

Published in final edited form as:

Neurobiol Aging. 2013 April ; 34(4): 1199–1209. doi:10.1016/j.neurobiolaging.2012.10.021.

Age associated changes in gene expression in human brain and isolated neurons

Azad Kumar, PhD^{1,7}, J. Raphael Gibbs, BS^{1,2}, Alexandra Beilina, PhD¹, Allissa Dillman, BS^{1,5}, Ravindran Kumaran, PhD¹, Daniah Trabzuni, MSc^{2,6}, Mina Ryten, PhD², Robert Walker, PhD⁴, Colin Smith, MD⁴, Bryan J. Traynor, MD¹, John Hardy, PhD², Andrew B. Singleton, PhD¹, and Mark R Cookson, PhD^{1,CA}

¹Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA

²Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK

³Department of Medical & Molecular Genetics, King's College London, UK

⁴Department of Neuropathology, The University of Edinburgh, Wilkie Building, Teviot Place, Edinburgh, UK

⁵Department of Neuroscience, Karolinska Institutet, 171 77 Stockholm, Sweden

⁶Department of Genetics, King Faisal Specialist Hospital and Research Centre, PO Box 3354, Riyadh 11211, Saudi Arabia

Abstract

Previous studies have suggested that there are genes whose expression levels are associated with chronological age. However, which genes show consistent age association across studies, and which are specific to a given organism or tissue remains unresolved. Here, we re-assessed this question using two independently ascertained series of human brain samples from two anatomical regions, the frontal lobe of the cerebral cortex and cerebellum. Using microarrays to estimate gene expression, we found sixty associations between expression and chronological age that were statistically significant and were replicated in both series in at least one tissue. There were a greater number of significant associations in the frontal cortex compared to the cerebellum. We then repeated the analysis in a subset of samples using laser capture microdissection to isolate purkinje neurons from the cerebellum. We were able to replicate five gene associations from either frontal cortex or cerebellum in the Purkinje cell dataset, suggesting that there is a subset of genes have robust changes with aging. Of these, the most consistent and strongest association was with expression of *RHBDL3*, a rhomboid protease family member. We confirmed several hits using an independent technique (qRT-PCR) and in an independent published sample series that used a different array platform. We also interrogated larger patterns of age related gene expression using weighted gene correlation network analysis (WGCNA). We found several modules that showed

^{CA}Correspondence to: Mark R. Cookson, Cell Biology and Gene Expression Unit, Laboratory of Neurogenetics, National Institute on Aging, 35 Convent Drive, Bethesda MD 20892-3707, USA. Phone: +1 301 451 3870. Fax: +1 301 451 7295. cookson@mail.nih.gov.

⁷Current address: Vaccine Research Center, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA

Datasets

Expression data is available at GEO, accession number GSE36192 for brain region samples, and accession number GSE37205 for laser captured purkinje cell data.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

significant associations with chronological age and, of these, several that showed negative associations were enriched for genes encoding components of mitochondria. Overall, our results show that there is a distinct and reproducible gene signature for aging in the human brain.

Introduction

Aging is a multidimensional phenomenon where many aspects of function and phenotype in organisms change over time and appears to be a product of both programmed, ie genetic, aspects and stochastic events [1]. The molecular mechanisms underpinning the biology of aging remain poorly defined. Identifying molecular events that vary along the lifespan has implications for understanding the causes of age-related phenotypes. For example, telomeres shorten with aging [2,3] and it has been suggested that telomere shortening plays a causal role in aging [4]. Robust molecular events that occur with aging might also act as markers of the aging process.

Previous studies of aging have examined gene expression using systems approaches, such as microarrays. There are several genes that show linear relationships between expression levels and chronological age in different tissues and species [5–14]. It has also been shown that there are biologically related groups of transcripts that robustly change with age. For example, mitochondrial gene expression is age responsive [15,16] and mitochondria may play a causal role in aging [17].

However, it is often not clear whether gene expression correlations with age are robust when comparing different tissues and organisms. One reason for this is that the magnitude of gene expression changes during aging in adults are relatively small compared to those seen, for example, during development [13]. It additionally appears that different tissues age at different rates and there may be species differences in gene expression in aging [7,10,12,18], which limits the ability to look for concordance across tissues and organisms. For example, a previous study suggested that there are more robust age-related changes in human cerebral cortex compared to the cerebellum and that the patterns of changes differed between humans and chimpanzees [7]. Other studies in human brain have identified differences in age-related gene expression in men and women [19], further demonstrating the complexity of identifying expression markers of aging.

In human tissues, genetic diversity also contributes to variation in gene expression [20]. Furthermore, in some tissues cellular heterogeneity may also impact measurements of gene expression, especially if the cellular composition of tissue changes with aging. In the brain where there are many types of neurons and glia, age-dependent loss of neurons would lead to apparent associations with gene expression. Some previous studies have addressed this by examining general markers for neurons and glia [5] while others have used cell isolation techniques including laser capture microdissection (LCM) to clarify gene expression patterns in Parkinson's disease [21] and in Alzheimer's disease [22]. However, it remains to be clarified to what extent any specific gene expression change with aging can be replicated in neurons, especially in the human brain.

Overall, these considerations suggest that while gene expression may be a marker of the aging process, it would be of interest to further examine age-associated expression in robust well-powered datasets. Our primary aim in the current study was to identify single gene or group of genes that are correlated with aging, preferentially those that show a linear relationship with age. To identify such genes, we analyzed two large series of human brain samples [23,24] for association of age and mRNA expression in the frontal cortex and cerebellum using a single microarray platform by re-arraying previously collected samples. These brain areas were chosen because of previous data in smaller sample series [7]

suggesting differences in the number of age-related associations between the two regions. Additionally, in order to understand whether such changes can be detected specifically in neurons, we repeated the analyses in Purkinje cells isolated by LCM. Finally, we also examined gene networks associated with aging in order to test the hypothesis that networks might robustly detect changes in expression not seen at the level of single genes [25]. Overall, our results support the hypothesis that there are reliable associations between gene expression and chronological age in the human brain.

Materials and Methods

Samples

We performed mRNA expression analyses from two human brain regions, the frontal region of the cerebral cortex and the cerebellum. Two independent series were obtained, a discovery set of 249 subjects from the University of Maryland, USA [20,23] and a replication set from the Edinburgh Sudden Death Brain and Tissue Bank [24]. These brain samples have previously been examined by microarrays [20,24], but on different platforms. Therefore, to provide a consistent dataset, we generated data using a single microarray platform (see below). After quality control, we used 203 samples in the discovery dataset and 73 in the replication dataset. Both sample series included male and female subjects with an age range of 15–91 (mean 35.2, median 30) for the discovery set and a range of 16–83 (mean 50.4, median 52.5) for the replication dataset. Demographic details of the cases are listed in Supplementary table S1.

RNA extraction and processing

Frozen tissue samples of the posterior lobes of the cerebellar cortex, or the superior frontal cerebral cortex (Brodmann area 8 and 46) were obtained from neurologically normal Caucasian subjects. Aliquots of 100–200mg of frozen tissue (predominantly grey matter) were carefully sub-dissected from each tissue for all subjects and used for expression assays. RNA was extracted using Trizol, biotinylated and amplified using the Illumina® TotalPrep-96 RNA Amplification Kit.

For laser-capture microdissection (LCM), we took frozen sections of cerebellum from a subset (n=98; age range 14–72) of samples from the discovery dataset. Tissue was immersed in Shandon M-1 embedding matrix (Thermo Electron Corporation, Rockford, IL) and stored at –80°C until use. Cryostat sections (7–8 µm thick) were cut and stained with Cresyl Violet (Ambion, Austin, TX). Laser-capture microdissection was performed with ArcturusXT microdissection system (Arcturus, Mountain View, CA). Between 70 and 150 excised Purkinje cells were selected from the slide surface and captured on LCM Macro Caps. High-quality cellular RNA was recovered from the collected cells using PicoPure™ RNA isolation kit (Arcturus) and treated with RNase-free DNase (Qiagen, Valencia, CA). The quality of RNA was analyzed using an Agilent 2100 bioanalyzer (Agilent, Foster City, CA). For arrays, we used samples with an RNA integrity number (RIN) greater than 5.0 (range 5–8.5) with two samples excluded because of low RIN. Two rounds of amplification were carried out with the Ambion MessageAmp II aRNA kit.

Microarrays

Amplified RNA from either bulk tissue extracts or LCM Purkinje cells were hybridized onto HumanHT-12_v3 Expression BeadChips (Illumina). These arrays contain 48,804 probes estimating expression of ~25,000 annotated genes from the RefSeq (Build 36.2, release 22) and Unigene (Build 199).

Correlations between gene expression and age

Each sample used in this series has been genotyped for common DNA variants as described elsewhere [20,23]. We used genotype data to perform a principal components analysis for identity by state of genotype and took the first two principal components, PC1 and PC2, from this analysis to estimate overall genetic distance within the sample series. Cubic spline normalization was applied to raw output from array scans, then expression values were corrected for known covariates of gender, post-mortem interval (PMI), principal components PC1 and PC2 from genotyping as above and hybridization batch using multivariate regression as outlined previously [20] except that age was not specified in the model. The values for each of these parameters for each sample are listed in Table S1.

The residuals for expression after covariate correction were then tested against age using linear regression. *P* values were adjusted for multiple testing using a false discovery rate (FDR) correction set at 0.05. Probes were included if detected in >95% of samples in each series; for the frontal cortex this was 8926 usable probes whereas for the cerebellum this was 8824 probes. For the Purkinje cells dataset, only probes that had shown prior evidence of association in the above datasets were tested.

To compare our datasets with an additional sample series, we used data from Colantuoni *et al* [13] downloaded from GEO (accession number GSE30272). We selected the cases (*n*=163) in that series with an overlapping age range to our own dataset (15–80; mean 42.3; median 44.5) and also to the NIMH brain bank. For each probe in the dataset, we performed regressions of age against normalized, surrogate variable corrected [13] expression. *P* values were corrected for multiple testing using the FDR method ($\alpha=0.05$) for all tested probes. To compare these values against our own data, we identified all probes in each sample series that mapped to a unique gene using official gene symbols then matched all available values of *R* and *P* in the three sample series by official gene symbol. A total of 4863 genes could be unambiguously mapped. To confirm that we had accurately mapped to the same gene, we performed BLAT on the probe sequences from confirmed associated hits.

Analysis at a network level was performed using the WGCNA (Weighted gene correlation network analysis) package in R [27]. A set of consensus modules for co-expression were identified using all five datasets, ie for discovery and replication in the frontal cortex and cerebellum as well as the purkinje cell dataset, with one cerebellum sample in the discovery dataset that had excess missing values. We constructed a signed hybrid network using the normalized but non-residualized values for the 6767 probes shared across all datasets, setting a soft power of 6. Covariates as above (PMI, Gender, Age, PC1, PC2 and hybridization batch) were tested for correlation with module membership. Enrichment for Gene Ontology (GO) terms within modules was performed using DAVID [28,29].

Quantitative Reverse transcriptase PCR (qRT-PCR)

RNA from a subset of frontal cortex samples in the discovery series were converted into cDNA using the SuperScript® III First-Strand Synthesis System (Life Technologies). Quantitative RT-PCR was performed using Power Sybr Green (Life Technologies) and primer pairs that amplified each of three genes of interest and the housekeeper gene β -actin (Supplementary table S2). All reactions were run in quadruplicate on a 7900HT Fast Real-time PCR system (Applied Biosystems). Expression levels for genes of interest were calculated following normalisation to β -actin taking into account primer efficiency as previously described [26]. Relative expression levels were then plotted against age after correction for other known covariates as described above.

Results

Linear associations between age and expression for a restricted number of genes

We analyzed the relationship between chronological age and gene expression in two regions of the human brain, the frontal portion of the cerebral cortex and the cerebellum using samples we have previously described [20,23,24]. For consistency between sample series, we generated array data on a single platform. For the frontal cortex, samples were taken from Brodman areas 8 and 46 although it should be noted that expression differences between subregions of the frontal cortex are minimal, with correlation coefficients from expression being estimated as >0.99 in a recent study [30]. Therefore, the results here are likely representative of the frontal cortex as a whole.

We first considered the numbers of, and distribution of effect sizes, for probes that show an association with chronological age in two datasets. Within the two tissue types that we considered, frontal cortex and cerebellum, the estimated standardized effect sizes ranged from -0.6 to 0.6 with similar distributions and direction of effect in both datasets (Fig 1). There was a positive correlation between effect sizes in the discovery and replication datasets in both the frontal cortex (Pearson's $r=0.556$) and cerebellum ($r=0.3293$) that was highly significant ($p < 2 \times 10^{-16}$ in either dataset).

Fifty-six probes in the frontal cortex and five probes in the cerebellum showed significant (FDR adjusted $P < 0.05$) associations with aging that were replicated in both datasets (Supplementary Table S3). Of the 56 probes that were significant in the frontal cortex in both datasets, seven were not detected in cerebellum. The remaining 49 probes were detected but did not show significant associations in one or both of the cerebellum datasets. For example, *C2CD2* was associated with aging in the frontal cortex ($R^2=0.341$, $P_{\text{adjusted}} < 10^{-17}$ in the discovery set; $R^2=0.167$, $P_{\text{adjusted}}=0.048$ replication) but not in the cerebellum ($R^2=0.014$ and $R^2 < 0.01$ in discovery and replication) despite adequate detection of expression. For the cerebellum, two additional probes (*PENK* and *COL8A2*) were significant that were not detected in the cortex. Two probes, mapping to *NR3C2* and *TMEM158* were replicated in the cerebellum and in the discovery dataset in frontal cortex but not in the replication dataset for frontal cortex. Considering each tissue separately, for all of the probes that were significant and replicated in both datasets (56 for the frontal cortex and 5 for the cerebellum), the direction of the correlations with age was identical for both discovery and replication datasets. We repeated this analysis with a less stringent set of probes, namely those that were significant ($P_{\text{adjusted}} < 0.05$) in the discovery dataset without reference to the replication dataset, comparing the direction of association for probes greater than the least significant probe in the discovery dataset (an effect size of 0.15; see Supplementary Table S4). Again, this indicated an excess of correlations in similar directions compared to opposite direction results (Supplementary Fig. S1), suggesting consistent effects for associations that did not reach the formal level of statistical support for significance in both datasets.

Collectively, these results suggest that there are a number of genes that are significantly associated with chronological age in the brain with both positive and negative associations identified. In agreement with previous studies [7], detection of age associations is more robust in the frontal cortex than in the cerebellum. We next performed a series of independent experiments to validate our initial findings.

Replication of age-associated effects in isolated neurons

One difficulty with examining gene expression in a heterogeneous tissue such as brain is that different cells have different gene expression patterns. To address whether changes with age occur specifically in neurons, we isolated Purkinje cells from the cerebellum of a subset

($n=98$; age range 14–72) of cases in the discovery dataset, collecting >70 individual cells per case. We chose Purkinje cells because they can be easily identified using simple rapid staining in frozen sections. Because the quality of RNA may be lower from these cresyl violet stained sections, we measured RNA integrity using a Biolalyzer and only used those with an RNA integrity number (RIN) of more than 5.0 (Supplementary Fig. S2). The range of RNA quality is likely less than in bulk tissue and, although sufficient to support array experiments, may introduce additional variance in our estimates of gene expression (see Discussion). Due to the limited recoveries of RNA from small numbers of cells we used two rounds of amplification for the LCM samples compared to a single round for the bulk tissue extracts. We performed gene expression arrays and despite differences in RNA processing protocols, the overall distribution of gene expression was similar between Purkinje cells and bulk cerebellum but we were able to enrich for Purkinje cell markers such as *PCP2* [31] and diminish glial markers such as GFAP whilst retaining expression of housekeeping genes (Supplementary Fig. S2).

We then tested the association of age with gene expression. To minimize the effects of decreased power due to the lower numbers of cases used, we only tested those genes that had been nominated as associated with age in the experiments using bulk extracted tissue. From the list of 60 genes associated with aging in either frontal cortex or cerebellum but that were replicated in both series, we were able to confirm five significant associations in Purkinje cells (Fig. 2; values for R^2 and P_{adjusted} are shown in Table 1; Supplementary file 1 shows values for all age correlations in the Purkinje cell samples corrected for the larger numbers of tests). Of these, the highest correlation was for *RHBDL3* ($R^2=0.134$, $P_{\text{adjusted}}=0.0012$). Four additional genes (*NR3C2*, *GPX3*, *VPS18* and *SGSH*) showed significant associations, ten probes were not detected and the remaining 45 were not significantly associated with age in the Purkinje cells. However, as in the analysis of bulk tissue, the direction of effect was generally congruent between the Purkinje cell dataset and each of the bulk tissue samples, although estimates of effect size were lower in the isolated cells (Supplementary Fig. S3). These data suggest that some of the age-related changes in gene expression can be replicated in isolated neurons.

Technical and biological replication of age associations of gene expression

To provide validation by an additional technique, we took a subset of samples from the discovery series of frontal cortex and used quantitative reverse transcriptase-PCR (qRT-PCR) to estimate expression levels of three genes that showed a range of strengths of association with age, *RHBDL3*, *SGSH* and *NR3C2*. We found a significant correlation between age and covariate corrected expression of *RHBDL3* (Fig. 3A; $R = 0.497$, $P=9.92 \times 10^{-5}$) or *SGSH* (Fig. 3B; $R=0.339$, $P=0.0132$). We saw a positive correlation between age and covariate corrected expression of *NR3C2* (Fig. 3C; $R=0.231$) although this did not reach statistical significance ($P=0.087$). However, there was a clear relationship between estimated effect size by array and by RT-PCR (Fig. 3D).

Additionally, we compared our results with data from Colantuoni et al [13] who used a different array platform and independently ascertained sample series. Although matched for brain region (BA46/9), age range and with similar power (see Materials and methods), this study therefore represents both a technical and a biological replicate of our own datasets and therefore might be more compelling than technical replication alone. By matching probes to unique genes, we were able to confirm that the overall distribution of associations was similar between Colantuoni et al and either our discovery (Fig. 4A) and replication (Fig. 4B) datasets. Four of the five validated genes from table 1, *RHBDL3*, *NR3C2*, *VPS18* and *SGSH* were significant ($P_{\text{adjusted}} < 0.05$ for correlations between age and expression) in all three frontal cortex datasets (Figs. 4C–E and Supplementary table S5). *GPX3* was not assayed in the Colantuoni et al dataset as a no probe mapped to this gene.

Collectively, these data validate our initial observations that there are a number of genes that show associations between age and expression and in some cases can be validated across multiple independent studies and sample series using different techniques.

Network level associations with age are distinct from single gene associations

The analysis above provided single probes that we considered as replicated candidates for association with human brain aging, but the number that survived our replication approach were too few to identify specific biological themes. One possible reason for this is that while networks of genes may change in response to aging, each individual gene may contribute a small amount that could be variable in each individual. Thus, examining networks of related genes may provide additional power to examine underlying responses of the brain to age. This might be important in detecting gene families such as mitochondrial genes that are associated as a group with age, but where each specific transcript does not reach statistical significance.

To address this, we used weighted gene correlation network analysis (WGCNA) [27] to define a consensus set of modules of genes with similar expression patterns that were shared across the five datasets described above using 6767 probes with detectable expression in all five series. Twenty seven distinct consensus modules were detected, which are identified here by numbers (Fig. 5A). We then generated correlations for each probe against each of the covariates we had considered above, i.e. Age, PMI, Gender, PC1 and PC2 from the genetic analysis and hybridization batch. These gene significance measures were then correlated with module membership for each gene within a module such that a correlation coefficient, and associated p value, was assigned for each module and variable (Fig. 5B). Most of the tested relationships were not statistically significant. Only one module showed association with PMI, none were associated with gender or the two genetic measures PC1 and PC2. Only hybridization batch and age showed a number of significant correlations (Fig. 5B).

Of the modules that were associated with age, a subset (modules numbered 5,7,8,9 and 10 in Fig. 5B) had negative correlations with module membership and gene significance for age (R between -0.13 and -0.25 ; P from 0.05 to 3×10^{-4}) while a separate group (16,17,18, 19 and 20) showed positive correlations ($0.21 < R < 0.28$, $0.003 < P < 5 \times 10^{-5}$). In general, these were not contaminated by associations with other covariates that we examined, although there were significant correlations between gene significance and module membership for modules 8 and 9 ($R = -0.32, -0.31$; $P = 2 \times 10^{-6}, 4 \times 10^{-6}$ respectively) and modules 16 and 19 ($R = 0.18, 0.16$; $P = 0.009, 0.02$ respectively) with hybridization batch, suggesting some possible overlap between these two variables. We next examined the individual genes that contributed to each module, using the DAVID tool to annotate Gene Ontology categories for each list of genes in the five modules significantly ($P < 0.05$) negatively correlated with age compared to the five with positive correlations and five randomly selected modules that showed no correlation with age (Table 2). In four out of five of the negatively correlated modules, the cellular component GO category with lowest P value was mitochondrion or mitochondrial lumen. The GO terms for biological processes included two models (light yellow and green) of generation of precursor metabolites and energy. In contrast, nuclear lumen or intracellular lumen was the most significant cellular component GO term for the modules that were positively correlated with age. In this set of modules, the most significant GO terms for biological processes pointed to different aspects of gene expression, including transcription, histone acetylation, chromatin modification, DNA metabolic processes and RNA processing. The randomly selected modules that did not show association with aging had varied GO terms without a clear theme. Overall, these results suggest that there are biologically relevant gene expression networks within the human brain and that

mitochondrial gene expression tends to decrease with chronological age while there is a positive association with control of gene expression from the nucleus.

Discussion

In the current study, we reassessed the associations between gene expression and chronological age in the human brain. We have previously shown that such human brain samples can be used to demonstrate genetic effects on expression and DNA methylation [19] and also that there are robust associations between the extent of DNA methylation and chronological age [23]. Specifically, we designed the experiment to use two relatively well-powered sample series to examine age associations, and also considered the effects of regional variability in the brain.

The main outcome of the work is that a set of specific associations between age and expression are robust enough to survive comparison across two replicate series. The observed effect sizes are often modest, which indicates that associations with age may be difficult to identify smaller sample series. We saw fewer significant associations in the replication compared to the discovery dataset, in accordance with the fewer samples tested. This contention is largely consistent with the small number of age vs. gene expression associations nominated by previous studies [5,6,10,12,13]. However, we note that we were specifically focused on linear patterns of gene expression change with chronological age as we predict that these will be the most readily replicated across studies. It has been suggested that some genes show more complex patterns of change with age [13,19], perhaps having higher or lower expression in midlife. The current, covariate adjusted datasets, may be helpful in testing for such additional patterns of association in the future. However, overall, our data supports the idea that in the human brain there are distinct and measureable patterns of gene expression changes with age.

We were able to replicate several age vs. expression associations using samples where neurons were enriched using LCM. This is consistent with previous suggestions that at least some markers of neurons and glia do not change with age [5]. One limitation is that the series for LCM was less well powered than the original discovery dataset and it might be interesting to see if more associations are recovered with additional samples. Additionally, it should be noted that while LCM enriches for cell types, the separation is imperfect and tightly associated cells such as astrocytes may also contribute some signal in the LCM series. We also note that the expression data from the LCM tissue series was generated with a different amplification protocol and that one measure of RNA quality (RNA Integrity number/RIN) showed a wide range in the recovered cell samples. While these technical aspects did not appear to bias measurement of housekeeping genes such as GAPDH, and normalization of the overall datasets produced a relatively consistent estimate of gene expression, it remains possible that individual probes might behave differently under different amplification protocols and may have more variance in isolate cells than in bulk samples. Despite these limitations, we found associations with age and showed that some were congruent with those in bulk tissue. However, we note that it is likely that the false negative rate, i.e. true associations between age and gene expression, may be higher in the Purkinje cell data than in bulk tissue leading to an underestimate of the true aging signal.

An open question is whether different tissues and cells age at different rates, which has been suggested previously [10]. We did not directly test this idea, but we were able to compare two different brain regions in the evolutionarily newer frontal cortex compared to the cerebellum. Published data suggest that there is a difference in the rate of aging in these two brain regions [7] and, supporting this, more genes were detected as having age associations in the frontal cortex than the cerebellum. However, age vs. gene expressions associations

that were replicated in the Purkinje cells, a principal neuronal population in the cerebellum, included some that were not significant in the bulk cerebellum (e.g. *GPX3*). It is therefore possible that some age-associations are masked in bulk tissue but revealed in some neurons, although this remains speculative without further work on other cellular populations. It is also important to note that there are samples with expression values that appear to be outliers from the rest of the population (e.g. *GPX3* and *VPS18* in the cerebellum in Fig. 2). Sampling variability may be difficult to control in human *post mortem* studies such as the one presented here, and, along with other technical difficulties with such sample series, probably limits power to detect modest associations. Despite these difficulties, we were able to confirm that there was a similar overall pattern of age-related associations in the frontal cortex of an independently ascertained and assayed series from Colantuoni et al [13], suggesting our approach can identify robust associations that are predictive of other series. Further testing this proposal will require the development of additional large, well-powered series in the human brain.

Of interest is that a recent discovery/replication study performed using gene expression in human blood samples [14] also identified relatively few robust associations with age. Sixteen transcripts were significant in both datasets with correlation coefficients in the range of 0.35–0.5. However, the list of nominated transcripts in the current study and the work of Nakamura et al do not overlap. This might either be due to differences in overall genes expression between tissues, to differences in the way genes change with age in different tissues or to technical differences between studies. We were able to examine the top hit in the blood study, *NEFL*, as this gene encodes the light chain of neurofilaments, which are abundantly expressed in the brain as well as in naive memory T and CD4 T cells. In contrast to the data from Nakamura et al., who reported a statistically significant positive correlation between *NEFL* expression and age in two independent cohorts, we were not able to replicate this result in our brain samples using probe ILMN_1659086 (Supplementary Fig. S4). Outside of technical and population differences of the samples, this supports the idea that the brain may age differently from other tissues. Along with the replication of associations of age in Purkinje cells, the lack of significant association between age and a marker of neurons in the brain supports the idea discussed above that differences in cellularity with age are not a primary driving factor for our observed associations. However, accumulation of reactive glia or vascular gene expression changes are still likely to occur with age and may contribute to some signals.

The five genes for which we have the greatest level of support for a significant association with chronological age were *RHBDL3*, *NR3C2*, *GPX3*, *VPS18* and *SGSH*. *RHBDL3* was cloned previously as a homologue of the *Drosophila melanogaster* rhomboid protease that plays a role in epidermal growth factor signaling and is expressed in the mouse brain as well as in other organs [32], and has no known connections to aging. Given the robust associations seen here, we propose that *RHBDL3* might have value as a marker of aging, although this should be confirmed or refuted in other tissues and species. *NR3C2* encodes a mineralocorticoid receptor that may play a role in sarcopenia and other age-related phenotypes [33]. The relevance of *NR3C2* in brain function is not clear but the expression of this gene in the brain and in isolated neurons suggests that it may have an unidentified role in neuronal function that might be explored in future studies. *GPX3* appears to play a role in antioxidant defenses and knockout increases volume of stroke in mice [34] and may therefore support the hypothesized contribution of oxidative stress to aging. *VPS18* is a human homologue [35] of a yeast gene involved in vacuolar protein sorting that may play a role in the endosome/lysosome system in mammalian cells. Finally, *SGSH* encodes a lysosomal enzyme, N-sulfoglucosamine sulfohydrolase that degrades heparin sulfate and mutations in which cause the lysosomal storage disease Sanfilippo syndrome A [36]. Therefore, these different genes do not appear to be biologically related to each other,

although *VPS18* and *SGSH* are both related to lysosomal processing, which may have lesser capacity in aging tissues.

The lack of biological similarity between the single genes that survived replication initially appears surprising. However, it is possible that looking for correlations at the level of single genes does not allow us to detect subtler, but potentially meaningful, changes in transcripts across classes. Our data would appear to support this contention. Using network analysis, which should be sensitive to consistent but small (i.e. sub-significant) events across related transcripts, we were able to identify mitochondrial pathways as showing negative correlations with age. Of interest is that no single mitochondrial gene was identified that was replicated across tissues; the network changes are therefore distinct from the single gene expression changes.

Whether any of the observed changes are causally related to aging phenotypes or are consequential to, or even epiphenomena of, the aging process cannot be determined from these data. Some of the changes in mitochondrial genes have been suggested to play a central role in aging and hence it is possible that some of the signals reported here are causal. However, to confirm this would require manipulation of those genes in appropriate model systems that would allow measurement on the effects on longevity. This would be most interesting if some of the proposed associations, such as *RHBDL3*, can be further replicated. The current data may therefore have utility for developing new hypotheses about aging and perhaps in the generation of biomarkers of the aging process.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (Z01-AG000947 and Z01-AG000185) and by the UK Medical Research Council Biomedical Informatics Postdoctoral Training Fellowship (G0802462 to M.R) and Medical Research Council Project Grant (grant G0901254 to J.H). DT was supported by the King Faisal Specialist Hospital and Research Centre, Saudi Arabia. We would like to thank the LCM core facility at the National Cancer Institute for help with microdissection.

References

1. Martin GM. The biology of aging: 1985–2010 and beyond. *FASEB J*. 2011; 25:3756–3762. [PubMed: 22046003]
2. Shammass MA. Telomeres, lifestyle, cancer, and aging. *Curr Opin Clin Nutr Metab Care*. 2011; 14:28–34. [PubMed: 21102320]
3. Kirkwood TBL. Systems biology of ageing and longevity. *Philos Trans R Soc Lond, B, Biol Sci*. 2011; 366:64–70. [PubMed: 21115531]
4. Sahin E, Colla S, Liesa M, Moslehi J, Müller FL, Guo M, et al. Telomere dysfunction induces metabolic and mitochondrial compromise. *Nature*. 2011; 470:359–365. [PubMed: 21307849]
5. Lu T, Pan Y, Kao SY, Li C, Kohane I, Chan J, et al. Gene regulation and DNA damage in the ageing human brain. *Nature*. 2004; 429:883–891. [PubMed: 15190254]
6. Rodwell GEJ, Sonu R, Zahn JM, Lund J, Wilhelmy J, Wang L, et al. A transcriptional profile of aging in the human kidney. *PLoS Biol*. 2004; 2:e427. [PubMed: 15562319]
7. Fraser HB, Khaitovich P, Plotkin JB, Pääbo S, Eisen MB. Aging and gene expression in the primate brain. *PLoS Biol*. 2005; 3:e274. [PubMed: 16048372]
8. Zahn JM, Sonu R, Vogel H, Crane E, Mazan-Mamczarz K, Rabkin R, et al. Transcriptional profiling of aging in human muscle reveals a common aging signature. *PLoS Genet*. 2006; 2:e115. [PubMed: 16789832]

9. Bahar R, Hartmann CH, Rodriguez KA, Denny AD, Busuttil RA, Dollé MET, et al. Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature*. 2006; 441:1011–1014. [PubMed: 16791200]
10. Zahn JM, Poosala S, Owen AB, Ingram DK, Lustig A, Carter A, et al. AGEMAP: a gene expression database for aging in mice. *PLoS Genet*. 2007; 3:e201. [PubMed: 18081424]
11. Kadish I, Thibault O, Blalock EM, Chen KC, Gant JC, Porter NM, et al. Hippocampal and cognitive aging across the lifespan: a bioenergetic shift precedes and increased cholesterol trafficking parallels memory impairment. *J Neurosci*. 2009; 29:1805–1816. [PubMed: 19211887]
12. Oh S, Tseng GC, Sibille E. Reciprocal phylogenetic conservation of molecular aging in mouse and human brain. *Neurobiol Aging*. 2011; 32:1331–1335. [PubMed: 19729224]
13. Colantuoni C, Lipska BK, Ye T, Hyde TM, Tao R, Leek JT, et al. Temporal dynamics and genetic control of transcription in the human prefrontal cortex. *Nature*. 2011; 478:519–523. [PubMed: 22031444]
14. Nakamura S, Kawai K, Takeshita Y, Honda M, Takamura T, Kaneko S, et al. Identification of blood biomarkers of aging by transcript profiling of whole blood. *Biochem Biophys Res Commun*. 2012; 418:313–318. [PubMed: 22266314]
15. Hamatani T, Falco G, Carter MG, Akutsu H, Stagg CA, Sharov AA, et al. Age-associated alteration of gene expression patterns in mouse oocytes. *Hum Mol Genet*. 2004; 13:2263–2278. [PubMed: 15317747]
16. Liu LF, Shen WJ, Ueno M, Patel S, Kraemer FB. Characterization of age-related gene expression profiling in bone marrow and epididymal adipocytes. *BMC Genomics*. 2011; 12:212. [PubMed: 21545734]
17. Tranah GJ. Mitochondrial-nuclear epistasis: implications for human aging and longevity. *Ageing Res Rev*. 2011; 10:238–252. [PubMed: 20601194]
18. Miller JA, Horvath S, Geschwind DH. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc Natl Acad Sci USA*. 2010; 107:12698–12703. [PubMed: 20616000]
19. Berchtold NC, Cribbs DH, Coleman PD, Rogers J, Head E, Kim R, et al. Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proc Natl Acad Sci USA*. 2008; 105:15605–15610. [PubMed: 18832152]
20. Gibbs JR, van der Brug MP, Hernandez DG, Traynor BJ, Nalls MA, Lai SL, et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. *PLoS Genet*. 2010; 6:e1000952. [PubMed: 20485568]
21. Zheng B, Liao Z, Locascio JJ, Lesniak KA, Roderick SS, Watt ML, et al. PGC-1 α , a potential therapeutic target for early intervention in Parkinson's disease. *Sci Transl Med*. 2010; 2:52ra73.
22. Blalock EM, Buechel HM, Popovic J, Geddes JW, Landfield PW. Microarray analyses of laser-captured hippocampus reveal distinct gray and white matter signatures associated with incipient Alzheimer's disease. *J Chem Neuroanat*. 2011; 42:118–126. [PubMed: 21756998]
23. Hernandez DG, Nalls MA, Moore M, Chong S, Dillman A, Trabzuni D, et al. Integration of GWAS SNPs and tissue specific expression profiling reveal discrete eQTLs for human traits in blood and brain. *Neurobiol Dis*. 2012; 47:20–28. [PubMed: 22433082]
24. Trabzuni D, Ryten M, Walker R, Smith C, Imran S, Ramasamy A, et al. Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J Neurochem*. 2011; 119:275–282. [PubMed: 21848658]
25. Miller JA, Oldham MC, Geschwind DH. A systems level analysis of transcriptional changes in Alzheimer's disease and normal aging. *J Neurosci*. 2008; 28:1410–1420. [PubMed: 18256261]
26. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res*. 2001; 29:e45. [PubMed: 11328886]
27. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*. 2008; 9:559. [PubMed: 19114008]
28. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57. [PubMed: 19131956]

29. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009; 37:1–13. [PubMed: 19033363]
30. Kang HJ, Kawasawa YI, Cheng F, Zhu Y, Xu X, Li M, et al. Spatio-temporal transcriptome of the human brain. *Nature.* 2011; 478:483–489. [PubMed: 22031440]
31. Oberdick J, Levinthal F, Levinthal C. A Purkinje cell differentiation marker shows a partial DNA sequence homology to the cellular sis/PDGF2 gene. *Neuron.* 1988; 1:367–376. [PubMed: 2483097]
32. Jászai J, Brand M. Cloning and expression of Ventrhold, a novel vertebrate homologue of the *Drosophila* EGF pathway gene rhomboid. *Mech Dev.* 2002; 113:73–77. [PubMed: 11900977]
33. Burton LA, McMurdo MET, Struthers AD. Mineralocorticoid antagonism: a novel way to treat sarcopenia and physical impairment in older people? *Clin Endocrinol (Oxf).* 2011; 75:725–729. [PubMed: 21699555]
34. Jin RC, Mahoney CE, Coleman Anderson L, Ottaviano F, Croce K, Leopold JA, et al. Glutathione peroxidase-3 deficiency promotes platelet-dependent thrombosis in vivo. *Circulation.* 2011; 123:1963–1973. [PubMed: 21518981]
35. Huizing M, Didier A, Walenta J, Anikster Y, Gahl WA, Krämer H. Molecular cloning and characterization of human VPS18, VPS 11, VPS16, and VPS33. *Gene.* 2001; 264:241–247. [PubMed: 11250079]
36. Scott HS, Blanch L, Guo XH, Freeman C, Orsborn A, Baker E, et al. Cloning of the sulphamidase gene and identification of mutations in Sanfilippo A syndrome. *Nat Genet.* 1995; 11:465–467. [PubMed: 7493035]

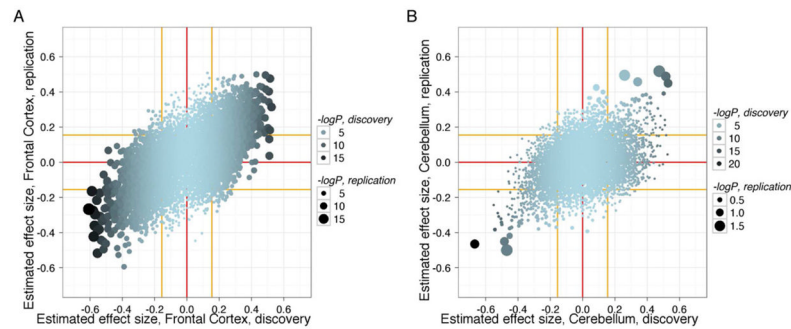


Figure 1.

Consistency of probes that show association with chronological age across the two datasets in frontal cortex (left) and cerebellum (right). Each point shows the estimated effect size for the association between age and expression of a given probe that was detected in both the discovery (x axis) and replication (y axis) datasets. Points are colored from blue to black by FDR adjusted P value in the discovery datasets and sized by FDR adjusted P value in the replication datasets and thus larger, darker points are those that are more highly significant in both datasets. There is an overall correlation in each dataset between effect sizes that is highly significant ($P < 2 \times 10^{-16}$; see text for details). The red horizontal and vertical lines indicate an effect size of 0 in each dataset and the orange lines indicate the range of effect sizes more than 0.115, which is the least significant effect size in the discovery dataset for frontal cortex. The majority of the probes that were significant outside this threshold were congruent in discovery and replication datasets.

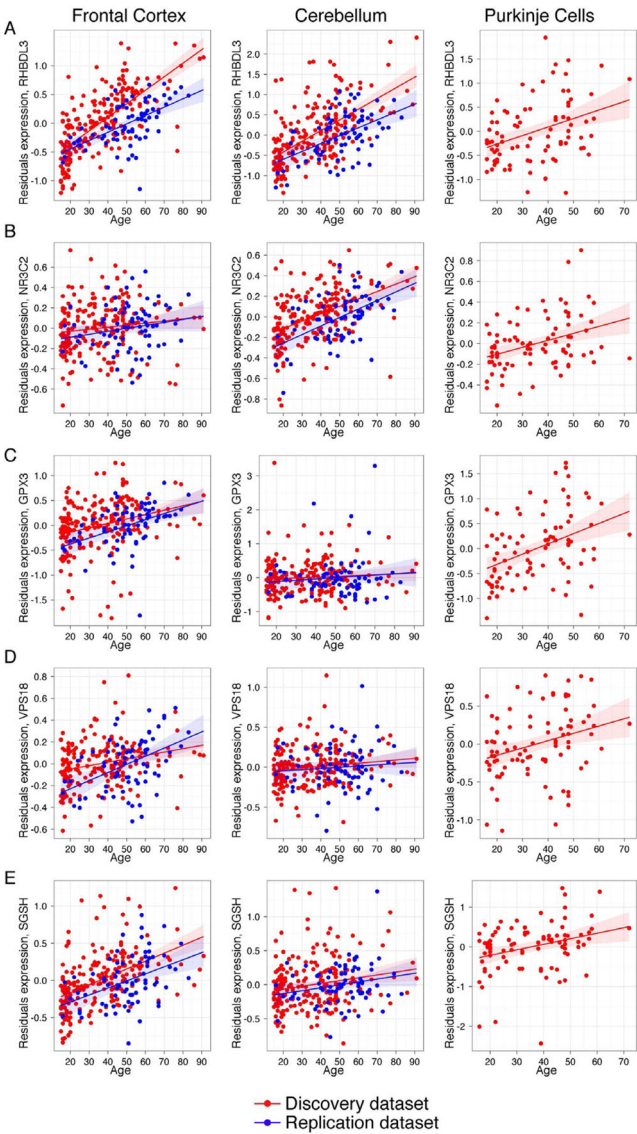


Figure 2. Replicated age associations. Each plot shows the association between age (x axes) and residuals of expression (y axes) for discovery (red) and replication (blue) datasets. Genes that were significant and replicated in either the Frontal Cortex (left hand panels) or the Cerebellum (center panels) were also tested in laser-captured Purkinje cells from a subset of the discovery dataset (right panels). These included the genes *RHBDL3* (A), *NR3C2* (B), *GPX3* (C), *VPS18* (D) and *SGSH* (E). Values of R^2 and associated P values are given in table 1.

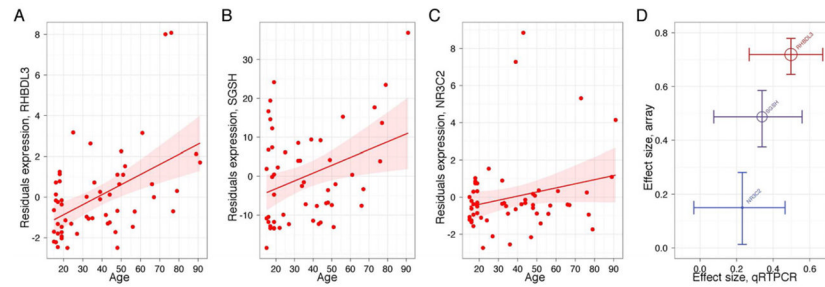


Figure 3.

Technical replication of age-associations by qRT-PCR. (A–C) We performed quantitative reverse-transcriptase PCR (qRT-pCR) for RHDBL3 (A), SGSH (F) and NR3C2 (C) and plotted the residuals of expression after correction for co-variates on the y-axes against age on the x-axes. In each panel, the red lines show a linear regression and the pale red line shows the 95% confidence interval of the fit to the line. (D) Comparing all three genes tested here for their association (estimated as R) between age and residuals of expression for each of the genes using qRT-PCR (x axis) or in the array study (y axis). Error bars in both directions indicate the 95% confidence interval for effect size estimate with each technique.

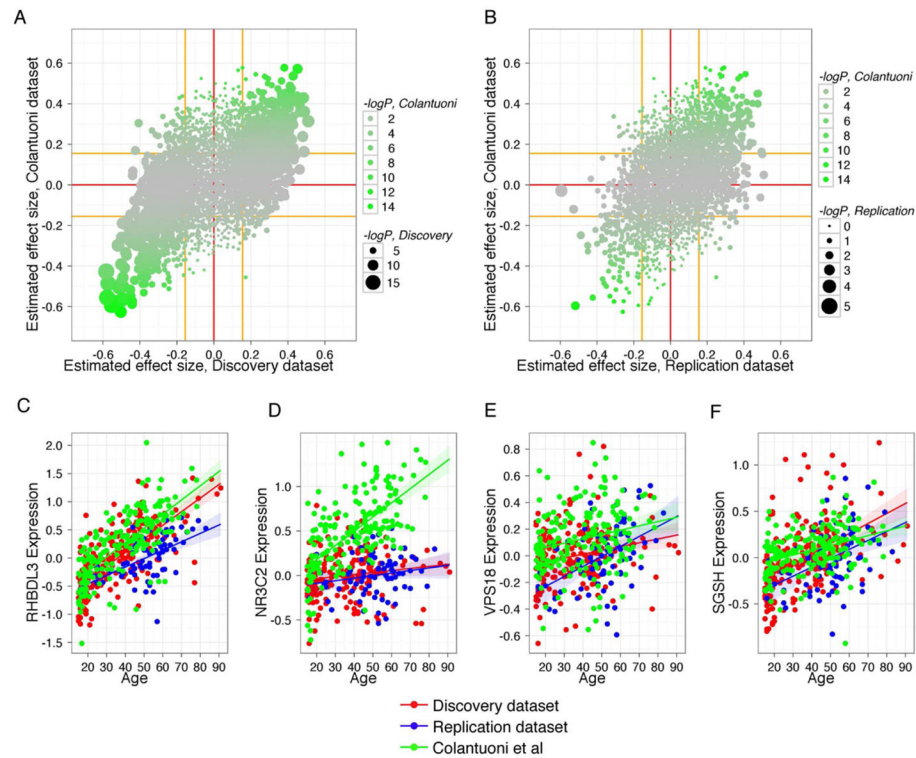


Figure 4.

Validation in an independent sample series. (A,B) Comparison of estimated effect sizes for the relationships between age and gene expression. Each point shows the estimated effect size for the relationship between age and expression of a detected gene in the work of Colantuoni et al, 2011 (y axes) compared to the frontal cortex data from this work (x axes) in the discovery (A) or replication (B) datasets. Data points are colored by $-\log$ of FDR adjusted P values in the Colantuoni dataset and sized by $-\log$ of FDR adjusted P values in our datasets and thus larger, darker datapoints represent more highly significant associations. (C–F) Normalized, covariate adjusted expression of age associated genes were plotted for each individual sample in each of three datasets; the discovery (red) and replication (blue) datasets from this study and the data from Colantuoni et al (green). These included RHBDL3 (C), NR3C2 (D), VPS18 (E) and SGSH (F) (see table 1), all of which showed significant age-associations in all tested datasets. GPX3 (shown in figure 2) was not included, as a probe for this gene was not annotated in the Colantuoni et al dataset.

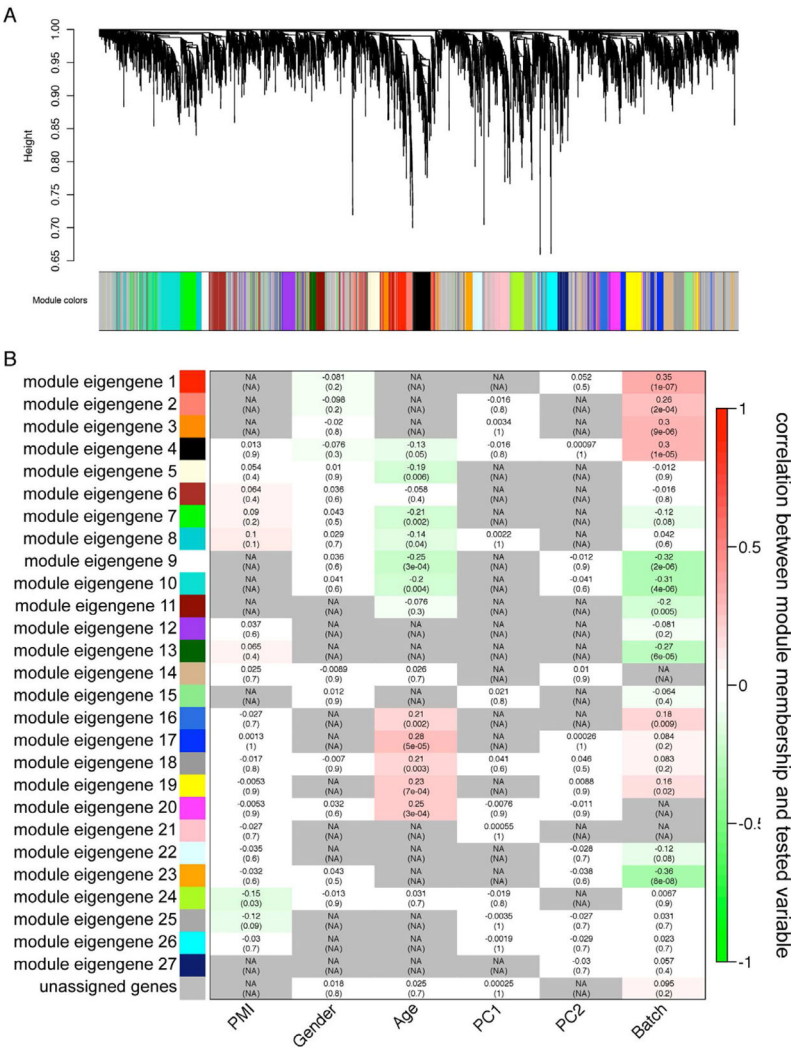


Figure 5. Weighted gene correlation network analysis. (A) Dendrogram for consensus modules across five tested datasets (discovery and replication datasets in Frontal Cortex and Cerebellum, plus LCM enriched purkinje cells). Each line is a single gene where height (y-axis) indicates dissimilarity. Branches of the dendrogram with similar expression patterns are captured into modules identified with different colors as indicated below the dendrogram. Grey indicates genes that were not assigned to any specific module. (B) For each module (numbered on the left hand side of the table where colors match those in panel A), module membership for individual genes was correlated with gene significance for each of the covariates of *post mortem interval* (PMI), Gender, Age, principal component 1 (PC1) and 2 (PC2) of the genetic analysis and hybridization batch as indicated along the bottom. Each value shows the correlation coefficient (with *P* values in parentheses below each), color-coded as shown on the scale on the right of the plot. NA; Not applicable.

Table 1

Replicated associations between chronological age and gene expression

Probe	RefSeq	Gene	Frontal Cortex				Cerebellum				Purkinje cells (n=98)	
			Discovery (n=203)		Replication (n=73)		Discovery (n=203)		Replication (n=73)		Purkinje cells (n=98)	
			r ²	P ^a	r ²	P	r ²	P	r ²	P	r ²	P
ILMN_1716019	NM_138328	<i>RHBDL3</i>	0.506	$<1 \times 10^{-17}$	0.331	4.43×10^{-4}	0.375	$<1 \times 10^{-12}$	0.257	1.97×10^{-2}	0.134	1.21×10^{-2}
ILMN_2133675	NM_000199	<i>SGSH</i>	0.239	1.19×10^{-11}	0.173	4.47×10^{-2}	0.032	3.32×10^{-2}	0.045	NS	0.091	4.34×10^{-2}
ILMN_1726666	NM_002084	<i>GPX3</i>	0.080	2.27×10^{-4}	0.188	3.38×10^{-2}	0.006	NS	0.008	NS	0.167	3.49×10^{-3}
ILMN_1721575	NM_020857	<i>VPS18</i>	0.048	4.74×10^{-3}	0.203	2.28×10^{-2}	0.019	NS	0.007	NS	0.086	4.34×10^{-2}
ILMN_1661777	NM_000901	<i>NR3C2</i>	0.024	4.96×10^{-2}	0.036	NS	0.225	1.96×10^{-10}	0.268	1.97×10^{-2}	0.129	1.22×10^{-2}

Footnotes

^aReported *P*-values adjusted for multiple testing using a false discovery rate (FDR) of 0.05. For the frontal cortex and cerebellum, all detected probes (8824 and 8926 respectively) were considered whereas for the purkinje cells, only the 60 probes that showed evidence of association with age in the prior datasets were tested.

\$watermark-text

\$watermark-text

\$watermark-text

GO enrichment in age-associated gene networks

Table 2

Module Eigengene ^a	$r(P)^b$	GO: Biological Processes		GO: Cellular Component		GO: Molecular Function	
		Term	P ^c	Term	P	Term	P
5	-0.19 (0.006)	GO:0006091~generation of precursor metabolites and energy	8.82×10 ⁻⁵	GO:0031980~mitochondrial lumen	1.261 × 10 ⁻⁶	GO:0043566~structure-specific DNA binding	0.02
7	-0.21 (0.002)	GO:0006091~generation of precursor metabolites and energy	6.07 × 10 ⁻⁸	GO:0005739~mitochondrion	1.48 × 10 ⁻⁶	GO:0000166~nucleotide binding	2.50 × 10 ⁻⁵
8	-0.14 (0.04)	GO:0046907~intracellular transport	1.24 × 10 ⁻⁴	GO:0005739~mitochondrion	0.00186	GO:0000166~nucleotide binding	0.00540
9	-0.25 (3×10 ⁻⁴)	GO:0008380~RNA splicing	0.00677	GO:0030529~ribonucleoprotein complex	1.75 × 10 ⁻⁴	GO:0003723~RNA binding	NS
10	-0.2 (0.004)	GO:0006412~translation	1.14 × 10 ⁻¹⁵	GO:0005739~mitochondrion	3.53 × 10 ⁻²⁸	GO:0003735~structural constituent of ribosome	2.90 × 10 ⁻¹³
16	0.21 (0.002)	GO:0006350~transcription	0.0106	GO:0031981~nuclear lumen	0.0239	GO:0003677~DNA binding	0.0104
17	0.28 (5×10 ⁻⁵)	GO:0016573~histone acetylation	3.09 × 10 ⁻⁴	GO:0070013~intracellular organelle lumen	3.66 × 10 ⁻⁶	GO:0000166~nucleotide binding	1.96 × 10 ⁻⁴
18	0.21 (0.003)	GO:0016568~chromatin modification	1.37 × 10 ⁻⁶	GO:0070013~intracellular organelle lumen	6.06 × 10 ⁻⁶	GO:0003723~RNA binding	5.21 × 10 ⁻⁵
19	0.23 (7×10 ⁻⁴)	GO:0006259~DNA metabolic process	1.03 × 10 ⁻⁴	GO:0070013~intracellular organelle lumen	5.25 × 10 ⁻⁵	GO:0000166~nucleotide binding	7.02 × 10 ⁻⁴
20	0.24 (3×10 ⁻⁴)	GO:0006396~RNA processing	2.41 × 10 ⁻⁴	GO:0031981~nuclear lumen	0.0127	GO:0051427~hormone receptor binding	0.00618
1	NS	GO:0051259~protein oligomerization	2.45 × 10 ⁻⁴	GO:0031981~nuclear lumen	5.79 × 10 ⁻⁴	GO:0003723~RNA binding	4.29 × 10 ⁻⁸
12	NS	O:0043543~protein amino acid acylation	0.0125	GO:0005891~voltage-gated calcium channel complex	0.0139	GO:0030695~GTPase regulator activity	3.53 × 10 ⁻⁴
13	NS	GO:0006886~intracellular protein transport	2.15 × 10 ⁻⁴	GO:0030131~clathrin adaptor complex	6.91 × 10 ⁻⁷	GO:0008565~protein transporter activity	0.00137
23	NS	GO:0006414~translational elongation	1.22 × 10 ⁻³⁹	GO:0033279~ribosomal subunit	9.41 × 10 ⁻³⁵	GO:0003735~structural constituent of ribosome	3.92 × 10 ⁻³⁵
26	NS	GO:0007265~Ras protein signal transduction	2.01 × 10 ⁻⁴	GO:0044437~vacuolar part	2.01 × 10 ⁻⁴	GO:0042802~identical protein binding	0.00547

Footnote:

^aModule Eigengenes from the weighted gene correlation network analysis (WGCNA) are numbered as in figure 5.

^bValues for correlation coefficient are listed for each network module between module membership and gene significance for aging and P values are in parentheses. Note that the first five modules showed negative correlation with age, the second five showed positive correlation and the final set were randomly chosen from modules that did not show significant association (NS)

^cP values for enrichment of GO terms within the indicated modules