Transcriptomic changes highly similar to Alzheimer's disease are observed in a subpopulation of individuals during normal brain aging

Shouneng Peng[1,2,3], Lu Zeng[1,2,3], Jean-vianney Haure-mirande[1,2,3], Minghui Wang[1,2,3], Derek M. Huffman[4,5,6], Vahram Haroutunian[7,8], Michelle Erlich[9,10], Bin Zhang[1,2,3], Zhidong Tu[1,2,3*]

[1] Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York City, NY 10029, USA

[2] Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York City, NY 10029, USA

[3] Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, One Gustave L. Levy Place, New York, NY 10029, USA

[4] Departments of Medicine, Albert Einstein College of Medicine, 1300 Morris Park Avenue, New York, NY 10461, USA

[5] Institute for Aging Research, Albert Einstein College of Medicine, 1300 Morris Park Avenue, New York, NY 10461, USA

[6] Departments of Molecular Pharmacology, Medicine and Institute for Aging Research Albert Einstein College of Medicine, 1300 Morris Park Avenue, New York, NY 10461, USA

[7] Departments of Psychiatry and Neuroscience, Icahn School of Medicine at Mount Sinai, New York City, NY, USA

[8] Mental Illness Research, Education and Clinical Center (MIRECC), James J. Peters VA Medical Center, Bronx, NY, USA

[9] Department of Neurology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[10] Departments of Pediatrics and Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

[*]Corresponding author: Zhidong Tu
IMI 3-70F, 1425 Madison Ave
New York City, NY 10029
E-mail: zhidong.tu@mssm.edu
Office Phone: 212-659-8508

Abstract

Aging is a major risk factor for late-onset Alzheimer's disease (LOAD). How aging contributes to the development of LOAD remains elusive. In this study, we examine multiple large-scale human brain transcriptomic data from both normal aging and LOAD to understand the molecular interconnection between aging and LOAD. We find that shared gene expression changes between aging and LOAD are mostly seen in the hippocampus and several cortical regions. In the hippocampus, phosphoprotein, alternative splicing and cytoskeleton are the commonly dysregulated biological pathways in both aging and AD, while synapse, ion transport, and synaptic vesicle genes are commonly down-regulated. Aging-specific changes are associated with acetylation and methylation, while LOAD-specific changes are related to glycoprotein (both up- and down-regulations), inflammatory response (up-regulation), myelin sheath and lipoprotein (down-regulation). We also find that normal aging brains from relatively young donors (45-70 years old) cluster into subgroups and some subgroups show gene expression changes highly similar to those seen in LOAD brains. Using brain transcriptome data from older individuals (>70 years), we find that samples from cognitive normal older individuals cluster with the "healthy aging" subgroup while AD samples mainly cluster with the AD similar subgroups. This implies that individuals in the healthy aging subgroup will likely remain cognitive normal when they become older and vice versa. In summary, our results suggest that on the transcriptome level, aging and LOAD have strong interconnections in some brain regions in a subpopulation of cognitive normal aging individuals. This supports the theory that the initiation of LOAD occurs decades earlier than the manifestation of clinical phenotype and it may be essential to closely study the "normal brain aging" in a subgroup of individuals in their 40s-60s to identify the very early events in LOAD development.

Keywords: Aging, LOAD, gene expression, brain transcriptome, brain aging subgroups, Alzheimer's disease.

Introduction

Alzheimer's disease (AD) is the most common cause of dementia and about 5.5 million Americans live with the disease based on the Alzheimer's Association 2018 report. Aging is the major risk factor for the late-onset AD (LOAD), which represents over 95% of the AD cases that occurred at age 65 or older (Caselli et al., 2009). A study analyzing 1,246 subjects aged 30–95 years found that the risk of developing AD dramatically increases in APOE ε4 carriers who are 70 years or older (Jack et al., 2015). It has been well-recognized that normal brain aging and LOAD share multiple common features, e.g., aging brains often manifest certain degrees of cognitive impairment, memory loss, metabolic disturbances, bioenergetic deficits, and inflammation. Even though aging increases the risk of AD and the two processes share similarities in multiple aspects, a detailed brain-region-specific view of their interconnection at the molecular level is not fully available. It is also unclear which aging mechanisms are playing major contributions to AD development and why some individuals may age without major cognitive deficits while others develop AD (Herrup, 2010, Koivisto et al., 1995). To address these issues, we perform a global comparison of the transcriptomes from normal aging brains and LOAD brains across multiple regions in a hope to gain new insights into the molecular interconnection between aging and AD.

Despite the fact that many transcriptomic studies have been performed to investigate aging and LOAD independently, only a few studies have compared normal aging and LOAD transcriptomic data in a systematic way (Berchtold et al., 2013, Cribbs et al., 2012, Lanke et al., 2018, Mastroeni et al., 2017). For example, Berchtold et al. used microarray to profile 81 aging and AD brains and found that synapse-related genes showed progressive down-regulation in both aging and AD (Berchtold et al., 2013). Using the same dataset, Lanke et al. performed an integrated analysis on young, aging and AD brains by constructing co-expression network models, and found that modules associated with astrocytes, endothelial cells and microglial cells were up-regulated while modules associated with neurons, mitochondria and endoplasmic reticulum were down-regulated, and these modules significantly correlated with both AD and aging. All these studies greatly helped our understanding of the interconnections between

aging and AD; however, the previous studies were limited in the brain regions examined and more importantly, they all treated aging and AD samples as uniform groups, and did not consider the possible heterogeneity in either aging or AD.

In this study, we compare the gene expression profiles of normal brain aging (age ≤ 70) with LOAD (age ≥ 60) (Hodes and Buckholtz, 2016, Yang et al., 2015) to understand the similarity and difference between aging and AD across multiple brain regions. We also consider brain aging subgroups and compare the aging subgroups with LOAD.

Materials and Methods

Data collection and pre-processing

We compiled and processed multiple large-scale human brain aging and AD gene expression datasets. We summarize and describe each dataset below and list all the data we studied in Table 1.

GTEx brain data

GTEx brain gene expression data (v7 and v8) from 13 brain regions were downloaded from the GTEx (Genotype-Tissue Expression) portal (Consortium, 2015) and NIH dbGaP database. Donor ages ranged between 20-70 years and samples from donors annotated with any brain diseases were removed from further analysis. We corrected gender, collection center, RIN (RNA Integrity Number), ischemic time, and top 3 genotype principal components (PCs) to calculate gene expression associated with donors' chronological age.

UK brain data

Data from 134 post-mortem brain donors free of neurological conditions were obtained from the UKBEC dataset (UK Brain Expression Consortium) (Millar et al., 2007, Trabzuni et al., 2011). 10 brain regions were included, namely, cerebellum (CRBL, n = 131 samples), frontal cortex (FCTX, n = 128), hippocampus (HIPP, n = 123), medulla (specifically inferior olivary

nucleus, MEDU, = 120), occipital cortex (specifically primary visual cortex, OCTX, n = 130), putamen (PUTM, n= 130), substantia nigra (SNIG, n= 102), temporal cortex (TCTX, n = 120), thalamus (THAL, n = 125) and intralobular white matter (WHMT, n = 132). To make the donor age distribution comparable between GTEx and UK, we only considered samples from donors whose age ≤ 70. The sample size for each brain region after filtering by age ≤ 70 is listed in Table 1.

Mount Sinai brain data

364 human brains (238 females and 126 males) were accessed from the Mount Sinai/JJ Peters VA Medical Center Brain Bank (MSBB–Mount Sinai NIH Neurobiobank) cohort. These samples represented the full spectrum of cognitive and neuropathological disease severity in the absence of discernable non-AD neuropathology. Donor ages for the samples ranged 61–108. For microarray profiling of 19 brain regions, a linear model was adopted to identify genes differentially expressed among different disease stage groups using R package Limma with default parameters and corrected for covariates including sex, postmortem interval (PMI), pH, and race (Wang et al., 2016). For each brain regions, DEGs (Differential Expression Genes) with FDR ≤ 0.05 in any of the 6 traits (CDR, Braak, CERAD, PLQ_Mn, NPrSum, NTrSum in contrast of High vs. Low, Low vs. Normal or High vs. Normal) were combined as the DEGs for that brain region (see Table S1)(Wang et al., 2016).

Other brain data

The AMP-AD knowledge portal (Hodes and Buckholtz, 2016) hosts datasets from three studies: Mayo Clinic Brain Bank (Mayo), the Religious Orders Study and Memory and Aging Project (ROSMAP) and Mount Sinai School of Medicine (MSSM). Mayo dataset included temporal cortex (80 AD, 71 controls) and cerebellum (79 AD, 72 controls). ROSMAP dataset contained samples from dorsolateral prefrontal cortex BA9 region (155 AD, 86 controls). The MSSM cohort included samples from four brain regions: parahippocampal gyrus, inferior frontal gyrus, superior temporal gyrus, and the frontal pole (n = 476, Table 1). DEGs were downloaded from the portal and filtered for genes with FDR ≤ 0.05.

Jager AD gene signatures

Jager et al. performed analyses on 478 ROSMAP dorsal lateral prefrontal cortex (DLPFC) tissue samples (Mostafavi et al., 2018). 5 gene lists were considered containing genes whose expression was associated with AD-related traits including clinical diagnosis of AD at the time of death, cognitive decline, tau, amyloid, and pathologic diagnosis of AD.

Annse2018 and Rooij2019 AD gene expression signatures

Annse2018 profiled hippocampus CA1 gene expression in 10 males with age between 60 ~ 81(5 AD and 5 controls) (Annese et al., 2018). DEGs with |log2(Fold Change)| > 1 and FDR ≤ 0.05 were selected. Rooij2019 profiled gene expression of hippocampus samples from 18 AD and 10 controls. DEGs with differential expression score ≥ 0.1 and FDR ≤ 0.05 were considered (van Rooij et al., 2019).

Differential expression analysis and age-associated gene expression identification

Differential expression (DE) analysis in aging and AD was performed using the R package edgeR and Limma (Law et al., 2014, Robinson et al., 2010) and we adjusted batch, RIN, sex, and post-mortem interval (PMI) in GTEx data and brain source, gender, batch effects and PMI in UK data. A linear regression model was applied to identify gene expression changes associated with age and we adjusted the same covariates as we did in the DE analysis (Yang et al., 2015, Zeng et al., 2020).

Subgroup identification using hierarchical clustering

We used hierarchical clustering method to identify subgroups in normal aging GTEx and UK brain samples. We used the top 5000 most variable genes which showed similar clustering result with using all genes (ARI>0.9) for the hierarchical clustering of GTEx data.Ward.D2 method in the R hclust function was used (Murtagh and Legendre, 2014).

Compare PHG with GTEx hippocampus samples by clustering their transcriptomes

We obtained gene expression data from 215 parahippocampal gyrus (PHG) samples which was profiled at Mount Sinai and full details of the data is previously described(Wang et al., 2018) We merged and compared the PHG and GTEx hippocampus gene expression data. We first

calculated log2(tmp+1) for both datasets and then merged these datasets and removed the batch effects using R ComBat package with AGE, PMI, SEX, RIN as covariates. Negative values were assigned to 0 followed by sample-wide quantile normalization. The merged data matrix contained 271 samples (215 from PHG, 56 from GTEx) and we then adjusted the AGE, PMI, SEX, RIN for each gene for the down-stream clustering analysis. For the PHG data, we defined 22 normal (CDR = 0, braak score: bbscore ≤ 3, CERAD = "NL") and 63 AD (CDR ≥ 1, bbscore ≥ 5, CERAD = "definite AD") samples. We further selected the PHG samples with age > 70 and obtained 78 samples (19 normal, 59 AD). Based on the same GTEx top 5000 variance genes and hierarchical clustering method, we clustered PHG samples and found the normal and AD samples were mixed to some degree (Figure 7A). After removing the mixed samples, we obtained 51 samples showing clear separation into two groups corresponding to donors' AD status (14 normal samples and 37 AD samples).

Functional enrichment analysis

We annotated the biological functions of each gene list using DVAID tool (Huang et al., 2009, Sherman and Lempicki, 2009). We also performed pathway analysis using MetaCore integrated software suite (https://portal.genego.com/; last accessed in November 2019) to determine enriched biological processes. Signal transduction gene regulation network analysis was based on SIGNOR 2.0 (Licata et al., 2020) from the SIGnaling Network Open Resource (downloaded on Aug. 2020) using networkanalyst web tool (Zhou et al., 2019).

Results

1. A global comparison of aging and AD signatures across brain regions

1.1. Derive brain aging and AD signatures from multiple transcriptomic data

We collected multiple large-scale brain aging and AD gene expression datasets (Figure 1) as summarized in Tables 1 and S1. It is of note that to obtain "normal" aging gene expression signatures, samples from donors annotated with any brain-related diseases were removed in the GTEx data. We use a linear regression model to identify gene expression associated with donors' chronological age and consider these genes as brain aging genes (we call each list of brain aging genes as an aging signature for the corresponding brain region). We derive aging

signatures in 11 out of 13 brain regions from GTEx data; two regions, i.e., the brain spinal cord and substantia nigra show no apparent age-associated genes (0 and 1 gene, respectively) and are not considered for further analysis. Similarly, we derive aging signatures from 8 out of the 10 brain regions profiled in UK data; substantia nigra and medulla tissues show no apparent aging signatures and are removed from further analysis.

Genes differentially expressed between AD and normal control samples are extracted from previously published work (similarly we call them as AD signatures). These AD signatures are from AMP-AD knowledge portal (Hodes and Buckholtz, 2016), MS AD sets from Mount Sinai Medical Center Brain Bank (MSBB) AD cohort (Wang et al., 2016), 5 Jager gene lists in the dorsal lateral prefrontal cortex (DLPFC) region (Mostafavi et al., 2018), Annese2018 DEG list from hippocampus (HIPP) CA1 (denoted as Annese2018 HIPP)(Annese et al., 2018) and Rooij2019 DEG list in HIPP(van Rooij et al., 2019) (denoted as Rooij2019 HIPP) (see Table 1 and Method).

1.2. A global comparison of aging and AD signatures across brain regions

We first identified genes that show similar gene expression regulation across multiple brain regions in aging and AD datasets, respectively (we call them global aging and AD signatures). We found that 91 aging genes are consistently up or down-regulated in more than 4 out of 19 aging gene lists (11 GTEx and 8 UK brain regions). Among them, 27 are up-regulated and 64 are down-regulated (Table S2) with age. Similarly, 86 AD signature genes show consistent regulation in at least 9 out of 32 AD gene signature lists, among which 51 are consistently up-regulated and 35 are consistently down-regulated in AD brains (Table S2). We then perform function annotation of these global aging and AD signatures.

1.2.1. The global aging and AD signatures show up-regulation of immune complement pathway and down-regulation of synaptic related signaling

We use MetaCore to annotate the function of consistently regulated aging and AD genes. The 27 up-regulated aging genes are enriched for regulation of lipid metabolism, insulin regulation

of glycogen metabolism (PHK gamma gene; FDR = 2.9E-2), immune/inflammation complement pathway (*C4B*, *C4*, *C4A*; FDR = 1.9E-06), response to metal ion/transport categories (e.g., cellular response to copper ion; FDR = 1.7E-06), and dopamine metabolic process (*MAOB*, *MAO*; FDR= 3.8E-3) (see Table S3.1); the 64 down-regulated aging genes are enriched for *MAPK* (FDR = 1.2E-02) / *ASK1* (FDR = 2.4E-02) pathway, synaptic vesicle / calcium ion exocytosis categories, synaptic related signaling (NMDA, glutamate) (see Table S3.2). The 51 up-regulated AD genes are involved in immune complement pathway (GPCRs), TGF-beta receptor signaling (FDR = 4.6E-02), cytoskeleton remodeling, and L-glutamate import/neurotransmitter transport (see Table S3.3); 35 down-regulated AD genes are involved in synapse and GABAergic neurotransmission, glutamate secretion/glutamatergic pathways, vesicle fusion/recycling and transport, cytoskeleton remodeling, cell adhesion and protein phosphorylation (Table S3.4). Therefore, both AD and aging global signature genes show up-regulation in immune complement pathway and down-regulation in synapse (especially glutamate, synaptic vesicle recycling) related pathways; while aging genes are more involved in metal ion, transport, glycogen metabolism and AD genes are more involved in cytoskeleton remodeling, cell adhesion and synaptic categories.

1.2.2. *SST* and *SVOP* are down-regulated and *FOXJ1*, *SLC44A1* are up-regulated in the global aging and AD signatures

Two genes, *SST* and *SVOP* are consistently down-regulated in both global aging and AD gene signatures, while *FOXJ1* and *SLC44A1* are consistently up-regulated.

SST (somatostatin) is a neuropeptide hormone that maintains permeability and integrity of the blood-brain barrier (BBB) by regulating *LRP1* and *RAGE* expression. It abrogates Aβ-induced JNK phosphorylation and expression of *MMP2* to maintain permeability and integrity of BBB (Paik et al., 2019). Reduction of SST levels in the CSF and brain tissue is associated with impaired cognitive function and memory loss (Solarski et al., 2018). *SVOP* (synaptic vesicle 2-related protein) can bind to adenine nucleotide (particularly NAD) via its C-terminal extremity(Yao and Bajjalieh, 2009). Although the functions of *SVOP* remain obscure, the

evolutionary conservation and homology to transporters support it may play a role in molecular trafficking in synaptic vesicles(Janz et al., 1998).

In mammalian cells, *FOXJ1* is a member of the Forkhead/winged helix (FOX) family of transcription factors that is involved in ciliogenesis (Yu et al., 2008). It is shown to suppress *NFκB*, a key regulator in the immune response (Lin et al., 2004). It is hypothesized that *FOXJ1* may play a protective role involved in the pathophysiology of brain injury and may be required for the differentiation of the cells (Jacquet et al., 2009). The solute carrier 44A1 (*SLC44A1*) is a plasma membrane choline transporter. It is also a mitochondrial protein and acts as a choline transmembrane transporter. Choline is essential for the synthesis of the neurotransmitter ACh by cholinergic neurons for regulating neuronal activity. Choline deprivation in the central nervous system reduces acetylcholine (ACh) release, memory retention, and spatial cognition in the hippocampus (Canty and Zeisel, 1994, Nakamura et al., 2001, Zeisel, 2007).

2. Brain region-specific comparison of aging and AD signatures

In addition to comparing the global aging and AD signatures, we also compare aging and AD signatures in a brain region-specific manner.

2.1. Hypothalamus, hippocampus, and certain cortical regions show stronger age-related gene expression changes compared to other brain regions

As can be seen in Figure S1, the number of aging genes varies across brain regions. The regions showing a large number of aging genes in the GTEx data are hypothalamus (HYPO: 490 UP and 1160 DN genes), hippocampus (HIPP: 705 UP and 698 DN genes), and brain anterior cingulate cortex BA24 (BA24_AC: 320 UP and 888 DN genes). For the UK data, the brain regions showing a large number of aging genes are hippocampus (HIPP:432 UP and 324 DN genes) and temporal cortex (TCTX: 91 UP and 419 DN genes). It is of note that GTEx and UK do not cover the same brain regions, e.g., the hypothalamus was profiled by GTEx but not in the UK brain data. Among the 4 regions profiled by both GTEx and UK, (i.e., HIPP, CRBL, PUTM, FCTX), HIPP and CRBL show a higher number of aging genes and strong overlap between GTEx and UK aging signatures than the other two brain regions. Interestingly, the

hippocampus aging signature significantly overlaps with aging signatures from several cortical regions, while the cerebellum aging signature has very little overlap with other regions. The cerebellum show the most distinguishable gene expression patterns from all other brain regions which was also reported by the Allen Brain Atlases study (Mahfouz et al., 2015).

In summary, brain aging signatures are highly region-specific. Hippocampus, hypothalamus, and cortex TCTX and BA24_AC are more affected by aging on the transcriptome level than other brain regions surveyed by the GTEx and UK brain data.

2.2. Common hippocampus aging genes between GTEx and UK data

In general, the aging signatures from GTEx and UK show large reproducibility in matched brain regions. Using hippocampus as an example, the 705 up-regulated GTEx aging genes significantly overlap with 431 up-regulated UK aging genes by 77 genes (FDR= 2.6E-31); and the 698 down-regulated GTEx aging genes overlap with 324 down-regulated UK aging genes by 42 genes (FDR=3.6E-12). In addition to the immune/inflammation categories found in previous global aging signatures, the 77 common up-regulated aging genes are also enriched for *TGF-β* signal, transcription regulation (such as mRNA transcription by RNA polymerase II with FDR of 2.2E-03, ATP-dependent chromatin remodeling with FDR of 7.3E-03), blood vessel remodeling, membrane protein intracellular domain proteolysis, protein transport, and metabolic process (Table S4.1). Furthermore, from signal transduction network analysis on the KEGG database (Table S4.3 and S4.4), the top signal network contains 67 up-regulated genes and is involved in many metabolic-related pathways such as endocrine resistance (FDR = 4.0E-15), *foxo* signaling pathway (FDR = 1.4E-13), thyroid hormone signaling pathway (FDR = 8.7E-12), sphingolipid signaling pathway (FDR = 1.8E-10), neurotrophin signaling pathway (FDR = 1.8E-10), autophagy – animal (FDR = 3.6E-10), and insulin resistance (FDR = 1.5E-09).

The 42 common down-regulated aging genes between GTEx and UK are related to insulin-like growth factor 1 (*IGF-1*) signaling, *ERK1/2* pathway (for caveolin-mediated endocytosis or signal transduction), methylation, cell adhesion (e.g., synaptic contact), and long-term

potentiation pathways and previous reported energy-dependent pathways such as synaptic categories (e.g. synapse vesicle), oxidative categories (Table S4.2) based on MetaCore analysis. *IGF-1* (insulin-like growth factor) is a growth factor and neurohormone with some evidence suggesting its involvement in neurocognitive functions, neuroinflammation, and amyloid-β clearance. Furthermore, from signal transduction network analysis on the KEGG database (Tables S4.5 and S4.6), the top signal network is involved in circadian entrainment (FDR = 9.2E-07) and multiple synaptic-related pathways.

2.3. Brain aging signatures from the hypothalamus, the hippocampus, and BA24 anterior cingulate cortex significantly overlap with AD signatures.

After we evaluate the aging signatures between GTEx and UK data, we then compare the aging signatures with AD signatures in a brain-region specific manner. As can be seen in Figure 2, strong overlap between aging and AD signatures are observed in hippocampus, hypothalamus, and several cortical regions. For example, the GTEx hippocampus aging signature strongly overlaps with AD signatures derived from either hippocampus or several cortex regions. For the 698 GTEx_HIPP_DN aging genes, they overlap with Rooj2019_HIPP_DN by 288 genes (adjusted P-value = 2.78E-130); and the 705 GTEx_HIPP_UP genes overlap with Mayo_TCTX _UP (1578) by 150 genes (adjusted P-value = 6.62E-28). It is of note that as we divide aging and AD signatures into up- and down-regulated genes, in most cases, the down-regulated aging genes only significantly overlap with down-regulated AD genes, and vice versa, which supports that these signatures represent real biological signals. The overlap pattern is highly brain region specific. For example, GTEx_HIPP_UP (705) significantly overlap with AD signatures from hippocampus, while it only overlaps with AD signatures from non-hippocampus regions such as MS_BM44_IFG_UP (513) by 21 genes (adjusted P-value = 0.72). In both GTEx and UK data, the aging signature in cerebellum shows relatively weak overlap with our AD signatures.

We also observe that the GTEx hippocampus and UK hippocampus aging signatures show similar overlap pattern with AD signatures across different brain regions. Similarly, different AD signatures such as Annesse2018 and Rooj2019 show similar overlap pattern with aging

signatures across different brain regions, which further suggests that the aging and AD signatures from different studies are biologically meaningful and contain real signals of aging and AD.

2.4. Functional enrichment of aging specific, aging/AD common and AD specific genes in hippocampus

Since aging and AD gene signatures show strong overlap in the hippocampus as seen in Figure 2, we focus on this brain region to investigate the functional enrichment of aging and AD genes. To achieve better robustness, we annotate hippocampus aging genes derived from both GTEx and UK with hippocampus AD genes derived from both Annese2018 and Rooij2019 (Figure 3). We divide genes in the aging and AD signatures into three categories: aging specific signature genes (denoted as "ASGs"), aging-AD common signature genes (denoted as "AADGs"), and AD specific signature genes (denoted as "ADSGs"). Phosphoprotein and alternative splicing are enriched in both up- and down-regulated ASGs, AADGs and ADSGs; while glycosylation/glycoprotein, immune/inflammatory, stress response are more enriched in ADSGs; acetylation, nucleus, and Ubl conjugation are mostly enriched in ASGs. Down-regulated process of ASGs, AADGs, and ADSGs are all associated with membrane/cytoskeleton categories. Down-regulated AADGs are enriched for synapses (including cell junction, cholinergic/GABAergic/dopaminergic/glutamatergic synapse, AMPA etc.) and ion/transport (Figure 3 and Table S5); The comparison suggests that different from the normal aging process, the AD-specific and aging/AD common pathways are more enriched for glycoprotein, inflammatory response, synapse and mitochondrial dysfunction, while the aging process are more related to the nucleus, acetylation, coiled coil, which are less directly related to neurodegenerative phenotype. The functional annotation of AADGs suggests that dysregulation in transcription regulation, energy metabolism, membrane remodeling, extracellular vesicles (EV) and synapse pathways have already initiated and developed to some degrees in normal brain aging even though these individuals remain cognitive normal, while other biological processes such as inflammation/immune response are further escalated in AD patients.

3. Subgroup identification in normal aging brain hippocampus samples

3.1 Subgroups can be found in both GTEx and UK hippocampus datasets

As we have shown that in the hippocampus, aging and AD share multiple gene expression changes. However, not everyone will develop AD in advanced age, and this suggests that heterogeneity may exist in normal aging brains representing different levels of cognitive resilience. To examine the potential heterogeneity of normal aging brains, we explore subgrouping in both GTEx and UK hippocampus aging brain samples; and we observe that both GTEx and UK samples can be divided into major subgroups named as A, B, and C (see Figure 4).

3.2 DEGs derived from comparing aging brain subgroups highly overlap with AD signatures

We perform pair-wise comparisons to derive DEGs between subgroups in both GTEx and UK datasets, respectively. We jointly consider these DEGs (FDR < 0.01) with 3 hippocampus AD signatures (MS, Rooij2019, and Annese2018). As shown in Figure 5, DEGs derived from comparing GTEx subgroups B vs A (denoted as GTEx DEG BvsA) show strong overlap with AD signatures from Rooij2019 and Annese2018. For the GTEx_DEG_BvsA down-regulated genes (3018 genes), it overlaps with Rooij2019 down-regulated genes (1622 genes) by 1274 genes while they only overlap with Rooij2019 up-regulated genes (1045 genes) by 7 genes. This is highly significant, as Rooij2019 down-regulated genes (1622 genes) overlap with another AD signature, Annese2018 down-regulated genes (1183 genes) by 614 genes, which is comparable to the overlap we see with aging subgroup DEGs. It is of note that GTEx samples used in the subgroup analysis are from donors 45-70 years old without AD or other types of brain diseases. Simply by performing subgrouping and calculating DEGs from these subgroups, the DEGs highly overlap with AD signatures from independent studies, which further supports that gene expression changes in a subpopulation of the cognitive normal brains have very strong interconnection with AD. Similarly, GTEx DEG CvsA also strongly overlaps with AD signatures. Based on the pattern of overlap, we infer that GTEx subgroup A is more likely to be “healthy” aging subgroups while GTEx subgroups B and C are more similar to AD (here denoted as “AD similar group”). Since GTEx DEG CvsB show reverse overlap with AD

signatures from Annese2018 and Rooij2019, this indicates that subgroup B has relatively more similarity with AD samples compared to subgroup C. Therefore, the order of subgroups judged by how similar they are with AD samples can be inferred as AD samples > subgroup B > subgroup C > subgroup A, which is also consistent with the observation that GTEx DEG CvsA overlap with GTEx DEG BvsA in the same gene regulation direction. Similarly, for the UK data, we also observe very strong overlap between DEGs from comparing various subgroups and AD signatures. For example, the UK_DEG_AvsC_DN (3144 genes) highly overlap with Rooij2019_DN signature (1622 genes) by 762 genes. Similarly, we infer the order of UK subgroups judged by how similar they are with AD signatures as AD > UK subgroup A >> UK subgroup B > UK subgroup C.

3.3 Function annotation of GTEx and UK subgroup DEGs with AD signatures

We next used the DAVID tool to annotate the functional enrichment for GTEx and UK subgroup DEGs (Figure 6, Table S6). We focus on DEGs (FDR < 0.01) derived from comparing GTEx subgroup B (a subgroup that is more similar to AD) with a relatively healthy subgroup A (named GTEx DEGs BvsA) and similarly for UK DEGs AvsC with two AD signatures (Annese2018 and Rooij2019). Similar to previous annotations, such as aging specific genes (ASGs), we denote the aging subgroup DEGs as ABGs and annotate the functional enrichment of ABSGs (aging subgroup DEG specific genes), ADSGs, and ABADGs (common genes between aging subgroup DEGs and AD signature). As shown in Figure 6, certain post translational modifications (PTMs) (e.g., phosphoprotein, alternative splicing, Ubl conjugation, acetylation), membrane and cytoskeleton categories are significantly enriched in both up- and down-regulated ABSGs, ADSGs, ABADGs. ADSGs (both up- and down-regulation) are enriched for glycoprotein, while up-regulated ADSGs are more enriched for immunity, inflammatory response categories and down-regulated ADSGs are enriched for calcium/ion categories. The up-regulated conserved ABSGs show enrichment in cell cycle category and down-regulated conserved ABSGs show enrichment in proteostasis such as proteasome, protein transport and protein binding. For ABADGs, up-regulated genes are mainly enriched for transcription regulation and metabolism categories (e.g. PI3K-Akt

signaling pathway, MAPK signaling pathway, insulin resistance), while down-regulated genes are mainly enriched for synapse, lipoprotein, and circadian entrainment categories.

In summary, aging subgroup DEG analysis further suggests that remodeling of PTMs (e.g., phosphoprotein and alternative splicing), proteostasis, cytoskeleton, and metabolism are initiated in normal aging subgroups that show highly similar gene expression changes with AD patients.

3.4 GTEx "healthy aging subgroup" individuals will more likely remain cognitive normal in older ages as implied by another brain transcriptomic dataset

In both GTEx and UK datasets, we observe a subgroup that is different from AD and subgroups that are more similar to AD at the transcriptome level. We hypothesize that the subgroup less similar to AD is a "healthy aging subgroup" which will likely remain cognitive normal when they live into 70s and 80s. In contrast, the "AD similar group" will likely have a higher chance of having cognitive deficits in older ages. To test this hypothesis, we may need to perform a longitudinal study to follow-up each subgroup individuals for several decades. Despite the huge cost and long experimental time, this is not very feasible mainly because brain tissues are only available from postmortem donors in most cases; and therefore, we can only profile the brain tissue transcriptome at a single time point right after donor's death. To evaluate the hypothesis, we have to rely on an alternative approach. In one of such approaches, we ask for older individuals (>70 yrs) who have cognitive normal, if their brain transcriptome will be more similar to the GTEx healthy aging subgroup or AD similar subgroup. If most of them show similar transcriptome with the healthy aging subgroup, that implies the GTEx healthy aging subgroup will likely to remain cognitive normal when they become older.

To achieve this goal, we considered another large brain transcriptome dataset covering more than 200 brain tissues in the parahippocampal gyrus (PHG) region(Wang et al., 2018). Although this dataset is not from the same brain region (hippocampus), it is a region next to hippocampus and the dataset contains relatively large sample size and is well-annotated. From

these PHG samples, we identified two groups of samples, i.e., control and AD samples. The normal control samples are from donors with age > 70, CDR = 0, braak score (bbscore) ≤ 3, CERAD = 'NL' which has 19 samples. For this group, the donor age ranged from 73 to 103, with mean and standard deviation 83.3 ± 8.7 (yrs). The second group is the AD group, which we require donors to have age > 70 yrs, CDR ≥ 1, bbscore ≥ 5, CERAD = 'definite AD', and it has 59 samples based on these criteria. For this group, the donor age ranged from 71 to 104, with mean and standard deviation 86.966 ± 8.4 (yrs). We merged the PHG data with the GTEx hippocampus RNA-seq data and corrected the batch effect using R ComBat package (see method). We further corrected gene expression by donors' AGE, PMI, SEX, RIN for the following clustering analysis. As we processed GTEx gene expression together with PHG data to unify the two datasets, this introduced some changes to the GTEx gene expression. Therefore, we performed the hierarchical clustering on the 56 GTEx samples again and observed largely consistent subgrouping structure (subgroup B [n=11] > C [n=10] > A [n=35]) similar to the previous clustering results (subgroup B' [n=9] > C' [n=11] > A' [n=36] in the same order related to AD gene expression). For example, the new subgroup A (35 samples) overlaps with the previous subgroup A' (36 samples) by 32 samples. We also performed hierarchical clustering on just the 78 PHG samples and found some normal and AD samples were mixed in the clustering (see Figure 7A). To obtain clear separation between control and AD samples, we removed those mixed samples and obtained 14 normal (age 73 ~ 103, 5 male and 9 female) and 37 AD samples (age 73 ~ 104, 11 male and 26 female) that formed two clearly separated groups (see Figure 7B). We then assign each PHG sample to the GTEx subgroups by adding one PHG sample at a time to the 56 GTEx samples and run hierarchical clustering. By doing so, we find that all the 14 control samples cluster with the GTEx normal aging subgroup A. For the 37 PHG AD samples, 5 (3 APOE ε3/ ε4 and 2 APOE ε3/ ε3) of them cluster with GTEx subgroup B and 26 cluster with GTEx subgroup C (the two "AD similar" subgroups), only 6 samples (all APOE ε3/ ε3) cluster with the "healthy aging" subgroup A. This result suggests that the gene expression of "healthy" aging" subgroup is associated with cognitive normal status even in older individuals while the gene expression of the "AD similar" subgroup is more associated with AD in older population.

Discussion

To better understand the interconnection between aging and LOAD, we collected gene expression profiles from several large-scale transcriptomic datasets covering multiple brain regions and systematically compared aging and LOAD gene expression signatures. Different brain regions show varied levels of gene expression changes in aging and AD, respectively. Among all the brain regions we studied, hippocampus is one of the top regions that show a very strong interconnection between aging and AD on the transcriptomic level. We observe common functional enrichment in aging and LOAD related to PTMs (especially for alternative splicing and phosphoprotein), neurotransmission (especially glutamate), membrane, cytoskeleton, and lipid metabolism. We also show that gene expression changes in aging brain has more association with acetylation while AD-specific gene expression changes are more related to inflammatory response, glycoprotein, mitochondria and synapse. We also demonstrate that cognitively normal brains are not homogeneous in their transcriptomes and several major subgroups could be identified. By comparing gene expression among different subgroups, we show that gene expression changes in a subpopulation have strong overlap with the AD signature, suggesting that although aging is a general risk factor for AD, only a subset of individuals may experience gene expression changes to a level that significantly facilitates the onset of the disease.

From our study, several biological processes shared between LOAD and normal aging could point to the mechanisms of how aging and AD interconnect. First, PTMs related genes such as alternative splicing and phosphoprotein are enriched in both up- and down-regulated ASGs, AADGs/ABADGs and ADSGs (e.g., top function terms in Figures 3 and 6). PTMs regulation plays critical role for synaptic plasticity at several levels (Kiltschewskij and Cairns, 2017) by increasing proteome diversity through alternative splicing, or by enabling activity-dependent regulation of mRNA localization, translation or degradation in the dendrite. It will be interesting to investigate what events trigger the PTMs that are observed in aging and AD. Second, the immune response strongly showed-up in AD signatures as can be seen in our

results and previous studies (Bordner et al., 2011, Cribbs et al., 2012, Reichwald et al., 2009, Verbitsky et al., 2004). Since it is less significantly up-regulated in either ASGs/AADGs or ABSGs/ABADGs, this supports that the escalated inflammation is a hallmark of AD, relative to normal brain aging which is also known to be associated with low-grade neuroinflammation. Whether the inflammatory response plays a key causative role in the early stage of AD development or it is more of a reactive response to the upstream pathology requires further investigation.

Since it is infeasible to directly test if individuals in the "healthy aging" subgroup A will remain cognitive normal decades later, we rely on an alternative strategy in which we compare transcriptomes between GTEx brain aging subgroups with older individuals who are either AD or cognitive normal. For the 51 PHG samples (14 control and 37 AD samples), we observe that 31 out of 37 AD samples clustered with GTEx subgroups B and C (AD similar subgroups); while all the 14 normal control samples clustered with GTEx subgroup A, which presumably represents a healthy brain aging group. This result suggests that the gene expression in subgroup A is strongly associated with normal cognitive functions in both relatively young and older individuals, which supports that GTEx subgroup A is a healthy aging subgroup and individuals in this subgroup will likely to remain cognitive normal when they get older. However, we do observe that several AD patients have their PHG transcriptome clustered with GTEx subgroup A (6 out of 37 AD samples). This indicates that an individual with a healthy aging gene expression pattern in the hippocampus or PHG can still have AD. We think this could be related to the heterogeneity of AD which is increasingly known to have multiple subtypes. For example, at least four subtypes of AD (typical, limbic-predominant, hippocampal-sparing, and minimal atrophy AD) have been reported based on distribution of tau related pathology and regional brain atrophy (Ferreira et al., 2020). It is possible that some subtypes of AD will have their gene expression similar to our healthy aging subgroup. For example, it has been reported that hippocampal-sparing AD subtype has a lower frequency of APOE ε4 compared with typical and limbic-predominant AD (Ferreira et al., 2020); interestingly, all the 6 AD samples clustered with the GTEx subgroup A are from donors of APOE ε3/ε3, which may suggest that these AD patients belong to hippocampal-sparing AD

subtype. Since AD is not a single brain-region disease, to accurately predict AD development based on gene expression or any type of brain region-specific data, multiple regions should be examined and studied together.

In summary, our study suggests that a combined analysis of aging and LOAD (e.g. in other omics data) can help us to understand how aging may contribute to the development of LOAD. Since currently most genomic studies on AD focused on AD samples from donors in 70s to 90s, such samples may not provide the information regarding the molecular events occurred in the very early stage of the disease initiation and development. As we demonstrate in this work, the brains from a subpopulation of cognitive normal individuals in their 40s-70s already show gene expression changes similar to AD, this supports that the initiation of LOAD could occur decades earlier than the manifestation of clinical phenotype and it may be critical to closely study cognitive normal individuals in their 40s-60s to identify the triggering events in AD development.

Acknowledgements

This work is supported by NIH grants R01AG055501, R01AG067312 to Z.T. and D.M.H. and R01AG057907 to B.Z. The content is solely the responsibility of the authors and does not represent the official views of the National Institutes of Health. This work was also supported in part through the computational resources and staff expertise provided by Scientific Computing at the Icahn School of Medicine at Mount Sinai.

Author Contributions

ZT conceived the main concept of the work and BZ proposed the comparison with PHG data for “validation”. ZT and SP contributed to conceptualization, methodology, analysis, and preparation of the manuscript. SP, LZ, JH, MW, DH, VH, ME, BZ helped to discuss and improve the work. All authors have read and approved the manuscript for submission.

Competing interests

No competing interests declared.

Reference

1. Annese, A., Manzari, C., Lionetti, C., Picardi, E., Horner, D.S., Chiara, M., . . . D'Erchia, A.M. (2018) Whole transcriptome profiling of Late-Onset Alzheimer's Disease patients provides insights into the molecular changes involved in the disease. *Sci Rep* 8, 4282.
2. Berchtold, N.C., Coleman, P.D., Cribbs, D.H., Rogers, J., Gillen, D.L., and Cotman, C.W. (2013) Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiol Aging* 34, 1653-1661.
3. Bordner, K.A., Kitchen, R.R., Carlyle, B., George, E.D., Mahajan, M.C., Mane, S.M., . . . Simen, A.A. (2011) Parallel declines in cognition, motivation, and locomotion in aging mice: association with immune gene upregulation in the medial prefrontal cortex. *Exp Gerontol* 46, 643-659.
4. Canty, D.J. and Zeisel, S.H. (1994) Lecithin and choline in human health and disease. *Nutr Rev* 52, 327-339.
5. Caselli, R.J., Dueck, A.C., Osborne, D., Sabbagh, M.N., Connor, D.J., Ahern, G.L., . . . Reiman, E.M. (2009) Longitudinal modeling of age-related memory decline and the APOE epsilon4 effect. *N Engl J Med* 361, 255-263.
6. Consortium, G.T. (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348, 648-660.
7. Cribbs, D.H., Berchtold, N.C., Perreau, V., Coleman, P.D., Rogers, J., Tenner, A.J., and Cotman, C.W. (2012) Extensive innate immune gene activation accompanies brain aging, increasing vulnerability to cognitive decline and neurodegeneration: a microarray study. *J Neuroinflammation* 9, 179.
8. Ferreira, D., Nordberg, A., and Westman, E. (2020) Biological subtypes of Alzheimer disease: A systematic review and meta-analysis. *Neurology* 94, 436-448.
9. Herrup, K. (2010) Reimagining Alzheimer's disease--an age-based hypothesis. *J Neurosci* 30, 16755-16762.
10. Hodes, R.J. and Buckholtz, N. (2016) Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) Knowledge Portal Aids Alzheimer's Drug Discovery through Open Data Sharing. *Expert Opin Ther Targets* 20, 389-391.
11. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic acids research* 37, 1-13.
12. Jack, C.R., Jr., Wiste, H.J., Weigand, S.D., Knopman, D.S., Vemuri, P., Mielke, M.M., . . . Petersen, R.C. (2015) Age, Sex, and APOE epsilon4 Effects on Memory, Brain Structure, and beta-Amyloid Across the Adult Life Span. *JAMA Neurol* 72, 511-519.
13. Jacquet, B.V., Salinas-Mondragon, R., Liang, H., Therit, B., Buie, J.D., Dykstra, M., . . . Ghashghaei, H.T. (2009) FoxJ1-dependent gene expression is required for

differentiation of radial glia into ependymal cells and a subset of astrocytes in the postnatal brain. *Development* 136, 4021-4031.

14. Janz, R., Hofmann, K., and Sudhof, T.C. (1998) SVOP, an evolutionarily conserved synaptic vesicle protein, suggests novel transport functions of synaptic vesicles. *J Neurosci* 18, 9269-9281.
15. Kiltschewskij, D. and Cairns, M.J. (2017) Post-Transcriptional Mechanisms of Neuronal Translational Control in Synaptic Plasticity. *Synaptic Plasticity*, 15.
16. Koivisto, K., Reinikainen, K.J., Hanninen, T., Vanhanen, M., Helkala, E.L., Mykkanen, L., . . . Riekkinen, P.J., Sr. (1995) Prevalence of age-associated memory impairment in a randomly selected population from eastern Finland. *Neurology* 45, 741-747.
17. Lanke, V., Moolamalla, S.T.R., Roy, D., and Vinod, P.K. (2018) Integrative Analysis of Hippocampus Gene Expression Profiles Identifies Network Alterations in Aging and Alzheimer's Disease. *Front Aging Neurosci* 10, 153.
18. Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014) voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol* 15, R29.
19. Licata, L., Lo Surdo, P., Iannuccelli, M., Palma, A., Micarelli, E., Perfetto, L., . . . Cesareni, G. (2020) SIGNOR 2.0, the SIGnaling Network Open Resource 2.0: 2019 update. *Nucleic Acids Res* 48, D504-D510.
20. Lin, L., Spoor, M.S., Gerth, A.J., Brody, S.L., and Peng, S.L. (2004) Modulation of Th1 activation and inflammation by the NF-kappaB repressor Foxj1. *Science* 303, 1017-1020.
21. Mahfouz, A., van de Giessen, M., van der Maaten, L., Huisman, S., Reinders, M., Hawrylycz, M.J., and Lelieveldt, B.P. (2015) Visualizing the spatial gene expression organization in the brain through non-linear similarity embeddings. *Methods* 73, 79-89.
22. Mastroeni, D., Khdour, O.M., Delvaux, E., Nolz, J., Olsen, G., Berchtold, N., . . . Coleman, P.D. (2017) Nuclear but not mitochondrial-encoded oxidative phosphorylation genes are altered in aging, mild cognitive impairment, and Alzheimer's disease. *Alzheimers Dement* 13, 510-519.
23. Millar, T., Walker, R., Arango, J.C., Ironside, J.W., Harrison, D.J., MacIntyre, D.J., . . . Bell, J.E. (2007) Tissue and organ donation for research in forensic pathology: the MRC Sudden Death Brain and Tissue Bank. *J Pathol* 213, 369-375.
24. Mostafavi, S., Gaiteri, C., Sullivan, S.E., White, C.C., Tasaki, S., Xu, J., . . . De Jager, P.L. (2018) A molecular network of the aging human brain provides insights into the pathology and cognitive decline of Alzheimer's disease. *Nat Neurosci* 21, 811-819.
25. Murtagh, F. and Legendre, P. (2014) Ward’s hierarchical agglomerative clustering method: which algorithms implement Ward’s criterion? *Journal of classification* 31, 274-295.
26. Nakamura, S., Murayama, N., Noshita, T., Annoura, H., and Ohno, T. (2001) Progressive brain dysfunction following intracerebroventricular infusion of beta(1-42)-amyloid peptide. *Brain Res* 912, 128-136.
27. Paik, S., Somvanshi, R.K., and Kumar, U. (2019) Somatostatin Maintains Permeability and Integrity of Blood-Brain Barrier in beta-Amyloid Induced Toxicity. *Mol Neurobiol* 56, 292-306.

28. Reichwald, J., Danner, S., Wiederhold, K.H., and Staufenbiel, M. (2009) Expression of complement system components during aging and amyloid deposition in APP transgenic mice. *J Neuroinflammation* 6, 35.
29. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.
30. Sherman, B.T. and Lempicki, R.A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4, 44.
31. Solarski, M., Wang, H., Wille, H., and Schmitt-Ulms, G. (2018) Somatostatin in Alzheimer's disease: A new Role for an Old Player. *Prion* 12, 1-8.
32. Trabzuni, D., Ryten, M., Walker, R., Smith, C., Imran, S., Ramasamy, A., . . . Hardy, J. (2011) Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies. *J Neurochem* 119, 275-282.
33. van Rooij, J.G.J., Meeter, L.H.H., Melhem, S., Nijholt, D.A.T., Wong, T.H., Netherlands Brain, B., . . . van Swieten, J.C. (2019) Hippocampal transcriptome profiling combined with protein-protein interaction analysis elucidates Alzheimer's disease pathways and genes. *Neurobiol Aging* 74, 225-233.
34. Verbitsky, M., Yonan, A.L., Malleret, G., Kandel, E.R., Gilliam, T.C., and Pavlidis, P. (2004) Altered hippocampal transcript profile accompanies an age-related spatial memory deficit in mice. *Learn Mem* 11, 253-260.
35. Wang, M., Beckmann, N.D., Roussos, P., Wang, E., Zhou, X., Wang, Q., . . . Zhang, B. (2018) The Mount Sinai cohort of large-scale genomic, transcriptomic and proteomic data in Alzheimer's disease. *Sci Data* 5, 180185.
36. Wang, M., Roussos, P., McKenzie, A., Zhou, X., Kajiwara, Y., Brennand, K.J., . . . Zhang, B. (2016) Integrative network analysis of nineteen brain regions identifies molecular signatures and networks underlying selective regional vulnerability to Alzheimer's disease. *Genome Med* 8, 104.
37. Yang, J., Huang, T., Petralia, F., Long, Q., Zhang, B., Argmann, C., . . . Consortium, G.T. (2015) Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Sci Rep* 5, 15145.
38. Yao, J. and Bajjalieh, S.M. (2009) SVOP is a nucleotide binding protein. *PloS one* 4, e5315.
39. Yu, X., Ng, C.P., Habacher, H., and Roy, S. (2008) Foxj1 transcription factors are master regulators of the motile ciliogenic program. *Nat Genet* 40, 1445-1453.
40. Zeisel, S.H. (2007) Gene response elements, genetic polymorphisms and epigenetics influence the human dietary requirement for choline. *IUBMB Life* 59, 380-387.
41. Zeng, L., Yang, J., Peng, S., Zhu, J., Zhang, B., Suh, Y., and Tu, Z. (2020) Transcriptome analysis reveals the difference between "healthy" and "common" aging and their connection with age-related diseases. *Aging Cell* 19, e13121.
42. Zhou, G., Soufan, O., Ewald, J., Hancock, R.E.W., Basu, N., and Xia, J. (2019) NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. *Nucleic Acids Res* 47, W234-W241.

Figures Legends

Figure 1. Flowchart of the comparison between normal brain aging and AD. We collect gene expression profiles from a large number of normal aging brain and AD brain samples across multiple brain regions. We perform both a global comparison (Q1) and region-specific comparison (Q2) of aging and AD transcriptomes. We also consider the subgrouping in aging brain samples and compare the aging subgroups with AD (Q3).

Figure 2. Overlap between aging and AD signatures in different brain regions. AD signatures are plotted in rows and aging signatures are plotted in columns. We separate each signature into up- and down- regulated genes and the number of genes in each signature is listed after its ID. The number in the heatmap indicates how many genes are common in the

corresponding aging and AD signatures while the color indicates the significance of the overlap.

Figure 3. Function annotation of hippocampus aging signatures (GTEx, UK) compared with AD signatures (Annese2018, Rooij2019). We consider the functional enrichment of aging specific genes (ASGs), conserved AD specific genes (ADSGs) between two AD signatures, and the overlap between aging and AD genes (AADGs) which represents genes shared by at least one aging list and one AD list. We list the most representative function categories with FDR < 0.05.

Figure 4. Three major subgroups can be identified from either GTEx or UK hippocampus gene expression data. (**A).** Hierarchical clustering of GTEx hippocampus samples. 56 samples with donor age between 45 and 70 are plotted. Three subgroups are labeled in different colors in the dendrogram. The top color bar indicates that samples from different age groups are relatively evenly distributed into all the subgroups. (**B).** Hierarchical clustering of 70 UK hippocampus samples (45 ≤ age ≤ 75) also suggests these samples can be divided into three major subgroups.

Figure 5. Comparison of DEGs between GTEx subgroups and AD signatures from hippocampus. "GTEx DEG CvsB" represents DEGs derived from comparing subgroups C vs. B in GTEx hippocampus, similar naming format is used for the rest gene lists.

Figure 6. Function annotation of ABSGs, ABADGs, and ADSGs for a comparison of aging subgroup DEGs and AD signatures. The overlap of ABSGs between GTEx and UK is denoted as "Conserved ABSGs". The union of Rooij2019 and Annese2018 subtracting any ABGs is denoted as "ADSGs". We list the most representative function categories with FDR < 0.05.

Figure 7. Three major subgroups before and after mixed sample removal. (A). Hierarchical clustering of 78 PHG samples with age > 70. Three subgroups are labeled in different colors in the dendrogram. The color bar on the top indicates from the AD status of the sample: normal or AD. (**B).** Hierarchical clustering of 51 PHG samples with age > 70 after mixed samples being removed.

Tables

Table 1. List of brain transcriptomic datasets used for obtaining aging and AD gene signatures.

Datasets	Gene List ID	List Details (Tissue, traits)	# of genes (UP/DOWN)	# of protein coding genes	Sample Size
GTEx "Aging" sets					
GTEx	G_AMY	Amygdala	50(19/31)	46(19/27)	46
GTEx	G _BA24_AC	Anterior_cingulate_cortex (BA24)	1460(473/987)	1208(320/888)	56
GTEx	G_CAU	Caudate (basal_ganglia)	77(50/27)	55(34/21)	75
GTEx	G_CRBH	Cerebellar_Hemisphere	322(138/184)	277(108/169)	73
GTEx	G_CRBL	Cerebellum	427(213/214)	340(162/178)	87
GTEx	G_CTX	Cortex	97(34/63)	78(30/48)	71
GTEx	G_ BA9 DLPFC	Frontal_Cortex (BA9)	15(8/7)	12(5/7)	61
GTEx	G_HIPP	Hippocampus	1804(950/854)	1403(705/698)	57
GTEx	G_HYPO	Hypothalamus	2140(822/1318)	1650(490/1160)	53
GTEx	G_NC	Nucleus_accumbens (basal_ganglia)	11(6/5)	9(5/4)	69
GTEx	G_PUTM	Putamen (basal_ganglia)	18(13/5)	17(13/4)	67
UK "Aging" sets					
UK	UK_FCTX	Frontal cortex (FCTX)	170(50/120)	145(42/103)	94
UK	UK_TCTX	Temporal cortex (TCTX)	678(169/509)	510(91/419)	82
UK	UK_OCTX	Occipital cortex (OCTX)	54(30/24)	44(23/21)	95
UK	UK_WHMT	Intralobular white matter (WHMT)	464(267/197)	373(204/169)	93
UK	UK_CRBL	Cerebellum (CRBL)	186(101/85)	156(87/69)	92
UK	UK_PUTM	Putamen (PUTM)	65(30/35)	52(22/30)	95
UK	UK_THAL	Thalamus (THAL)	12(8/4)	9(5/4)	91
UK	UK_HIPP	Hippocampus (HIPP)	959(544/415)	755(431/324)	93
Other "AD" sets					
MSSM	MSSM_FP	BM10(MSSM_FP)	54(20/34)	47(19/28)	135
MSSM	MSSM_IFG	BM44(MSSM_IFG)	71(24/47)	61(24/37)	116
MSSM	MSSM_PHG	BM36(MSSM_PHG)	1577(936/641)	1391(868/523)	103
MSSM	MSSM_STG	BM22(MSSM_STG)	269(150/119)	242(142/100)	122

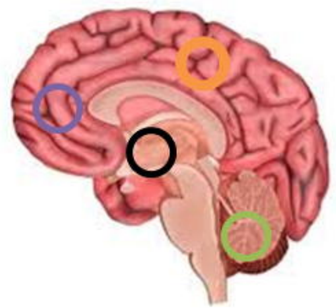
Mayo	Mayo_CRBL_Source *	Cerebellum	2632(1360/1272)	2282(1231/1051)	151
Mayo	Mayo_ CRBL	Cerebellum	1880(1008/872)	1600(901/699)	151
Mayo	Mayo_TCTX_Source *	Temporal cortex	3842(2034/1808)	3437(1933/1504)	151
Mayo	Mayo_TCTX	Temporal cortex	2951(1663/1288)	2628(1578/1050)	151
ROSMAP	ROSMAP_DLPFC	Dorsolateral prefrontal cortex (DLPFC)	162(89/73)	141(80/61)	241
Annese2018	Annese2018_HIPP	Hippocampus	2122(808/1314)	1925(742/1183)	10
Rooij2019	Rooij2019_HIPP	Hippocampus	2840(1109/1731)	2667(1045/1622)	28
Jager	Jager_Clinical_AD	Clinical_AD in DLPFC	855(466/389)	740(412/328)	478
Jager	Jager_Cognitive_decline	Cognitive_decline in DLPFC	3035(1481/1554)	2662(1281/1381)	478
Jager	Jager_Tau_tangles	Tau_tangles in DLPFC	238(155/83)	209(133/76)	478
Jager	Jager_B_amyloid	B_amyloid in DLPFC	2315(1158/1157)	2020(1060/960)	478
Jager	Jager_Patho_AD	Patho_AD in DLPFC	98(58/40)	85(51/34)	478
MS “AD” sets					
MS	MS_BM10_FP	Frontal Pole	19(5/14)	17(5/12)	63
MS	MS_BM17_OCTX	Occipital Visual Cortex	345(325/21)	297(279/19)	53
MS	MS_BM20_ITG	Inferior Temporal Gyrus	397(397/0)	349(349/0)	58
MS	MS_BM21_MTG	Middle Temporal Gyrus	635(81/565)	571(66/516)	58
MS	MS_BM22_STG	Superior Temporal Gyrus	25(0/25)	24(0/24)	60
MS	MS_BM23_PCC	Posterior Cingulate Cortex	34(7/27)	29(7/22)	58
MS	MS_BM32_AC	Anterior Cingulate	239(33/207)	214(30/184)	59
MS	MS_BM36_PHG	Parahippocampal Gyrus	55(7/49)	48(6/42)	60
MS	MS_BM38_TP	Temporal Pole	44(28/16)	39(25/14)	58
MS	MS_BM4_PCG	Precentral Gyrus	131(8/123)	122(7/115)	49
MS	MS_BM44_IFG	Inferior Frontal Gyrus	874(589/566)	764(513/495)	53
MS	MS_BM46_DLPFC	Dorsolateral Prefrontal Cortex	58(11/47)	53(8/45)	57
MS	MS_BM7_SPL	Superior Parietal Lobule	346(240/110)	314(220/97)	50
MS	MS_BM8_SFG	Superior Prefrontal Gyrus	953(426/670)	869(386/617)	56
MS	MS_CD	Caudate Nucleus	31(30/1)	29(28/1)	52
MS	MS_HIPP	Hippocampus	107(25/82)	95(22/73)	55
MS	MS_PUTM	Putamen	94(52/42)	82(45/37)	52

* AMP-AD Mayo gene signature is based on SourceDiagnosis.

Supporting Information

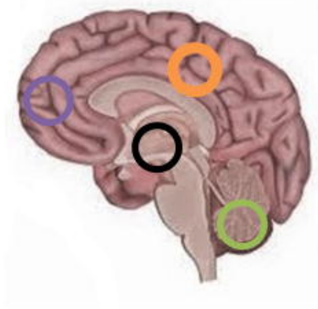
Supplementary Figures
Supplementary Tables

Normal aging brains



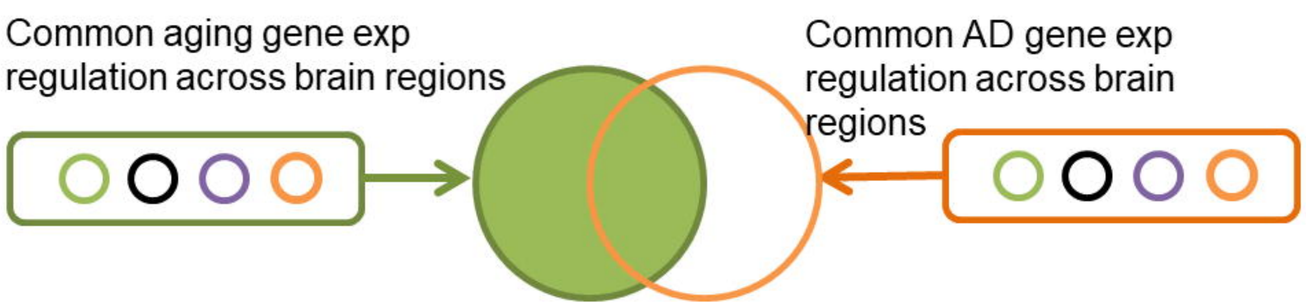
GTEx brain 13 regions
UK brain 10 regions

AD brains



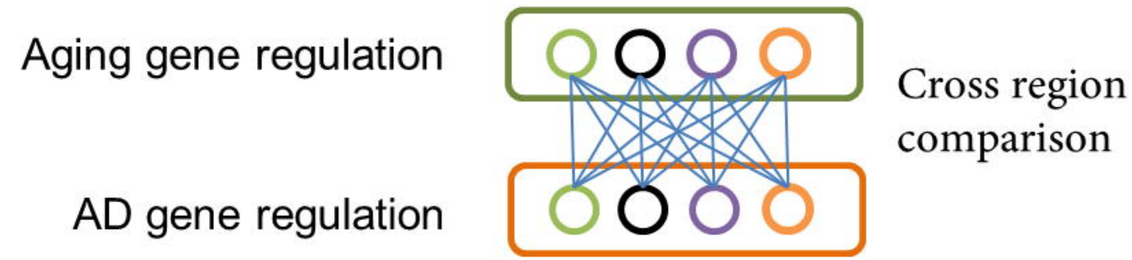
AMP-AD 13 regions
MS Brain 19 regions
& Others

Q1: Identify commonly regulated aging and AD genes across brain regions and how do they compare?

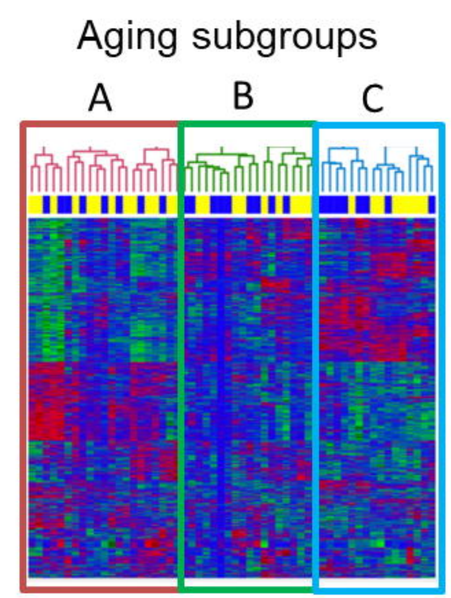


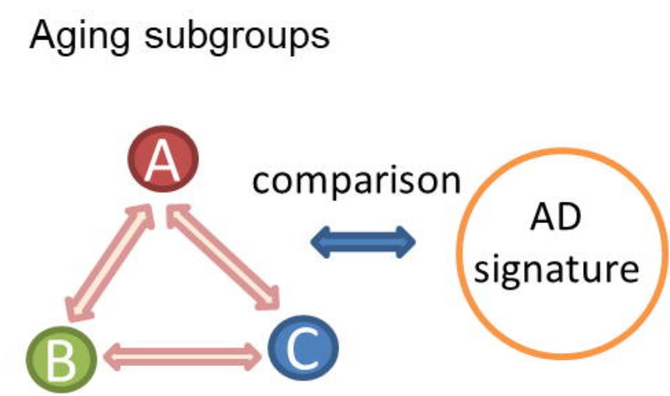
Brain region specific gene expression changes

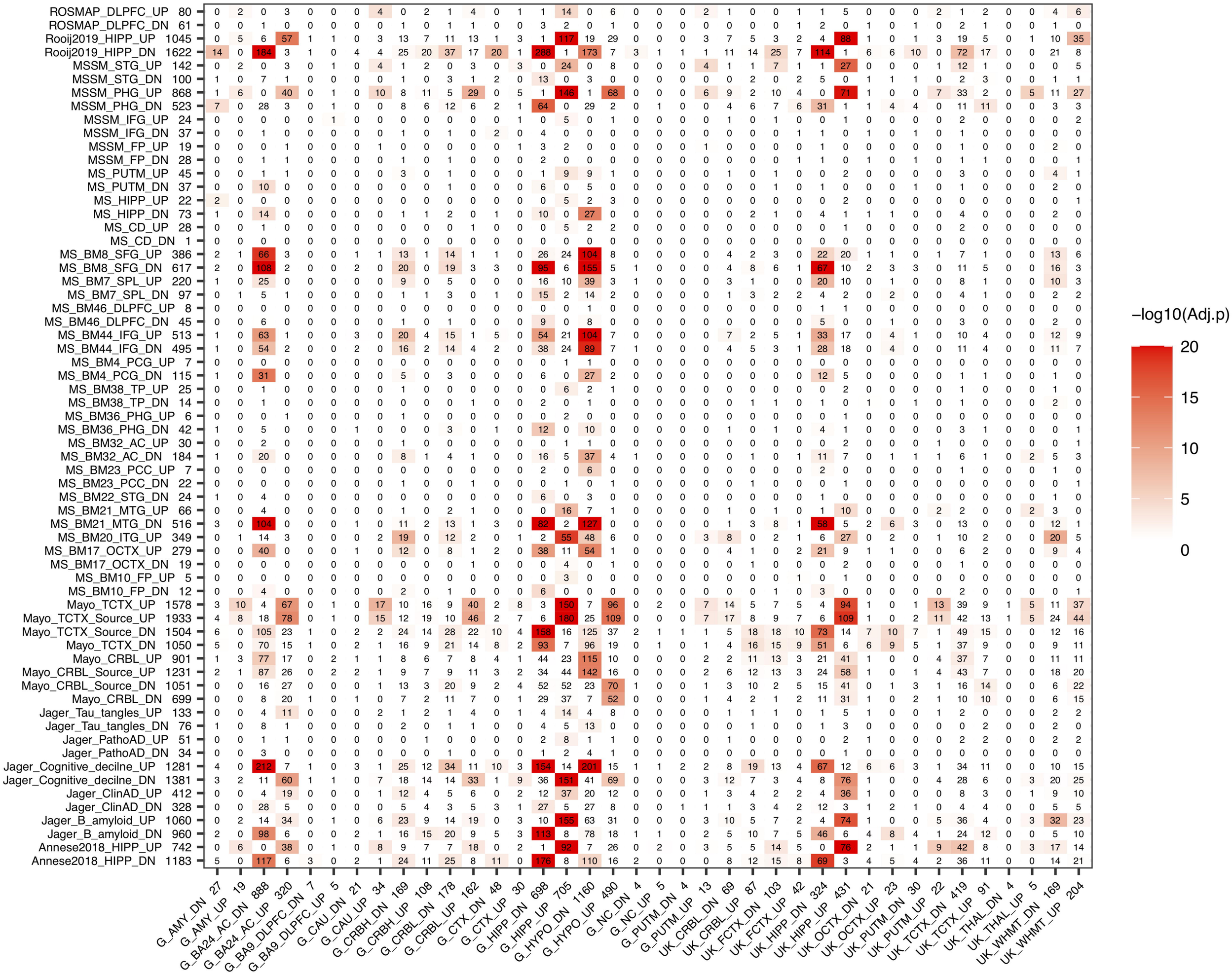
Q2: Brain region-specific comparison between aging and AD gene expression changes across brain regions



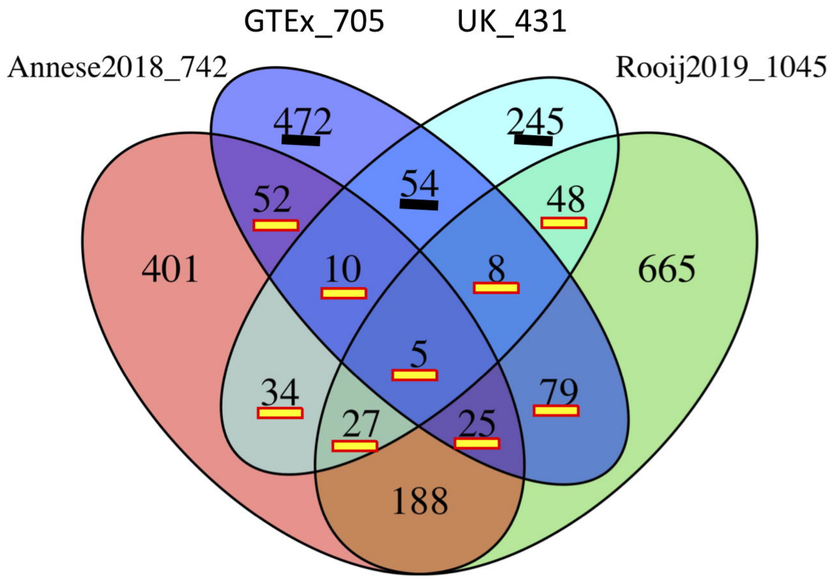
Q3: Identify subgroups of aging and compare subgroup-DEGs with AD genes in hippocampus






ROSMAP_DLPFC_UP 80
ROSMAP_DLPFC_DN 61
Rooij2019_HIPP_UP 1045
Rooij2019_HIPP_DN 1622
MSSM_STG_UP 142
MSSM_STG_DN 100
MSSM_PHG_UP 868
MSSM_PHG_DN 523
MSSM_IFG_UP 24
MSSM_IFG_DN 37
MSSM_FP_UP 19
MSSM_FP_DN 28
MS_PUTM_UP 45
MS_PUTM_DN 37
MS_HIPP_UP 22
MS_HIPP_DN 73
MS_CD_UP 28
MS_CD_DN 1
MS_BM8_SFG_UP 386
MS_BM8_SFG_DN 617
MS_BM7_SPL_UP 220
MS_BM7_SPL_DN 97
MS_BM46_DLPFC_UP 8
MS_BM46_DLPFC_DN 45
MS_BM44_IFG_UP 513
MS_BM44_IFG_DN 495
MS_BM4_PCG_UP 7
MS_BM4_PCG_DN 115
MS_BM38_TP_UP 25
MS_BM38_TP_DN 14
MS_BM36_PHG_UP 6
MS_BM36_PHG_DN 42
MS_BM32_AC_UP 30
MS_BM32_AC_DN 184
MS_BM23_PCC_UP 7
MS_BM23_PCC_DN 22
MS_BM22_STG_DN 24
MS_BM21_MTG_UP 66
MS_BM21_MTG_DN 516
MS_BM20_ITG_UP 349
MS_BM17_OCTX_UP 279
MS_BM17_OCTX_DN 19
MS_BM10_FP_UP 5
MS_BM10_FP_DN 12
Mayo_TCTX_UP 1578
Mayo_TCTX_Source_UP 1933
Mayo_TCTX_Source_DN 1504
Mayo_TCTX_DN 1050
Mayo_CRBL_UP 901
Mayo_CRBL_Source_UP 1231
Mayo_CRBL_Source_DN 1051
Mayo_CRBL_DN 699
Jager_Tau_tangles_UP 133
Jager_Tau_tangles_DN 76
Jager_PathoAD_UP 51
Jager_PathoAD_DN 34
Jager_Cognitive_decilne_UP 1281
Jager_Cognitive_decilne_DN 1381
Jager_ClinAD_UP 412
Jager_ClinAD_DN 328
Jager_B_amyloid_UP 1060
Jager_B_amyloid_DN 960
Annese2018_HIPP_UP 742
Annese2018_HIPP_DN 1183
G_AMY_DN 27
G_AMY_UP 19
G_BA24_AC_DN 888
G_BA24_AC_UP 320
G_BA9_DLPFC_DN 7
G_BA9_DLPFC_UP 5
G_CAU_DN 21
G_CAU_UP 34
G_CRBH_DN 169
G_CRBH_UP 108
G_CRBL_DN 178
G_CRBL_UP 162
G_CTX_DN 48
G_CTX_UP 30
G_HIPP_DN 698
G_HIPP_UP 705
G_HYPO_DN 1160
G_HYPO_UP 490
G_NC_DN 4
G_NC_UP 5
G_PUTM_DN 4
G_PUTM_UP 13
UK_CRBL_DN 69
UK_CRBL_UP 87
UK_FCTX_DN 103
UK_FCTX_UP 42
UK_HIPP_DN 324
UK_HIPP_UP 431
UK_OCTX_DN 21
UK_OCTX_UP 23
UK_PUTM_DN 30
UK_PUTM_UP 22
UK_TCTX_DN 419
UK_TCTX_UP 91
UK_THAL_DN 4
UK_THAL_UP 5
UK_WHMT_DN 169
UK_WHMT_UP 204
−log10(Adj.p)
20
15
10
5
0

A. Up hippocampus gene lists



Conserved 188 ADSGs UP

Term	N	FDR_{BH}
Phosphoprotein	111	3.86E-05
Repressor	21	6.12E-05
inflammatory response	18	5.11E-04
Activator	20	6.35E-04
Cell membrane	53	6.35E-04
Glycoprotein	65	0.00386
Transcription regulation	40	0.00411
Stress response	7	0.00797
Polymorphism	132	0.01592
glycosylation site:N-linked (GlcNAc...)	63	0.03114
Cytoplasm	62	0.04739

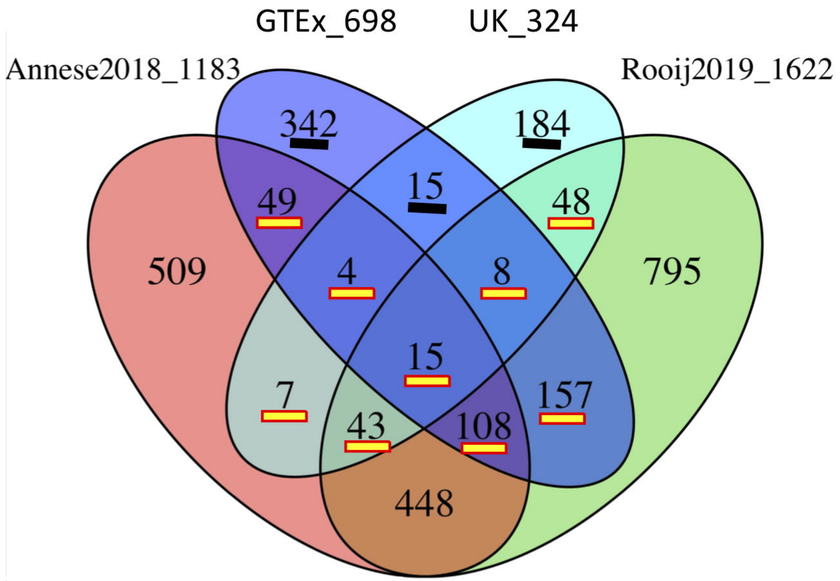
288 AADGs UP (▭)

Term	N	FDR_{BH}
Phosphoprotein	163	3.65E-06
Alternative splicing	192	1.89E-05
Staphylococcus aureus infection	9	7.82E-04
Coiled coil	70	0.001906
Actin-binding	15	0.002436
Cytoskeleton	33	0.007146
Polymorphism	199	0.007146
Cytoplasm	92	0.030856

771 ASGs UP (▬)

Term	N	FDR_{BH}
Alternative splicing	500	6.90E-12
Phosphoprotein	401	1.59E-09
Coiled coil	181	1.00E-08
splice variant	384	8.13E-08
Acetylation	194	8.87E-08
Polymorphism	522	5.48E-06
Nucleus	261	7.68E-06
Disease mutation	134	0.001606
Metal-binding	178	0.003041
Endoplasmic reticulum	65	0.004370
Zinc-finger	97	0.005189
Protein transport	42	0.006566
Cytoskeleton	67	0.006720
Ciliopathy	15	0.006720
Cytoplasm	220	0.011919
Ubl conjugation	89	0.023305
Transcription regulation	115	0.030039
Repressor	38	0.030039

B. Down hippocampus gene lists



Conserved 448 ADSGs DN

Term	N	FDR_{BH}
Membrane	268	8.92E-22
Synapse	43	5.96E-17
myelin sheath	28	1.09E-13
Transport	98	1.22E-12
Synaptic vesicle cycle	19	7.76E-12
Phosphoprotein	238	8.88E-07
Ion transport	39	1.28E-06
mitochondrion	66	1.35E-06
Lipoprotein	44	1.09E-05
Alternative splicing	281	2.70E-05
Cytoplasmic vesicle	29	8.55E-05
Glycolysis	8	8.55E-05
Oxidative phosphorylation	16	3.34E-04
Metabolic pathways	60	3.34E-04
Glycoprotein	137	4.69E-04
Retrograde endocannabinoid signaling	13	8.44E-04
extracellular exosome	91	0.02873
Methylation	36	0.03057

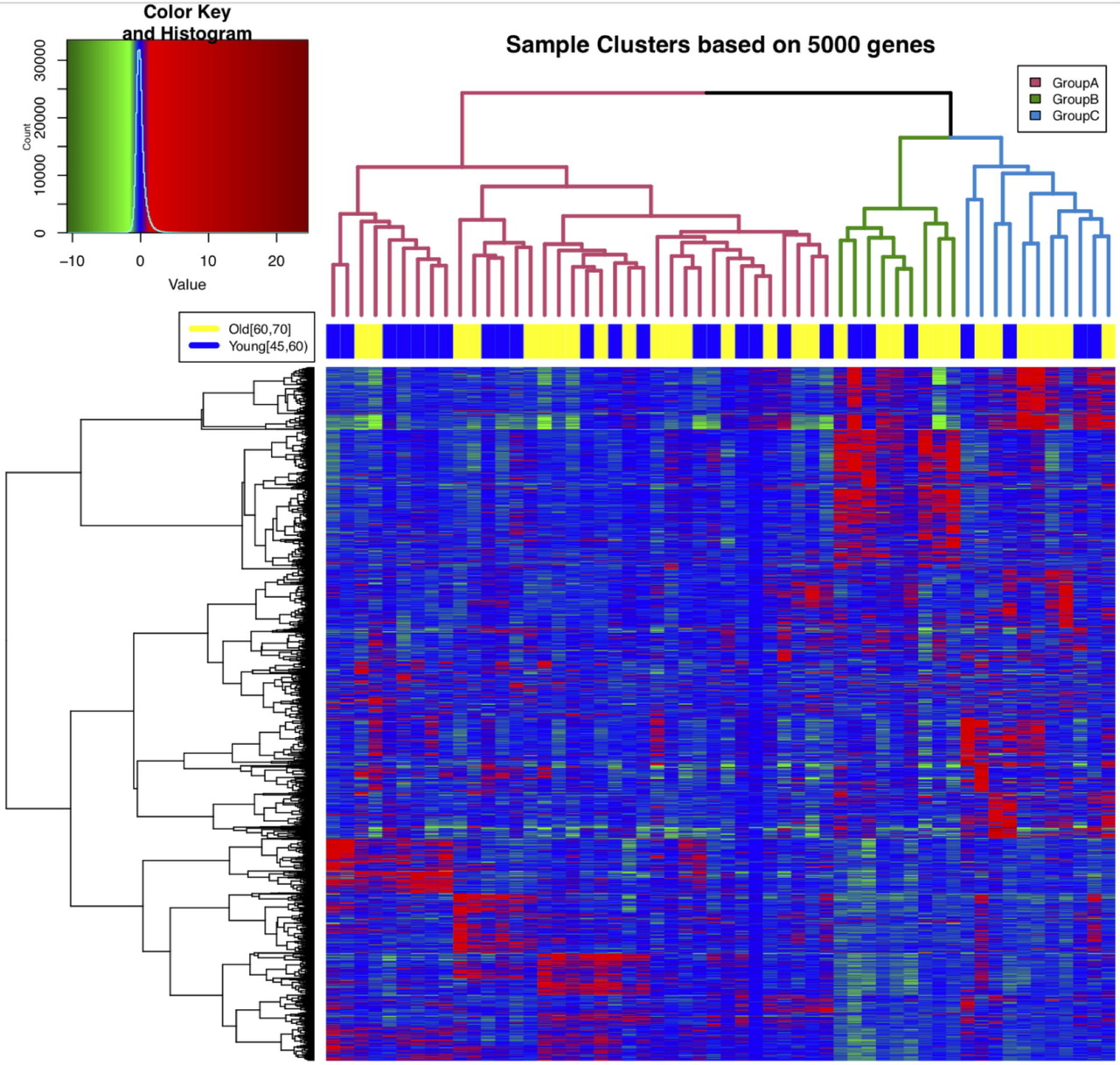
439 AADGs DN (▭)

Term	N	FDR_{BH}
Synapse	59	7.44E-32
plasma membrane	173	4.55E-16
Phosphoprotein	254	1.73E-12
Ion transport	43	4.15E-09
synaptic vesicle	15	6.77E-07
Kinase	41	1.41E-06
Alternative splicing	271	1.08E-04
cAMP signaling pathway	17	2.40E-04
Phosphatidylinositol signaling system	12	2.50E-04
Circadian entrainment	11	9.33E-04
Cytoskeleton	39	0.03103

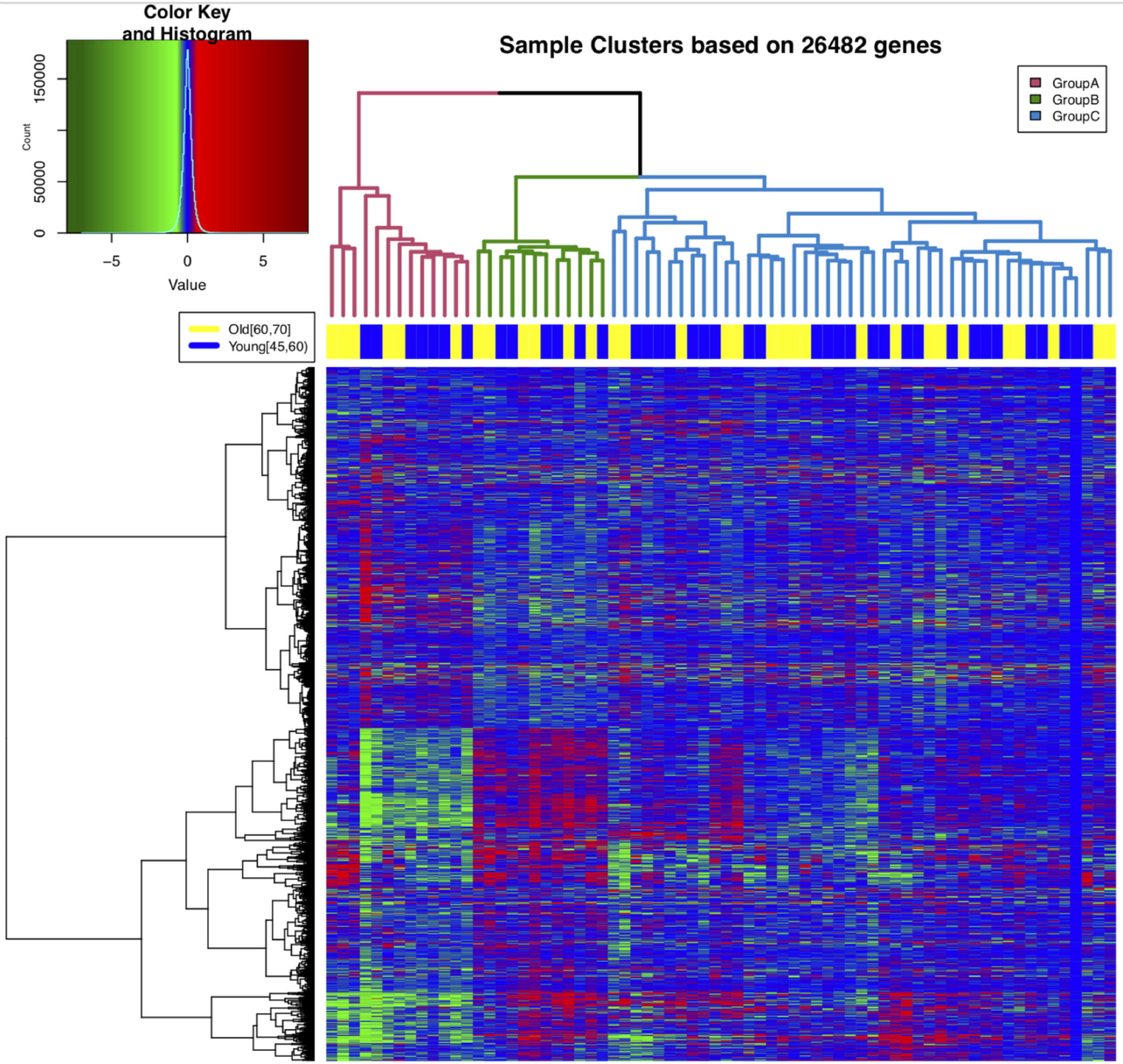
541 ASGs DN (▬)

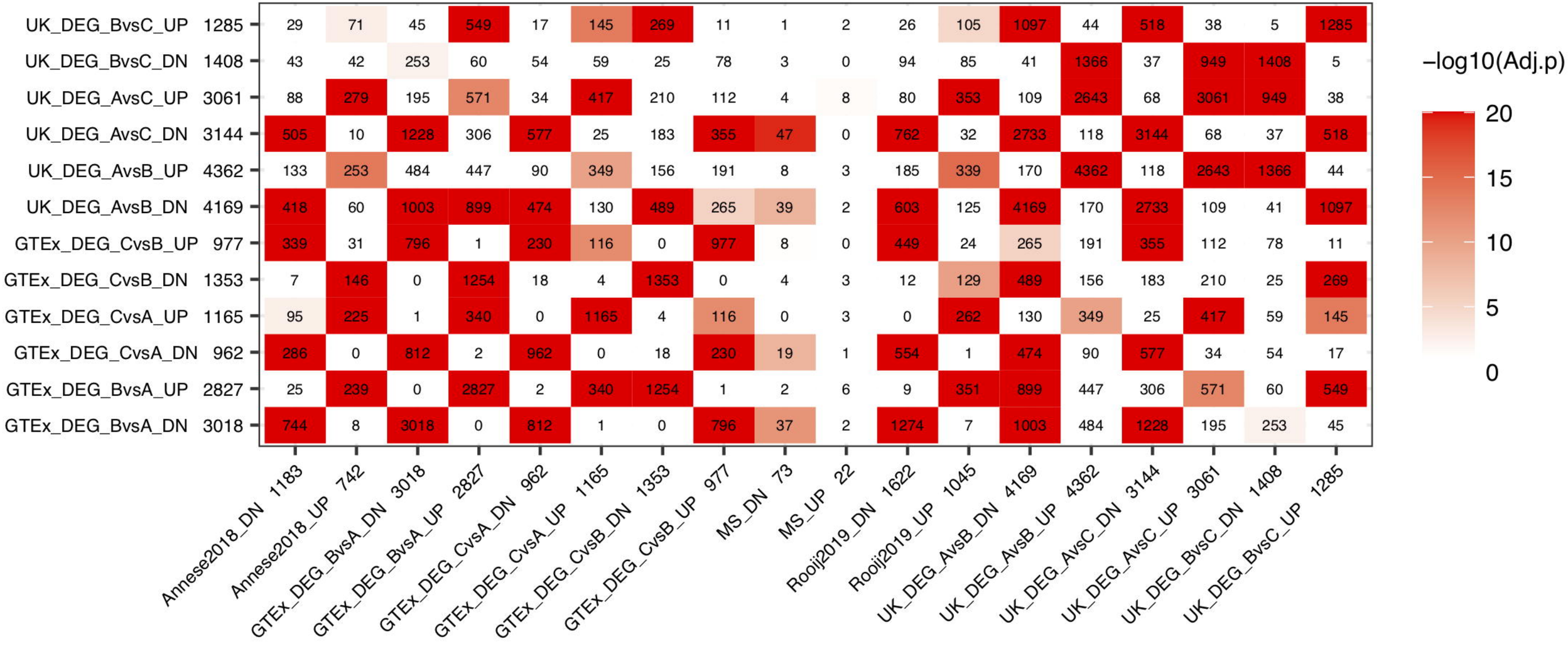
Term	N	FDR_{BH}
Phosphoprotein	288	1.07E-07
Methylation	59	2.13E-06
protein binding	317	2.57E-06
Cytoplasm	182	4.08E-06
Kinase	47	4.08E-06
Ubl conjugation pathway	43	1.81E-05
Nucleotide-binding	82	4.40E-05
Transferase	79	4.53E-05
Neurogenesis	20	0.001289
Serine/threonine -protein kinase	26	0.001329
Acetylation	126	0.001329
Nucleus	179	0.001329
Endocytosis	21	0.004188

A. GTEx hippocampus 3 subgroups



B. UK hippocampus 3 subgroups

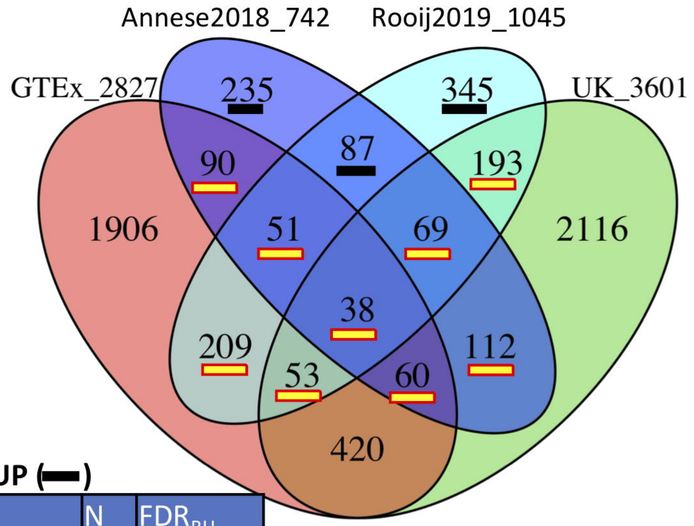



UK_DEG_BvsC_UP 1285
UK_DEG_BvsC_DN 1408
UK_DEG_AvsC_UP 3061
UK_DEG_AvsC_DN 3144
UK_DEG_AvsB_UP 4362
UK_DEG_AvsB_DN 4169
GTEx_DEG_CvsB_UP 977
GTEx_DEG_CvsB_DN 1353
GTEx_DEG_CvsA_UP 1165
GTEx_DEG_CvsA_DN 962
GTEx_DEG_BvsA_UP 2827
GTEx_DEG_BvsA_DN 3018
Annese2018_DN 1183
Annese2018_UP 742
GTEx_DEG_BvsA_DN 3018
GTEx_DEG_BvsA_UP 2827
GTEx_DEG_CvsA_DN 962
GTEx_DEG_CvsA_UP 1165
GTEx_DEG_CvsB_DN 1353
GTEx_DEG_CvsB_UP 977
MS_DN 73
MS_UP 22
Rooij2019_DN 1622
Rooij2019_UP 1045
UK_DEG_AvsB_DN 4169
UK_DEG_AvsB_UP 4362
UK_DEG_AvsC_DN 3144
UK_DEG_AvsC_UP 3061
UK_DEG_BvsC_DN 1408
UK_DEG_BvsC_UP 1285
−log10(Adj.p)
20
15
10
5
0

A. UP genes between aging and AD

875 ABADGs UP ( )

Term	N	FDR_{BH}
Phosphoprotein	537	4.96E-36
Activator	72	1.25E-10
Polymorphism	611	1.76E-10
protein binding	512	2.16E-10
Alternative splicing	550	3.01E-10
Repressor	64	1.58E-09
Cytoplasm	283	3.83E-08
Transcription	160	2.44E-07
Nucleus	298	2.99E-07
angiogenesis	37	3.87E-07
Ubl conjugation	120	1.44E-06
focal adhesion	47	1.88E-06
sequence variant	623	3.35E-06
Extracellular matrix	32	6.11E-06
Coiled coil	185	6.11E-06
Actin-binding	33	6.86E-06
PI3K-Akt signaling pathway	42	1.71E-05
Cell adhesion	45	3.90E-05
Cytoskeleton	81	1.37E-04
extracellular exosome	178	1.99E-04
Rho GTPase activation protein	18	2.24E-04
MAPK signaling pathway	30	9.90E-04
Insulin resistance	18	9.90E-04
membrane raft	25	0.00131



668 ADSGs UP ( )

Term	N	FDR_{BH}
Glycoprotein	244	5.63E-16
Immunity	60	6.10E-16
Polymorphism	475	2.67E-10
Membrane	328	4.22E-10
Signal	207	1.05E-09
Receptor	102	1.20E-08
cell surface	52	2.23E-08
leukocyte migration	23	2.39E-07
Secreted	109	7.26E-07
sequence variant	479	9.80E-07
receptor activity	29	3.00E-06
Apoptosis	41	2.45E-05
Tuberculosis	26	3.54E-05
Osteoclast differentiation	22	3.54E-05
extracellular space	82	8.29E-05
Cell adhesion	36	1.74E-04
membrane raft	22	6.38E-04
Phagosome	21	7.32E-04

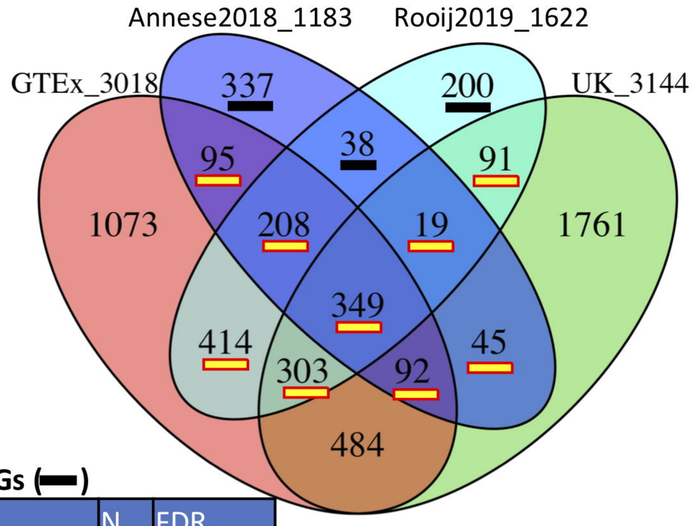
Conserved 420 ABSGs UP

Term	N	FDR_{BH}
Phosphoprotein	247	2.03E-12
Alternative splicing	271	6.80E-06
Cytoplasm	145	2.15E-05
splice variant	216	5.07E-05
Cytoskeleton	49	1.04E-04
Coiled coil	99	1.04E-04
Methylation	44	1.67E-04
Cell cycle	30	0.00282
Polymorphism	284	0.00282
Cell projection	32	0.00282
Actin-binding	17	0.00536
Microtubule	17	0.00630
Cell division	20	0.01103
Motor protein	11	0.01119
Nucleus	138	0.01162
Endocytosis	10	0.01814

B. DOWN (DN) genes between aging and AD

1616 ABADGs DN( )

Term	N	FDR_{BH}
Synapse	156	2.62E-74
cell junction	159	5.76E-54
Membrane	877	1.75E-50
Transport	306	2.48E-31
Ion transport	144	1.67E-29
dendrite	100	3.08E-27
Ion channel	96	3.35E-25
axon	69	1.10E-19
Synaptic vesicle cycle	36	1.34E-18
Phosphoprotein	819	2.24E-18
Alternative splicing	995	8.21E-17
perikaryon	42	7.06E-16
myelin sheath	49	1.68E-14
Retrograde endocannabinoid signaling	39	1.55E-13
Oxidative phosphorylation	44	9.77E-13
neurotransmitter secretion	27	4.87E-12
Lipoprotein	126	1.03E-10
Synaptosome	23	2.09E-09
growth cone	35	2.55E-09
Oxytocin signaling pathway	41	2.67E-09
Circadian entrainment	31	6.13E-09
Long-term potentiation	25	1.17E-08



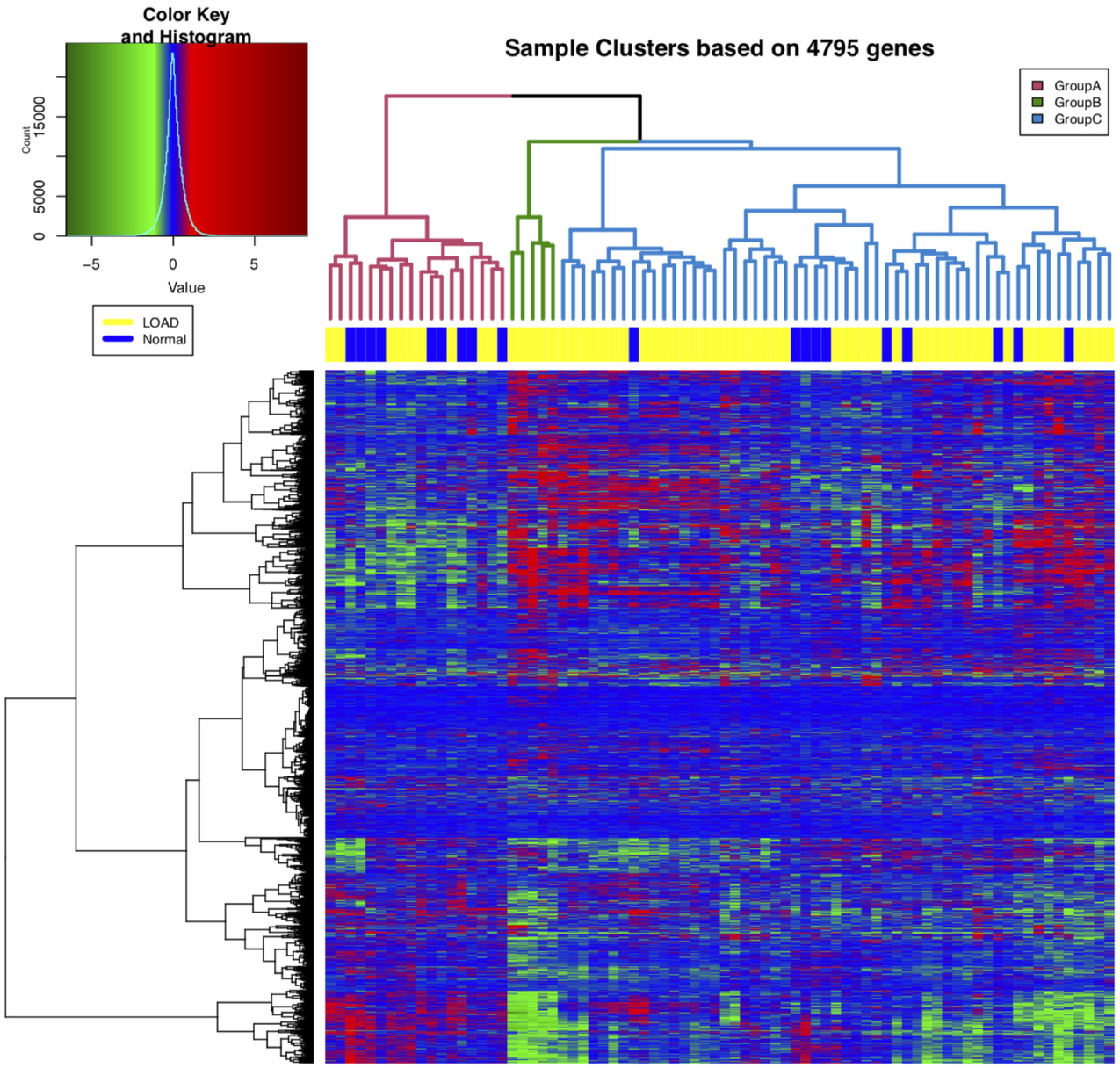
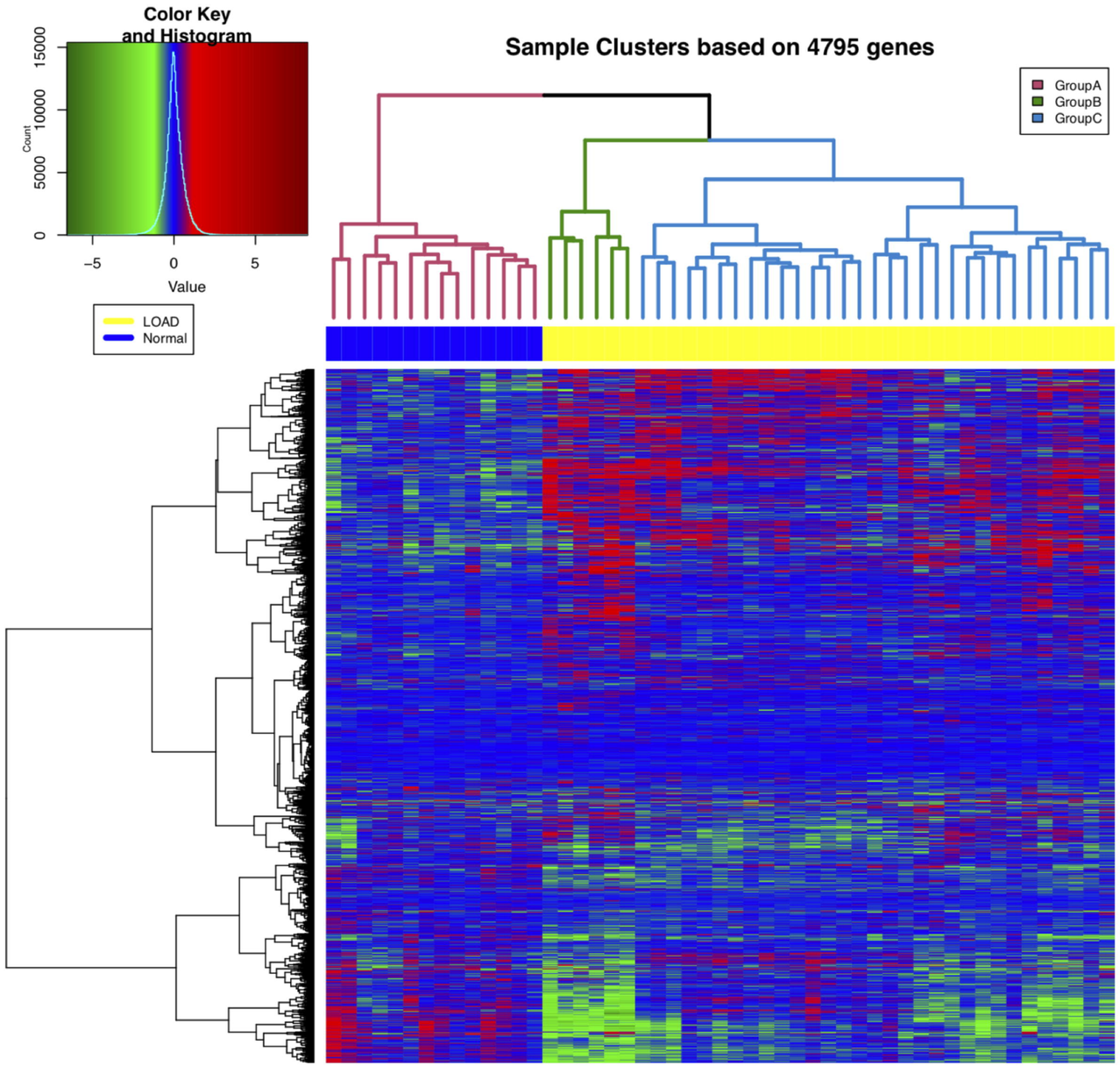
576 ADSGs ( )

Term	N	FDR_{BH}
Calcium	62	9.35E-09
Ciliopathy	19	1.74E-06
Glycoprotein	182	7.06E-06
Disease mutation	115	1.30E-05
plasma membrane	76	3.07E-05
Dynein	9	1.42E-04
Cytoskeleton	7	5.39E-04
Signal	157	0.00109
Ion transport	37	0.00215
axoneme	12	0.00261
Polymorphism	380	0.00319
Membrane	253	0.00319
Secreted	83	0.00379
Ion channel	24	0.00519
extracellular region	72	0.00808
Alternative splicing	331	0.03113

Conserved 484 ABSGs DN

Term	N	FDR_{BH}
cytosol	154	4.40E-13
Phosphoprotein	264	1.62E-08
Acetylation	137	1.62E-08
Cytoplasm	174	3.65E-08
Protein transport	42	1.39E-07
Alternative splicing	305	1.10E-05
Nucleotide-binding	76	2.98E-05
mitochondrion	63	2.82E-04
GTP binding	27	0.00209
protein binding	268	0.00441
Transferase	65	0.00467
Transport	72	0.00640
Proteasome	8	0.00678
Nucleus	158	0.00764
Ubl conjugation pathway	32	0.00922

A. PHG 3 subgroups before mixed sample removal



B. PHG 3 subgroups after mixed sample removal