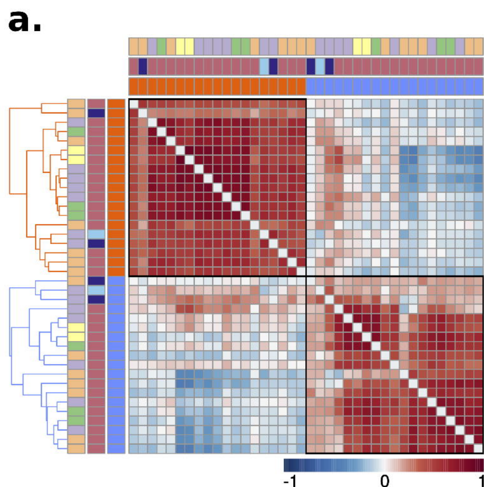
856 https://doi.org/10.1093/bioinformatics/btg405
857 Gorbunova, V., Seluanov, A., Mao, Z., & Hine, C. (2007). Changes in DNA repair during aging.
858 *Nucleic Acids Research*, *35*(22), 7466–7474. https://doi.org/10.1093/nar/gkm756
859 Grillo, F. W., Song, S., Teles-Grilo Ruivo, L. M., Huang, L., Gao, G., Knott, G. W., … De Paola, V.
860 (2013). Increased axonal bouton dynamics in the aging mouse cortex. *Proceedings of the*
861 *National Academy of Sciences*, *110*(16), E1514–E1523.
862 https://doi.org/10.1073/pnas.1218731110
863 Herndon, L. A., Schmeissner, P. J., Dudaronek, J. M., Brown, P. A., Listner, K. M., Sakano, Y., …
864 Driscoll, M. (2002). Stochastic and genetic factors influence tissue-specific decline in ageing C.
865 elegans. *Nature*, *419*(6909), 808–814. https://doi.org/10.1038/nature01135
866 Johnson, S. C., Dong, X., Vijg, J., & Suh, Y. (2015). Genetic evidence for common pathways in
867 human age-related diseases. *Aging Cell*, *14*(5), 809–817. https://doi.org/10.1111/acel.12362
868 Johnson, S. C., Rabinovitch, P. S., & Kaeberlein, M. (2013). mTOR is a key modulator of ageing and
869 age-related disease. *Nature*, *493*(7432), 338–345. https://doi.org/10.1038/nature11861
870 Kanehisa, M., Sato, Y., Furumichi, M., Morishima, K., & Tanabe, M. (2019). New approach for
871 understanding genome variations in KEGG. *Nucleic Acids Research*, *47*(D1), D590–D595.
872 https://doi.org/10.1093/nar/gky962
873 Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., … Sestan, N. (2011). Spatio-temporal
874 transcriptome of the human brain. *Nature*, *478*(7370), 483–489.
875 https://doi.org/10.1038/nature10523
876 Kassambara, A. (2018). ggpubr: “ggplot2” Based Publication Ready Plots. *R Package Version 0.1.8.*
877 Retrieved from https://cran.r-project.org/package=ggpubr
878 Kedlian, V. R., Donertas, H. M., & Thornton, J. M. (2019). The widespread increase in inter-individual
879 variability of gene expression in the human brain with age. *Aging*, *11*(8), 2253–2280.
880 https://doi.org/10.18632/aging.101912
881 Kibbe, W. A., Arze, C., Felix, V., Mitraka, E., Bolton, E., Fu, G., … Schriml, L. M. (2015). Disease
882 Ontology 2015 update: an expanded and updated database of human diseases for linking
883 biomedical knowledge through disease data. *Nucleic Acids Research*, *43*(D1), D1071–D1078.
884 https://doi.org/10.1093/nar/gku1011
885 Lodato, M. A., Rodin, R. E., Bohrson, C. L., Coulter, M. E., Barton, A. R., Kwon, M., … Walsh, C. A.
886 (2018). Aging and neurodegeneration are associated with increased mutations in single human
887 neurons. *Science*, *359*(6375), 555–559. https://doi.org/10.1126/science.aao4426
888 Lodato, M. A., Woodworth, M. B., Lee, S., Evrony, G. D., Mehta, B. K., Karger, A., … Walsh, C. A.
889 (2015). Somatic mutation in single human neurons tracks developmental and transcriptional
890 history. *Science*, *350*(6256), 94–98. https://doi.org/10.1126/science.aab1785
891 Lombard, D. B., Chua, K. F., Mostoslavsky, R., Franco, S., Gostissa, M., & Alt, F. W. (2005). DNA
892 Repair, Genome Stability, and Aging. *Cell*, *120*(4), 497–512.
893 https://doi.org/10.1016/j.cell.2005.01.028
894 Lu, T., Pan, Y., Kao, S.-Y., Li, C., Kohane, I., Chan, J., & Yankner, B. A. (2004). Gene regulation and

bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

895 DNA damage in the ageing human brain. *Nature*, *429*(6994), 883–891.
896 https://doi.org/10.1038/nature02661
897 Maheshri, N., & O'Shea, E. K. (2007). Living with Noisy Genes: How Cells Function Reliably with
898 Inherent Variability in Gene Expression. *Annual Review of Biophysics and Biomolecular*
899 *Structure*, *36*(1), 413–434. https://doi.org/10.1146/annurev.biophys.36.040306.132705
900 Martinez-Jimenez, C. P., Eling, N., Chen, H.-C., Vallejos, C. A., Kolodziejczyk, A. A., Connor, F., …
901 Odom, D. T. (2017). Aging increases cell-to-cell transcriptional variability upon immune
902 stimulation. *Science*, *355*(6332), 1433–1436. https://doi.org/10.1126/science.aah4115
903 Martins, R., Lithgow, G. J., & Link, W. (2016). Long live FOXO: unraveling the role of FOXO proteins
904 in aging and longevity. *Aging Cell*, *15*(2), 196–207. https://doi.org/10.1111/acel.12427
905 Matys, V., Fricke, E., Geffers, R., Gössling, E., Haubrock, M., Hehl, R., … Wingender, E. (2003).
906 TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Research*, *31*(1),
907 374–378. https://doi.org/10.1093/nar/gkg108
908 Miller, J. A., Ding, S.-L., Sunkin, S. M., Smith, K. A., Ng, L., Szafer, A., … Lein, E. S. (2014).
909 Transcriptional landscape of the prenatal human brain. *Nature*, *508*(7495), 199–206.
910 https://doi.org/10.1038/nature13185
911 Morrison, J. H., & Baxter, M. G. (2012). The ageing cortical synapse: hallmarks and implications for
912 cognitive decline. *Nature Reviews. Neuroscience*, *13*(4), 240–250.
913 https://doi.org/10.1038/nrn3200
914 Polleux, F., Ince-Dunn, G., & Ghosh, A. (2007). Transcriptional regulation of vertebrate axon guidance
915 and synapse formation. *Nature Reviews Neuroscience*, *8*(5), 331–340.
916 https://doi.org/10.1038/nrn2118
917 Rouillard, A. D., Gundersen, G. W., Fernandez, N. F., Wang, Z., Monteiro, C. D., McDermott, M. G., &
918 Ma'ayan, A. (2016). The harmonizome: a collection of processed datasets gathered to serve and
919 mine knowledge about genes and proteins. *Database*, *2016*, baw100.
920 https://doi.org/10.1093/database/baw100
921 Rubinsztein, D. C., Mariño, G., & Kroemer, G. (2011). Autophagy and Aging. *Cell*, *146*(5), 682–695.
922 https://doi.org/10.1016/j.cell.2011.07.030
923 Schratt, G. (2009). microRNAs at the synapse. *Nature Reviews Neuroscience*, *10*(12), 842–849.
924 https://doi.org/10.1038/nrn2763
925 Somel, M., Guo, S., Fu, N., Yan, Z., Hu, H. Y., Xu, Y., … Khaitovich, P. (2010). MicroRNA, mRNA,
926 and protein expression link development and aging in human and macaque brain. *Genome*
927 *Research*, *20*(9), 1207–1218. https://doi.org/10.1101/gr.106849.110
928 Somel, M., Khaitovich, P., Bahn, S., Pääbo, S., & Lachmann, M. (2006). Gene expression becomes
929 heterogeneous with age. *Current Biology*, *16*(10), R359–R360.
930 https://doi.org/10.1016/j.cub.2006.04.024
931 Sowell, E. R., Thompson, P. M., & Toga, A. W. (2004). Mapping Changes in the Human Cortex
932 throughout the Span of Life. *The Neuroscientist*, *10*(4), 372–392.
933 https://doi.org/10.1177/1073858404263960

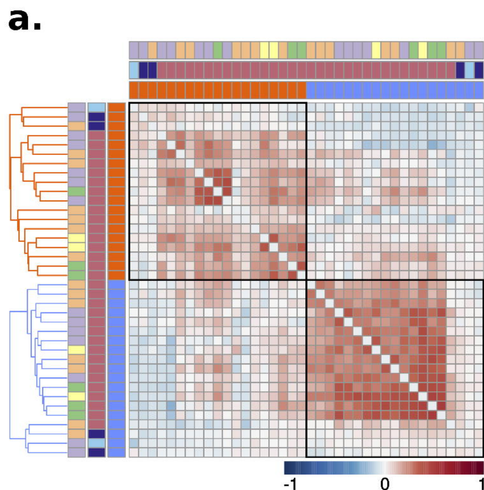
bioRxiv preprint doi: https://doi.org/10.1101/595249; this version posted July 25, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

934 Stefani, G., & Slack, F. J. (2008). Small non-coding RNAs in animal development. *Nature Reviews*
935 *Molecular Cell Biology*, *9*(3), 219–230. https://doi.org/10.1038/nrm2347
936 Tacutu, R., Thornton, D., Johnson, E., Budovsky, A., Barardo, D., Craig, T., … De Magalhães, J. P.
937 (2018). Human Ageing Genomic Resources: New and updated databases. *Nucleic Acids*
938 *Research*, *46*(D1), D1083–D1090. https://doi.org/10.1093/nar/gkx1042
939 Tebbenkamp, A. T. N., Willsey, A. J., State, M. W., & Šestan, N. (2014). The developmental
940 transcriptome of the human brain. *Current Opinion in Neurology*, *27*(2), 149–156.
941 https://doi.org/10.1097/WCO.0000000000000069
942 The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing
943 strong. *Nucleic Acids Research*, *47*(D1), D330–D338. https://doi.org/10.1093/nar/gky1055
944 Vijg, J. (2004). Impact of genome instability on transcription regulation of aging and senescence.
945 *Mechanisms of Ageing and Development*, *125*(10–11), 747–753.
946 https://doi.org/10.1016/j.mad.2004.07.004
947 Viñuela, A., Brown, A. A., Buil, A., Tsai, P.-C., Davies, M. N., Bell, J. T., … Small, K. S. (2018). Age-
948 dependent changes in mean and variance of gene expression across tissues in a twin cohort.
949 *Human Molecular Genetics*, *27*(4), 732–741. https://doi.org/10.1093/hmg/ddx424
950 von Mering, C., Jensen, L. J., Snel, B., Hooper, S. D., Krupp, M., Foglierini, M., … Bork, P. (2005).
951 STRING: known and predicted protein-protein associations, integrated and transferred across
952 organisms. *Nucleic Acids Research*, *33*(Database issue), D433-7.
953 https://doi.org/10.1093/nar/gki005
954 Walker, R., Gurven, M., Hill, K., Migliano, A., Chagnon, N., De Souza, R., … Yamauchi, T. (2006).
955 Growth rates and life histories in twenty-two small-scale societies. *American Journal of Human*
956 *Biology*, *18*(3), 295–311. https://doi.org/10.1002/ajhb.20510
957 Wickham, H. (2017). ggplot2 – Elegant Graphics for Data Analysis. *Journal of Statistical Software*,
958 *77*(April), 3–5. https://doi.org/10.18637/jss.v077.b02
959 Ximerakis, M., Lipnick, S. L., Simmons, S. K., Adiconis, X., Innes, B. T., Dionne, D., … Rubin, L. L.
960 (2018). Single-cell transcriptomics of the aged mouse brain reveals convergent, divergent and
961 unique aging signatures. *BioRxiv*, 440032. https://doi.org/10.1101/440032
962 Zhu, Y., Wang, L., Yin, Y., & Yang, E. (2017). Systematic analysis of gene expression patterns
963 associated with postmortem interval in human tissues. *Scientific Reports*, *7*(1), 5435.
964 https://doi.org/10.1038/s41598-017-05882-0
965

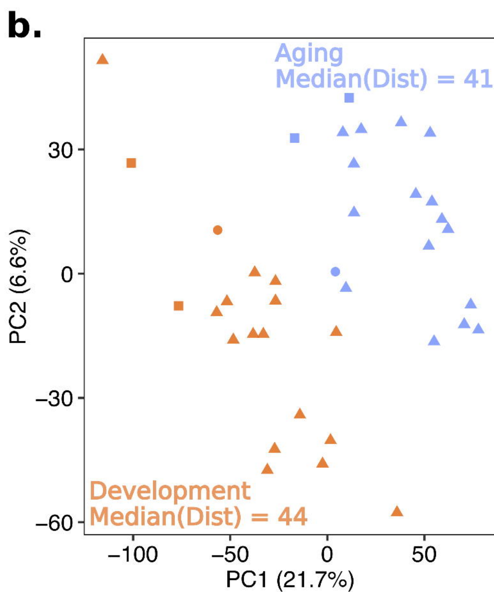

a.
Brain Regions
Frontal Cortex
Parietal Cerebral Wall
Temporal Cerebral Wall
Other
Data Source
Colantuoni2011
Kang2011
Somel2011
Period
Development
Aging
-1
0
1

b.
Development
Median(Dist) = 21
Aging
Median(Dist) = 77
PC2 (13.7%)
PC1 (32%)
50
0
−50
−100
−50
0
50
100
Data Source
Colantuoni2011
Kang2011
Somel2011
Period
Development
Aging
c.
Development
Aging
Somel2011_PFC 1441 1132 1 1
Somel2011_CBC 752 776 0 0
Kang2011_VFC 323 235 126 341
Kang2011_V1C 212 206 29 91
Kang2011_STR 134 82 2 20
Kang2011_STC 183 130 101 152
Kang2011_S1C 864 1101 8 9
Kang2011_OFC 412 295 6 10
Kang2011_MFC 177 116 0 0
Kang2011_MD 80 42 0 0
Kang2011_M1C 301 285 6 7
Kang2011_ITC 368 256 46 91
Kang2011_IPC 652 859 106 169
Kang2011_HIP 25 34 0 1
Kang2011_DFC 447 311 2 11
Kang2011_CBC 619 565 0 0
Kang2011_AMY 95 47 0 0
Kang2011_A1C 238 207 13 14
Colantuoni2011_PFC 2732 2775 857 936
3000 200 10 1
1 10 200 3000
No. of genes (in log scale)
No. of genes (in log scale)
Change in Expression
Decrease
Increase

a.

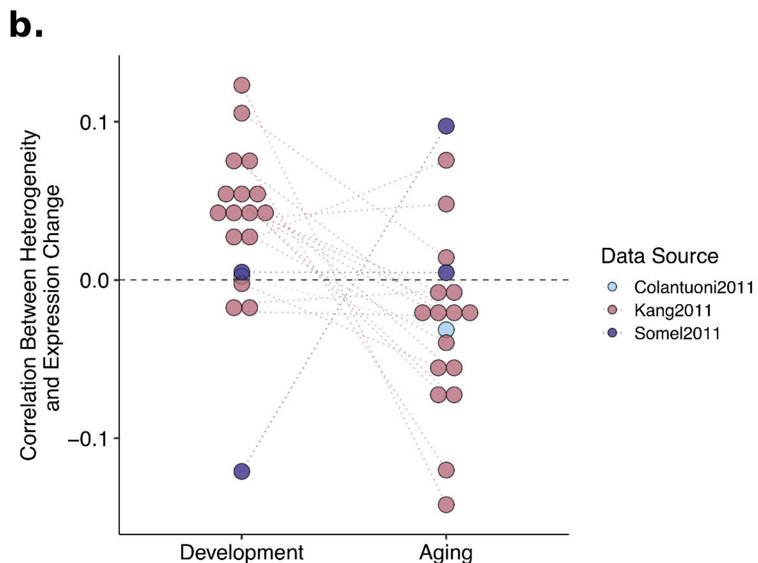


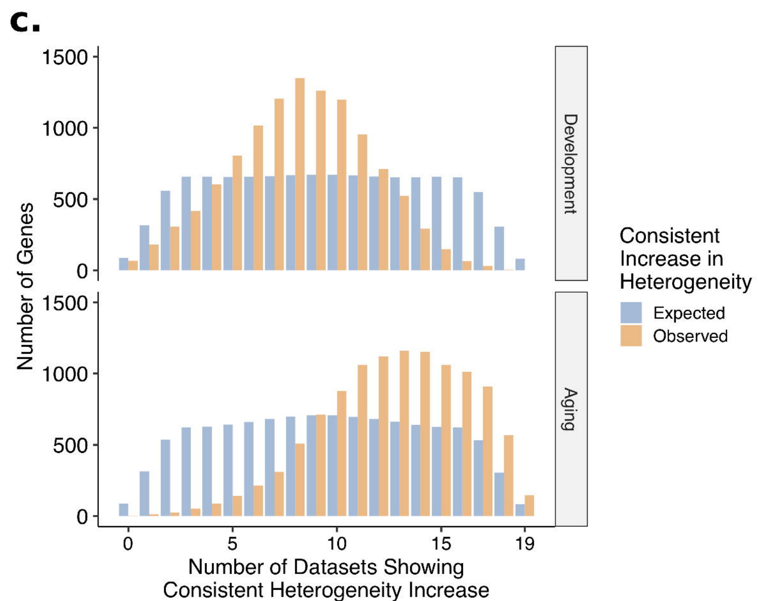
b.



c.

a.
Time Period
Development
Aging
Kang2011_STR
Somel2011_CBC
Kang2011_VFC
Kang2011_CBC
Kang2011_IPC
Kang2011_MD
Kang2011_V1C
Colantuoni2011_PFC
Kang2011_STC
Somel2011_PFC
Kang2011_ITC
Kang2011_OFC
Kang2011_M1C
Kang2011_AMY
Kang2011_HIP
Kang2011_S1C
Kang2011_DFC
Kang2011_A1C
Kang2011_MFC
−0.6
−0.3
0.0
0.3
0.6
Heterogeneity Change
0.0
0.2
0.4
Difference
between medians


b.
Correlation Between Heterogeneity
and Expression Change
0.1
0.0
−0.1
Development
Aging
Data Source
Colantuoni2011
Kang2011
Somel2011


c.
Number of Genes
1500
1000
500
0
Development
Aging
0
5
10
15
19
Number of Datasets Showing
Consistent Heterogeneity Increase
Consistent
Increase in
Heterogeneity
Expected
Observed

Cluster 1
# of genes=15

Cluster 2
# of genes=15

Cluster 3
# of genes=28

Cluster 4
# of genes=24

Cluster 5
# of genes=16

Cluster 6
# of genes=18

Cluster 7
# of genes=16

Cluster 8
# of genes=15

Scaled heterogeneity level

3
2
1
0
−1

25 40 60 85

Age in years

Data Source
Kang2011
Somel2011
Colantuoni2011

Brain Region
Frontal Cortex
Parietal Cerebral Wall
Temporal Cerebral Wall
Other

a.

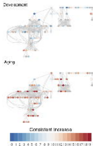
Development

Aging

Notch signaling pathway
Renal cell carcinoma
RNA polymerase
mRNA surveillance pathway
Autophagy – other
AMPK signaling pathway
Ubiquitin mediated proteolysis
Phospholipase D signaling pathway
Longevity regulating pathway – multiple species
Longevity regulating pathway
FoxO signaling pathway
Autophagy – animal
Adherens junction
Apelin signaling pathway
mTOR signaling pathway
Herpes simplex infection
Axon guidance
Protein digestion and absorption
Primary immunodeficiency
Linoleic acid metabolism
Fat digestion and absorption

0 5 10 15 20 5 10 15 20

Number of Datasets with Increase in Heterogeneity



miRNA (MiRTarBase)
Transcription Factor (TRANSFAC)
p=0.007
p=0.045
Correlation between heterogeneity change and the number of features
0.15
0.10
0.05
0.00
-0.05
-0.10
Development
Aging
Development
Aging
Data Source
Colantuoni2011
Kang2011
Somel2011