



Aging Shapes the Population-Mean and -Dispersion of Gene Expression in Human Brains

Candice L. Brinkmeyer-Langford^{1†}, Jinting Guan^{2†}, Guoli Ji^{2,3*} and James J. Cai^{1,4*}

¹ Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX, USA, ² Department of Automation, Xiamen University, Xiamen, China, ³ Innovation Center for Cell Signaling Network, Xiamen University, Xiamen, China, ⁴ Interdisciplinary Program in Genetics, Texas A&M University, College Station, TX, USA

OPEN ACCESS

Edited by:

Ashok Kumar,
University of Florida, USA

Reviewed by:

Nuno Raimundo,
University of Göttingen, Germany
Jeremy J. Day,
University of Alabama at Birmingham,
USA

*Correspondence:

Guoli Ji
gjji@xmu.edu.cn
James J. Cai
jcai@tamu.edu

[†]These authors have contributed
equally to this work.

Received: 19 May 2016

Accepted: 15 July 2016

Published: 03 August 2016

Citation:

Brinkmeyer-Langford CL, Guan J, Ji G
and Cai JJ (2016) Aging Shapes the
Population-Mean and -Dispersion of
Gene Expression in Human Brains.
Front. Aging Neurosci. 8:183.
doi: 10.3389/fnagi.2016.00183

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 to 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 1446 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.

Keywords: expression dispersion, Genotype-Tissue Expression (GTEx), expression quantitative trait loci, brain transcriptome, gene expression, gene expression profiling, factor analysis, statistical

INTRODUCTION

Aging is a natural process, 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 particularly affect certain sub-regions of the brain (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 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 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:

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

TABLE 1 | Numbers of age-related genes in 13 GTEx brain regions.

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

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.

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 2000 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 $j = 1, \dots, n_k$ observations in each of $k = 1, \dots, g$ age groups. Levene's test statistic is the ANOVA F -ratio comparing the g groups, calculated on the

absolute deviations $z_{kj} = |x_{kj} - \bar{x}_k|$, where $\bar{x}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} x_{kj}$ is the group means. To extend the analytical framework to multiple 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 $j = 1, \dots, n_k$ observations in each of $k = 1, \dots, g$ age groups, Levene's test literally performs ANOVA on MD_{kj} , given the absolute deviation $z_{kj} = |MD_{kj} - \overline{MD}_k|$ with the group means $\overline{MD}_k = \frac{1}{n_k} \sum_{j=1}^{n_k} MD_{kj}$.

All methods are summarized in **Supplementary Figure S3**.

RESULTS

Identification of Brain Region-Specific Age-Related Genes

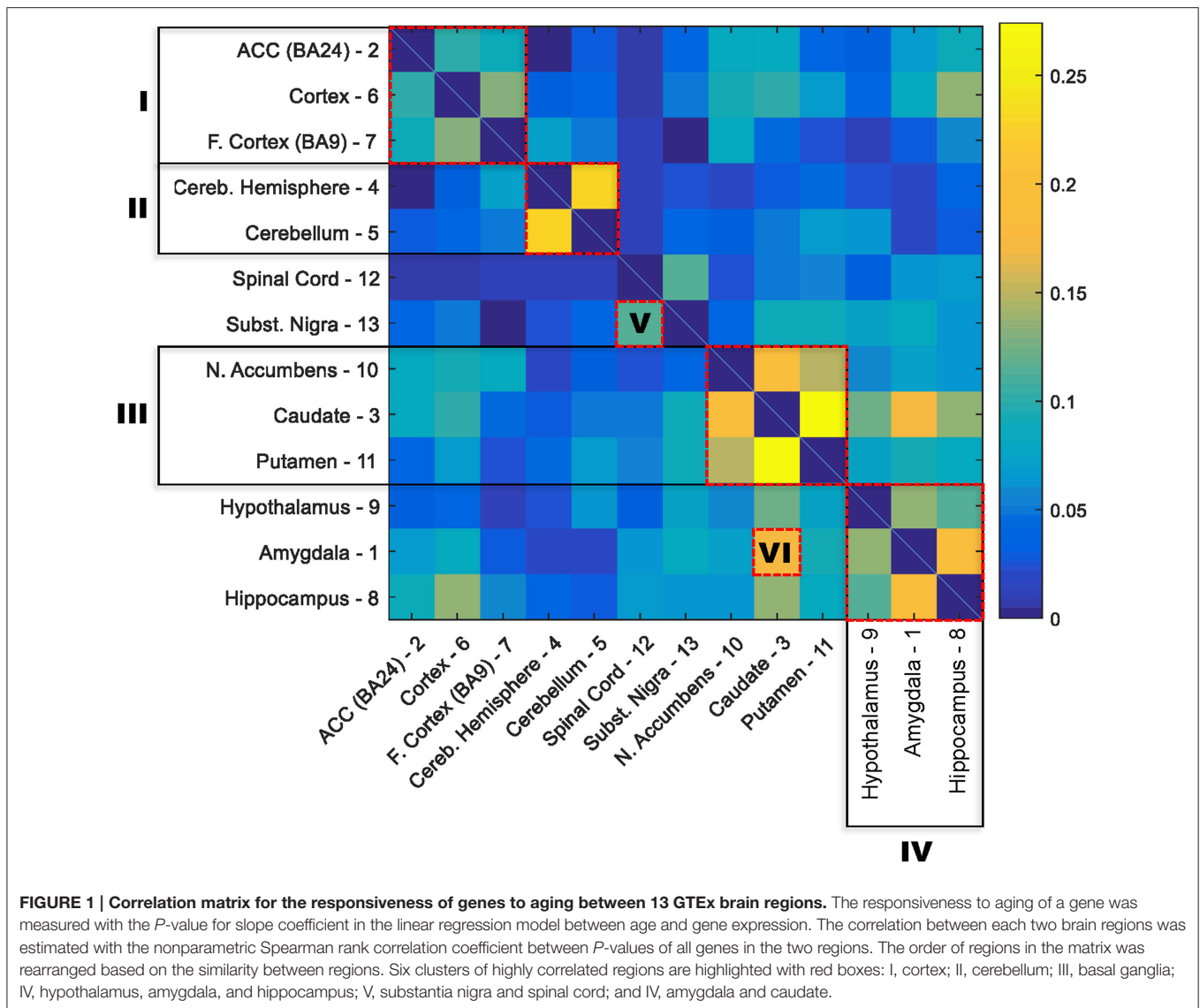
The data of transcriptomic profiles for brain tissues of 169 European ancestry donors aged 20–70 years at death (GTEx Consortium, 2015) was downloaded from the GTEx portal website. The tissue samples were collected from 13 regions (or subareas) of the human brain, namely (1) amygdala, (2) anterior cingulate 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 anterior cingulate cortex and spinal cord where no genes were identified (**Supplementary Figure S4**). At FDR of 5%, 1446 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 10 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*, *MTIG*, *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 cerebellar hemisphere and cerebellum (PAXgene), which is unexpected because tissue samples of these subareas 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 cortex (PAXgene) and frontal cortex (BA9), because both were sampled from the same region of cortex. Indeed, clustering analysis based on the Euclidean distance between gene expression profiles confirmed that cerebellar hemisphere and cerebellum, as well as cortex and frontal cortex, are more similar to each other, respectively, than to other brain regions (**Supplementary Figure S1**). We consider that the markedly fewer genes identified in cerebellar hemisphere than cerebellum, and in frontal cortex than cortex, 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 cerebellum and cortex 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 cerebellar hemisphere and frontal cortex], 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 cerebellar hemisphere and cerebellum and between cortex and frontal cortex were high, then we considered that the discrepancies in age-related gene numbers between cerebellar hemisphere and cerebellum, and between cortex and frontal cortex, were simply due to the effect of freezing on the statistical power of age-related gene detection, rather than on the gene expression regulation. **Figure 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 caudate and putamen, which both belong to basal ganglia, is the highest

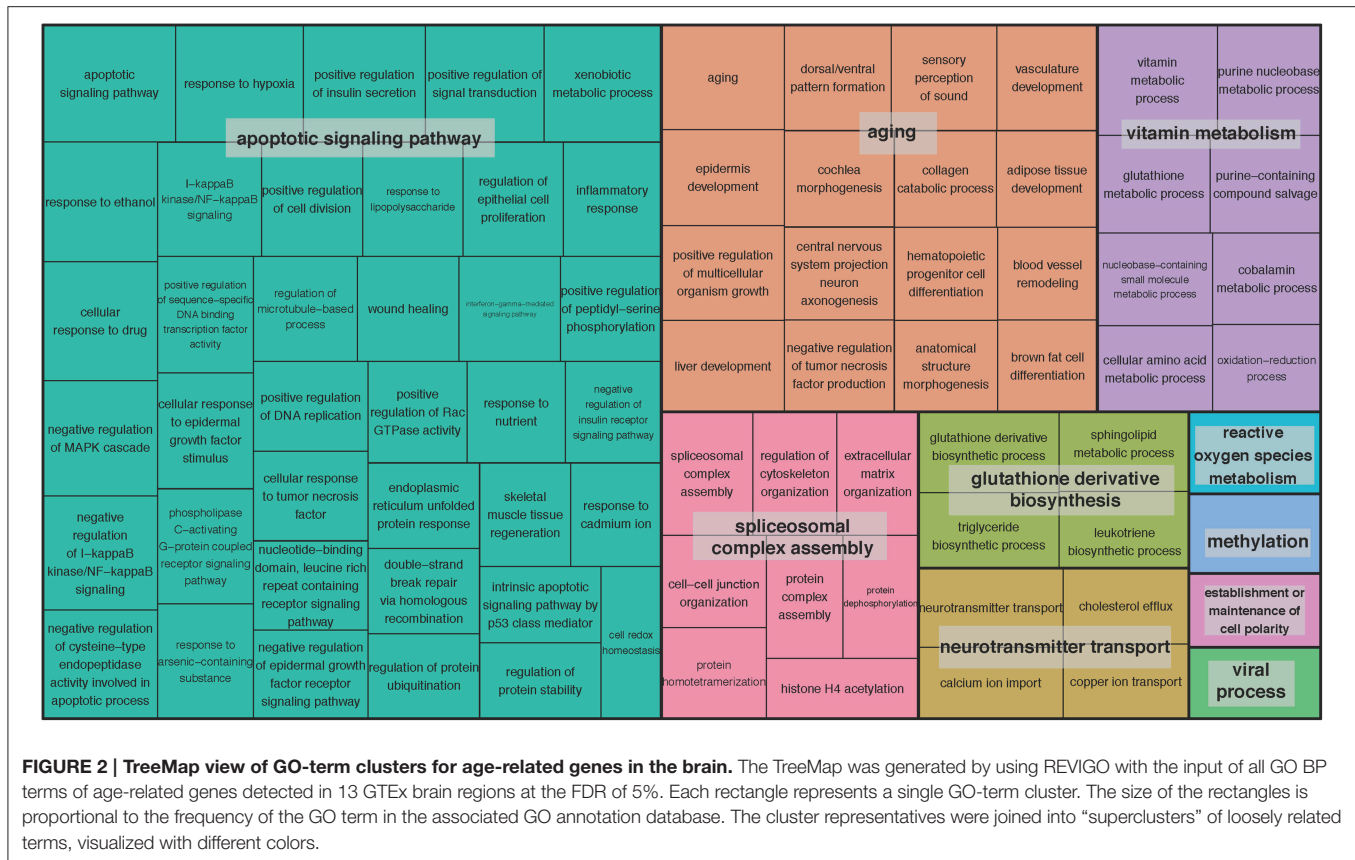


among all region pairs. Intriguingly, the second highest is between cerebellar hemisphere and cerebellum, despite of the considerable discrepancy in identified genes between the frozen and unfrozen cerebellar samples, reinforcing the point that samples from the same brain region are indeed more similar to each other with respect to the genes' responsiveness to aging. Likewise, cortex, frontal cortex and 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 caudate, nucleus accumbens, and putamen; **Figure 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 5787 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) (**Figure 2**) and molecular functions (MF) (**Supplementary Figure S2**). For example, the top-level

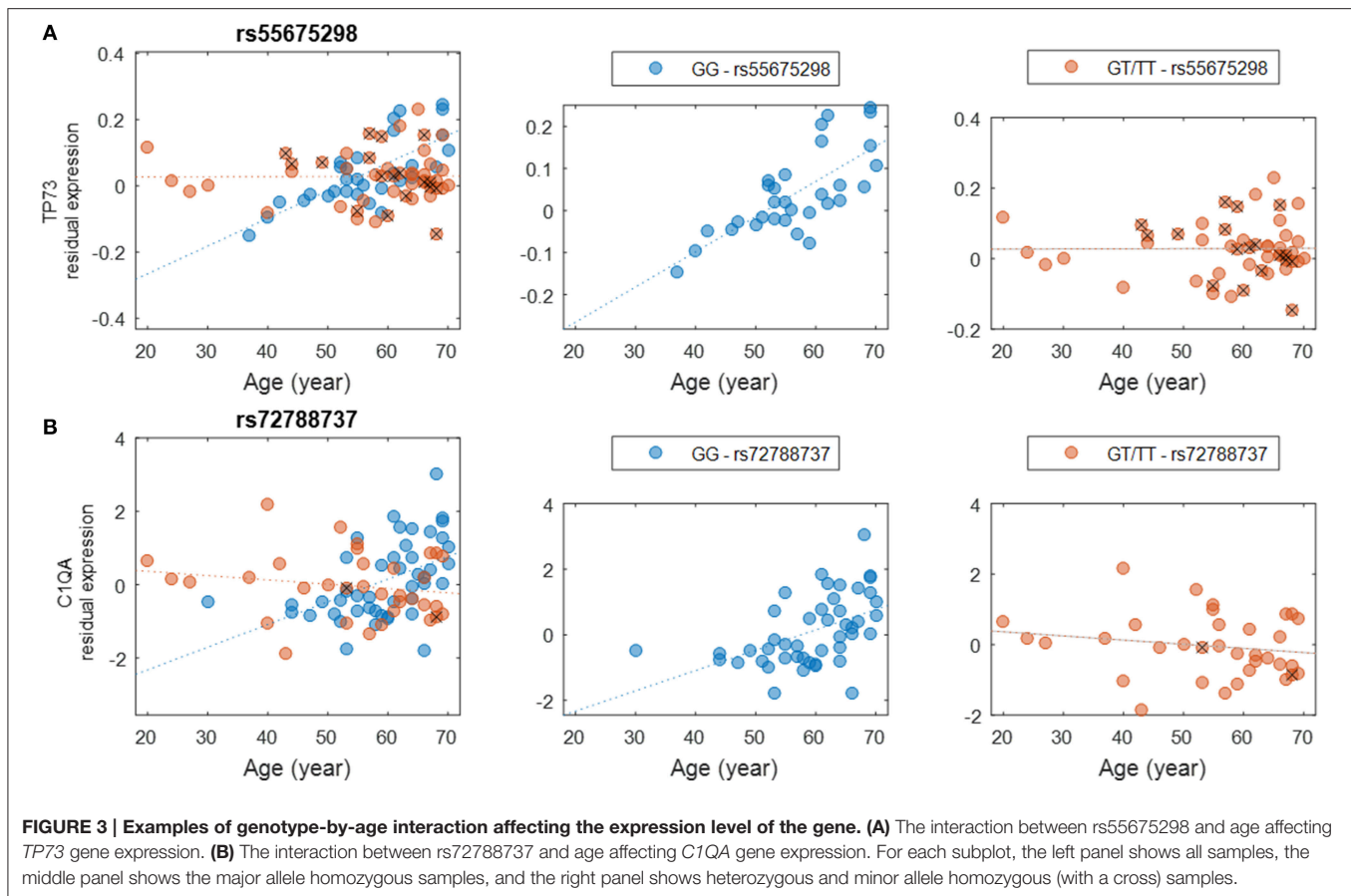


BP GO term clusters are represented by the terms apoptotic signaling pathway, aging, spliceosomal complex assembly, glutathione derivative biosynthesis, neurotransmitter transport, vitamin metabolism, reactive oxygen species metabolism, methylation, establishment or maintenance of cell polarity, and viral process (Figure 2). The top-level MF GO term are listed in Supplementary Figure 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 Figure 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* (Figure 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, 2002; Yang et al., 2000; Li et al., 2004; Wetzel et al., 2008). Another example of a relationship between SNP genotype and age-related gene expression involves *CIQA* and the SNP rs72788737 (Figure 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 *CIQA*), 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). Figures 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 (Figure 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 (Figure 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 MD (Methods). Compared to old specimens, young samples show an increased level of scattering in their

TABLE 2 | GO-term clusters for DV genes showing the most differential variability in expression between age groups.

GO term	Count	Enrichment fold	FDR	Expression variance 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	

TERF1-TERF2 expression, indicated by the higher level of MD (Figure 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 (1446 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

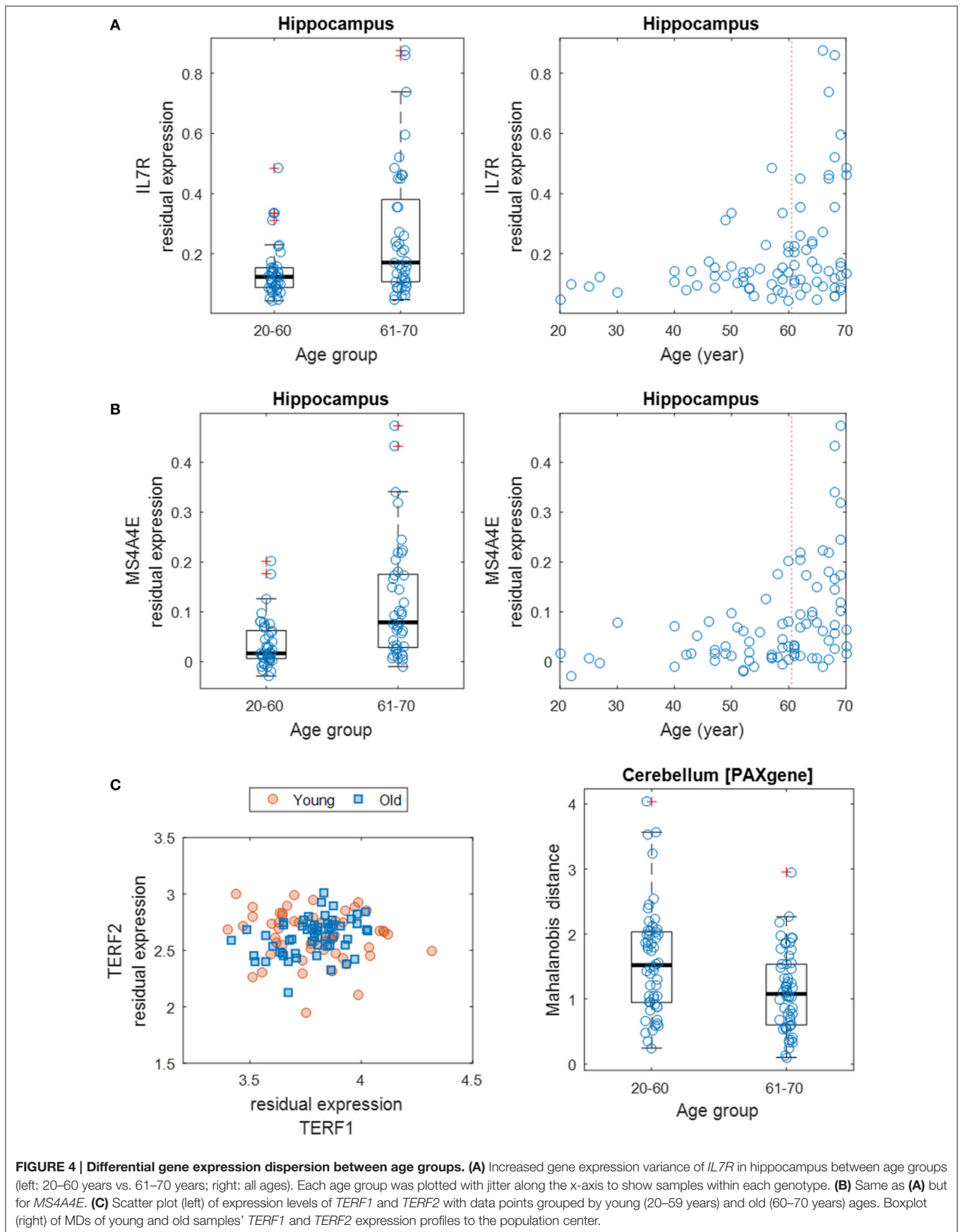


TABLE 3 | DV gene sets showing significant differential variability in brain region-specific expression between age groups (FDR <5%).

GO Term	Gene Set	Expression variance change with age	P-value (Levene's test)	FDR
(5) CEREBELLUM				
GO:0001309~age-dependent telomere shortening	<i>TERF1, TERF2</i>	Decrease	3.80E-06	0.0094
GO:0032214~negative regulation of telomere maintenance via semi-conservative replication	<i>TERF1, TERF2</i>	Decrease	3.80E-06	0.0094
GO:0003691~double-stranded telomeric DNA binding	<i>PURA, TERF1, TERF2, XRCC5, XRCC6</i>	Decrease	2.51E-05	0.0416
(6) CORTEX				
GO:0044062~regulation of excretion	<i>NPHS1, SLC9A3R1</i>	Decrease	6.13E-06	0.0182
GO:0009756~carbohydrate mediated signaling	<i>CLEC7A, COLEC12</i>	Increase	7.77E-06	0.0242
GO:0003840~gamma-glutamyltransferase activity	<i>GGT1, GGT2, GGT5, GGT6, GGT7, GGTL1, GGTL2, GGTL3</i>	Increase	8.07E-05	0.035
GO:0036374~glutathione hydrolase activity	<i>GGT1, GGT5, GGT6, GGT7</i>	Increase	8.07E-05	0.035
(13) SUBSTANTIA NIGRA				
GO:2000146~negative regulation of cell motility	<i>AP1AR, CTNNA1, GATA3, PIN1, TACSTD2</i>	Decrease	5.20E-06	0.036

Genes are grouped into gene sets based on GO terms shared in the functional annotations of these genes.

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 BP and MF. 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 (Rodríguez et al., 2016), 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; 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., 2016). 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 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.

AUTHOR CONTRIBUTIONS

JC conceived and designed the study. CB, JG, GJ, and JC conducted the analyses. CB, JG, GJ, and JC wrote the manuscript. CB and JG contributed equally to this work.

ACKNOWLEDGMENTS

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 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 supplement 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.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fnagi.2016.00183>

Supplementary Figure S1 | Clustering based on the Euclidean distance between gene expression profiles.

Supplementary Figure S2 | TreeMap view of GO-term clusters generated by REVIGO, showing top-level molecular function GO terms of age-related gene sets at 5% FDR significance.

Supplementary Figure S3 | Flowchart summary of methods described in this paper. Output can be found in the figures and tables indicated on the far right-hand side of the figure.

Supplementary Figure S4 | Diagram showing regions of the brain with age-related gene expression differences indicated in parentheses. Here, we show numbers of age-related genes that are up- or down-regulated at FDR 5%. Approximate locations of some regions are delineated with a dashed-line border, as not all regions are located within the same plane.

Supplementary Table S1 | Age-related genes whose expression is correlated or anti-correlated with chronological age, across all regions investigated.

Supplementary Table S2 | Age-related genes identified by randomly subsampling samples to 54 (the minimal sample size) for all regions.

Supplementary Table S3 | Pathway-based factor analysis output showing highly significant gene sets across the 13 brain regions investigated ($P < 0.05$, corrected using Bonferroni procedure for the total number of tested gene sets).

Supplementary Table S4 | SNPs interrogated for interactions between genotype and the effects of aging on the expression of functionally-significant genes.

Supplementary Table S5 | Differentially variable genes, identified using Levene's test, across brain regions (at significance level FDR 5%). The far right-hand column indicates whether expression dispersion decreases with age.

Supplementary Table S6 | Distribution of differentially variable genes across brain regions.

REFERENCES

- Akiyama, H., Barger, S., Barnum, S., Bradt, B., Bauer, J., Cole, G. M., et al. (2000). Inflammation and Alzheimer's disease. *Neurobiol. Aging* 21, 383–421. doi: 10.1016/S0197-4580(00)00124-X
- Anand Brown, A., Ding, Z., Vinuela, A., Glass, D., Parts, L., Spector, T., et al. (2015). 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
- Anderson, M. J. (2006). Distance-based tests for homogeneity of multivariate dispersions. *Biometrics* 62, 245–253. doi: 10.1111/j.1541-0420.2005.00440.x
- Bahar, R., Hartmann, C. H., Rodriguez, K. A., Denny, A. D., Busuttill, R. A., Dolle, M. E., et al. (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* 441, 1011–1014. doi: 10.1038/nature04844
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B-Methodol.* 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. doi: 10.1074/jbc.272.33.20313
- Busuttill, R. A., Garcia, A. M., Reddick, R. L., Dolle, M. E., Calder, R. B., Nelson, J. F., et al. (2007b). Intra-organ variation in age-related mutation accumulation in the mouse. *PLoS ONE* 2:e876. doi: 10.1371/journal.pone.0000876
- Busuttill, R., Bahar, R., and Vijg, J. (2007a). Genome dynamics and transcriptional deregulation in aging. *Neuroscience* 145, 1341–1347. doi: 10.1016/j.neuroscience.2006.09.060
- Carithers, L. J., Ardlie, K., Barcus, M., Branton, P. A., Britton, A., Buia, S. A., et al. (2015). A novel approach to high-quality postmortem tissue procurement: the GTEx project. *Biopreserv. Biobank.* 13, 311–319. doi: 10.1089/bio.2015.0032
- Curtis, H. J. (1963). Biological mechanisms underlying the aging process. *Science* 141, 686–694. doi: 10.1126/science.141.3582.686
- Dennis, G. Jr., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4:P3. doi: 10.1186/gb-2003-4-5-p3
- Ebbert, M. T. W., Boehme, K. L., Wadsworth, M. E., Staley, L. A., Alzheimer's Disease Neuroimaging Initiative, Alzheimer's Disease Genetics Consortium, et al. (2016). Interaction between variants in *CLU* and *MS4A4E* modulates Alzheimer's disease risk. *Alzheimers Dement.* 12, 121–129. doi: 10.1016/j.jalz.2015.08.163
- Erikson, G. A., Bodian, D. L., Rueda, M., Molparia, B., Scott, E. R., Scott-Van Zeeland, A. A., et al. (2016). Whole-genome sequencing of a healthy aging cohort. *Cell* 165, 1002–1011. doi: 10.1016/j.cell.2016.03.022
- Fearnley, J. M., and Lees, A. J. (1991). Ageing and Parkinson's disease: substantia nigra regional selectivity. *Brain* 114 (Pt 5), 2283–2301. doi: 10.1093/brain/114.5.2283
- Feser, J., and Tyler, J. (2011). Chromatin structure as a mediator of aging. *FEBS Lett.* 585, 2041–2048. doi: 10.1016/j.febslet.2010.11.016
- 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. Psychiatr.* 66, 137–147. doi: 10.1136/jnnp.66.2.137
- 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. doi: 10.1371/journal.pbio.0030274
- Glass, D., Vinuela, A., Davies, M. N., Ramasamy, A., Parts, L., Knowles, D., et al. (2013). Gene expression changes with age in skin, adipose tissue, blood and brain. *Genome Biol.* 14:R75. doi: 10.1186/gb-2013-14-7-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. doi: 10.1126/science.3155875
- 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. doi: 10.1016/j.yexmp.2012.07.002
- GTEx Consortium, (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648–660. doi: 10.1126/science.1262110
- Guerreiro, R. J., Beck, J., Gibbs, J. R., Santana, I., Rossor, M. N., Schott, J. M., et al. (2010). Genetic variability in *CLU* and its association with Alzheimer's disease. *PLoS ONE* 5:e9510. doi: 10.1371/journal.pone.0009510
- Harman, D. (1992). Role of free radicals in aging and disease. *Ann. N.Y. Acad. Sci.* 673, 126–141. doi: 10.1111/j.1749-6632.1992.tb27444.x
- Healy, D. G., Falchi, M., O'sullivan, S. S., Bonifati, V., Durr, A., Bressman, S., et al. (2008). 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
- 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. doi: 10.1038/nrn3880
- Hernandez, D. G., Nalls, M. A., Moore, M., Chong, S., Dillman, A., Trabzuni, D., et al. (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. doi: 10.1016/j.nbd.2012.03.020
- Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J. C., Carrasquillo, M. M., et al. (2011). 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
- Hong, S., Beja-Glasser, V. F., Nfonoyim, B. M., Frouin, A., Li, S., Ramakrishnan, S., et al. (2016). Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science* 352, 712–716. doi: 10.1126/science.aad8373
- Hulse, A. M., and Cai, J. J. (2013). Genetic variants contribute to gene expression variability in humans. *Genetics* 193, 95–108. doi: 10.1534/genetics.112.146779
- 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. doi: 10.1126/science.6474172
- Jellinger, K. (1986). "Pathology of parkinsonism," in *Recent Developments in Parkinson's Disease*, eds S. Fahn, C. D. Marsden, P. Jenner and P. Teychenne (New York, NY: Raven), 33–36.
- Kang, H. J., Kawasawa, Y. I., Cheng, F., Zhu, Y., Xu, X., Li, M., et al. (2011). Spatio-temporal transcriptome of the human brain. *Nature* 478, 483–489. doi: 10.1038/nature10523

- Karagiannis, S. N., Wang, Q., East, N., Burke, F., Riffard, S., Bracher, M. G., et al. (2003). 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
- Khaitovich, P., Weiss, G., Lachmann, M., Hellmann, I., Enard, W., Muetzel, B., et al. (2004). A neutral model of transcriptome evolution. *PLoS Biol.* 2:E132. doi: 10.1371/journal.pbio.0020132
- Leek, J. T., and Storey, J. D. (2007). Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* 3, 1724–1735. doi: 10.1371/journal.pgen.0030161
- 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. doi: 10.1111/j.0014-3820.2005.tb00900.x
- Li, Q., Athan, E. S., Wei, M., Yuan, E., Rice, S. L., Vonsattel, J. P., et al. (2004). 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
- 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. doi: 10.1093/toxsci/kfn215
- 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. doi: 10.1007/s002510100339
- Lin, J., Countryman, P., Buncher, N., Kaur, P. E. L., Zhang, Y., Gibson, G., et al. (2014). TRF1 and TRF2 use different mechanisms to bind 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
- Lu, T., Pan, Y., Kao, S. Y., Li, C., Kohane, I., Chan, J., et al. (2004). Gene regulation and DNA damage in the ageing human brain. *Nature* 429, 883–891. doi: 10.1038/nature02661
- Miller, J. A., Ding, S. L., Sunkin, S. M., Smith, K. A., Ng, L., Safer, A., et al. (2014). Transcriptional landscape of the prenatal human brain. *Nature* 508, 199–206. doi: 10.1038/nature13185
- 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. doi: 10.1371/journal.pgen.1002144
- Naito, A. T., Sumida, T., Nomura, S., Liu, M. L., Higo, T., Nakagawa, A., et al. (2012). Complement C1q activates canonical Wnt signaling and promotes aging-related phenotypes. *Cell* 149, 1298–1313. doi: 10.1016/j.cell.2012.03.047
- Naj, A. C., Jun, G., Beecham, G. W., Wang, L. S., Vardarajan, B. N., Buross, J., 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. doi: 10.1038/ng.801
- Nelson, P. T., Wang, W. X., Wilfred, B. R., Wei, A., Dimayuga, J., Huang, Q., et al. (2015). Novel human ABCC9/SUR2 brain-expressed transcripts and an eQTL relevant to hippocampal sclerosis of aging. *J. Neurochem.* 134, 1026–1039. doi: 10.1111/jnc.13202
- Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., et al. (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* 314, 130–133. doi: 10.1126/science.1134108
- Nuttall, S. L., Martin, U., Sinclair, A. J., and Kendall, M. J. (1998). Glutathione: in sickness and in health. *Lancet* 351, 645–646. doi: 10.1016/S0140-6736(05)78428-2
- Oh, G., Ebrahimi, S., Wang, S. C., Cortese, R., Kaminsky, Z. A., Gottesman, I. I., et al. (2016). Epigenetic assimilation in the aging human brain. *Genome Biol.* 17, 76. doi: 10.1186/s13059-016-0946-8
- Oldham, M. C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., et al. (2008). Functional organization of the transcriptome in human brain. *Nat. Neurosci.* 11, 1271–1282. doi: 10.1038/nn.2207
- 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. doi: 10.1371/journal.pgen.1001276
- Passtoors, W. M., Van Den Akker, E. B., Deelen, J., Maier, A. B., Van Der Breggen, R., Jansen, R., et al. (2015). IL7R gene expression network associates with human healthy ageing. *Immun. Ageing* 12, 21. doi: 10.1186/s12979-015-0048-6
- Peters, M. J., Joehanes, R., Pilling, L. C., Schurmann, C., Conneely, K. N., Powell, J., et al. (2015). The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* 6, 8570. doi: 10.1038/ncomms9570
- 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. doi: 10.1126/science.289.5477.304
- Proitsi, P., Lee, S. H., Lunnon, K., Keohane, A., Powell, J., Troakes, C., et al. (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. doi: 10.1016/j.neurobiolaging.2013.08.002
- Rodríguez, S. A., Grochová, D., McKenna, T., Borate, B., Trivedi, N. S., Erdos, M. R., et al. (2016). Global genome splicing analysis reveals an increased number of alternatively spliced genes with aging. *Aging Cell* 15, 267–278. doi: 10.1111/acel.12433
- 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. doi: 10.1016/s0140-6736(96)90715-1
- Rousseuw, P. J., and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41, 212–223. doi: 10.1080/00401706.1999.10485670
- Segovia, G., Porras, A., Del Arco, A., and Mora, F. (2001). Glutamatergic neurotransmission in aging: a critical perspective. *Mech. Ageing Dev.* 122, 1–29. doi: 10.1016/S0047-6374(00)00225-6
- Somel, M., Khaitovich, P., Bahn, S., Paabo, S., and Lachmann, M. (2006). Gene expression becomes heterogeneous with age. *Curr. Biol.* 16, R359–360. doi: 10.1016/j.cub.2006.04.024
- Sood, S., Gallagher, I. J., Lunnon, K., Rullman, E., Keohane, A., Crossland, H., et al. (2015). A novel multi-tissue RNA diagnostic of healthy ageing relates to cognitive health status. *Genome Biol.* 16, 185. doi: 10.1186/s13059-015-0750-x
- 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. doi: 10.1371/journal.pgen.1000776
- 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. doi: 10.1038/nprot.2011.457
- Stephan, A. H., Madison, D. V., Mateos, J. M., Fraser, D. A., Lovelett, E. A., Coullier, L., et al. (2013). A dramatic increase of C1q protein in the CNS during normal aging. *J. Neurosci.* 33, 13460–13474. doi: 10.1523/JNEUROSCI.1333-13.2013
- Supek, F., Bosnjak, M., Skunca, N., and Smuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* 6:e21800. doi: 10.1371/journal.pone.0021800
- Szilard, L. (1959). On the nature of the aging process. *Proc. Natl. Acad. Sci. U.S.A.* 45, 30–45. doi: 10.1073/pnas.45.1.30
- 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. doi: 10.1097/WCO.000000000000069
- Vijg, J. (2004). 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
- 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. doi: 10.1097/00005072-198511000-00003
- Walter, S., Atzmon, G., Demerath, E. W., Garcia, M. E., Kaplan, R. C., Kumari, M., et al. (2011). A genome-wide association study of aging. *Neurobiol. Aging* 32, e2115–e2128. doi: 10.1016/j.neurobiolaging.2011.05.026
- 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. doi: 10.1534/genetics.113.157503

- 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*. doi: 10.1101/026393
- 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. doi: 10.1111/j.1474-9726.2007.00337.x
- Wetzel, M. K., Naska, S., Laliberte, C. L., Rymar, V. V., Fujitani, M., Biernaskie, J. A., et al. (2008). 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
- Yang, A., Walker, N., Bronson, R., Kaghad, M., Oosterwegel, M., Bonnin, J., et al. (2000). p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. *Nature* 404, 99–103. doi: 10.1038/35003607
- Yang, J., Huang, T., Petralia, F., Long, Q., Zhang, B., Argmann, C., et al. (2015). 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
- 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. doi: 10.1038/nn.4063
- 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. doi: 10.1371/journal.pgen.1004942
- Zhao, J., Akinsanmi, I., Arafat, D., Cradick, T. J., Lee, C. M., Banskota, S., et al. (2016). 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
- Zhu, M., Tao, Y., He, Q., Gao, H., Song, F., Sun, Y. M., et al. (2014). Common GSAP promoter variant contributes to Alzheimer's disease liability. *Neurobiol. Aging* 35, e2651–e2657. doi: 10.1016/j.neurobiolaging.2014.05.018

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Brinkmeyer-Langford, Guan, Ji and Cai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.