Aging shapes the population-mean and -dispersion of gene expression in human brains

Candice L. Brinkmeyer-Langford[1]☯, Jinting Guan[1,2]☯, Guoli Ji[2,3]*, James J. Cai[1,4]*

[1]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, Texas, USA.

[2]Department of Automation, Xiamen University, Xiamen, Fujian, China.

[3]Innovation Center for Cell Signaling Network, Xiamen University, Xiamen, Fujian, China.

[4]Interdisciplinary Program in Genetics, Texas A&M University, College Station, Texas, USA.

☯ These authors contributed equally to this work.

* Correspondence to: glji@xmu.edu.cn (GJ); jcai@tamu.edu (JJC)

1

Abstract

Human aging is associated with cognitive decline and an increased risk of neurodegenerative disease. Our objective for this study was to evaluate potential relationships between age and variation in gene expression across different regions of the brain. We analyzed the Genotype-Tissue Expression (GTEx) data from 54 and 101 tissue samples across 13 brain regions in post-mortem donors of European descent aged between 20 and 70 years at death. After accounting for the effects of covariates and hidden confounding factors, we identified 1,446 protein-coding genes whose expression in one or more brain regions is correlated with chronological age at a false discovery rate of 5%. These genes are involved in various biological processes including apoptosis, mRNA splicing, amino acid biosynthesis, and neurotransmitter transport. The distribution of these genes among brain regions is uneven, suggesting variable regional responses to aging. We also found that the aging response of many genes, e.g., *TP37* and *C1QA*, depends on individuals' genotypic backgrounds. Finally, using dispersion-specific analysis, we identified genes such as *IL7R*, *MS4A4E*, and *TERF1*/*TERF2* whose expressions are differentially dispersed by aging, i.e., variances differ between age groups. Our results demonstrate that age-related gene expression is brain region-specific, genotype-dependent, and associated with both mean and dispersion changes. Our findings provide a foundation for more sophisticated gene expression modeling in the studies of age-related neurodegenerative diseases.

Introduction

Aging is a natural process bestowed upon those fortunate enough to participate in the journey, and the progression of age has profound impacts on physical and mental health. The mechanisms underlying age-related cognitive decline and increased risk of neurodegenerative disease remain unclear, though both decline and disease are universally common; therefore, it is critically important to understand the effects of aging on the human brain. One way to approach this goal is to detect the gene expression changes in the human brain during the aging process (Lu et al., 2004). Brain transcriptomic studies hold promise for better understanding the role of aging in both normal brain activity and the development of neurodegenerative disease. The advent of high-throughput sequencing has allowed the study of genome-wide patterns of change in gene expression associated with aging.

In recent years a number of studies on age-related gene expression have been published, such as (Glass et al., 2013;Peters et al., 2015;Sood et al., 2015;Yang et al., 2015), though some have either overlooked the central nervous system or focused on the brain *in toto* for comparisons with other organs and tissues. Meanwhile, mounting evidence shows that the human brain is functionally heterogeneous, with different sub-regions showing distinct functions, cell-type compositions, and gene expression patterns (Fraser et al., 2005;Oldham et al., 2008). In fact, the development and functions of separate anatomical regions of the brain are guided by specific, and often independent, networks of gene expression, e.g., (Kang et al., 2011;Miller et al., 2014;Tebbenkamp et al., 2014). Not surprisingly, therefore, different regions of the brain differ in their susceptibilities to diseases. For example, the hippocampus is affected in Alzheimer's disease, while Parkinson's disease affects the substantia nigra (Hyman et al., 1984;Jellinger, 1986). The striatum of the basal ganglia is a primary region affected in Huntington's disease (Vonsattel et al., 1985). Neurodegenerative diseases tend to be unique to the affected brain sub-regions (Graveland et al., 1985;Fearnley and Lees, 1991;Francis et al., 1999;Neumann et al., 2006;Yao et al., 2015).

In the present study, we focused on the differential gene expression associated with age in multiple brain regions. Rather than evaluating only single genes, we also applied factor analysis (Anand Brown et al., 2015) to identify functional gene sets that are associated with age. More important, our analysis was directed to gene expression dispersion (e.g., variance in gene expression across samples or other measures of dispersion) as a metric to reveal a new model of age-related gene expression patterns. Several studies in humans and model organisms have suggested that age may influence the level of phenotypic dispersion of the population. Thus, to

3

achieve a comprehensive picture of brain aging, we included the dispersion-specific analyses of gene expression with age.

Methods

GTEx brain tissues and expression data

The Genotype-Tissue Expression (GTEx) project was established to determine how genetic variation affects normal gene expression in human tissues, ultimately to inform the study of human diseases (GTEx_Consortium, 2015). The project has collected multiple different human tissues from each of hundreds of donors to isolate nucleic acids from the tissues and perform genotyping, gene expression profiling, whole genome sequencing, and RNA sequencing (RNA-seq) analyses. Among these many tissues, there is a plethora of samples from a handful of sub-regions of the brain, from which expression data sets were generated for the GTEx project and used in the present study.

The expression data (v6, October 2015 release) for brain specimens of post-mortem donors were obtained from the GTEx portal website (http://www.gtexportal.org/). The data was generated using RNA-seq with tissues initially sampled from two brain regions: cerebellum and cortex, preserved using the PAXgene tissue preservation system (Groelz et al., 2013), and with tissues subsequently sampled from frozen brains in following regions: amygdala, anterior cingulate cortex (BA24), caudate (basal ganglia), cerebellar hemisphere, frontal cortex (BA9), hippocampus, hypothalamus, nucleus accumbens (basal ganglia), putamen (basal ganglia), spinal cord (cervical c-1), and substantia nigra (Carithers et al., 2015;GTEx_Consortium, 2015). From the downloaded data, we extracted the whole-gene level RPKM (Reads Per Kilobase of transcript per Million mapped reads) values for protein-coding genes. The data for different brain regions was quantile normalized, and log2 transformed, separately. For each region, 10% lowly expressed genes were excluded from data analysis based on their mean expression level across samples. The donor's information of gender, body mass index (BMI), and sample's ischemic time were also downloaded. Tissue samples from donors of non-European ancestry were excluded from the subsequent data analyses. The number of remaining samples of 13 brain regions ranged between 54 and 101 (**Table 1**).

Accounting for confounding factors using PEER algorithm

Prior to the regression analysis, we used a two-step approach based on the PEER algorithm (Stegle et al., 2012) to control for known covariates as well as hidden data structures in the GTEx expression data. For each region, PEER was first used to discover patterns of common

4

variation across the entire data set and create up to 15 assumed global hidden factors. In doing so, the known covariates, including the donors' age, gender and BMI for all samples from the 13 regions, were included in the PEER models. Also, for samples from (5) cerebellum (PAXgene) and (6) cortex (PAXgene), ischemic time was included as one additional covariate. Note that, at this step, the age of donors was included to enable the PEER to discover correlated patterns across global structured data (O. Stegle, personal communication, November 11, 2015). Next, the correlation between each of the 15 constructed factors and age was tested with a data set of each region. The factor(s) showing a Pearson's correlation test *P*-value smaller than 0.05 were excluded. The remaining factors (denoted *PC_k*, where $1 \leq k \leq N$ and *N* is the number of factors), along with non-age covariates (i.e., known covariates excluding age), were used as a new set of covariates in the regression analysis. Furthermore, in the pathway-based factor analysis (described below), the remaining factors and non-age covariates were supplied to PEER as a new set of covariates and were regressed out. In this way, the effects of all known covariates other than age and hidden data structures, which could potentially confound the subsequent analyses, were removed. The residual values of the regression were used as a new, corrected gene expression data in the subsequent analyses.

Analysis of gene expression with age using linear regression model

For each region, we modeled gene expression using the following linear regression model:

$$Y_i = \mu + Age_i\alpha + Sex_i\beta + BMI_i\gamma + \sum_{k=1}^{N} PC_{ki}\delta + \epsilon_i$$

where Y_i is the expression level of a given gene in sample *i*; *Age_i*, *Sex_i* and *BMI_i* are the age, sex and BMI of sample *i* with regression coefficients *α*, *β* and *γ*, respectively; PC_{ki} ($1 \leq k \leq N$) is the value of the *k*-th hidden factors for the *i*-th sample with regression coefficient *δ*; *N* is the total number of factors uncorrelated with age; ϵ_i is the error term, and *μ* is the regression intercept.

We fitted the model in Matlab with the `fitlm` function in the Statistics toolbox. For each gene, a least square approach was used to estimate the regression coefficients. If *α* was significantly deviated from 0, the gene was considered to be age-associated. A gene was considered up-regulated with age if $\alpha > 0$ and down-regulated if $\alpha < 0$.

Throughout this study, the GO term enrichment analysis was carried out using the DAVID Bioinformatics Resource server (Dennis et al., 2003). The FDR adjustment on the *P*-values was made using the Benjamini–Hochberg procedure (Benjamini and Hochberg, 1995).

Pathway-based factor analysis of gene expression associated with age

The rationale behind pathway-based factor analysis is that a statistical factor analysis (e.g., PEER) can not only remove noise components from high-dimensional data but also derive factors summarizing pathway expression to analyze the relationships between expression and aging (Anand Brown et al., 2015). We used the pathway-based factor analysis to analyze the correlation between age and gene expression of GO-term defined gene sets. We first applied PEER to the whole gene expression matrix for each brain region to regress out global factors. The residual expression levels were treated as new expression data sets; for a given GO-term gene set, PEER was used to construct factors. The constructed factors on the gene sets were taken as concise summaries of common expression variation across each set. These factor values were considered as phenotypes and referred to as phenotype factors. Subsequently, by looking for associations between these new phenotype factors and age, we discovered groups of functionally related genes with a common response to aging.

Detecting effect of genotype-by-age interaction on gene expression

To investigate the genotype-by-age interaction contribution to gene expression, we included the genotype-by-age interaction term to the linear regression model described above. As a contributing factor to the gene expression variance, the significance of this interaction term was assessed for each gene after the model was fitted. Sample donors' genotype data was downloaded from dbGaP under accession number phs000424.v6.p1 (October 2015). At each polymorphic site, an individual's genotype was denoted with 0, 1, or 2 based on the number of non-reference alleles, respectively. SNPs with minor allele frequency greater than 15% were included for the test. To keep the overall computing time feasible, we randomly selected 2,000 genes genome-wide and also included only SNPs with minor allele frequency greater than 15% in the analysis, which was run on the high-performance computing cluster of Texas A&M Institute for Genome Sciences and Society (TIGSS).

Test for expression heteroscedasticity between age groups

To compare the level of gene expression dispersion between age groups, we used Levene's tests. The test examines if the gene expression levels of different age groups have equal deviations from the group means. Let x_{kj} be a set of $j=1,…,n_k$ observations in each of $k=1,…,g$ age groups. Levene's test statistic is the ANOVA F-ratio comparing the g groups, calculated on the absolute deviations $z_{kj} = \left| x_{kj} - \bar{x}_{k\cdot} \right|$, where $\bar{x}_{k\cdot} = \frac{1}{n_k} \sum_{j=1}^{n_k} x_{kj}$ is the group means. To extend the analytical framework to muliple genes, we used the Mahalanobis distance (MD)-based

generalization of Levene's test (Anderson, 2006). A robust version of MD was used to quantify the distance from individual sample i to the multivariate centroid of all samples: $MD(x_{i.}, x_c) = \sqrt{(x_{i.} - x_c)^T \Psi^{-1}(x_{i.} - x_c)}$, where x_i is the vector of expression of genes in sample i; x_c is the location estimator based on the minimum covariance determinant (Rousseeuw and Van Driessen, 1999); and ψ is the scattering estimator. Let MD_{kj} be a set of $j=1,…,n_k$ observations in each of $k=1,…,g$ age groups, Levene's test literally performs ANOVA on MD_{kj}, given the absolute deviation $z_{kj} = \left| MD_{kj} - \overline{MD}_{k.} \right|$ with the group means $\overline{MD}_{k.} = \frac{1}{n_k} \sum_{j=1}^{n_k} MD_{kj}$.

Results

Identification of brain region-specific age-related genes

The data of transcriptomic profiles for brain tissues of 169 European ancestry donors aged 20 – 70 years at death (GTEx_Consortium, 2015) was downloaded from the GTEx portal website. The tissue samples were collected from 13 regions (or subareas) of the human brain, namely (1) amygdala, (2) anterior cingulate cortex [Brodmann area 24 (BA24)], (3) caudate (basal ganglia), (4) cerebellar hemisphere, (5) cerebellum (PAXgene), (6) cortex (PAXgene), (7) frontal cortex (BA9), (8) hippocampus, (9) hypothalamus, (10) nucleus accumbens (basal ganglia), (11) putamen (basal ganglia), (12) spinal cord (cervical c-1), and (13) substantia nigra. We excluded samples derived from non-European donors. The final data matrices, including 54 – 101 samples all from European donors, were normalized separately by different brain regions (**Methods**).

We used linear regression models, controlling for covariates and hidden confounding factors (**Methods**), to identify genes whose expression is correlated or anti-correlated with chronological age. **Table 1** shows the numbers of age-related genes in the 13 brain regions at the false discovery rate (FDR) of 1%, 5%, and 10%. The number of genes identified in most regions increases with the relaxation of FDR cutoff except in (2) anterior cingulate cortex and (12) spinal cord where no genes were identified. At FDR of 5%, 1,446 distinct age-related genes across all regions were identified (**Supplementary Table S1**). Of these, 155 were found in more than one region of the brain: seven were identified in four, 21 in three, and 127 in two brain regions. For each of these "multi-hit" genes, the directions of expression response to aging were the same in the different brain regions where the gene was identified. For comparison, in a previous study, using microarray data from ten regions of 100 post-mortem brains aged from 16 to 83 years, Glass et al. (2013) identified 14 age-related genes. Out of the 14 genes, six (*HSD11B1*, *MS4A6A*, *MT1G*, *PTPN3*, *SLC7A5*, and *WWC2*) are among our 5% FDR age-

related genes, showing consistent directions of expression response to aging. In another previous study, Lu et al. (2004) compared frontal cortical samples from young and old adult individuals and identified 416 age-related genes whose expression differs by at least 1.5-fold. Out of the 416 genes, 61 (14.7%) are among our 5% FDR age-related genes.

Table 1 also shows that some regions demonstrate significantly more age-related genes than others, suggesting distinct brain regions might have different levels of sensitivity or responsiveness to aging. To show that such regional specificities are not completely due to the difference in the number of samples from different brain regions used in the analysis, we repeated the identification of age-related genes by randomly subsampling samples to 54 (the minimal sample size) for all regions. We found that the number of identified genes decreased substantially as the sample size decreases, but the differences in numbers of identified genes between regions largely remained (**Supplementary Table S2**).

The numbers of age-related genes identified show a significant discrepancy between (4) cerebellar hemisphere and (5) cerebellum (PAXgene), which is unexpected because tissue samples of (4) and (5) were essentially from the same region of the cerebellum. Similarly, it is unexpected to see a great discrepancy in the numbers of age-related genes identified between (6) cortex (PAXgene) and (7) frontal cortex (BA9), because (6) and (7) were sampled from the same region of cortex. Indeed, clustering analysis based on the Euclidean distance between gene expression profiles confirmed that (4) and (5), as well as (6) and (7), are more similar to each other, respectively, than to other brain regions (**Supplementary Fig. S1**). We consider that the markedly fewer genes identified in (4) than (5), and in (7) than (6), may be attributed to whether or not the samples were subject to frozen storage before RNA-seq was performed. Among all GTEx brain specimens, only (5) and (6) were initially sampled "on site" from the post-mortem donors, while the rest were subsequently resampled after the brains were frozen and stored (Carithers et al., 2015;GTEx_Consortium, 2015). Thus, it is likely that the frozen-thaw cycle introduced extra expression variability to the samples [e.g., in (4) and (7)], resulting in the identification of fewer genes. To illustrate this further, we sought to examine the cross-region correlation between genes' responsiveness to aging. We used each gene's *P*-value against age in the linear regression model as the measure of the gene's responsiveness to aging. We ranked genes by their *P*-values and then compared the ranks of genes across regions. If the correlations between (4) and (5) and between (6) and (7) were high, then we considered that the discrepancies in age-related gene numbers between (4) and (5), and between (6) and (7), were simply due to the effect of freezing on the statistical power of age-related gene detection,

8

rather than on the gene expression regulation. **Fig. 1** shows the correlation matrix with Spearman correlation coefficient (SCC) between regions. Firstly, we found that the correlation (i.e., the similarity in gene rank) between anatomically closely related regions is higher. For example, the SCC between (3) caudate and (11) putamen, which both belong to basal ganglia, is the highest among all region pairs. Intriguingly, the second highest is between (4) and (5), despite of the considerable discrepancy in identified genes between the frozen and unfrozen cerebellar samples, reinforcing the point that samples from the same brain region are indeed more similar to each other with respect to the genes' responsiveness to aging. Likewise, (6), (7) and (2) anterior cingulate cortex are more correlated with each other than with other regions. Thus, the human brain appears to have different aging patterns in the cerebellum, cortex, and basal ganglia [including (3) caudate, (10) nucleus accumbens, and (11) putamen] (**Fig. 1**). These results are consistent with the findings of a previous microarray-based gene expression study (Fraser et al., 2005).

Analysis of gene expression pathway factors associated with age

We set out to detect gene sets, in addition to single genes, with expression associated with age. We adopted the pathway-based factor analysis (Anand Brown et al., 2015) and applied it to 14,825 functional gene sets defined by gene ontology (GO) terms (**Methods**). As a result, 239 highly significant gene sets across the 13 brain regions were identified ($P < 0.05$, corrected using Bonferroni procedure for the total number of tested gene sets) (**Supplementary Table S3**). The related GO terms included: *neurogenesis* (GO:0022008), *neuron projection* (GO:0043005), *memory* (GO:0007613), and *regulation of synaptic plasticity* (GO:0048167). To obtain a broader functional overview of gene sets, we used the clustering method implemented in REVIGO (Supek et al., 2011) to summarize as many as 5,787 GO terms associated with age-related genes at 5% FDR significance. With REVIGO, these GO terms were evaluated against each other and clustered based on their context similarity. The TreeMap plots for the clusters were then generated, showing that the function of age-related gene sets points to a large collection of biological processes (BP) (**Fig. 2**) and molecular functions (MF) (**Supplementary Fig. S2**). For example, the top-level BP GO term clusters are represented by the terms apoptotic signaling pathway, aging, spliceosomal complex assembly, glutathione derivative biosynthesis, neurotransmitter transport, vitamin metabolism, reactive oxygen species metabolism, methylation, establishment or maintenance of cell polarity, and viral process (**Fig. 2**). The top-level MF GO term are listed in **Supplementary Fig. S2**. The largest clusters represented include growth factor binding, peptidase activity, phosphotransferase activity

(alcohol group as acceptor), copper ion binding, symporter activity, heme binding, virus receptor activity, poly(A) RNA binding, and beta-amyloid binding (**Supplementary Fig. S2**).

Genotype-by-age interactions

We evaluated the interactions between genotype and age to assess how genetic background influences gene expression in the brain at different ages (**Methods**). These interactions could be used to develop strategies to predict individuals' risks for specific conditions based on the associations between each individual's genotype and expression changes anticipated for genes of interest. Other studies have identified age-related eQTL such as those connected with longevity (Walter et al., 2011;Erikson et al., 2016), AD [(Proitsi et al., 2014;Zhu et al., 2014); see also (Guerreiro et al., 2010) though this study did not identify any significant eQTL], and neurological conditions such as PD (Hernandez et al., 2012) and hippocampal sclerosis of aging (Nelson et al., 2015). In this study, we detected a number of interactions with high significance (nominal $P < 10^{-5}$). A comprehensive list of SNPs and genes is provided (**Supplementary Table S4**), albeit none of these interactions survived multiple testing corrections due to the sheer large number of tests performed. We found it intriguing that certain genotypes seem to be more susceptible to the effects of aging on the expression of functionally significant genes. For example, genotypes of SNP rs55675298 can have different effects on expression of tumor protein p73 gene, *TP73* (**Fig. 3A**). For individuals with the GG genotype, there is an age-associated increase in *TP73* expression; individuals with GT or TT genotypes do not experience this increase, which could have profound health implications based on the potential roles of TP73 in conditions related to aging. TP73 is a member of the p53 transcription factor family and is located in a region that is frequently deleted in tumors, particularly neuroblastomas. Furthermore, *TP73* has been found to be critical for normal neuronal development and survival, making it a potential candidate gene for susceptibility to Alzheimer's disease (AD) (Pozniak et al., 2000;Yang et al., 2000;Pozniak et al., 2002;Li et al., 2004;Wetzel et al., 2008). Another example of a relationship between SNP genotype and age-related gene expression involves *C1QA* and the SNP rs72788737 (**Fig. 3B**). Here again, the GG genotype seems to confer increased expression with age while GT/TT genotypes are not correlated with increased expression. Normal aging is associated with an increase in C1q protein (encoded by *C1QA*), particularly in certain regions of the brain that are especially prone to degenerative diseases related to aging (Stephan et al., 2013). C1q contributes to an aging-related decrease in the regenerative capacity of certain tissues (Naito et al., 2012). Less C1q, on the other hand, may

confer some protection against synapse loss and aging-related dysfunction of the hippocampus, e.g., (Stephan et al., 2013;Hong et al., 2016).

Aging affects the population-level dispersion of gene expression in brain

Next, we focused on the identification of differentially variable (DV) genes. These genes show differences in the degree of dispersion in expression levels between age groups. Using Levene's test, we identified 970 DV genes across brain regions (including 848 distinct genes) showing a significant difference in expression variance between young (20–60 years) and old (61–70 years) individuals at the significance level of FDR of 5% (**Supplementary Table S5**). About half of these genes show an increased expression variance in the old while the other half show a decreased variance. Across brain regions, the distribution of DV genes is more balanced (**Supplementary Table S6**), compared to the distribution of age-related genes. The region (6) cortex contains 108 DV genes, which is the most; while the (12) spinal cord contains as little as 40, which is the least. The top GO-term clusters for these DV genes include those related to sensory perception, peptide receptor activity, chemotaxis, peptidase inhibitor activity, and neurotransmitter binding (**Table 2**). **Fig. 4A, B** show two examples of DV genes—*IL7R* and *MS4A4E*. The expression dispersion of *IL7R* in the hippocampus is more pronounced in old than young adults (**Fig. 4A**). This gene is known for its possible role as a determinant of the rate of aging (Passtoors et al., 2015). In the other example, the expression dispersion of *MS4A4E* in the hippocampus also increases with age (**Fig. 4B**). This gene, as a member of the membrane-spanning four domains subfamily A gene cluster, plays a role in embryogenesis, oncogenesis, and the development of AD (Liang et al., 2001;Karagiannis et al., 2003;Hollingworth et al., 2011;Naj et al., 2011).

Furthermore, we expanded the utilization of Levene's test, coupled with a distance measure, to a multivariate setting (Anderson, 2006) to identify age-related DV gene sets, i.e., sets of genes with significant differential expression dispersion between young and old age groups (**Methods**). Eight GO term-defined gene sets (seven distinct gene contents) were identified at the 5% FDR significance level in three brain regions (**Table 3**). These include a set of two genes, *TERF1* and *TERF2*, with the function of age-dependent telomere shortening. Lin et al. (2014) showed that TERF1 and TERF2 use different mechanisms to find telomeric DNA but share a novel mechanism to search for protein partners at telomeres. The deviation of expression profiles of the two genes from individual samples to the population mean centroid was measured with Mahalanobis distance (MD) (**Methods**). Compared to old specimens, young samples show an

increased level of scattering in their *TERF1-TERF2* expression, indicated by the higher level of MD (**Fig. 4C**).

Discussion

Using the GTEx data, we assessed the associations between gene expression and chronological age in different neuroanatomical regions of the human brain. The main findings of our work include: (1) the gene expression responsiveness to aging in various brain regions varies widely, and (2) the gene expression dispersion is a biologically relevant parameter for characterizing the age-related expression alteration.

Transcriptomic assays of the GTEx project generate high-dimensional structured data sets in which there are correlated patterns across large numbers of genes. Some of these are due to the known technical or biological effects, which can be removed by fitting them as covariates. However, even after this, there is typically substantial structural correlation that can potentially confound the subsequent analyses (Leek and Storey, 2007;Parts et al., 2011). Therefore, correcting hidden confounding factors along with covariates is indispensable in revealing the true relationship between gene expression change and the effect under consideration—which, in our case, is aging. Thus, we have carefully controlled for the structural correlations in different brain regions by inferring the hidden confounding factors using the method of factor analysis (Stegle et al., 2012) and then regressing them out. As a result of a rigorous control of input data, we detected a significant number of age-related genes (1,446 distinct genes at the 5% FDR level) across brain regions, which is more than what have previously reported elsewhere (Lu et al., 2004;Glass et al., 2013).

Aging of the brain occurs in a region-specific manner

Our analysis of age-related genes specifically focused on each sub-region of the human brain. We detected the most genes with significant age associations in the cerebellum, which plays an important role in adapting and fine-tuning motor programs to make accurate movements, as well as the cortex, which plays a major role in many complex brain functions such as memory and awareness.

At this stage, we were aware of the effect of freezing—the frozen storage seemed to have a profound impact on the age-related gene detection. With the expression data from cerebellum and cortex samples subjected to frozen storage, we were unable to detect as many age-related genes as we identified using data from unfrozen "fresh" cerebellum and cortex samples. That is

to say, the linear regression-based method for detecting age-related genes was unpowered when applied to the frozen brain samples.

Nevertheless, samples of the majority of brain regions (11 out of the 13) analyzed in the present study were collected from frozen brains. We considered them to be processed in a uniformly consistent manner and thus the results generated from these brain regions are comparable to each other. The relative abundance of age-related genes detected in these regions suggests that different regions have different age-related gene expression changes as a result. To the best of our knowledge, this is the first time that expression data was analyzed for so many brain regions in large numbers of samples processed similarly across locations and times for a single study. Overall, our results support the idea that in the human brain there are measurable patterns of gene expression changes associated with age, and these patterns are distinct from one region of the brain to another. Given that the effect of freezing tends to weaken the overall differential expression signal, our results of the number of age-related genes derived from frozen samples of the 11 regions should be considered as a lower bound of the real number of age-related genes.

Pathway-based factor analysis identifies functional gene sets related to age

We adopted a newly developed pathway-based factor analysis (Anand Brown et al., 2015) to identify age-related gene sets. The analysis is a two-step approach. The factor analysis method, implemented in PEER (Stegle et al., 2012), was first used to discover patterns of common variation across the entire data set. Then newly derived factors summarizing expression of pathways or gene sets were used to analyze the relationships between expression and aging. This analysis allowed us to identify functionally related genes with a common response to aging. Our results support that aging is associated with a large number of biological processes and molecular functions. Many of these associations are consistent with our current knowledge. For example, aging is related to chromatin modulation (Feser and Tyler, 2011), apoptotic signaling pathway (Harman, 1992), glutathione and vitamin (Nuttall et al., 1998), oxidation-reduction process (Berlett and Stadtman, 1997), spliceosome complex assembly (Rodriguez et al., 2015), and neurotransmitter transport (Segovia et al., 2001).

In addition to focusing on linear patterns of gene expression change with chronological age, we also observed extensive interactions between the aging effect and the influence of background regulatory variants. These findings are important for the in-depth analysis of aging effects from the perspective of personal genomics.

Evidence for the aging effect on gene expression dispersion

There is a substantial body of evidence for the impacts of aging on gene expression dispersion. In mice, for example, Southworth et al. (2009) observed a decrease in the correlated expression between normally co-expressed genes, which was associated with aging. Also in mice, Bahar et al. (2006) demonstrated an age-related increase in cell-to-cell gene expression variation in the heart [but see (Warren et al., 2007)]. Data from both humans and rats indicate that gene expression becomes more heterogeneous with age (Somel et al. (2006), and Li et al. (2009) further showed that gene expression variability in male rats is age-dependent. More recently, using human twin data, Oh et al. (2016) found that gene expression levels as well as epigenetic modifications increased in similarity in brain tissues of older individuals. Furthermore, a large study by Peters et al. (2015) revealed age-related gene expression levels that decreased with age; these findings could be attributable to dysregulation of transcriptional and translational systems.

In the present study, we identified a large number of genes whose population-level expression dispersion is age-related. For example, the variance in *MS4A4E* expression in the hippocampus was greatly increased in individuals over age 60. Such an increase in expression variability may have resulted from a decrease in normal regulation of cell growth and inflammation, which may be related to an increase in AD risk (Akiyama et al., 2000;Hollingworth et al., 2011;Naj et al., 2011;Heppner et al., 2015). The increased gene expression variability may also be due to the interaction between *MS4A4E* with other genes, e.g., *CLU* (Ebbert et al., 2015). Furthermore, we argue that the incomplete penetrance observed in neurodegenerative diseases (Rossor et al., 1996;Healy et al., 2008) may be attributed to differences in phenotypic robustness, which may be associated with or reflected in the age-related gene expression variability among individuals who are susceptible to these diseases.

Also, it is interesting to explore possible mechanisms underlying the increase or decrease in gene expression variability, as global gene expression is under stabilizing selection (Khaitovich et al., 2004;Lemos et al., 2005). Previously, we have shown that both common and rare genetic variants may confer regulatory function to contribute to gene expression dispersion (Hulse and Cai, 2013;Wang et al., 2014;Zeng et al., 2015). In particular, common genetic variants contribute to gene expression variability via distinct modes of action—e.g., epistasis and destabilizing mutations (Wang et al., 2015). Rare and private regulatory variants have been found to be responsible for extreme gene expression in outlier samples (Montgomery et al., 2011;Zeng et al., 2015;Zhao et al., 2016). Given this background information about the genetic

14

regulation of gene expression, we argue that aging may be associated with gene expression through age-related genome instability. Mutations accumulate with age in a tissue-specific manner. The major components of the mutation spectrum include point mutations and genome rearrangements such as translocations and large deletions (Busuttil et al., 2007a). The accumulation of somatic mutations over time in various tissues and organs has been suggested as a general explanation of aging (Szilard, 1959;Curtis, 1963;Vijg, 2004). Different organs or tissues show greatly different rates of mutations that accumulate with age. The brain as a whole does not seem to accumulate mutations with age at all, but certain regions of the brain (e.g., hippocampus and hypothalamus) are much more susceptible to mutagenesis and do show increased mutational loads at old age (Busuttil et al., 2007b).

In conclusion, we demonstrate that age-related gene expression is brain region-specific, genotype-dependent, and both mean and dispersion changes in expression level are associated with the aging process. These findings provide a necessary foundation for more sophisticated gene expression modeling in the studies of age-related neurodegenerative diseases.

Acknowledgements

Geneva (MH090941 & MH101814), the University of Chicago (MH090951, MH090937, MH101820, MH101825), the University of North Carolina - Chapel Hill (MH090936 & MH101819), Harvard University (MH090948), Stanford University (MH101782), Washington University St Louis (MH101810), and the University of Pennsylvania (MH101822). The data used for the analyses described in this manuscript were obtained from: the GTEx Portal on 10/29/2015 and dbGaP accession number phs000424.v6.p1 on 10/30/2015.

Author contributions

JJC conceived and designed the study. CLB, JG, GJ and JJC conducted the analyses. CLB, JG, GJ, and JJC wrote the manuscript.

Competing interests

The authors declare no competing interests.

References

Akiyama, H., Barger, S., Barnum, S., Bradt, B., Bauer, J., Cole, G.M., Cooper, N.R., Eikelenboom, P., Emmerling, M., Fiebich, B.L., Finch, C.E., Frautschy, S., Griffin, W.S., Hampel, H., Hull, M., Landreth, G., Lue, L., Mrak, R., Mackenzie, I.R., Mcgeer, P.L., O'banion, M.K., Pachter, J., Pasinetti, G., Plata-Salaman, C., Rogers, J., Rydel, R., Shen, Y., Streit, W., Strohmeyer, R., Tooyoma, I., Van Muiswinkel, F.L., Veerhuis, R., Walker, D., Webster, S., Wegrzyniak, B., Wenk, G., and Wyss-Coray, T. (2000). Inflammation and Alzheimer's disease. *Neurobiol Aging* 21**,** 383-421.

Anand Brown, A., Ding, Z., Vinuela, A., Glass, D., Parts, L., Spector, T., Winn, J., and Durbin, R. (2015). Pathway-based factor analysis of gene expression data produces highly heritable phenotypes that associate with age. *G3 (Bethesda)* 5**,** 839-847.

Anderson, M.J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62**,** 245-253.

Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttil, R.A., Dolle, M.E., Calder, R.B., Chisholm, G.B., Pollock, B.H., Klein, C.A., and Vijg, J. (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* 441**,** 1011-1014.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* 57**,** 289-300.

Berlett, B.S., and Stadtman, E.R. (1997). Protein oxidation in aging, disease, and oxidative stress. *J Biol Chem* 272**,** 20313-20316.

Busuttil, R., Bahar, R., and Vijg, J. (2007a). Genome dynamics and transcriptional deregulation in aging. *Neuroscience* 145**,** 1341-1347.

Busuttil, R.A., Garcia, A.M., Reddick, R.L., Dolle, M.E., Calder, R.B., Nelson, J.F., and Vijg, J. (2007b). Intra-organ variation in age-related mutation accumulation in the mouse. *PLoS One* 2**,** e876.

Carithers, L.J., Ardlie, K., Barcus, M., Branton, P.A., Britton, A., Buia, S.A., Compton, C.C., Deluca, D.S., Peter-Demchok, J., Gelfand, E.T., Guan, P., Korzeniewski, G.E., Lockhart, N.C., Rabiner, C.A., Rao, A.K., Robinson, K.L., Roche, N.V., Sawyer, S.J., Segre, A.V.,

Shive, C.E., Smith, A.M., Sobin, L.H., Undale, A.H., Valentino, K.M., Vaught, J., Young, T.R., Moore, H.M., and Consortium, G.T. (2015). A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank* 13**,** 311-319.

Curtis, H.J. (1963). Biological mechanisms underlying the aging process. *Science* 141**,** 686-694.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4**,** P3.

Ebbert, M.T., Boehme, K.L., Wadsworth, M.E., Staley, L.A., Alzheimer's Disease Neuroimaging, I., Alzheimer's Disease Genetics, C., Mukherjee, S., Crane, P.K., Ridge, P.G., and Kauwe, J.S. (2015). Interaction between variants in CLU and MS4A4E modulates Alzheimer's disease risk. *Alzheimers Dement*.

Erikson, G.A., Bodian, D.L., Rueda, M., Molparia, B., Scott, E.R., Scott-Van Zeeland, A.A., Topol, S.E., Wineinger, N.E., Niederhuber, J.E., Topol, E.J., and Torkamani, A. (2016). Whole-Genome Sequencing of a Healthy Aging Cohort. *Cell* 165**,** 1002-1011.

Fearnley, J.M., and Lees, A.J. (1991). Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain* 114 (Pt 5)**,** 2283-2301.

Feser, J., and Tyler, J. (2011). Chromatin structure as a mediator of aging. *FEBS Lett* 585**,** 2041-2048.

Francis, P.T., Palmer, A.M., Snape, M., and Wilcock, G.K. (1999). The cholinergic hypothesis of Alzheimer's disease: a review of progress. *J Neurol Neurosurg Psychiatry* 66**,** 137-147.

Fraser, H.B., Khaitovich, P., Plotkin, J.B., Paabo, S., and Eisen, M.B. (2005). Aging and gene expression in the primate brain. *PLoS Biol* 3**,** e274.

Glass, D., Vinuela, A., Davies, M.N., Ramasamy, A., Parts, L., Knowles, D., Brown, A.A., Hedman, A.K., Small, K.S., Buil, A., Grundberg, E., Nica, A.C., Di Meglio, P., Nestle, F.O., Ryten, M., Consortium, U.K.B.E., Mu, T.C., Durbin, R., Mccarthy, M.I., Deloukas, P., Dermitzakis, E.T., Weale, M.E., Bataille, V., and Spector, T.D. (2013). Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol* 14**,** R75.

Graveland, G.A., Williams, R.S., and Difiglia, M. (1985). Evidence for degenerative and regenerative changes in neostriatal spiny neurons in Huntington's disease. *Science* 227**,** 770-773.

Groelz, D., Sobin, L., Branton, P., Compton, C., Wyrich, R., and Rainen, L. (2013). Non-formalin fixative versus formalin-fixed tissue: a comparison of histology and RNA quality. *Exp Mol Pathol* 94**,** 188-194.

Gtex_Consortium (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348**,** 648-660.

Guerreiro, R.J., Beck, J., Gibbs, J.R., Santana, I., Rossor, M.N., Schott, J.M., Nalls, M.A., Ribeiro, H., Santiago, B., Fox, N.C., Oliveira, C., Collinge, J., Mead, S., Singleton, A., and Hardy, J. (2010). Genetic variability in CLU and its association with Alzheimer's disease. *PLoS One* 5**,** e9510.

Harman, D. (1992). Role of free radicals in aging and disease. *Ann N Y Acad Sci* 673**,** 126-141.

Healy, D.G., Falchi, M., O'sullivan, S.S., Bonifati, V., Durr, A., Bressman, S., Brice, A., Aasly, J., Zabetian, C.P., Goldwurm, S., Ferreira, J.J., Tolosa, E., Kay, D.M., Klein, C., Williams, D.R., Marras, C., Lang, A.E., Wszolek, Z.K., Berciano, J., Schapira, A.H., Lynch, T., Bhatia, K.P., Gasser, T., Lees, A.J., Wood, N.W., and International, L.C. (2008). Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol* 7**,** 583-590.

Heppner, F.L., Ransohoff, R.M., and Becher, B. (2015). Immune attack: the role of inflammation in Alzheimer disease. *Nat Rev Neurosci* 16**,** 358-372.

Hernandez, D.G., Nalls, M.A., Moore, M., Chong, S., Dillman, A., Trabzuni, D., Gibbs, J.R., Ryten, M., Arepalli, S., Weale, M.E., Zonderman, A.B., Troncoso, J., O'brien, R., Walker,

R., Smith, C., Bandinelli, S., Traynor, B.J., Hardy, J., Singleton, A.B., and Cookson, M.R. (2012). Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiol Dis* 47, 20-28.

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J.C., Carrasquillo, M.M., Abraham, R., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Jones, N., Stretton, A., Thomas, C., Richards, A., Ivanov, D., Widdowson, C., Chapman, J., Lovestone, S., Powell, J., Proitsi, P., Lupton, M.K., Brayne, C., Rubinsztein, D.C., Gill, M., Lawlor, B., Lynch, A., Brown, K.S., Passmore, P.A., Craig, D., Mcguinness, B., Todd, S., Holmes, C., Mann, D., Smith, A.D., Beaumont, H., Warden, D., Wilcock, G., Love, S., Kehoe, P.G., Hooper, N.M., Vardy, E.R., Hardy, J., Mead, S., Fox, N.C., Rossor, M., Collinge, J., Maier, W., Jessen, F., Ruther, E., Schurmann, B., Heun, R., Kolsch, H., Van Den Bussche, H., Heuser, I., Kornhuber, J., Wiltfang, J., Dichgans, M., Frolich, L., Hampel, H., Gallacher, J., Hull, M., Rujescu, D., Giegling, I., Goate, A.M., Kauwe, J.S., Cruchaga, C., Nowotny, P., Morris, J.C., Mayo, K., Sleegers, K., Bettens, K., Engelborghs, S., De Deyn, P.P., Van Broeckhoven, C., Livingston, G., Bass, N.J., Gurling, H., Mcquillin, A., Gwilliam, R., Deloukas, P., Al-Chalabi, A., Shaw, C.E., Tsolaki, M., Singleton, A.B., Guerreiro, R., Muhleisen, T.W., Nothen, M.M., Moebus, S., Jockel, K.H., Klopp, N., Wichmann, H.E., Pankratz, V.S., Sando, S.B., Aasly, J.O., Barcikowska, M., Wszolek, Z.K., Dickson, D.W., Graff-Radford, N.R., Petersen, R.C., et al. (2011). Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 43, 429-435.

Hong, S., Beja-Glasser, V.F., Nfonoyim, B.M., Frouin, A., Li, S., Ramakrishnan, S., Merry, K.M., Shi, Q., Rosenthal, A., Barres, B.A., Lemere, C.A., Selkoe, D.J., and Stevens, B. (2016). Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science*.

Hulse, A.M., and Cai, J.J. (2013). Genetic variants contribute to gene expression variability in humans. *Genetics* 193, 95-108.

Hyman, B.T., Van Hoesen, G.W., Damasio, A.R., and Barnes, C.L. (1984). Alzheimer's disease: cell-specific pathology isolates the hippocampal formation. *Science* 225, 1168-1170.

Jellinger, K. (1986). "Pathology of parkinsonism," in *Recent developments in Parkinson's disease,* eds. S. Fahn, C.D. Marsden, P. Jenner & P. Teychenne. (Raven, New York), 33-36.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M.B., Krsnik, Z., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S.N., Vortmeyer, A., Weinberger, D.R., Mane, S., Hyde, T.M., Huttner, A., Reimers, M., Kleinman, J.E., and Sestan, N. (2011). Spatio-temporal transcriptome of the human brain. *Nature* 478, 483-489.

Karagiannis, S.N., Wang, Q., East, N., Burke, F., Riffard, S., Bracher, M.G., Thompson, R.G., Durham, S.R., Schwartz, L.B., Balkwill, F.R., and Gould, H.J. (2003). Activity of human monocytes in IgE antibody-dependent surveillance and killing of ovarian tumor cells. *Eur J Immunol* 33, 1030-1040.

Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., Wirkner, U., Ansorge, W., and Paabo, S. (2004). A neutral model of transcriptome evolution. *PLoS Biol* 2, E132.

Leek, J.T., and Storey, J.D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 3, 1724-1735.

Lemos, B., Meiklejohn, C.D., Caceres, M., and Hartl, D.L. (2005). Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* 59, 126-137.

Li, Q., Athan, E.S., Wei, M., Yuan, E., Rice, S.L., Vonsattel, J.P., Mayeux, R.P., and Tycko, B. (2004). TP73 allelic expression in human brain and allele frequencies in Alzheimer's disease. *BMC Med Genet* 5**,** 14.

Li, Z., Wright, F.A., and Royland, J. (2009). Age-dependent variability in gene expression in male Fischer 344 rat retina. *Toxicol Sci* 107**,** 281-292.

Liang, Y., Buckley, T.R., Tu, L., Langdon, S.D., and Tedder, T.F. (2001). Structural organization of the human MS4A gene cluster on Chromosome 11q12. *Immunogenetics* 53**,** 357-368.

Lin, J., Countryman, P., Buncher, N., Kaur, P., E, L., Zhang, Y., Gibson, G., You, C., Watkins, S.C., Piehler, J., Opresko, P.L., Kad, N.M., and Wang, H. (2014). TRF1 and TRF2 use different mechanisms to find telomeric DNA but share a novel mechanism to search for protein partners at telomeres. *Nucleic Acids Res* 42**,** 2493-2504.

Lu, T., Pan, Y., Kao, S.Y., Li, C., Kohane, I., Chan, J., and Yankner, B.A. (2004). Gene regulation and DNA damage in the ageing human brain. *Nature* 429**,** 883-891.

Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., Arnold, J.M., Bennet, C., Bertagnolli, D., Brouner, K., Butler, S., Caldejon, S., Carey, A., Cuhaciyan, C., Dalley, R.A., Dee, N., Dolbeare, T.A., Facer, B.A., Feng, D., Fliss, T.P., Gee, G., Goldy, J., Gourley, L., Gregor, B.W., Gu, G., Howard, R.E., Jochim, J.M., Kuan, C.L., Lau, C., Lee, C.K., Lee, F., Lemon, T.A., Lesnar, P., Mcmurray, B., Mastan, N., Mosqueda, N., Naluai-Cecchini, T., Ngo, N.K., Nyhus, J., Oldre, A., Olson, E., Parente, J., Parker, P.D., Parry, S.E., Stevens, A., Pletikos, M., Reding, M., Roll, K., Sandman, D., Sarreal, M., Shapouri, S., Shapovalova, N.V., Shen, E.H., Sjoquist, N., Slaughterbeck, C.R., Smith, M., Sodt, A.J., Williams, D., Zollei, L., Fischl, B., Gerstein, M.B., Geschwind, D.H., Glass, I.A., Hawrylycz, M.J., Hevner, R.F., Huang, H., Jones, A.R., Knowles, J.A., Levitt, P., Phillips, J.W., Sestan, N., Wohnoutka, P., Dang, C., Bernard, A., Hohmann, J.G., and Lein, E.S. (2014). Transcriptional landscape of the prenatal human brain. *Nature* 508**,** 199-206.

Montgomery, S.B., Lappalainen, T., Gutierrez-Arcelus, M., and Dermitzakis, E.T. (2011). Rare and common regulatory variation in population-scale sequenced human genomes. *PLoS Genet* 7**,** e1002144.

Naito, A.T., Sumida, T., Nomura, S., Liu, M.L., Higo, T., Nakagawa, A., Okada, K., Sakai, T., Hashimoto, A., Hara, Y., Shimizu, I., Zhu, W., Toko, H., Katada, A., Akazawa, H., Oka, T., Lee, J.K., Minamino, T., Nagai, T., Walsh, K., Kikuchi, A., Matsumoto, M., Botto, M., Shiojima, I., and Komuro, I. (2012). Complement C1q activates canonical Wnt signaling and promotes aging-related phenotypes. *Cell* 149**,** 1298-1313.

Naj, A.C., Jun, G., Beecham, G.W., Wang, L.S., Vardarajan, B.N., Buros, J., Gallins, P.J., Buxbaum, J.D., Jarvik, G.P., Crane, P.K., Larson, E.B., Bird, T.D., Boeve, B.F., Graff-Radford, N.R., De Jager, P.L., Evans, D., Schneider, J.A., Carrasquillo, M.M., Ertekin-Taner, N., Younkin, S.G., Cruchaga, C., Kauwe, J.S., Nowotny, P., Kramer, P., Hardy, J., Huentelman, M.J., Myers, A.J., Barmada, M.M., Demirci, F.Y., Baldwin, C.T., Green, R.C., Rogaeva, E., St George-Hyslop, P., Arnold, S.E., Barber, R., Beach, T., Bigio, E.H., Bowen, J.D., Boxer, A., Burke, J.R., Cairns, N.J., Carlson, C.S., Carney, R.M., Carroll, S.L., Chui, H.C., Clark, D.G., Corneveaux, J., Cotman, C.W., Cummings, J.L., Decarli, C., Dekosky, S.T., Diaz-Arrastia, R., Dick, M., Dickson, D.W., Ellis, W.G., Faber, K.M., Fallon, K.B., Farlow, M.R., Ferris, S., Frosch, M.P., Galasko, D.R., Ganguli, M., Gearing, M., Geschwind, D.H., Ghetti, B., Gilbert, J.R., Gilman, S., Giordani, B., Glass, J.D., Growdon, J.H., Hamilton, R.L., Harrell, L.E., Head, E., Honig, L.S., Hulette, C.M., Hyman, B.T., Jicha, G.A., Jin, L.W., Johnson, N., Karlawish, J., Karydas, A., Kaye, J.A., Kim, R., Koo, E.H., Kowall, N.W., Lah, J.J., Levey, A.I., Lieberman, A.P., Lopez, O.L., Mack, W.J., Marson, D.C., Martiniuk, F., Mash, D.C., Masliah, E., Mccormick, W.C., Mccurry, S.M., Mcdavid, A.N., Mckee, A.C., Mesulam, M., Miller, B.L., et al. (2011). Common variants at

MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 43**,** 436-441.

Nelson, P.T., Wang, W.X., Wilfred, B.R., Wei, A., Dimayuga, J., Huang, Q., Ighodaro, E., Artiushin, S., and Fardo, D.W. (2015). Novel human ABCC9/SUR2 brain-expressed transcripts and an eQTL relevant to hippocampal sclerosis of aging. *J Neurochem* 134**,** 1026-1039.

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M., Mccluskey, L.F., Miller, B.L., Masliah, E., Mackenzie, I.R., Feldman, H., Feiden, W., Kretzschmar, H.A., Trojanowski, J.Q., and Lee, V.M. (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 314**,** 130-133.

Nuttall, S.L., Martin, U., Sinclair, A.J., and Kendall, M.J. (1998). Glutathione: in sickness and in health. *Lancet* 351**,** 645-646.

Oh, G., Ebrahimi, S., Wang, S.C., Cortese, R., Kaminsky, Z.A., Gottesman, Ii, Burke, J.R., Plassman, B.L., and Petronis, A. (2016). Epigenetic assimilation in the aging human brain. *Genome Biol* 17**,** 76.

Oldham, M.C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., and Geschwind, D.H. (2008). Functional organization of the transcriptome in human brain. *Nat Neurosci* 11**,** 1271-1282.

Parts, L., Stegle, O., Winn, J., and Durbin, R. (2011). Joint genetic analysis of gene expression data with inferred cellular phenotypes. *PLoS Genet* 7**,** e1001276.

Passtoors, W.M., Van Den Akker, E.B., Deelen, J., Maier, A.B., Van Der Breggen, R., Jansen, R., Trompet, S., Van Heemst, D., Derhovanessian, E., Pawelec, G., Van Ommen, G.J., Slagboom, P.E., and Beekman, M. (2015). IL7R gene expression network associates with human healthy ageing. *Immun Ageing* 12**,** 21.

Peters, M.J., Joehanes, R., Pilling, L.C., Schurmann, C., Conneely, K.N., Powell, J., Reinmaa, E., Sutphin, G.L., Zhernakova, A., Schramm, K., Wilson, Y.A., Kobes, S., Tukiainen, T., Consortium, N.U., Ramos, Y.F., Goring, H.H., Fornage, M., Liu, Y., Gharib, S.A., Stranger, B.E., De Jager, P.L., Aviv, A., Levy, D., Murabito, J.M., Munson, P.J., Huan, T., Hofman, A., Uitterlinden, A.G., Rivadeneira, F., Van Rooij, J., Stolk, L., Broer, L., Verbiest, M.M., Jhamai, M., Arp, P., Metspalu, A., Tserel, L., Milani, L., Samani, N.J., Peterson, P., Kasela, S., Codd, V., Peters, A., Ward-Caviness, C.K., Herder, C., Waldenberger, M., Roden, M., Singmann, P., Zeilinger, S., Illig, T., Homuth, G., Grabe, H.J., Volzke, H., Steil, L., Kocher, T., Murray, A., Melzer, D., Yaghootkar, H., Bandinelli, S., Moses, E.K., Kent, J.W., Curran, J.E., Johnson, M.P., Williams-Blangero, S., Westra, H.J., Mcrae, A.F., Smith, J.A., Kardia, S.L., Hovatta, I., Perola, M., Ripatti, S., Salomaa, V., Henders, A.K., Martin, N.G., Smith, A.K., Mehta, D., Binder, E.B., Nylocks, K.M., Kennedy, E.M., Klengel, T., Ding, J., Suchy-Dicey, A.M., Enquobahrie, D.A., Brody, J., Rotter, J.I., Chen, Y.D., Houwing-Duistermaat, J., Kloppenburg, M., Slagboom, P.E., Helmer, Q., Den Hollander, W., Bean, S., Raj, T., Bakhshi, N., Wang, Q.P., Oyston, L.J., Psaty, B.M., Tracy, R.P., Montgomery, G.W., Turner, S.T., et al. (2015). The transcriptional landscape of age in human peripheral blood. *Nat Commun* 6**,** 8570.

Pozniak, C.D., Barnabe-Heider, F., Rymar, V.V., Lee, A.F., Sadikot, A.F., and Miller, F.D. (2002). p73 is required for survival and maintenance of CNS neurons. *J Neurosci* 22**,** 9800-9809.

Pozniak, C.D., Radinovic, S., Yang, A., Mckeon, F., Kaplan, D.R., and Miller, F.D. (2000). An anti-apoptotic role for the p53 family member, p73, during developmental neuron death. *Science* 289**,** 304-306.

Proitsi, P., Lee, S.H., Lunnon, K., Keohane, A., Powell, J., Troakes, C., Al-Sarraj, S., Furney, S., Soininen, H., Kloszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Lovestone, S., Hodges, A., and Addneuromed, C. (2014). Alzheimer's disease susceptibility variants in the

MS4A6A gene are associated with altered levels of MS4A6A expression in blood. *Neurobiol Aging* 35**,** 279-290.

Rodriguez, S.A., Grochova, D., Mckenna, T., Borate, B., Trivedi, N.S., Erdos, M.R., and Eriksson, M. (2015). Global genome splicing analysis reveals an increased number of alternatively spliced genes with aging. *Aging Cell*.

Rossor, M.N., Fox, N.C., Beck, J., Campbell, T.C., and Collinge, J. (1996). Incomplete penetrance of familial Alzheimer's disease in a pedigree with a novel presenilin-1 gene mutation. *Lancet* 347**,** 1560.

Rousseeuw, P.J., and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41**,** 212-223.

Segovia, G., Porras, A., Del Arco, A., and Mora, F. (2001). Glutamatergic neurotransmission in aging: a critical perspective. *Mech Ageing Dev* 122**,** 1-29.

Somel, M., Khaitovich, P., Bahn, S., Paabo, S., and Lachmann, M. (2006). Gene expression becomes heterogeneous with age. *Curr Biol* 16**,** R359-360.

Sood, S., Gallagher, I.J., Lunnon, K., Rullman, E., Keohane, A., Crossland, H., Phillips, B.E., Cederholm, T., Jensen, T., Van Loon, L.J., Lannfelt, L., Kraus, W.E., Atherton, P.J., Howard, R., Gustafsson, T., Hodges, A., and Timmons, J.A. (2015). A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome Biol* 16**,** 185.

Southworth, L.K., Owen, A.B., and Kim, S.K. (2009). Aging mice show a decreasing correlation of gene expression within genetic modules. *PLoS Genet* 5**,** e1000776.

Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* 7**,** 500-507.

Stephan, A.H., Madison, D.V., Mateos, J.M., Fraser, D.A., Lovelett, E.A., Coutellier, L., Kim, L., Tsai, H.H., Huang, E.J., Rowitch, D.H., Berns, D.S., Tenner, A.J., Shamloo, M., and Barres, B.A. (2013). A dramatic increase of C1q protein in the CNS during normal aging. *J Neurosci* 33**,** 13460-13474.

Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6**,** e21800.

Szilard, L. (1959). On the Nature of the Aging Process. *Proc Natl Acad Sci U S A* 45**,** 30-45.

Tebbenkamp, A.T., Willsey, A.J., State, M.W., and Sestan, N. (2014). The developmental transcriptome of the human brain: implications for neurodevelopmental disorders. *Curr Opin Neurol* 27**,** 149-156.

Vijg, J. (2004). Impact of genome instability on transcription regulation of aging and senescence. *Mech Ageing Dev* 125**,** 747-753.

Vonsattel, J.P., Myers, R.H., Stevens, T.J., Ferrante, R.J., Bird, E.D., and Richardson, E.P., Jr. (1985). Neuropathological classification of Huntington's disease. *J Neuropathol Exp Neurol* 44**,** 559-577.

Walter, S., Atzmon, G., Demerath, E.W., Garcia, M.E., Kaplan, R.C., Kumari, M., Lunetta, K.L., Milaneschi, Y., Tanaka, T., Tranah, G.J., Volker, U., Yu, L., Arnold, A., Benjamin, E.J., Biffar, R., Buchman, A.S., Boerwinkle, E., Couper, D., De Jager, P.L., Evans, D.A., Harris, T.B., Hoffmann, W., Hofman, A., Karasik, D., Kiel, D.P., Kocher, T., Kuningas, M., Launer, L.J., Lohman, K.K., Lutsey, P.L., Mackenbach, J., Marciante, K., Psaty, B.M., Reiman, E.M., Rotter, J.I., Seshadri, S., Shardell, M.D., Smith, A.V., Van Duijn, C., Walston, J., Zillikens, M.C., Bandinelli, S., Baumeister, S.E., Bennett, D.A., Ferrucci, L., Gudnason, V., Kivimaki, M., Liu, Y., Murabito, J.M., Newman, A.B., Tiemeier, H., and Franceschini, N. (2011). A genome-wide association study of aging. *Neurobiol Aging* 32**,** 2109 e2115-2128.

Wang, G., Yang, E., Brinkmeyer-Langford, C.L., and Cai, J.J. (2014). Additive, epistatic, and environmental effects through the lens of expression variability QTL in a twin cohort. *Genetics* 196**,** 413-425.

Wang, G., Yang, E., Yang, J., Zhou, B., Tian, Y., and Cai, J.J. (2015). Epistasis and decanalization shape gene expression variability in humans via distinct modes of action. *bioRxiv*.

Warren, L.A., Rossi, D.J., Schiebinger, G.R., Weissman, I.L., Kim, S.K., and Quake, S.R. (2007). Transcriptional instability is not a universal attribute of aging. *Aging Cell* 6**,** 775-782.

Wetzel, M.K., Naska, S., Laliberte, C.L., Rymar, V.V., Fujitani, M., Biernaskie, J.A., Cole, C.J., Lerch, J.P., Spring, S., Wang, S.H., Frankland, P.W., Henkelman, R.M., Josselyn, S.A., Sadikot, A.F., Miller, F.D., and Kaplan, D.R. (2008). p73 regulates neurodegeneration and phospho-tau accumulation during aging and Alzheimer's disease. *Neuron* 59**,** 708-721.

Yang, A., Walker, N., Bronson, R., Kaghad, M., Oosterwegel, M., Bonnin, J., Vagner, C., Bonnet, H., Dikkes, P., Sharpe, A., Mckeon, F., and Caput, D. (2000). p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. *Nature* 404**,** 99-103.

Yang, J., Huang, T., Petralia, F., Long, Q., Zhang, B., Argmann, C., Zhao, Y., Mobbs, C.V., Schadt, E.E., Zhu, J., Tu, Z., and Consortium, G.T. (2015). Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci Rep* 5**,** 15145.

Yao, P., Lin, P., Gokoolparsadh, A., Assareh, A., Thang, M.W., and Voineagu, I. (2015). Coexpression networks identify brain region-specific enhancer RNAs in the human brain. *Nat Neurosci* 18**,** 1168-1174.

Zeng, Y., Wang, G., Yang, E., Ji, G., Brinkmeyer-Langford, C.L., and Cai, J.J. (2015). Aberrant gene expression in humans. *PLoS Genet* 11**,** e1004942.

Zhao, J., Akinsanmi, I., Arafat, D., Cradick, T.J., Lee, C.M., Banskota, S., Marigorta, U.M., Bao, G., and Gibson, G. (2016). A Burden of Rare Variants Associated with Extremes of Gene Expression in Human Peripheral Blood. *Am J Hum Genet* 98**,** 299-309.

Zhu, M., Tao, Y., He, Q., Gao, H., Song, F., Sun, Y.M., Li, H.L., Wu, Z.Y., and Saffen, D. (2014). Common GSAP promoter variant contributes to Alzheimer's disease liability. *Neurobiol Aging* 35**,** 2656 e2651-2657.

Figure Legends

Fig. 1. Correlation matrix for the responsiveness of genes to aging between 13 GTEx brain regions. The responsiveness to aging of a gene was measured with the *P*-value for slope coefficient in the linear regression model between age and gene expression. The correlation between each two brain regions was estimated with the nonparametric Spearman rank correlation coefficient between *P*-values of all genes in the two regions. The order of regions in the matrix was rearranged based on the similarity between regions. Six clusters of highly correlated regions are highlighted with red boxes: I – cortex; II – cerebellum; III – basal ganglia; IV – hypothalamus, amygdala, and hippocampus; V – substantia nigra and spinal cord; and IV – amygdala and caudate.

Fig. 2. TreeMap view of GO-term clusters for age-related genes in the brain. The TreeMap was generated by using REVIGO with the input of all GO BP terms of age-related genes detected in 13 GTEx brain regions at the FDR of 5%. Each rectangle represents a single GO-term cluster. The size of the rectangles is proportional to the frequency of the GO term in the associated GO annotation database. The cluster representatives were joined into 'superclusters' of loosely related terms, visualized with different colors.

Fig. 3. Examples of genotype-by-age interaction affecting the expression level of the gene. (**A**) The interaction between rs55675298 and age affecting *TP73* gene expression. (**B**) The interaction between rs72788737 and age affecting *C1QA* gene expression. For each subplot, the left panel shows all samples, the middle panel shows the major allele homozygous samples, and the right panel shows heterozygous and minor allele homozygous (with a cross) samples.

Fig. 4. Differential gene expression dispersion between age groups. (**A**) Increased gene expression variance of *IL7R* in hippocampus between age groups (left: 20-60 years vs. 61-70 years; right: 20-55, 56-64, and 64-70 years). Each age group was plotted with jitter along the *x*-axis to show samples within each genotype. (**B**) Same as (**A**) but for *MS4A4E*. (**C**) Scatter plot (left) of expression levels of *TERF1* and *TERF2* with data points grouped by young (20-59 years) and old (60-70 years) ages. Boxplot (right) of MDs of young and old samples' *TERF1* and *TERF2* expression profiles to the population center.

23

Tables

Table 1. Numbers of age-related genes in 13 GTEx brain regions. Columns "Up" and "Down" list the numbers of up-regulated and down-regulated aging genes, respectively. Results derived from using three different FDR cutoffs (1%, 5%, and 10%) are shown. Two regions, (5) Cerebellum (PAXgene) and (6) Cortex (PAXgene), for which samples were preserved using the PAXgene tissue preservation system, are highlighted in bold.
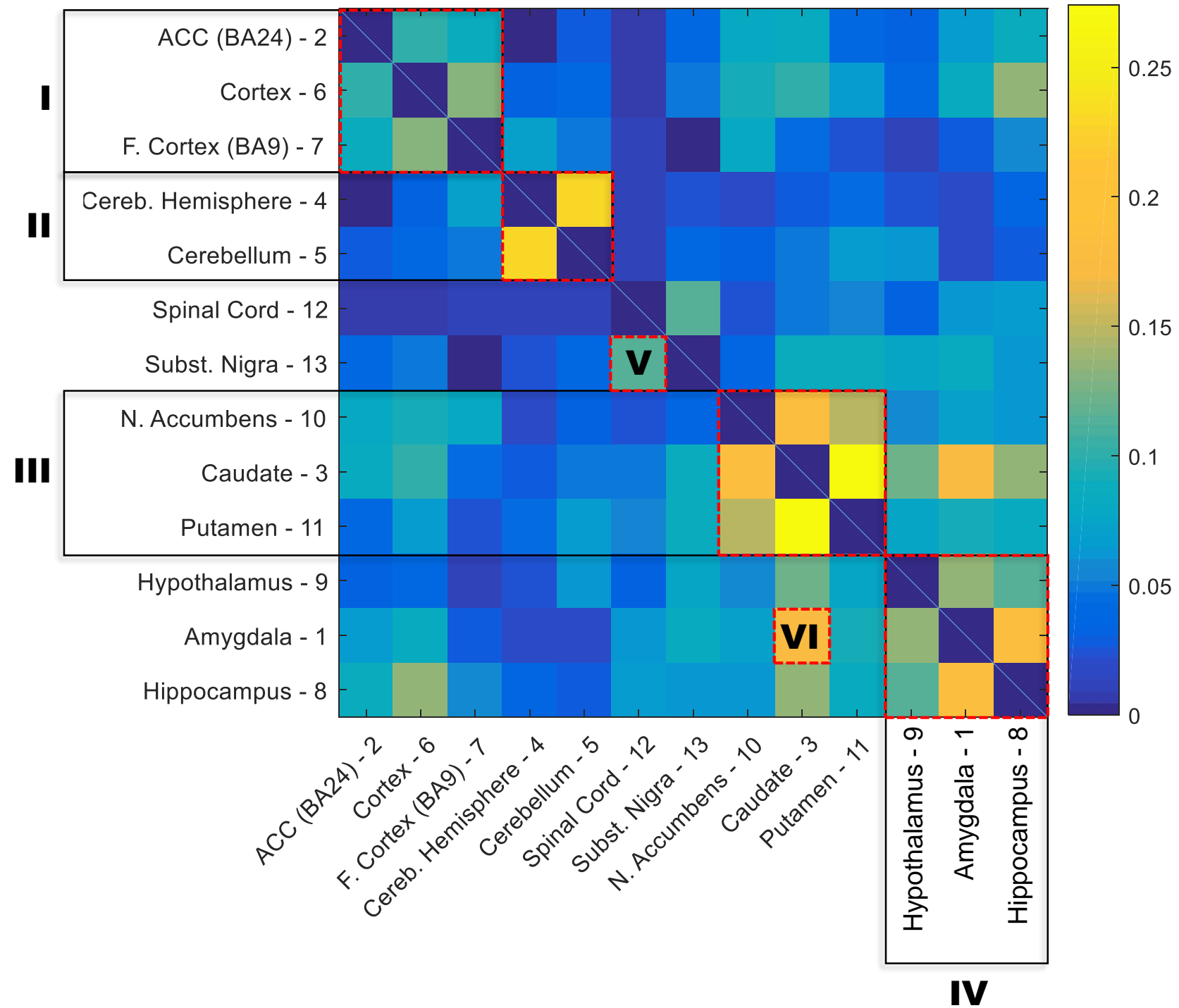
| Brain Region | FDR | | | | | | | | |
| | 1% | | | 5% | | | 10% | | |
	Up	Down	**Total**	Up	Down	**Total**	Up	Down	**Total**
(1) Amygdala (*n*=67)	0	0	**0**	0	1	**1**	8	9	**17**
(2) Anterior Cingulate Cortex (BA24) (*n*=77)	0	0	**0**	0	0	**0**	0	0	**0**
(3) Caudate (basal ganglia) (*n*=101)	23	26	**49**	171	176	**347**	355	390	**745**
(4) Cerebellar Hemisphere (Frozen) (*n*=93)	0	0	**0**	2	8	**10**	23	35	**58**
(5) Cerebellum (PAXgene) (*n*=110)	54	86	**140**	358	466	**824**	688	788	**1476**
(6) Cortex (PAXgene) (*n*=97)	2	4	**6**	52	64	**116**	139	184	**323**
(7) Frontal Cortex (BA9) (Frozen) (*n*=95)	0	0	**0**	2	2	**4**	2	3	**5**
(8) Hippocampus (*n*=82)	0	0	**0**	5	2	**7**	87	42	**129**
(9) Hypothalamus (*n*=86)	0	0	**0**	20	22	**42**	84	81	**165**
(10) Nucleus Accumbens (basal ganglia) (*n*=97)	6	5	**11**	34	33	**67**	104	91	**195**
(11) Putamen (basal ganglia) (*n*=85)	11	9	**20**	110	105	**215**	257	268	**525**
(12) Spinal Cord (cervical c-1) (*n*=64)	0	0	**0**	0	0	**0**	0	0	**0**
(13) Substantia Nigra (*n*=54)	0	0	**0**	1	2	**3**	1	2	**3**

Table 2. Top five GO-term clusters for DV genes showing significant differential variability in their expression between age groups.

GO term	Count	Enrichment fold	FDR	Dispersion change with age
Annotation Cluster 1				**Decrease**
GO:0007186~G-protein coupled receptor protein signaling pathway	106	2.29	7.87E-13	
GO:0007606~sensory perception of chemical stimulus	53	2.69	1.30E-07	
GO:0007166~cell surface receptor linked signal transduction	131	1.71	1.66E-07	
GO:0004984~olfactory receptor activity	47	2.56	5.17E-06	
GO:0007608~sensory perception of smell	46	2.59	4.50E-06	
GO:0050890~cognition	74	1.98	9.18E-06	
GO:0007600~sensory perception	67	2.01	2.33E-05	
GO:0050877~neurological system process	89	1.78	2.11E-05	
Annotation Cluster 2				**Increase**
GO:0008528~peptide receptor activity, G-protein coupled	20	4.11	1.02E-04	
GO:0001653~peptide receptor activity	20	4.11	1.02E-04	
GO:0042277~peptide binding	25	2.89	0.001	
Annotation Cluster 3				**Increase**
GO:0042330~taxis	19	2.88	0.02	
GO:0006935~chemotaxis	19	2.88	0.02	
GO:0007626~locomotory behavior	26	2.30	0.03	
GO:0042379~chemokine receptor binding	10	4.75	0.02	
GO:0007610~behavior	37	1.91	0.04	
GO:0008009~chemokine activity	9	4.59	0.05	
GO:0005125~cytokine activity	16	1.92	0.54	
Annotation Cluster 4				**Decrease**
GO:0004867~serine-type endopeptidase inhibitor activity	14	3.57	0.01	
GO:0030414~peptidase inhibitor activity	17	2.60	0.06	
GO:0004866~endopeptidase inhibitor activity	16	2.59	0.09	
GO:0004857~enzyme inhibitor activity	18	1.56	0.77	
Annotation Cluster 5				**Increase**
GO:0030594~neurotransmitter receptor activity	12	2.96	0.13	
GO:0042165~neurotransmitter binding	12	2.73	0.21	
GO:0008188~neuropeptide receptor activity	6	3.59	0.57	
GO:0042923~neuropeptide binding	6	3.43	0.59	

Table 3. Gene sets whose gene expression dispersion differs significantly (FDR<5%) between young and old adults. Genes of each set were grouped based on the shared GO term in their functional annotations.

GO Term	Gene Set	Dispersion change with age	*P*-value (Levene's test)	FDR
(5) Cerebellum				
GO:0001309~age-dependent telomere shortening	*TERF1, TERF2*	Decrease	3.80E-06	0.0094
GO:0032214~negative regulation of telomere maintenance via semi-conservative replication	*TERF1, TERF2*	Decrease	3.80E-06	0.0094
GO:0003691~double-stranded telomeric DNA binding	*PURA, TERF1, TERF2, XRCC5, XRCC6*	Decrease	2.51E-05	0.0416
(6) Cortex				
GO:0044062~regulation of excretion	*NPHS1, SLC9A3R1*	Decrease	6.13E-06	0.0182
GO:0009756~carbohydrate mediated signaling	*CLEC7A, COLEC12*	Increase	7.77E-06	0.0242
GO:0003840~gamma-glutamyltransferase activity	*GGT1, GGT2, GGT5, GGT6, GGT7, GGTLC1, GGTLC2, GGTLC3*	Increase	8.07E-05	0.035
GO:0036374~glutathione hydrolase activity	*GGT1, GGT5, GGT6, GGT7*	Increase	8.07E-05	0.035
(13) Substantia nigra				
GO:2000146~negative regulation of cell motility	*AP1AR, CTNNA1, GATA3, PIN1, TACSTD2*	Decrease	5.20E-06	0.036

26