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, TX 77843, 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)

Running title: Gene expression of aging brains

Keywords: aging, gene expression, human brain, age-related gene, genotype-by-age interaction, mean and dispersion, expression variability.

Abstract

Human aging is associated with cognitive decline and an increased risk of neurodegenerative disease. We assessed age-related brain gene expression by analyzing the Genotype-Tissue Expression (GTEx) data from 2,522 tissue samples across 13 brain regions in 169 donors of European descent aged between 20 and 70 years. After controlling for covariates and hidden confounding factors, we identified 1,446 protein-coding genes whose expression in one or more brain regions is correlated or anti-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 markedly uneven, suggesting a variable brain region-specific response to aging. The aging response of many genes, e.g., *TP37* and *C1QA*, is dependent on individuals' genotypic backgrounds. Finally, using dispersion-specific analysis, we identified genes, e.g., *IL7R*, *MS4A4E*, and *TERF1*/*TERF2*, whose expression is 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 during an individual's lifespan. The process of aging has profound impacts on physical and mental health. It is well known that aging is associated with cognitive decline and an increased risk of neurodegenerative disease; therefore, understanding the effects of aging on the human brain is of particular interest. One of the ways to approach this goal is to detect the gene expression changes in the human brain during the aging process [e.g., ref[1]]. Brain transcriptomic assay holds promise for a better understanding of the role of aging in both normal brain activity and the development of neurodegenerative disease. The advent of high-throughput sequencing technology has allowed the study of genome-wide patterns of gene expression changes associated with aging.

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[2]. 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.

Most recent studies on age-related gene expression, such as refs.[3,4], have either overlooked the central nervous system or focused on the brain as a whole to compare with other organs and tissues. Meanwhile, mounting evidence shows that the human brain is highly heterogeneous with different sub-regions showing distinct functions, cell-type composition, and gene expression patterns[5,6]. Therefore, it is necessary to treat the different sub-regions of the human brain distinctly so that the region-specific properties of the brain can be revealed. This is important as neurodegenerative diseases such as amyotrophic lateral sclerosis, Parkinson’s disease, Alzheimer’s disease, and Huntington’s disease tend to be unique to the affected brain sub-regions[7-10].

In the present study, we focused on the differential gene expression associated with the aging process in multiple brain regions. In addition to single genes, we also applied newly developed factor analysis[11] to identify functional gene sets that are associated with age. More important, our analysis was directed to gene expression dispersion (e.g., the variance in gene expression across samples) as a metric to reveal a new paradigm of aging-related gene expression patterns. Several studies in humans and model organisms have suggested that aging may

influence the level of phenotypic dispersion of the population. The increase of dispersion may be manifested as increases in between-individual variability[12], heterogeneity of residual expression[13], and cell-to-cell expression variation[14], or as a decreased correlation of gene expression within genetic modules[15]. Thus, to achieve a comprehensive picture of brain aging, we included the dispersion-specific analyses of gene expression with aging.

Methods

GTEx brain tissues, genotype, and expression data

The expression data (v6, October 2015 release) for brain specimens of 169 post-mortem donors aged between 20 and 70 years were obtained from the GTEx portal website. 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[16], 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[2,17]. 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**).

Correcting for confounding factors using PEER algorithm

Prior to the regression analysis, we used a two-step approach based on the PEER algorithm[18] 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 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 $1 \le k \le N$ 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} ($1 \le k \le N$) 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 gene ontology (GO) term enrichment analysis was carried out using the DAVID Bioinformatics Resource server[19]. The false discovery rate (FDR) adjustment on the *P*-values was made using the Benjamini-Hochberg procedure[20].

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[11]. 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.

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 $j=1,\ldots,n_k$ observations in each of $k=1,\ldots,g$ 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 [21]. A robust version of MD was used to quantify the distance from individual sample *i* to the multivariate centroid of all samples: $MD(x_{i\cdot}, x_c) = \sqrt{(x_{i\cdot} - x_c)^T \Psi^{-1} (x_{i\cdot} - x_c)}$, 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[22]; and ψ is the scattering estimator. Let MD_{kj} be a set of $j=1,\ldots,n_k$ observations in each of $k=1,\ldots,g$ age groups, Levene's test literally performs ANOVA on MD_{kj}, given the absolute deviation $z_{kj} = \left|MD_{kj} - \overline{MD}_{k\cdot}\right|$ with the group means $\overline{MD}_{k\cdot} = \frac{1}{n_k}\sum_{j=1}^{n_k} MD_{kj}$.

Results

Identification of brain region-specific age-related genes

From the GTEx portal website, we obtained transcriptomic profiles determined from brain tissues of 169 individuals aged between 20 and 70 years[2]. 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 (**Materials and Methods**).

After controlling for covariates and hidden confounding factors, we used linear regression models to identify genes whose expression is correlated or anti-correlated with chronological age. At the level of FDR of 5%, 1,446 distinct age-related genes were identified across brain regions (**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.

Distinct brain regions exhibit specific “signatures” of sensitivity to aging. Some regions—such as (5) cerebellum—demonstrate a more fluid transcriptome in response to aging, while others—such as (2) anterior cingulate cortex—are less susceptible to gene expression changes associated with aging. **Table 1** presents the numbers of age-related genes in the 13 brain regions at the FDR of 1%, 5%, and 10%. In general, the number of genes increases with the relaxation of FDR cutoff, except that, for (2) anterior cingulate cortex and (12) spinal cord, there were no genes to be identified no matter what FDR cutoff was used. For comparison, in a previous study, using microarray data from ten brain regions of 100 postmortem brains aged 16 – 83 years, Glass, et al. [4] 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 largely consistent directions of expression response to aging. In another previous study, Lu, et al. [1] 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 are among our 5% FDR age-related genes and 98 are among our 10% FDR age-related genes.

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 Figure 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 postmodern donors, while the rest were subsequently resampled after the brains were frozen and stored[2,17]. 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 corrections 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, 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), reinforcing the point that, although there is a considerable discrepancy in identified genes between frozen and unfrozen samples, samples from the same brain region are indeed more similar to each other with respect to the genes' responsiveness to aging. Likewise, (6) and (7) are also correlated with each other and with (2) anterior cingulate cortex, forming a module. Furthermore, the human brain appears to have different aging patterns in the cerebellum, cortex, and basal ganglia (including caudate,

nucleus accumbens, and putamen) (**Fig. 1**). These results are consistent with the findings of a previous microarray-based gene expression study in multiple brain regions[5].

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[11] and applied it to 14,825 functional gene sets defined by GO terms (**Materials and 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 S2**). 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[23] 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 clusters are represented by 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, beta-amyloid binding, and so on (**Supplementary Fig. S2**).

Genotype-by-age interactions

To assess the influence of individuals' genetic background of on gene expression in the brain in response to aging, we conducted the genotype-by-age interaction analysis (**Materials and Methods**). We detected a number of interactions with high significance (nominal $P < 10^{-5}$). A comprehensive list of relevant SNPs and genes is provided (**Supplementary Table S3**), although none of these interactions survived multiple testing corrections due to the sheer large number of tests performed. Nevertheless, 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)[24-28]. 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[29]. C1q contributes to an aging-related decrease in the regenerative capacity of certain tissues[30]. Less C1q, on the other hand, seems to confer some protection against an aging-related decline in function for regions of the brain vulnerable to aging.

Aging affects the population-dispersion of gene expression in brain

Next, we focused on the differences in expression dispersion between age groups to identify differentially variable (DV) genes. Using Levene's test, we identified 848 such 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 S4**). Compared to the distribution of age-related genes in different brain regions (**Table 1**), the distribution of DV genes is more balanced across regions (**Supplementary Table S5**). 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. **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[31]. 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[32-35].

Furthermore, we expanded the utilization of Levene's test, coupled with a distance measure, to a multivariate setting[21] to identify age-related DV gene sets, i.e., sets of genes with significant differential expression dispersion between young and old age groups (**Materials and 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.[36] 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) (**Materials and 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[37,38]. 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[18] 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[1,4].

Region-specific response to aging may reveal the brain region functions

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 in parallel. 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 to identify age-related functional gene sets

We adopted a newly developed pathway-based factor analysis[11] to identify age-related gene sets. The analysis is a two-step approach. The factor analysis method, implemented in PEER[18], 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[39], apoptotic signaling pathway[40], glutathione and vitamin[41], oxidation-reduction process[42], spliceosome complex assembly[43], and neurotransmitter transport[44].

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 has been evidence of the impact of aging on gene expression dispersion. Southworth, et al. [15] showed that the correlation of gene expression within genetic modules decreases in aging mice. Somel, et al. [13] showed that gene expression becomes heterogeneous with age. Li, et al. [12] showed that gene expression variability in the male is age-dependent. Bahar, et al. [14] showed that the cell-to-cell variation in gene expression in heart increases in aging mice [but see ref[45]]. In the present study, we identified a large number of genes whose expression dispersion is age-related. For example, we found that the variance in *MS4A4E* expression in the hippocampus was greatly increased in individuals over age 60; increased expression dispersion could already be appreciated starting around age 56. 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[33,35,46,47]. The increased gene expression variability may also be due to the interaction between *MS4A4E* with other genes, e.g., *CLU*[48]. Furthermore, we argue that the incomplete penetrance observed in neurodegenerative diseases[49,50] may be attributed to differences in the phenotypic robustness, which may be associated with or reflected in the age-related gene expression variability, among individual who carry a mutation develop the disease.

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[51,52]. Previously, we have shown that both common and rare genetic variants may confer regulatory function to contribute to gene expression dispersion[53-55]. In particular, common genetic variants contribute to gene expression variability via distinct modes of action—e.g., epistasis and decanalization[56]. Rare and private regulatory variants have been found to be responsible for extreme gene expression in outlier samples[55,57,58]. Given this background information about the genetic 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[59]. The accumulation of somatic mutations over time in various tissues and organs has been suggested as a general explanation of aging[60-62]. 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[63].

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

We thank Oliver Stegle and Leopold Parts for technical help with PEER. We acknowledge the Texas A&M Institute for Genome Sciences and Society (TIGSS) for providing computing resources and system administration support. This work was supported by the fund of China Scholarship Council to JG, and the National Natural Science Foundation of China (No. 61573296), the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20130121130004), the Fundamental Research Funds for the Central Universities in China (Xiamen University: Nos. 201412G009, 201510384106) to GJ. The Genotype-Tissue Expression (GTEx) Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health. Additional funds were provided by the NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. Donors were enrolled at Biospecimen Source Sites funded by NCI\SAIC-Frederick, Inc. (SAIC-F) subcontracts to the National Disease Research Interchange (10XS170), Roswell Park Cancer Institute (10XS171), and Science Care, Inc. (X10S172). The Laboratory, Data Analysis, and Coordinating Center (LDACC) was funded through a contract (HHSN268201000029C) to The Broad Institute, Inc. Biorepository operations were funded through an SAIC-F subcontract to Van Andel Institute (10ST1035). Additional data repository and project management were provided by SAIC-F (HHSN261200800001E). The Brain Bank was supported by a supplements to University of Miami grants DA006227 & DA033684 and to contract N01MH000028. Statistical Methods development grants were made to the University of 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.

References

1 Lu, T. *et al.* Gene regulation and DNA damage in the ageing human brain. *Nature* **429**, 883-891, doi:10.1038/nature02661 (2004).

2 GTEx_Consortium. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648-660, doi:10.1126/science.1262110 (2015).

3 Yang, J. *et al.* Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci Rep* **5**, 15145, doi:10.1038/srep15145 (2015).

4 Glass, D. *et al.* Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol* **14**, R75, doi:10.1186/gb-2013-14-7-r75 (2013).

5 Fraser, H. B., Khaitovich, P., Plotkin, J. B., Paabo, S. & Eisen, M. B. Aging and gene expression in the primate brain. *PLoS biology* **3**, e274, doi:10.1371/journal.pbio.0030274 (2005).

6 Oldham, M. C. *et al.* Functional organization of the transcriptome in human brain. *Nat Neurosci* **11**, 1271-1282, doi:10.1038/nn.2207 (2008).

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

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

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

10 Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130-133, doi:10.1126/science.1134108 (2006).

11 Anand Brown, A. *et al.* Pathway-based factor analysis of gene expression data produces highly heritable phenotypes that associate with age. *G3 (Bethesda)* **5**, 839-847, doi:10.1534/g3.114.011411 (2015).

12 Li, Z., Wright, F. A. & Royland, J. Age-dependent variability in gene expression in male Fischer 344 rat retina. *Toxicol Sci* **107**, 281-292, doi:10.1093/toxsci/kfn215 (2009).

13 Somel, M., Khaitovich, P., Bahn, S., Paabo, S. & Lachmann, M. Gene expression becomes heterogeneous with age. *Curr Biol* **16**, R359-360, doi:10.1016/j.cub.2006.04.024 (2006).

14 Bahar, R. *et al.* Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**, 1011-1014, doi:10.1038/nature04844 (2006).

15 Southworth, L. K., Owen, A. B. & Kim, S. K. Aging mice show a decreasing correlation of gene expression within genetic modules. *PLoS genetics* **5**, e1000776, doi:10.1371/journal.pgen.1000776 (2009).

16 Groelz, D. *et al.* Non-formalin fixative versus formalin-fixed tissue: a comparison of histology and RNA quality. *Exp Mol Pathol* **94**, 188-194, doi:10.1016/j.yexmp.2012.07.002 (2013).

17 Carithers, L. J. *et al.* A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank* **13**, 311-319, doi:10.1089/bio.2015.0032 (2015).

18 Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500-507, doi:10.1038/nprot.2011.457 (2012).

19 Dennis, G., Jr. *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* **4**, P3 (2003).
20 Benjamini, Y. & Hochberg, Y. 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, doi:10.2307/2346101 (1995).
21 Anderson, M. J. Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* **62**, 245-253, doi:10.1111/j.1541-0420.2005.00440.x (2006).
22 Rousseeuw, P. J. & Van Driessen, K. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**, 212-223, doi:Doi 10.2307/1270566 (1999).
23 Supek, F., Bosnjak, M., Skunca, N. & Smuc, T. REVIGO summarizes and visualizes long lists of gene ontology terms. *PloS one* **6**, e21800, doi:10.1371/journal.pone.0021800 (2011).
24 Pozniak, C. D. *et al.* p73 is required for survival and maintenance of CNS neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **22**, 9800-9809 (2002).
25 Pozniak, C. D. *et al.* An anti-apoptotic role for the p53 family member, p73, during developmental neuron death. *Science* **289**, 304-306 (2000).
26 Yang, A. *et al.* p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. *Nature* **404**, 99-103, doi:10.1038/35003607 (2000).
27 Li, Q. *et al.* TP73 allelic expression in human brain and allele frequencies in Alzheimer's disease. *BMC Med Genet* **5**, 14, doi:10.1186/1471-2350-5-14 (2004).
28 Wetzel, M. K. *et al.* p73 regulates neurodegeneration and phospho-tau accumulation during aging and Alzheimer's disease. *Neuron* **59**, 708-721, doi:10.1016/j.neuron.2008.07.021 (2008).
29 Stephan, A. H. *et al.* A dramatic increase of C1q protein in the CNS during normal aging. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **33**, 13460-13474, doi:10.1523/JNEUROSCI.1333-13.2013 (2013).
30 Naito, A. T. *et al.* Complement C1q activates canonical Wnt signaling and promotes aging-related phenotypes. *Cell* **149**, 1298-1313, doi:10.1016/j.cell.2012.03.047 (2012).
31 Passtoors, W. M. *et al.* IL7R gene expression network associates with human healthy ageing. *Immun Ageing* **12**, 21, doi:10.1186/s12979-015-0048-6 (2015).
32 Liang, Y., Buckley, T. R., Tu, L., Langdon, S. D. & Tedder, T. F. Structural organization of the human MS4A gene cluster on Chromosome 11q12. *Immunogenetics* **53**, 357-368 (2001).
33 Hollingworth, P. *et al.* Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* **43**, 429-435, doi:10.1038/ng.803 (2011).
34 Karagiannis, S. N. *et al.* Activity of human monocytes in IgE antibody-dependent surveillance and killing of ovarian tumor cells. *Eur J Immunol* **33**, 1030-1040, doi:10.1002/eji.200323185 (2003).
35 Naj, A. C. *et al.* Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* **43**, 436-441, doi:10.1038/ng.801 (2011).
36 Lin, J. *et al.* 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, doi:10.1093/nar/gkt1132 (2014).
37 Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS genetics* **3**, 1724-1735, doi:10.1371/journal.pgen.0030161 (2007).

38 Parts, L., Stegle, O., Winn, J. & Durbin, R. Joint genetic analysis of gene expression data with inferred cellular phenotypes. *PLoS genetics* **7**, e1001276, doi:10.1371/journal.pgen.1001276 (2011).
39 Feser, J. & Tyler, J. Chromatin structure as a mediator of aging. *FEBS Lett* **585**, 2041-2048, doi:10.1016/j.febslet.2010.11.016 (2011).
40 Harman, D. Role of free radicals in aging and disease. *Annals of the New York Academy of Sciences* **673**, 126-141 (1992).
41 Nuttall, S. L., Martin, U., Sinclair, A. J. & Kendall, M. J. Glutathione: in sickness and in health. *Lancet* **351**, 645-646 (1998).
42 Berlett, B. S. & Stadtman, E. R. Protein oxidation in aging, disease, and oxidative stress. *J Biol Chem* **272**, 20313-20316 (1997).
43 Rodriguez, S. A. *et al.* Global genome splicing analysis reveals an increased number of alternatively spliced genes with aging. *Aging Cell*, doi:10.1111/acel.12433 (2015).
44 Segovia, G., Porras, A., Del Arco, A. & Mora, F. Glutamatergic neurotransmission in aging: a critical perspective. *Mech Ageing Dev* **122**, 1-29 (2001).
45 Warren, L. A. *et al.* Transcriptional instability is not a universal attribute of aging. *Aging Cell* **6**, 775-782, doi:10.1111/j.1474-9726.2007.00337.x (2007).
46 Akiyama, H. *et al.* Inflammation and Alzheimer's disease. *Neurobiol Aging* **21**, 383-421 (2000).
47 Heppner, F. L., Ransohoff, R. M. & Becher, B. Immune attack: the role of inflammation in Alzheimer disease. *Nature reviews. Neuroscience* **16**, 358-372, doi:10.1038/nrn3880 (2015).
48 Ebbert, M. T. *et al.* Interaction between variants in CLU and MS4A4E modulates Alzheimer's disease risk. *Alzheimers Dement*, doi:10.1016/j.jalz.2015.08.163 (2015).
49 Rossor, M. N., Fox, N. C., Beck, J., Campbell, T. C. & Collinge, J. Incomplete penetrance of familial Alzheimer's disease in a pedigree with a novel presenilin-1 gene mutation. *Lancet* **347**, 1560 (1996).
50 Healy, D. G. *et al.* Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol* **7**, 583-590, doi:10.1016/S1474-4422(08)70117-0 (2008).
51 Lemos, B., Meiklejohn, C. D., Caceres, M. & Hartl, D. L. Rates of divergence in gene expression profiles of primates, mice, and flies: stabilizing selection and variability among functional categories. *Evolution* **59**, 126-137 (2005).
52 Khaitovich, P. *et al.* A neutral model of transcriptome evolution. *PLoS Biol* **2**, E132, doi:10.1371/journal.pbio.0020132 (2004).
53 Hulse, A. M. & Cai, J. J. Genetic variants contribute to gene expression variability in humans. *Genetics* **193**, 95-108, doi:10.1534/genetics.112.146779 (2013).
54 Wang, G., Yang, E., Brinkmeyer-Langford, C. L. & Cai, J. J. Additive, epistatic, and environmental effects through the lens of expression variability QTL in a twin cohort. *Genetics* **196**, 413-425, doi:10.1534/genetics.113.157503 (2014).
55 Zeng, Y. *et al.* Aberrant gene expression in humans. *PLoS genetics* **11**, e1004942, doi:10.1371/journal.pgen.1004942 (2015).
56 Wang, G. *et al.* Epistasis and decanalization shape gene expression variability in humans via distinct modes of action. *bioRxiv* **doi: http://dx.doi.org/10.1101/026393**, doi:10.1101/026393 (2015).
57 Zhao, J. *et al.* A Burden of Rare Variants Associated with Extremes of Gene Expression in Human Peripheral Blood. *Am J Hum Genet* **98**, 299-309, doi:10.1016/j.ajhg.2015.12.023 (2016).
58 Montgomery, S. B., Lappalainen, T., Gutierrez-Arcelus, M. & Dermitzakis, E. T. Rare and common regulatory variation in population-scale sequenced human genomes. *PLoS genetics* **7**, e1002144, doi:10.1371/journal.pgen.1002144 (2011).

59 Busuttil, R., Bahar, R. & Vijg, J. Genome dynamics and transcriptional deregulation in aging. *Neuroscience* **145**, 1341-1347, doi:10.1016/j.neuroscience.2006.09.060 (2007).
60 Szilard, L. On the Nature of the Aging Process. *Proc Natl Acad Sci U S A* **45**, 30-45 (1959).
61 Curtis, H. J. Biological mechanisms underlying the aging process. *Science* **141**, 686-694 (1963).
62 Vijg, J. Impact of genome instability on transcription regulation of aging and senescence. *Mech Ageing Dev* **125**, 747-753, doi:10.1016/j.mad.2004.07.004 (2004).
63 Busuttil, R. A. *et al.* Intra-organ variation in age-related mutation accumulation in the mouse. *PloS one* **2**, e876, doi:10.1371/journal.pone.0000876 (2007).

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.

Supplementary Figure Legends

Fig. S1. Distance matrix between 13 GTEx brain regions in gene expression. The Euclidean distance was calculated between average levels of protein-coding genes of different regions.

Fig. S2. TreeMap view of MF GO-term clusters for age-related genes in the brain.

Supplementary Table Legends

Table S1. List of age-related genes in GTEx brain regions at the level of FDR < 5%.

Table S2. List of GO-term defined gene sets most significantly associated with age together with the total number of genes in the set, the number of genes within sets significantly associated with age ($P < 0.05$, corrected using Bonferroni for the total number of genes in the pathway).

Table S3. List of SNPs and genes whose expressions in (6) cortex (PAXgene) show genotype-by-age interaction at the significant level of a nominal $P < 1E-6$.

Table S4. List of genes and gene sets showing significant changes in gene expression dispersion between age groups at the significance level of FDR < 5%.

Table S5. Numbers of genes showing significant changes in gene expression dispersion between age groups in GTEx brain regions.

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.

	FDR								
	1%			5%			10%		
Brain Region	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	Fold Enrichment	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

Fig. 1.

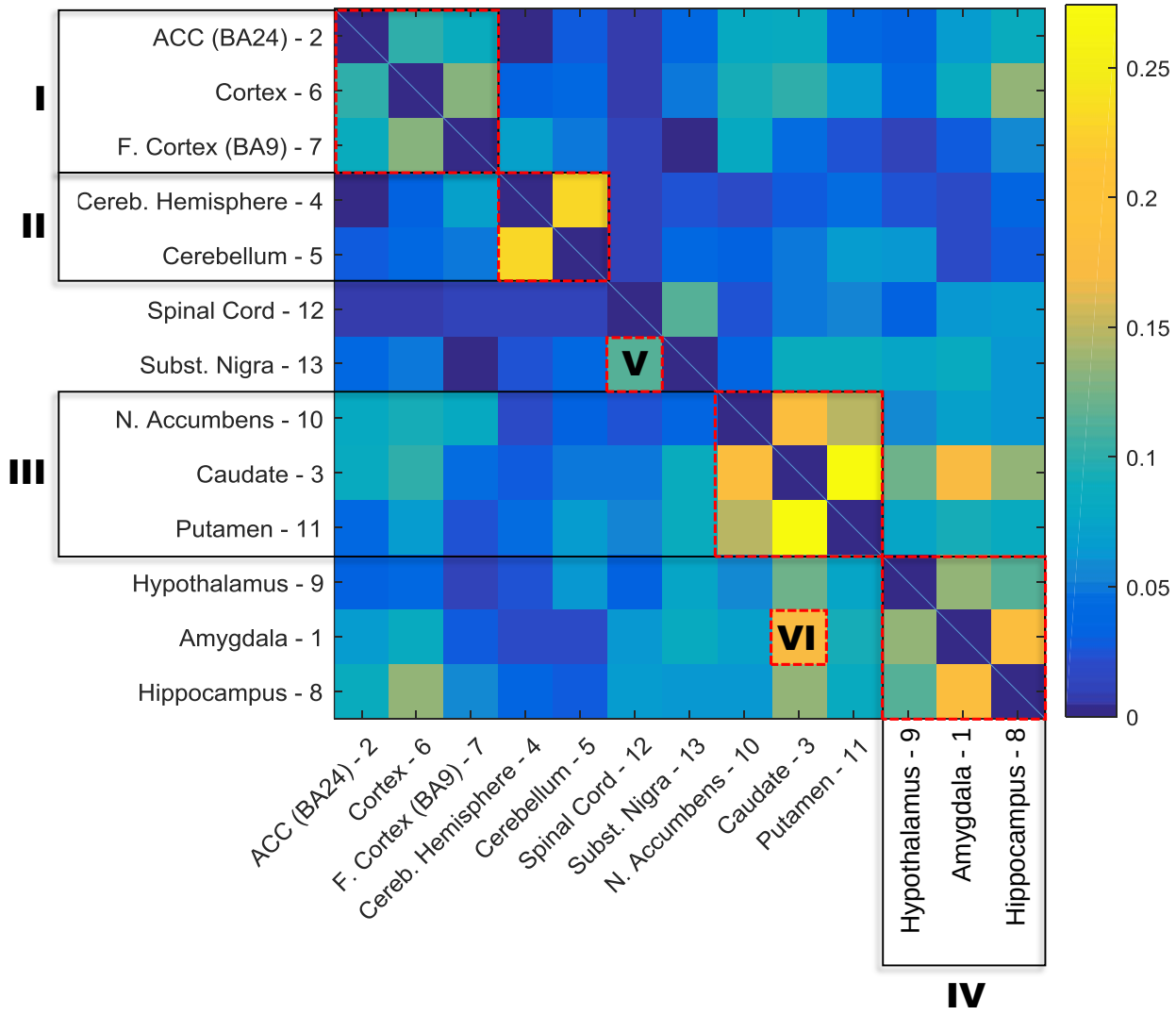

I
II
III
IV
V
VI
ACC (BA24) - 2
Cortex - 6
F. Cortex (BA9) - 7
Cereb. Hemisphere - 4
Cerebellum - 5
Spinal Cord - 12
Subst. Nigra - 13
N. Accumbens - 10
Caudate - 3
Putamen - 11
Hypothalamus - 9
Amygdala - 1
Hippocampus - 8
0.25
0.2
0.15
0.1
0.05
0

Fig. 2.

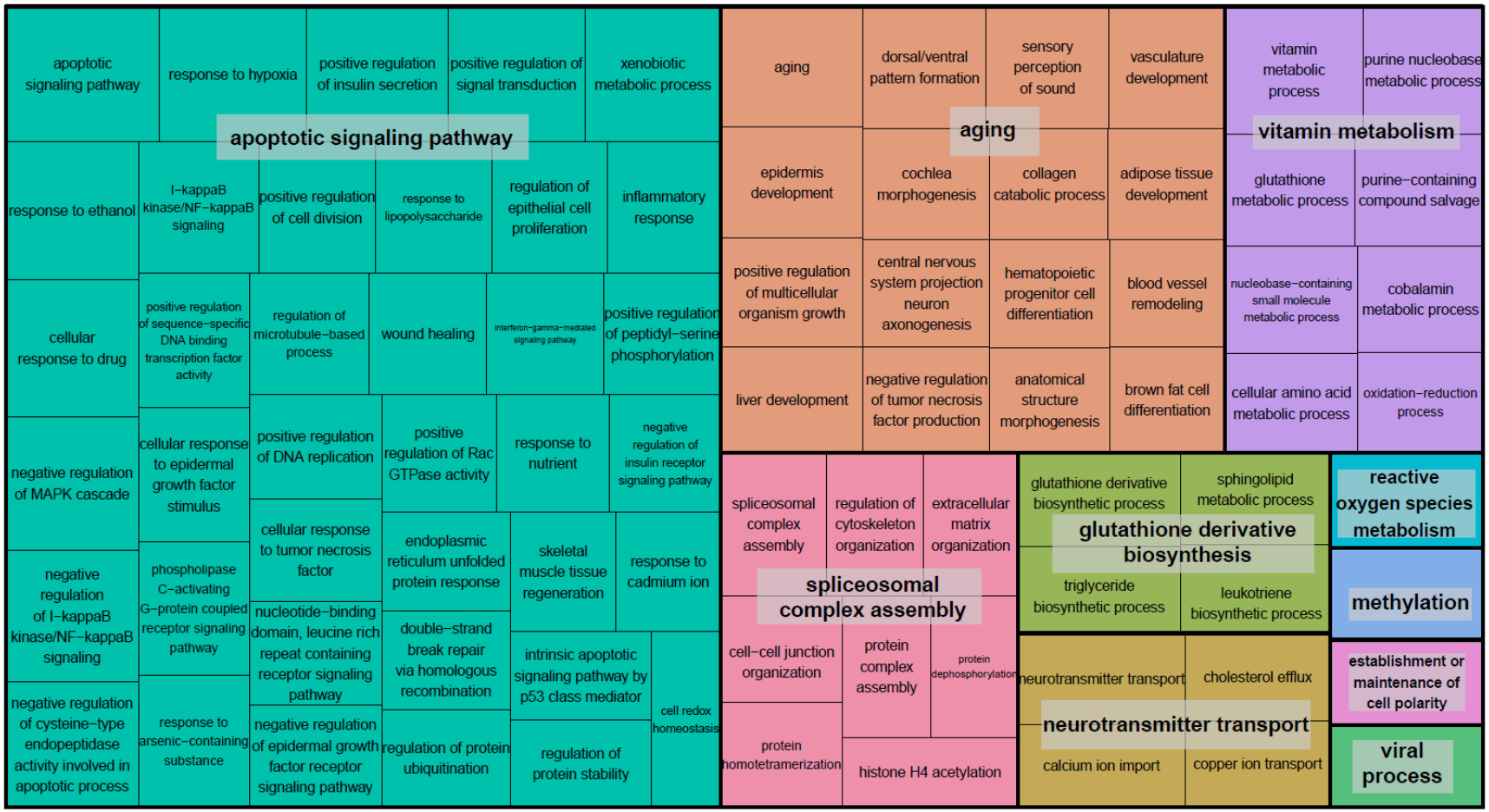

apoptotic signaling pathway
response to hypoxia
positive regulation of insulin secretion
positive regulation of signal transduction
xenobiotic metabolic process
apoptotic signaling pathway
response to ethanol
I−kappaB kinase/NF−kappaB signaling
positive regulation of cell division
response to lipopolysaccharide
regulation of epithelial cell proliferation
inflammatory response
cellular response to drug
positive regulation of sequence−specific DNA binding transcription factor activity
regulation of microtubule−based process
wound healing
interferon−gamma−mediated signaling pathway
positive regulation of peptidyl−serine phosphorylation
negative regulation of MAPK cascade
cellular response to epidermal growth factor stimulus
positive regulation of DNA replication
positive regulation of Rac GTPase activity
response to nutrient
negative regulation of insulin receptor signaling pathway
negative regulation of I−kappaB kinase/NF−kappaB signaling
phospholipase C−activating G−protein coupled receptor signaling pathway
cellular response to tumor necrosis factor
endoplasmic reticulum unfolded protein response
skeletal muscle tissue regeneration
response to cadmium ion
negative regulation of cysteine−type endopeptidase activity involved in apoptotic process
response to arsenic−containing substance
nucleotide−binding domain, leucine rich repeat containing receptor signaling pathway
double−strand break repair via homologous recombination
intrinsic apoptotic signaling pathway by p53 class mediator
cell redox homeostasis
negative regulation of epidermal growth factor receptor signaling pathway
regulation of protein ubiquitination
regulation of protein stability
aging
dorsal/ventral pattern formation
sensory perception of sound
vasculature development
aging
epidermis development
cochlea morphogenesis
collagen catabolic process
adipose tissue development
positive regulation of multicellular organism growth
central nervous system projection neuron axonogenesis
hematopoietic progenitor cell differentiation
blood vessel remodeling
liver development
negative regulation of tumor necrosis factor production
anatomical structure morphogenesis
brown fat cell differentiation
vitamin metabolic process
purine nucleobase metabolic process
vitamin metabolism
glutathione metabolic process
purine−containing compound salvage
nucleobase−containing small molecule metabolic process
cobalamin metabolic process
cellular amino acid metabolic process
oxidation−reduction process
spliceosomal complex assembly
regulation of cytoskeleton organization
extracellular matrix organization
spliceosomal complex assembly
cell−cell junction organization
protein complex assembly
protein dephosphorylation
protein homotetramerization
histone H4 acetylation
glutathione derivative biosynthetic process
sphingolipid metabolic process
glutathione derivative biosynthesis
triglyceride biosynthetic process
leukotriene biosynthetic process
neurotransmitter transport
cholesterol efflux
neurotransmitter transport
calcium ion import
copper ion transport
reactive oxygen species metabolism
methylation
establishment or maintenance of cell polarity
viral process

Fig. 3.

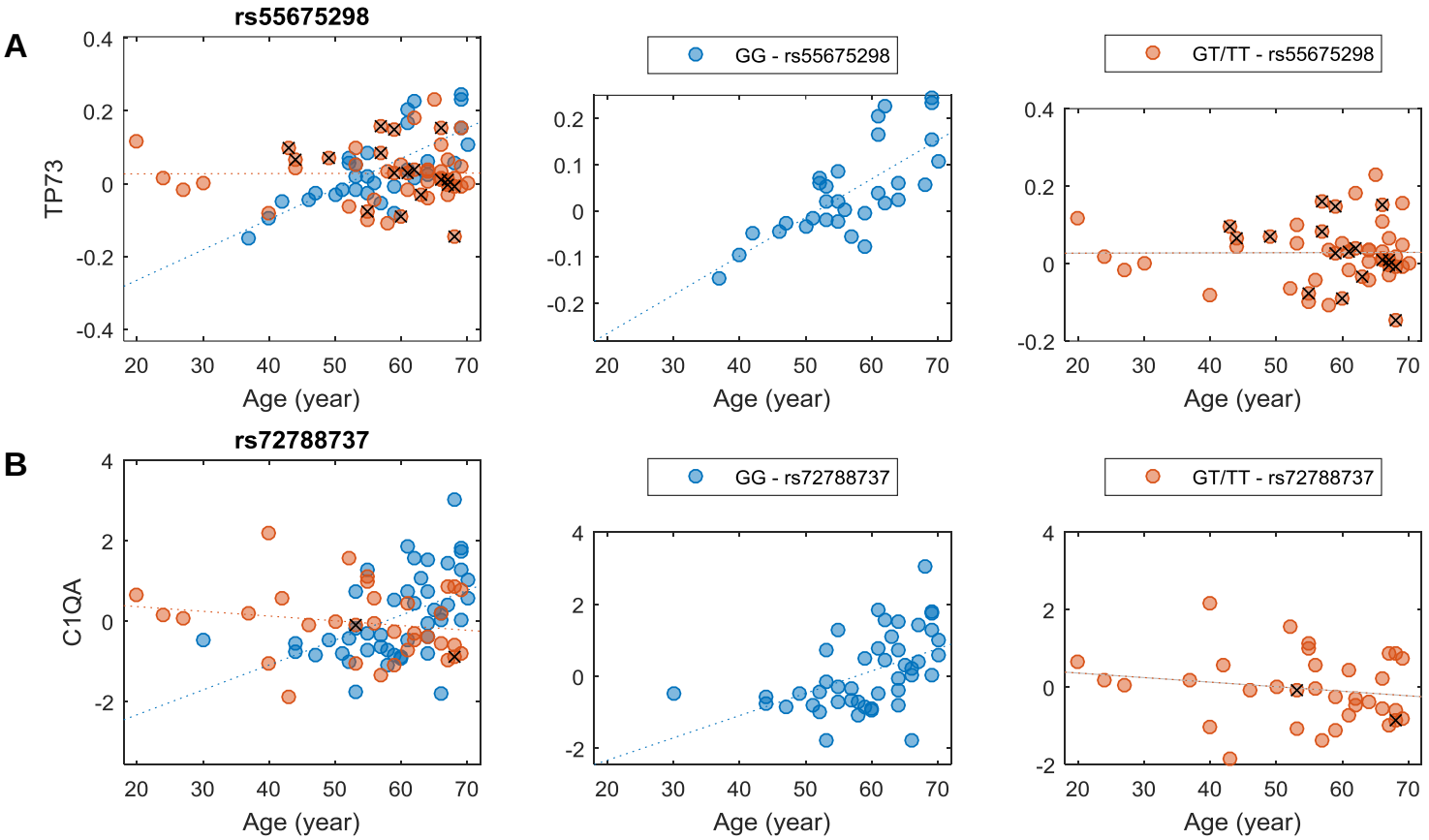
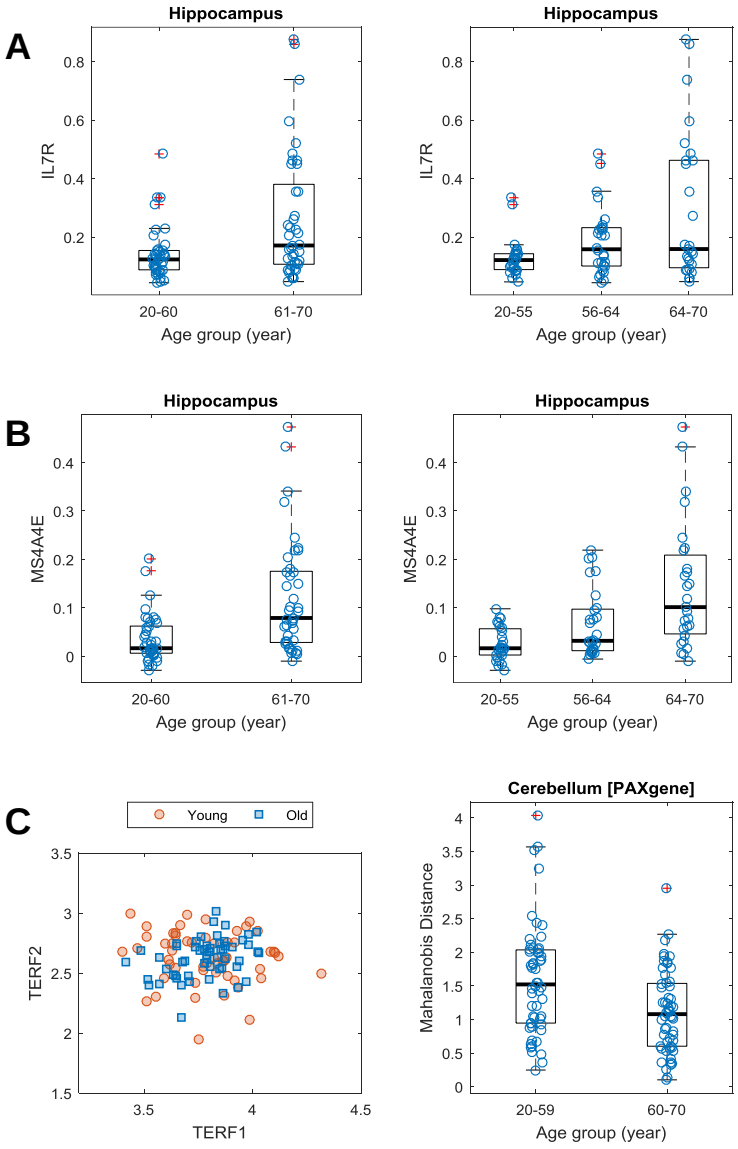

A
rs55675298
TP73
Age (year)
GG - rs55675298
GT/TT - rs55675298
B
rs72788737
C1QA
Age (year)
GG - rs72788737
GT/TT - rs72788737

Fig. 4.


A
Hippocampus
IL7R
0.8
0.6
0.4
0.2
20-60
61-70
Age group (year)
Hippocampus
IL7R
0.8
0.6
0.4
0.2
20-55
56-64
64-70
Age group (year)
B
Hippocampus
MS4A4E
0.4
0.3
0.2
0.1
0
20-60
61-70
Age group (year)
Hippocampus
MS4A4E
0.4
0.3
0.2
0.1
0
20-55
56-64
64-70
Age group (year)
C
Young
Old
TERF2
3.5
3
2.5
2
1.5
3.5
4
4.5
TERF1
Cerebellum [PAXgene]
Mahalanobis Distance
4
3.5
3
2.5
2
1.5
1
0.5
0
20-59
60-70
Age group (year)